High Resolution Analysis of Halftone Prints
- A Colorimetric and Multispectral Study

Daniel Nyström
High Resolution Analysis of Halftone Prints
- A Colorimetric and Multispectral Study

© Daniel Nyström 2008

Digital Media Division
Department of Science and Technology
Campus Norrköping, Linköping University
SE-601 74 Norrköping, Sweden

ISBN 978-91-7393-727-6  ISSN 0345-7524

Printed by LIU-Tryck, Linköping, Sweden, 2008
To Madelein
Abstract

To reproduce color images in print, the continuous tone image is first transformed into a binary halftone image, producing various colors by discrete dots with varying area coverage. In halftone prints on paper, physical and optical dot gains generally occur, making the print look darker than expected, and making the modeling of halftone color reproduction a challenge. Most available models are based on macroscopic color measurements, averaging the reflectance over an area that is large in relation to the halftone dots. The aim of this study is to go beyond the macroscopic approach, and study halftone color reproduction on a micro-scale level, using high resolution images of halftone prints. An experimental imaging system, combining the accuracy of color measurement instruments with a high spatial resolution, opens up new possibilities to study and analyze halftone color prints.

The experimental image acquisition offers a great flexibility in the image acquisition setup. Besides trichromatic RGB filters, the system is also equipped with a set of 7 narrowband filters, for multi-channel images. A thorough calibration and characterization of all the components in the imaging system is described. The spectral sensitivity of the CCD camera, which cannot be derived by direct measurements, is estimated using least squares regression. To reconstruct spectral reflectance and colorimetric values from the device response, two conceptually different approaches are used. In the model-based characterization, the physical model describing the image acquisition process is inverted, to reconstruct spectral reflectance from the recorded device response. In the empirical characterization, the characteristics of the individual components are ignored, and the functions are derived by relating the device response for a set of test colors to the corresponding colorimetric and spectral measurements, using linear and polynomial least squares regression techniques.

Micro-scale images, referring to images whose resolution is high in relation to the resolution of the halftone, allow for measurements of the individual halftone dots, as well as the paper between them. To capture the characteristics of large populations of halftone dots, reflectance histograms are computed as well as 3D histograms in CIEXYZ color space. The micro-scale measurements reveal that the reflectance for the halftone dots, as well as the paper between the dots, is not constant, but varies with the dot area coverage. By incorporating the varying micro-reflectance in an expanded Murray-Davies model, the nonlinearity caused by optical dot gain can be accounted for without applying the nonphysical exponentiation of the reflectance values, as in the commonly used Yule-Nielsen model.

Due to their different intrinsic nature, physical and optical dot gains need to be treated separately when modeling the outcome of halftone prints. However, in measurements of reflection colors, physical and optical dot gains always co-exist, making the separation a difficult task. Different methods to separate the physical and optical dot gain are evaluated, using spectral reflectance measurements, transmission scans and micro-scale images. Further, the relation between the physical dot gain and the halftone dot size is investigated, demonstrated with FM halftones of various print resolutions. The physical dot gain exhibits a clear correlation with the dot size and the dot gain increase is proportional to the increase in print resolution. The experimental observations are followed by discussions and a theoretical explanation.
Acknowledgements

During the years of work leading to this dissertation, I have been surrounded by a number of people who have contributed to the outcome, directly or indirectly, and should be acknowledged.

First, I would like to thank my supervisor Professor Björn Kruse, who has been the most influential person on the direction of my work. He introduced me to research within the fields of color science and graphic arts during my master thesis work; and has ever since supported me in my Ph.D studies with valuable ideas, encouragement and guidance.

Associate professor Li Yang, who has acted as my co-supervisor during the later half of my Ph.D. studies, is gratefully acknowledged for all his ideas, suggestions and valuable discussions. Somehow, he has always managed to find his time for discussions and last minute proof readings, even during summer holidays and after leaving academia for work in industry.

All my friends and colleagues in the Digital Media group are thanked for enjoyable coffee breaks, and for creating such a friendly and inspiring working atmosphere. This includes both the current members of the group, as well as former members, some of them now spread over the world. I would also like to thank Ivan Rankin for the fast and valuable linguistic reading.

The first part of this work, leading up to my Licentiate thesis, has been carried out within the Swedish national research program T2F. The later part has been funded by The Lundberg Foundation for Research and Education, and by Vinnova, through the research program PaperOpt. All financial support is gratefully acknowledged. Besides financing, being part of the research programs T2F and PaperOpt has also provided me with a network of contacts and colleagues, giving me valuable experiences and new acquaintances.

On a personal level, I would like to express my deepest gratitude to my friends and my family for all their support and encouragement. Especially to my mother, who has always encouraged me and believed in me. Finally, I would like to thank Madelein, the most important person in my life, for all her love and support, and for sharing and enriching my life during these years. From Skiathos to Långkärr.

Thank you.

Norrköping, December 2008

Daniel Nyström
List of publications


Contents

Abstract v
Acknowledgements vii
List of publications ix
Contents xi

1 Introduction 1
1.1 Introduction 3
1.2 Background 3
1.3 Aim of the study 4
1.4 Method 5
1.5 Structure of the dissertation 5

2 Color fundamentals 7
2.1 Introduction 9
2.2 Colorimetry 9
  2.2.1 Light, surfaces and observers 9
  2.2.2 CIE Standard observer 11
  2.2.3 Chromaticity diagram 13
  2.2.4 CIE Standard illuminants 14
  2.2.5 Color matching and metamerism 14
  2.2.6 CIELAB color space 15
  2.2.7 Color difference formulae 17
2.3 Color measurements 18
  2.3.1 Instruments 18
  2.3.2 Measurement geometry 19
  2.3.3 Precision and accuracy in color measurements 20
2.4 Color imaging 20
  2.4.1 Color image acquisition 20
2.4.2 Color reproduction 21
2.4.3 Color management 22

2.5 Multispectral imaging 23
2.5.1 Background 23
2.5.2 Terminology 23
2.5.3 The multispectral approach 23
2.5.4 Previous work 24

3 Device characterization 27
3.1 Introduction 29
3.2 Calibration and characterization 29
3.3 Characterization approaches 30
3.4 Input devices 31
3.4.1 Model-based input device characterization 31
3.4.2 Empirical input device characterization 32
3.5 Output devices 32
3.5.1 Model-based output device characterization 33
3.5.2 Empirical output device characterization 33
3.6 Least-squares regression techniques 34
3.6.1 Linear least squares regression 34
3.6.2 Polynomial regression 35
3.7 Metrics for evaluating device characterization 36
3.8 Color target design 36

4 Halftone color reproduction characterization 39
4.1 Introduction 41
4.2 Digital halftoning 42
4.2.1 Color halftoning 43
4.3 Dot gain 44
4.4 Modeling halftone color reproduction 45
4.4.1 The Murray-Davies model 45
4.4.2 The Yule-Nielsen model 47
4.4.3 The Neugebauer model 48
4.4.4 The Clapper-Yule model 50
4.4.5 Modeling the paper spread function 52
4.4.6 The Probability approach 52

5 Spectral sensitivity estimation 55
5.1 Introduction 57
5.2 The Image acquisition system 58
5.2.1 Image acquisition setup 58
5.2.2 Image acquisition model 58
5.3 Theory of spectral sensitivity estimation 60
5.3.1 The spectral sensitivity function 60
8 High resolution analysis of halftone prints 101
  8.1 Introduction 103
  8.2 Methodology 104
    8.2.1 Reflectance histograms 105
    8.2.2 Estimating the physical dot area coverage 107
    8.2.3 Color histograms 108
    8.2.4 Extension to color 109
  8.3 Experimental setup 110
    8.3.1 Printed samples 110
    8.3.2 Micro-scale images 111
  8.4 Experimental results 111
    8.4.1 Reflectance values 111
    8.4.2 Modeling $R_p(a)$ and $R_i(a)$ 113
    8.4.3 Tristimulus values 115
  8.5 Summary and discussion 117
9 Dot gain and screen resolution 121
  9.1 Introduction 123
  9.2 Methodology 124
    9.2.1 Physical dot gain from spectral reflectance 124
    9.2.2 Physical dot gain from transmission scans 125
    9.2.3 Physical dot gain from micro-scale images 126
  9.3 Experimental setup 126
  9.4 Experimental results 127
    9.4.1 Estimations of physical dot gain 127
    9.4.2 Spectral estimations 128
    9.4.3 Dot gain and print resolution 130
    9.4.4 Physical and optical dot gain 132
  9.5 Summary and discussion 134
10 Summary and future work 137
  10.1 Summary 139
  10.2 Future work 140
Bibliography 143
Appendices 153
  Appendix A: Calibration of the image acquisition system 155
    Spectral measurements 155
    Illumination 156
    Color filters 159
    Optics 162
    CCD camera 165
    Summary and discussion 168
Appendix B: Empirical characterization data

- The polynomial approximation functions 169
- Experimental results 171
- The size of the training set 173
- Cross-media characterization 174

Appendix C: High resolution analysis data

- Experimental setup 177
- Experimental results 178
Chapter 1

Introduction

1.1 Introduction
1.2 Background
1.3 Aim of the study
1.4 Method
1.5 Structure of the dissertation
1.1 Introduction

This aim of this dissertation is high resolution analysis of halftone prints. This involves characterization of an experimental imaging system to allow for colorimetric and multispectral image acquisition. Combining the accuracy of color measurement instruments with a high spatial resolution opens up new possibilities to study and analyze halftone color prints, allowing measurements of the individual halftone dots as well as the paper between them. The high resolution analysis reveals properties of halftone color prints which can not be derived by conventional macroscopic color measurements.

1.2 Background

Halftone color reproduction, i.e. producing various colors by discreet dots with varying area coverage, is by far the most common way of reproducing color images in print. In halftone prints on paper, physical and optical dot gains generally occur, causing a tone value increase, and making the modeling of halftone color reproduction a challenge. Physical dot gain is closely related to the printing process, and refers to the fact that the size of the printed halftone dots differs from their nominal size. Optical dot gain originates from light scattering inside the substrate, causing light exchange between different chromatic areas, and making the dot appear bigger than its physical size when it is perceived or measured. Due to their different intrinsic nature, physical and optical dot gains need to be treated separately in order to accurately model the outcome of halftone prints. However, in reflection measurements of halftone prints, the physical and optical dot gains always co-exist, making the separation of one type from another a difficult task.

Many models predicting the color output from halftone prints have been proposed, starting from the 1930s. The task is difficult, mainly because of the co-existence of physical and optical dot gain. Most available models are based on macroscopic color measurements. Generally, the spectral reflectance or the colorimetric values CIEXYZ or CIELAB are measured by averaging the reflectance over an area that is large in relation to the size of the halftone dots. Such macroscopic measurements correspond well to the color appearance at normal viewing distance, but do not reveal the underlying microstructure of the printed halftone. The practical motivation for a better...
understanding and characterization of physical and optical dot gain is to be able to predict and account for the effects with the aim of improving print quality.

Microscopic images of the prints allow the study of the microstructure of the printed halftones in detail. Color images are typically acquired in a device-dependent format, specific for the imaging device. In analysis of halftone color prints, there is a need for accurate mappings to device-independent color representations, preferably the colorimetric representations CIEXYZ and CIELAB. The functions describing such mappings are derived through the process of device characterization. However, since the sensitivities of typical imaging devices are different from the CIE color matching functions, this relationship is usually not trivial.

Even when accurate transformations to device-independent colorimetric representations can be derived, colorimetric imaging still suffers from some limitations. Colorimetric imaging is by its nature always metameric, i.e. based on metameric matching rather than spectral matching. When the sensitivities of the imaging device differ from the CIE color matching functions, two images that are metamerically identical for an imaging device may appear different for human observers. Hence, the ideal would be to represent color by its spectral power distributions, using multispectral imaging. Multispectral imaging allows for the separation of the spectral properties of the object from the illumination, thus representing the color by its spectral reflectance. As the physical representation of color, spectral reflectance is independent of the characteristics of the image acquisition system, and the multispectral images can be transformed to any color space and be rendered under any illumination.

1.3 Aim of the study

The aim of this study is high resolution analysis of halftone prints, i.e. to go beyond the macroscopic approach and study halftone color prints on a micro-scale level, using high resolution images. This approach requires high quality acquisition of colorimetric and multispectral images. Accurate methods for computing colorimetric and spectral data from the recorded device-dependent signals, to obtain colorimetric and multispectral images are thus needed.

Multispectral image acquisition, i.e. recovering spectral properties of the sample, requires the computation of spectral reflectance data from a relatively small number of channels. This work will try to answer what colorimetric and spectral accuracy can be achieved, by combining knowledge of all parts of the system, a thorough calibration, and employing different methods for device characterization. Is the conventional trichromatic principle of image acquisition sufficient, or is multi-channel imaging required, to achieve satisfactory colorimetric and spectral accuracy? Can the spatial resolution of digital images be combined with the spectral resolution of color measurement instruments to allow for accurate colorimetric and spectral measurements in each pixel of the image?

Furthermore the aim is to use such images for high resolution analysis of halftone prints. Can microscopic measurements reveal characteristics and properties of halftone prints not apparent in macroscopic color measurements that can be used to gain a better understanding of halftone color reproduction?
1.4 Method

The image acquisition system used is an experimental system with great flexibility for
the user providing numerous ways to control and alter the image acquisition setup.
Besides the conventional trichromatic RGB filters, the system also provides the
possibility of acquiring multi-channel images using 7 narrowband filters.

To ensure stability and repeatability, a thorough calibration of the image
acquisition system is carried out. All components involved in the image acquisition
system are calibrated with respect to repeatability, spatial uniformity and temporal
stability. The spectral sensitivity of the camera, which can not be derived by direct
measurements, is estimated by relating the camera response to the spectral reflectance
for a set of carefully selected color samples.

To derive mappings to colorimetric and spectral representations, two conceptually
different approaches are used: model-based and empirical characterization. In model-
based characterization, the physical model describing the image acquisition process is
inverted to reconstruct spectral reflectance from the recorded camera response. A priori
knowledge on the smooth nature of spectral reflectances is exploited, by representing
the spectra as linear combinations of different basis functions. Empirical device
characterization is a “black box” approach ignoring the characteristics of the system.
The characterization functions are derived by relating the device response for a set of
test colors to colorimetric and spectral measurements using linear and polynomial least
squares regression.

After the device calibration and characterization, the experimental image
acquisition system is used to acquire micro-scale images of halftone color prints. Micro-
scale images, referring to images whose resolution is high in relation to the resolution of
the halftone, allow for measurements of the individual halftone dots as well as the paper
between them. The primary focus is on studies on how the micro-reflectance of the
halftone dots and the paper between them (which in most models is assumed to be
constant) varies with the dot area fraction. To capture the characteristics of large
populations of halftone dots, reflectance histograms are computed as well as 3D
histograms in CIEXYZ color space.

To estimate the physical dot area, i.e. separating the physical and optical dot
gains, three different methods are described and evaluated. Physical dot gain is
estimated from spectral reflectance measurements, from transmission scans using a flat-
bed scanner, and from microscopic images of the prints, respectively.

1.5 Structure of the dissertation

This dissertation is written as a monograph based on contents which in most cases, but
not all, have been previously published in the papers listed on page ix. Rewriting the
content into a monograph has given me the opportunity to present the ideas and the
work without the limitations imposed on the previous publications regarding templates
and number of pages. I hope, and believe, that the structure of the dissertation has
benefited from this format compared to presenting a number of separate papers, some
with considerable overlap.
The first part of the dissertation, Chapters 2, 3 and 4, provides a brief theoretical background of concepts and methods used in the dissertation. Chapter 2 presents an overview of color science, including brief introductions to colorimetry, color measurements, and color imaging. The concept of multispectral imaging is also introduced. Chapter 3 focuses on device characterization, describing different approaches and explaining definitions and terminology associated with the topic. It also provides a description of least squares regression techniques that are used in the following chapters. Chapter 4 provides a brief introduction to halftone color reproduction, introducing the concepts of digital halftoning and dot gain, as well as an overview of models predicting the outcome of halftone prints.

The second part of the dissertation, Chapters 5 to 7, focuses on colorimetric and multispectral image acquisition, describing the calibration and characterization of an experimental imaging system. Chapter 5 introduces the image acquisition system and explains the estimation of the spectral sensitivity function for the camera. By relating the camera response to the spectral reflectance for a set of carefully selected color samples the camera sensitivity function is estimated using least-squares regression techniques. The model-based approach to reconstruct spectral reflectance, i.e. by inverting the spectral image acquisition model, is described in Chapter 6. Multispectral images are computed from trichromatic and multi-channel images, respectively. The chapter also includes discussions on appropriate metrics for evaluating the reconstructed spectra. The results for the empirical approach, using regression-based characterization, for colorimetric and spectral reconstructions, are given in Chapter 7. The different performances in reconstructing spectral and colorimetric data from trichromatic and multi-channel images are examined.

The third part, including Chapters 8 and 9, focuses on studies of halftone color prints. In Chapter 8, the image acquisition system is used to acquire micro-scale images used for high resolution analysis of halftone prints. Micro-scale images allow for measurements of the individual halftone dots and analysis on how the reflectance for the halftone dots as well as the paper between them, varies with the dot area fraction. Chapter 9 investigates different methods for estimating the physical dot gain in color halftones, a task that requires the separation of the physical and optical dot gain. Further, the relation between the physical dot gain and the halftone dot size is investigated and demonstrated with FM halftones of various print resolutions.

Finally, Chapter 10 provides a short summary of the work and results, as well as introducing some ideas on directions for future work.

Results that are not essential for the discussion have been placed in appendices. The technical specifications are described in Appendix A. The calibration of the image acquisition system is described in Appendix A. The technical specifications are given for all the components in the system, along with measurement results and discussions on the demands on each component for high quality image acquisition. Appendix B contains additional data from the regression-based characterization in Chapter 7, including experiments on the influence of the size of the training set and the performance of the derived functions when applied to color samples of different media and colorants. Appendix C collects some additional results from a preliminary study on high resolution analysis of halftone prints, based on the ideas presented in Chapter 8.
Chapter 2

Color fundamentals

2.1 Introduction
2.2 Colorimetry
2.3 Color measurements
2.4 Color imaging
2.5 Multispectral imaging
2.1 Introduction

The phenomenon of color is a complex visual sensation, involving physical properties of light, but also physiological and psychological properties of the human observer. This chapter provides a brief description of the basics of color science. There is no ambition to give a complete review of this complex topic, merely describing the very basics, which in most cases involves simplifications. The aim is to provide the necessary background, along with definitions and terminologies, for the concepts used throughout the dissertation.


2.2 Colorimetry

Colorimetry is the science of measuring, representing, and computing color in a way that takes into account the interaction between the physical aspects of color and the physiological aspects of human vision. The basis of colorimetry is a set of standards, defined by Commission Internationale de l’Eclairage (CIE), the primary organization for the standardization of color metrics and terminology.

2.2.1 Light, surfaces and observers

The basic, physical stimulus of color is electromagnetic radiation in the visible band of the spectrum, usually referred to as light. The visible band of the spectrum is typically defined by the wavelengths between approximately 380 and 780 nm (Hunt, 1998). Below the visible band lies the ultraviolet region of radiation and above the visible band the infrared region. The properties of light are physically characterized by their spectral power distribution (SPD), i.e. the distribution of power as a function of wavelength.

The color of an object depends on its spectral reflectance properties, i.e. the amount of the incident light that is reflected by the illuminated object at different
wavelengths. If we represent the spectral radiance from an illuminant as \( I(\lambda) \) and the spectral reflectance for an object as \( R(\lambda) \), then the radiance reflected by the object, \( E(\lambda) \), is given by:

\[
E(\lambda) = I(\lambda)R(\lambda)
\]

(2.1)

This spectral interaction between light and surfaces defines the basis for all representations of color. However, even though the spectral power distribution, \( E(\lambda) \), characterizes the color properties of the light source and the object, the light on its own has no color unless it is observed by a human observer, converting the spectral properties of the light into a color sensation, see Fig. 2.1.

When light reaches the human eye it is detected by two different types of light sensitive cells, responsible for the human vision: rods and cones. The information is further processed by the neural system and the brain into a visual color sensation. The rods are essentially monochromatic and responsible for night (scotopic) vision, and do not contribute to color vision. The sensation of color comes from the three different types of cones, usually denoted \( L \), \( M \) and \( S \) cones, providing photopic vision under normal levels of light. The three types of cones are sensitive to light of long, medium and short wavelengths respectively. The stimulus from the incoming radiation for each type of cone is given by:

\[
\begin{align*}
L_{tot} &= \int \lambda E(\lambda)L(\lambda)d\lambda \\
M_{tot} &= \int \lambda E(\lambda)M(\lambda)d\lambda \\
S_{tot} &= \int \lambda E(\lambda)S(\lambda)d\lambda
\end{align*}
\]

(2.2)

where \( L(\lambda) \), \( M(\lambda) \), and \( S(\lambda) \) are the spectral sensitivity functions for the cones, and \( E(\lambda) \) the SPD of the light reaching the eye. The resulting stimuli of the cones, \( L_{tot}, M_{tot} \) and
The sensory expression of color is thus dependent on the interaction of three different elements: a light source, an object and an observer. This involves both physical aspects of color, such as the spectral interaction between the light and the object, and physiological aspects of human vision. The interaction between these two aspects, the psychophysical aspect, dealing with the relation between physical attributes and the resulting sensations, is defined by colorimetry (Hardeberg, 2001).

Note that the given model is very limited and contains simplifications in several respects. The interaction between light and object is in reality far more complicated than just surface reflection, and may also involve for example refraction, absorption and scattering in the interior of the object. Furthermore, the geometrical effects such as directional specular reflections are not mentioned, nor are effects such as fluorescence or polarization. Moreover, the human visual system is more complicated than implied, and the perceived color will also be affected by the surroundings and the state of chromatic adaptation of the observer (Fairchild, 1998). However, with these limitations in mind, the model serves as a basis for the upcoming discussions and definitions.

### 2.2.2 CIE Standard observer

The exact forms of the spectral sensitivity functions for the cones used in Eq. 2.2 are difficult to measure directly, and may vary between individuals. To have an agreement between different measurements it is desirable to define a standard set of color matching functions, (CMFs) representing the characteristics of the average human response to light spectra, thus representing a “standard observer”.

In 1931 CIE defined the CIE 1931 XYZ color-matching functions, defining the color matching properties of the CIE 1931 standard colorimetric observer. The X, Y and Z tristimulus values, forming the basis for all colorimetry, are given by:

\[
\begin{align*}
X &= k \int \lambda I(\lambda) R(\lambda) \tilde{x}(\lambda) d\lambda \\
Y &= k \int \lambda I(\lambda) R(\lambda) \tilde{y}(\lambda) d\lambda \\
Z &= k \int \lambda I(\lambda) R(\lambda) \tilde{z}(\lambda) d\lambda
\end{align*}
\]

where \(\tilde{x}(\lambda), \tilde{y}(\lambda)\) and \(\tilde{z}(\lambda)\) are the CIEXYZ color-matching functions, see Fig 2.2. In absolute colorimetry, the normalization factor \(k\) is set to a constant, expressed in terms of the maximum efficacy of radiant power, equal to 683 lumens/W (Sharma, 2003). In relative colorimetry, the normalization factor, \(k\), is chosen such that \(Y = 100\) for a chosen reference white, usually a perfect diffuse reflector, with spectral reflectance equal to unity for all wavelengths, i.e.:
Figure 2.2. The CIE 1931 color matching functions for the 2° standard colorimetric observer.

Note that the XYZ color matching functions do not correspond to a set of physical primaries, but to linear transformations of the physical primaries, chosen to eliminate the negativity of the physical primaries, and normalized to yield equal tristimulus values for the equi-energy spectrum. Furthermore, \( \pi(\lambda) \) is chosen to coincide with the luminous efficiency function for photopic vision, i.e. the tristimulus value \( Y \) represents the perceived luminance.

The CIE 1931 standard colorimetric observer is sometimes referred to as the 2° observer, since the color-matching functions are based on a visual field of 2°. Later, the CIE 1964 Supplementary Standard Colorimetric Observer was defined, using a visual field of 10°. All colorimetric computations in this dissertation are based on the CIE 1931 XYZ color-matching functions, representing the 2° observer.

Practically, measurements of spectral power distributions and spectral reflectance will be sampled, using some wavelength interval, and the integrals of Eq. 2.3 will be replaced by summations. If we use vector notation and represent the spectral signal as the discrete \( N \)-component vector \( f \), sampled at wavelengths \( \lambda_1, \ldots, \lambda_N \), Eq. 2.3 can be rewritten as:

\[
e = A^c f
\]  

(2.5)

where \( e \) is the colorimetric 3-component vector of the resulting tristimulus response, XYZ, and \( A_c \) is the \( N \times 3 \) matrix with columns representing the color-matching functions, \( x(\lambda) \), \( y(\lambda) \) and \( z(\lambda) \).
2.2.3 Chromaticity diagram

To graphically visualize color, the *CIE x,y chromaticity diagram* is often used, providing a two-dimensional representation. Chromaticity diagrams are based on the relative magnitudes of the tristimulus values, called *chromaticity coordinates*, as:

\[
x = \frac{X}{X + Y + Z} \\
y = \frac{Y}{X + Y + Z} \\
z = \frac{Z}{X + Y + Z}
\]  

(2.6)

It is clear that \( x + y + z = 1 \) and hence that the chromaticity can be represented using only two variables, usually \( x \) and \( y \). The two variables \( x \) and \( y \) form a two-dimensional chromaticity diagram, representing a projection of the three-dimensional XYZ color space onto a plane. The chromaticity diagram provides a sort of color map on which the chromaticities of all colors can be plotted, see Fig 2.3. The curved line, representing the chromaticities of monochromatic light, is called the *spectral locus* and is a continuous convex hull enclosing the domain of all colors. The line that connects the ends of the spectral locus is called the *purple boundary*. Note that chromaticity diagrams show only projections of tristimulus values, hence bright and dim colors can be projected onto exactly the same point.

![CIE x,y chromaticity diagram](image-url)

*Figure 2.3. The CIE x,y chromaticity diagram.*
2.2.4 CIE Standard illuminants

Because the appearance of color strongly depends on the illumination, there is a need for accurate definitions of the illuminants involved. To fulfill this, CIE has introduced a number of standard illuminants, defined in terms of spectral power distributions.

In 1931 CIE defined the standard illuminants \( A \) (representing a tungsten filament lamp), \( B \) (representing sunlight) and \( C \) (representing average daylight). The standard illuminants \( B \) and \( C \), representing daylight, had too little power in the UV region, and with the increasing use of fluorescing agents there was a need for standard illuminants mimicking the daylight in the UV region. In 1963, CIE recommended the new standard illuminants \( D_{50} \) and \( D_{65} \) to represent average daylight of different color temperature in the visible and in the UV spectra down to 300 nm. These standard illuminants are commonly used in color systems and standards, e.g. in television, where \( D_{65} \) is the reference white for the PAL system. \( D_{65} \) is commonly used to represent daylight within the paper industry, while the more yellowish \( D_{50} \) has become a standard in the graphic arts industry. The spectral power distributions of the CIE standard illuminants \( A, D_{50} \) and \( D_{65} \) are displayed in Fig. 2.4. (Hunt, 1998)

![Figure 2.4. The spectral power distributions for the CIE standard illuminants A, D50 and D65.](image)

2.2.5 Color matching and metamerism

Two spectra, represented by the \( N \)-components vectors \( \mathbf{f} \) and \( \mathbf{g} \), chromatically match each other when they produce identical tristimulus values, i.e.

\[
A_i' \mathbf{f} = A_i' \mathbf{g}
\]  

(2.7)

Because \( A_i \) is an \( N \times 3 \) matrix with \( N > 3 \), the equation system may have multiple solutions, implying that different spectra may produce the same color stimulation. This
The phenomenon is called *metamerism* which means that two different spectra result in the same tristimulus values, i.e. appearing to have the same color, under a given illumination. The pair of distinct spectra producing the same tristimulus values are called *metamers* and the match is referred to as a *metameric match*, as opposed to a spectral match.

One effect of metamerism is that two colors that match under a given illuminant may differ when they are viewed under another illumination. This sometimes causes practical problems, for example when clothes that match perfectly in a dressing room may appear completely different in outdoor environments. Besides being problematic sometimes, metamerism is the basis for conventional color reproduction, using three primary colors to achieve a colorimetric match to the target color rather than a spectral match (Sec. 2.4).

To describe the various types of metamerism, CIE has recommended the use of metamerism indices. The *Illuminant Metamerism Index* considers the color difference between a metameric pair, caused by substituting a reference illuminant (preferably D65) by a test illuminant. The *Observer Metamerism Index* measures the color difference between a metameric pair caused by substituting the reference observer (either the 2° observer or the 10° observer) by a *Standard Deviate Observer* (SDO), having different spectral sensitivities. (Hunt, 1998)

### 2.2.6 CIELAB color space

The CIEXYZ tristimulus values provide the basis for Colorimetry, representing colors in the three-dimensional XYZ color space. It is natural and practically useful to associate the differences in the XYZ color space to the perceived difference. Unfortunately, the visual system is complex and the eyes’ sensitivity to light is nonlinear, in contrast to the tristimulus values that are linearly related to the spectral power of the light (Eq. 2.3). Therefore, the XYZ color space is perceptually non-uniform, i.e. the Euclidian differences in XYZ color space between colors do not correspond to the perceived color differences.

To linearly map the Euclidian distance in a color space into the perceptual color difference, a perceptually *uniform color space* is needed. In 1976 CIE proposed the *CIE 1976 (L*a*b*) color space* (CIELAB), which is approximately uniform. The CIELAB coordinates are computed using non-linear transformations from the tristimulus XYZ values, as:

\[
L^* = \begin{cases} 
116 \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 & \text{if } \frac{Y}{Y_n} > 0.008856 \\
903.3 \left( \frac{Y}{Y_n} \right) & \text{if } \frac{Y}{Y_n} \leq 0.008856
\end{cases}
\]

\[
a^* = 500 \left[ f \left( \frac{X}{X_n} \right) - f \left( \frac{Y}{Y_n} \right) \right]
\]

\[
b^* = 200 \left[ f \left( \frac{Z}{Z_n} \right) - \frac{Y}{Y_n} \right]
\]
\[ b^* = 200 \left[ f \left( \frac{Y}{Y_n} \right) - f \left( \frac{Z}{Z_n} \right) \right] \]  

(2.10)

where

\[
f(x) = \begin{cases} 
  x^3 & x > 0.008856 \\
  7.787x + \frac{16}{116} & x \leq 0.008856 
\end{cases}
\]

(2.11)

The values \(X_n, Y_n, Z_n\) refers to the CIEXYZ tristimulus values of a reference white, usually represented by one of the CIE standard illuminants. The use of the reference white is an attempt to account for the adaptive characteristics of the visual system. The purpose of using a linear model at lower light levels is because the cones become less sensitive while the rods become active at low levels of light (Trussel, et al., 2005).

The CIELAB color space is defined by the three coordinates, \(L^*\), representing lightness, \(a^*\), representing the red-green axis, and \(b^*\), representing the yellow-blue axis, see Fig. 2.5. The scale of \(L^*\) is 0 to 100, with 0 representing black and 100 the reference white. A variant of the CIELAB representation is given by the cylindrical coordinates, defined by the CIE 1976 chroma,

\[ C_{ab}^* = \sqrt{(a^*)^2 + (b^*)^2} \]

(2.12)

and the CIE 1976 hue-angle,

\[ h_{ab}^* = \arctan \left( \frac{b^*}{a^*} \right) \]

(2.13)

Figure 2.5. Interpretation of the \(L^*, a^*\) and \(b^*\) axes in CIELAB color space.
It is important to point out that the CIELAB color space is only approximately uniform, sometimes referred to as being a pseudo-uniform color space. There are still significant differences in the correspondence to the perceived color differences in the different parts of the CIELAB space, with the blue region being especially problematic (Sharma, 2003). Since no ideal uniform color space is available, the CIELAB color space probably represents the most important colorimetric system at present (Kipphan, 2001).

### 2.2.7 Color difference formulae

When comparing colors it is desirable to define a measure for the perceived difference in color appearance. A color difference formula is designed to give a quantitative measure of the perceived color difference between a pair of color samples under a given set of conditions. One simple and commonly used color difference formula is the CIE 1976 \( L^*a^*b^* \) color difference, \( \Delta E_{ab}^* \), corresponding to the Euclidian distance in CIELAB color space, i.e.

\[
\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}
\]  

(2.14)

where \( \Delta L^* \), \( \Delta a^* \) and \( \Delta b^* \) are the differences in \( L^* \), \( a^* \) and \( b^* \), respectively, between the pair of samples. An alternative formulation of the CIE 1976 color difference is expressed in terms of the lightness difference, \( \Delta L^* \), chroma difference, \( \Delta C_{ab}^* \), and hue difference, \( \Delta H_{ab}^* \), as:

\[
\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta C_{ab}^*)^2 + (\Delta H_{ab}^*)^2}
\]

(2.15)

Note that the hue difference is not defined as the difference in hue angle, \( h_{ab}^* \), but as:

\[
\Delta H_{ab}^* = \sqrt{(\Delta E_{ab}^*)^2 - (\Delta L^*)^2 - (\Delta C_{ab}^*)^2}
\]

(2.16)

The perceptual interpretation of the color difference \( \Delta E_{ab}^* \) is not straightforward and there are a number of different interpretations proposed. The just noticeable difference (JND) is about 1 \( \Delta E_{ab}^* \) (Hunt, 1995) but varies for different parts of CIELAB space. A “rule of thumb” for the interpretation of the \( \Delta E_{ab}^* \) color difference, in graphic arts applications, is given in Table 2.1 (Hardeberg, 2001):

<table>
<thead>
<tr>
<th>( \Delta E_{ab}^* )</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3</td>
<td>Hardly perceptible</td>
</tr>
<tr>
<td>3 - 6</td>
<td>Perceptible but acceptable</td>
</tr>
<tr>
<td>&gt; 6</td>
<td>Not acceptable</td>
</tr>
</tbody>
</table>

*Table 2.1. Perceptual impact of \( \Delta E_{ab}^* \) color difference between two color samples, in side by side comparison.*
In a perceptually uniform space, the Euclidian distance would provide a good metric of the perceived color difference. However, the non-uniformities of CIELAB result in variations in the perceptual correspondence to $\Delta E_{ab}$ in different parts of the color space. In 1994, CIE proposed a revised color difference formula, the CIE 1994 color difference, which incorporates corrections for the non-uniformity of CIELAB (CIE, 1995). The CIE94 color difference, $\Delta E_{94}$, is given by:

$$\Delta E_{94} = \sqrt{\left( \frac{\Delta L^*}{k_L S_L} \right)^2 + \left( \frac{\Delta C^*_{ab}}{k_C S_C} \right)^2 + \left( \frac{\Delta H^*_{ab}}{k_H S_H} \right)^2}$$

(2.17)

where the weighting functions $S_L$, $S_C$ and $S_H$ vary with the chroma of the reference sample, as:

$$S_L = 1, \quad S_C = 1 + 0.045C^*, \quad S_H = 1 + 0.015C^*$$

(2.18)

The parametric factors $k_L$, $k_C$, $k_H$ are included to account for the influence on viewing and illumination conditions. Under reference conditions they are set to:

$$k_L = k_C = k_H = 1$$

(2.19)

A given color difference, represented by equally sized spheres of $\Delta E_{ab}$, is in $\Delta E_{94}$ represented by elliptical volumes, with the size and shape varying throughout the color space. For neutral colors $\Delta E_{94}$ equals $\Delta E_{ab}$, while $\Delta E_{94}$ becomes smaller for more saturated colors.

Later attempts to improve the uniformity of color difference formula include the CIEDE2000 (Lou, et al., 2000; CIE, 2001). Beside the chroma and hue weighting functions used in CIE94, CIEDE2000 include a number of additional parameters to further compensate for the non-uniformity of CIELAB. However, the improvements achieved by incorporating the more advanced corrections in the formula are found to be small compared to the improvements of the CIE94 formula (Melgosă, et al., 2004). A later study even claims that the current CIEDE2000 formula has systematic errors and is not an improvement to the CIE94 color difference (Granger, 2008).

### 2.3 Color measurements

#### 2.3.1 Instruments

Color measurement instruments, fall into two general categories: broadband and narrowband instruments. Broadband instruments use broadband filters to filter the incoming radiation and deliver up to three color signals. Photometers measure luminance only, densitometers give the optical density for red, green and blue. Colorimeters record CIE tristimulus values directly, by using photocells combined with color filters designed to match the CIE color matching functions. They are fast and relatively inexpensive, but their accuracy is limited because it is difficult to design
filters that exactly match the color matching functions. Accurate colorimetric measurements require computation from spectral power data, delivered by narrow band instruments.

Measuring spectral data involves *spectroradiometry* or *spectrophotometry*. In spectroradiometry, the incoming radiation is measured in narrow bands of wavelengths throughout the spectrum, using *spectroradiometers*. In spectrophotometry, the amount of reflected light from an object is compared to the incident light, thus delivering a measure of the spectral reflectance for the sample.

