PUBLICATIONS OF
THE UNIVERSITY OF EASTERN FINLAND

Dissertations in Forestry and Natural Sciences

ANA GEBEJES

SPECTRAL VIDEO

Application in human eye analysis and tracking

SPECTRAL VIDEO

APPLICATION IN HUMAN EYE ANALYSIS AND TRACKING

Ana Gebejes

SPECTRAL VIDEO

APPLICATION IN HUMAN EYE ANALYSIS AND TRACKING

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences
No 275

University of Eastern Finland
Joensuu
2017

Academic dissertation
To be presented by permission of the Faculty of Science and Forestry
for public examination in the Temple of Challenge in Science Park at the
University of Eastern Finland, Joensuu, on June, 30, 2017,
at 12 o'clock noon

Grano Oy
Jyväskylä, 2017
Editors: Pertti Pasanen, Matti Vornanen,
Jukka Tuomela, Matti Tedre
Distribution: University of Eastern Finland / Sales of publications
www.uef.fi/kirjasto
ISBN: 978-952-61-2550-3 (Print)
ISBN: 978-952-61-2551-0 (PDF)
ISSNL: 1798-5668
ISSN: 1798-5668
ISSN: 1798-5676 (PDF)

Author's address: Ana Gebejes
University of Eastern Finland
School of Computing
P.O. Box 111
80101 JOENSUU, FINLAND
email: ana.gebejes@uef.fi

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

Docent Roman Bednarik, Ph.D.
University of Eastern Finland
School of Computing
P.O. Box 111
80101 JOENSUU, FINLAND
email: roman.bednarik@uef.fi

Reviewers: Associate professor Meritxell Vilaseca Ricart, Ph.D
Universitat Politècnica de Catalunya, Centre for Sensors, Instruments and Systems Development
Rambla Sant Nebridi, 10
08222 TERRASSA, SPAIN
email: meritxell.vilaseca@upc.edu

Professor Pavel Zemčík, Ph.D
Faculty of Information Technology, Department of Computer Graphics and Multimedia
Božetěchova 2
612 66 BRNO, CZECH REPUBLIC
email: zemcik@fit.vutbr.cz

Opponent: Professor Dietrich Paulus, Ph.D
Koblenz University
Active Vision Group
Rhabanusstraße 3
55118 MAINZ, GERMANY
email: paulus@uni-koblenz.de

Gebejes, Ana
Spectral video: application in human eye analysis and tracking
Joensuu: University of Eastern Finland, 2017
Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences 2017; 275
ISBN: 978-952-61-2550-3 (print)
ISSNL: 1798-5668
ISSN: 1798-5668
ISBN: 978-952-61-2551-0 (PDF)
ISSN: 1798-5676 (PDF)

ABSTRACT

In many applications there is a need for monitoring via spectral video – for example, in industrial production-line inspections and medical procedures. This thesis concentrates on exploring the potential of a novel technique for spectral acquisition – i.e., spectral video acquisition. In addition, the thesis suggests the possibility of applying spectral video to spectral object tracking (i.e., spectral eye tracking). Here, a seven-channel spectral video system, the FluxData (FD) model FD 1665-MS7 (FluxData Inc., USA), was used. First, the FD system's performance was examined, and then the system was applied in order to capture spectral videos of a fast-moving object: the human eye. Moreover, a novel, first of its kind, publicly available combined spectral video/spectral image database was created: the SPectral Eye vidEo Database (SPEED).

In a spectral video, each frame of the video stream contains a spectral image cube. In this way, spectral video systems capture spatial, spectral, and temporal information, and combine the advantages of spectral imaging and video capture. What is of utmost importance for spectral video systems is the ability to compare frame sequences in order to ascertain whether or not they have similar spectral content, as well as if the changes in content happen understand and interpret them correctly. To ensure that the changes do not occur due to noise, reliable characterization methods are needed. In this study, a set of methods for characterizing a spectral video camera has been compiled and implemented. The characterization data collected was then used to calibrate the FD spectral video system and to prepare this system for practical application.

The unique database, SPEED, which has been created for this project was motivated by the challenges faced in eye-tracking, especially under harsh conditions (such as when eyewear reflections, make-up, and extreme eye positions interfere). The FD spectral video system was used to record the human eye during the performance of observational tasks. In addition, to compile a spectral ground truth, fifty-one-channel spectral images were collected using a Nuance EX (CRi, Inc., USA)

Liquid Crystal Tunable Filter (LCTF). It is proposed that spectral signatures can be exploited to create new methodologies for the imaging, training, analysis, and interpretation of eye-tracking data, especially in harsh conditions. At the moment, SPEED consists of 30 fifty-one-channel spectral images, 60 seven-channel spectral images, and 180 seven-channel spectral videos of 30 voluntary subjects. The database continues to grow, and is publicly available to all researchers (e.g., for teaching-related needs or creating and testing new methods of dynamic eye analysis via spectra).

This thesis demonstrates the possibility of using the SPEED database in multiple novel dynamic computer-vision applications, such as spectral-reflectance information-based classification and segmentation, reflectance-based object detection, illumination-independent object tracking, and temporal spectral analysis. In addition to its use in the field of eye tracking, SPEED has the potential to be applied to many other eye-related areas of research (e.g., medicine, biometrics, and eye/vision-based studies). As the most significant contribution of the thesis, SPEED provides a platform for creating novel spectral-data-based methodologies for the imaging, training, analysis, and interpretation of spectral video. Moreover, it provides valuable data for the creation of new spectral-data-based enhanced devices, software, and algorithms for use in all eye-related research fields.

Universal Decimal Classification: 004.932, 621.397, 681.586.5

Library of Congress Subject Headings (LCSH): Digital video; Spectral imaging; Image processing — Digital techniques; Detectors; Cameras — Calibration; Computer vision; Eye tracking; Digital video – Databases.

Keywords: spectral video, sensor characterization, video characterization, spectral image processing, spectral video processing, eye-tracking, hyperspectral imaging, human eye, spectral image and spectral video database.

ACKNOWLEDGEMENTS

Since our parents hear us for the first time an almost overwhelming amount of love comes our way. They advise us to be careful, to make good choices, to choose the right path in our life. Even though it might be annoying at times, they become the voice in our head pushing us to do better and get further.

You see, my mother taught me some really awesome things like, for instance, making baklava. For a very long time my only concern was how to make the most tasty baklava and I can proudly say today I mastered that craft. My mother also taught me that if I want to master any other craft I have to listen to older, more experienced people. Listen to what they can teach me but not to follow their footsteps, not to make the same mistakes they made but learn and benefit from those mistakes. She taught me how to pay attention to others and memorize information that can be useful for me in the future.

My father, on the other hand, was a typical man. He taught me how to fix my broken bicycle, how to assemble machinery and most importantly he taught me how to kick a ball. He was also one of my chess coaches and honestly he was the strictest one. From him I learned that it is not a problem if you lose a game but it is a problem if you don't learn from the mistakes you made. Only accepting that you were defeated can open your mind to see what the mistake you made was and help you remember not to make the same one again.

All these values that I have gained at early age helped and guided me through my life and education. They helped me to absorb knowledge and be patient to wait for the moment when I will know enough to start researching, deepening my knowledge with practical work and teaching. With this desire I started the PhD program at UEF. It was an eventful four-year-long journey here at the Spectral Color Research Laboratory at the School of Computing. Today I can say with a lot of pride and honor that what I received from this journey was much more than what I have expected. For these four years I was privileged to study, work, learn from, laugh with, cry, be angry with and love so many different and unique people. This gave me energy to get where I am now. Critiques and grades built me professionally, good lectures inspired me and the bad ones motivated me to build better teaching skills myself.

During my PhD studies I had a privilege to work in a TEKES - Olympus project, Spectral imaging and analysis in environmental and industrial applications, with an amazing team of local and international researchers and a company, Olympus. Even now as I am writing I can still hear the summer sound of Japan and remember the hospitality. Endless brainstorming and exchange of knowledge and experience enriched this period and the first half of this thesis was possible thanks to those times.

From Japan my studies took me all the way to the International Space Station, well not literally but via a Skype connection. I had a chance to work on improving

my teaching skills with astronauts on Earth and in Space. This Space experience took me to a new world of innovation where I could explore the potentials of team creativity. This experience gave me the courage to create my own project and thanks to the Faculty of Forestry and Natural Sciences this project became a reality and the second part of this thesis was created. I wish to acknowledge all the colleagues I have had the opportunity to work with and all the people that participate in the thesis experiments.

My supervisor, Prof. Markku Hauta-Kasari, was along the way all this time providing the support and encouragement. Paying attention to him and his advices helped me to really mature as a researcher and a manager of research. He was there to hear out all the ideas that came to my mind and constructively answer to those.

My second supervisor, Doc. Roman Bednarik, brought visionary energy into my research that fed and facilitated the final stages of this thesis. That energy helped me see that everything is in fact possible.

This work would not have been the same without the support and encouragement from my friends and colleagues. Maja and Anita, I can still clearly hear you say 'Just finish it and come home already'. Pauli, your guidance will always be highly appreciates. Ville, I have learned a lot from our endless brainstorming sessions. All my friends from 'Forever', thank you for the countless squats and jump kicks that kept my body and mind fit for completing this endeavor. Mareena, thank you for your kindness and friendship and for always bringing the warmest feelings in the freezing winter.

Thank you is simply not enough to express how grateful I am to all of you for taking part in this journey with me. I will cherish this time and use it as fuel to continue with my work because "working hard always pays off and there is never enough of knowledge".

Thank you all for your time and patience.

Joensuu, 25th March 2017
Ana Gebejes

LIST OF ABBREVIATIONS

AOTF	Acousto–Optic Tunable Filter
B	Blue
CASSI	Coded Aperture Snapshot Spectral Imagers
CC24	ColorChecker Classic
CCD	Charge-Coupled Device
CFA	Color Filter Array
CMOS	Complementary Metal Oxide Semiconductor
CTIS	Computed Tomographic Imaging Spectrometry
DSG96	Digital ColorChecker SG
FD	FluxData model: FD-1665-MS7, FluxData, Inc., USA
fps	Frames Per Second
FWHM	Full Width Half Maximum
G	Green
GFC	Goodness-of-fit Coefficient
HPA	HyperPixel Array™
LCTF	Liquid Crystal Tunable Filter
LED	Light-Emitting Diode
LS	Line Scan
NIR	Near-InfraRed
PD	Pearson Distance
R	Red
RAD	spectroradiometer
RMSE	Root-Mean-Square Error
SAM	Spectral Angle Mapping
SPEED	SPectral Eye vidEo Database
VIS	VISible

LIST OF ORIGINAL PUBLICATIONS

The thesis consists of the present review of the author's work in the field of spectral color science, spectral video acquisition and application. This thesis is based on data presented in the following scientific publications, referred to by the Roman numerals I–III.