Both spectroradiometers and spectrophotometers require means of dispersing the light into a spectrum such that light at different wavelengths can be measured. Usually the light is dispersed using gratings, but prisms and narrowband interference filters can also be used. The dispersed radiation is then detected by photoelectric cells. In the case of spectrophotometers, a light source is also required, most commonly tungsten-halogen lamps or xenon flash lamps. When the samples contain fluorescent whitening agents, FWA, the UV content of the light source will affect the measurements. Studies have shown that even small changes in the UV content have effect on the measured reflectance spectra for papers including FWA (Andersson & Norberg, 2007).

For most purposes, it is considered sufficient to sample the spectrum at 5 nm intervals, but in some cases 10 nm or 20 nm intervals are also appropriate. The CIE color matching functions are tabulated in the range 360 to 830 nm, but for most colorimetric purposes it is considered sufficient to use the range 380 to 780 nm. Some instruments use a smaller range of wavelengths, commonly 400 to 700 nm. (Hunt, 1998)

### 2.3.2 Measurement geometry

An important consideration in color measurements is the geometry of viewing and illumination. CIE has recommended 6 different geometries for colorimetric measurements of reflective samples and another 6 for transmission measurements (Hunt, 1998).

A common arrangement for reflective measurements within the graphic arts industry is the 45°/0° geometry, denoting an illumination with 45° angle of incidence and with the detector normal to the surface. The geometry is intended to reduce the effect of specular reflection and to represent typical viewing conditions. The disadvantage is that the result is dependent on the structure of the surface topology because of the directed illumination.

The d/0° geometry, denoting diffuse illumination and measurement from surface normal, is commonly used for color measurements within the paper industry. The diffuse illumination is provided by an integrating sphere whose inside is coated with a highly reflective material, usually barium sulfate. The sample is placed against an opening in the sphere, and the illumination is arranged so that neither the sample nor the detector is directly illuminated, i.e. so that only diffuse illumination strikes the sample, and so that no light from the illuminant directly reaches the detector. (Pauler, 1998)
2.3.3 Precision and accuracy in color measurements

By precision is meant the consistency with which measurements can be made, i.e. the ability to deliver stable and repeatable results. Precision is affected by random errors and the most common sources are variation in sensitivity, electronic noise and sample preparation.

By accuracy is meant the degree to which measurements agree with those made by a standard instrument or procedure in which all possible errors are minimized. Accuracy is affected by systematic errors and common sources in modern instruments are wavelength calibration, detector linearity, measurement geometry and polarization.

The importance of precision and accuracy depends on the application. For example, when the same instrument is used to monitor the consistency of a product, good precision is vital but great accuracy is not. When colorimetric results from different instruments are to be compared, high accuracy is crucial. Furthermore, for any comparison to be meaningful, it is essential that the illuminant, standard observer, and the measurement geometry must all be the same (Hunt, 1998).

2.4 Color imaging

In the real world, colors exist as spatial variations of spectral distributions of radiance and reflectance. To capture these scenes digitally using a color recording device, the images must be sampled both spatially and spectrally. The captured color images are reproduced from recorded data, typically by using additive or subtractive color mixing of a set of primary colors.

2.4.1 Color image acquisition

Digital color recording devices consist mainly of digital cameras or color scanners, operating on similar principles. The color information is recorded by optical-electronic sensors that spatially sample the image. Light from the image passes through a number of color filters of different spectral transmittance before it reaches the sensors. The transmission filters typically consist of a set of red, green and blue filters, producing RGB-images.

The sensors in digital cameras are typically arranged as two-dimensional arrays, allowing for the image to be captured in a single exposure. There exist different schemes for separating the RGB color-channels. The most common scheme is color filter arrays (CFAs), where each cell of the sensor is covered with a transmissive filter of one of the primary colors. The most commonly used mosaic pattern for the filters is the Bayer pattern, with 50% green cells in a checker board arrangement, and alternating lines of red and blue cells. Other methods for color separation include color sequential, where the image is composed of sequential exposures while switching the filters, and multi-sensor color, where the light is separated into red, green and blue colors using a beam splitter and detected by three separate monochrome sensors (Paraluskii & Spaulding, 2003).

Scanners are typically designed to scan images on paper or film using its inbuilt light source. There is no need to capture the stationary object in a single exposure and
typically linear sensor arrays are used to scan the image, moving along the direction perpendicular to the sensor array. Usually three sensor arrays are used, corresponding to the three color channels, R, G and B, but there are also arrangements using three different lamps, obtaining the color image from three successive measurements with a single array. To improve accuracy and reduce color reproduction errors due to metamerism, six color scanners are being developed, using two successive scans with different light sources (Hunter, et al., 2008).

2.4.2 Color reproduction

Generally speaking, all the color reproduction techniques can be classified into two groups: additive and subtractive. In additive color reproduction, color is produced on a dark background as an additive mixture of light of different wavelengths, known as primary colors. Typically, the additive primaries are red, green and blue (RGB). The principle of additive color mixture is illustrated in Fig. 2.6(a), where mixing red with green light produces yellow, red and blue produces magenta, blue and green forms cyan and the mixture of all three primaries gives white. Additive color reproduction is typically used for emissive displays, such as CRT and LCD displays.

Subtractive color reproduction, typically used for transparent or reflective media, produces colors by blocking/removing spectral components from “white” light through light absorption. The most common subtractive primaries are cyan, magenta and yellow (CMY), colorants that absorb light in the red, green and blue spectral bands of the spectrum, respectively. The principle is illustrated in Fig. 2.6(b), where the overlay of cyan and magenta producing blue, cyan and yellow produces green, magenta and yellow produces red and the overlay of all three colorants results in black. In color printing, it is common to add a fourth black (K) colorant to improve the reproduction of gray tones and allowing for darker colors and a higher dynamic range. Halftone color printing, which will be further described in Chapter 4, can be seen as a hybrid system, because the colorants are subtractively mixed, but the perceived color is the average of the different colored regions over a small area. (Sharma, 2003)

![Figure 2.6. The principle of additive (a) and subtractive (b) color mixing.](image)
2.4.3 Color management

The principles of color image acquisition and reproduction described in the previous sections rely on device-dependent color representations, specific for each device. For example, the RGB sensitivities of a digital camera or a flatbed scanner are generally different to the primaries of a CRT or a LCD display. In other words, a color image will look differently when displayed on different devices. To achieve consistent color representations with different devices, it is necessary to map the device-dependent color representations into a device-independent space, which is the key of color management.

In digital color management, the device-independent color space is called the profile connection space (PCS). CIEXYZ and CIELAB are the commonly used profile connection spaces. The transformations between device-dependent data and the PCS are described by device profiles, for input and output devices. The device profiles defining the relationship between device-dependent and device-independent color spaces are derived by device characterization, as described in Chapter 3. A widely adopted standard for storing device profiles is the International Color Consortium (ICC) profile (ICC, 2004). A color management module (CMM) is responsible for interpreting the device profiles and performing the transformation to and from the device-independent profile connection space. When the gamut, i.e. the range of reproducible colors, differs between different devices, a gamut mapping algorithm must be used (Morovic, 2008). The principle of ICC color management is illustrated in Fig. 2.7.

![Diagram of ICC color management system](image)

Figure 2.7. The ICC color management system. The relationship between device colors and the profile connection space (PCS) are described by ICC-profiles, for input and output devices.

Ideally, the color management system should perform accurate color transformations between different types of media and devices, but to achieve this is not a trivial task. First, there are significant differences in the gamuts of reproducible colors for different color devices. Furthermore, the differences in viewing conditions for different media imply that a simple colorimetric match does not necessarily give an appearance match. (Sharma, 2003)
2.5 Multispectral imaging

2.5.1 Background

The trichromatic nature of human vision was first discovered in the 19th century, and has later been verified by studies of the human eye. This three-dimensionality of color has formed the basis for colorimetry and for color imaging using three channels, including e.g. modern television, computer displays, as well as film-based photography and digital photography (Fairchild et al., 2001). However, three-channel color imaging has several limitations in color image acquisition and reproduction.

Three-channel imaging is by its nature always metameric, i.e. based on metameric matching rather than spectral matching. However, the sensitivities of typical imaging devices differ from the CIE color matching functions, thus producing metameric matches that differs from those of a human observer. As an example, consider the ICC workflow, relating device values to colorimetric values, as described in Sec. 2.4.3. When two non-identical spectra are metameric for the input device, they will always map to the same colorimetric values, even though they are not metameric with respect to a human observer. Conversely, when two colors are metameric to an observer but not to the input device, the CMM will in error treat the two as having different colorimetric values (Rosen, et al., 2000).

The limitations of metameric imaging are further expanded when the effects of illumination are considered. For example, it is possible for a metameric imaging system to be unable to distinguish a white object under a red light from a red object under a white light (Fairchild et al., 2001). With these limitations in mind, to represent color by its spectral power distributions, i.e. multispectral imaging, is clearly advantageous.

2.5.2 Terminology

The terminology and definitions referring to the concepts of multispectral imaging and multi-channel imaging are sometimes confusing, with different meanings by different authors. Throughout this dissertation, we will refer to multi-channel images as images containing more than the conventional three color channels (with the exception of CMYK-images). By multispectral images we mean images where each pixel contains information about the spectral properties of the samples being imaged. Even though multispectral images are typically derived using multi-channel systems, they can also be derived using conventional trichromatic systems (see for example Connah, et al., 2001; Imai & Berns, 1999). Another terminology referring to multispectral imaging is sometimes called multi-channel visible spectrum imaging (MVSI), or simply spectral imaging (Imai, et al., 2002).

2.5.3 The multispectral approach

It is well known that the only way to assure a color match for all observers across changes in illumination is to achieve a spectral match (Imai & Berns, 1999). By representing color as spectral power distributions, metamerism in color imaging can be avoided. Furthermore, it allows for the separation of the spectral properties for the object from the illumination, thus representing the color of an object by its spectral
reflectance. This is because the physical representation of color, spectral reflectance is completely independent of the characteristics of the image acquisition system. Therefore, the differences between any spectra will be distinguished, independently if they are metameric with respect to any illumination, observer or image capturing device. The multispectral images can be transformed to any color space and be rendered under any illumination. Further more, the gamut will not be limited by the set of primaries of a specific imaging device. The concept of multispectral imaging involves capturing, processing and reproducing images with a high number of channels.

2.5.4 Previous work

The interest in multispectral imaging is rapidly increasing and research is ongoing in several laboratories around the world, focusing on the acquisition, processing and reproduction of multispectral images. Recently, with the decrease in price and sizes of the devices, multispectral technologies are also beginning to find their way into industrial applications (Hill, 2007; Hauta-Kasari, 2007).

There are numerous studies on the different topics related to multispectral imaging and the references listed below are limited to a small selection only, with the intention to give a good overview of the concept. For some general descriptions of multispectral imaging, including aspects on the workflow and processing of the data, we refer to: Berns, et al. (1998), Hauta-Kasari (1999), Rosen et al. (2000), Rosen et al. (2001), Willert, et al. (2006) and Imai (2007). The increasing amount of spectral images available also requires effective methods of storing and retrieving spectral images from databases (Kohonen, 2007).

Multi spectral image acquisition

The acquisition of multispectral images involves recovering spectral properties of the sample being imaged. Typically, the image is captured using multi-channel systems of narrowband characteristics, but trichromatic systems are also used. Hence, multispectral imaging requires for the computation of spectral reflectance data from a relatively small number of channels. It is possible simply because the spectral properties of most surfaces are smooth functions of wavelength (Cheung, et al., 2005).

Works that are more specifically focused on the acquisition of multispectral images, one of the main focuses of this dissertation, include: Imai & Berns (1999), Sugiura, et al. (1999), Imai et al. (2000), Haneishi, et al. (2000), Hardeberg (2001) and Connah, et al. (2001).

Multi spectral image reproduction

On the reproduction side, spectral imaging systems are capable of producing images that are robust to changes in illumination. If a printer has a large set of inks to choose from, it should be possible to select a subset of inks that achieve a spectral match to the multispectral images (Imai & Berns, 1999). When a printed image has the same reflectance properties as the original object, the original and the reproduction will match under any illumination and for any observer or imaging device (Fairchild, 2001). One of the applications in focus for the research on spectral color reproductions is spectral imaging of paintings, avoiding illuminant metamerism (see e.g. Berns, et al., 2008). However, spectral color reproduction includes a number of new considerations and
requirements compared to conventional color reproduction. For example, spectral color reproduction requires new approaches for colorant separation to achieve a spectral match, rather than a colorimetric match (Gerhardt & Hardeberg, 2007). Furthermore, spectral gamut mapping (Urban, et al., 2008) and spectral halftoning (Hardeberg & Gerhardt, 2007) are topics that need to be addressed in spectral print reproduction.

Chapter 3

Device characterization

3.1 Introduction
3.2 Calibration and characterization
3.3 Characterization approaches
3.4 Input devices
3.5 Output devices
3.6 Least-squares regression techniques
3.7 Metrics for evaluating device characterization
3.8 Color target design
3.1 Introduction

Device characterization is the process of deriving the relationship between device-dependent and device-independent color representations, for a device. This chapter intends to provide a background of the concept of device characterization, as well as providing the definitions and terminology associated with the topic.

The difference between device calibration and device characterization is defined as well as the concept of forward and inverse characterization. The two conceptually different approaches for device characterization, model-based and empirical characterization are described for input and output devices.

Among the different mathematical techniques used for data fitting or data interpolation in device characterization, the focus is on the least squares regression techniques. This is to provide readers with the theoretical background for the work described in following chapters of the dissertation.

3.2 Calibration and characterization

It is necessary to distinguish device calibration from device characterization. Device calibration is the process of maintaining a device with a fixed known characteristic color response and should be carried out prior to device characterization. Calibration may involve simply ensuring that the controls of a device are kept at fixed nominal settings, but it may also include linearization of the individual color channels or gray-balancing.

Device characterization derives the relationship between device-dependent and device-independent color data for a calibrated device. For input devices, the signal captured by the input device is first processed through a calibration function while output devices are addressed through a final calibration function, see Fig. 3.1. Typically, device characterization is carried out relatively infrequently compared to calibration, which is done more frequently to compensate temporal changes and maintain a device in a fixed known state. The two form a pair, so if the characteristic color response of the device is altered by a new calibration, the characterization should be re-derived. (Bala, 2003)
The characterization function can be defined in two directions. The *forward characterization* function defines the response of a device to a known input, thus describing the color characteristics of the device. For input devices, this corresponds to the mapping from a device-independent color stimulus, $c$, to the device-dependent signals, $d$, recorded when exposed to that stimulus. For output devices it corresponds to the mapping from the input device-dependent color, $d$, to the resulting rendered color, in device-independent coordinates, $c$. The *inverse characterization* function compensates for the device characteristics and determines the input that is required to obtain a desired response. For an input device the inverse characterization function can be used to compute the device independent representation for a stimulus (e.g. colorimetric values or spectral reflectance) from the recorded device response. For an output device, the inverse function gives the input (e.g. RGB or CMYK) required to obtain a specific colorimetric output.

The output of the calibration and characterization processes is a set of mappings between device-independent and device-dependent color data. These can be implemented as some combination of power-law mapping, matrix conversion, white-point normalization and look-up tables. The widely adopted standard to store this information is the ICC profile, used in color management (Sec. 2.4.3).

### 3.3 Characterization approaches

There are generally two different approaches to derive the characterization function: *model-based* and *empirical characterization*. For input devices, the two approaches are sometimes referred to as *spectral sensitivity based* and *color target based* (Hong & Lou, 2000) or *spectral models* and *analytical models* (Hardeberg, 2001).

Model-based characterization uses a physical model that describes the process by which the device captures or renders color. Access to raw device data is generally required, and the quality of the characterization is dependent on how well the model
reflects the real behavior of the device. Model-based approaches have better generality and can provide insights into device characteristics.

In empirical characterization, the mapping is made using a black box approach, i.e. without explicitly modeling the device characteristics. By correlating the device response for a number of test colors to the corresponding device-independent values, the characterization function is derived using mathematical fitting. Empirical approaches often provide more accurate characterizations for end-user applications, but the functions derived will be optimized only for a specific set of conditions, including the illuminant, the media and the colorant. There are also hybrid techniques, which combine strengths from both model-based and empirical approaches. (Bala, 2003)

3.4 Input devices

The two main types of digital color input devices are scanners and digital cameras, recording the incoming radiation through a set of color filters (typically RGB). The calibration of input devices typically involves establishing all settings, such as aperture, exposure time and internal settings, and to determine the relationship between the input radiance and the device response. The main difference between the characterization of scanners and digital cameras is that scanners employ a fixed illumination as part of the system, while digital cameras may capture images under varying and uncontrolled conditions.

3.4.1 Model-based input device characterization

The basic model that describes the response of an image capturing device with \( m \) color filters is given by:

\[
d_k = \int_{\lambda \in V} E(\lambda) S_k(\lambda) \, d\lambda + \varepsilon_k \quad (3.1)
\]

where \( k = 1, \ldots, m \) (\( m = 3 \) for RGB devices), \( d_k \) is the sensor response, \( E(\lambda) \) is the input spectral radiance, \( S_k(\lambda) \) is the spectral sensitivity of the \( k \)-th sensor, \( \varepsilon_k \) is the measurement noise for channel \( k \), and \( V \) is the spectral sensitivity domain of the device.

If we represent the spectral signal as the discrete \( N \)-component vector, \( \mathbf{f} \), uniformly sampled at wavelengths \( \lambda_1, \ldots, \lambda_N \), Eq. 3.1 can be rewritten as:

\[
\mathbf{d} = \mathbf{A} \mathbf{f} + \mathbf{\varepsilon} \quad (3.2)
\]

where \( \mathbf{d} \) is an \( m \)-component vector of device signals (e.g. [R,G,B]), \( \mathbf{f} \) is the \( N \)-component vector of the input spectral signal, \( \mathbf{A} \) is the \( N \times M \) matrix with columns representing the sensor sensitivities and \( \mathbf{\varepsilon} \) is the measurement noise term. Note that if the wavelength sample interval is larger than 1 nm, then Eq. 3.2 should be completed with the \( \Delta \lambda \) factor.

According to Sec. 2.2, we know that colorimetric signals can be computed by:
\[ \mathbf{c} = \mathbf{A} \mathbf{f} \]  

(3.3)

where \( \mathbf{c} \) is the colorimetric vector \([X, Y, Z]\) and \( \mathbf{A} \), the \( N \times 3 \) matrix with columns representing the XYZ color-matching functions.

From Eq. 3.2 and 3.3, it can be seen that, in the absence of noise, there exists a unique mapping between device-dependent signals \( \mathbf{d} \) and device independent signals \( \mathbf{c} \), when there is a unique transformation from \( \mathbf{A}_d \) to \( \mathbf{A}_c \). In the case of three-channel devices (RGB), the sensor’s response \( \mathbf{A}_d \) must be a non-singular transformation of the color matching functions, \( \mathbf{A}_c \). Devices that fulfill this so-called Luther-Ives condition are referred to as colorimetric devices (Bala, 2003).

Unfortunately, practical considerations make it difficult to design sensors that meet this criterion. The assumption of a noise-free system is unrealistic and most filter sets are designed to have more narrow-band characteristics than the color matching functions, to improve efficiency. For the majority of input devices that do not fulfill the conditions for being colorimetric devices, the relationship between XYZ and device RGB is typically more complex than a linear \( 3 \times 3 \) matrix. (Bala, 2003)

### 3.4.2 Empirical input device characterization

The workflow for empirical input device characterization is illustrated in Fig. 3.2. After the device has been calibrated, the characterization is performed using a target of color patches that spans the color gamut of the device. The device-dependent coordinates \( \{d_i\} \) (e.g. RGB) are extracted for each color patch and correlated with the corresponding device-independent values \( \{c_i\} \) (typically XYZ or L*a*b*), obtained using a spectroradiometer or spectrophotometer. Then the characterization functions, relating the device response to colorimetric representations, are computed e.g. by least squares regression techniques, as described in Sec. 3.6.

![Figure 3.2. A schematic diagram for the workflow of empirical input device characterization (Bala, 2003).](image)

### 3.5 Output devices

Recall from Chapter 2 that output color devices generally can be broadly categorized into two groups: emissive devices producing colors via an additive mixing of light

32
(typically RGB), and devices that produce reflective prints or transparencies via subtractive color mixing.

3.5.1 Model-based output device characterization

In the case of emissive displays, the characterization is greatly simplified by the assumptions of *channel independence* and *chromaticity constancy*. This means that each of the RGB channels operates independently of the other and that the spectral radiance from a given channel has the same basic shape and is only scaled as a function of the signal driving the display. With these assumptions, model-based display characterization is a fairly simple process, resulting in a $3 \times 3$ matrix conversion relating XYZ to RGB values. The calibration of emissive displays involves finding the relation between the RGB input values and the resulting displayed luminance. This relationship is typically modeled using a power-law relationship (gamma correction) for CRT-displays, while LCD-displays are better modeled using a sigmoidal S-shaped function.

Output devices based on subtractive color mixing, such as color printers, exhibit more complex nonlinear characteristics, which makes the model-based approach a challenge. For example, the halftoning introduces additional optical and spatial interactions, with the co-existing physical and optical dot gain giving non-linear increases in tone values. This will be further treated in Chap. 4, together with an overview of physical models of color printing, focusing on predicting the outcome of halftone prints.

3.5.2 Empirical output device characterization

The workflow for output device characterization is illustrated in Fig. 3.3. A digital target of color patches spanning the gamut of the device, with known device values, $d_i$, (e.g. RGB or CMYK) is sent to the device and the resulting (displayed or printed) colors are measured in device-independent coordinates, $c_i$. A relationship is established between device-dependent and device-independent color representations which can be used to derive the characterization function. To evaluate the characterization functions, a validation set of test targets that are independent from the training set should be used.

![Figure 3.3. The workflow for Empirical output device characterization (Bala, 2003).](image)

The main drawback of the empirical approach is the strong media-dependency. The characterization functions (or ICC profiles) will be optimal only for the specific set
of conditions used to generate them. In color printing, this means that new characterization functions must be derived as soon as one component, such as the paper, is altered. The impact of different paper properties on profile generation has been studied in Norberg, 2006.

3.6 Least-squares regression techniques

In empirical approaches, the characterization functions are derived by relating the device response, \( d_i \), to the corresponding device-independent values, \( c_i \), for a set of color samples. In the following, let \( \{d_i\} \) be the \( m \)-dimensional device-dependent coordinates for a set of \( T \) color samples, and \( \{c_i\} \) be the corresponding set of \( n \)-dimensional device-independent coordinates \((i = 1, \ldots, T)\). For most characterization functions, \( n = 3 \) (typically XYZ or L*a*b*) and \( m = 3 \) or \( 4 \) (typically RGB or CMYK). Note however, that for multi-channel imaging systems \( m > 3 \), depending on the number of color channels utilized. The pair \((\{d_i\}, \{c_i\})\) refers to the set of training samples and is used to evaluate one or both of the following functions:

\[
 f : F \in \mathbb{R}^n \rightarrow \mathbb{R}^m ,\text{ mapping the device-dependent data with the domain } F \text{ to device independent color space.}
\]

\[
 g : G \in \mathbb{R}^n \rightarrow \mathbb{R}^m ,\text{ mapping device-independent data with the domain } G \text{ to device-dependent color space.}
\]

To derive the characterization functions, mathematical techniques as for instance data fitting or interpolation are used. A common approach is to parameterize \( f \) (or \( g \)) and to find the parameters by minimizing the mean squared error metric using least-squares regression.

Besides least squares regression, other approaches for data fitting and interpolation used in device characterization include e.g. lattice-based interpolation, spline fitting and the use of neural networks. (Bala, 2003)

3.6.1 Linear least squares regression

The simplest form of least squares regression is linear least squares regression. The characterization function is approximated by a linear relationship \( c = d \cdot A \), where \( d \) is the \( 1 \times m \) input color vector and \( c \) is the \( 1 \times n \) output color vector.

The optimal \( m \times n \) matrix \( A \) is derived by minimizing the mean squared error of the linear fit to the set of training samples \((\{d_i\}, \{c_i\})\), as:

\[
 A_{opt} = \arg\min_A \left\{ \frac{1}{T} \sum_{i=1}^{T} \lVert c_i - d_i A \rVert^2 \right\} \quad (3.3)
\]

If the samples \( \{c_i\} \) are collected into a \( T \times n \) matrix \( C = [c_1, \ldots, c_T] \) and \( \{d_i\} \) into a \( T \times m \) matrix \( D = [d_1, \ldots, d_T] \), then the linear relationship is given by:

\[
 c_i = d_i A_{opt} \]
(3.4)

and the optimal $A$ is given by:

$$A = (D'D)^{-1}D'C = (D)^{-}C$$

(3.5)

where $(D) = (D'D)^{-1}D'$ denotes the Moore-Penrose pseudo-inverse of $D$ (sometimes referred to as the generalized inverse).

### 3.6.2 Polynomial regression

Polynomial regression is a special case of least squares regression, where the characterization function is approximated by a polynomial. For example, the inverse characterization function of a 3-channel system, mapping RGB values to XYZ tristimulus values, is obtained by expressing XYZ as a polynomial function of $R$, $G$ and $B$. The regression from $\mathbb{R}^3$ to $\mathbb{R}^3$, i.e. from RGB to XYZ, is equivalent to three independent regressions from $\mathbb{R}^3$ to $\mathbb{R}$, i.e. from RGB to each of the XYZ components (Hardeberg, 2001). As an example, a second order polynomial approximation is given by:

$$[X Y Z] = \begin{bmatrix} 1, R, G, B, R^2, RG, RB, G^2, GB, B^2 \end{bmatrix} \begin{bmatrix} w_{X,1} & w_{Y,1} & w_{Z,1} \\ w_{X,2} & w_{Y,2} & w_{Z,2} \\ \vdots \\ w_{X,10} & w_{Y,10} & w_{Z,10} \end{bmatrix}$$

(3.6)

or, generally:

$$c = pA$$

(3.7)

where $c$ is the $n$-component colorimetric output vector, $p$ is the $Q$-component vector of polynomial terms derived from the input vector $d$, and $A$ is the $Q \times n$ matrix of polynomial weights to be optimized.

After arranging the input device data $d$ into the polynomial vector $p$, the polynomial regression is reduced into a linear least squares regression problem, with the optimal matrix of polynomial weights, $A$, given by:

$$A_{opt} = \arg\min_A \left\{ \frac{1}{T} \sum_{i=1}^{T} \left\| c_i - p_i A \right\|^2 \right\}$$

(3.8)

Again, collecting the samples $\{c_i\}$ into a $T \times n$ matrix $C = [c_1, ..., c_T]$ and the polynomial terms $\{p_i\}$ into a $T \times Q$ matrix $P = [p_1, ..., p_T]$, the relationship given in Eq. 3.7 can be rewritten as:
\[ C = PA \]  

(3.9)

Then, the optimal solution for \( A \) is given by:

\[ A = (P'P)^{-1}P'C = (P)^\dagger C \]  

(3.10)

where \((P)^\dagger\) denotes the pseudo-inverse of \( P \). In order that the \( Q \times Q \) matrix \((P'P)\) to be invertible, it requires \( T \geq Q \), i.e. the number of training colors should be at least as many as the number of polynomial terms. For the preferred case, \( T > Q \), we have an over-determined system of equations for which we derive the least-squares solution. Unfortunately, the number of the polynomial terms increases rapidly with the order of the polynomial. For example, a third-order polynomial in the complete form, including all the cross-product terms, gives \( Q=64 \). Therefore, simplified approximations are commonly used with a smaller number of polynomial terms than in the complete form. In general it is recommended to use the smallest number of polynomial terms that adequately fits the function while still smoothening out the noise (Bala, 2003).

### 3.7 Metrics for evaluating device characterization

Generally speaking, the minimized quantitative error metrics of the regression provide an indicator of the overall accuracy of the characterization. However, using the regression error alone is not sufficient to evaluate the characterization, for several reasons. First, the error is available only for the training samples used to derive the characterization function. Hence the accuracy of the characterization should be evaluated using independent test targets, not the training colors. Furthermore, the error used in the fitting or interpolation techniques is not always calculated in a visually meaningful color space. Evaluation of errors should preferably be within visually relevant metrics, such as the color difference formulae \( \Delta E^{*a*b*} \) or \( \Delta E^{*94} \) previously described in Sec. 2.2.

The statistics usually reported in the literature, when evaluating device characterization, are the mean, maximum and standard deviation, of the \( \Delta E \) values for a set of test samples. The 95th percentile of \( \Delta E \) values (i.e. the value below which 95% of the \( \Delta E \) values lie) is also sometimes computed. To judge whether the characterization error is satisfactory small or not, using the chosen error metric, depends very strongly on the application and the expectations from the user. The accuracy of the characterization will also be limited by the inherent stability and uniformity of a given device. (Bala, 2003)

### 3.8 Color target design

The first factor to consider in color target design is the set of colorants and underlying medium of the target. For input devices, where the target is not created as part of the characterization process, the colorants and media should be representative of what is likely to be used for the device. For output color devices, the generation of the target is part of the characterization process, and should be carried out using the same colorants and media as in the final color rendition.
Another factor is the choice of color patches. Typically, they are chosen to span the gamut, i.e. the range of colors to be captured (input devices) or rendered (output devices). Considering the spatial layout of the patches, it may be desirable to generate targets with the same colors in different spatial layouts. In such a manner, measurements from multiple targets can be averaged to reduce the effects of non-uniformity and noise in the characterization process. (Bala, 2003)

A few targets have been adopted as industry standards, e.g. the GretagMacbeth ColorChecker chart and the Kodak photographic Q60 target for input device characterization, and the IT8.7/3 CMYK target for printer characterization.
Chapter 4

Halftone color reproduction

4.1 Introduction
4.2 Digital halftoning
4.3 Dot gain
4.4 Modeling halftone color reproduction
4.1 Introduction

Halftone color reproduction, i.e. producing various colors by small dots with varying area coverage, is by far the most common way of reproducing color images in print. In halftone prints on paper, physical and optical dot gains generally occur, making the resulting print darker than expected, and making the modeling and characterization of halftone color reproduction a challenge. The aim of this chapter is to provide a short background, introducing the concept of digital halftoning and dot gain. In order to put the content of later chapters in some perspective, an overview of previous work on modeling the output of halftone prints is also given. Even though many of today’s digital printing devices use multilevel halftoning, the discussion has been limited to binary halftoning, printing with two levels of ink only.

The model overview takes on a historical approach, focusing mainly on the milestones of models predicting the outcome of halftone prints, from the 1930s and on. The classical, well known models are presented, along with their limitations and the later proposed developments and refinements over the years. More recent approaches, such as modeling halftone reproduction by point spread functions and probability models, are mentioned briefly for completeness, along with references for the interested reader. The model overview has been limited to models dealing directly with halftone reproduction, therefore ignoring well-known models such as the Kubelka-Munk model (Kubelka & Munk, 1931), which in its general form deals with uniform colorant layers.

In order to keep the length of the background part manageable, several topics that are not directly necessary for the discussion in the remaining chapters have been left out, even though being important in the field of halftone print reproduction. For example, no printing technologies are discussed and neither are any halftone methods or considerations in halftone design. The interested reader can refer to e.g. Kipphan (2001) for printing technologies and to Kang (1999) for digital halftoning. Furthermore, the light-paper interaction, which is of great importance in halftone color reproduction on paper, is in fact a far more complex topic than what is made apparent in the simplified examples in this overview (Yang, 2003). For a textbook on the field of paper optics refer e.g. to Pauler (1998) and for a detailed description on the field of modeling and simulation of light scattering in paper to Edström (2007).
4.2 Digital halftoning

Most printing devices are only bilevel, meaning that they are able of printing ink only at a certain fixed density, or leaving the substrate unprinted. In order to reproduce a continuous tone image on such printing device, it must first be transformed into a binary image, usually called a bitmap, consisting of zeros and ones only. The transformation from continuous tone into a binary bitmap image is referred to as halftoning, or screening. The basic idea of halftoning is that when the dots are small enough, the halftone pattern will not be visible at normal viewing distance, but visually integrated and interpreted as varying shades of gray or color.

The different approaches for halftoning can generally be divided into two main categories: Amplitude Modulated (AM) and Frequency Modulated (FM) halftoning. In AM halftoning, the frequency of the halftone dots is constant, and the different tonal values are reproduced by varying the size of the halftone dots. In FM halftoning, the dot size is constant and the number (and thus the frequency) of dots varies. Generally, FM methods are better for reproducing fine details, while the AM method is better in reproducing slowly varying tones. Figure 4.1 illustrates the concept of AM halftoning for different tone levels in the case of 8×8 halftone cells (upper part), and the corresponding tone levels halftoned using a FM technique (lower part).

![Figure 4.1. Examples of AM and FM halftoning for the gray levels 12.5%, 25%, 50% and 75%.

In AM halftoning, the image is built from a number of halftone cells, each consisting of a number of smaller microdots, the smallest reproducible dot size. The screen ruling or screen frequency denotes the number of halftone cells per inch, in lpi. The print resolution denotes the number of micro dots per inch, in dpi. The ratio of the print resolution and the screen frequency determines the number of micro dots per halftone cell, and thus the number of reproducible gray levels, as:

\[
\text{Graylevels} = \left(\frac{\text{dpi}}{\text{lpi}}\right)^2 + 1
\] 

(4.1)
Clearly, there has to be a tradeoff between the spatial and the tonal resolution. For a fixed print resolution (in dpi), increasing the screen frequency (in lpi) will increase the spatial resolution, but at the cost of the number of reproducible gray levels. FM halftoning methods generally do not use halftone cells and therefore only the print resolution (in dpi) is relevant, while the screen frequency (in lpi) is not used in the case of FM halftones.

4.2.1 Color halftoning

In order to reproduce a color image, it must first be separated into three or more color separations. As described in Chap. 2, the subtractive primary colors cyan, magenta and yellow (CMY) are most often used in three color printing, with the additional black (K) colorant added in four color printing. Printing three or four color separations on top of each other introduces additional considerations regarding interference between the color separations. Only a slight misregistration of the halftone screens can introduce color errors and unwanted Moiré patterns. To reduce the sensitivity to Moiré effects in AM halftones, different screen angles are commonly used, printing the color separations at different angles. Usually, the four angles correspond to 75°, 15°, 0° and 45° for cyan, magenta, yellow and black, respectively. Rotated screens reduces the effect of misregistration but, on the other hand, introduce a new type of pattern, commonly referred to as rosette patterns, visible for lower screen resolutions, see Fig. 4.2(a).

In FM color halftoning there is no need for rotated screens and the rosette and Moiré patterns are generally avoided, since the screen patterns are no longer regular and periodic. However, FM color halftones can sometimes give a somewhat “grainy” appearance; see Fig. 4.2(b). Normally, the different color separations are halftoned independently of each other. It has, however, been shown that dependent FM color halftoning can increase the print quality and at the same time reduce the amount of ink (Gooran, 2001; 2004).

![Figure 4.2. Tints of 20% cyan, magenta and yellow, halftoned using AM halftoning with rotated screens (a), and FM halftoning (b).]
4.3 Dot gain

After printing, the halftone dots generally appear larger than their nominal size in the bitmap image, resulting in increased tonal values than expected, a phenomenon usually referred to as dot gain. Dot gain in halftone prints actually encompasses two fundamentally different phenomena: Physical dot gain (also known as mechanical dot gain) and Optical dot gain (also known as the Yule-Nielsen effect).

Physical dot gain refers to the fact that the size of the printed halftone dots differs from their nominal size in the bitmap. Physical dot gain is closely related to the printing process, including the operation settings on the press, the ink-transfer and ink-setting processes. For example, sources of physical dot gain could be ink spreading due to a low ink viscosity, the film exposure to plate, or the print cylinder pressure.

Optical dot gain, on the other hand, originates from light scattering inside the substrate, causing light exchanges between different chromatic areas. The result is that the dot area appears larger when the reflective light is perceived/measured, compared to the physical dot size, and hence a darker tone value. Figure 4.3 gives a simplified example of the causes of optical dot gain, illustrating different paths for a photon entering a halftone print on paper. Photon $A$ is reflected through the ink layer, photon $B$ is absorbed in the ink, photon $C$ is directly reflected at the paper surface and photon $D$ illustrates diffuse bulk reflection within the paper. The reason for optical dot gain is illustrated by photon $E$, entering the bare paper and is then scattered within the paper under a halftone dot, and partially absorbed by the ink on its way back.

![Figure 4.3](image)

Figure 4.3. Different paths for a photon entering a halftone print. Optical dot gain is caused by light exchange of different chromatic areas, illustrated by photon $E$.

The effect of optical dot gain depends on the ratio of the length of the lateral scattering of light within the substrate to the size of the halftone dots. When the dot size becomes small in relation to the lateral scattering length, the optical dot gain will increase. The lateral scattering length in paper is typically in the order of 0.1 mm (Arney, et al., 1996a; Rogers, 1998b). Because of optical dot gain, the relationship between the printed dot area fraction and the perceived/measured reflectance is non-linear, making the modeling of halftone color reproduction a challenging task. It has, however, been shown that the existence of optical dot gain actually increases the color gamut, i.e. the range of reproducible colors (Gustavson, 1997; Yang, 2003).
Due to their different intrinsic nature, physical and optical dot gain need to be treated separately in order to accurately model the outcome of halftone prints. However, in measurements of reflection colors, physical and optical dot gains always co-exist, making the separation of one type from another a difficult task, a topic that will be further addressed in Chap. 9.