I Gebejes A., Orava J., Penttinen N., Heikkinen V., Hiltunen J., Hauta-Kasari M. (2014, June). Color and image characterization of a three CCD seven band spectral camera. *In International Conference on Image and Signal Processing. Springer International Publishing.* 96-105

II Gebejes A., Fält P., Bednarik R., Hauta-Kasari M. (2017, Feb). Procedures to analyze and evaluate the output quality of spectral video systems. *IEEE Transactions on Image Processing.* SUBMITTED

III Gebejes A., Fält P., Bednarik R., Hauta-Kasari M. (2016, Sep). SPEED: SPectral Eye vidEo Database. *In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM.* 1666-1675

Throughout this thesis, these publications will be referred to by their Roman numerals. These original papers have been included at the end of this thesis with the permission of the publishers.

AUTRHOR'S CONTRIBUTION

I) The research presented in this thesis is the results of the author's original ideas on spectral video application.

II) The author initiated and established a collaboration between two research groups at UEF - Spectral Color Research and Eye-tracking, in order to investigate the possibilities of using a spectral video device as an eye tracker. This work is pioneering as spectral techniques have not been tested in this field previously.

III) The author used a unique experimental setup to create what can be thought of as a remote spectral eye tracker.

IV) The publications discussed in this thesis are original research papers on spectral video characterization and its application. The author's contribution to these papers (as well as the contributions of the co-authors) can be summarized as follows:
Publications I and **II** provide a systematic collection of camera-characterization methods and their applications using a novel seven-channel spectral video system. These methods were researched and applied by the main author. All the seven-channel still image and video data used in **Publications I** and **II** was collected by the main author. The ground truth data for **Publication I** was collected by co–author Dr. Jouni Hiltunen. Advanced reflectance estimation models used in **Publication I** were applied by co–author Dr. Ville Heikkinen. Programming for the seven-channel system control was done by co–author, Dr. Niko Penttinen. All the computations in **Publication II** were undertaken by the main author.
In **Publication III** the seven-channel video system was used to create a novel spectral video/spectral image database named SPEED: SPectral Eye vidEo Database. The main author created the unique experimental setup and performed all the measurements. The setup is original, it can be considered the very first spectral eye tracker. The co-authors contributed by supervising the development of the experimental setup. The data analysis and all the numerical computations were done by the author. The MATLAB code used to obtain examples of pupil detection and feature classification (in Section 4.3.3) was written by Dr. Pauli Fält. The manuscripts of **Publications I–III** were written by the main author; the co-authors contributed to their improvement. **Publication I** was mainly written by the author together with the co-authors. In all publications the main author has the biggest contribution in the original research ideas, implementation and writing.

CONTENTS

1 INTRODUCTION

This thesis explores the potential of spectral video acquisition and its application in spectral object tracking. A seven-channel spectral video camera (FluxData model FD 1665-MS7, FluxData, Inc., USA) was charaterized and utilized to capture spectral videos of a fast-moving object: the human eye. In addition, a novel, publicly available combined spectral video/spectral image database was created: the SPectral Eye vidEo Database (SPEED).

Spectral still camera systems are often too slow to capture spectral images of dynamic objects at high-enough frame rates. This constrains such systems to use in static and controlled applications only. As such, the logical extension of spectral still imaging is the recording of spectral video. Recent technological advances in optics and electronics have led to the development of spectral-video sensors that are able to capture spectral information at video rates. In a spectral video, each frame of the video stream contains a spectral image cube. These systems are thus able to capture spatial, spectral, and temporal information, combining the advantages of spectral imaging and video capture. Spectral data provides unique spectral signatures that indicate the inherent properties of materials and allow objects to be located, identified, and analyzed more easily. Video in general enables surveillance to take place over time; it can be used in motion detection, the tracking of moving objects, and computer vision. One would expect, therefore, that spectral video systems could be applied in novel ways to assist with computer vision — e.g., spectral-reflectance information-based surveillance, illumination-independent object tracking, and temporal spectral analysis. This thesis takes a critical approach to this assumption, aiming to provide an understanding of the importance of, and need for, spectral video systems in practice.

Numerous research publications in various areas of science and industry demonstrate the importance of spectral information by replacing conventional imaging devices with spectral systems in real–life applications – e.g., in medical imaging, art paintings, industrial-production line-quality control, food safety, and the wood and pulp industry (Haneishi et al., 2000; Fält et al., 2009; Hirvonen et al., 2014; Feng et al., 2012). Several of these applications also rely on video monitoring. For instance, the real-time visualization of spectral images during surgery and the monitoring of temporal changes in organs and tissues is of great import to surgeons (Hasnat et al., 2016).

Currently, there are a number of difficulties related to spectral video systems. For instance, a compromise between spatial, spectral, and temporal resolutions is always necessary due to hardware limitations (Cao et al., 2011). In addition, increased resolutions usually result in the creation of a bulky instrument (Feng et al., 2014). In practice, therefore, spectral video output is still effectively possible only for sensors with relatively small number of spectral bands. The utilization of a smaller number

of bands also allows for shorter exposure times, which tend to be suitable and desirable, where video acquisition is concerned. In this thesis, one such system – the FluxData (FD) seven-channel spectral video system – is explored (as described in more detail in Section 3.2.2). As the FD system is a novel one, which, to the author's knowledge, has not been evaluated in the public domain, the first task was to characterize it in order to understand its performance and shortcomings. For this reason, the first part of the thesis is devoted to the characterization of the FD system in relation to both still-image and video output. The results of this characterization effort will be discussed in Sections 4.1 and 4.2. Following this step, the FD system was applied practically to record, analyze, and track the human eye.

The human eye is an organ that provides the sense of sight; up to 80% of all impressions are perceived via this organ (Kaplan, 2015). Vision is a vitally important sense as the eyes are active in almost every activity that we perform: working, reading, writing, driving, communicating, exercising, cooking, and eating, to name but a few (Land and Hayhoe 2001). By processing the light reflected or emitted from an object, the eyes facilitate the interpretation of the object's shape, color, texture, and dimension. Human beings use this organ to observe and learn about the surrounding world; arguably, sight is used more than any of the other senses. It is no surprise, therefore, that the eye is considered such an interesting object to study, observe, monitor, and analyze. The works summarized in Section 3.3 present details about the main spectral signatures of the eye region and can be said to motivate spectral research into applications concerning the human eye. Indeed, researchers in many fields (including medicine, biometrics and eye-vision research) have explored the potential of using spectral imaging systems to measure the activity of the human eye (see Section 3.4).

All of the aforementioned points have served to inspire the spectral-video-based study performed for this thesis. To the author's knowledge, the potential of spectral eye tracking has not yet been investigated; the database offered here – SPEED – is the first, therefore, to serve such a purpose. SPEED is the first publicly available combined spectral image/spectral video database to introduce the potential for spectral video data to track eye features (i.e., the skin, iris, pupil, blood vessels, sclera, and hair). The database will be introduced in detail in Section 4.3.1 and its potential will be discussed in Sections 4.3.2, 4.3.3, and 5. SPEED provides a platform for creating new spectral-data-based enhanced devices, software, and algorithms that could also be used in other eye-related research areas (e.g., medicine and eye/vision research).

This thesis is organized as follows: first, in Section 2, the main aims of the study are explained as are the methods employed to achieve them. Section 3 provides brief introductions to spectral color theory (Section 3.1), the most typically used spectral measurement methods (Section 3.2.1), and state-of-the-art spectral video acquisition (Section 3.2.2). Section 3.2.3 presents the main working principles for the FD system used in this thesis. This is followed by descriptions of the spectral signatures of the

human eye (Section 3.3) and state-of-the-art spectral eye research (Section 3.4). Section 4 offers a summary of contributions to this thesis, as well as a critical overview of the FD system and its applications in recording the human eye. An overview of SPEED is provided in Section 4.3. Finally, the main claims and findings of this thesis are discussed in Section 5.

2 AIMS OF THE STUDY

The main aim of this study is to research the usability and potential of the FD spectral video camera in practical applications. Attempts have been made to discover whether or not there is a practical need for such a system, as well as to identify the trade-offs it comes with. Finally, the system was challenged with a fast moving object: the human eye.

Throughout the thesis the following aims were addressed:

1. To characterize a spectral video system in terms of single frame output

   ⇒ The characterization was performed in terms of spectral sensitivities, linearity, spatial uniformity, spatial sensor alignment and noise. In addition spectral reflectance estimation was evaluated for 400-700 nm range using empirical regression methods and the Digital ColorChecker SG and ColorChecker charts.

2. To characterize a spectral video system in terms of video output

   ⇒ The characterization was performed in terms of noise, frame mutuality/correlation, and spectral and color performance. It was executed in two main parts – monochromatic video analysis and spectral video analysis. Channel-wise statistics and spectral analysis via spectral reflectance estimation were the applied analytic tools.

3. To gather a spectral video database of the human eye

   ⇒ A novel database consisting of 180 seven-channel spectral eye motion videos and 30 still fifty-one-channel spectral images of the left eyes of 30 voluntary human subjects was created. Unfavorable conditions, such as eyewear reflections, extreme angles and make-up, were also incorporated.

4. To use the gathered spectral color data and analyze the main features of the human eye while performing an observational task

   ⇒ The performed spectral image and video analysis reveals six features with unique spectral signatures: skin, iris, blood vessels, pupil, sclera and hair.

5. To run preliminary tests with existing computational tools on the collected eye data for spectral analysis, reflectance-based object segmentation and multiple eye feature tracking.

⇒ Seven-channel spectral data was analyzed over the recorded video frames for both standard conditions and unfavorable conditions. First Derivative and Spectral Angle Mapper (SAM) were tested for feature detection and segmentation.