Throughout this dissertation, the Nominal dot area will denote the intended dot area, calculated from the binary image sent to the printer. The Physical dot area is the actually printed dot area, including the change in physical dot size. The Effective dot area is an estimated value, including the effect of both physical and optical dot gain.

4.4 Modeling halftone color reproduction

Many models predicting the color output from halftone prints have been proposed and later further refined and developed, starting from the 1930s. There is no intention here to present a complete overview of every model ever proposed, merely to give a historical overview, focusing on well-known models that have been applied in many contexts and been subject to further development over the years. All the models describe the forward function, predicting the resulting color output, in terms of spectral reflectance or tristimulus values, given the input to the printer in terms of the dot area fractions of the colorants.

Due to the limitation in instruments available at the time, many of the older models were first presented in the form of print density values. Here, they are all presented in their spectral reflectance form, predicting the output spectral reflectance. However, unless otherwise noted, the models are also applicable to predict the spectrally averaged mean reflectance, or the tristimulus values CIEXYZ, given mean reflectance or tristimulus values as input.

4.4.1 The Murray-Davies model

In the 1930s Murray and Davies published the first model to predict the output reflectance of a monochrome halftone print (Murray, 1936). The mean reflectance $R_\text{a}$ is simply given by linear interpolation of the reflectance of the bare paper, $R_p$, and the full tone ink, $R_i$, weighted by the dot area fraction, $a$, as:

$$R(a) = aR_i + (1-a)R_p$$  \hspace{1cm} (4.2)

Since the reflected light from different areas is added to predict the overall reflectance, the Murray-Davies (MD) model makes a good intuitive sense and preserves the linearity of photon additivity. Figure 4.4 illustrates the principle of the MD formula, where the total reflectance is given by the sum of the ink reflectance, $R_\text{a}$, with area $a$ and the paper reflectance, $R_p$, with area $(1-a)$. 

45
Figure 4.4. The principle of the Murray-Davies formula. The total reflectance is given by the reflectance of the paper, $R_p$, with area $(1-a)$, and the reflectance from the printed ink, $R_i$, with area $a$.

It is, however, well known that the applicability of the MD model is limited, and that it over predicts the reflectance when the nominal dot area is used in the equation. The relationship of $R$ versus $a$ is in fact non-linear, due to light scattering in the paper substrate, causing optical dot gain. Figure 4.5(a) displays an example of the predicted mean reflectance of a 600 dpi FM halftone print using the MD formula, compared to the measured reflectance for varying dot area, $a$. Figure 4.5(b) displays the spectral reflectance predicted by the MD formula, compared to experimental data, for a magenta tint of 50% nominal dot area coverage.

Another use of the Murray-Davies formula is to predict the effective dot area by rearranging Eq. 2 into:
The effective dot area can then be used to compute the total dot gain as the difference to the nominal dot area coverage. If spectral reflectance data is used, Eq 4.3 is most often computed at a single wavelength, typically the wavelength of minimum reflectance, because it varies the most with varying area coverage (Wyble and Berns, 2000). Nevertheless, one has to keep it in mind that the optical dot gain is spectral dependent (Yang, 2003), hence using only a single wavelength in Eq. 4.3 may not be sufficient.

4.4.2 The Yule-Nielsen model

In the 1950s, Yule and Nielsen published their famous work on light penetration and scattering in paper (Yule & Nielsen, 1951). It was then shown that the nonlinear relationship between the dot area fraction and the reflectance could be approximated by a power function. To the reflectance values in the Murray-Davies formula, the exponent $1/n$ was added, as:

$$R(a) = \left[ aR_1^{1/n} + (1-a)R_p^{1/n} \right]^n$$

(4.4)

where $n$ is the Yule-Nielsen $n$-factor, accounting for light scattering in the paper, and all other variables are the same as in the MD formula. The $n$-factor is an empirically derived constant, selected to provide the best fit to experimental data.

One argument that has been made is that the Yule-Nielsen (YN) model does not physically describe the phenomenon of optical dot gain, since the conservation of energy is lost when the nonlinear transform is applied to the reflectance values (Arney, et al., 1995a; 1995b). Nevertheless, the Yule-Nielsen model is still commonly used because it is simple, works relatively well, and predicts the printer output better than Murray-Davies.

The Yule-Nielsen $n$-factor

Much work has been published on the physical meaning of the Yule-Nielsen $n$-factor. Ruckdeschel & Hauser (1978) presented a physical analysis relating $n$ to light scattering in paper, and concluded that unless other factors than optical scattering of light are involved, only $n$-values of $1 \leq n \leq 2$ are physically meaningful. For $n=1$ there is no light scattering (and YN reduces to the MD formula), and for $n=2$ complete light scattering occurs in the substrate. It has been suggested that an average $n$ value of 1.7, should be satisfactory when the real value is unknown (Pearson, 1980). However, with modern high-resolution printers, values of $n$ greater than 2 are often required (Wyble & Berns, 2000). Many studies report the best results for $n>2$, for which there is no direct physical interpretation (see e.g. Shiraiwa & Mizuna, 1993; Balasubramanian, 1999). One explanation is that the contribution of physical dot gain has not been properly taken into account (Yang, 2003). When modelling optical dot gain, it is necessary to use the physical dot area coverage, including the physical dot gain instead of the nominal dot area coverage.
area. This topic will be further discussed in Chap. 9 along with methods for estimating the physical dot gain, thus separating it from the optical dot gain.

Another regression-based model allows \( n \) to vary with wavelength in the spectral version of Yule-Nielsen (Inno & Berns, 1998a; 1998b):

\[
R(a) = \left[ aR_i^{1/n(\lambda)} + (1-a_i)R_p^{1/n(\lambda)} \right]^{n(\lambda)}
\]

(4.5)

The spectrally varying \( n \)-value can improve the model significantly, but also requires more computation to optimize \( n \) for each wavelength. Note that there is no physical meaning of \( n(\lambda) \), it is merely a fitted parameter to provide the best fit to experimental data (Wyble and Berns, 2000).

4.4.3 The Neugebauer model

The famous Neugebauer model (Neugebauer, 1937) is a relatively straightforward extension of the monochrome Murray-Davies formula to handle the multiple colorants in color printing. The reflectance is given as the summation of the reflectance for the different colorants (including overlapping primary colors), weighted by their fractional area coverage, as:

\[
R = \sum_{j=1}^{2k} a_j R_j
\]

(4.6)

where \( k \) is the number of printing inks and \( a_j \) is the fractional area for each of the Neugebauer primaries, with reflectance \( R_j \). For a three color print, the eight Neugebauer primaries correspond to: white (the bare substrate), cyan, magenta and yellow (the primary inks) red, green blue (the secondary colors) and three color black. A common assumption when computing the area fractions is that the halftone dots are randomly distributed on the substrate which can then be calculated by the Demichel (1924) equations. For example, the area fractions for the eight Neugebauer primaries in a three color print are given by:

\[
a_w = (1-c)(1-m)(1-y) \\
a_c = c(1-m)(1-y) \\
a_m = (1-c)m(1-y) \\
a_y = (1-c)(1-m)y \\
a_{my} = (1-c)my \\
a_{cy} = c(1-m)y \\
a_{cm} = cm(1-y) \\
a_{my} = cmy
\]

(4.7)
where \(c, m, \) and \(y\) are the dot area fractions for the primary colors cyan, magenta and yellow (or, statistically, the probabilities for a given point to be covered by these inks). The extension to the 16 Neugebauer primaries in four color printing is then straightforward. Note that the assumption of randomly placed halftone dots requires that the halftone screens of the color separations are statistically independent of each other. For more detailed descriptions and considerations regarding the randomness of halftone screens refer to: Rogers (1998a) and Amidor & Hersch (2000).

Besides the area coverage computed by the Demichel equations, the only input to the Neugebauer model is the reflectance measurements of the Neugebauer primaries. Being a rather straightforward extension to the Murray-Davies model, the same assumptions and limitations also applies to the Neugebauer model.

**Yule-Nielsen modified Neugebauer**

A natural extension to the Neugebauer formula is to combine it with the Yule-Nielsen formula as proposed by Yule and Colt (1951) in terms of print density, and by Viggiano (1995) in its spectral reflectance form. The Yule-Nielsen modified Neugebauer is commonly referred to as the \(n\)-modified Neugebauer (Emmel, 2003):

\[
R^{1/n} = \sum_{j=1}^{k} a_j R_j^{1/n}
\]

The Yule-Nielsen modified Neugebauer is widely used for the characterization of printers and plays a significant role for print characterization in color management (Hersch, 2005; Hersch & Hébert, 2006).

**Cellular Neugebauer**

The classical Neugebauer model can be seen as interpolating through the printer gamut from only a few points on its surface. For example, for a three-color print, the printer gamut forms a cube with the eight Neugebauer primaries corresponding to the combinations of cyan, magenta and yellow of either 0 or 100% area coverage, located at the corners (Fig. 4.6a). In the Cellular Neugebauer model, the printer gamut is subdivided into smaller cells to provide more intermediate “primaries” inside the printer gamut (Heuberger, 1992; Rolleston & Balasubramanian, 1993). For example, if 50% area coverage of cyan, magenta and yellow are included, each colorant has then three levels and there are \(3^3 = 27\) Neugebauer primaries (Fig. 4.6b). Note that the cellular Neugebauer actually models a multi-level halftone process, with more than two ink levels, and that the justification for using the cellular model for binary halftones is empirical rather than physical (Bala, 2003). Adding more intermediate primaries gives a finer cellular division in CMY-space and thus a finer interpolation of the measured data, hence yielding a greater accuracy. However, more Neugebauer primaries naturally requires more measurements, and also for additional computations to determine which primaries to use in the interpolation, usually done by iterative search methods (Wyble and Berns, 2000).
4.4.4 The Clapper-Yule model

After formulating the Yule-Nielsen equation, Yule worked with Clapper to develop a model for halftone prints, taking into account surface reflection, internal reflection and ink transmission (Clapper & Yule, 1953). The Clapper-Yule (CY) formula is an extension of the Yule-Nielsen formula in its complete light scattering form \((n=2)\), taking into account the Fresnel reflections, occurring at the interface of two media of different refraction indices, as:

\[
R(a) = K r_s + (1-r_s)(1-r_i) - \frac{R_0(1-a+at)^2}{1-r_s R_0(1-a+at^2)} \tag{4.9}
\]

In the equation, \(a\) is the physical dot area coverage, \(t\) is the spectral transmittance of the ink, \(R_0\) is the bulk reflectance of the paper substrate, \(r_s\) and \(r_i\) are the Fresnel reflectance values of external and internal reflections in the interface between air and paper. The factor \(K\), ranging from 0 to 1, gives the fraction of specular reflected light reaching the detector and depends on measurement geometry. For instruments using the d/0° or the 45°/0° measurement geometry, the specular component can be discarded, i.e. setting \(K=0\) (Hersch, et al., 2005).

The meaning of the parameters in the Clapper-Yule model is illustrated in Fig. 4.7. Incident light reaches a halftone print with area coverage \(a\) and ink transmittance \(t(\lambda)\). The specular reflection, \(r_s\), occurs at the air-paper interface and only the portion \((1-r_s)\) enters the paper. With the probability \((1-a)\) that the light enters the paper without traversing the ink layer, the light reaching the paper substrate is reduced by a factor of \((1-r_s)(1-a+at)\). It is then diffusely reflected by the paper, with bulk reflection \(R_0\). When exiting the substrate it is internally reflected at the paper-air interface according to the reflection factor \(r_i\). The remaining part, \((1-r_i)\), exits the paper with the portion \(a\) through the ink layer and the portion \((1-a)\) without traversing the ink. Thus, at the first reflection, the spectral attenuation of the incident light is:

\[
(1-r_s)(1-r_i)R_0(1-a+at)^2 \tag{4.10}
\]
The part reflected at the paper-air interface is then again diffusely reflected by the paper and travels upwards again. At the \( n \):th reflection, the spectral attenuation is given by:

\[
(1-r_j)(1-r_f)R_0(1-a+at)^2 \left( r_j(1-a+at^2)R_0 \right)^{n-1}
\]  

(4.11)

The sum of this geometrical series for an infinite number of internal reflections then leads to the Clapper-Yule equation, according to Eq. 4.9 (Hersch, et al., 2005).

\[
R = K r_j + (1-r_j)(1-r_f)R_0 \left( \sum_{j=1}^{s} a_j t_j \right)^2 \left( 1 - r_j R_0 \sum_{j=1}^{s} a_j t_j^2 \right) 
\]  

(4.12)

where \( a_j \) is the relative amount of colorant \( j \) and \( t_j \) is the corresponding spectral transmittance.

Similar to the Yule-Nielsen modified Neugebauer, the Clapper-Yule model can also be combined with the Neugebauer formula to predict the reflectance of multicolorant halftones (Hersch, et al., 2005):

\[
Figure 4.7. Attenuation of light by multiple reflections in a halftone print, according to the Clapper-Yule model.
\]
and independent of the dot area fraction. As will be shown in later chapters, this is rarely the case in reality.

4.4.5 Modeling the paper spread function

In more recent years, halftone print modeling has taken on a different approach, addressing the theoretical physics of light scattering within the paper and the halftone dots. By measuring, or simulating, the paper point spread function (PSF), a mathematical description of the light spreading properties of the paper, the output reflectance can be computed by a convolution to the halftone image. Among others, the approach was adopted by Kruse & Wedin (1995) and has then been further developed and refined by Gustavson (1997):

\[
R(x, y, \lambda) = [I(x, y, \lambda)T(x, y, \lambda) * P_{xy}(x, y, \lambda)]T(x, y, \lambda) \tag{4.13}
\]

where \(R(x, y, \lambda)\) is the predicted reflectance, \(I(x, y, \lambda)\) is the distribution of the incident light, \(T(x, y, \lambda)\) is the transmittance of the ink and \(P_{xy}(x, y, \lambda)\) is the point spread function, describing the diffuse reflection in the paper. Note that both the predicted reflectance and the ink transmittance are now functions of spatial location, and that the color of the halftone dots is thus no longer necessary assumed to be uniform.

Examples of other studies incorporating the point spread function (PSF) of paper, or the corresponding modulation transfer function (MTF) in the Fourier domain, include: Rogers (1998a; 1998b; 2000a), Agar (2000) and Yang, et al. (2001b).

Compared to the earlier models, the computational complexity is now greatly increased and the data collection and model fitting is considerably more difficult. The main focus is now to accurately model and learn more about the physical process. However, such models are generally not the best choice for predicting halftone print output in printer profiling. (Wybble & Berns, 2000)

4.4.6 The Probability approach

Since optical dot gain is caused by light exchanges between different chromatic areas, one approach aims at determining the probability that light enters and exits any given pair of colorant areas. The approach was first applied to monochrome prints (Arney, 1997), and later expanded to handle multi-color prints (Arney, et al., 1998). In the following, the subscripts \(i\) and \(j\) denote the colorant through which the light enters and exits, respectively. First, the area coverage of each primary is measured or computed (e.g. using the Demichel equations). Then, the probability that the light will exit the same area, \(j\), that it enters is modeled as:

\[
P_{ij} = 1 - (1 - a_j) \left[1 - (1 - f_i)^w + (1 - f_j^w)\right] \tag{4.14}
\]

where \(a_j\) is the area coverage of colorant \(j\), and \(w\) is an empirically derived parameter.

The probabilities that the light exits and enters different chromatic areas are then given by:
\[ P_{ji} = (1 - P_{jj}) \left( \frac{a_i}{a - a_j} \right) \]  

(4.15)

Note that for a three-color print, i.e. with eight different chromatic areas, there are 64 different light paths to be considered and hence 64 probabilities that need to be computed. The probabilities computed by Eq. 4.15 are then used as weights to determine the reflectance of each primary, as:

\[ R_i = R_p T \sum_j T_j P_{ji} \left( \frac{a_i}{a_j} \right) \]  

(4.16)

where \( R_p \) is the reflectance of the paper and \( T_i \) is the ink transmittance for colorant \( i \). Eq. 4.16 basically sums up the light entering at different chromatic areas, \( j \), reflects the light with the substrate reflectance, \( R_p \), and then filter it with the ink transmittance, \( T_i \), when exiting. Once the reflectance, \( R_i \), of each primary has been derived according to Eq. 4.16, it can be used in the Neugebauer equation (Eq. 4.6) to compute the output reflectance. Like Murray-Davies and Neugebauer, the probability model preserves the linearity of photon additivity, since the reflected light from different areas is added to predict the overall reflectance. It does, however, require the empirical exponent, \( w \), and the computation of the probabilities of all the possible light paths between the colorants (i.e. 64 for a three color print). However, the model was evaluated using only two colors, using a fixed amount of magenta and varying only the cyan colorant (Arney, et al., 1998).

The probability approach has later been refined and developed to handle the effect of ink penetration, for monochrome and multicolor prints (Yang, 2003). Instead of using the empirical parameter \( w \), the light-exchange probability function has been related to the point spread function (PSF) for the substrate (Yang et al., 2001a). This was demonstrated for multicolor prints, by employing Gaussian types of PSFs (Yang et al., 2001b).
Chapter 5

Spectral sensitivity estimation

5.1 Introduction
5.2 The Image acquisition system
5.3 Theory of spectral sensitivity estimation
5.4 Color targets
5.5 Experimental setup
5.6 Results
5.7 Summary and discussion
5.1 Introduction

Deriving the forward characterization function describing the device response for a known spectral input requires the spectral characteristics of all the components in the system. This chapter focuses on the spectral characterization of an experimental image acquisition system. The spectral power for the illuminant is derived by measurements using a spectroradiometer, along with the spectral characteristics of the color filters. The main focus of the chapter is the estimation of the spectral sensitivity function for the camera, which is not as straightforward as the properties of the illuminant and color filters since it cannot be derived by direct measurements.

There are mainly two different classes of methods to determine the spectral sensitivity function of an image acquisition device such as a CCD-camera. The first class of methods is based on direct recording the device response to monochromatic light for each wavelength of the visible spectrum (see e.g. Martínez-Verdú, et al., 2000). The method is highly accurate, but requires a monochromator, an instrument that can filter a broadband light source into narrow-bands of wavelengths. The second type of approach is to estimate the spectral sensitivity, by relating the recorded device response to the known reflectance for a set of test samples (Hardeberg, 2001).

The second, empirical approach is adopted, estimating the spectral sensitivity using carefully selected color samples as a training set. By relating the camera response to the spectral reflectance, the camera sensitivity function is estimated using least-squares regression techniques. To reduce the sensitivity of the estimation to noise, the principal eigenvector method is used. To further improve the accuracy, a priori information on properties of the camera sensitivity function (such as smoothness, positive-valued and uni-modality) is incorporated as linear constraints in the regression.

To obtain high quality imaging also requires repeatability and stability in the acquisition, which is commonly referred to as obtaining precision in color measurements. To ensure this, all components have been calibrated, prior to the work of spectral characterization of the imaging system. The calibration of the image acquisition system is described in Appendix A, which also includes detailed descriptions of all the components in the system along with discussions on the demands placed on each component.
5.2 The Image acquisition system

The image acquisition system is an experimental system with great flexibility for the user and numerous ways to control and alter the image acquisition setup. While it is possible to capture images of arbitrary objects, the primary usage is to capture macro images of flat objects, such as substrates.

5.2.1 Image acquisition setup

The substrate is placed on a table which allows for motor-driven translation in x-y directions and rotation around the optical axis. The illumination is provided using a tungsten halogen lamp through optical fibers, which offers directed light with an adjustable angle of incidence for reflectance images. By using a backlight setup, imaging with transmitting illumination is also possible. The images are captured using a monochrome CCD camera combined with macro optics, consisting of enlarging lenses and various extension rings.

Color images are captured sequentially, using filters mounted in a filter wheel in the optical path. By using this color sequential method, full color information is acquired in every single pixel and there is no need for any interpolation or de-mosaicing, as is the case for conventional digital cameras using color filter arrays. The filter wheel contains 20 color filters, including RGB-filters, CMY-filters and a set of neutral density filters of various optical densities. Furthermore, a set of 7 interference filters allow for the acquisition of multi-channel images. The functionality of all components in the image acquisition system is controlled using a Matlab interface. Figure 5.1 illustrates the image acquisition setup, in the mode of reflective imaging.

5.2.2 Image acquisition model

The linear model describing the image acquisition process, in the case of reflectance imaging, is given in Eq. 5.1. The device response, $d_k$, for the $k$:th channel is, for each pixel, given by:

$$
F_k(\lambda)O(\lambda)QE(\lambda)R(\lambda)E_k(\lambda)d_kS(\lambda)
$$
\[
d_k = \int_{\lambda} I(\lambda)F_k(\lambda)R(\lambda)O(\lambda)QE(\lambda)d\lambda + \varepsilon_k
\]  

(5.1)

where \(I(\lambda)\) is the spectral irradiance of the illumination (including the spectral characteristics of lamp as well as the optical fibers used for light transportation), \(F_k(\lambda)\) is the spectral transmittance of filter \(k\), \(R(\lambda)\) is the spectral reflectance of the object, \(O(\lambda)\) is the spectral transmittance of the optics, \(QE(\lambda)\) is the spectral quantum efficiency for the CCD sensor, \(\varepsilon_k\) is the measurement noise for channel \(k\), and \(V\) is the spectral sensitivity region of the CCD sensor.

The validity of the image acquisition model (Eq. 5.1) requires a linear response of the camera, i.e. a camera response proportional to the energy of the incident light, which has been verified (Appendix A). The amplitude of the camera response value, \(d_k\), will also depend on the aperture size, the exposure time, and on the focal length used. However, this will only affect the total level of incoming radiance to the CCD sensor, i.e. it is not wavelength dependent, and need not be included in the spectral image acquisition model.

The spectral characterization of the illuminant and the color filters is straightforward since it can be can be obtained by direct measurements using a spectroradiometer. Figure 5.2 displays the spectral power distribution for the illumination, and the spectral transmittance for the RGB, CMY and interference filters, respectively. A detailed description of the spectral measurements and the components in the image acquisition system is given in Appendix A.

![Figure 5.2. Spectral power for the illuminant (a), spectral transmittance for the RGB filters (b), the CMY filters (c) and the 7 multi-channel filters (d).](image-url)
5.3 Theory of spectral sensitivity estimation

Deriving the spectral sensitivity function for the camera is not as straightforward as the characterization of the illumination and color filters since it cannot be derived by direct measurements. The manufacturer of the CCD camera and the optics has provided data representing ‘typical’ devices, which may not apply to each individual device. The aim of this section is to examine if a more accurate estimate of the spectral sensitivity can be derived, compared to the manufacturer data. Deriving the spectral sensitivity function is the final step in completing the spectral image acquisition model.

5.3.1 The spectral sensitivity function

For a fixed combination of CCD sensor and optics, it is not necessary to separate the spectral transmittance of the optics from the spectral quantum efficiency of the CCD. They can be represented by the combined spectral sensitivity function, \( S(\lambda) \), as:

\[
S(\lambda) = O(\lambda)QE(\lambda) \tag{5.2}
\]

The irradiance of the illumination together with the spectral transmission of filter \( k \), and the object reflectance gives the incoming radiance to the CCD sensor for the \( k \)-th channel, \( E_k(\lambda) \), as (see Fig. 5.1):

\[
E_k(\lambda) = L(\lambda)F_k(\lambda)R(\lambda) \tag{5.3}
\]

The spectral image acquisition model according to Eq. 5.1 then reduces to:

\[
d_k = \int_{\lambda_1}^{\lambda_N} E_k(\lambda)S(\lambda)d\lambda + \epsilon_k \tag{5.4}
\]

The objective of this chapter is to derive the spectral sensitivity function, \( S(\lambda) \), combing the quantum efficiency for the CCD with the spectral transmittance of the optics. Note that with a monochrome CCD sensor, with the color filters placed along the path of the illumination, the spectral sensitivity estimation involves only deriving a single function, \( S(\lambda) \), instead of three different functions that also include filter characteristics, which is the case for conventional RGB-cameras.

5.3.2 Pseudo-inverse (PI) estimation

Recall the linear model of image acquisition from Eq 5.4. If we represent the spectral power distributions \( E(\lambda) \) from a set of \( T \) color patches as discrete \( N \)-component vectors, \( e_i \), sampled at wavelengths \( \lambda_1, \ldots, \lambda_N \), Eq. 5.4 can be rewritten as:

\[
d_i = e_i^T s + \epsilon \tag{5.5}
\]

where \( d_i \) is the camera response values of the color patches, \( i = 1 \ldots T \), \( e_i \) is the spectral power distribution for color patch \( i \) (including the illuminant, the reflectance of the
sample and any filter in the optical path), \( s \) is the spectral sensitivity of the camera (including the CCD-sensor and the optics) and \( \varepsilon \) is the measurement noise term.

The optimal estimate of the spectral sensitivity function, \( \hat{s}_\text{opt} \), can then be obtained by minimizing the mean square error metric of the linear fit to a set of samples, as:

\[
\hat{s}_\text{opt} = \arg \min_s \left\{ \frac{1}{T} \sum_{i=1}^{T} \left\| \mathbf{e}_i \mathbf{s} - \mathbf{d}_i \right\|^2 \right\}
\]  

(5.6)

If the spectral data of the samples are collected into the \( N \times T \) matrix \( \mathbf{E} \), with the columns corresponding to spectral measurements of each color sample, and the camera response values collected into the \( T \)-component vector \( \mathbf{d} \), then the linear relationship is given by:

\[
\mathbf{d} = \mathbf{E}\mathbf{s} + \varepsilon
\]  

(5.7)

and the optimal spectral sensitivity estimation, \( \hat{s} \), is given by:

\[
\hat{s} = (\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{d}
\]  

(5.8)

where \( (\mathbf{E}'\mathbf{E})^{-1} \) denotes the Moore-Penrose pseudo-inverse of \( \mathbf{E}' \). In the absence of noise and with rank(\( \mathbf{E} \)) \( \geq N \) (i.e. with \( T \geq N \)), the solution is accurate up to the working precision. Under real conditions, however, the assumption of noise-free image acquisition is unrealistic and the pseudo-inverse solution without additional constraints is not often useful (Hardeberg, 2001).

### 5.3.3 Principal Eigenvector (PE) method

The problem with the pseudo-inverse solution is that, although the spectral data in \( \mathbf{E}' \) is \( N \)-dimensional, the effective dimension of real reflectance spectra (i.e. the number of components needed to describe a spectrum) is usually significantly less. This means that \( \mathbf{E}' \) contains only a small number of significant eigenvalues, resulting in an unstable and noise-sensitive estimation of \( \hat{s} \) (Bala, 2003). With the Principal Eigenvector (PE) method (also known as the rank-deficient pseudo-inverse), the noise sensitivity of the system inversion is reduced, by including only the singular vectors corresponding to the most significant singular values in the spectral autocorrelation function of the training samples (Hardeberg, 2001).

If a Singular Value Decomposition (SVD) is applied to the matrix \( \mathbf{E}' \) containing the spectral data of the test patches then:

\[
\mathbf{E}' = \mathbf{U}\mathbf{W}\mathbf{V}'
\]  

(5.9)

where \( \mathbf{U} \) and \( \mathbf{V} \) are orthonormal matrices and \( \mathbf{W} \) is a \( T \times N \) diagonal matrix of rank \( R \) with the diagonal elements \( w_i \), \( i = 1 \ldots R \), corresponding to the singular values of \( \mathbf{E}' \) (all positive and monotonically decreasing). Since the singular values of a matrix of spectral reflectance data, such as \( \mathbf{E} \), decrease rapidly with \( i \), the reflectance spectra can be
represented by a smaller number of parameters. By only taking into account the first $r < R$ singular values, the spectral sensitivity can be approximated by:

$$\hat{s} = VW^{(r)^{-1}}U^t d$$  \hspace{1cm} (5.10)

where:

$$W^{(r)^{-}} = \begin{cases} w_i^{-1}, & \text{for } i \leq r \\ 0, & \text{otherwise} \end{cases}$$ \hspace{1cm} (5.11)

The choice of $r$, the number of principal eigenvectors to be included in the inversion, controls the degree of freedom and should be adjusted depending on the noise level. With too few eigenvectors included, the degree of freedom will be too limited to fit the real sensitivity curve; while with too many, the noise will be dominant in the inversion. (Hardeberg, 2001)

An alternative formulation is to define a tolerance value, $\alpha$, and include only the principal eigenvectors with singular values that are bigger than this constant, when normalized by the first singular value. $W^{(\alpha)^{-}}$ in Equation 5.10 is replaced by $W^{(\alpha)^{0}}$, where:

$$W^{(\alpha)^{-}} = \begin{cases} w_i^{-1}, & \text{for } w_i / w_1 > \alpha \\ 0, & \text{otherwise} \end{cases}$$ \hspace{1cm} (5.12)

For a fixed number of $\alpha$ it is desirable that the reflectance data contains as many singular values as possible which can be included in the inversion. A smaller $\alpha$ allows for a greater number of principal eigenvectors being included, but will also amplify the noise in the estimation, since estimation variance grows inversely with the singular values squared. (DiCarlo, et al., 2004)

5.3.4 Additional constraints

One common drawback with the unconstrained PI and PE methods in the case of spectral sensitivity estimation is that a priori information of the known nature of the spectral sensitivity function is not exploited. For example, the spectral sensitivity must be a positive-valued function which is usually uni-modal and smooth. These assumptions can be incorporated into the regression as a set of linear constraints, which can help to improve the accuracy of the estimate (Finlayson & Hubel, 1998).

**Positivity**

The spectral sensitivity function must be positive at all wavelengths, since no real sensor can give a negative response to a stimulus. To ensure a positive sensitivity function estimate, the linear constraint:

$$\hat{s}(\lambda_n) \geq 0, \text{ for } n = 1...N$$ \hspace{1cm} (5.13)
is used in the regression.

**Smoothness**

When the function only contains the detector sensitivity (i.e. the illuminant is not included) it is generally a smooth function of wavelength. This makes it possible to represent the sensitivity function as linear combinations of a set of basis functions, such as the Fourier basis. We represent the spectral sensitivity function as linear combinations of sine and cosine functions, as:

\[
\hat{s} = \sigma_1 B_1 + \sigma_2 B_2 + \ldots + \sigma_l B_l
\]  

(5.14)

where \( B_1 = k \), \( B_2 = \sin(x) \), \( B_3 = \cos(x) \), \( B_4 = \sin(2x) \), \ldots; \( l \) is the number of basis functions, controlling how smooth the estimated spectral sensitivity function will be, \( x = (\lambda - \lambda_{min}) \pi / c \), and \( c \) is the cycle length used.

By collecting the Fourier basis functions into the \( N \times l \) matrix, \( B \), the spectral sensitivity estimate is now expressed as:

\[
\hat{s} = B \sigma
\]  

(5.15)

Then the objective is to find the \( l \)-component vector of weights, \( \sigma \), which minimizes the objective function:

\[
\| \mathbf{e}' B \sigma - \mathbf{d} \|^2
\]  

(5.16)

An alternative approach to ensure a smooth spectral sensitivity function is to add a regularization term. As suggested by Barnard and Funt (2002) it is better to promote smoothness by adding a regularization term, since the Fourier basis puts constraints on the sensor function, which are not necessary for simple smoothness with the result that many good candidates for the sensor function are excluded. With the regularization term added, the objective function becomes:

\[
\hat{s}_{\text{reg}} = \arg \min \left\{ \frac{1}{T} \sum_{i=1}^{T} \left\| \mathbf{e}' i \hat{s} - \mathbf{d} \right\|^2 + \gamma \sum_{i=1}^{N} \| \mathbf{Q} \hat{s} \|^2 \right\}
\]  

(5.17)

where \( \mathbf{Q} \) is an \( N \times N \) matrix, which provides an approximation of the second derivative:

\[
\mathbf{Q} = \begin{bmatrix}
-1 & 2 & -1 \\
-1 & 2 & -1 \\
& & \ddots & \ddots & \ddots \\
& & & -1 & 2 & -1 \\
& & & & -1 & 2 & -1
\end{bmatrix}
\]  

(5.18)
The first term of Eq. 5.17 expresses the RMS error and the second term the non-smoothness of the sensitivity function, thus providing a regularization term. The coefficient $\gamma$ specifies the relative weight of the regularization term and is found by trial and error. (Barnard & Funt, 2002)

**Modality**

The modality refers to the number of peaks in the spectral sensitivity function. Many sensor functions appear to be uni-modal, i.e. having only one single peak, a constraint expressed as:

$$s(\lambda_{m+1}) \geq s(\lambda_n), \quad n = 1, \ldots, m - 1$$

$$s(\lambda_{m+1}) \leq s(\lambda_n), \quad n = m, \ldots, N$$

(5.19)

The method requires trying all possible peak locations, $m$, and choosing the ones that produces the least error. If some assumptions can be made regarding the peak locations, only the “plausible” combinations of peak positions need to be included. If the uni-modality assumption does not hold, $n$-modality can be used as a set of similar linear constraints. (Finlayson & Hubel, 1998)

### 5.3.5 Alternative objective function

An alternative to the approach of minimizing the absolute RMS error, using Eq. 5.6, is to instead minimize the relative RMS error. The motivation is that the variation of the pixels increases with increasing magnitude and that minimizing the relative error better reduces the error in chromaticity (Barnard & Funt, 2002). To minimize the relative error, the objective function is replaced by:

$$s_{opt} = \arg \min_s \left\{ \frac{1}{T} \sum_{i=1}^{T} \left[ e_i \left( \frac{s_{opt} - d_i}{d_i} \right)^2 \right] \right\}$$

(5.20)

The minimization of the relative error may need some modifications to deal with very small values $d_i$ which are likely to be inaccurate. Since Eq. 5.20 can be considered as a weighted version of Equation 5.6 with $(1/d)^2$ as the weighting factor, it is easy to put an upper bound on this weighting, to safeguard against very small values of $d_i$. (Barnard & Funt, 2002)

### 5.4 Color targets

The selection of color patches for the test target plays an important role in the spectral sensitivity estimation. It is desirable to obtain reliable results without having to perform time-consuming measurements on very large numbers of target patches. A method for selecting those reflective targets which are most significant for the estimation of the spectral sensitivity is proposed by Hardeberg (2001):
Starting with the full set of available reflective spectra \( r_p, p = 1 \ldots P \), the first is to select \( r_{s_1} \), with the maximal RMS value, as:

\[
\| r_{s_1} \| \geq \| r_p \| \quad \text{for } p = 1 \ldots P
\]  

(5.21)

Next select \( r_{s_2} \) that minimizes the condition number, i.e. the ratio of the largest singular value to the smallest singular value. Denoting the minimum and maximum singular values of a matrix \( X \) as \( w_{\min}(X) \) and \( w_{\max}(X) \), the minimization of the condition number is expressed by:

\[
\frac{w_{\max}\left(\left\{ r_{s_1}, r_{s_2} \right\}\right)}{w_{\min}\left(\left\{ r_{s_1}, r_{s_2} \right\}\right)} \leq \frac{w_{\max}\left(\left\{ r_{s_1}, r_{s_2}, \ldots, r_{s_i} \right\}\right)}{w_{\min}\left(\left\{ r_{s_1}, r_{s_2}, \ldots, r_{s_i} \right\}\right)}, \quad \text{for } p = 1 \ldots P, \ p \neq s_i
\]  

(5.22)

Further patches are selected by the same rule, i.e. select sample \( r_{s_i} \) as:

\[
\frac{w_{\max}\left(\left\{ r_{s_1}, r_{s_2}, \ldots, r_{s_i} \right\}\right)}{w_{\min}\left(\left\{ r_{s_1}, r_{s_2}, \ldots, r_{s_i} \right\}\right)} \leq \frac{w_{\max}\left(\left\{ r_{s_1}, r_{s_2}, \ldots, r_{s_{i-1}}, r_{s_i}, r_{s_{i+1}}, \ldots, r_p \right\}\right)}{w_{\min}\left(\left\{ r_{s_1}, r_{s_2}, \ldots, r_{s_{i-1}}, r_{s_i}, r_{s_{i+1}}, \ldots, r_p \right\}\right)}, \quad \text{for } p = 1 \ldots P, \ p \neq s_1, s_2, \ldots, s_{i-1}
\]  

(5.23)

The motivation behind the method is that for each iteration step, the reflectance target that is most different from the other targets is selected, thus obtaining the most significant targets for the spectral sensitivity estimation. It has been shown that for spectral sensitivity estimation, a set of 20 colors from the Munsell atlas selected using this method was comparable to that using the complete set of 1269 Munsell colors (Hardeberg, 2001).

### 5.4.1 NCS colors

Available reference colors to be used as test target were a set of color patches from the Natural Colour System, NCS (Hård, 1995). The samples have the size of 150 × 105 mm and have been provided by the Scandinavian Colour Centre. The color patches are created using different pigments, making them a better choice than printed color patches in which the halftone pattern will be visible in the images due to the large magnification of the optics. The reason for not using a standard target, such as the Macbeth ColorChecker chart or the Kodak Q60 target, is that the measurement field and minimum distance for the spectroradiometer, make it difficult to measure the patches of these small dimensions.