3 BACKGROUND

The theoretical background is organized into four main sections. In Section 3.1, the physical phenomenon of electromagnetic radiation reflected from an object is discussed and reflectance is defined as a desirable, measurable, physical quantity. This is followed by an explanation of the main principles of reflectance measurement in Section 3.2, with special attention being devoted to methods for gathering spectral video data. In Section 3.3, the meaning of spectral signatures is introduced and a discussion is commenced about how these signatures vary between different features in the human eye region – namely, the pupil, the iris, the sclera, blood vessels, hair, and skin. The chapter closes with Section 3.4, in which the different fields in which the human eye is already examined spectrally are introduced and the benefits of using spectral video for eye analysis are highlighted.

3.1 SPECTRAL COLOR THEORY

Human beings refer to what they see via the concept of color. However, there is no absolute definition of color as such because, ultimately, interpretations of color are subjective (Ohta and Robertson, 2005; Hardeberg, 2001; Wyszecki and Stiles, 1982; Kaiser and Boynton, 1996). An interpretation is made in the brain as the result of a visual sensation that stimulates light-sensitive cells. One could even think of color as an illusion. For this reason, it cannot be described reliably and it is not directly measurable. The visual sensation that forms the illusion of color, however, can be described and measured. This sensation is termed the "color signal." Figure 3.1 demonstrates the process of the color signal's formation schematically. In more detail, the color signal can be defined as the radiant power of a given magnitude and spectral composition that enter the eye and produce the sensation of color (Wyszecki and Stiles, 1982). In other words, color signal is, in fact, colored light that is formed due to varying amounts of radiant energy for different wavelengths (Hardeberg, 2001). A general term *spectrum* will be used throughout this thesis for any spectral distribution that describes how radiation is distributed over the electromagnetic spectrum. The radiant energy in question is called electromagnetic radiation and it consists of quanta called photons (Wyszecki and Stiles, 1982). The energy of a photon is inversely proportional to its wavelength (unit nanometers nm) and the wavelength defines the range of electromagnetic spectrum. Based on wavelength electromagnetic radiation can be classified into the following groups (from shortest to longest wavelength): gamma-rays, X-rays, ultraviolet, visible, infrared, microwaves, and radio waves. However, the human eye is sensitive only to a relatively narrow band of wavelengths ranging from 380 nm to 780 nm (1 nm = 10^{-9} m) (Wyszecki and Stiles, 1982). This wavelength band is called the visible spectrum (VIS).

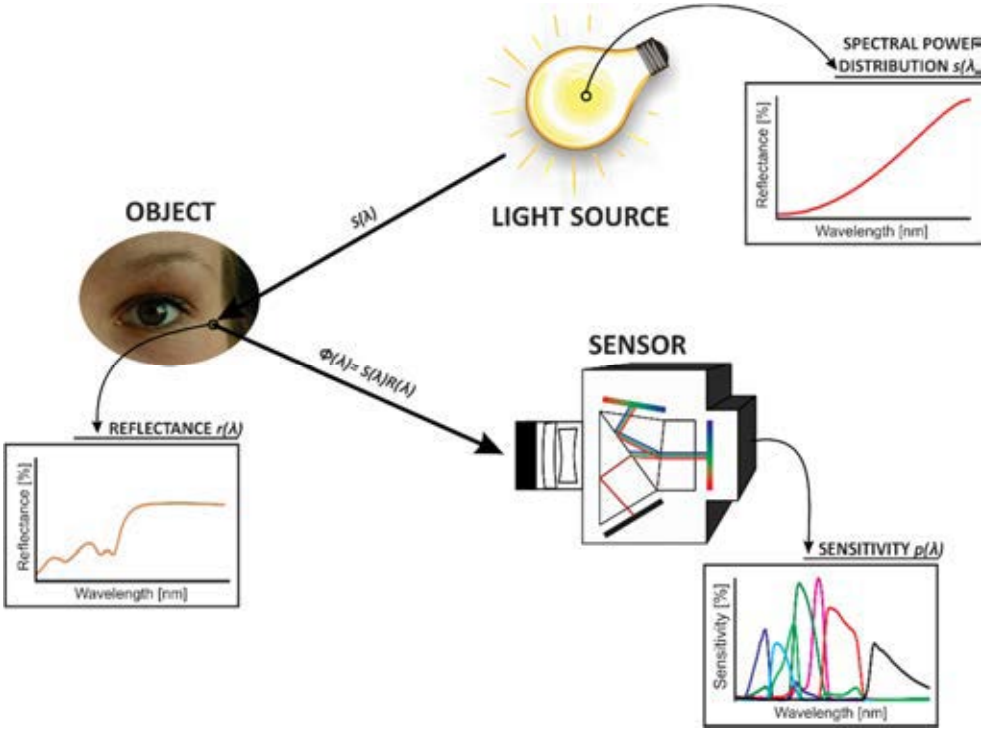

Figure 3.1 Schematics of color signal formation and acquisition. $s(\lambda)$ corresponds to the spectral power distribution of the light emitted by the light source, $r(\lambda)$ corresponds to the spectral reflectance of the object,$\Phi(\lambda)$ corresponds to the color signal, and $\rho(\lambda)$ t corresponds to the spectral sensitivity of the detector (in order to relate this scheme clearly to the current thesis, the actual eye from the SPEED database is shown here as the object and the FD seven-channel system is shown as the detector).

In order to measure a spectrum three main elements are needed: illumination, an object and a sensor. Illumination provides the electromagnetic radiation initially needed for a sensory response (human or artificial). Once this radiation hits the surface of a material spectral interactions between the incoming radiation and the object's surface material happen. Observing this interaction is a key to finding spectral signatures that reveal crucial material properties. In general, radiation can be transmitted, absorbed and reflected. Following Figure 3.1 let's focus on the reflected radiation. The reflected radiation here is a color signal that reaches the sensory system. This signal has a lot of terms in literature. It can be understood as radiance coming from the surface, irradiance falling on the sensor, luminous flux falling on the sensor but all of these terms have the same meaning – they refer to the color signal. In this work color signal will be notated as $\Phi(x, y, \lambda)$. The color signal is highly dependent on the spectral distribution of the light that illuminated the surface. More precisely, it is the combination (product) of the illumination spectrum of the light source and the reflectance spectrum of the object. Reflectance is a true,

illumination independent property of a material that determines which percentage of the incoming illumination spectrum will be reflected at each wavelength. Since spectral reflectance is a characteristic property of an object it allows to investigate the inherent object properties and it is a desired feature to be measure and explored.

3.1.1 Mathematical model of reflectance measurement

In order to measure reflectance the process of sampling and quantization must be performed. Also, the reflectance data must be "extracted" from the color signal. The following section will present the main equations that model these processes.

The components that are included in a system for reflectance measurement are: the illumination source, the sensor and the object (see Figure 3.1). We denote $s(\lambda)$ as the spectral distribution of the illumination and $\rho_i(\lambda)$ as the spectral sensitivity function of the sensor's i^{th} detector. If a point (x, y) on the object's surface has reflectance $r(x, y; \lambda)$, then the signal detected by the sensor from this point can be mathematically modelled as (Wyszecki and Stiles, 1982):

$$v_i(x,y) = \int_0^\infty s(\lambda) r(x,y;\lambda) \rho_i(\lambda) d\lambda + v_{i,\text{dark}}(x,y) \qquad (3.1)$$

where $v_{i,\text{dark}}(x,y)$ represents the dark noise for the sensor's i^{th} detector.

Equation 3.1 is limited as it does not take into account geometrical effects (i.e. angles of incident light and observation; specular reflection) and effects such as fluorescence, polarization, sub-surface penetration, etc. Even with these limitations in mind, Equation 3.1 remains very useful as a model for the analysis of the interaction between light and objects (Hardeberg, 2001).

All of the listed functions are considered to be discrete n-dimensional vectors within a certain wavelength range, i.e., $s(\lambda), r(x, y; \lambda)$ and $\rho_i(\lambda)$, and in practice become column-vectors $\boldsymbol{s}, \boldsymbol{r}, \boldsymbol{\rho}_i \in \mathbb{R}^n$, respectively. For example, reflectance is expressed as:

$$\mathbf{r}(x,y) = [r(x,y,\lambda_1), r(x,y,\lambda_2), \ldots, r(x,y,\lambda_n)]^{\mathrm{T}} \qquad (3.2)$$

where T denotes transpose. Now, Equation 3.1 can be written in matrix form, as follows:

$$v_i(x,y) = \mathbf{w}_i^{\mathrm{T}} \mathbf{r}(\mathrm{x},\mathrm{y}) + v_{i,dark}(x,y) \qquad (3.3)$$

where the vector

$$\mathbf{w}_i = diag(\mathbf{s})\boldsymbol{\rho}_i \qquad (3.4)$$

which describes the combined spectral effect of the illumination and the i^{th}detector.

If the sensor has m spectrally unique detectors, one obtains m values for a point (x, y). One can now define a vector $\mathbf{v} \in \mathbb{R}^m$ which contains the m detected values $v_i(x, y), i = 1, \ldots, m$, as follows:

$$\mathbf{v}(\mathrm{x}, \mathrm{y}) = \mathbf{W}^{\mathrm{T}}\mathbf{r}(\mathrm{x}, \mathrm{y}) + \mathbf{v}_{\mathrm{dark}}(\mathrm{x}, \mathrm{y}) \tag{3.5}$$

where $\mathbf{W}$ is a $n \times m$ matrix and has the vectors $\mathbf{w}_i$ on its columns, and vector $\mathbf{v}_{dark} \in \mathbb{R}^m$ contains the measured dark noise values $v_{i,dark}$.

From Equation 3.1 it can be noted that the response of the sensor is in fact the response to the color signal. In order to obtain reflectance from this color signal post-processing is needed i.e. flat-field correction.

If measurements are made from a perfectly reflecting white diffuser sample, for which the reflectance is $r_{white}(x, y; \lambda) = 1$ within the wavelength range of interest, one gets:

$$\mathbf{v}_{\mathrm{white}}(x, y) = \mathbf{W}^{\mathrm{T}}\mathbf{r}_{\mathrm{white}} + \mathbf{v}_{\mathrm{dark}}(x, y) \tag{3.6}$$

Now, the flat-field correction is done by computing the ratio of the measured sample spectrum and the white reference spectrum:

$$\mathbf{r}(x, y) = \frac{\mathbf{v}(x, y) - \mathbf{v}_{\mathrm{dark}}(x, y)}{\mathbf{v}_{\mathrm{white}}(x, y) - \mathbf{v}_{\mathrm{dark}}(x, y)} \tag{3.7}$$

This operation removes the effect of the illumination from the color signal so that one can obtain reflectance data and it also removes the effect of the dark noise. If this operation is performed for every spatial location within the measured frame, then spatial non-uniformities that might occur due to the spatial non-uniformity of the light source or the detector are also removed. Data obtained in this way is spectral reflectance factor, a ratio of the reflected light from a specimen to the reflected light from a perfect reflecting diffuser under identical specified conditions.

In practice flat-field correction is performed in such a way that consecutive measurements (point or frame-based) of a desired object, a perfectly reflecting white reference plate and a dark reference measurement (with the device closed from any incoming light) are made under the exact same illumination conditions and device settings. Then the object measurement is post-processed following Equation 3.7 where $\mathbf{v}_{\mathrm{white}}(x, y)$ would be the white reference data and $\mathbf{v}_{\mathrm{dark}}(x, y)$ would be the dark reference data.

Although a perfectly reflecting diffuser is assumed to be a perfect white in practice it is not. Therefore, to compensate for these imperfections one can introduce

to the post-processing the calibrated reflectance of the perfectly reflecting diffuser, $\boldsymbol{r}_{calibrated}(x, y)$, as folows:

$$\mathbf{r}(x, y) = \frac{\mathbf{v}(x, y) - \mathbf{v}_{\text{dark}}(x, y)}{\mathbf{v}_{\text{white}}(x, y) - \mathbf{v}_{\text{dark}}(x, y)} * \boldsymbol{r}_{calibrated}(x, y) \qquad (3.8)$$

This calibrated reflectance is usually provided by the manufacturer or if needed can be measured with a high precision spectral measurement device (i.e. spectrophotometer).

3.2 SNAPSHOT SPECTRAL VIDEO SYSTEMS

Spectral imaging (or imaging spectrometry) is a growing area of research that aims to measure three-dimensional data of the object space (Sellar and Glenn, 2005). As the use of the word "imaging" in the name suggests, two of the three dimensions (x and y) are spatial dimensions, while the third dimension (λ) is the spectral, (spectrometry) dimension (see Figure 3.2). Three-dimensional data formed in this way is commonly known as a spectral image or a spectral cube. Practically speaking a spectral cube is a stack of images of the same scene acquired at multiple wavelengths of light, which are called channels. This method thus allows the spectral signature to be measured for every point in the scene being imaged.

Despite rapid technological developments, simultaneous capture of all three dimensions of the spectral cube is still a challenging task. Where practical applications are concerned, therefore, the majority of spectral-imaging instruments rely on some sort of scanning. One must first capture two dimensions at a time and stack those 2D slices in a sequence. Recently however, several so called snapshot spectral imaging systems have become commercially available (Hagen and Kudenov, 2013). The name "snapshot" refers to their capability of capturing the whole spectral cube in a single detector integration period. These novel systems remove the need for scanning and enable much faster imaging to take place – in some cases, even spectral video recording is possible. In this thesis, one such novel system is presented, analysed, and applied practically: the FD seven-channel snapshot spectral video system

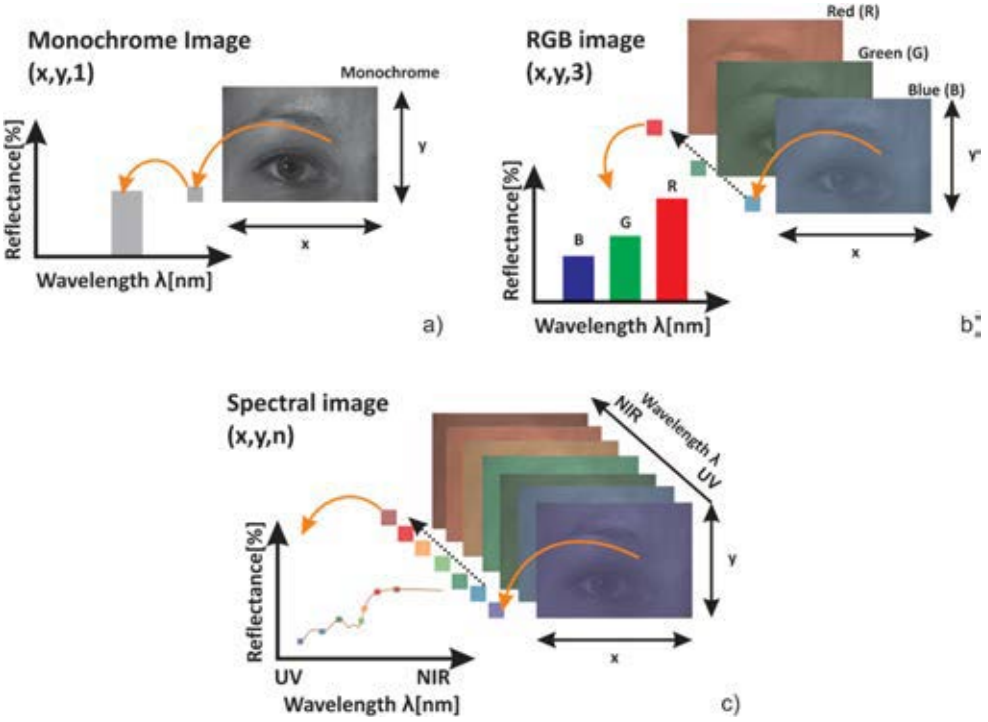

Figure 3.2 Comparison between a) a monochrome image, b) an RGB image and c) a spectral image. At each spatial location (x, y) monochrome image has one numerical intensity value only. An RGB image has three intensity values - red, green and blue channels respectively. A spectral image contains many spectral channels, enabling a more detailed spectral signature for the point to be acquired.

.

3.2.1 Spectral systems – general overview

Spectral systems can be divided into two large groups: spectral measuring systems and spectral imaging systems. Spectral measuring systems are used primarily to quantify the spectral characteristics of objects. They include both spectral measuring devices (i.e., spectroradiometers [RAD], spectrophotometers, bispectrophotometers) and filter-based instruments that directly measure color (i.e., colorimeters) (Sharma, 2003). All of these devices perform point-based measurements, where the size of the measurement point depends on the construction of the device and can reach a radius of few millimeters. The output for such a measurement, therefore, is a value averaged over the area of the measurement point. Consequently, in practice, these devices can be used to measure spatially uniform scenes only, one measurement point at a time. Figure 3.3 below presents one of the typical methods used for spectral point measurements.

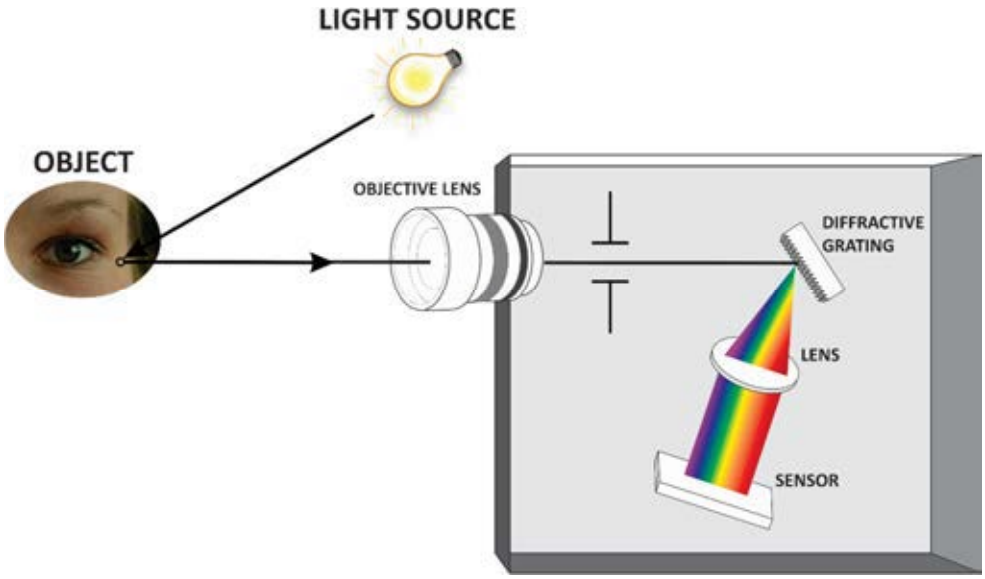

Figure 3.3. Spectral point measurement system (spectroradiometer). The device measures a spectrum from a single point on the object. Light enters the device through the objective lens and disperses into its spectral components via a dispersive element (in this case diffractive grating). The dispersed light is then collected using a line-detector element.

Spectral imaging systems, on the other hand, acquire both spatial and spectral information. They record spatial information by sampling the scene using two-dimensional electronic sensors (a Charge-Coupled Device [CCD] or a Complementary Metal Oxide Semiconductor [CMOS]) and they record spectral information by capturing images at different wavelengths on the electromagnetic spectrum. The output is a three-dimensional image cube of a certain spatial and spectral resolution. These devices can be used to measure spatially varying spectral information.

Based on the acquisition principle of the three-dimensional cube, spectral imaging systems can be divided further into line-based and frame-based systems. Line-based systems capture spatial and spectral information one line at a time and require the inclusion of a moving stage in order to record the whole scene line by line. As such, the spatial scanning and stacking of the scanned lines are required in order to form the spectral cube. Figure 3.4 below displays one of the typical methods used for spectral line measurements.