50 colors were selected from the 365 available colors from the following pages in the NCS Color atlas 96: B, B50G, G, G50Y, Y, Y50R, R and R50B. The colors have been selected according to Eqs. 5.21 - 5.23, using the spectral reflectance data available from NCS. Since all measurements will be made using the same illuminant, the 25 most spectrally varying colors under the illuminant of the system were selected, based on spectral measurements. The 25 most significant color patches used for the spectral sensitivity estimation are listed in Nyström (2006). The remaining colors are used for evaluating the estimated spectral sensitivity function.
To investigate the dimensionality of the 25 selected color patches, singular-value decomposition is performed on the matrix $\mathbf{E}$ containing the spectral data under the illumination of the system. Figure 5.3 displays the relative strength of the 25 first singular values of $\mathbf{E}$, normalized by the first singular value. The singular values decrease rapidly due to correlation between color patches. When compared to the reference tolerance value $\alpha = 0.1$ only 3 of the singular values pass the limit when no color filter is used. A comparison with other reflective charts commonly used for calibration shows that both the Macbeth ColorChecker (24 patches) and the Macbeth DC chart (over 200 patches) also give only 3 singular values that are greater than $\alpha = 0.1$ (DiCarlo, et al., 2004). Taking these into consideration, the effective dimension of the test targets must be considered as satisfying. It could, nevertheless, be improved if more color samples had been available. Since all available colors come from only 9 different pages of the NCS Color atlas, correlation between spectral reflectance is inevitable.

The image acquisition system has the advantage of having the color filters placed along the optical path of the light source, providing the possibility to capture the test target under different illuminations. When 14 different color channels are used to illuminate the test target, the relative strength of the singular values increases and there are now 5 singular values that exceed the threshold, $\alpha = 0.1$. The 14 channels used correspond to: no filter, RGB-filters, CMY-filters and the 7 interference filters, i.e. all available filters except for the neutral density filters.

Figure 5.3. The 25 first singular values from the 25 NCS reflectance spectra, using a logarithmic scale.
5.5 Experimental setup

The spectral sensitivity estimation requires spectral measurements of the reference patches and recording of the corresponding camera response values. All conditions for the spectral measurements and the image acquisition should be identical.

5.5.1 Spectral measurements

The spectral measurements of the NCS-patches are performed using a spectroradiometer, mounted in the same optical axis as the CCD-camera, using the 45°/0° measurement geometry. For each color patch, the mean reflectance spectrum from 5 sequential measurements is computed. The spectral data are in the interval 380 to 780 nm sampled at 4 nm intervals, giving \( N = 101 \).

The reflectance measurements are performed using the illuminant only, without any filter. From the measured reflectance of the NCS patches, the resulting spectral data \( E_k(\lambda) \) for channel \( k = 2, \ldots, 14 \) is then computed using the known spectral transmittance for each filter (Fig. 5.2). The 14 channels correspond to: no filter, the RGB-filters, CMY-filters and the 7 interference filters. The spectral data of the 25 reference color patches, using all 14 color-channels, is collected into the 101×350 matrix \( E \), with each column containing the spectral data of color patch \( p \), filter \( k \).

5.5.2 Image acquisition

For each color target, 14-chanel images have been captured, using the same 14 color filters as for the spectral computations, and the same 45°/0° measurement geometry. The mean is computed from 5 sequential images, captured at slightly different locations, by moving the xy-table on which the image is placed. Because of the high magnification of the optics, the images have been captured slightly out of focus, to prevent small irregularities in the surface to influence the result. The exposure times used have been individually calibrated for each channel.

5.5.3 Processing the images

As a pre-processing step, all images are first corrected for dark current and electronic gain, as:

\[
I_{ck}(x, y) = a_k \frac{I_k(x, y) - Dc_k(x, y)}{W_k(x, y) - Dc_k(x, y)}
\]

(5.24)

where \( Dc_k(x, y) \) refers to dark current images, captured with the illumination turned off and the lens cap on, \( W_k(x, y) \) refers to white images captured using a reference white under uniform illumination, and \( a_k \) is a channel-dependent normalization factor which can be chosen to give the corrected and normalized image, \( I_{ck}(x, y) \), desired maximal values (Hardeberg, 2001). A more detailed description of the characteristics of dark current and electronic gain, and how to correct for the effects, is given in Appendix A.
The channel-dependent normalization factors, $a_k$, used in Eq. 5.24 are chosen as:

$$a_k = \frac{E_{t_1}}{E_{t_k}}$$  \hspace{1cm} (5.25)

where $E_{t_1}$ is the calibrated exposure time for channel 1 (i.e. the illuminant without any filter) and $E_{t_k}$ is the exposure time used for channel $k$. In this way, the camera response values are adjusted according to the channel-dependent exposure times, and normalized against the mean camera response for the white reflectance standard.

From the corrected and normalized images $I_{ck}(x,y)$, the mean values are computed for each channel $k$ and used as the camera response value, $d_{k,p}$ for NCS colors $p = 1, \ldots, 25$. When computing the mean values it has been checked that no pixel value deviates more than 5% from the mean, to prevent any defective pixels from affecting the results. The resulting camera response values $d_{k,p}$ are collected into the 350-component vector $\mathbf{d}$, used for the spectral sensitivity estimation.

## 5.6 Results

Since the real spectral sensitivity function is unknown, the estimated sensitivity is evaluated by using the forward characterization model to predict device response values, $\hat{\mathbf{d}}$, from the known spectral data, as:

$$\hat{\mathbf{d}} = \mathbf{E}\hat{s}$$  \hspace{1cm} (5.26)

where $\mathbf{E}$ is the $N \times P$ matrix with the columns corresponding to spectral measurements of each color sample, and $\hat{s}$ the estimated spectral sensitivity.

For the evaluation, the remaining NCS color patches have been used, predicting the device response for all 14 color channels. The predicted response values, $\hat{\mathbf{d}}$, are then compared with the measured device responses, $\mathbf{d}$, to compute the following error metrics:

$$\text{RMS error} = \frac{1}{P} \sum_{i=1}^{P} \left\| \hat{d}_i - d_i \right\|^2$$  \hspace{1cm} (5.27)

and the

$$\text{Relative RMS error} = \frac{1}{P} \sum_{i=1}^{P} \frac{\left\| \hat{d}_i - d_i \right\|^2}{d_i}$$  \hspace{1cm} (5.28)

### 5.6.1 Manufacturer data

The manufacturers of the optics and CCD camera have provided data representing ‘typical’ curves for the CCD quantum efficiency and spectral transmittance of the optics. The data is available only in graphical form and the resolution is limited. To evaluate the accuracy of the data provided by the manufacturers, data has been collected
manually from the graphs, and then cubically interpolated to the spectral range 380:4:780 nm. The data for the optics and CCD have been combined together into the camera sensitivity function according to Eq. 5.2. The curves provided by the manufacturers of the optics and the CCD camera are depicted in Nyström (2006).

The estimation errors when using the manufacturer data to predict the camera response values from real spectral data are relatively small. Clearly the data provided by the manufacturer is fairly accurate, with typical curves that better represent the actual devices than for consumer products, where the variance between the individual devices is typically large (DiCarlo, et al., 2004). Considering the good quality of the data provided by the manufacturers, the assumption was then made that the camera sensitivity function should have a similar form, i.e. it is most likely both uni-modal and smooth. The sensitivity function derived from the manufacturer data is used as a reference to see if it is possible to achieve a more accurate estimate of the spectral sensitivity function.

5.6.2 PI and PE solutions

The unconstrained pseudo-inverse (PI) solutions have been derived using absolute and relative RMS as objective function, according to Eqs. 5.6 and 5.20. As expected, the resulting estimations are very noisy. Even though providing low RMS-errors, their magnitude, negative values and non-smoothness make them practically useless as an approximation of the true camera sensitivity function.

The resulting spectral sensitivity function has also been estimated, using the principle eigenvector (PE) method for the first 1 to 9 principle eigenvectors included in the inversion. The smoothness of the estimated function decreases with increasing number of PEs due to the increasing sensitivity to noise. The estimates are neither uni-modal nor generally smooth enough to be considered as realistic candidates. They are however, all positive, without explicitly using positivity as a constraint. The spectral sensitivity estimations and the resulting prediction errors for the unconstrained PI and PE methods are given in Nyström, 2006.

5.6.3 Constrained solutions

To ensure smooth camera sensitivity functions, the estimations are represented as linear combinations of Fourier basis, according to Eq. 5.14, and to ensure positivity, the constraint according to Equation 5.13 is used. Both the number of basis and the cycle length used are allowed to vary, which gives a greater degree of freedom for the sensitivity estimation compared to previous works where the cycle length is fixed and equal to the wavelength interval used (Barnard & Funt, 2002; Finlayson & Hubel, 1998). The number of base-functions, \( l \), is varied between 1:15, and the cycle length, \( c \), is 100:10:490 nm, giving in total 600 possible estimations for each different number of PEs included.

To select the optimal estimates from all the candidates derived, selection criteria are needed. The best results were found by combining the two error metrics, using the absolute RMS error as the objective function in the regression, and the relative RMS error as criterion for selection. Furthermore, the uni-modality constraint was used as a criterion for selection. It was found that the uni-modality worked better as a criterion for
selection compared to when it was incorporated as a constraint in the regression, which produced unrealistic, flat curves. It seems with this approach that the number of principal eigenvectors included is of little importance.

As an alternative to ensuring smoothness by using the Fourier basis, the method of adding a regularization term to the objective function according to Eqs. 5.17 - 5.18, has been implemented. Barnard and Funt (2002) claimed that the value of the coefficient $\gamma$ is “easily found by trial and error”. However, it is clear that the RMS error will increase with increasing weight, $\gamma$, in the objective function. Hence, there must be a manual trade-off between low RMS-error and the acceptable smoothness.

The same approach as for the Fourier basis is adapted, i.e. using the absolute RMS as objective and searching for the lowest relative RMS-error. The number of eigenvectors included in the inversion has a greater impact on the result, compared to the Fourier basis approach. The RMS errors decreases with an increasing number of PE included, but so does the non-smoothness. The best result for the candidates that are smooth and uni-modal is found when PE = 3 and $\gamma = 30$. The spectral sensitivity estimates for different numbers of PEs and the resulting prediction errors are given in Nyström (2006) for the Fourier basis approach, as well as using the regularization term.

Table 5.1 lists the predicted errors of the spectral sensitivity estimations considered as the best candidates, using the Fourier basis and regularization term, respectively. For comparison, the results for the function derived from manufacturer data are included. The corresponding sensitivity functions are displayed in Fig. 5.4.

Table 5.1. The resulting prediction errors for the best estimates of the camera sensitivity function compared to the data provided by the manufacturer

<table>
<thead>
<tr>
<th></th>
<th>Manufacturer data</th>
<th>Fourier basis</th>
<th>Regularization term</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS-error</td>
<td>$5.45 \times 10^{-4}$</td>
<td>$2.26 \times 10^{-4}$</td>
<td>$2.33 \times 10^{-4}$</td>
</tr>
<tr>
<td>Relative RMS-error</td>
<td>0.038</td>
<td>0.0149</td>
<td>0.0184</td>
</tr>
<tr>
<td>Mean error</td>
<td>0.012</td>
<td>0.0070</td>
<td>0.0072</td>
</tr>
<tr>
<td>Max error</td>
<td>0.13</td>
<td>0.0965</td>
<td>0.0972</td>
</tr>
</tbody>
</table>
Figure 5.4. The best estimates of the camera sensitivity function compared to the estimate derived from data provided by the manufacturer. Clearly, both the estimated sensitivity functions show resemblance to the function derived from manufacturer data, being uni-modal and smooth. Table 5.1 reveals that both the estimated functions outperform the manufacturer data in terms of all the error metrics analyzed. The best choice is the estimation based on Fourier basis, where both absolute and relative RMS errors decreases by more than half compared to manufacturer data. The maximal error for the predicted device response to the validation colors, using 14 color-channels, is reduced to 0.097 (images normalized in the range [0,1]) and the mean is as low as 0.007.

5.7 Summary and discussion

The spectral sensitivity function, including the spectral transmittance of the optics and the quantum efficiency of the CCD, has been estimated using a set of 25 carefully selected NCS colors. By relating the measured reflectance spectra of the reference colors to the corresponding device responses, the sensitivity function was derived using least squares regression techniques. The principle eigenvector method was used to reduce the noise sensitivity, and a priori knowledge on the nature of camera sensitivity, such as positive valued, smoothness and uni-modality, was used as constraints in the regression.

The estimated camera sensitivity function considered as the best candidate was the result from minimizing the absolute RMS error, using the PE-method with positivity constraint and Fourier basis to ensure smoothness. Of the large number of estimations, the best candidate was selected using the relative RMS error and the uni-modality requirement as selection criterion. The selected sensitivity function is similar to the function derived from the manufacturer data, but reduces the prediction errors by more than a half. When the estimated sensitivity function was used to predict the response to a validation set of colors patches, independently of the ones used in the estimation, the
mean prediction error was 0.007 and the maxim 0.097 (camera values in the range [0,1]).

Besides the method employed in this study, there are other approaches to deriving the spectral sensitivity function, using either physical measurements or estimation techniques. A monochromator provides a light source with a narrow-band spectral power centered at tunable wavelengths, which can be used to accurately record the device response (see e.g. Wu & Allebach, 2000; MacDonald & Wei, 2002). Hawkins (2008) presented a prototype filter array, consisting of 36 narrow-band filters illuminated by LEDs, allowing for spectral camera characterization through a single exposure.

Another approach for sensitivity estimation uses a set theoretical formulation to express the constraints as convex sets and use the iterative technique projection onto convex sets (POCS) to generate the sensitivity function (see e.g. Sharma & Trussel, 1996; Alsam & Andersen, 2008). Alsam & Lenz (2006) proposed a method based on metameric blacks, to estimate the null-space of the sensor and derive projection operators to solve for the sensitivity function.

There have also been more recent proposals on methods for optimal color sample selection, for camera characterization (Cheung & Westland, 2006; Steder, et al., 2008). However, since these methods generally aim at maximizing the color differences between color samples in CIELAB color space, their main use should be in empirical characterization, relating device values to colorimetric data. When estimating the spectral sensitivity function, the method adopted in this study (Hardeberg, 2001), should have an advantage, since it maximizes the spectral differences between the color targets.

The spectral sensitivity function that has been estimated will be used in the following chapter for reconstructing spectral data from device response. It was the last component of the spectral image acquisition model, which is now complete.
Chapter 6

Model-based spectral reconstruction

6.1 Introduction
6.2 Theory of spectral reconstruction
6.3 Experimental setup
6.4 Metrics for spectral comparison
6.5 Experimental results
6.6 Summary and discussion
6.1 Introduction

After deriving the spectral characteristics of the illumination and the color filters, and estimating the spectral sensitivity function for the camera in Chapter 5, the spectral image acquisition model is complete. The model describes the forward characterization function predicting the device response to a known spectral input. However, the inverse characterization function is often required, deriving device-independent color data for an object using the recorded device response. In colorimetric imaging, the device-independent representations are typically CIEXYZ or CIELAB, and in multispectral imaging estimations of the spectral reflectance are derived.

The focus in this chapter is multispectral imaging, using model-based characterization. The aim is to reconstruct the spectral reflectance of objects from the recorded device response, by inverting the spectral model of the image acquisition system. As the physical representation of color of an object, spectral reflectance is completely independent of the characteristics of the image acquisition system. From the spectral reflectance data it is straightforward to convert the data to colorimetric representations, as well as to render the image under arbitrary illumination.

Compared to empirical characterization, model-based characterization is more general and allow for applications to arbitrary combinations of media and colorants. The accuracy of the spectral reconstructions depends on the noise levels, and on how well the spectral model reflects the real characteristics of the image acquisition system, i.e. it is not dependent on a training set of color samples.

Spectral reconstruction is performed using both trichromatic (RGB) and multi-channel images, to evaluate the difference in performance. A priori knowledge on the smooth nature of spectral reflectances is exploited by representing the spectra as linear combinations of different basis functions. The metrics used to evaluate the results include the spectral RMS difference as well as the Euclidian distance in CIEXYZ color space, and the CIE 1976 and 1994 color differences in CIELAB.

6.2 Theory of spectral reconstruction

Recall the forward spectral image acquisition model describing the response of the device by modeling the physical characteristics of all the components involved in the
image acquisition process (repeated in Eq. 6.1). The device response, \( d_k \), for the \( k \):th channel is, for each pixel, given by:

\[
d_k = \int_{\lambda \in \mathcal{V}} I(\lambda) F_k(\lambda) R(\lambda) S(\lambda) d\lambda + e_k
\]  

(6.1)

Having obtained the spectral characteristics of all the components, the characteristics of the image acquisition system can be represented by a spectral transfer function (Farell et al, 1999). The spectral transfer function, \( W_k(\lambda) \), describes the spectral characteristics for each channel \( k \), including the illuminant, the camera sensitivity and the filter transmittance, as:

\[
W_k(\lambda) = I(\lambda) F_k(\lambda) S(\lambda)
\]

(6.2)

Denote the spectral signal as a discrete \( N \)-component vector, sampled at wavelengths \( \lambda_1, \ldots, \lambda_N \), and let \( \mathbf{W} \) be the \( N \times K \) matrix in which each column describes the spectral transfer function of channel \( k \). Then the camera response vector, \( \mathbf{d} \), for a sample with spectral reflectance \( \mathbf{r} \) is given by:

\[
\mathbf{d} = \mathbf{W} \mathbf{r}
\]

(6.3)

When inverting the model, we seek the \( N \times K \) reconstruction matrix \( \mathbf{M} \) that best reconstructs the spectral reflectance, \( \tilde{\mathbf{r}} \), from the camera response \( \mathbf{d} \), as:

\[
\tilde{\mathbf{r}} = \mathbf{M} \mathbf{d}
\]

(6.4)

6.2.1 Pseudo-inverse solution

The most straightforward approach to derive the reconstruction matrix is to simply invert Eq. 6.4, using the pseudo-inverse approach. This gives the reconstruction operator:

\[
\tilde{\mathbf{r}} = (\mathbf{W}^\dagger) \mathbf{d}
\]

(6.5)

where \((\mathbf{W}^\dagger)^{\dagger}\) denotes for the More-Penrose pseudo-inverse of \( \mathbf{W}^\dagger \). Thus, the pseudo-inverse reconstruction operator is given by:

\[
\mathbf{M}_0 = (\mathbf{W}^\dagger)^{\dagger}
\]

(6.7)

Generally, the pseudo-inverse reconstruction is sensitive to noise, which makes the approach not always useful in practice. When \( K < N \), i.e. the number of color channels \( K \) is less than the number of spectral sampling points \( N \), the matrix \( \mathbf{W} \) is of insufficient rank and the algebraic equations are underdetermined. Furthermore, this method minimizes the Euclidian distance in the camera response domain (i.e. between \( \mathbf{d} \) and \( \mathbf{W}^\dagger \tilde{\mathbf{r}} \)), which does not necessarily mean that the reconstructed spectrum will be close to the real spectrum. (Hardeberg, 2001)
6.2.2 Basis functions

Another approach is to instead seek another reconstruction matrix, $M_1$, which minimizes the Euclidian distance between the reconstructed spectrum and the original spectrum (Hardeberg, 2001). The vast majority of reflectance spectra for real and man-made surfaces are smooth functions of wavelength (Maloney, 1986; Connah, et al., 2001). By exploiting this \textit{a priori} information, we can assume that the reflectance can be represented by a linear combination of a set of smooth basis functions, $B = \left[ b_1, b_2, \ldots, b_p \right]$. With this assumption, the reflectance can be represented using the vector of coefficients, $a$, as:

$$ r = Ba $$

Hence, $\tilde{r}$ can be obtained from $a$ (using Eqs. 6.3, 6.4, and 6.8), as:

$$ \tilde{r} = M_1 d = M_1 W'r = M_1 W'Ba $$

Then, the ideal expression $r = \tilde{r}$, becomes:

$$ M_1W'Ba = Ba $$

(6.10)

Assuming the basis $B$ represents a statistically significant representation of the reflectances likely to be encountered, Eq. 6.10 should be true for any $a$, and hence:

$$ M_1W'B = B $$

(6.11)

This gives the reconstruction operator $M_1$, minimizing the RMS spectral error of the reconstructed spectrum (Hardeberg, 2001).

$$ M_1 = BB'W(W'BB'W)^{-1} $$

(6.12)

The base functions, $B$, can consist of a set of real, measured spectral reflectances, which then should be representatives to the reflectance of samples that are likely to be encountered in the image acquisition system. To reduce the noise sensitivity, the principle eigenvector method can be applied to the set of spectral basis, as described in Sec. 5.3.3. An alternative to spectral basis is to simply let $B$ be a set of Fourier basis functions.

6.3 Experimental setup

In multispectral imaging, each pixel should contain the spectral information about the scene, combining the strengths of spectral and spatial resolution. However, when experimenting on spectral reconstruction, it is necessary to evaluate the results by comparing to spectral measurements. Hence, the comparison must be made using color samples which have a constant color over some area, which can be measured by a spectroradiometer or spectrophotometer.
For reference color samples to evaluate the results, 25 color patches from NCS are used. The spectral reflectance values of the patches have been measured using a spectroradiometer. Correspondingly, the camera response values have been acquired under identical conditions, using the 45°/0° geometry. The setup for the spectral measurements and the image acquisition is identical to Sec. 5.5, and so is the processing of the captured images.

### 6.4 Metrics for spectral comparison

To evaluate the quality of the reconstructed spectral reflectance, metrics are needed to compute the closeness to the real spectra. Depending on the shapes and magnitudes of the spectral reflectance curves, different metrics can result in very different results (Imai et al., 2000). The spectral RMS error corresponds to the Euclidian distance, in spectral reflectance space, between the original and the reconstructed spectra. It is simple, widely used and well suited for comparison of physical stimuli, but it does not correlate to the perceived color difference. Even though the CIE color difference formulas described in Sec. 2.2.7, $\Delta E_{ab}$ and $\Delta E_{94}$, are not directly designed for evaluating spectral matches, they provide a measure of the perceived color difference under a given illuminant, based on characteristics of the human visual system. Their main drawback, however, is that they provide poor correlation to the spectral similarity for metameric colors. Other metrics include the metamerism indices which compare to which extent two spectra differs under different illuminants and observers (Sec. 2.2.5).

For a thorough survey and comparison of different metrics used for spectral estimations, we refer to Imai et al. (2002). According to Imai there is no metric that suits for all purposes and combinations of different metrics should be used to take advantage of each metric.

We choose to present the RMS spectral reconstruction error together with the Euclidian difference in CIEXYZ color space, $\Delta XYZ$. Furthermore, the CIE color differences $\Delta E_{ab}$ and $\Delta E_{94}$ are computed, under the CIE standard illuminant D65. For each metric, the maximum and mean values of the reference samples are presented, as well as the 95th percentile, i.e. the value below which 95% of the samples lie.

### 6.5 Experimental results

Experiments on inverting the spectral model to reconstruct spectral reflectance have been performed for the NCS color patches. Spectral reflectance data is reconstructed using the operators $M_0$ and $M_1$ according to Eqs. 6.7 and 6.12, using both multi-channel images (from the 7 interference filters) and trichromatic (RGB) images. Concerning the basis functions, $B$, both Fourier basis and spectral basis are evaluated. With the Fourier basis, the best results were obtained using 5 base functions, which is used in all of the following results.

For the spectral basis, the database provided by NCS, containing spectral reflectance of 1750 different color samples are used. The spectral reflectance data available from NCS is in the range 400 to 700 nm. For the results to be directly comparable, the spectral data has been cropped to the range 400 to 700 nm even for the PI-method and Fourier basis. It is worth noting that due to the very low energy available...
from the illuminant at both ends of the visible spectrum, the data outside this interval is noisy and the reconstruction error decreases by excluding data outside this range. To reduce the sensitivity to noise, the principle eigenvector (PE) method has been used, as described in Sec. 5.2.3. The best results were obtained when representing the spectral basis by the singular vectors corresponding to the 5 most significant singular values.

The RMS spectral reconstruction errors for the evaluations patches are given in Table 6.1, together with their Euclidian distances in the CIEXYZ color space. The CIELAB differences $\Delta E_{ab}$ and $\Delta E_{94}$ are collected in Table 6.2. The corresponding error metrics using standard illuminant A are listed in Nyström (2006).

**Table 6.1. Spectral reconstruction errors, expressed as RMS difference in spectral space, and the Euclidian distance in CIEXYZ space, $\Delta XYZ$.**

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>RMS</th>
<th>$\Delta XYZ$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Mean 95%</td>
</tr>
<tr>
<td>RGB</td>
<td>PI</td>
<td>0.0706</td>
<td>0.0230 0.0598</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td><strong>0.0041</strong></td>
<td><strong>0.0014 0.0040</strong></td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>0.0155</td>
<td>0.0049 0.0141</td>
</tr>
<tr>
<td>Multi</td>
<td>PI</td>
<td>0.0092</td>
<td>0.0030 0.0072</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td><strong>0.0039</strong></td>
<td><strong>0.0012 0.0030</strong></td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>0.0040</td>
<td><strong>0.0011 0.0027</strong></td>
</tr>
</tbody>
</table>

**Table 6.2. Colorimetric reconstruction errors. CIE 1976 color difference, $\Delta E_{ab}$ and CIE 1994 color difference $\Delta E_{94}$.**

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>$\Delta E_{ab}$</th>
<th>$\Delta E_{94}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Mean 95%</td>
</tr>
<tr>
<td>RGB</td>
<td>PI</td>
<td>24.3 14.8 22.0</td>
<td>17.0 9.55 16.0</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td><strong>15.8 4.17 10.3</strong></td>
<td><strong>6.20 2.72 6.17</strong></td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>18.7 8.28 18.6</td>
<td>10.9 5.07 10.5</td>
</tr>
<tr>
<td>Multi</td>
<td>PI</td>
<td>4.36 1.52 3.14</td>
<td><strong>1.91 0.97 1.86</strong></td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td><strong>4.21 1.81 3.30</strong></td>
<td>2.05 1.13 1.98</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>7.27 2.11 5.78</td>
<td><strong>2.81 1.21 2.73</strong></td>
</tr>
</tbody>
</table>

Figure 6.1 displays the reconstructed spectral reflectance from RGB-images, using three NCS colors as examples. The corresponding results using multi-channel images are given in Fig. 6.2.

For the RGB images, the pseudo-inverse solutions are essentially useless (Fig. 6.1), resulting in mean errors close to $15 \Delta E_{ab}$. The results improved when using base functions, with the best results obtained for spectral basis. However, a mean $\Delta E_{ab}$ difference of 4 and a maximum of 14 are still too high for the method to be considered as useful for multispectral imaging.
Figure 6.1. Reconstructed spectral reflectance from RGB-images. Full lines are measured spectral reflectance, the dashed lines are reconstructed.

Figure 6.2. Reconstructed spectral reflectance from multi-channel images. Full lines are measured reflectance spectra, the dashed lines are reconstructed.

For the multi-channel images, the results improve dramatically (Fig. 6.2). Both spectral basis and Fourier basis approaches lead to equivalent results in terms of the RMS spectral difference, while the colorimetric results are in favor of the spectral basis. The pseudo-inverse solution is somewhat noisy and suffers from larger spectral RMS difference, compared to the base-function approaches. However, apart from the noisy fluctuations, the general shapes of the reconstructed spectra follow the real spectra well. This result in colorimetric differences being comparable to the result achieved for the spectral basis and even slightly better. Clearly, the pseudo-inverse method produces
spectral reconstructions that are close to metameric matches to the measured spectra, even though they are not good spectral matches. This good performance for the pseudo-inverse solution is somewhat surprising and clearly illustrates the accuracy of the image acquisition model. The result is also a reminder that poor performances in terms of spectral RMS difference do not necessarily mean poor performance using a conceptually meaningful color difference metric.

Histograms exhibiting the distributions of the $\Delta E_{ab}$ errors under the D65 illuminant for the different reconstruction methods are depicted in Fig. 6.3. This clearly illustrates the improvement when increasing the number of channels from 3 to 7 (note the different scale). With a mean color difference of 1.5 $\Delta E_{ab}$ and the maximum 4.4 $\Delta E_{ab}$ (standard illuminant D65), the accuracy of the characterization model should be considered as good. Note that the PI-method, which produces the smallest colorimetric errors, is very general, since no basis functions are used, which should be representative for real data. However, if the aim is to reconstruct spectral data, not just colorimetric, it is preferable to reconstruct the spectral reflectance using base functions. The best overall performance is given by the method using spectral basis.

Figure 6.3. The distribution of the $\Delta E_{ab}$ errors between measured and reconstructed spectral reflectance, using standard illuminant D65.

Figure 6.4 displays the spectral reconstruction error bands, i.e. the difference between the reconstructed and measured spectral reflectance values, $\tilde{r} - r$, employing the different methods. As seen, the spectral errors by the pseudo-inverse approach are periodic, with small errors mainly occurring at wavelengths where the interference filters have their peak transmittance (compare to Fig. 5.2). The error band representing spectral basis reveals that the largest errors occur at the two ends of the visible spectrum, especially at the red end where the reconstructed reflectance is generally overestimated. The results from the Fourier basis show large errors at the both ends of the visible spectrum as well, and also in the band around 650 nm. The fact that the spectral reconstruction errors often tend to be largest at the far ends of the spectrum.
have been noted elsewhere, for spectral basis (Cheung & Westland, 2003), as well as for Fourier basis (Mansouri, et al, 2007).

Figure 6.4. Spectral reconstruction error band for the 25 NCS colors, using the different reconstruction operators for multi-channel images.

6.6 Summary and discussion

The inverse characterization function has been derived using a model-based approach, i.e. by inverting the spectral model of the image acquisition system, derived in the previous chapter. The result of the inversion is a set of spectral reconstruction matrices, allowing for the acquisition of multispectral images, i.e. to reconstruct the spectral reflectance from the recorded device response. To exploit *a priori* knowledge on the smooth nature of spectral reflectance, basis functions have been used to improve the accuracy of the inversion. The basis functions evaluated consist of Fourier basis and a large set of real reflectance spectra from NCS, combined with the principle eigenvector method. The spectral reconstruction methods have been tested using multi-channel images as well as trichromatic RGB-images. The results have been evaluated by comparing the reconstructed spectral reflectance to measurements by a spectroradiometer for a set of 25 NCS color patches.

The results show that using trichromatic RGB imaging is insufficient in reconstructing the spectral and colorimetric data, using the model-based approach. The best results, using spectral basis, produced a mean CIE 1976 color difference of $4 \Delta E_{ab}$ and the maximum $14 \Delta E_{ab}$.

When using multi-channel images, the improvement was significant, both in terms of colorimetric differences and in the RMS spectral reconstruction error. The best results in terms of colorimetric differences were obtained for the simple pseudo-inverse solution, producing a mean color difference of $1.5 \Delta E_{ab}$ and the maximum $4.4 \Delta E_{ab}$. The results using spectral basis were comparable in terms of $\Delta E_{ab}$ and significantly better in
terms of the RMS spectral difference. Fourier basis produced good results in RMS spectral difference, but slightly worse in terms of the CIE color differences.

The results indicate that if the aim is to derive colorimetric data only, the PI-method is sufficient. The PI-method is very general, since no basis functions are used, which should then be good representatives for the spectra likely to be encountered. However, the PI-method produces spectral reconstructions that are good metameric matches, not good spectral matches. Hence, for multispectral imaging, i.e. when reconstructing spectral data from the device response, the methods using base functions are required. The best overall performance is given by the method using spectral basis.

It is worth noting that the reconstructed spectra are estimated in 4 nm sampling intervals, to be directly comparable with the spectral measurements by the spectroradiometer. For most purposes, 10 nm sampling intervals would be sufficient, which would then simplify the estimation and probably reduce the noise sensitivity, especially for the PI-method.

It would be desirable to verify the results using a larger validation set, preferably of various media and colorants. Note that the spectral basis used consists of principle eigenvectors from the database of NCS color, which should be favorable when reconstructing reflectance spectra of NCS patches. For other media, it would probably be beneficial to use other spectral bases, e.g. including the reflectance of the substrate and the printing inks, for multispectral imaging of color prints. In a more recent study, the use of wavelets as basis functions for spectral reconstructions has been proposed (Mansouri, et al., 2007). Wavelet basis was found to perform better than Fourier basis and is in addition more general compared to spectral basis, which is dependent on the selected set of spectral reflectance data.

Possible ways of improving the accuracy of the spectral reconstruction could be to use Wiener estimation techniques, to take into account the signal dependent acquisition noise (Haneishi, et al, 2000). However, effective use of Wiener estimation requires a good estimation of the noise variance under real conditions (Heikkinen, et al., 2008).
Chapter 7

Regression-based colorimetric and spectral reconstruction

7.1 Introduction
7.2 Media dependency
7.3 Methodology
7.4 Experimental setup
7.5 Experimental results
7.6 Summary and discussion
7.1 Introduction

To compare the results using model-based characterization to estimate spectral and colorimetric data described in Chapter 6, experiments are performed using empirical characterization. The spectral characteristics of the image acquisition system are ignored and the inverse characterization functions are derived based on the recorded device response for a training set of color samples. By relating the device response to spectral and colorimetric measurements, the characterization functions are derived using least squares regression techniques, as described in Sec. 3.6.

Besides from colorimetric regression, mapping device values directly to CIEXYZ and CIELAB using polynomial regression, experiments are also performed on reconstructing spectral reflectance by linear least squares regression. The characterization functions are derived using both RGB and multi-channel images. To evaluate the characterization functions, a validation set of 50 test colors is used, independent of the training set used to derive the mappings. The spectral reconstructions are evaluated using the spectral RMS difference and the colorimetric estimations by using the Euclidian distance in CIEXYZ color space, and the CIE 1976 and 1994 color difference formulas.

Experiments and results that are not essential for the discussion have been moved to an appendix. Appendix B contains the experimental results for some additional polynomials evaluated as approximation functions, as well as experiments on the influence of the size of the training set, and the performance of the derived characterization functions when used for color samples of different media and colorants.

7.2 Media dependency

The characterization functions derived using empirical approaches will be optimized only for a specific set of conditions, including the illuminant, the media and the colorant. Once the conditions change, e.g. capturing images of a different substrate or a different print method, the characterization has to be re-derived in order to obtain good accuracy (Pan et al., 2001; Andersson et al., 2003). The dependency on the illuminant is natural and is often not an issue when the light source is fixed and can be considered as a property of the system. However, the fact that the characterization function is also
media- and colorant dependent is a major drawback, preventing the characterization function from being applied to arbitrary combinations of media and colorants.

The media- and colorants dependency can be explained from the so-called eye-camera metamerism (Hong & Lou, 2000), also referred to as device metamerism (Cheung, et al., 2005). Recall from Chapter 2 that two color samples are metameric when giving the same tristimulus values for a specific illumination condition, even though they are spectrally different. When the spectral responses of an imaging device are different from the color matching functions, i.e. the device is not colorimetric; colors that are metameric in terms of tristimulus values are not necessarily metameric with respect to the imaging sensors. This means that metameric color samples, which give the same tristimulus values under a given illuminant, may generate different device response values under the same illuminant. Conversely, two spectrally different samples, which are metameric for the specific imaging system, will give the same device response and thus the same colorimetric values after applying the characterization function, even though they are not a metameric match in terms of tristimulus values, under the given illuminant.

The magnitude of the color mismatch is related to the difference in spectral reflectance properties of the samples and the consequence of this eye-camera metamerism is that the mappings derived in the characterization process will be media dependent, i.e. optimal only for the combination of media and colorants used for the characterization. Since empirical characterization generally requires a large number of measurements to obtain device data and the corresponding colorimetric values of the training samples, this media dependence is a drawback if the characterization function is to be used for arbitrary combinations of media and colorants.

7.3 Methodology

In the following, recall the workflow for empirical characterization of input devices, presented in Sec. 3.4. The aim is to derive the inverse characterization function, mapping device response values \{d_i\} to device-independent colorimetric representations \{c_i\} and spectral reconstructions \{\hat{r}_i\}, using least squares regression techniques.

7.3.1 Colorimetric regression

A common approach for empirical characterization is to use linear or polynomial regression from device RGB values to CIEXYZ (see for example Hong & Lou, 2000). From the XYZ values it is then straightforward to compute the CIELAB values using standard formulae. The drawback with using regression to CIEXYZ is that the RMS error in XYZ color space, which is minimized in the regression, is not closely related to the perceived color difference. If the final aim is to derive data in CIELAB color space, it could therefore be preferable to use regression directly in the CIELAB domain, i.e., to minimize \(\Delta E_{ab}\) in the regression, which provides a better correspondence to the visual color difference (Hardeberg, 2001). Since the relationship between device RGB data and CIELAB color space is not linear, a non-linear pre-processing step of the RGB values using a cubic root function has been proposed, i.e. using \(R^{1/3}, G^{1/3}, B^{1/3}\) in the regression (Hardeberg, 2001). The cubic root function originates from the CIELAB
transformation, which involves a cubic root function of the XYZ tristimulus values (Eqs 2.8 – 2.10).

As a complement to regression directly to CIEXYZ or CIELAB, we propose an alternative approach, using a polynomial characterization function from device data to CIEXYZ color space, but minimizing the perceptually more meaningful ΔEab color difference in the process. This cannot be achieved using least squares regression techniques alone, but requires non-linear optimization. The parameters for the polynomial relating device data to CIEXYZ color space, derived by least squares regression, are used as initial conditions. Then, the new parameters are derived by an unconstrained nonlinear optimization, using the Optimization toolbox for Matlab, with the mean ΔEab color difference for the training set as a cost function. The algorithm used is the BFGS Quasi-Newton method, with a mixed quadratic and cubic line search procedure (Shanno, 1970). The result is a function relating device data to CIEXYZ in a way that minimizes the CIE 1976 color difference, ΔEab.