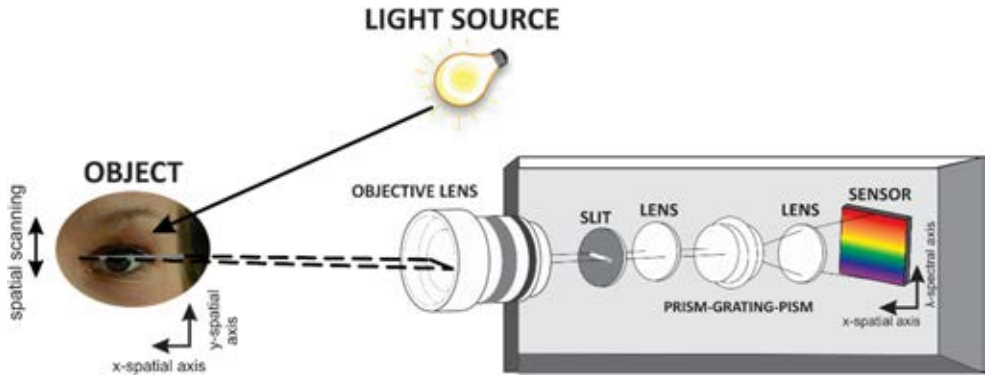

Figure 3.4 Line measuring spectral system (Line-Scan [LS]). This system records all the spectra from a single line on the object. Light enters the device through the objective lens. It is guided through a narrow slit and dispersed into its spectral components by a prism-grating-prism component. The dispersed light is then collected using a two-dimensional detector element.

Frame-based systems can be used to capture the whole scene and make the spatial acquisition process much faster. However, with the majority of available frame-based systems, spectral scanning is still required. Thus, we can divide the frame-based systems into scanning frame-based spectral imaging systems and snapshot spectral imaging systems. The former perform spectral scanning by consecutively transmitting the image through a number of color filters with different spectral transmittances. They can be band-pass filters (i.e., acousto–optic tunable filters, abbreviated AOTF – see Morris et al., 1994) or tunable filters (i.e., liquid crystal tunable filters, abbreviated LCTF – see Hardeberg et al., 2002). Scanning devices can have high spectral resolutions (tens and hundreds of bands) but the scanning process limits the application of these systems to static scenes – and therefore single frame acquisitions. Figure 3.5 demonstrates the LCTF method utilized for frame-based spectral scanning measurement.

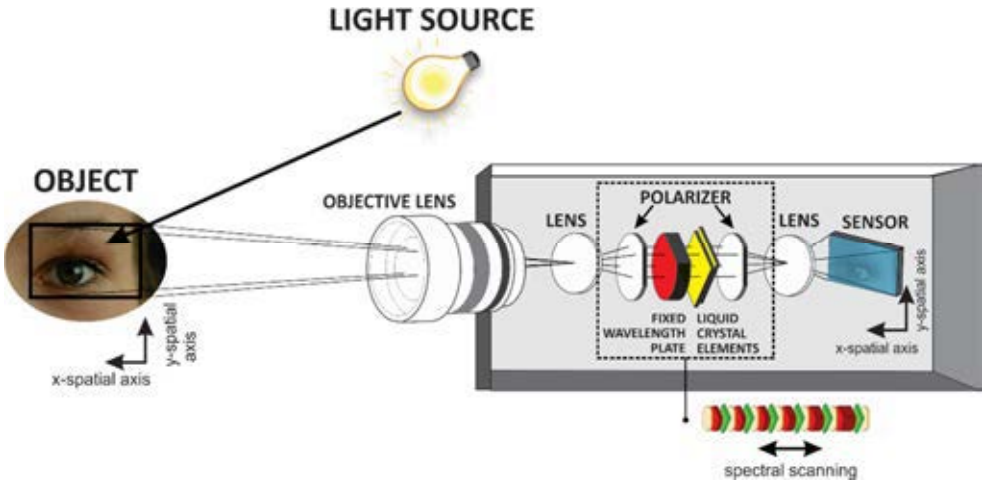

Figure 3.5. Frame-based spectral imaging system (LCTF). LCTF is comprised of several cascaded stages. The dashed line section shows an enlarged diagram of an individual stage of LCTF. The principle of operation is based on liquid crystal elements and retardation in phase of light rays to pass certain wavelengths of light. Liquid crystals are electronically controllable band–pass filters that can be tuned to a desired wavelength via the rotation of the crystals' axes when a voltage is applied. Polarizer elements ensure that only the wavelengths that are in phase are transmitted to the next stage. This allows for the capturing of full-frames one particular central wavelength at a time.

Unlike scanning, snapshot spectral imaging systems collect both spatial and spectral data during a single detector integration time. In the existing literature, 13 different snapshot spectral imaging techniques have been reported (Hagen and Kudenov, 2013). In some, the system can even allow for spectral video capture. One such system will be presented in more detail in Section 3.2.3, and will be analyzed and applied practically throughout the thesis.

Of those 13 existing techniques, two will be introduced here, as the system used in this thesis is, in fact, based on a combination of these two principles. The two principles in question are beamsplitting and color filter array (CFA) (McDuff et al., 2014; Monno et al., 2014; Losson et al., 2010). The working principle of beamsplitting is based on the separation of light into a certain number of spectral bands and the simultaneous detection of each band with an independent sensor (see Figure 3.6a). Even though scanning is avoided in this way, the trade-off that these systems face relates to the spectral resolution, as beamsplitter designs can allow for the separation of approximately six bands maximum. Moreover, it is not easy to split the light into more than four bands without compromising the system's performance (Hagen and Kudenov, 2013). A CFA is a filter mosaic created by placing color filters over the individual pixels of a sensor in a certain pattern in order to capture multiple bands simultaneously (see Figure 3.6b). In post-processing, this method requires demosaicing and results, therefore, in a loss of true spatial resolution for some

channels. Despite the spatial loss, this is a very popular technology in consumer cameras. The most widely used mosaic is the Bayer mosaic (see Figure 3.6b) (Losson et al., 2010). This type of system is not considered to be a spectral snapshot system in itself as it captures very sparse spectral information via three channels only (red, green, and blue). When combined with additional external filters or computational methods of spectrum estimation, however, it can become one. Recent research has even led to the development of multispectral filter arrays that will allow for the simultaneous capture of more than three channels (Shinoda et al., 2015).

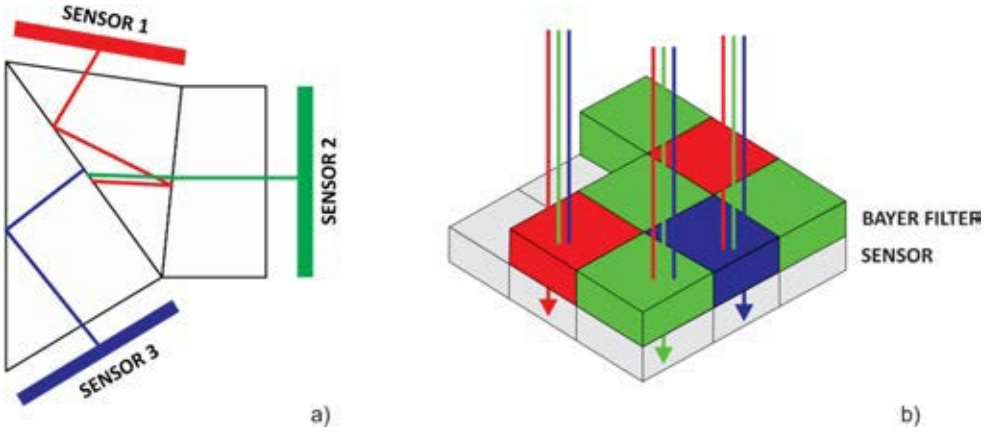

Figure 3.6 Schematics of two snapshot spectral imaging technologies: a) beamsplitting and b) CFA. Beamsplitting uses cemented cubes with filters placed between the inner surfaces. When light reaches an inner surface it is split into two directions. Thanks to the filter, only a specific spectral composition can travel in each direction until it reaches the sensor. Each split component is detected via an independent sensor. CFA on the other hand places color filters on individual sensor elements in a certain pattern that is called a mosaic. The most commonly used mosaic is the Bayer filter mosaic. Each pixel is filtered to record one of RGB colors. In order to obtain the missing two colors interpolation is needed. This process is called demosaicing.

As spectral still camera systems are often too slow to capture spectral images from dynamic objects, there is an evolutionary need for the extension of spectral imaging to the recording of spectral video. Nowadays, such system creation is a hot topic as spectral video acquisition with high spectral resolution is still a challenging task. When using the systems that are available for spectral video recording, a trade-off must be made in terms of spatial, spectral, and temporal resolutions, as well as physical size and complexity.

3.2.2 Spectral video systems – state-of-the-art

The beginnings of spectral imaging can be traced back to the early nineteenth century, when spectral imaging systems were developed mostly for use in astronomy (Hagen and Kudenov, 2013). Similarly, the roots of spectral video could be found in this era. Astronomers at the time faced problems with the devices that were available due to motion artifacts and poor light-collection efficiency, and so they began to look at non-scanning techniques for spectral image capture. Image slicing (Bowen, 1938), fiber reformatting (Vanderriest, 1980; Kapany, 2004), and lenslet-based pupil array dispersion (Lee, 2001) emerged as the first three techniques for doing exactly that. They were given the name *snapshot spectral imagers* as they are able to acquire a full spectral cube within a single detector integration time period. With the growth of airborne remote sensing, the use of spectral imaging expanded from astronomy to include agricultural assessment and management (Everitt, 1991). Additionally, after the launch of Landsat in 1972, field-portable spectral imaging systems started to be created in order to support the calibration of the satellite. This led to the realization that new approaches for system design are still needed to decrease scanning artifacts even further. One of these approaches involved the usage of beamsplitting elements (Stoffels, 1978) and multiple-camera capture systems. These were the first attempts to acquire videos with more than three spectral channels. Unfortunately, at the time, the technology could not accommodate more than four channels (King and Vlcek, 1990; Everitt et al., 1991; Phinn et al., 1996). As computers advanced, however, the option of computational sensing became feasible and new approaches to reconstruction-based imaging techniques were developed using computed tomographic imaging spectrometry (CTIS) (Descour and Dereniak, 1995). Toward the end of the nineteenth century and during the first decade of the twentieth century, the field of snapshot spectral imaging came to include a vast amount of novel acquisition technologies (Hagen and Kudenov, 2013): Multiple Filtered Camera, Tunable Echelle Imager, Spectrally Resolving Decatur Arrays, Image-Replicating Imaging Spectrometer, Coded Aperture Snapshot Spectral Imagers (CASSI), Image Mapping Spectrometry, Snapshot Hyperspectral Imaging Fourier Transform Spectrometer and Multispectral Saganc Interferometer. One of the reasons why such a large variety of techniques emerged was the need to overcome the same obstacles that the astronomers faced: motion artifacts and poor light throughput.