7.3.2 Spectral regression

Even though empirical approaches are mainly used to derive mappings to colorimetric data, CIEXYZ or CIELAB, the methodology can also be used to reconstruct spectral reflectance (see e.g. Solli, et al., 2005). Instead of seeking the relationship between device response values and colorimetric values, we now look at the linear relationship \( r = dA \), where \( r \) is the spectral reflectance \( N \)-component vector for a color sample, \( d \) is the \( m \)-component vector representing the corresponding device response, and \( A \) is an \( m \times N \) spectral reconstruction matrix. If the spectral reflectance for a set of \( T \) training samples \( \{r_i\} \) are collected into a \( T \times N \) matrix \( R = [r_1, \ldots, r_T] \) and the device responses \( \{d_i\} \) into a \( T \times m \) matrix \( D = [d_1, \ldots, d_T] \), then the linear relationship is given (Eq. 7.1) (compare to Eq. 3.4):

\[
R = DA
\]  

and the optimal reconstruction matrix \( A \) is then given by:

\[
A = (D'D)^{-1}D'R = (D')^{-1}R
\]  

where \((D')\) is the Moore-Penrose pseudo-inverse of \( D \). The spectral reflectance can then be reconstructed from camera response values as:

\[
\tilde{r} = dA
\]

Note that the spectral reconstruction matrix, \( A \), is now derived entirely based on the recorded device response to a set of training samples, i.e. ignoring the spectral characteristics of the imaging system.

7.3.3 Evaluation

Figure 7.1 summarizes the different approaches for deriving colorimetric and spectral reflectance values, and the way which the results are evaluated. From the recorded
device response, \( d \), (either RGB or multi-channel images), the spectral reflectance can be reconstructed using linear regression. The spectral estimation is compared to the spectral reflectance measured by a spectroradiometer to compute the RMS spectral difference.

From the spectral reflectance, the tristimulus values \( \text{CIEL}^{*}\text{XYZ} \) can be computed using the CIE color matching functions (Eq. 2.3) under a chosen illuminant. The XYZ values can alternatively be derived directly by colorimetric regression from the device response, or using the non linear optimization, minimizing \( \Delta E_{ab} \). The estimated tristimulus values are evaluated by computing the Euclidian distance to the XYZ values computed from the measured spectral reflectance.

From the XYZ tristimulus values it is then possible to compute CIELAB coordinates, according to Eqs 2.14 – 2.16. Alternatively, the CIELAB values can be derived using direct regression from device response, with or without the cubic root function as a pre-processing step. The CIELAB values are evaluated using the color difference formulas \( \Delta E_{ab} \) or \( \Delta E_{94} \), as described in Sec. 2.2.7.

![Figure 7.1. The different approaches for deriving colorimetric and spectral reflectance values from the recorded device response.](image)

### 7.4 Experimental setup

#### 7.4.1 The training set

Being rather time consuming in spectral measurements using a spectroradiometer, as well as in image acquisition, it is desirable to use as few test colors as possible, while achieving a satisfactory result. It has been suggested that a suitable number of test colors for polynomial regression is 40-60 training samples, with no significant
improvement of the accuracy obtained by further increasing the size of the training set
(Hong & Lou, 2000). Similarly, it has been shown that reducing the training set from
288 to 54 test colors gives only a very small increase in the estimation error (Hardeberg,
2001). The basic and most important demand is that the number of training colors is
larger than the number of parameters in the model, ensuring that the system is not
underdetermined.

As training set for the regression we use a set of 50 test colors, printed on HP
Photo-matte paper by an HP 10PS inkjet printer (Solli, 2004). The reason for not using
NCS colors, as with the model-based approach, was simply that the number of available
NCS color patches was not sufficient to be used both as training set and for evaluation.
To investigate the influence of the size of the training set on the characterization,
experiments are performed on reducing the number of training colors.

For the evaluation of the derived characterization function, 50 randomly selected
colors are used, printed using the same substrate and conditions as the training set.
Since characterization functions derived by least squares regression will always be
optimized for the specific training set, it is important to use a different set of test colors
for the evaluation. An independent set of evaluation colors guards against a model that
overfits the training set, giving unrealistically good results (Cheung, 2005).

7.4.2 Setup

To prevent the halftone pattern of the printed test colors from appearing in the images
because of the large magnification, the images are captured slightly out of focus, using
the shortest extension available (Tab. A.1). In all other aspects, the setup for the spectral
measurements as well as the acquisition and processing of the images, is identical to the
setup for the model-based characterization, as described in Sec. 5.5.

7.4.3 Choice of the approximation functions

Besides the basic linear regression, experiments are performed using polynomial
regression to CIEXYZ and CIELAB from RGB and multi-channel images, as described
in Sec. 3.6. The recorded device response is arranged into different polynomial $Q$
component vectors, $p$, and the optimal matrix of polynomial weights are found using
Eq. 3.10. The regression to CIELAB is performed using the device response directly, as
well as by using the cubic root function as a pre-processing step. Besides direct
regression, the polynomials for CIEXYZ conversion are also computed using non-linear
optimization techniques.

There are numerous ways to build the polynomial approximation functions, but to
limit the length of the chapter, only the two polynomials giving the best results, for
RGB and multi-channel data respectively, are presented. The results from experiments
using additional polynomials are presented in Appendix B, together with the results
using linear least squares regression.

RGB data

For polynomial regression using trichromatic RGB-images, we evaluate the following
two polynomials as approximation functions:
A second order polynomial completed with the additional RGB term, with the dimension $Q = 11$. Hong and Lou (2000) achieved the best result using this polynomial and especially emphasize the importance of including the ‘black’ $1$ and ‘white’ RGB terms in the polynomial functions.

$$\mathbf{p}_2 = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2, RGB]$$ (7.4)

A third-order polynomial with $Q = 20$. This polynomial has been successfully used for mapping RGB values to CIELAB, using the pre-processing step described earlier, for both scanners and digital cameras (Hardeberg, 2001; Andersson, et al., 2003; Andersson, 2004).

$$\mathbf{p}_3 = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2, R^3, R^2 G, R^2 B, R G^2, \quad
RB^2, G^3, G^2 B, GB^2, B^3]$$ (7.5)

Multi-channel data

Empirical characterization using multi-channel images is not as common as using RGB-images. Generally, model-based approaches are used for multi-channel imaging devices. To our knowledge, empirical characterization for multi-channel devices has been limited to linear regression from multi-channel data to CIEXYZ, for various number of color channels (Hardeberg, 2001).

For multi-channel images there are numerous ways to build the approximation functions and the number of terms increases rapidly for higher order polynomials, making it unfeasible to include all the cross-product terms. Experiments are performed on linear regression to CIEXYZ as well as directly to CIELAB, using 7 channel and 14-channel images. The 7 channel images correspond to the 7 interference filters, $M_1, ..., M_7$. The 14-channel images are completed with the RGB-filters, the CMY-filters and the white $W$-channel, i.e. with no filter in the optical path. For the conversions to CIELAB color space it is investigated whether the cubic root pre-processing step is useful even for multi-channel data. The two selected polynomials are:

$$\mathbf{p}_{m2} = [1, M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_1^2, M_2^2, M_3^2, \quad
M_4^2, M_5^2, M_6^2, M_7^2, M_1 M_2 M_3 M_4 M_5 M_6 M_7]$$ (7.6)

A second-order polynomial using the 7 interference filters, without cross-product terms. $Q = 16$

$$\mathbf{p}_{m5} = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2, W, M_1, \quad
M_2, M_3, M_4, M_5, M_6, M_7, C, M, Y, \quad
W R G B M_1 M_2 M_3 M_4 M_5 M_6 M_7 C M Y]$$ (7.7)

This polynomial includes second-order terms from the RGB-channels, together with first-order terms from the remaining 11 channels. The ‘white’ term is based on all 14 channels. $Q = 22$. 

92
7.5 Experimental results

7.5.1 Colorimetric regression

The characterization functions are derived using polynomial regression and the non-linear optimization routine, based on the training set of 50 colors patches. The derived functions are then used to predict the colorimetric values for 50 evaluation colors, independent of the training set. The estimated colorimetric values are compared to the colorimetric values from measurements using a spectroradiometer. For the regression to CIEXYZ, we report the results using the Euclidian distance in CIEXYZ color space, \( \Delta XYZ \), as well as the color differences in CIELAB color space, \( \Delta E_{ab} \) and \( \Delta E_{94} \). The results from regression directly to CIELAB are evaluated using the color differences \( \Delta E_{ab} \) and \( \Delta E_{94} \). The reported error metrics are computed using the standard illuminant D65. The corresponding results for standard illuminant A are given in Nyström, 2006. For each metric, the maximum and mean values for the 50 evaluation samples are presented, as well as the 95th percentile. The results for the linear least squares regression and a comparison of the results for a number of additional polynomial approximation functions are presented in Appendix B.

**RGB data**

The results for the characterization to CIEXYZ using trichromatic RGB data are given in Tab. 7.1, in terms of \( \Delta XYZ \), for polynomial regression to CIEXYZ, and the functions derived using non-linear optimization, respectively. The corresponding results in terms of the color differences \( \Delta E_{ab} \) and \( \Delta E_{94} \) are given in Tab. 7.2, together with the results for regression directly to CIELAB, with and without the pre-processing step.

| Regression to | Polynomial | \( \Delta XYZ \) | | | |
|--------------|------------|-----------------|---|---|
| XYZ          | \( p_2 \)  | Max 4.90        | Mean 1.01 | 95% 2.48 |
|              | \( p_3 \)  | Max 3.45        | Mean 0.90 | 95% 2.15 |
| (Min \( \Delta E_{ab} \)) | \( p_2 \)  | Max 6.19        | Mean 1.00 | 95% 2.35 |
|              | \( p_3 \)  | Max 4.21        | Mean 0.89 | 95% 2.25 |

**Table 7.1.** The results from the regression using RGB-data, in terms of the Euclidian distance in CIEXYZ color space, \( \Delta XYZ \), using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>( \Delta E_{ab} )</th>
<th>Max 2.51</th>
<th>Mean 1.23</th>
<th>95% 2.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ</td>
<td>( p_2 )</td>
<td>Max 3.29</td>
<td>Mean 1.68</td>
<td>95% 2.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p_3 )</td>
<td>Max 3.98</td>
<td>Mean 1.91</td>
<td>95% 3.36</td>
<td></td>
</tr>
<tr>
<td>(Min ( \Delta E_{ab} ))</td>
<td>( p_2 )</td>
<td>Max 5.83</td>
<td>Mean 2.82</td>
<td>95% 5.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p_3 )</td>
<td>Max 5.19</td>
<td>Mean 1.91</td>
<td>95% 3.37</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>( p_2 )</td>
<td>Max 6.31</td>
<td>Mean 1.76</td>
<td>95% 3.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p_3 )</td>
<td>Max 4.31</td>
<td>Mean 1.72</td>
<td>95% 3.36</td>
<td></td>
</tr>
<tr>
<td>(pre-proc.)</td>
<td>( p_2 )</td>
<td>Max 3.08</td>
<td>Mean 1.17</td>
<td>95% 2.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p_3 )</td>
<td>Max 4.04</td>
<td>Mean 1.11</td>
<td>95% 1.99</td>
<td></td>
</tr>
</tbody>
</table>

**Table 7.2.** The results from the regression using RGB-data, in terms of CIE 1976 color difference \( \Delta E_{ab} \) and CIE 1994 color difference \( \Delta E_{94} \), using standard illuminant D65.
The best results of the regression from RGB to CIEXYZ in terms of $\Delta$XYZ were obtained for the third order polynomial $p_3$ (Tab. 7.1). However, when converting the derived XYZ-values to CIELAB, the second order polynomial $p_2$ outperforms $p_3$ in terms of $\Delta E_{ab}$ and $\Delta E_{94}$ (Tab. 7.2). Even though producing the best results in terms of minimizing the $\Delta$XYZ distance, $p_3$ is not the optimal choice if the visual color difference is considered.

For the regression to CIELAB color space, it is clear that direct regression from RGB to CIELAB, without the pre-processing step, is not favorable. Better results are always obtained, both by computing $L^*a^*b^*$ values from the derived XYZ-values, and when using the cubic root function as a pre-processing step in the regression to CIELAB.

The best results in terms of the visually meaningful color differences, $\Delta E_{ab}$ and $\Delta E_{94}$, were generally obtained when using the functions converting RGB to CIEXYZ, derived by non-linear optimization to minimize the $\Delta E_{ab}$ color difference. However, the method produces slightly higher errors in terms of $\Delta$XYZ, and requires some optimization software, as well as extra computational time. Note that the second order polynomial, $p_2$, performs equally well as the third order polynomial $p_3$. Clearly, a higher order polynomial with a larger number of polynomial terms does not necessarily result in more accurate estimations.

Recall the results for the model-based approach, using RGB-images (Sec. 6.5) where the best results corresponded to a mean color difference exceeding 4.1 $\Delta E_{ab}$ and with the maximum close to 16 $\Delta E_{ab}$. Obviously, regression methods to colorimetric data are superior to model-based characterization when only RGB-data is available. At least, this is true when the functions are used to predict colorimetric values for samples of the same media and colorant that have been used to derive the characterization function.

### Multi-channel data

The results of the regression using multi-channel data are given in Tables 7.3 and 7.4. The first and perhaps the most obvious observation from the results is how little one can gain by using multi-channel data. Compared to the model-based approach where the results improved dramatically when extending from 3 to 7 channels, we see that with the empirical approach, multi-channel data lead to almost identical results as for RGB-data. Clearly, there is nothing to gain by increasing the number of color channels when polynomial regression is used to derive colorimetric data.

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>$\Delta$XYZ</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ $p_m2$</td>
<td>3.24</td>
<td>0.94</td>
<td>2.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XYZ $p_m5$</td>
<td>3.75</td>
<td>0.93</td>
<td>1.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XYZ $p_m2$</td>
<td>3.49</td>
<td>0.97</td>
<td>2.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XYZ $p_m5$</td>
<td>3.80</td>
<td>0.94</td>
<td>1.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.4. The results from the regression using multi-channel data, in terms of CIE 1976 color difference $\Delta E_{ab}$ and CIE 1994 color difference $\Delta E_{94}$, using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>$\Delta E_{ab}$ Max</th>
<th>$\Delta E_{ab}$ Mean</th>
<th>$\Delta E_{ab}$ 95%</th>
<th>$\Delta E_{94}$ Max</th>
<th>$\Delta E_{94}$ Mean</th>
<th>$\Delta E_{94}$ 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ</td>
<td>$p_{m2}$</td>
<td>3.76</td>
<td>1.94</td>
<td>3.22</td>
<td>2.24</td>
<td>1.16</td>
<td>2.13</td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>3.16</td>
<td>1.96</td>
<td>3.13</td>
<td>2.91</td>
<td>1.23</td>
<td>2.31</td>
</tr>
<tr>
<td>XYZ (Min $\Delta E_{ab}$)</td>
<td>$p_{m2}$</td>
<td>3.71</td>
<td>1.88</td>
<td>3.40</td>
<td>2.22</td>
<td>1.17</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>3.44</td>
<td>1.99</td>
<td>3.17</td>
<td>2.72</td>
<td>1.23</td>
<td>2.28</td>
</tr>
<tr>
<td>Lab</td>
<td>$p_{m2}$</td>
<td>8.43</td>
<td>2.34</td>
<td>4.52</td>
<td>5.11</td>
<td>1.47</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>8.97</td>
<td>2.69</td>
<td>5.11</td>
<td>5.03</td>
<td>1.64</td>
<td>3.69</td>
</tr>
<tr>
<td>Lab (pre-proc.)</td>
<td>$p_{m2}$</td>
<td>3.84</td>
<td>1.95</td>
<td>3.78</td>
<td>2.41</td>
<td>1.22</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>3.25</td>
<td>1.77</td>
<td>2.81</td>
<td>2.28</td>
<td>1.21</td>
<td>2.13</td>
</tr>
</tbody>
</table>

For the regression to CIELAB, the non-linear pre-processing step still improves the results, especially in terms of the maximum errors. Clearly, the pre-processing step plays an important role even for multi-channel images, the same way as for the RGB-images. The estimation errors for the regression to CIELAB, using the pre-processing step, are in the same order of magnitude as the errors for regression to CIEXYZ and for the functions derived using the non-linear optimization routine.

Distributions of the $\Delta E_{ab}$ errors for the polynomials regression from RGB and multi-channel images respectively are displayed in Fig. 7.2, using standard illuminant D65. Clearly, the results are very similar between RGB and multi-channel data, as well as between regressions to CIEXYZ and to CIELAB using pre-processing.

Figure 7.2. The distribution of the $\Delta E_{ab}$ errors for the colorimetric estimations, using the polynomial regression, under standard illuminant D65.

95
7.5.2 Spectral regression

The spectral reconstruction matrix $A$ has been derived according to Eq. 7.2, using the recorded device response for the set of 50 training colors. To be directly comparable with the results from the model-based characterization, the spectral reflectance is estimated in the range 400:4:700 nm, giving $N = 76$, i.e. $N > T$ and the system is thus underdetermined.

The spectral reflectance has been reconstructed according to Equation 7.3, for the set of 50 evaluation colors. As described in Chapter 6, we can utilize the a priori knowledge that spectral reflectance are typically smooth functions to represent the reconstructed spectra as linear combinations of smooth basis functions. As with the model-based approach, we evaluate spectral basis, using the NCS database, as well as Fourier basis. Five basis functions are used, corresponding to the first five Fourier basis functions and to the five singular vectors corresponding to the most significant singular values in the spectral autocorrelation function of the spectral database.

The results from the empirical spectral reconstruction, in terms of spectral RMS difference and $\Delta X\text{YZ}$ are given in Table 7.5. The corresponding colorimetric reconstruction errors in terms of CIE color differences, $\Delta E_{ab}$ and $\Delta E_{94}$, are listed in Tab. 7.6, computed using D65. Examples of estimated spectral reflectance for 3 different color samples are presented in Figs 7.3 and 7.4, derived from RGB and multi-channel images, respectively.

### Table 7.5. Spectral reconstruction errors, expressed as RMS difference in spectral space, and the Euclidian distance in CIEXYZ-space, $\Delta X\text{YZ}$.

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>RMS</th>
<th>$\Delta X\text{YZ}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>RGB</td>
<td>PI</td>
<td>0.0082</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>0.0062</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>0.0072</td>
<td>0.0035</td>
</tr>
<tr>
<td>Multi</td>
<td>PI</td>
<td>0.0030</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>0.0040</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>0.0052</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

### Table 7.6. Colorimetric reconstruction errors, in terms of CIE 1976 color difference, $\Delta E_{ab}$ and CIE 1994 color difference $\Delta E_{94}$ using D65 and Illuminant A.

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>$\Delta E_{ab}$</th>
<th>$\Delta E_{94}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>RGB</td>
<td>PI</td>
<td>13.2</td>
<td>7.53</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>12.4</td>
<td>7.44</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>13.8</td>
<td>6.89</td>
</tr>
<tr>
<td>Multi</td>
<td>PI</td>
<td>6.89</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>9.32</td>
<td>5.52</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>13.9</td>
<td>6.08</td>
</tr>
</tbody>
</table>

For the spectral reconstruction with the RGB-images, we see that the results for the pseudo-inverse solution are clearly superior to the corresponding results using the model-based approach. This is not surprising considering that the matrix used in the inversion based on the test samples (Eq. 7.2) is of rank 50, compared to the spectral
transfer matrix for the RGB channels (Eq. 6.5) which is of rank 3. However, the improvement when applying the different basis functions is not as evident as for the model-based approach and the best results for the model-based approach, using spectral basis, could not be achieved.

Figure 7.3. Empirically reconstructed spectral reflectance using RGB images. Full lines are measured reflectance spectra, the dashed lines are reconstructed.

Figure 7.4. Empirically reconstructed spectral reflectance using multi-channel images. Full lines are measured reflectance spectra, the dashed lines are reconstructed.
The results of the spectral reconstruction improve for multi-channel images compared to RGB-images, but not to the same degree as the corresponding improvement achieved for the model-based approach. In terms of spectral RMS difference, the results are comparable to the model-based results (compare to Tab. 6.1). However, in terms of colorimetric differences, the estimation errors are always larger. The best results obtained, in terms of $\Delta E_{ab}$ color difference, were for the PI method, which produced estimates with a mean color difference of 3.9 $\Delta E_{ab}$ and the maximum 6.9 $\Delta E_{ab}$. This is to be compared to the corresponding results for the model based approach, giving a mean color difference of 1.5 $\Delta E_{ab}$ and the maximum 4.4 $\Delta E_{ab}$ (Tab. 6.2).

Histograms of the distribution of the CIE 1976 $\Delta E_{ab}$ differences using D65 are given in Fig. 7.5, for the different reconstruction methods. The results improve for multi-channel images, but not to the same levels of accuracy obtained using the model-based approach (compare to Fig. 6.3). Restricting the reconstructed spectra to be linear combinations of basis functions does not improve the result for the multi-channel images.

If the aim is to estimate colorimetric data only, the polynomial regression directly to CIEXYZ and CIELAB is clearly superior to computing colorimetric values from the empirical spectral reconstructions (compare to Tables 7.3 and 7.4, Fig. 7.2). Cheung and Westland (2003) found that for a three-channel system, multispectral imaging is not an advantage when the aim is to compute colorimetric values. These results confirm that the same holds even for a multi-channel system. At least, this is true for samples of the same media and colorant that have been used to derive the characterization function. For cross-media characterization, however, colorimetric values from spectral regression seem to be more accurate compared to colorimetric regression (see Appendix B).

Figure 7.5. The distribution of the $\Delta E_{ab}$ errors between measured and reconstructed spectral reflectance, using empirical spectral reconstruction. Standard illuminant D65.
7.6 Summary and discussion

In this chapter, empirical characterization has been used to derive mappings from device dependent data to the colorimetric representations CIEXYZ and CIELAB, and to reconstruct spectral reflectance from the recorded device signals. In the empirical approach, the spectral characteristics of the system are ignored and the mappings are derived by relating the recorded device response for a set of training colors, using least squares regression techniques.

The results for the empirical characterization showed that good colorimetric accuracy could be obtained by using polynomial regression to the colorimetric representations CIEXYZ and CIELAB. It can be noted that the results using trichromatic imaging are comparable to the multi-channel results, an obvious difference to the results for the model-based approach. Clearly, there is nothing to gain by increasing the number of channels from 3 to 7 when using the colorimetric regression approach. This can be explained by the fact that the system of equations that is inverted is now based on the number of samples in the training set, not on the number of channels in the image acquisition system as is the case for the model-based characterization. The results for colorimetric regression using RGB-images offered the same colorimetric accuracy as for the model-based characterization using multi-channel images.

However, the accuracy is dependent on the mappings being derived using the same conditions, media and colorants. Experiments on cross-media characterization, using color samples of different media and colorants (presented in Appendix B) revealed that the eye-camera metamerism is severe, producing large errors when the characterization functions are used for different media and colorants. The more complex the estimation functions, i.e. the higher the order of the polynomial and the number of terms included, the stronger is the effect of the cross media characterization. This is the main drawback of regression techniques, that the characterization will be optimized only for the specific set of conditions, including the colorant and the media, used to derive the characterization functions. It has been demonstrated that even variations in a single paper property, such as fluorescence, can seriously damage the cross-media performance when using empirical characterization techniques for paper measurements (Andersson, 2007).

Empirical techniques were also used to reconstruct spectral reflectance, by computing a spectral reconstruction matrix based on measurements on the training set. The results revealed that for RGB-images, the pseudo-inverse method performed considerably better than the corresponding model-based approach. However, using multi-channel images and building the spectra from basis functions did not improve the results to the same degree, and the level of accuracy obtained for the model-based approach could not be reached for the empirical spectral estimations. The spectral reconstruction using regression could possibly be improved by increasing the training set used to derive the reconstruction matrix. Alternatively, the spectra could be reconstructed using a coarser interval, e.g. 400:10:700 nm, giving \( N = 31 \), and thus a reconstruction matrix of sufficient rank. However, the empirical approach will still suffer from the media dependence, making the more general model-based approach a better choice for multispectral imaging, at least when multi-channel data is available.

The training set used consisted of 50 color patches, which should be sufficient according to e.g. Hong & Lou (2000). This has later been verified by de Lasarte, et al.
(2008), claiming that the improvement in accuracy is very limited when increasing the training set, and that the results are practically independent of the number of samples from 40 samples on. However, other studies have shown that the maximum error decreases with a larger training set (Cheung & Westland, 2003) and that the size of the training set is a very important factor in the training process (Jetsu, et al., 2006). Clearly, the results are contradictory, and it is not easy to draw any general conclusions on the sufficient size of the training set. This will probably depend on a number of factors, such as the exact methodology used for the regression, the number of color channels in the imaging system and the strategy adopted for sample selection. Therefore, it would be desirable to conduct the experiments using a larger training set to examine if the results could improve. A larger validation set to evaluate the results would also be preferable.

Beside the regression techniques adopted in this study, other methods proposed for empirical characterization, mapping device data to colorimetric representations, include e.g. combinations of look-up-tables and interpolation (Hung, 1993). Deriving the characterization functions using neural networks has also been used in several studies (e.g. Cho, et al., 1995; Cheung & Westland, 2002). However, Cheung, et al. (2004) concluded that polynomial regression offers the better alternative, since neural networks can be difficult and time-consuming to train, and still only offer a similar performance.

It is worth mentioning that for empirical characterization techniques, the imaging system will always be calibrated against a specific instrument, i.e. the spectroradiometer or spectrophotometer used as a reference in the characterization process (Andersson, 2006). Several studies have shown that there can be considerable differences between color measuring instruments, such as spectrophotometers and spectrodensitometers, especially when there are differences in measurement geometry or illuminants (Spooner, 1994; Wyble & Rich, 2007; Radenic & Bohan, 2008). All experimental results from device characterization using empirical techniques should be interpreted with this in mind.
Chapter 8

High resolution analysis of halftone prints

8.1 Introduction
8.2 Methodology
8.3 Experimental setup
8.4 Experimental results
8.5 Summary and discussion
8.1 Introduction

As described in Chap. 4, many models have over the years been proposed to predict the outcome of halftone prints. The task is difficult, partly because of light scattering within the paper bulk, causing optical dot gain. Furthermore, the optical dot gain often co-exists with physical dot gain, caused by physical dot extension in the printing process. Understanding the physical nature of physical and optical dot gain in halftone prints is essential for accounting for the effect in order to improve print quality. Most available models are based on macroscopic color measurements, the average value over an area that is large relative to the halftone dot size. The aim of this study is to go beyond the macroscopic approach, and to study halftone print reproduction on a micro-scale level.

In the 1930s Murray and Davies published the first model to predict the output reflectance of a halftone print (Murray, 1936). The mean reflectance \( R \), is simply given by linear interpolation of the reflectance of the bare paper, \( R_p \), and the full tone, \( R_i \), weighted by the dot area fraction, \( a \) (Eq. 4.2). Since the reflected light from different areas is added to predict the overall reflectance, the Murray-Davies (MD) model preserves the linearity of photon additivity. It is, however, well-known that the applicability of the model is very limited. The relationship of \( R \) versus \( a \) is in fact nonlinear, due to light scattering in the paper substrate, causing optical dot gain.

In the 1950s, it was shown (Yule & Nielsen, 1951) that the nonlinear relationship could be approximated by a power function (Eq. 4.4). The Yule-Nielsen \( n \)-factor, accounting for light scattering in the paper, is an empirically derived constant, selected to provide the best fit to experimental data. However, the Yule-Nielsen (YN) model does not physically describe the phenomenon of optical dot gain, and the conservation of energy is lost when the nonlinear transform is applied to the reflectance values. The model was later developed (Clapper & Yule, 1953) to take into account surface reflection, internal reflection and ink transmission (Eq. 4.9).

Notice that the fundamental assumption in these models is that the color for the substrate and the ink is both uniform and constant. The inputs to the models are the reflectance values for the unprinted paper and for the ink at full ink area coverage. These reflection values are then assumed to be constant for all intermediate ink area percentages. However, in the 1990s it was shown that the color of the halftone dots and the paper between the dots is not constant, but dependent on the dot area fraction (Engeldrum, 1994). The reflectance of the printed halftone dots, as well as the paper between them, decreases with increasing dot area coverage, due to the light scattering in
the substrate. An expanded Murray-Davies model was later proposed, with the constants for paper and ink reflection replaced by the functions $R_i(a)$ and $R_p(a)$ (Arney, et al., 1995a; 1995b):

$$R(a) = aR_i(a) + (1 - a)R_p(a) \quad (8.1)$$

This model preserves the additivity of reflectance while the non-linear relationship between $R$ and $a$ caused by optical dot gain, is now accounted for by using the functions $R_i(a)$ and $R_p(a)$. Naturally, the difficulty with this approach is to derive $R_i(a)$ and $R_p(a)$, i.e. the way that the reflectance of the ink and paper shift with varying dot area coverage. It is clear that it is not possible to measure these components using macroscopic color measurements, giving the averaged reflectance of the print. Clearly there is a need for accurate color measurements also on a microscopic level.

Previous attempts to measure and characterize $R_i(a)$ and $R_p(a)$ have been made by point-wise measurements using a spectroradiometer equipped with magnification lenses (Engeldrum, 1994). Spectral and colorimetric values for the halftone dots and the paper between the dots were then measured, for a 25 lpi AM print, using a field of view corresponding to a circle of 0.13 mm in diameter. Clearly, the method with point wise measurements is not sufficient to derive measurements on a micro-scale level for prints with more reasonable print resolutions. Furthermore, the limited number of measurements will not necessarily be representative for a large population of halftone dots. Later on, histogram data from grayscale images have been used to compute the micro-reflectance values for the dots and the paper (Arney, et al., 1995a). The application was then limited to monochrome grayscale prints, and the maximal print resolution was only 330 dpi.

The experimental imaging system introduced and described in the previous chapters allows for micro-scale imaging of halftone prints, combining a high spatial resolution with colorimetric and spectral accuracy. The aim of this chapter is to investigate if histogram data from such images can be used to measure and characterize $R_i(a)$ and $R_p(a)$, for both grayscale and color halftones. The validity of the expanded Murray-Davies model (Eq. 8.1) is evaluated for AM and FM halftone prints of various resolutions, including resolutions considerably higher than in previous works. An extension of the model to predict tristimulus values is also proposed, using 3D histograms in CIELAB color space. The methodology also requires methods to accurately compute the physical dot area, a task that involves determining the physical dot gain.

### 8.2 Methodology

The possibility of acquiring colorimetric and multispectral images of prints opens for new approaches to properly characterize the microstructure of halftone color prints. Micro-scale images, i.e. when the resolution of the images is high in relation to the resolution of the printed halftone, allow for measurements of the individual halftone dots, as well as the paper between them. Examples of micro-scale images for 40% blue prints are displayed in Fig. 8.1, for a 100 lpi AM halftone, as well as a 600 dpi FM halftone.
8.2.1 Reflectance histograms

To capture the typical characteristics of a large population of halftone dots, which may differ in their appearance, histograms are computed from the micro-scale images. A histogram is merely a probability density function for the occurrence of different colors or reflectance values in the image, and contains nominally no spatial information. However, with the a priori knowledge that the image represents a halftone print captured in micro-scale, one can relate properties of the histogram to spatial properties of the halftone, such as the edges of the halftone dots (Arney & Wong, 1998; Nyström & Kruse, 2005).

A reflectance histogram is a plot of the frequency of occurrence of the reflectance values in the images, as a function of the reflectance, \( R \). For a perfectly reproduced halftone pattern, the histogram would be truly bimodal, with only two peaks corresponding to the reflectance of the ink and the paper. For a real print, however, the populations around \( R_i \) and \( R_p \) are typically spread out, due to the spread or blurring of the halftone dots, and to light scattering within the substrate. The closer the histogram corresponds to the ideal bimodal case, i.e. the higher the peaks representing ink and paper, and the deeper the valley representing the dot edges, the sharper and more well defined are the printed halftone dots.

Figure 8.2 displays normalized reflectance histograms for prints of 50% nominal dot area coverage, using frequency modulated halftones of various print resolutions on the same paper grade. In the histograms, the peaks correspond to the ink dots and the paper between the dots. The valley between the peaks corresponds to the edges between the halftone dots and the paper substrate. With the relatively large halftone dots of the 150 dpi print, the histogram is close to the ideal case, indicating uniform halftone dots with hard, well defined, edges. With increasing print resolution, the size of the halftone dots gets smaller in thus more difficult to reproduce sharply. Smaller halftone dots generally give soft edges, with the ink thickness at the edges smaller than that of the dot center (Azuma, et al., 2003). Furthermore, when the dot size gets small in relation to the lateral scattering length of the light, the optical dot gain will increase due to light scattering in the substrate. The effect is that with increasing print resolution, an increasing part of the reflectance values will correspond to the edges between the
printed halftone dots and the paper, represented by the valley between the peaks in the histograms, Fig. 8.2.

![Figure 8.2. Normalized reflectance histogram for prints of 50% nominal dot area coverage, using FM halftones of various print resolutions.](image)

Furthermore, from the reflectance histograms it is possible to study how the reflectance of the ink, $R_i(a)$, and the paper, $R_p(a)$, varies with the dot area coverage, $a$. Figure 8.3 displays an example of a normalized histogram of reflectance values, for the full tone ($a=1$); the bare paper ($a=0$), and a 60% tint. It is clear from the positions of the peaks for the 60% tint that the reflectance values of the ink, $R_i(0.6)$, and paper, $R_p(0.6)$, have shifted with the dot area coverage, $a$.

![Figure 8.3. Normalized reflectance histogram for halftones prints of $a=0$, $a=0.6$ and $a=1$.](image)
From the histogram it is possible to derive the mean reflectance of the print, $R$, as well as the micro reflectance for the dots, $R_i(a)$, and that of the paper between the dots, $R_p(a)$. The mean reflectance, $R$, corresponding to the macroscopic measurement, is given by integral:

$$R = \int_0^1 R \cdot H(R)dr$$

(8.2)

The integral can be divided into two parts, corresponding to the contribution from the ink and from the paper, respectively:

$$R = \int_0^{R_t} R \cdot H(R)dr + \int_{R_t}^1 R \cdot H(R)dr$$

(8.3)

where $R_t$ is a reflection threshold value, defining the transition between the printed dot and the paper. If the reflectance of the ink and the paper is approximated by the peak values from the histogram, $R_i(a)$ and $R_p(a)$, then Eq. 8.3 becomes:

$$R = R_i(F_i)\int_0^{R_t} H(R)dr + R_p(F_p)\int_{R_t}^1 H(R)dr$$

(8.4)

The integrals in Eq. 8.4 correspond to the area fractions for the ink, $a$, and the paper, $a_p = (1-a)$, and Eq. 8.4 thus reduces to the expanded Murray Davies according to Eq. 8.1. (Arney, et al., 1995a)

### 8.2.2 Estimating the physical dot area coverage

It is important that the dot area fraction, $a$, used in Eq. 8.1 corresponds to the physical dot area coverage, $a_{phy}$, which typically differs from the nominal dot area, $a_0$, due to physical dot gain in the printing process. The physical dot area fraction can be computed from the histogram, using a threshold value, $R_t$, as the limit between the ink and the paper, as:

$$a_{phy} = \frac{\int_0^{R_t} H(R)dr}{\int_0^1 H(R)dr}$$

(8.5)

The most straightforward approach to define the threshold, $R_t$, is to use the midpoint between the peaks corresponding to $R_i(a)$ and $R_p(a)$ in the histogram. Because no extra data are needed beside $R_i(a)$ and $R_p(a)$, this approach can also be used in cases when only these components are available.
When micro-scale images of the prints are available, another possibility is to derive the threshold from image data. By using vertical and horizontal line scans across the halftone dots, \( R_t \) can be defined as the region of maximum rate of change in reflectance values, \( dR/dx \) (Arney, et al., 1995a). An edge is a part of the image where the tone variation is large, and \( R_t \), representing the boundary between the dot and the paper, is thus defined as the position having the steepest slope in reflection between dot and paper. To ensure that the threshold value is representative for all halftone dots, every tenth line in the image, both horizontally and vertically, are used when computing \( R_t \). Figure 8.3 illustrates an example of a scan line through a few halftone dots, along with the corresponding value of \( R_t \) and the points of maximum rate of change, \( dR/dx \), for each edge between ink and paper.

![Figure 8.3. Illustration of a line scan (solid) across halftone dots, with the maximum rate of change \( dR/dx \) (+) and the resulting threshold value, \( R_t \) (dotted).](image)

### 8.2.3 Color histograms

For images of halftone color prints captured in micro-scale, the only colors present are the ones representing the paper, the printed primary inks, and different combinations of overlapping primary inks. In a 3D color histogram from such images, the paper and the different combinations of printed inks will appear as clusters in the histogram, with the transitions between the clusters corresponding to the edges of the halftone dots. This requires the resolution of the images to be sufficiently high in relation to the resolution of the halftone, to ensure that each individual halftone dot is represented by a sufficient number of pixels, and that the characteristics of the edges are accurately represented in the images. If the resolution of the image is reduced, the initial clusters will dissolve and relocate towards the position in color space corresponding to the perceived color and, correspondingly, to the macroscopic color measurement.

Figure 8.4 depicts an example of a visualization of a 3D color histogram in CIEXYZ color space, for a 60% blue halftone print. The four clusters correspond to the
paper, the primary inks cyan and magenta, and the secondary blue overlapping of the primary inks.

Figure 8.4. Visualization of 3D color histogram in CIEXYZ color space for a 60% blue halftone print.

8.2.4 Extension to color

When extending the expanded Murray-Davies to color prints, the reflectance values in Eq. 8.1 are replaced by the CIEXYZ tristimulus values for the ink and the paper, $XYZ_p(a)$ and $XYZ_i(a)$. After converting the micro-scale images into CIEXYZ color space, 3D color histograms are computed. In the 3D histograms, the paper and the inks appear as clusters, with the transitions between the clusters corresponding to the edges of halftone dots. The tristimulus values of the paper and the ink, $XYZ_p(a)$ and $XYZ_i(a)$, can then be computed as the centers of gravity of the clusters corresponding to the paper and ink, for each area coverage, $a$. The threshold value $R_t$ is now replaced by a threshold plane in XYZ color space, located at the midpoint, orthogonal to the vector between the clusters.

However, computing full 3D histograms and performing all computations in 3D, is computationally heavy. Furthermore, for some prints the clusters are not easily defined, leading to errors in the estimations of the values for $XYZ_p(a)$ and $XYZ_i(a)$, as well as the threshold plane. Since the transitions occurring between the clusters are found to be close to one-dimensional, we suggest a projection down to one-dimensional color distributions (Wedin & Kruse, 1995; Nyström & Kruse, 2005). The centers of gravity for the clusters of full-tone ink and bare paper are used to create a projection-vector between the clusters. Then, an orthogonal projection of all color coordinates in the XYZ images are made onto the vector. The resulting 1D color distributions are normalized so that 0 and 1 correspond to the centers of gravity for the unprinted paper and the full ink coverage, respectively. Figure 8.5 illustrates examples of projections to one-dimensional color distributions, for magenta halftones of various nominal ink coverage, $a$. 

109
Figure 8.5. One-dimensional color distributions for magenta halftones, projected from CIEXYZ color space.

After computing the one-dimensional color distributions, CDs, the same methodology as described in Sec. 8.2.1 for the reflectance case can be applied to find the peaks for the ink and paper in terms of the 1D color distribution, $CD_p(a)$ and $CD_i(a)$, and to compute the physical dot area fraction, $a_{phy}$. The resulting values in terms of the 1D color distribution are then simply converted back to CIEXYZ tristimulus values, using the known vector of projection.

8.3 Experimental setup

8.3.1 Printed samples

The printed samples used in the study consisted of offset prints on two different paper grades, matt and gloss, printed with amplitude modulated halftones (AMs), as well as frequency modulated halftones (FMs). The AM halftones used clustered dots and were generated for the screen frequencies 65, 100, 135 and 170 lpi. The FM halftones were generated for the print resolutions 150, 300, 600 and 1200 dpi, using an iterative method for optimal dot placement (Gooran, 2001). The nominal dot area coverage of the patches are 2, 5, 10, 15,..., 90, 95, 98 and 100% respectively. The test charts were printed with a commercial 4-color offset press (Heidelberg GT052), and all the patches were printed using the same plate.

Macroscopic measurements of the spectral reflectance values of the printed color patches were derived using a Gretag Machbeth Spectrolino spectrophotometer, equipped with a UV filter, using the $45^\circ/0^\circ$ measurement geometry. All colorimetric computations were made using the CIE standard illuminant D65.
8.3.2 Micro-scale images

Grayscale and RGB images of various test prints have been captured using the 45°/0° measurement geometry. The field of view was 2.7 × 2 mm, giving a resolution corresponding to 2μm/pixel. All images are first corrected for dark current and CCD gain, as described in Sec. 5.5. After calibration against a white reference, the pixel values for the grayscale images correspond to reflectance values.

The conversion to CIEXYZ color space was made by polynomial regression from RGB images, using characterization functions individually derived for each paper grade. The third order polynomial according to Eq. 7.5 was used as approximation function in regression.

8.4 Experimental results

The ideas proposed in this chapter were first tested in a preliminary study, using already available test prints (Nyström, 2008a). The results from this study are given in Appendix C. Since the results were promising, the experiments were repeated using new test prints, including more intermediate tone levels, and various print resolutions.

The preliminary experiments, given in Appendix C, also include different methods for estimating the physical dot area. Beside the midpoint between the peaks and the line scans from micro-scale images, the physical dot area was also computed using micro-scale images captured in transmission mode. Since the best results were obtained using line scans, this method has been adopted here.

8.4.1 Reflectance values

Tables 8.1 and 8.2 list the RMS errors between the measured mean reflectance and the predicted reflectance values, for all printed halftone patches. The expanded Murray-Davies model (Eq. 8.1) has been used with \( R_{d}(a) \) and \( A(a) \) computed from grayscale histograms. The physical dot area coverage, \( a_{\text{phys}} \), has been estimated using line scans from micro-scale images. Computations with the ordinary Murray-Davies (MD) model and the Yule-Nielsen (YN) model are also included for comparison. For Yule-Nielsen, the optimal \( n \)-value has been used, computed individually for each print. The best results are marked in bold in the tables.

<table>
<thead>
<tr>
<th>Halftone</th>
<th>Resolution</th>
<th>MD</th>
<th>YN</th>
<th>( n )</th>
<th>Exp. MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>65 lpi</td>
<td>0.0324</td>
<td>0.0082</td>
<td>1.64</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>100 lpi</td>
<td>0.0313</td>
<td>0.0060</td>
<td>1.54</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td>135 lpi</td>
<td>0.0362</td>
<td>0.0062</td>
<td>1.96</td>
<td>0.0051</td>
</tr>
<tr>
<td></td>
<td>170 lpi</td>
<td>0.0310</td>
<td>0.0058</td>
<td>1.68</td>
<td>0.0066</td>
</tr>
<tr>
<td>FM</td>
<td>150 dpi</td>
<td>0.0195</td>
<td>0.0052</td>
<td>1.28</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>300 dpi</td>
<td>0.0249</td>
<td>0.0064</td>
<td>1.37</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>600 dpi</td>
<td>0.0236</td>
<td>0.0045</td>
<td>1.49</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>1200 dpi</td>
<td>0.0266</td>
<td>0.0051</td>
<td>1.54</td>
<td>0.0029</td>
</tr>
</tbody>
</table>

*Table 8.1. RMS errors between measured and predicted mean reflectance values for prints on the glossy paper.*
Table 8.2. RMS errors between measured and predicted mean reflectance values for prints on the matt paper.

<table>
<thead>
<tr>
<th>Halftone</th>
<th>Resolution</th>
<th>MD</th>
<th>YN</th>
<th>n</th>
<th>Exp. MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>65 lpi</td>
<td>0.0245</td>
<td>0.0071</td>
<td>1.38</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>100 lpi</td>
<td>0.0258</td>
<td>0.0073</td>
<td>1.45</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>135 lpi</td>
<td>0.0394</td>
<td>0.0074</td>
<td>1.99</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td>170 lpi</td>
<td>0.0247</td>
<td><strong>0.0052</strong></td>
<td>1.62</td>
<td>0.0065</td>
</tr>
<tr>
<td>FM</td>
<td>150 dpi</td>
<td>0.0286</td>
<td>0.0094</td>
<td>1.39</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>300 dpi</td>
<td>0.0222</td>
<td>0.0051</td>
<td>1.34</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>600 dpi</td>
<td>0.0244</td>
<td>0.0051</td>
<td>1.45</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>1200 dpi</td>
<td>0.0216</td>
<td>0.0080</td>
<td>1.35</td>
<td>0.0030</td>
</tr>
</tbody>
</table>

As expedited, the expanded Murray-Davis clearly outperforms the ordinary Murray-Davis. The results when employing the expanded Murray-Davies model are also generally better than the Yule-Nielsen model, using optimal n-factors. The estimated physical dot area, $a_{phy}$, has also been used for the MD and YN computations. Using the nominal dot area, $a_0$, greatly increases the prediction errors for the MD-model and gives n-factors much larger than 2. Note that in Tables 8.1 and 8.2, all Yule-Nielsen n-factors belong to the physically meaningful range, $1 \leq n \leq 2$, due to the use of the correct physical dot area.

Figure 8.6 displays the measured reflectance values compared to the predicted reflectance using the model, and the reflection for the ink, $R_p(a)$, and the paper, $R_i(a)$, estimated from the histograms for AM prints of various screen frequency on the matt paper. The corresponding values for the FM prints are given in Fig. 8.7.

![Figure 8.6. Measured mean reflectance (solid), predicted reflectance (+), $R_p(a)$ (dashed) and $R_i(a)$ (dotted). AM prints of various screen frequency on the matt paper.](image-url)
It is clear from Figs. 8.6 and 8.7 that the micro-reflectance for the ink and the paper, $R_i(a)$ and $R_p(a)$, varies with the physical dot area coverage, $a_{phy}$. The variation of the reflectance of the paper, $R_p(a)$, is generally larger than that of the ink $R_i(a)$. The variation in reflectance values for both ink and paper increases with increasing screen frequency (AM prints) and print resolution (FM prints), and thus with decreasing dot size.

For the AM prints, the largest variations occur when $a_{phy}$ approaches 0 and 1, i.e. when size of the dots and the paper between the dots become small. For the FM prints, the variation of the ink and paper reflectance is more linear with the dot area coverage. This is because the effect of the light scattering is closely related to the dot size, giving $R_i(a)$ and $R_p(a)$ that is non-linear with respect to $a$ for the AM prints, with varying dot size, and a more linear behavior for the FM prints, using constant dot size. The experimental findings in Figs. 8.6 and 8.7 demonstrate a close resemblance to previous simulations using point spread functions (Yang, et al. 2001a).

Note that for the 1200 dpi FM print, the reflectance for ink and paper, $R_i(a)$ and $R_p(a)$, are close to the mean reflectance, $R$. Clearly, for this high resolution, when the dot size becomes small in relation to the scattering length, the optical behavior starts to resemble a continuous tone system, with ink at varying concentration. It is obvious that the MD and YN models, using only the reflectance values for the full tone and the bare paper, must be fundamentally wrong in this case.

### 8.4.2 Modeling $R_p(a)$ and $R_i(a)$

In previous attempts to model the functions for $R_i(a)$ and $R_p(a)$, it was first suggested that functions varied between a common reflectance limit (Arney, et al., 1995a). The reflectance corresponding to $R_0t$, where $R_0$ is the bulk reflectance and $t$ is the ink transmittance was said to imposed a common limit for $R_i(a)$ and $R_p(a)$. The following
functions were then proposed (found by trial and error) to model $R_i(a)$ and $R_p(a)$ (Arney, et al., 1995a):

$$R_i(a) = R_0 t \left(1 - (1 - t) a^w \right)$$  (8.6)

$$R_p(a_p) = R_0 \left(1 - (1 - t)(1 - a_p^w) \right)$$  (8.7)

where $a_p = (1 - a)$ is the area coverage for the paper and $w$ is a newly introduced empirical parameter, selected to provide best fit to experimental data. In the equations, $R_i$ is allowed to vary between $R_0$ and $R_0 t^2$, and $R_p$ is allowed to vary between $R_0$ and $R_0 t$ with increasing dot area coverage, $a$.

However, when increasing the print resolution, it was discovered that the assumption of the common reflectance limit, $R_0 t$, was no longer valid. This is also apparent from Figs. 8.6 and 8.7, for the AM prints of 135 and 170 lpi screen frequencies, and from the 600 dpi FM print. A new extension was then proposed, allowing both $R_i(a)$ and $R_p(a)$ to vary between the limits $R_0$ and $R_0 t$ (Arney, et al., 1995a):

$$R_i(a) = R_0 \left(1 - (1 - t) a^w \right) \left(1 - (1 - t) a^v \right)$$  (8.8)

$$R_p(a_p) = R_0 \left(1 - (1 - t)(1 - a_p^w) \right) \left(1 - (1 - t)(1 - a_p^v) \right)$$  (8.9)

where the newly introduced power factor, $v$, is an empirical factor said to be related to the softness of the dot edges, while $w$ represents the effect of light scattering.

Figure 8.8 displays the results when the parameters $w$ and $v$ are derived using non-linear optimization software to provide the best fit to experimental data, minimizing the RMS error of the predicted and measured mean reflectance.

![Figure 8.8. Mean reflectance: measured (solid) and predicted (+). $R_i(a)$: measured (dashed) and modeled (x). $R_p(a)$: measured (dotted) and modeled (*).](image-url)
It is clear from Fig. 8.8, that even though the predicted mean reflectance fits the measured reflectance, producing small RMS errors, the modeled functions for $R_i(a)$ and $R_p(a)$, do not show a close resemblance to the experimental data (possibly with the exception of the 600 dpi FM print). If $w$ and $v$ are instead optimized to fit the experimentally measured $R_i(a)$ and $R_p(a)$, the RMS error of the predicted mean reflectance increases dramatically. It is simply not possible to obtain a good fit to experimental data using the parameters $w$ and $v$, equal for both $R_i(a)$ and $R_p(a)$, according to Eqs. 8.8 and 8.9. The functions, proposed after experiments on low resolution digital prints, are clearly not general enough to deal with offset prints.

Furthermore, it is questionable if one can model the light scattering and dot shape effects separately, using the parameters $w$ and $v$, according to Eqs. 8.8 and 8.9. Based on only the mean reflectance, $R$, the physical dot area, $a_{phys}$, and the reflectance of the ink and paper, $R_i(a)$ and $R_p(a)$, it should not be possible to unambiguously separate the scattering and dot hardness effects. Possibly, the softness of the halftone dots could instead be derived from the reflectance histograms from the micro-scale images, as discussed in Sec. 8.2.1, Fig. 8.2.

### 8.4.3 Tristimulus values

Table 8.3 lists the results for the AM halftones, when the expanded Murray-Davies model is used to predict tristimulus values, based on the one-dimensional color distributions projected from CIEXYZ color space. The tristimulus values for the ink and paper are computed from the one-dimensional color distributions and the midpoint between the peaks is used to define the threshold. The results are evaluated using the Euclidian distance in CIEXYZ color space, $\Delta XYZ$, as well as the CIE 1976 color difference, $\Delta E_{ab}$, between the measured and predicted colorimetric values. The corresponding results for the FM halftones are given in Tab. 8.4.

<table>
<thead>
<tr>
<th>Halftone</th>
<th>Color</th>
<th>$\Delta XYZ_{max}$</th>
<th>$\Delta XYZ_{mean}$</th>
<th>$\Delta E_{ab max}$</th>
<th>$\Delta E_{ab mean}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM 65 lpi</td>
<td>Black</td>
<td>2.53</td>
<td>1.26</td>
<td>2.49</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>3.62</td>
<td>2.05</td>
<td>3.21</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>1.79</td>
<td>1.10</td>
<td>4.12</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>2.41</td>
<td>1.02</td>
<td>2.96</td>
<td>1.16</td>
</tr>
<tr>
<td>AM 100 lpi</td>
<td>Black</td>
<td>0.86</td>
<td>0.53</td>
<td>1.88</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>3.02</td>
<td>1.75</td>
<td>3.57</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>2.24</td>
<td>1.25</td>
<td>4.47</td>
<td>2.89</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>2.94</td>
<td>1.42</td>
<td>2.48</td>
<td>1.38</td>
</tr>
<tr>
<td>AM 135 lpi</td>
<td>Black</td>
<td>2.18</td>
<td>1.06</td>
<td>2.61</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>2.03</td>
<td>1.09</td>
<td>3.78</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>1.99</td>
<td>1.03</td>
<td>4.40</td>
<td>2.56</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>2.02</td>
<td>1.06</td>
<td>2.67</td>
<td>1.35</td>
</tr>
<tr>
<td>AM 170 lpi</td>
<td>Black</td>
<td>1.45</td>
<td>0.50</td>
<td>2.04</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>3.57</td>
<td>1.59</td>
<td>5.90</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>2.64</td>
<td>1.40</td>
<td>6.66</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>2.82</td>
<td>1.62</td>
<td>5.07</td>
<td>1.61</td>
</tr>
</tbody>
</table>
Table 8.3. Estimation errors between measured and predicted tristimulus values for paper Gloss, FM halftones.

<table>
<thead>
<tr>
<th>Halftone</th>
<th>Color</th>
<th>ΔXYZ max</th>
<th>ΔXYZ mean</th>
<th>ΔE&lt;sub&gt;ab&lt;/sub&gt; max</th>
<th>ΔE&lt;sub&gt;ab&lt;/sub&gt; mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>FM 150 dpi</td>
<td>Black</td>
<td>1.07</td>
<td>0.57</td>
<td>2.03</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>2.31</td>
<td>1.41</td>
<td>2.89</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>1.40</td>
<td>0.71</td>
<td>2.97</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>1.86</td>
<td>0.94</td>
<td>2.26</td>
<td>1.05</td>
</tr>
<tr>
<td>FM 300 dpi</td>
<td>Black</td>
<td>0.76</td>
<td>0.38</td>
<td>2.19</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>2.39</td>
<td>1.47</td>
<td>4.10</td>
<td>2.74</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>2.49</td>
<td>1.20</td>
<td>5.56</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>3.23</td>
<td>1.31</td>
<td>2.41</td>
<td>1.27</td>
</tr>
<tr>
<td>FM 600 dpi</td>
<td>Black</td>
<td>2.65</td>
<td>1.08</td>
<td>2.93</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>3.43</td>
<td>2.01</td>
<td>7.55</td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>4.48</td>
<td>2.29</td>
<td>9.30</td>
<td>5.28</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>2.49</td>
<td>1.12</td>
<td>3.24</td>
<td>1.72</td>
</tr>
<tr>
<td>FM 1200 dpi</td>
<td>Black</td>
<td>1.43</td>
<td>0.64</td>
<td>1.98</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>3.67</td>
<td>1.98</td>
<td>8.18</td>
<td>4.87</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>6.19</td>
<td>3.28</td>
<td>9.19</td>
<td>5.89</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>8.11</td>
<td>1.72</td>
<td>4.49</td>
<td>1.63</td>
</tr>
</tbody>
</table>

Note that the prediction errors in Tables 8.2 and 8.3 are computed as the difference between measurements using a spectrophotometer, and the predicted values computed from the micro-scale images. Hence, the color differences in Tables 8.3 and 8.4 also include the errors introduced in the conversion from RGB to XYZ values. Considering this, one must say that the results are reasonably good and that the method of using projections down to one-dimensional color distributions works well.

A closer inspection of the outliers, producing the largest estimation errors, reveals that these are generally caused by unrepresentative measures of the bare paper or the full-tone ink. Since the projection vectors are computed from the XYZ values of the bare paper and the full-tone, errors are introduced when any of these measurements are not representative, e.g. when the paper contains any misprinted ink.

Figure 8.9 displays the measured and predicted tristimulus values for black prints, using AM halftones of various screen frequencies, as well as how the tristimulus values for the ink and the paper, XYZ<sub>i</sub> and XYZ<sub>p</sub>, varies with the dot area coverage, a<sub>phy</sub>. Examples of measured and predicted tristimulus values for paper gloss, using 135 lpi AM halftones, are displayed in Fig. 8.10, for the cyan, magenta, yellow and black colorants.
8.5 Summary and discussion

This chapter focuses on high resolution analysis of halftone prints, especially how the reflectance of the printed dots and the paper between them vary with respect to the dot area fraction. The reflectance values of the ink, $R_i(a)$, and paper, $R_p(a)$, are estimated for
each area coverage, using histogram data for both grayscale and color halftones. The physical dot area fraction, including the physical dot gain, has been estimated using line scans in the micro-scale images. Alternative methods for estimating the physical dot area, thus separating the physical dot gain from the optical, will be further discussed and evaluated in Chap. 9.

For the color case, predicting tristimulus values, projections down to one-dimensional color distributions are used to overcome the difficulties when using the full 3D histograms to compute \(XYZ(a), X'YZ(a)\) and \(a_{phy}\). The projection vectors are computed from the centers of gravity for the clusters corresponding to the bare paper and to the fulltone ink. Experiments on using principle component analysis, PCA, defining the projection by the most significant eigenvector from the covariance matrix of all XYZ coordinates in the prints, has shown to result in only small improvements of the results. It provides a more robust method, compared to that using only the coordinates for the full tone and bare paper, for which the measurements must be ensured to be representative for the real values of the ink and paper. However, the simplicity of using only measurements for the bare paper and the fulltone inks, i.e. the Neugebauer primaries, for the projection vectors makes the method a more attractive alternative.

The validity of the expanded Murray-Davies model, utilizing the varying reflectance \(R_i(a)\) and \(R_p(a)\) to account for the non-linearity caused by optical dot gain, has been evaluated. It is shown that the model is valid, even for prints of considerably higher print resolution than those used in previous works. The prediction errors of the model were smaller than the results when employing the Yule-Nielsen model, using optimal \(n\)-values. However, unlike the Yule-Nielsen model, the extended Murray-Davies model preserves the linear additivity of reflectance, thus providing a better physical description of halftone color reproduction.

The variation in reflectance values for the ink and paper, \(R_i(a)\) and \(R_p(a)\), increases with increasing print resolution, and thus with decreasing dot size. For the AM prints, the largest variations occur when \(a_{phy}\) approaches 0 and 1, i.e. when size of the dots and the paper between the dots, respectively, becomes small. For the FM prints, using constant dot size, the variation of the ink and paper micro-reflectance is more linear with the dot area coverage. For high resolution 1200 dpi FM print, the reflectance for ink and paper, \(R_i(a)\) and \(R_p(a)\), are close to the mean reflectance, \(R\), acting almost like a continuous tone system, using ink at varying ink concentrations. From these findings, it is obvious that the Murray-Davies and Yule-Nielsen models, using only the reflectance values for the full tone and the bare paper must be fundamentally wrong.

A previously proposed model for \(R_i(a)\) and \(R_p(a)\), using the empirical parameters \(w\) and \(v\), was applied to experimental data. When fitted against the mean reflectance, i.e. the macroscopic measurement, the model produces good predictions. However, when examining the modeled functions for \(R_i(a)\) and \(R_p(a)\) using microscopic measurements, they do not provide a good fit to the experimental data. By using the previously proposed model, incorporating the parameters \(w\) and \(v\), which are equal for both \(R_i(a)\) and \(R_p(a)\), it was not possible to obtain a good fit to experimental data.

Future work should be devoted to finding better models for \(R_i(a)\) and \(R_p(a)\), i.e. how the reflectance of the ink and the paper between the dots varies with dot area coverage. Especially, focus should be on relating these functions to physical properties of the paper substrate and to properties of the halftoning employed, such as the halftone geometry, print resolution and the hardness of the printed dots. Furthermore, the method
should be further developed to handle the case of multiple, overlapping, colorants, i.e. into an expanded Neugebauer model, incorporating the varying reflectance of the ink and paper.

The combination of micro-scale imaging with colorimetric and spectral accuracy, provide a powerful tool for future development of a deeper understanding of the complex process of halftone color printing. An understanding, crucial in the development of more sophisticated models of halftone color printing, which can benefit the print quality.
Chapter 9

Physical got gain and print resolution

9.1 Introduction
9.2 Methodology
9.3 Experimental setup
9.4 Experimental results
9.5 Summary and discussion
9.1 Introduction

Recall from Chap. 4 that the tone value increase referred to as dot gain in halftone printing actually encompasses two fundamentally different phenomena. Physical dot gain is closely related to the printing process and refers to the fact that the size of the printed halftone dots differs from their nominal size. Optical dot gain originates from light scattering inside the substrate, causing light exchanges between different chromatic areas, and making the dot appear bigger than its physical size when it is perceived/measured. The effect of optical dot gain depends on the ratio of the length of the lateral light scattering within the substrate to the size of the printed halftone dots. When the dot size becomes small in relation to the lateral scattering length, the optical dot gain will increase. The lateral scattering length in paper is typically in the order of 0.1 mm (Arney, et al., 1996a; Rogers, 1998b).

Due to their different intrinsic nature, physical and optical dot gains need to be treated separately in order to accurately model the outcome of halftone prints. The methodology discussed in Chap. 8, as well as the models for optical dot gain presented in Chap. 4, all require the use of the correct physical dot area coverage, instead of the nominal one. However, in reflection measurements of halftone prints the physical and optical dot gains always co-exist, making the separation of one type from another a difficult task.

In this chapter different methods to determine the physical dot gain in halftone prints, thus separating it from the optical dot gain, are investigated. Physical dot gain is computed in three different ways: from spectral reflectance measurements, from transmission scans using a flat-bed scanner, and from microscopic images of the prints, respectively. Further, the relation between the physical dot gain and the size of the halftone dots is investigated by using frequency modulated halftones of various print resolutions and thus different dot sizes. The validity of the complete-light-scattering approximation is also evaluated, by using screen resolutions where the printed dot size is no longer small in relation to the lateral scattering of the light.
9.2 Methodology

9.2.1 Physical dot gain from spectral reflectance

When the size of the printed halftone dots is small enough in relation to the lateral scattering length, complete light scattering within the substrate is considered, also known as complete photon diffusion (Rogers, 2000b). In such cases, the reflectance of a halftone print can be described by the Clapper-Yule model, modeling optical dot gain by taking into account lateral scattering within the substrate as well as multiple internal reflections (see Sec. 4.6 for details):

\[ R(a) = K_r + (1 - 1)(1 - r_i) \frac{R_0(1-a) + at}{1-r_iR_0(1-a + at^2)} \]  (9.1)

In the equation, \( a \) is the physical dot area coverage, \( t \) is the spectral transmittance of the ink, \( R_0 \) is the bulk reflectance of the paper substrate, \( r_s \) and \( r_i \) are the Fresnel reflectance values of external and internal reflections. Both the specular reflection, \( r_s \) and the internal reflection, \( r_i \), depend on the refraction indices of the air (\( n=1 \)) and the paper. In the current study the general assumption that the paper and ink share the common refraction index \( n=1.53 \) is adopted, giving \( r_s=0.054 \) and \( r_i=0.641 \), for light at an incident angle of 45° (Hersch, et al., 2005). Note that when the refraction index for paper/ink is set to 1, the specular reflection within the substrate is ignored and Eq. 9.1 reduces to the Yule-Nielsen formula, with the \( n \)-factor set to 2 for complete light scattering (Eq. 4.4).

The factor \( K \), ranging from 0 to 1, gives the fraction of the specular surface reflection that reaches the detector and depends on the measurement geometry. Ideally, measurements should be performed using an integrated sphere, i.e. using the \( d/0° \) geometry where there is no specular reflection. However, it has been shown that for spectrophotometers using the 45°/0° measurement geometry, the specular component can be discarded in measurements on paper with refraction index \( n=1.53 \) (Hersch, et al., 2005). In the current study, all reflectance measurements are performed using the 45°/0° measurement geometry and the specular reflection component has been discarded, i.e. being set as \( K=0 \).

The bulk reflection of the paper, \( R_0=R(0) \), is computed from the measured reflectance, \( R_p \), of a single sheet of paper with a black background, by setting \( a=0 \) in Eq. 9.1:

\[ R_0 = \frac{R_p}{R_p r_i + (1-r_s)(1-r_i)} \]  (9.2)

Similarly, the ink transmittance, \( t \), is estimated from the reflectance measurement of a full tone print, \( R_1=R(1) \), by setting \( a=1 \) in Eq. 9.1:
\[
\begin{align*}
t = \frac{R_i}{\sqrt{R_i R_0 r_i + R_0 (1 - r_i)(1 - r_i)}} \tag{9.3}
\end{align*}
\]

With the measured spectrum of a test patch, one may obtain the physical dot area coverage, \(a_{phy}\), by fitting the computed spectrum to the measurement, using \(a\) as the single fitting-parameter in Eq. 9.1. The physical dot gain, \(\Delta a_{phy}\), is then simply given by the difference between the physical dot area, \(a_{phy}\), and the nominal one, \(a_0\), as:

\[
\Delta a_{phy} = a_{phy} - a_0 \tag{9.4}
\]

### 9.2.2 Physical dot gain from transmission scans

An alternative method to compute the physical dot gain is to use transmission images of the printed halftones where no optical dot gain is present. The relationship between the physical dot area and the transmission intensity has been derived, taking into account the light scattering within the substrate and the internal and external surface reflections (Yang & Lundström, 2007). Under the same assumptions as the Clapper-Yule formula, the transmittance of the print can be described by:

\[
T(a) = (1 - r_i)(1 - r_i) \frac{(1 - a + at)T_0}{1 - r_i R_0 (1 - a + at^2)} \tag{9.5}
\]

where \(T_0\) is the paper bulk transmittance.

For the bare paper, the transmittance \(T_p = T(0)\) is given by setting \(a = 0\) in Eq. 9.5:

\[
T_p = T(0) = (1 - r_i)(1 - r_i) \frac{T_0}{1 - r_i R_0} \tag{9.6}
\]

For transmission measurements of halftone prints with area coverage \(a\), the measured transmittance intensity, \(I_t(a)\), is given by:

\[
I_t(a) = I_{0t} T(a) \tag{9.7}
\]

where \(T(a)\) is the transmission of the halftone patch, as given in Eq. 9.5, and \(I_{0t}\) is the intensity of the illuminating light source. For the unprinted paper, the corresponding intensity is given by:

\[
I_t(0) = I_{0t} T_p \tag{9.8}
\]

where \(T_p\) is the transmittance of the bare paper, as given in Eq. 9.6.

Combining Eq. 9.8 with Eq. 9.7, gives:
Finally, by replacing $T(a)$ and $T_p$ in Eq. 9.9 by their corresponding expression according to Eqs. 9.5 and 9.6, one obtains:

$$\frac{(1-a + at)}{1-r_i R_0 (1-a + at^2)} = \frac{I_a(a)}{I_a(0)} \frac{1}{1-r_i R_0}$$

(9.10)

After deriving the paper bulk reflection, $R_0$, and the ink transmittance, $t$, the physical dot area coverage, $a_{phy}$, can be computed from transmission images of the prints according to Eq. 9.10 (Yang & Lundström, 2007). The corresponding physical dot gain is then given by Eq. 9.4.

### 9.2.3 Physical dot gain from micro-scale images

Another possibility to estimate the physical dot area coverage is to use micro-scale images of the printed halftones, as described in Sec. 8.2.2. From reflectance histograms computed from the micro-scale images, the physical dot area can be computed according to Eq. 8.5, integrating the histogram up to a threshold, $R_t$. The reflectance threshold value, $R_t$, representing the boundary between the dot and the paper, is defined as the region of maximum rate of change in reflectance values $dR/dx$, see Fig. 8.4.

### 9.3 Experimental setup

The test prints used in this study consisted of offset prints on coated and uncoated paper grades. To study the dependency of the dot size, test patches of varying print resolutions, 150, 300, 600 and 1200 dpi, were generated using frequency modulated (FM) halftoning. The halftone algorithm employed uses an iterative method for optimal dot placement (Gooran, 2001). The nominal dot area coverage of the patches are 2, 5, 10, 15, ..., 90, 95, 98 and 100% respectively, and all the patches were printed using the same plate. The test charts were printed with a commercial 4-color offset press (Heidelberg GT052) on both coated and uncoated substrates.

The spectral reflectance values of the printed test charts were measured using a Gretag Machbeth Spectrolino spectrophotometer. The scanner used is a FujiFilm FineScan 7250 flatbed scanner. The printed test charts were scanned in the resolution 300 ppi in transmission mode, and stored as 16-bit tiff images. The transmission intensities were then computed as the mean value over the area for each color patch, from the transmission images. Since the color channel in the image that is complementary to the printed color (e.g. the green channel for a magenta print) is most sensitive to variations, it has been used when computing the transmittance intensities, $I_a(a)$.

The micro-scale images of the printed test charts have been captured using the 45°/0° measurement geometry. The field of view was 2.7 × 2 mm, giving a resolution...
corresponding to 2μm/pixel. All images are calibrated against a white reference and corrected for dark current and CCD gain. For details on the experimental setup refer to Sec. 8.3.

9.4 Experimental results

9.4.1 Estimations of physical dot gain

Figure 9.3 displays examples of computed physical dot gain for magenta prints of different print resolutions on coated paper. Solid lines correspond to physical dot gain computed from spectral reflectance measurements, dashed lines from the transmission scans, and the dotted lines from micro-scale images. The corresponding results for uncoated paper are displayed in Fig. 9.4.

![Figure 9.3: Physical dot gain of magenta prints on coated paper. Solid lines: estimated from reflectance spectra. Dashed lines: estimated from transmission scans. Dotted lines: estimated from microscopic images of the prints.](image)

Clearly, the physical dot gain computed using the three different methods agree fairly well (Figs. 9.3 and 9.4). It is obvious that the method using micro-scale images fails to predict the physical dot gain for the dark tones in the 1200 dpi prints. The reason is that with such small dots, it is difficult to distinguish the inked from the non-inked areas when ink coverage is high. The combination of physical and optical dot gain in the prints result in images where the paper between the dots is no longer visible and the threshold, $R_t$, cannot be accurately computed from the images (see App. C).
The method using transmission scans gives a somewhat better result for the coated paper grade compared to the uncoated paper, where the paper fibers are more clearly visible and causes some instability. In previous work (Nyström, 2008a), attempts were made to use transmission images captured in micro-scale. However, in such high magnification, paper fibers become clearly visible, while the imaged area is too small to be statistically significant. Consequently, the computed intensity values become unstable and depend strongly on the local structure of the fibers captured in the image, which varies over the substrate area. Using transmission scans from a flatbed scanner averaging the intensity over a larger area clearly produces more stable and reliable results.

### 9.4.2 Spectral estimations

Figure 9.5 displays the computed spectra for coated paper using Eq. 9.1 with the estimated physical dot percentage, $a_{phy}$, compared to the experimentally measured spectra, for 20, 40 and 60% nominal dot percentage, $a_0$. The corresponding results for the uncoated paper are displayed in fig. 9.6. Clearly, the estimated spectra are in very good agreement with the corresponding measurements. The middle tones are used for comparison because they are far from the constraints, 0 and 100%, and are thus most difficult to predict. The mean and maximum spectral RMS difference between the measured and predicted spectra for all color patches are presented in Tab. 9.1 together with the corresponding CIE1976 color differences.

The good agreement of the spectral predictions even for the lower print resolutions further shows that the assumption of complete light scattering holds surprisingly well, even for a resolution as low as 150 dpi. Recall that the basic assumption in the Clapper Yule model is that the lateral light propagation inside the paper bulk is large compared to the size of the halftone dots, and thus that the
probability of light exiting a specific colorant is equal to the colorant coverage (Hersch, et al., 2005). A 150 dpi halftone gives a dot size of about 0.17 mm, i.e. halftone dots that are no longer small in relation to the length of the lateral scattering of light (typically about 0.1 mm). Nevertheless, the results when employing the Clapper-Yule formula according Eq. 9.1, are still reasonably good.

Figure 9.5. Computed (dots) and measured (solid line) reflectance spectra for magenta prints on coated paper, with 20, 40 and 60% nominal dot percentages.

Figure 9.6. Computed (dots) and measured (solid line) reflectance spectra for magenta prints on uncoated paper, with 20, 40 and 60% nominal dot percentages.
Table 9.1. Spectral RMS difference and CIE 1976 color difference between predicted and measured spectra.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Resolution</th>
<th>RMS $\times 10^{-4}$</th>
<th>$\Delta E_{ab}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Mean</td>
<td>Max</td>
</tr>
<tr>
<td>Coated</td>
<td>150 dpi</td>
<td>4.51</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>300 dpi</td>
<td>1.46</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>600 dpi</td>
<td>0.57</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>1200 dpi</td>
<td>1.03</td>
<td>0.35</td>
</tr>
<tr>
<td>Uncoated</td>
<td>150 dpi</td>
<td>2.51</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>300 dpi</td>
<td>0.52</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>600 dpi</td>
<td>0.86</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>1200 dpi</td>
<td>2.44</td>
<td>0.78</td>
</tr>
</tbody>
</table>

9.4.3 Dot gain and print resolution

From the experimental results, it is also obvious that the physical dot gain demonstrates a clear correlation with the print resolution, namely the higher the print resolution (and the smaller the dot), the greater the physical dot gain. When studying the physical dot gain for the different print resolutions, it becomes clear that the increase in physical dot gain is proportional to the increase in resolution. Increasing the print resolution from 150 to 300 dpi and from 300 to 600 dpi, doubles the physical dot gain, while an increase from 600 to 1200 dpi causes about 50-60% gain in physical dot gain. A comparison of the physical dot gain for the different print resolutions on coated paper is displayed in Fig. 9.7.

![Figure 9.7. Physical dot gain of magenta prints on coated paper estimated from reflectance spectra, for 150, 300, 600 and 1200 dpi.](image)

For simplicity, we explain the observation with FM screening and square halftone dots, which have been used in the test prints employed in this study. It is easy to prove
mathematically that the argument also holds for round dots. With FM screening, all the
dots are identical and different tone values are achieved by applying different numbers
of dots. For a nominal tone value, \( a_0 \), the dot area coverage can be expressed as:

\[
a_0 = N d_0^2
\]  

(9.11)

where \( d_0 \) is the nominal side-length of a single halftone dot and \( N \) is the number of dots.