Some of the snapshot technologies mentioned above have been adopted for use in spectral video systems. Recent applications of spectral video have involved the use of Lenslet Arrays (Bodkin et al., 2009), CTIS (Descour and Dereniak, 1995), CASSI (Wagadarikar et al., 2009), and beamsplitting (**Publication II**). Other approaches have also been taken involving hybrid systems (Du et al., 2009; Cao et al., 2011a; Ma et al., 2014; Cao et al., 2011b; Feng et al., 2014) and systems that employ spectral reflectance estimation (Morovič and Haneishi, 2006; Hasnat et al, 2016).

CTIS and CASSI are methods that treat multispectral imaging as a reconstruction problem. In the 1990s, Descour and Dereniak (1995) presented an improved design for a visible-spectrum CTIS (470–770 nm). In order to capture multispectral frames at video rates, CTIS uses diffraction gratings to generate multiple, spectrally dispersed projections of the scene on the camera sensor in a single shot. These projections are then reconstructed into the final 3D datacube. Since then, there has been significant progress in the use of CTIS. In 2001 the first high-resolution CTIS became available, providing a 203 × 203 × 55 datacube on a 2048 × 2048 CCD camera. In 2005, successful demonstrations of this system's reflective design were made (Hagen and Kudenov, 2013). There are still many remaining problems, however, regarding the computational complexity of CTIS, as well as limitations related to calibration difficulties and measurement artifacts. For these reasons, CTIS can be used only in very simple scenes, as reported in the literature (Hagen and Kudenov, 2008).

CASSI was the first attempt to use compressive sensing theory within snapshot spectral imaging (Hagen and Kudenov, 2013). In the last decade, Wagadarikar et al. (2009) have used CASSI for the spectral imaging of a dynamic scene at video rate, as demonstrated with a video of lit candles. The spectral range of the system was 450–650 nm, with 33 acquisition channels and a 30 fps video rate. The problem is that, although CASSI provides a high spectral and temporal resolution, the spatial resolution is low, at just 256×280 pixels.

Other types of system that can achieve high temporal resolutions have also been presented. Bodkin et al. (2009) present the HyperPixel Array™ (HPA) Imager that works on the principle of a pinhole or lenslet array and acquires a 20-channel datacube using up to a 30 fps video rate. The spatial resolution is recorded, however, as 180x180 pixels only. Park et al. (2007) have developed a system that captures a six-channel video at 30 fps via multiplexed coded illumination. The drawback of this system is that it cannot be used effectively outside of the laboratory as it is limited by the need for controlled lighting. In Du et al. (2009) and Cao et al.'s (2011b) work, a prism-mask multispectral video system (PMVIS) is demonstrated. This system samples the light coming from a scene using an occlusion mask, then disperses it into its constituent spectra via a triangular prism, before capturing the data with a monochrome camera. Here, due to the occlusion mask, a decent spatial resolution is exchanged for gain in spectral resolution. In one way or another, every system seems to face challenges.

In order to combat trade-offs where spatial, spectral, and temporal resolutions are concerned, researchers have begun to develop hybrid system solutions (Cao et al., 2011a; Ma et al., 2014; Feng et al., 2014). Cao et al. (2011a) and Ma et al. (2014) present one such example, in which a high-spatial-resolution RGB video camera is added to an earlier presented PMVIS system. As PMVIS sacrifices spatial resolution for spectral resolution, the addition of the RGB camera compensates for the loss. The two camera systems are exposed simultaneously by the incorporation of a mirror into the system apparatus. The frames from the two cameras are aligned so that each scene

point that is measured by the multispectral camera has a corresponding RGB pixel. However, optical distortions caused by the prism and blur caused by the occlusion mask were identified as drawbacks, and so the system was improved by Feng et al. (2014), who added an objective lens and an AMICI prism element (AHVIS). In this way, the system can capture objects at far-off distances with improved light throughput and less optical distortion. One limitation that remains, though, is the fixed nature of the sampling points on the occlusion mask, which leads to both a lack of flexibility and data loss, in addition to the problem of the system's bulky size.

In practical applications, one can find even more diverse spectral video systems. As early as 2005, Iwama et al. (2005) presented a real-time multispectral and multiprimary video system that was later applied to telemedicine, digital archives, electronic commerce, and computer graphics by Yamaguchi et al. (2006a, 2006b). The system was composed of a six-channel HDTV camera system and six-primary display system. The camera system consisted of two conventional three-CCD cameras, with special color filters placed in front of each. The incoming light was divided into two optical paths by a half mirror in order to expose the two cameras simultaneously. In this way, a six-channel video can be acquired. Also, the display system used was a six primary DLP display. It was shown that more accurate colors can be reproduced with a multispectral, multiprimary video system than with a conventional HDTV system due to the higher fidelity and wider color-gamut reproduction. The main drawbacks to this system are its bulky size and the complicated operational principles involved. In addition, as these studies were focused on color reproduction, only six channels in the visible range were captured.

Recently, spectral video systems have been applied in the field of endoscopy (Leitner et al., 2013; Arnold et al., 2013; Hohmann et al., 2011; Hohmann et al., 2016), with researchers taking advantage of the development of fast spectral scanning techniques. Leitner et al., (2013) use an AOTF to achieve switching times below 1 ms, thus making it possible to acquire an eight-channel, five fps video for the detection of cancerous tissue. In Arnold et al.'s (2013) study, full spatial resolution (1002x1004 pixels) and 10-channel videos have been achieved but with four fps only. Hohmann et al. (2016) have even attempted to modify an endoscope's light-source unit to make it capable of using six wavelength bands. The resolution in this system is approximately 350x370 pixels (also limited by the normal resolution of the traditional endoscope).

Other aspect to be considered when acquiring a spectral video is the data size. First attempts to compress temporal sequences of spectral images were done in early 2000 (Koponen et al., 2000). At a time the spectral image sequence was generated from a computer animation in RGB format and then converted to spectral images as direct acquisition of spectral video was not possible. The work highlighted the need to separately perform compression in both spatial and spectral dimensions exploring the potential of motion estimation and Principal Component Analysis.

As these trade-offs regarding spatial, spectral, and temporal resolutions seem to be persistent and inevitable, some researchers have proposed solutions that rely on computational spectrum estimations, rather than proposing device-oriented solutions. Such approaches usually use simple acquisition systems to acquire data with high temporal and spatial resolutions, and apply spectral reflectance-estimation algorithms in post-processing in order to gain a higher spectral resolution. Morovič and Haneishi (2006) utilize a six-channel video camera to test the performance of multiple known techniques for estimating the reflectance from device response. They conclude that multiple factors influence the choice of reflectance estimation algorithm for spectral videos, such as computational, noise, and accuracy requirements. Recently, Hasnat et al. (2016) have reported on a framework for generating spectral video using an RGB video only. They apply this technique using neurosurgical operational microscope RGB videos. The paper discusses the importance of selecting an appropriate training set for the estimation algorithms and the need for a compromise between the frame rate and the speed of estimation when generating a near-real-time spectral video.

Table 1 summarizes the preferences for the spectral systems mentioned above in sections 3.2.1 and 3.2.2. This table offers a review so that the trade-offs discussed above can be overviewed. This data can be considered valuable for making decisions about the applications for which a system could be suitable. Moreover, where this thesis is concerned, the information enables a basic comparison to be made between the FD system utilized in this study and other state-of-the-art spectral systems. The following section describes the FD system in detail and reflects on the data in Table 3.1.

Table 3.1. Summary of the state-of-the-art spectral acquisition systems

	System	Reference	Wavelength Range [nm]	Number of Channels	Acquisition Mode	Spatial Resolution [px]	Frame Rate [fps]
POINT	RAD	Photo Research, Inc. 1999	380 - 780	201	point-based	point (1/8° - 2° area)	NA
SCANNING	LS	Specim	380 - 1000	240	line-based	1032 per line	NA
	LCTF	Publication II	450 - 950	51	frame – based spectral scan	1024 x 1392	NA
	AOTF	Leitner et al., 2013	400 - 650	8	frame – based spectral scan	1Mpx	5
SNAPSHOT	CTIS	Hagen and Kudenov, 2013; Ford et al, 1999	VIS	~50	snapshot CTIS	~ 203 x 203	~ 15
	CASSI	Wagadarikar et al., 2009	450 - 650	33	snapshot CASSI	256 x 280	30
	HPA	Bodkin et al., 2009	420 - 670	~20	snapshot lenslet	~ 180 x 180	30
	6-band HDTV	Yamaguchi et al., 2006a	VIS	6	snapshot beamsplit	HDTV	NA
	PMVIS	Cao et al, 2011b	400 - 1000	~ 100 depends on fo-cal length	snapshot PMVIS	~ 121 x 98 mask dependent	~12
	AHVIS	Feng et al., 2014	400 - 800	80	snapshot hybrid	1024 x 768	30
	RGB	Hasnat et al., 2016	VIS	3 via estimation up to 31	snapshot CFA	720×576	25
	FD	Publication II	380 - 1000	7	snapshot beamsplit CFA	1040 x 1392	30

3.2.3 Seven-channel snapshot spectral video system FluxData 1665-MS7

Scanning based spectral imaging devices due to either wavelength scanning or spatial scanning are not able to record a spectral video of dynamic targets. First attempts of using fast scanning AOTF were only recently reported in the field of endoscopy (Leitner et al., 2013) however, the achieved frame rates are still low. Snapshot spectral imaging devices, on the other hand, are able to achieve higher frame rates, however it comes with either lower spatial or spectral resolution. Some solutions for providing spectral video systems with high spatial, spectral, and temporal resolutions have been proposed by Ma et al. (2014) but the complexity of these systems is still a limitation, as is their bulkiness. All the methods mentioned above shows how the field of spectral video still faces challenges and must trade-off spatial, spectra and temporal resolution, and systems tend to be complex and large. There is plenty of room for improvement, therefore, including the development of new hardware, software, analysis methods, and practical applications. In this thesis, one kind of a commercially available snapshot spectral video system is analyzed and applied. The system preferences seem promising: optical elements facilitate fast video acquisition, and its compactness and size ease its usage and application. This section presents the snapshot spectral video system in question, FluxData (FD) 1665-MS7, giving an overview of it pros and cons. Table 3.2 summarizes the system specifications.