According to the definition, the print resolution, \( l \), is inversely proportional to the side
length, i.e.

\[
l = \frac{1}{d_0}
\]  

(9.12)

In the case of a physical dot gain with a side-length extension, \( \Delta d \), the real tone value, \( a \),
becomes:

\[
a = N(d_0 + \Delta d)^2
\]

\[
\approx N d_0^2 + 2 N d_0 \Delta d
\]

\[
= a_0 + \Delta a_{phy}
\]  

(9.13)

where the physical dot gain, \( \Delta a_{phy} \), corresponds to:

\[
\Delta a_{phy} = 2 N d_0 \Delta d
\]

\[
= 2 a_0 \Delta d
\]  

(9.14)

Equation 9.14 indicates that the physical dot gain, \( \Delta a_{phy} \), is proportional to the
nominal tone value, \( a_0 \), the print resolution, \( l \), and the side length extension, \( \Delta d \). The
relation of the physical dot gains for two halftone prints of resolutions \( l_1 \) and \( l_2 \), printed
with the same nominal dot area, \( a_0 \), will be:

\[
\frac{\Delta a_{phy1}}{\Delta a_{phy2}} = \frac{2 a_0 l_1 \Delta d_1}{2 a_0 l_2 \Delta d_2}
\]

\[
= \frac{l_1 \Delta d_1}{l_2 \Delta d_2}
\]  

(9.15)

Equation 9.15 indicates that the ratio of the physical dot gains is proportional to
that of the screen resolutions, as demonstrated in Fig. 9.7. The relative side-length
extensions, \( \Delta d_1/\Delta d_2 \), serves as a modulate factor.
9.4.4 Physical and optical dot gain

Figure 9.8 depicts the physical dot gain, $\Delta a_{\text{phy}}$, plotted together with the optical dot gain, $\Delta a_{\text{opt}}$, and the resulting total dot gain, $\Delta a_{\text{tot}}$, for coated paper using the various print resolutions. The effective dot area, $a_{\text{eff}}$, has first been computed using the Murray-Davies formula, according to Eq. 4.3. Because of the spectral dependence of optical dot gain, the effective dot area, $a_{\text{eff}}$, was computed as the best fit to spectral reflectance data, instead of using only the wavelength of maximum light absorption in Eq. 4.3. The total dot gain, $\Delta a_{\text{tot}}$, and the optical dot gain, $\Delta a_{\text{opt}}$, are then given by:

\[
\Delta a_{\text{tot}} = a_{\text{eff}} - a_0
\]  

(9.16)

\[
\Delta a_{\text{opt}} = a_{\text{eff}} - a_{\text{phy}}
\]  

(9.17)

From Fig. 9.8 it is clear that, with the exception of the 150 dpi print where the physical dot gain is almost negligible, the optical dot gain curves are asymmetric with respect to the nominal dot area, $a_0$. For the higher print resolutions, 600 and 1200 dpi, the optical dot gain curves reach their maxima already at nominal dot areas about 20-25%. The simple reason is that the optical dot gain responds to the physical dot area, including the physical dot gain, not to the nominal dot area. In Fig. 9.9, where the optical dot gain is plotted against the estimated physical dot area, $a_{\text{phy}}$, the curves are clearly symmetric around the maxima at the mid tone $a_{\text{phy}}=50\%$.

Figure 9.8. Total (solid line), physical (dashed line) and optical (dotted line) dot gain for magenta prints on coated paper, using the resolutions 150, 300, 600 and 1200 dpi.
The symmetric form of the optical dot gain vs. the physical dot area around its maximum at \(a_{\text{phy}}=50\%\) (Fig. 9.9) is a further indication to the applicability of complete-light-scattering assumption, even for print resolution of 150 dpi. As earlier reported (Yang et al, 2001a), in the case of complete light scattering, the optical dot gain can be expressed as:

\[
\Delta R_{\text{opt}} = P(1-t)^2 a_{\text{phy}}(1-a_{\text{phy}})
\]  

(9.18)

where \(P\) is the quantity describing the probability of light exchanges between the imaged and non-imaged areas and \(t\) is the transmittance of the ink. From Fig. 9.9, one clearly observes that the light-exchange probability, \(P\), intensifies with increasing screening-resolution, or decreasing dot sizes. When the ratio of the length of the lateral scattering of light compared to the size of the printed halftone dots becomes larger, the probability for light exchange between the inked and non-inked areas increases, and so does the effect of the optical dot gain. The light-exchange probability, \(P\), can be computed if the point spread function (PSF) of the paper substrate is known (Yang et al, 2001a). This was demonstrated (Yang et al, 2001b) for multicolor prints by employing Gaussian type of PSFs.

The results from Figs. 9.8 and 9.9 clearly illustrate the importance of using the correct physical dot area, \(a_{\text{phy}}\), instead of the nominal one, \(a_0\), in all attempts of describing or modeling optical dot gain. Thus, accurate methods for separating the physical dot gain from the optical dot are essential in all models coping with halftone printing.
9.5 Summary and discussion

In this chapter three different approaches to determining the physical dot gain in halftone prints, thus separating it from the optical dot gain, have been evaluated. The different methods use spectral reflectance measurements, transmission scans and microscopic images of halftone prints, respectively. The three different methods to estimate the physical dot gain produces similar results, and the good correspondence to the experimental data in terms of spectral estimations confirm the accuracy of the methodology. However, it should be pointed out that the methods based on reflectance measurements (Eq. 9.1) and on transmission scans (Eq. 9.10), both assume that the ink transmittance is constant and independent of the dot area coverage, an assumption which may not always hold in practice.

Using line-scans in micro-scale images of the prints to estimate the physical dot gain suffered from some problems for the darker tone values in the highest print resolution, when it is no longer possible to accurately determine the edges between dot and paper. Estimating physical dot gain using transmission scans gives a somewhat better result for the coated paper grade compared to uncoated, where the visible paper fibers caused a minor instability. However, compared to spectral measurements using spectrophotometers or using microscopic images (where special equipment is required), using a flatbed scanner is a fast and inexpensive alternative to determining the physical dot gain.

The fundamental assumption of the Clapper-Yule model (Eq. 9.1) is complete light scattering within the substrate, i.e. halftone dots which are small in relation to the lateral scattering length (typically about 0.1 mm). Otherwise, the probability of light exiting from a specific colorant is no longer equal to the colorant area coverage (Hersch, et al., 2005). Nevertheless, the spectral estimations when applying the model were reasonably accurate even for a FM halftone print with a print resolution as low as 150 dpi, giving a dot size of about 0.17 mm.

Another observation made is that when setting the refraction index to \( n=1 \) for the ink and substrate, the resulting predicted spectra are equally good as when using \( n=1.53 \), where the internal reflections in the substrate are taken into account. Note that with \( n=1, r_f=0 \) and Eq. 9.1 reduces to the Yule-Nielsen formula with the \( n \)-factor set to 2 for complete light scattering, and Eq. 9.10 reduces to the Murray-Davies formula in terms of transmission intensity values.

To investigate the relation between physical dot gain and halftone dot size, offset prints of FM halftones with different print resolutions were examined. The experimental results revealed a clear correlation between physical dot gain and halftone dot size, with an increase in physical dot gain proportional to the increase in screen resolution. The experimental findings were discussed and explained theoretically, showing that the physical dot gain is proportional to the nominal dot area, the side-length extension and the print resolution. For a more in-depth analysis and discussion on the causes and the physics behind physical dot gain in offset prints, refer to Yang & Lundström (2007) and Yang (2008).

The optical dot gain responds to the physical dot area (including the physical dot gain) not to the nominal dot area. This was shown by the optical dot gain curves which were highly asymmetric when plotted against the nominal dot area, \( a_0 \). When plotted against the estimated physical dot area, \( a_{phy} \), the optical dot gain curves were symmetric around the mid tone \( a_{phy}=50\% \), with increasing optical dot gain for increasing print
resolutions (or decreasing halftone dot size) This clearly illustrates the importance of using the correct physical dot area, $a_{\text{phys}}$, instead of the nominal one when describing or modeling optical dot gain. Hence, accurate methods to first estimate the physical dot area, thus separating the physical dot gain from the optical, are essential in all models coping with halftone color prints.
Chapter 10

Summary and future work

10.1 Summary
10.2 Future work
10.1 Summary

The aim of the study has been high resolution analysis of halftone prints. This first requires for high quality acquisition of colorimetric and multispectral images, which has been the focus of Chapters 5 to 7. The spectral image acquisition model describing the device response to a known spectral input has been derived, including the spectral characteristics of the illumination, the color filters and the CCD camera. The spectral sensitivity function for the camera, which could not be obtained by direct measurements, has been estimated. By relating measured reflectance spectra to the corresponding device responses for a set of carefully selected training colors, the sensitivity function was derived using least squares regression. The principle eigenvector method was used to reduce the noise sensitivity and a priori knowledge on the nature of camera sensitivity was used as constraints in the regression.

To reconstruct colorimetric and spectral data from the recorded device response, two conceptually different characterization methods have been employed: model-based and empirical characterization. In the model-based approach, the spectral image acquisition model is inverted, to reconstruct spectral reflectance data from the recorded device responses. A priori knowledge on the smooth nature of spectral reflectance was utilized by representing the reconstructed spectra as linear combinations of basis functions, using Fourier basis and a set of real reflectance spectra. The results revealed that a satisfactory spectral or colorimetric accuracy could not be obtained from trichromatic images. When using multi-channel images, the improvement was significant, both in terms of smaller colorimetric differences and in the RMS spectral reconstruction error. The best overall performance was obtained by the method using spectral basis. However, if the aim is only to derive colorimetric data, the simple pseudo-inverse method is sufficient, producing reconstructions that are good metameric matches.

In the empirical approach, the spectral characteristics of the system are ignored and the mappings are derived by relating the recorded device response to colorimetric and spectral data for a set of training colors, using least squares regression techniques. Empirical characterization was used to derive mappings to the colorimetric representations CIEXYZ and CIELAB, and to reconstruct spectral reflectance. The results showed that good colorimetric accuracy could be obtained by using polynomial regression to CIEXYZ and CIELAB. The results for trichromatic images were as good as for the multi-channel images and offered the same colorimetric accuracy as for the
model-based approach. However, experiments on cross-media characterization revealed that the accuracy is limited to the case when the mappings are used for the same media and colorant from which they are derived.

To conclude the part focusing on colorimetric and multispectral imaging, the results have shown that in cases when only trichromatic imaging is available, the best method for colorimetric imaging is the empirical approach using polynomial regression. However, because of the media-dependency, this requires the characterization functions to be derived for each combination of media and colorants. For multispectral imaging, multi-channel images are required to obtain the highest accuracy. The best results were obtained with the model-based approach and multi-channel images, combined with spectral basis. The model-based approach provides the additional advantage of being general, since it is based on the spectral characteristics of the image acquisition system not on the characteristics of a set of color samples. The accuracy achieved in the model-based characterization also indicates that the image acquisition model is accurate, reflecting the real characteristics of the image acquisition system well.

After the device characterization, the imaging system has been used for high resolution analysis of halftone prints. Micro-scale images, i.e. images whose resolution is high in relation to the resolution of the halftone, allow for measurements of the individual halftone dots, as well as the paper between them. To capture the characteristics of large populations of halftone dots, reflectance histograms are computed as well as 3D histograms in CIEXYZ color space. The micro-scale measurements revealed that the reflectance for the halftone dots, as well as the paper between the dots, is not constant but varies with the dot area coverage. By incorporating the varying micro-reflectance in an expanded Murray-Davies model, the nonlinearity caused by optical dot gain can be accounted for without applying the nonphysical exponentiation of the reflectance values, as in the Yule-Nielsen model.

Further, different methods for separating the physical and optical dot gain have been evaluated using spectral reflectance measurements, transmission scans and micro-scale images, respectively. Due to their different intrinsic nature, physical and optical dot gains need to be treated separately, but the task is difficult because the physical and optical dot gains always co-exist in reflection measurements. The three different methods produced similar results and the good correspondence to the experimental data confirms the accuracy of the methodology. The relation between the physical dot gain and the halftone dot size was further investigated, revealing that the physical dot gain exhibits a clear correlation with the dot size, and that the dot gain increase is proportional to the increase in print resolution.

10.2 Future work

High resolution analysis of halftone prints offers new way to measure, characterize and understand different properties associated with halftone color printing. The flexibility of the image acquisition system, such as varying the angle of incident illumination, and using backlight for transmission imaging, opens up many possibilities for future experiments. The knowledge of the system and the high level of spectral and colorimetric accuracy gained in this work will be of great importance for future work.

The experiments in Chap. 8 revealed that the reflectance values of the ink and paper, $R_i(a)$, and $R_p(a)$, vary with dot area fraction, and that the variation is closely
related to the print resolution. When incorporating the varying micro-reflectance values
in an expanded Murray-Davies model, the prediction errors were smaller than the
corresponding Yule-Nielsen results. However, microscopic measurements revealed that
a previously proposed model for $R_i(a)$ and $R_p(a)$ could not provide a good fit to the
experimental data. Hence, future work should be devoted to developing better models
for $R_i(a)$ and $R_p(a)$, i.e. how the reflectance of the halftone dots and the paper between
the dots varies with dot area coverage. In particular, focus should be on relating these
functions to physical properties of the printing process, including the printing method,
the paper substrate and properties of the halftoning employed, such as the halftone
geometry, print resolution and the hardness of the printed dots. If these functions can be
modeled properly, the expanded Murray-Davies model could be used for predicting the
output of halftone prints without the need for micro-scale measurements to derive $R_i(a)$
and $R_p(a)$.

The expanded Murray-Davies model was successfully extended to handle color
halftones, predicting tristimulus values, but the study was so far limited to single color
prints. A natural continuation would be to develop the method to handle the case of
multiple, overlapping, colorants. The expanded Murray-Davies model would then be
extended to an expanded Neugebauer model, incorporating the varying reflectance of
the paper and the Neugebauer primaries.

Further, the high resolution analysis and measurements have so far been limited to
mean reflectance and tristimulus values. With the opportunity to acquire highly resolved
multispectral images, a natural continuation would be to study how the spectral
reflectance of the halftone dots and the paper between them vary with dot area, with the
aim of developing spectral models predicting the spectral output for color halftones.
Spectral and colorimetric measurements on a micro-scale level will also provide a
powerful tool for evaluation and verification of different models, and for estimating and
evaluating the parameters used in models such as the probability model and point spread
functions.

We believe that the possibility to accurately estimate the spectral reflectance in
each pixel of high resolution images will open up many possibilities for future research
with the focus on gaining a deeper understanding of the complex process of color
halftone reproduction. An understanding, which may lead to the development of more
accurate and sophisticated models of halftone color reproduction, which can benefit the
print quality.
Bibliography


Appendices

A  Calibration of the image acquisition system
B  Empirical characterization data
C  High resolution analysis data
Appendix A: Calibration of the image acquisition system

To obtain repeatable and reliable image data of high quality, it is necessary to know the characteristics of all the components involved in the image acquisition process. This appendix includes the details of the calibration of the image acquisition system, with the aim of ensuring stable and repeatable results.

For all the components in the system, the technical specifications are given, along with some general background and a discussion on the demands placed on each component for high-quality image acquisition. Recommendations for settings and calibration routines are given, based on measurement results and on the demands on and properties of each component.

The emphasis in this appendix is on obtaining device-dependent data in a repeatable and stable way, commonly referred to as obtaining precision in color measurements. It has formed the necessary base for the work presented in Chapters 6-7, focusing on methods on how to transform the acquired device-dependent data into device-independent color representations of high accuracy.

Spectral measurements

The spectral characterization of the illuminant and the color filters is straightforward, since it can be obtained by direct measurements. Spectral measurements are performed using the spectroradiometer PR-650 SpectraScan from Photo research. The spectrophotometer is mounted using a special holder, placing it in the same optical axis as the CCD camera. The spectral data obtained are in the interval 380 to 780 nm sampled at 4 nm intervals (Photo Research, 1999).

Except for the measurements on the transmissive plate, all measurements are performed against a diffuse white reflectance standard, using the 45°/0° measurement geometry. The reflectance standard gives 99% reflectance ± 1% over the photopic region, and is placed in a special holder placing it in the same image plane as the object, to share the same illumination characteristics.
Illumination

The illumination plays a very important role in an image acquisition system and needs to fulfill the following requirements (Hardeberg, 2001):

The geometry should ensure a good spatial uniformity over the scan area, or otherwise be corrected for. If the camera is to be used for spectrophotometric or colorimetric measurements, a lighting/viewing geometry recommended by the CIE should be used (typically 45°/0° geometry).

The power of the lamp should be high enough to give a sufficient signal even through the most narrow-band filters used in the system. Using longer exposure times can compensate for low intensity, but could pose additional problems, such as slower acquisitions and amplified noise.

The spectral properties of the lamp should ensure that there is sufficient power over the visible spectrum. Furthermore, it is desirable to use lamps with a smooth spectral power distribution without any sharp peaks.

Stability and repeatability of the irradiance from the light source is of great importance for reliable and accurate measuring results.

The Illumination set-up

The illumination is provided by a tungsten halogen lamp through optical fibers. The lamp has been equipped with an infrared (IR) cut-off filter, to block non visible IR radiation. After passing through the color filters in the filter wheel, the light is led by the fibers in one of two possible directions: to provide either backlight for transmission images, or directed light from above, for reflection images. When the backlight setup is used, the light passes through a transmissive plate, which scatters the light to give a diffused and uniform illumination over the image area. For reflection measurements, the angle of incidence can be adjusted from 10° to 80°. The default angle is 45° which, along with the camera normal to the image plan, provides the standard lighting/viewing geometry 45°/0°, as recommended by CIE.

The lamp

The light source used in the image acquisition system consists of a 150W tungsten halogen lamp from Schott-Fostec. The lamp ‘DCR III’ runs on direct current with a regulated voltage supply adjustable from 0.5 to 20.5 VDC, remotely controlled via the Matlab interface. The specific lamp type used, ‘DDL’, has a nominated rated lifetime of 500 hours at 3150K color temperature when operated at full lamp intensity. When operated at lower intensity, the lifetime increases, while the color temperature decreases (Schott-Fostech, 2002).

Generally, tungsten halogen lamps can run at a higher color temperature, and with better efficacy, than ordinary lamps. The envelope is made of pure fused silica (‘quartz’); it is smaller than ordinary lamps and therefore runs at a much higher temperature. The envelope contains halogen gas. When tungsten evaporates, it forms tungsten halide, a gas that migrates back to the filament in a continuous cycle. Therefore, the dark deposits on the envelope can be avoided (Hunt, 1998).
Recommended lamp intensity

The spectral properties of the lamp are strongly dependent on the intensity level it is operated at. For the lower lamp intensity levels, the color temperature decreases dramatically and the irradiance from the lamp contains almost no power in the blue region of the spectrum. Figure A.1 displays the power distributions for different lamp intensity levels, from 10% to 100%, normalized with respect to their maximum. Furthermore, the relation between the radiant power is not linear with respect to lamp intensity level. For example, a lamp intensity level of 50% generates only 15% of the radiant power compared to full intensity level (see Nyström, 2006, for details).

Figure A.1. Relative spectral power distributions for the illumination at different lamp intensity levels.

The lamp intensity level to operate the lamp at must be a trade-off between the contradictory demands of increasing the lifetime and providing sufficient radiant power throughout the visible spectrum. After taking the cost of replacement lamps and possible degradation effects for an ageing lamp into consideration, we recommend that the lamp should be operated at 90% intensity level. The spectral power at 90% intensity level is very close to that of 100% while the nominal lifetime is prolonged for about 80%, i.e. 900 hours lifetime instead of 500. In the following, all calibration and characterization will be based upon the recommended 90% intensity level.

Stability and repeatability

According to the manufacturer, the output voltage of the power supply stabilizes immediately, while a cold lamp takes typically 25 minutes to stabilize within 1% or better. This stabilization period is only slightly affected by the lamp voltage used. Measurements verify that this promise holds, but at the same time indicate that 25 minutes warm-up time may not always be sufficient when the highest precision is required.
Figure 4.2 displays camera response values during the warm-up cycle for the lamp over a period of 120 min. The measurements have been repeated on 4 different occasions, each time starting with a cold lamp. Clearly, the most dramatic changes occur during the very first minutes, with up to 5% decrease in power. After 25 minutes, the difference between the measurements is still over 0.5% and the power continues to decrease with up to an additional 0.8% during the following hour. Notice that the lamp stabilization cycle differs between the different measurements, with the consequence that repeatability between measurements is hard to attain before the lamp is sufficiently warmed-up and stable.

![Figure A.2. Mean camera response values during the warm up time for the lamp. Measured on 4 different occasions, starting with a cold lamp.](image)

The warm-up time required depends on the demands on accuracy of the acquired images. For highest accuracy, the warm-up time should be extended to 60, or preferably, 90 minutes. After 60 minutes the maximum difference between the measurements is 0.25% and after 90 minutes it is down to 0.1%. From this point, the lamp can be considered as fairly stable, with only small changes during the following hours. Measurements over longer periods of time have revealed that the power continues to decrease, but at a much slower pace. The change in lamp intensity after the first 90 minutes of burning time, and over the following 10 hours, is in the order of magnitude 0.1%.

**Uniform illumination**

To obtain spatial uniformity in the acquired images, the illumination should be uniform over the entire scan area, or any non-uniformity should otherwise be compensated for. Considering the small field of view, due to the large magnification of the optics, uniform illumination over the scan area should not be hard to achieve. However, for shorter focal lengths giving a larger field of view, non-uniform illumination was detected, due to the directed properties of the light and the rather small light cone. This
was accounted for by reconstructing the holder for the optical fibers, increasing the distance to the image plane and yielding a larger light cone area.

Descriptions of how to compensate for non uniform illumination are given in Sec. 4.6, together with corrections for dark current and CCD gain offset.

**Spectral properties**

Figure A.3 displays the normalized spectral distribution of the illumination system, including the lamp and the optical fibers, measured against the white reflectance standard, using a warmed up lamp at 90% intensity. The spectral distribution is relatively smooth, without sharp peaks. The spectral power decreases in the blue domain of the spectrum and is low in the spectral band below 400 nm. Due to the IR cut-off filter, the spectral power is very low in the band above 750 nm. Blocking out IR radiation is important since the CCD sensor is sensitive even in the IR-band, which is a common source of errors when non-visible light affects the camera response (Hardeberg, 2001; Bala, 2003).

The spectral transmittance of the transmissive plate, which should be included in computations from transmission images, is displayed in Nyström (2006).

![Figure A.3. Spectral power distribution for the illumination at 90% lamp intensity level.](image)

**Color filters**

The filter wheel in the image acquisition system offers the possibility of using up to 20 different filters in the optical path of the light source. Any filter with the standard diameter 25 mm can be placed in the filter wheel and the filters can easily be replaced by the user for different purposes. The standard set of filters used in the filter wheel consists of sets of RGB and CMY filters, a set of 7 broadband interference filters, and a
set of 6 neutral density (ND) filters of various optical densities. The first position is empty, allowing for measurement without any filter, i.e. \( I(\lambda) \) directly.

The spectral transmittance for all filters in the filter wheel has been measured. For each channel, \( k \), the spectral irradiance \( E_k(\lambda) \) is measured against the reflectance standard, and the filter transmission \( F_k(\lambda) \) is given by:

\[
F_k(\lambda) = \frac{E_k(\lambda)}{I(\lambda)}
\]

(A.1)

**RGB-filters**

The sets of RGB and CMY filters consist of high saturation dichroic glass filters. The spectral transmittance for the red, green and blue filters, used for RGB-images are displayed in Fig. A.4. The spectral properties of the cyan, magenta and yellow (CMY) filter set are given in Nyström (2006) along with the details on the neutral density filters.

Figure A.4. Spectral transmittance for the RGB-filters.

Figure A.5 displays the CIE xy-chromaticities for the RGB channels (including the illumination and the RGB filters). The white-point plotted corresponds to the case of equal exposure times, illustrating the lower energy for the blue color channel. For comparison, the RGB chromaticities adopted in the sRGB standard are provided (Susstrunk et al., 1999). Clearly, the RGB chromaticities of the system differ to a certain extent from sRGB (which is based on CRT phosphor properties), especially for the blue and green primaries. This not unexpected, nor is it a problem, since the purpose of the image acquisition system is not to capture images directly optimized for displaying on screen.
Interference filters

The filters used for multi-channel image acquisition are 7 broadband interference filters with their center wavelengths (CWL) at 400 nm, 450 nm, 500 nm, 550 nm, 600 nm, 650 nm and 700 nm, respectively. All the interference filters have a bandwidth of 80 nm full width-half maximum (FWHM), except for the filter with 400 nm CWL, having 50 nm FWHM. The center wavelength tolerance is ±15 nm and the peak transmittance ranges from 55% to 60%, according to manufacturer’s specification. Note that the center wavelength corresponds to the midpoint of the pass-band specified by the FWHM which, for interference filters, is normally different from the peak transmittance wavelength. The spectral transmittance for the interference filters used for multi-channel imaging is displayed in Fig. A.6. The CIE xy-chromaticities for the interference filters are depicted in Nyström (2006).

The filters for multi-channel image acquisition have been selected from commercially available filters, with the intention to cover the whole visible spectrum with equally spaced pass bands. This strategy of filter selection is sometimes referred to as equi-spacing of filter central wavelengths. For different strategies for filter selection from sets of available filters, refer e.g. to Hardeberg (2001) and Helling (2008). Studies on filter design for multispectral imaging include Hauta-Kasari, et al. (1998), Imai, et al. (2001) and Quan (2002).

Figure A.5. CIE xy-chromaticities for the RGB channels (*), compared to the chromaticities according to the sRGB standard (x).
Optics

The optics used in the image acquisition system is the APO-CPN 2.8/40 macro system from Schneider Kreuznach. The macro system consists of enlarging lenses, various extension rings and a camera C-Mount adapter, and is designed for macro imaging in scientific applications. While normal lenses for photography are designed to give good images in the range from $1:\infty$ to about 1:10, this macro system is designed to give high quality images for close-up scales of about 1:20 to 1:1 (Schneider Kreuznach, 2004).

The spectral transmittance of the optics, $O(\lambda)$, is difficult to measure directly due to the mechanical dimensions of the optics and the spectroradiometer. Instead it has been estimated together with the spectral sensitivity of the CCD, as described in Chapter 5.

Magnification and flange-to-image distance

The magnification, $\beta'$, is defined as the negative ratio of the image size to object size. For enlarged images ($\beta' < -1$) it is recommended that the lens system is used in reverse position, see Fig. 4.7. For $\beta' = 1$, (1:1 scale), the object size equals the size of the CCD-sensor, 6.3 $\times$ 4.7 mm, which corresponds to 4.65 $\times$ 4.65 $\mu$m/pixel.

Figure A.6. Spectral transmittance for the 7 interference filters, denoted by their center wavelength (CWL).
The flange-to-image distance is the distance from the mounting surface of the lens to the CCD-sensor. It is given by the sum of the length of the extension rings, the back focal length of the lens (17.52 mm) and the camera C-Mount adapter (6.5 mm). From the relationship between the flange-to-image distance and the object-to-image distance it is possible to approximately determine the magnification $\beta'$ (Nyström, 2006).

**Resolution**

To establish the resolution more precisely, it has been measured for a number of different combinations of extension rings using a calibration test target. The test target from MLA is a multifunctional target, designed for calibration and analysis of optical characteristics of optical systems (Hess, 2002). Among other things, the test target includes a linear microscale of high accuracy, which has been used to measure the resolution for different extensions.

Table 4.1 lists the resolution of a number of different extensions, along with the correct focus distances. The ten recommended extensions are selected to cover the range of available resolutions as uniformly as possible.

<table>
<thead>
<tr>
<th>Extension (mm)</th>
<th>Scan area W (mm)</th>
<th>H (mm)</th>
<th>Pixel (μm)</th>
<th>ppi</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25.25</td>
<td>18.98</td>
<td>18.50</td>
<td>1373</td>
</tr>
<tr>
<td>33</td>
<td>14.33</td>
<td>10.79</td>
<td>10.50</td>
<td>2419</td>
</tr>
<tr>
<td>41</td>
<td>9.98</td>
<td>7.49</td>
<td>7.32</td>
<td>3470</td>
</tr>
<tr>
<td>50</td>
<td>7.43</td>
<td>5.60</td>
<td>5.46</td>
<td>4652</td>
</tr>
<tr>
<td>56</td>
<td>5.34</td>
<td>4.02</td>
<td>3.93</td>
<td>6463</td>
</tr>
<tr>
<td>68</td>
<td>4.30</td>
<td>3.24</td>
<td>3.16</td>
<td>8038</td>
</tr>
<tr>
<td>83</td>
<td>3.45</td>
<td>2.60</td>
<td>2.54</td>
<td>10000</td>
</tr>
<tr>
<td>106</td>
<td>2.65</td>
<td>1.99</td>
<td>1.94</td>
<td>13093</td>
</tr>
<tr>
<td>135</td>
<td>2.05</td>
<td>1.55</td>
<td>1.51</td>
<td>16821</td>
</tr>
<tr>
<td>174</td>
<td>1.58</td>
<td>1.19</td>
<td>1.16</td>
<td>21897</td>
</tr>
</tbody>
</table>
Aperture

The aperture of the optics can be varied between f1 and f6, where larger f-values mean smaller aperture openings. However, the aperture is continuously variable, i.e. it has no fixed aperture stops, which makes it very hard to assure repeatability. After changing the aperture, it is virtually impossible to bring the aperture back to its former setting. The system does, however, contain a lock screw, making it possible to lock the aperture in a fixed position.

With respect to repeatability, our first attempt was to recommend the use of aperture f6, for which repeatability can be assured because of its end position. The smallest aperture, f6, provides the maximum depth of focus and eliminates the risk of optical vignetting (Sec. 4.5.5). However, measurements of the modulation transfer function (MTF) reveal considerable loss in image sharpness for aperture f6, compared that of images taken with larger apertures. The details of the results and the measurements of the modulation transfer function are given in Nyström (2006).

As a trade-off between image sharpness, depth of field and eliminating optical vignetting, aperture f4 was selected. In the following, all calibration is based on aperture f4 and the optics has been locked in this position, using the lock screw.

Vignetting

Vignetting, the phenomenon of unintended darkening of the image corners, is caused by three different mechanisms. Mechanical vignetting is caused by mechanical extensions attached to the lens (e.g. a lens hood) that covers the field of view. Optical vignetting (also known as artificial or physical vignetting) occurs when the aperture is fully open because the entrance pupil is partly shielded by the lens barrel. The effect can be eliminated simply by reducing the aperture a few steps. Natural vignetting is associated with the \( \cos^4 \) law of illumination falloff. The illumination falloff due to natural vignetting is most troublesome for wide-angle lenses, but it is inherent to each lens design. The natural vignetting is given by:

\[
\text{Natural vignetting} = \cos(\alpha)\cos^3(\beta) 
\]

where \( \alpha \) is the opening angle in object space and \( \beta \) is the angle in image space. (Van Walree, 2006)

For the macro system that we use, mechanical vignetting is not an issue. Optical vignetting is largest for wide-angle and zoom lenses, and with the lens focused at infinity. With the focus close enough, as in the case of the macro-system, the vignetted area is pushed outwards and will be outside the sensor area. Furthermore, by avoiding using the largest aperture openings, the risk of optical vignetting is eliminated.

Hence, the only type of vignetting one has to consider is the natural vignetting which is inevitable. However, considering the focal lengths used, this will be relatively small. For example, the maximal darkening from natural vignetting occurring at the image corners will, according to Eq. 4.4, be approximately 0.56% for 50 mm extension, 0.18% for 106 mm, and 0.09% for 174 mm extension. Therefore, there is practically no need to explicitly compensate for the natural vignetting. The procedure of using reference white images, described in the CCD camera section, and the correction according to Eq. A.3 will implicitly compensate for the vignetting, along with the CCD gain and non-uniform illumination.
CCD camera

The camera used in the image acquisition system is the Pixelfly HiRes monochrome CCD ‘ICX205AL’, from PCO.imaging. It has a resolution of 1360 × 1024 pixels, a dynamic range of 12 bit, and the available exposure times range from 10μs to 10s. The CCD is of grade 0, which means no defective pixels, and uses digital temperature compensation to reduce noise. The 12bit data delivered from the AD converter is stored using 16 bits, with the higher 4 bits set to zero. The spectral range covers 280 to 1000 nm and the non-linearity of the camera response with respect to the incoming light is less than 1.5%. (PCO, 2003)

The sensor used in the Pixelfly camera is a full-frame CCD, with sensor elements of size 4.65 × 4.65 μm. Full-frame CCD sensors use photocapacitors as photodetectors, arranged as a parallel, vertical charged-coupled device (CCD). The charge collected by the photocapacitors is shifted into a serial horizontal CCD register at the bottom of the array, and then serially transferred to an output amplifier. Because the number of electrons collected in each sensor is proportional to the incident illumination level, the sensor typically has a linear response curve. Full-frame CCDs are primarily used in professional cameras and provide a larger dynamic range due to the larger cell size, compared to interline CCDs, where the photodiode covers only a small fraction of the cell area. Another popular sensor type is the CMOS image sensor which uses photodiodes as photodetectors and suffers from higher noise levels compared to CCD-sensors. (Parulski & Spaulding, 2003)

The calibration of a CCD-camera involves obtaining the camera sensitivity function to ensure linearity, testing for repeatability, and measuring the dark current and CCD-gain of the sensor. The spectral quantum efficiency of the CCD sensor, $QE(\lambda)$, sometimes referred to as spectral responsitivity (Wyszecki & Stiles, 2000) cannot be derived by direct measurements using a spectroradiometer and has been estimated in Chap. 5.

Linearity

The validity of the spectral image acquisition model according to Eq. 5.1 requires a linear response of the camera, i.e. a camera response proportional to the energy of the incident light. A CCD sensor is inherently a linear device, but for real devices the assumption may not always hold (Hardeberg, 2001). To test for CCD linearity, images were captured, out of focus, of a monitor with varying number of black and white pixels. In this way, the amount of incident light is controlled, using a halftoning approach. Figure 4.8 displays the camera response function (CRF) with respect to incident light, sometimes referred to as the opto-electric conversion function (Bala, 2003). The measured data fit reasonably well to a straight line, verifying that the camera response is close to linear and no further linearization of the camera response values is necessary.
Repeatability

To test for the repeatability of the CCD-sensor, 50 images were sequentially captured using identical camera settings. The maximum difference between the mean values of the images is 0.0016 and the standard deviation is 0.00031. Hence, the repeatability of the CCD-sensor is very good. It is hard to tell whether the small variations that do occur originate from the CCD or from variations in the illumination, originating from the light source.

Exposure time

The spectral power of the radiation differs widely for the different color channels due to the spectral properties of the light source and the different transmittance of the filters. To optimize the dynamic range and the signal to noise ratio, it is desirable to use exposure times individually calibrated for each color channel to give maximum digital signal without causing signal clipping. The integration times used must then naturally be included in computations from camera response values.

The linearity in camera response with respect to exposure time has been verified by capturing series of images with varying exposure time. The linearity was expected since the exposure time is controlled electronically, without the use of a mechanical shutter, which can cause variations.

Since there is no simple relationship between camera response and focal length (i.e. the extension used), calibration for the correct exposure times must be carried out for each different combination of extension rings used. However, the focal length will only affect the total level of power, not the spectral characteristics. This means that the relative camera responses of the different channels remain the same, and that the camera only needs to be calibrated without any color filter. The exposure times for the individual channels can then be obtained by a simple scaling, using known factors of the relation between the channels.
Two different approaches can be used when calibrating the exposure times media absolute or media relative (Bala, 2003). When calibrating with media absolute measurements, the exposure times are set to give an equal response for a reference white reflectance standard (e.g. R = G = B = 1), and the object reflectance is measured relative to this standard. When using media relative coordinates, the exposure times are set based on the reflectance of a specific medium, e.g. a substrate, and all reflectance measurements will then be relative to this reference medium.

**Dark current and electronic gain**

**Dark current**

Dark current (sometimes referred to as dark signal or black level) is the output signal from the CCD sensor with zero illumination. This is typically due to leaking current, imperfect charge transport, or thermally generated electrons within the semiconductor material of the CCD. The dark current will depend on the integration time used, as well as on the temperature at which the CCD is operated. (Hardeberg, 2001)

The dark current can be corrected for by subtracting the dark current values from the captured images. However, this must be done for each element, since it will vary significantly over the CCD sensor matrix. Furthermore, it must be done for each individual exposure time used (i.e. for each color channel), since it will depend on the integration time. By capturing images without any illumination (i.e. with the lamp turned off and the lens cap on), using identical exposure times as the real images, these dark current images can be used as offset correction data. Since the dark current level will vary with the temperature in the room, and possibly over time, capturing reference dark current images prior to each image acquisition is recommended.

To investigate the dark current level and how it varies, both spatially and with integration time, series of black images were captured with exposure times ranging from 1 to 4000 ms. The increase in mean dark current with integration time is small, increasing from 0.0087 for 1 ms images to 0.0089 for 4000 ms images (camera values normalized in the range [0,1]). The details on the dark current characteristics are given in Nyström (2006).