Table 3.2. Device specification of FluxData 1665-MS7 (FluxData, 2011)

Image device	Sony ICX285 – Basler scA1400 – 30gm/gc
Number of image devices	3 (2 Bayer BG color and 1 monochrome)
Number of channels	7
Sensor size	1040 x 1392 px (Pixel Size: 6.45 x 6.45 µm)
Bit depth	12-bit
Frame rate	30 f/s
Wavelength range	380 – 1000 nm
Lens	Carl Zeiss Planar 1,4/50 ZF-IR
Output	image and video
Size	9.1 x 11.7 x 11.2 cm
Weight	1.25 kg (without lens)

The system was custom-made in such a way that it provides a seven-channel capture with six VIS channels and one NIR channel. This seven-channel capture is achieved by employing a beamsplitting dichroic prism element and three CCD sensors, of which two are CFA sensors. Simultaneous exposure of the three sensors allows for spectral video recording at a speed of up to 30 fps.

The FD system uses the idea of splitting the light via a dichroic prism assembly, yet controlling the spectrum with interference filters that are coupled between the prism elements (see Fig. 3.7). These elements are placed at the core of the system to

split the light three ways. One CCD sensor is placed at each of the three sides of the prism element. In order to describe the working principle of the device, we shall refer to Figure 3.7. Let us assume the incoming radiance is an ideal, flat, equi-energy spectrum. After passing through the lens system, the light hits a low-pass filter located at the first inner surface of the prism element. This filter allows all of the wavelengths lower than 750 nm to pass through; wavelengths of 750 nm and upwards are directed to the bottom sensor (marked as CCD 3 in Figure 3.7). The sensor is a monochromatic sensor and, as such, measures only the near-infrared part of the incoming spectrum. The transmitted light continues to travel inside the prism until it hits the second inner surface. This surface contains a band-pass filter, with one band for wavelengths of 470 – 540 nm and one band for wavelengths of 620–730 nm. The light that passes through here is directed to the middle sensor (marked as CCD 2 in Figure 3.7). The remaining two bands (380–460 nm and 530–610 nm) are directed to the top sensor (marked as CCD 1 in Figure 3.7). It should be noted that the values mentioned are ideal theoretical values; the real filters do not have such sharp cut-offs for wavelengths.

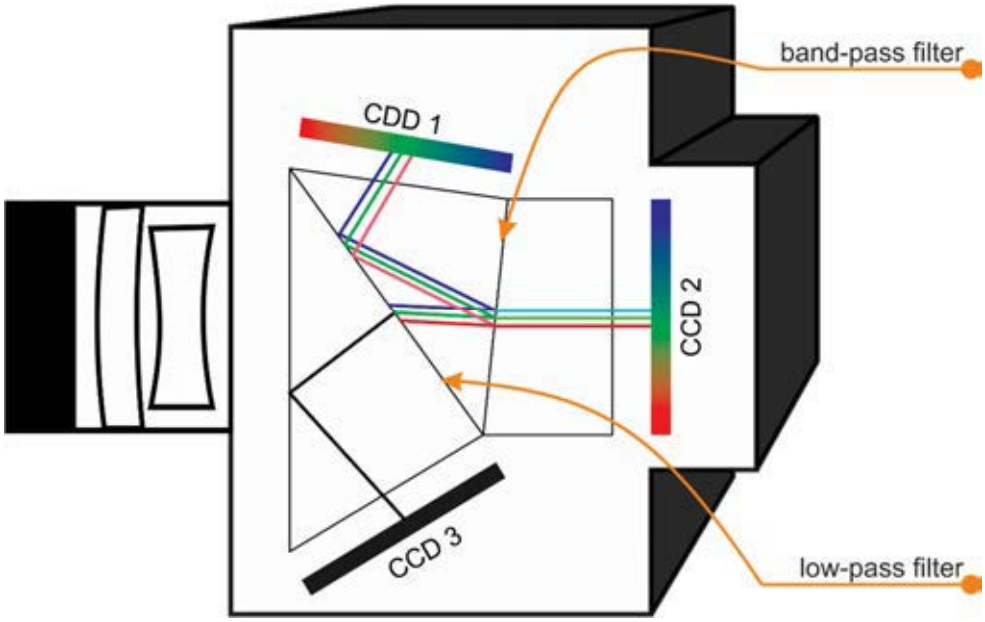

Figure 3.7. Schematics of FluxData 1665-MS7, showing its working principle.

CCD 1 and CCD 2 are color sensors of the same kind. They contain red (R), green (G), and blue (B) filters, which are organized in a Bayer pattern (Basler scA1400 - 30gc) (Basler 2017a). CCD 3 is a monochromatic sensor with the sensitivity of a standard silicon sensor (Basler, 2017b). Unlike other visible sensors this sensor does not have a NIR cut-off filter making it suitable for capturing NIR information. The

final shape of the sensitivities of the seven channels, therefore, can be understood as a combination of the filters used in the dichroic prism, the filters used for the Bayer pattern, and the sensitivity of the silicon sensor. The procedure for measuring the sensitivities and their shape is reported in **Publication I** and will be discussed in Section 4.1.1. Throughout this thesis, the sensors utilized will be referred to using the notation given in Figure 3.7. The channels will be referred to by letter and number notation, where the letter suggests the part of the spectrum to which the channel is sensitive (R, G, and B) and the number suggests to which of the two color sensors the channel belongs. The third sensor is sensitive to near-infrared light only, and the channel that belongs to it will be referred to as NIR.

The sensors have individual Ethernet connections, facilitating the separate, sensor-wise adjustment of exposure times. This feature might seem handy at first; in its practical application, however, it is also a limitation, in a way. As the color sensors host three different spectral bands with different levels of sensitivity, the bands must share the exposure time that is inherent in the sensor to which they belong. In this case, sensor-wise exposure-time control does not imply channel-wise exposure-time control, which would be more desirable, in terms of maximizing the dynamics of every channel. Additionally, sensor-wise exposure-time control limits the selection of light sources available for recording with the FD system – for example, if an incandescent light source is used, the short wavelength channels will receive a very small amount of light, while the long wavelength channels will receive too much. As a solution, one can use "daylight-like" sources that emit equal amounts of light for every wavelength, or one can design one's own light source to custom-emit light to suits the FD system best. Another limitation of sensor-wise exposure-time control is reflected in the temporal content of the recorded video. If the separate sensors had different exposure times during video recording, each of them would capture slightly different content, temporally speaking, which would lead to spatially different content as well. Such a result would affect the pixel-wise accuracy of the recorded spectra, leading to the need for additional post-processing for every frame in order to match the content.

When working with FD one must take into account size of the output data. One seven-channel image is about 20 MB in size, which means that one five second seven-channel video can be as large as 3 GB, depending on the used frame rate. For off-line uses of FD, this is not a big limitation; however, in on-line applications, it poses operational challenges.

Noise in the system – especially during video acquisition – is also an aspect that must be investigated. Both **Publications I** and **II** report the results of noise analysis, and will be discussed in Section 4.2.2.

Table 3.1 allows for a simple comparison of the FD system to be made with existing state-of-the-art spectral imaging and video systems. Compared with scanning spectral imaging systems, the FD system provides coarser spectral sampling. However, the 30 fps video output enables the acquisition of spectral

information for dynamic targets, which, in general, is not yet possible with scanning spectral imaging systems. In contrast with some other snapshot video systems, the FD system trades spectral resolution for gains in spatial and temporal resolution. On the plus side, the system is compact in size, and can be used for both VIS and NIR information capture. Also, if needed, one could estimate denser spectra in computational post-processing via reflectance factor estimation.

The overview given above of the main features of the FD system provide the motivation for examining the model in more detail. In this thesis, three publications are presented, in which the FD system is studied in more detail. Two of these publications explore the behavior of the system, while the third publication tests its applicability for recording and analyzing a dynamic object – i.e., the human eye.

3.3 SPECTRA OF THE HUMAN EYE REGION

In this thesis, the key focus is on the dynamic spectral measurement and analysis of the external features of the human eye region – i.e., the skin, the iris, blood vessels, the pupil, the sclera, and hair. In order to understand the spectral signatures of these features, both the internal and external structures must be examined, as well as their interactions with light. In this section of the thesis, Figure 3.8 is the core reference point for the graphical representation of the cross section of the right eye and its internal structure (Figure 3.8a), the external structure of the eye (Figure 3.8b), and an example of the reflectance spectra for the six external features, as measured in SPEED (Figure 3.8c).

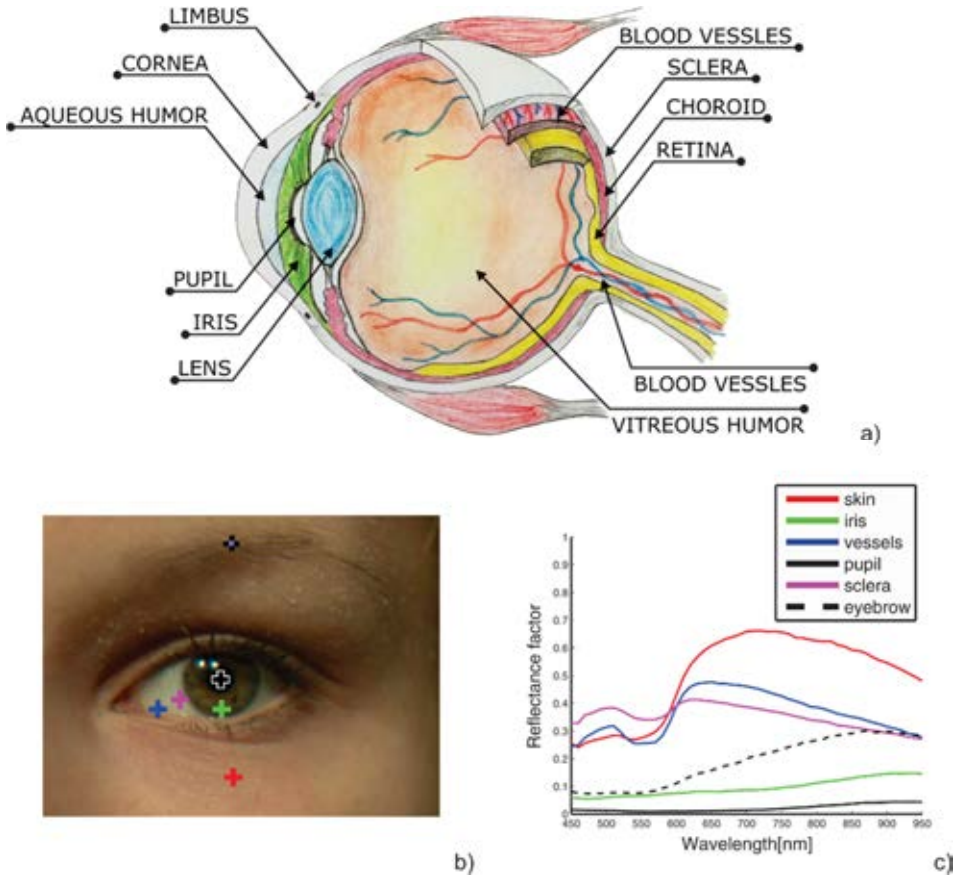

Figure 3.8. a) Internal features of the human eye; b) External features of the human eye – an RGB image generated from the LCTF 51-channel spectral cube; c) An example of the reflectance spectra for the six external features measured in SPEED (EID31 LCTF 51-channel data in the 450–950 nm range, with 10nm sampling). The representative spectra have, on average, a 10x10 pixel area.