**Electronic gain**

The electronic gain refers to the gain of the analog signal registered by each of the CCD elements. Electronic gain is characterized by capturing images of a diffuse white reflector under uniform illumination. The camera should be slightly out of focus to eliminate the effect of any local inhomogeneities of the reflecting surface. The electronic gain can then be measured for each individual CCD element, and compensated.

**Correction for dark current and electronic gain**

The correction for dark current and electronic gain is performed for each individual element of the CCD, and for each individual color channel because of the different exposure times used. Prior to image acquisition, reference images are captured for each color channel, $k$. The dark current images, $D_{ck}(x,y)$, are captured with the illumination turned off and the lens cap on. The white images, $W_{ck}(x,y)$, are captured using a reference white reflection standard under uniform illumination. For each channel, $k$, in the acquired images, $I_{ck}(x,y)$, the correction for dark current and CCD gain is given by:
\[ I_{c_k}(x,y) = a_k \frac{I_k(x,y) - Dc_k(x,y)}{W_k(x,y) - Dc_k(x,y)} \] (A.3)

where \( a_k \) is a channel-dependent normalization factor which can be chosen to give the corrected and normalized image, \( I_{c_k}(x,y) \), desired maximal values (Hardeberg, 2001).

Apart from correction for dark current and CCD gain offset, this equation implicitly includes correction for non-uniform illumination and vignetting. The reference white images should be captured slightly out of focus and the reference reflectance standard must be placed in the same image plane as the object to share the same illumination characteristics. For this purpose, a special holder for the reflectance standard has been constructed.

Summary and discussion

In this appendix, the different components of the image acquisition system have been described and calibrated. The focus has been to gain an understanding of the performance of all components, and how they meet their respective demands, to ensure stability and repeatability in the image acquisition. Recommendations have been given regarding the intensity level and the warm-up time for the lamp, and the aperture for the optics. The spectral characteristics of the illumination and the color filters have been derived by measurements, using a spectroradiometer. The results and recommendations form the basis for ensuring stability and repeatability, which has been crucial for the work presented in Chapters 5-7, focusing on camera sensitivity estimation and on reconstructing spectral and colorimetric data from the acquired images.
Appendix B:
Empirical characterization data

Appendix B collects additional experiments and results that were omitted from Chapter 7, focusing on empirical characterization of the image acquisition system. For the polynomial regression to CIEXYZ and CIELAB, additional polynomials are evaluated as approximation function, both for RGB and multi-channel image data. The influence of the size of the training set is examined by reducing the number of training samples used to derive the polynomial characterization functions. Further, cross-media characterization is performed to evaluate the performance of the characterization functions when used for color samples of different media and colorants, for colorimetric regression and spectral reconstructions, respectively.

The polynomial approximation functions

This section collects the additional polynomials evaluated as approximation functions in the polynomial regression for RGB and multi-channel data. For completeness, the polynomials presented in Chapter 7 are repeated.

RGB data

Besides the basic linear regression, we evaluate the following polynomials as approximation functions for trichromatic RGB-images:

\[ p_1 = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2] \]  \hspace{1cm} (B.1)

This is the basic second-order polynomial, with \( Q = 10 \).

\[ p_2 = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2, RGB] \]  \hspace{1cm} (B.2)

A second-order polynomial completed with the additional RGB term, giving \( Q = 11 \).
\[ \mathbf{p}_3 = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2, R^2G, R^2B, RG^2, \] 
\[ RGB, RB^2, G^3, G^2B, GB^2, B^3] \] (B.3)

A third-order polynomial with \( Q = 20 \).

**Multi-channel data**

For multi-channel images, experiments are performed on linear regression to CIEXYZ as well as directly to CIELAB, using 7-channel and 14-channel images. The 7-channel images correspond to the 7 interference filters with equally spaced pass bands, \( M_1, \ldots, M_7 \). The 14-channel images are completed with the RGB-filters, the CMY-filters and the white W-channel, i.e. with no filter in the optical path. Besides the linear regression, the following polynomial functions for multi-channel images are tested:

\[ \mathbf{p}_{m1} = [1, M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_1M_2M_3M_4M_5M_6M_7] \] (B.4)

This corresponds to linear regression from the 7-channels \( M_1, \ldots, M_7 \), completed with the ‘black’ \( 1 \) and ‘white’ \( M_1M_2M_3M_4M_5M_6M_7 \) term. \( Q = 9 \).

\[ \mathbf{p}_{m2} = [1, M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_1^2, M_2^2, M_3^2, M_4^2, M_5^2, M_6^2, M_7^2, M_1M_2M_3M_4M_5M_6M_7] \] (B.5)

A second-order polynomial using the 7 interference filters, without cross-product terms. \( Q = 16 \).

\[ \mathbf{p}_{m3} = [1, M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_1^3, M_2^3, M_3^3, M_4^3, M_5^3, M_6^3, M_7^3, M_1M_2M_3M_4M_5M_6M_7] \] (B.6)

A third-order polynomial using the 7 interference filters, without cross-product terms. \( Q = 23 \).

\[ \mathbf{p}_{m4} = [1, M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_1^2, M_2^2, M_3^2, M_4^2, M_5^2, M_6^2, M_7^2, M_1M_2M_3M_4M_5M_6M_7, M_1M_2M_3M_4, M_1M_3M_4, M_1M_2M_4, M_1M_5M_4, M_1M_6M_4, M_1M_7M_4, M_2M_3M_4, M_2M_4M_5, M_2M_5M_6, M_2M_6M_7, M_3M_4M_5, M_3M_5M_6, M_3M_6M_7, M_4M_5M_6, M_4M_5M_7, M_4M_6M_7, M_5M_6M_7, M_5M_6M_7, M_6M_7M_7] \] (B.7)

A second-order polynomial using the 7 interference filters, including a number of first-order cross-product terms. \( Q = 37 \).
\[ p_{m5} = [1, R, G, B, R^2, RG, RB, G^2, GB, B^2, W, M_1, M_2, M_1M_4, M_5, M_6, M_7, C, M, Y, WRGMBM_2M_4M_5M_6M_7CMY] \]

(B.8)

Second-order terms from the RGB-channels, together with first-order terms from the remaining 11 channels. The ‘white’ term is based on all 14 channels. \( Q = 22 \).

Experimental results

RGB data

The results for linear and polynomial regression using trichromatic RGB data are given in Tables B.1 and B.2. This completes the results presented in Tables 7.1 and 7.2.

Table B.1. The results from the regression using RGB-data, in terms of the Euclidian distance in CIEXYZ color space, \( \Delta XYZ \), using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>( \Delta XYZ )</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ Linear</td>
<td>7.67</td>
<td>3.24</td>
<td>6.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p_1 )</td>
<td>6.02</td>
<td>1.17</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p_2 )</td>
<td>4.90</td>
<td>1.01</td>
<td>2.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p_3 )</td>
<td>3.45</td>
<td>0.90</td>
<td>2.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XYZ (Min ( \Delta E_{ab} )) Linear</td>
<td>9.00</td>
<td>3.15</td>
<td>6.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p_1 )</td>
<td>5.72</td>
<td>1.25</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p_2 )</td>
<td>6.19</td>
<td>1.00</td>
<td>2.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p_3 )</td>
<td>4.21</td>
<td>0.89</td>
<td>2.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B.2. The results from the regression using RGB-data, in terms of CIE 1976 color difference \( \Delta E_{ab} \) and CIE 1994 color difference \( \Delta E_{94} \), using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>( \Delta E_{ab} )</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ Linear</td>
<td>13.0</td>
<td>7.11</td>
<td>11.6</td>
<td>8.38</td>
<td>4.24</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>9.80</td>
<td>2.41</td>
<td>6.94</td>
<td>7.80</td>
<td>1.67</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>4.29</td>
<td>1.94</td>
<td>3.22</td>
<td>2.51</td>
<td>1.23</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>4.55</td>
<td>2.08</td>
<td>3.94</td>
<td>2.62</td>
<td>1.18</td>
</tr>
<tr>
<td>XYZ (Min ( \Delta E_{ab} )) Linear</td>
<td>11.9</td>
<td>5.80</td>
<td>8.84</td>
<td>5.77</td>
<td>3.75</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>4.34</td>
<td>2.08</td>
<td>3.83</td>
<td>3.83</td>
<td>1.44</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>3.29</td>
<td>1.68</td>
<td>2.82</td>
<td>2.44</td>
<td>1.11</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>3.98</td>
<td>1.91</td>
<td>3.36</td>
<td>2.51</td>
<td>1.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>( \Delta E_{94} )</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ Linear</td>
<td>28.0</td>
<td>15.3</td>
<td>26.0</td>
<td>23.5</td>
<td>12.1</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>28.0</td>
<td>7.2</td>
<td>5.55</td>
<td>5.72</td>
<td>1.75</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>6.83</td>
<td>2.82</td>
<td>5.62</td>
<td>5.61</td>
<td>1.74</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>5.19</td>
<td>1.91</td>
<td>3.37</td>
<td>4.32</td>
<td>1.25</td>
</tr>
<tr>
<td>XYZ (Min ( \Delta E_{ab} )) Linear</td>
<td>28.0</td>
<td>15.3</td>
<td>26.0</td>
<td>23.5</td>
<td>12.1</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>28.0</td>
<td>7.2</td>
<td>5.55</td>
<td>5.72</td>
<td>1.75</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>6.83</td>
<td>2.82</td>
<td>5.62</td>
<td>5.61</td>
<td>1.74</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>5.19</td>
<td>1.91</td>
<td>3.37</td>
<td>4.32</td>
<td>1.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>( \Delta E_{ab} )</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab Linear</td>
<td>10.9</td>
<td>6.62</td>
<td>10.7</td>
<td>9.82</td>
<td>5.17</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>4.27</td>
<td>1.85</td>
<td>3.29</td>
<td>3.67</td>
<td>1.25</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>3.63</td>
<td>1.76</td>
<td>3.06</td>
<td>3.08</td>
<td>1.17</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>4.31</td>
<td>1.72</td>
<td>3.36</td>
<td>3.04</td>
<td>1.11</td>
</tr>
</tbody>
</table>
It is clear that polynomial regression is superior to linear regression for the characterization from RGB to CIEXYZ, as well as CIELAB. The best results in terms of $\Delta$XYZ were obtained for the third order polynomial $p_3$ (Tab. B.1). However, in terms of $\Delta$E$_{ab}$ and $\Delta$E$_{94}$ (Tab. B.2) the second order polynomial $p_2$ generally gives the best results.

**Multi-channel data**

The results for linear and polynomial regression using multi-channel data are given in Tables B.3 and B.4. This completes the results presented in Tables 7.3 and 7.4.

*Table B.3. The results from the regression using multi-channel data, in terms of the Euclidian distance in CIEXYZ color space, $\Delta$XYZ, using standard illuminant D65.*

<table>
<thead>
<tr>
<th>Regression to</th>
<th>Polynomial</th>
<th>$\Delta$XYZ</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ</td>
<td>Linear (7)</td>
<td>3.45</td>
<td>1.02</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>Linear (14)</td>
<td>5.83</td>
<td>1.31</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td>$p_{m1}$</td>
<td>3.78</td>
<td>0.97</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>$p_{m2}$</td>
<td><strong>3.24</strong></td>
<td>0.94</td>
<td>2.39</td>
</tr>
<tr>
<td></td>
<td>$p_{m3}$</td>
<td>3.62</td>
<td>1.04</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>$p_{m4}$</td>
<td>7.95</td>
<td>1.57</td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>3.75</td>
<td><strong>0.93</strong></td>
<td><strong>1.73</strong></td>
</tr>
</tbody>
</table>

| XYZ (Min $\Delta$E$_{ab}$) | Linear (7) | 4.26 | 1.40 | 2.96 |
|                            | Linear (14)| 5.40 | 1.54 | 3.36 |
|                            | $p_{m1}$   | 3.72 | 1.15 | 2.89 |
|                            | $p_{m2}$   | **3.49** | 0.97 | 2.43 |
|                            | $p_{m3}$   | 3.45 | 1.06 | 2.55 |
|                            | $p_{m4}$   | 7.96 | 1.58 | 4.14 |
|                            | $p_{m5}$   | 3.80 | **0.94** | **1.74** |

For the linear regression, we see that there is little to gain by doubling the number of channels from 7 to 14. With the single exception of regression directly to CIELAB without pre-processing, linear regression using the 7 interference filters performs better than the linear regression using all 14 color filters. For the regression to CIELAB, the pre-processing using the cubic root function still improves the results and clearly plays an important role even for multi-channel images, the same way as for the RGB-images. Generally, the best results for the polynomial regression are obtained for the polynomial vectors, $p_{m2}$ and $p_{m5}$, i.e. the second-order polynomial using the 7-channel data without cross product terms, and the polynomial using second-order RGB terms, together with first-order terms from the remaining 11 channels. The poor performance of the polynomial vector $p_{m4}$, the second-order polynomial including cross product terms, with $Q = 37$ is noticeable.
Table B.4. The results from the regression using multi-channel data, in terms of CIE 1976 color difference ΔE_{ab} and CIE 1994 color difference ΔE_{94}, using standard illuminant D65.

| Regression to | Polynomial | ΔE_{ab} | | | ΔE_{94} | | |
|--------------|------------|---------|---------|---------|---------|---------|---------|---------|
|              |            | Max     | Mean    | 95%     | Max     | Mean    | 95%     |
| XYZ          | Linear (7) | 8.33    | 2.94    | 6.87    | 6.53    | 1.75    | 5.36    |
|              | Linear (14)| 4.68    | 2.57    | 4.19    | 4.33    | 1.73    | 3.51    |
|              | pm1        | 10.0    | 2.94    | 7.18    | 7.75    | 1.74    | 5.67    |
|              | pm2        | 3.76    | 1.94    | 3.22    | 2.24    | 1.16    | 2.13    |
|              | pm3        | 3.93    | 1.84    | 3.29    | 3.10    | 1.25    | 2.27    |
|              | pm4        | 10.1    | 3.70    | 7.75    | 5.91    | 2.22    | 4.43    |
|              | pm5        | 3.16    | 1.96    | 3.13    | 2.91    | 1.23    | 2.31    |
| XYZ (Min ΔE_{ab}) | Linear (7) | 6.40    | 2.89    | 5.22    | 4.94    | 1.87    | 3.97    |
|              | Linear (14)| 3.88    | 2.35    | 3.73    | 3.56    | 1.73    | 3.31    |
|              | pm1        | 3.71    | 1.88    | 3.40    | 2.22    | 1.17    | 2.07    |
|              | pm2        | 4.63    | 1.91    | 3.69    | 3.38    | 1.31    | 2.79    |
|              | pm3        | 10.3    | 3.65    | 7.03    | 5.95    | 2.23    | 4.38    |
|              | pm5        | 3.44    | 1.99    | 3.17    | 2.72    | 1.23    | 2.28    |
| Lab          | Linear (7) | 16.7    | 6.76    | 16.1    | 13.2    | 4.45    | 13.1    |
|              | Linear (14)| 13.7    | 6.67    | 13.0    | 10.7    | 4.61    | 10.0    |
|              | pm1        | 9.25    | 5.03    | 9.13    | 7.61    | 3.12    | 5.44    |
|              | pm2        | 8.43    | 2.34    | 4.52    | 5.11    | 1.47    | 2.64    |
|              | pm3        | 7.51    | 2.11    | 3.65    | 4.02    | 1.33    | 2.50    |
|              | pm4        | 13.4    | 3.75    | 9.10    | 5.09    | 1.99    | 4.11    |
|              | pm5        | 8.97    | 2.69    | 5.11    | 5.03    | 1.64    | 3.69    |
| Lab (pre-proc.) | Linear (7) | 6.72    | 3.07    | 5.67    | 4.05    | 1.96    | 3.47    |
|              | Linear (14)| 5.58    | 1.94    | 3.63    | 3.63    | 1.35    | 2.72    |
|              | pm1        | 7.08    | 3.27    | 5.53    | 4.30    | 1.95    | 3.06    |
|              | pm2        | 3.84    | 1.95    | 3.78    | 2.41    | 1.22    | 1.98    |
|              | pm3        | 4.45    | 2.16    | 4.30    | 3.23    | 1.34    | 2.26    |
|              | pm4        | 20.8    | 4.54    | 11.1    | 9.91    | 2.69    | 6.59    |
|              | pm5        | 3.25    | 1.77    | 2.81    | 2.28    | 1.21    | 2.13    |

The size of the training set

To investigate the influence of the number of color samples in the training set used to derive the characterization function, we perform the same experiments using only 25 of the 50 training colors. The results for the polynomials giving the best result for RGB and multi-channel images, i.e. pm2, pm3, pm4 and pm5, are given in Tables B.5 and B.6. The results are computed using the standard illuminant D65 and should be compared to the results in Tables 7.1 - 7.4 (or, conversely, A.1 – A.4).
Table B.5. The results when reducing the training set to 25 test colors, in terms of the Euclidian distance in CIE XYZ color space, ΔXYZ, using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression</th>
<th>Polynomial</th>
<th>ΔXYZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
</tr>
<tr>
<td>RGB to XYZ</td>
<td>p₂</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td>p₃</td>
<td>5.91</td>
</tr>
<tr>
<td>Multi to XYZ</td>
<td>pₘ₂</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>pₘ₅</td>
<td>8.26</td>
</tr>
</tbody>
</table>

Table B.6. The results when reducing the training set to 25 test colors, in terms of CIE 1976 color difference ΔEₐb and CIE 1994 color difference ΔE₉₄ using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression</th>
<th>Polynomial</th>
<th>ΔEₐb</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
<th>ΔE₉₄</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Max</td>
<td>Mean</td>
<td>95%</td>
<td></td>
<td>Max</td>
<td>Mean</td>
<td>95%</td>
</tr>
<tr>
<td>RGB to XYZ</td>
<td>p₂</td>
<td>5.24</td>
<td>2.57</td>
<td>5.09</td>
<td>4.28</td>
<td>1.52</td>
<td>2.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p₃</td>
<td>20.7</td>
<td>4.13</td>
<td>10.1</td>
<td>14.5</td>
<td>2.49</td>
<td>5.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi to XYZ</td>
<td>pₘ₂</td>
<td>6.21</td>
<td>2.45</td>
<td>4.37</td>
<td>2.91</td>
<td>1.37</td>
<td>2.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pₘ₅</td>
<td>11.3</td>
<td>5.72</td>
<td>11.1</td>
<td>8.99</td>
<td>3.48</td>
<td>7.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB¹/₃ to Lab</td>
<td>p₂</td>
<td>5.55</td>
<td>2.20</td>
<td>4.61</td>
<td>3.09</td>
<td>1.27</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p₃</td>
<td>6.21</td>
<td>2.79</td>
<td>5.53</td>
<td>5.09</td>
<td>1.89</td>
<td>3.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi¹/₃ to Lab</td>
<td>pₘ₂</td>
<td>8.25</td>
<td>3.30</td>
<td>6.73</td>
<td>5.18</td>
<td>2.19</td>
<td>4.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pₘ₅</td>
<td>15.1</td>
<td>4.08</td>
<td>8.82</td>
<td>11.0</td>
<td>2.78</td>
<td>6.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As could be expected the errors increase when the training set is reduced by half. It is clear that the more terms included in the polynomials, the larger the increase of estimation errors. For example, for RGB-images, p₂ (Q=11) now clearly outperforms p₃ (Q=20). Similarly, for the multi-channel images, pₘ₂ (Q=16) performs better than pₘ₅ (Q=22). This is because higher-order polynomials, including more polynomial terms, need larger training sets for optimal performance. The best performance is obtained for the second-order polynomial p₂, using regression from RGB to CIELAB with the preprocessing step. It should be noted that 25 color samples included in the training set are still sufficient for the system of equations to be over-determined for the polynomials in question.

Graphs of the mean and maximal estimation errors in terms of ΔEₐb when varying the number of training colors included in the training set from 50 to 20 are given in Nyström (2006). The results further verify that the simpler polynomials are more robust when decreasing the training set.

Cross-media characterization

Experiments on cross-media characterization is performed to investigate to what extent the characterization is affected by eye-camera metamerism, i.e. how the derived functions perform when using color samples of different media and colorants.

Colorimetric regression

Tables B.7 and B.8 collect the results when the characterization functions derived for the printed color patches are used to estimate colorimetric data for NCS colors, used in Chap. 6.
Table B.7. The results for cross-media characterization, in terms of the Euclidian distance in CIEXYZ color space, $\Delta_{XYZ}$, using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression</th>
<th>Polynomial</th>
<th>$\Delta_{XYZ}$</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB to XYZ</td>
<td>$p_2$</td>
<td>24.9</td>
<td>13.0</td>
<td>22.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_3$</td>
<td>25.5</td>
<td>12.9</td>
<td>21.1</td>
<td></td>
</tr>
<tr>
<td>Multi to XYZ</td>
<td>Linear</td>
<td>10.4</td>
<td>4.17</td>
<td>9.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_{m2}$</td>
<td>13.7</td>
<td>5.36</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>23.8</td>
<td>16.1</td>
<td>23.1</td>
<td></td>
</tr>
</tbody>
</table>

Table B.8. The results for cross-media characterization, in terms of CIE 1976 color difference $\Delta_{E_{ab}}$ and CIE 1994 color difference $\Delta_{E_{94}}$, using standard illuminant D65.

<table>
<thead>
<tr>
<th>Regression</th>
<th>Polynomial</th>
<th>$\Delta_{E_{ab}}$</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
<th>$\Delta_{E_{94}}$</th>
<th>Max</th>
<th>Mean</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB to XYZ</td>
<td>$p_2$</td>
<td>58.5</td>
<td>11.3</td>
<td>31.3</td>
<td>19.5</td>
<td>7.30</td>
<td>14.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_3$</td>
<td>16.6</td>
<td>9.85</td>
<td>16.2</td>
<td>13.4</td>
<td>7.43</td>
<td>11.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi to XYZ</td>
<td>Linear (7)</td>
<td>27.7</td>
<td>4.85</td>
<td>9.63</td>
<td>15.6</td>
<td>3.25</td>
<td>5.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_{m2}$</td>
<td>26.0</td>
<td>9.78</td>
<td>22.2</td>
<td>15.0</td>
<td>6.31</td>
<td>11.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>40.2</td>
<td>18.1</td>
<td>37.6</td>
<td>25.9</td>
<td>10.1</td>
<td>22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB to Lab</td>
<td>$P_2$</td>
<td>18.2</td>
<td>8.41</td>
<td>14.6</td>
<td>11.6</td>
<td>6.82</td>
<td>10.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$</td>
<td>27.2</td>
<td>10.5</td>
<td>22.6</td>
<td>20.6</td>
<td>7.98</td>
<td>16.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi to Lab</td>
<td>Linear (7)</td>
<td>18.7</td>
<td>8.87</td>
<td>17.8</td>
<td>15.3</td>
<td>5.94</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_{m2}$</td>
<td>63.3</td>
<td>22.7</td>
<td>54.2</td>
<td>27.2</td>
<td>13.5</td>
<td>26.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_{m5}$</td>
<td>80.9</td>
<td>20.5</td>
<td>61.3</td>
<td>36.6</td>
<td>12.0</td>
<td>27.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is evident from the results that the eye-camera metamerism is severe and that the estimation errors increase dramatically. For the polynomial regressions, the RGB-images give better performance than the multi-channel images. The best result is actually obtained using the basic linear regression from multi-channel data to CIEXYZ. It is evident from the result that direct polynomial regression is optimized only for the specific combination of media and colorants used to derive the characterization function. It seems that the more complex the estimation function, i.e. the order of the polynomial and the number of terms included, the stronger the effect of the cross media characterization.

**Spectral regression**

To investigate how the empirical spectral reconstruction performs when used to reconstruct spectral reflectance for color samples of different media and colorants, we perform experiments on cross-media characterization. In the following, the spectral reconstruction matrix derived for the printed training set has been used to reconstruct spectral reflection for the set of NCS color patches.

The results from the cross-media spectral reconstruction are given in Tables B.9 and B.10, using D65. Examples of estimated spectral reflectance using RGB and multi-channel data are presented in Nyström (2006).
Table B.9. Spectral reconstruction errors for cross-media characterization expressed as RMS difference in spectral space, and the Euclidian distance in CIEXYZ space, \( \Delta XYZ \).

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>RMS</th>
<th>AXYZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max  Mean 95%</td>
<td>Max  Mean 95%</td>
</tr>
<tr>
<td>RGB</td>
<td>PI</td>
<td>0.0245 0.0076 0.0223</td>
<td>24.6 10.7 20.2</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>0.0238 0.0069 0.0221</td>
<td>24.6 11.5 21.6</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>0.0215 0.0067 0.0183</td>
<td>23.2 11.0 20.8</td>
</tr>
<tr>
<td>Multi</td>
<td>PI</td>
<td>0.0187 0.0024 0.0058</td>
<td>11.0 4.24 9.61</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>0.0151 0.0016 0.0038</td>
<td>12.8 4.79 9.95</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>0.0134 0.0022 0.0052</td>
<td>13.6 5.49 10.7</td>
</tr>
</tbody>
</table>

Table B.10. Spectral reconstruction errors for cross-media characterization, in terms of CIE 1976 color difference, \( \Delta E_{ab} \) and CIE 1994 color difference \( \Delta E_{94} \), using D65.

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>( \Delta E_{ab} )</th>
<th>( \Delta E_{94} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max  Mean 95%</td>
<td>Max  Mean 95%</td>
</tr>
<tr>
<td>RGB</td>
<td>PI</td>
<td>43.2 9.23 20.2</td>
<td>20.4 6.08 9.30</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>41.4 8.80 19.3</td>
<td>21.4 6.10 9.63</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>26.2 7.50 12.8</td>
<td>17.4 5.73 8.80</td>
</tr>
<tr>
<td>Multi</td>
<td>PI</td>
<td>22.5 4.66 9.70</td>
<td>14.6 3.20 5.81</td>
</tr>
<tr>
<td></td>
<td>Spectral</td>
<td>16.4 3.66 7.81</td>
<td>11.2 2.75 4.60</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td>24.5 5.35 15.1</td>
<td>14.2 3.30 5.32</td>
</tr>
</tbody>
</table>

Clearly, the reconstruction errors increase for the cross-media applications (compare to Tables 7.5 and 7.6), especially in terms of the maximal errors. When the reconstruction matrix is used for other media and colorants, the improvement using multi-channel data compared to RGB-data is more obvious. The pseudo-inverse method, which provided the best results reconstructing spectral reflectance for the same media and colorant, is now clearly outperformed by the methods using base functions. The Fourier basis provides the best results for RGB-data, while the spectral basis produces the best results for multi-channel data. Notice that the spectral basis is computed from the NCS spectral database, which should have the advantage when used to reconstruct reflectance spectra of other NCS color patches. The best results for the cross-media characterization are obtained using multi-channel data and spectral basis, producing mean errors comparable to the results using the same media and colorants, but with the maximal errors more than doubled.

Note that, when compared to the results for cross-media characterization using direct regression to CIEXYZ and CIELAB (Tables B.7 and B.8), the estimation errors decrease when the colorimetric data is computed from the reconstructed spectra. This is probably due to the fact that the eye-camera metamerism is avoided when the regression is used to estimate spectral reflectance. Even though the spectral reconstruction method could not compete with the direct colorimetric regression to estimate colorimetric data using the same media and colorant, it is clearly more robust when estimating colorimetric values for color samples of different-media and colorants.
Appendix C:
High resolution analysis data

Appendix C collects additional experiments and results that were omitted from Chapter 8, focusing on high resolution analysis of halftone prints. The ideas presented in Chap. 8 were first tested in a preliminary study (Nyström, 2008a), from which some of the results are presented here.

Besides the line scan method, introduced in Chap. 8, the physical dot area has been estimated using the midpoint between the peaks in the histogram, and by using micro-scale images captured in transmission mode. The physical dot gain has been estimated from the transmission images as described in Sec. 9.2.2, with the differences that the images are now captured in micro-scale.

Experimental setup

The printed samples used in the preliminary study consisted of already available offset prints on different paper grades, with both amplitude modulated (AM) and frequency modulated (FM) halftones. Test patches of the primary colors and black were printed with nominal dot area coverage, $a_0$, ranging from 0 to 100%, in 10% intervals. The details on the print samples are given in Tab. C.1. Macroscopic measurements of the spectral reflectance values of the printed color patches were derived using a spectrophotometer, equipped with a UV filter. All colorimetric computations were performed using the CIE standard illuminant D65.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Type</th>
<th>Dot Area</th>
<th>Resolution</th>
<th>Line Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper A</td>
<td>Silk, 130 gr/m²</td>
<td>1200 dpi</td>
<td>175 lpi</td>
<td></td>
</tr>
<tr>
<td>Paper B</td>
<td>Matt, 100 gr/m²</td>
<td>800 dpi</td>
<td>150 lpi</td>
<td></td>
</tr>
<tr>
<td>Paper C</td>
<td>Matt, 100 gr/m²</td>
<td>1200 dpi</td>
<td>150 lpi</td>
<td></td>
</tr>
<tr>
<td>Paper D</td>
<td>Gloss, 130 gr/m²</td>
<td>1200 dpi</td>
<td>175 lpi</td>
<td></td>
</tr>
<tr>
<td>Paper E</td>
<td>Uncoated, 150 gr/m²</td>
<td>800 dpi</td>
<td>150 lpi</td>
<td></td>
</tr>
</tbody>
</table>
Experimental results

Reflectance values

Table C.2 lists the RMS errors between the measured mean reflectance and the predicted reflectance values, for all printed halftone patches. The expanded Murray-Davies model (Eq. 8.1) has been used with $R_p(a)$ and $R_i(a)$ computed from grayscale histograms, as described in Sec. 8.2. The physical dot area coverage, $a_{phy}$, has been estimated using the midpoint between $R_p(a)$ and $R_i(a)$ (Mid), using line scans from micro-scale images (Line) and by using transmission images of the prints (Trans). Computations with the Murray-Davies (MD) model and the Yule-Nielsen (YN) model are also included for comparison. For Yule-Nielsen, the optimal $n$-value has been used, computed individually for each print. The best results are marked in bold in the table.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Halftone</th>
<th>MD</th>
<th>YN</th>
<th>Expanded MD</th>
<th>Mid</th>
<th>Line</th>
<th>Trans</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>AM</td>
<td>0.0685</td>
<td>0.0095</td>
<td>0.0084</td>
<td>0.0052</td>
<td>0.0539</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.1066</td>
<td>0.0135</td>
<td>0.0113</td>
<td>0.0100</td>
<td>0.0380</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>AM</td>
<td>0.0636</td>
<td>0.0103</td>
<td>0.0065</td>
<td>0.0034</td>
<td>0.0144</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.0942</td>
<td>0.0183</td>
<td>0.0116</td>
<td>0.0136</td>
<td>0.0092</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>AM</td>
<td>0.0619</td>
<td>0.0094</td>
<td>0.0098</td>
<td>0.0088</td>
<td>0.0145</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.1081</td>
<td>0.0128</td>
<td>0.0154</td>
<td>0.0170</td>
<td>0.0216</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>AM</td>
<td>0.0805</td>
<td>0.0098</td>
<td>0.0093</td>
<td>0.0073</td>
<td>0.0254</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.1323</td>
<td>0.0174</td>
<td>0.0116</td>
<td>0.0190</td>
<td>0.0148</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>AM</td>
<td>0.0769</td>
<td>0.0177</td>
<td>0.0100</td>
<td>0.0081</td>
<td>0.0207</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.0981</td>
<td>0.0196</td>
<td>0.0246</td>
<td>0.0297</td>
<td>0.0280</td>
<td></td>
</tr>
</tbody>
</table>

The results by employing the expanded Murray-Davies model are equivalent to, or better, than the Yule-Nielsen results using optimal $n$-value. As expected, the expanded Murray-Davis clearly outperforms the results for the ordinary Murray-Davis. The line scan method generally gives the best results among the different approaches to compute the physical dot area fraction, $a_{phy}$. The simple approach of defining the threshold as the midpoint of $R_i(a)$ and $R_p(a)$ performs surprisingly well, and generally gives better results than using transmission images. Computing the physical dot area from transmission images should in theory work well, as shown in Chap. 9, but the method suffers from some practical difficulties. In the transmission images captured in micro-scale, the paper fibers become clearly visible and strongly affect the computed intensity values. The result is that the method becomes unstable, with results that depend strongly on the local structure of fibers captured in the images, which varies over the paper area.

Figure C.1 displays the measured reflectance values compared to the predicted reflectance using the model, and the reflection for the ink, $R_i(a)$, and the paper, $R_p(a)$, estimated from the histograms. Clearly, the largest variation for both $R_i(a)$ and $R_p(a)$ occurs for the FM halftones, giving the strongest optical dot gain. The variation of the reflectance of the paper, $R_p(a)$ is generally larger than that of the ink $R_i(a)$.
The largest errors when applying the model clearly occur for the prints with FM halftones (Tab. C.1, Fig. C.1). The reason is that with the small halftone dots used in the FM halftones, it is difficult to distinguish the paper from the closely spaced dots when ink coverage is high. The combination of both physical and optical dot gain in the prints results in images where the paper between the dots is no longer visible. The result is that the bimodal properties of the histogram is lost, making it difficult to find the proper values of $R_p(a)$ and $R_i(a)$ using histogram data. This is apparent in Fig. C.1(d) (Paper B - FM), where it is clear that the paper reflectance, $R_p(a)$ has been overestimated for the higher dot percentages, $a_{phy}$.

The reason that the FM prints in this preliminary study produced larger estimation errors compared to the results presented in Chap. 8 is because they suffered from a much larger physical dot gain, making it difficult to locate the peaks in the histograms, as well as estimating the physical dot area.

As an example, images of 80% black tints using AM and FM halftoning for the same paper grade (Paper D) are displayed in Fig. C.2, along with the corresponding reflectance histograms. Clearly, the FM halftone appears more like a homogeneous gray even in micro-scale, as also suggested by its histogram where there is no peak corresponding to the reflectance of the paper.
Tristimulus values

Table C.3 lists the results when the model is used to predict tristimulus values based on the one-dimensional color distributions projected from CIEXYZ color space, as described in Sec. 8.2.4. The tristimulus values for the ink and paper are computed from the one-dimensional color distributions and the midpoint is used to define the threshold. The results are evaluated using the Euclidian distance in CIEXYZ color space, ΔXYZ, as well as the CIE 1976 color difference ΔEab, between the measured and predicted colorimetric values. Again, the largest estimation errors occur for the prints using the FM halftone, where the correct values for XYZ(a) and XYZp(a) are difficult to derive from histogram data. Generally, the black prints give the smallest estimation errors, and the largest errors appear for the yellow prints. The reason is that the tristimulus values for the yellow ink are much closer to those for the paper, making it harder to clearly separate the ink from the paper in histogram.

Again, the estimation errors are larger than the corresponding results presented in Chap. 8, especially for the FM prints. This is due to the larger physical dot gain in the prints, making it difficult to properly locate the correct peaks and thresholds from the histograms.

Figure C.3 illustrates how the CIE chromaticity coordinates, x=X/(X+Y+Z) and y=Y/(X+Y+Z), vary with the dot area coverage for the paper and the ink. The central points correspond to the chromaticity for the paper and the cluster of points at the extremes to the ink chromaticity for the primary colors cyan, magenta and yellow. It is noted that the chromaticity coordinates vary almost along straight lines, both for the ink and the paper.
Table C.3. Estimation errors between measured and predicted tristimulus values.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta XYZ$ max</th>
<th>$\Delta XYZ$ mean</th>
<th>$\Delta E_{ab}$ max</th>
<th>$\Delta E_{ab}$ mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paper A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>Black</td>
<td>3.37</td>
<td>1.34</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>2.22</td>
<td>1.09</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>4.31</td>
<td>2.51</td>
<td>6.67</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>4.50</td>
<td>1.89</td>
<td>4.85</td>
</tr>
<tr>
<td><strong>Paper A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FM</td>
<td>Black</td>
<td>3.07</td>
<td>1.21</td>
<td>6.03</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>8.80</td>
<td>2.97</td>
<td>8.00</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>5.14</td>
<td>3.54</td>
<td>10.85</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>3.59</td>
<td>2.25</td>
<td>10.76</td>
</tr>
<tr>
<td><strong>Paper D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>Black</td>
<td>2.40</td>
<td>0.99</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>5.71</td>
<td>3.46</td>
<td>7.54</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>3.43</td>
<td>2.07</td>
<td>7.58</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>6.57</td>
<td>3.67</td>
<td>11.53</td>
</tr>
<tr>
<td><strong>Paper D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FM</td>
<td>Black</td>
<td>2.50</td>
<td>0.99</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td>Cyan</td>
<td>10.02</td>
<td>4.38</td>
<td>12.63</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>12.35</td>
<td>3.97</td>
<td>14.71</td>
</tr>
<tr>
<td></td>
<td>Yellow</td>
<td>17.85</td>
<td>4.11</td>
<td>14.92</td>
</tr>
</tbody>
</table>

Figure C.3. CIE chromaticity coordinates for the ink and for the paper between the dots, with varying dot area coverage. Cyan, magenta and yellow colorants, Paper A – AM.