Light enters the eye through the *cornea*. The cornea is a transparent refractive element that transmits over 99% of visible light (refractive index of n = 1.376 at 555 nm) (Edelhauser and Ubels, 2003; Liou and Brennan, 1997). Its role is to bend light rays towards the pupil. It transmits radiation from 300 nm in ultraviolet light to 2500 nm in infrared, with typical water-absorption peaks at 1430 and 1950 nm (Boettner and Wolter, 1962). From 300 nm, transmittance increases rapidly, reaching about 80% at 380 nm; from 500–1300 nm, it is greater than 90%. While most of the light enters the eye, a small part reflects back into the environment due to the convex surface of the cornea, which acts as a mirror. Images formed in this reflection are called Purkinje images (Jóźwik et al., 2014).

After passing the cornea, light passes through a chamber filled with a transparent liquid called the *aqueous humor* (n = 1.336, Liou and Brennan, 1997). The aqueous humor has a high water content (about 98%) and its optical and mechanical properties are a result of the specific organization of collagen fibers within the liquid (Vecino and Sharma, 2011). It begins transmitting at 220 nm in ultraviolet light, with a strong absorption band at 265 nm, and continues transmitting at up to 2400 nm in infrared (Boettner and Wolter, 1962). In the visible region, transmittance is high, being slightly less than that in an equal thickness of water, and having absorption peaks at 980, 1200, 1430, and 1950 nm.

Before entering the eye, light is focused by the elastic *crystalline lens* (n = 1.408, Uhlhorn et al., 2008). Its transmission starts in ultraviolet and continues to 1900 in infrared, with the usual water-absorption bands (980, 1200 and 1430 nm). Both the starting point and the amount of transmission in ultraviolet and in short wavelength VIS vary with the age of the eye. In a young eye, transmission begins increasing rapidly at about 390 nm and reaches 90% at 450 nm, while in a 63-year-old lens, although transmission begins at 400 nm, it does not reach 90% until 540 nm (Boettner and Wolter, 1962).

Inside the eye, light travels through a transparent gelatinous substance called the *vitreous humor* (n = 1.336, Liou and Brennan, 1997). It is over 99% water, and contains small amounts of collagen fibrils, hyaluronic acid, and glucose (Lund-Andersen and Sander, 2003). The vitreous humor transmits from 300 nm in ultraviolet light, and no transmittance is noted beyond 1400 nm in infrared, There are strong water-absorption bands at 980 and 1200nm (Boettner and Wolter, 1962).

The final surface that light hits inside the eye is called the *ocular fundus*, which contains the *retina* – the layer containing photosensitive cells, rods, and cones, which absorb the light and convert it into nerve impulses, thus facilitating image formation. There is a layer of blood vessels and connective tissue lying over the retina called the *choroid*. There is also a third layer, the *sclera*, which is the so-called "white" of the eye. The function of the sclera is more protective than optical (Meek, 2008). The sclera surrounds the eye ball, apart from at the very front where it connects to the cornea at the *limbus*. Unlike the cornea, which is transparent, the sclera is totally opaque, which is due to the much-more interwoven collagen and elastic fiber within this layer, as

well as the much wider fibrils that are present. The magenta line in Figure 3.8c indicates the measured spectra for the sclera in SPEED. The sclera exhibits the highest reflectance of all the features in the VIS part of the spectrum, while in NIR, its reflectance becomes lower due to the water content.

The anterior part of the sclera contains the *blood vessels* that supply the iris and sclera. This network of vessels consists of arterial (red lines in Figure 3.8a) and venous (blue lines in Figure 3.8a) vessels. Spectrally speaking, arterial and venous blood differ from one another due to differences in oxygen content. Chromophores are the chemical compounds that determine the spectral shape of tissue. Hemoglobin, the main chromophore in blood, can be oxygenated or deoxygenated, depending on whether it is observed in arterial blood or in venous blood. Oxyhemoglobin's spectral signature exhibits absorption maxima at 415, 542, and 577 nm, while deoxyhemoglobin exhibits absorption maxima at 430, 555 and 760 nm (Zonios et al., 2001). This creates a very characteristic "W" shape in the spectrum of the blood, which is also recognizable in the spectrum of other blood-containing features in the eye region – e.g., the skin.

The last two anterior features of the eye are the *iris* and the *pupil*. The *iris* is a radial muscle structure that adaptively adjusts the size of the *pupil* and thus controls the amount of light that can reach the bottom of the eye. The iris is usually strongly pigmented with melanin. This chromophore contains two distinct heterogeneous macromolecules: the dark-brown eumelanin and the reddish-yellow pheomelanin (Hosseini et al., 2010). Eumelanin has a slightly higher absorbance than phaeomelanin, but both decrease exponentially with the wavelength, and become very low in the NIR region – especially after 750 nm (Anderson and Parrish, 1981; Hosseini et al., 2010). The ratio of the two melanins varies in different features and has been shown to have an impact on the appearance of the eye (Thody et al., 1991). Darkly pigmented irises have been found to contain higher amounts of eumelanin and higher eumelanin/phaeomelanin ratios than light-colored irises (Medina et al., 2011). Keratin, carotene, bilirubin, lycopenic acid, and licorice are secondary factors that affect the overall appearance of the iris (Anderson and Parrish, 1981).

Melanin also contributes to the spectral signatures of *skin* (Angelopoulo et al., 2001) and *hair* (Cesarini, 1990). In the eye region, hair is found in the form of eyelashes and eyebrows. Eyelashes grow at the edges of the eyelid, protecting the eye from objects that might enter. Eyebrows grow above the eye to prevent sweat, water, and other debris from falling into the eye.

The *pupil* is a hole in the center of the iris that allows light to enter the eye. It appears to be black because of the absorption of light by tissues inside and within the eye.

The eye is surrounded by skin. The most common chromophores in skin are hemoglobin and melanin. In addition, water is an element that contributes to the spectral shape of skin (and of other features as well). The spatial distribution of these chromophores determines the appearance of skin. Eumelanin plays a fundamental

role here, in terms of appearance and photoprotection. It is found in the top layer of the skin (the epidermis). Hemoglobin is found in the bottom layer (the vascular network of the dermis). These chromophores create a typical spectral signature for skin with a persistent "W" pattern at around 560 nm (Angelopoulo et al., 2001). In the NIR part of the spectrum the spectrum of skin is determined by both chromophores and water content.

This overview demonstrates that there are six main spectrally different features in the eye region: the pupil, the iris, the sclera, blood vessels, skin (eyelids and the area around the eye), and hair (eyelashes and eyebrows). These features contain light-absorbing chemical compounds called chromophores, which contribute to the spectral signature of each feature. The most significant chromophores here are melanin, hemoglobin, and water. The spectral measurement of these features is presented in **Publication III**. A more detailed analysis of the measured spectra is given in Section 4.3. These signatures have been exploited for many applications, facilitating the detection, differentiation, segmentation and analysis of human eyes. The following section will describe them state-of-the-art research areas in which spectral imaging techniques have already been implemented in applications involving the human eye. Furthermore, the section will introduce the potential to implement spectral analysis in a new field: eye tracking.

3.4 SPECTRAL EYE RESEARCH – STATE OF THE ART

In many fields, researchers have explored the notion of using spectral imaging systems to measure the human eye. Vilaseca et al. (2008) present a six- and seven-channel multispectral system for measuring the reflectance and color of the human iris. Their work indicates the feasibility of the spectral characterization of the iris, as well as suggesting its potential use in applications such as prosthesis and contact-lens manufacturing. Medina et al. (2011) demonstrate how 20-channel spectral imaging can be used to characterize melanin and noninvasively diagnose ocular diseases. In the field of biometrics, extensive research has been conducted into enriching iris-image databases with spectral information. As Zhang et al. al. (2015) mention, existing iris-recognition systems operate predominantly in a single band of the NIR spectrum, using mostly narrow wavelength bands that peak at around 850 nm, which limits their use. The publicly available databases BATH, CASIA, ICE2005, ICE2006, MMU, and WVU, which are used for biometric purposes, all contain NIR images only. In addition, the UPOL and UBIRIS databases contain VIS images only (Proenca et al., 2010). Only the UTIRIS database collects both VIS and NIR images (Hossein et al, 2010). While the majority of researchers focus on interpreting the NIR patterns of the iris, studies have shown that important information is also contained in the complex textures that are formed from pigments (Hosseini et al., 2010).

The eye region consists of many features – namely, the pupil, the iris, the sclera, blood vessels, skin (eyelids and the area around the eye), and hair (eyelashes and

eyebrows). In addition to having important structural properties, all of these features contain chromophores, which carry important spectral signatures. These signatures are usually lost when single-band capture is used. The different features could be captured, however, at multiple wavelengths. Many researchers, therefore, have designed their own spectral imaging systems to be able to capture both structural and pigment-related information. For instance, Boyce et al. (2006) use just four spectral channels to demonstrate the potential for spectral imaging to enhance the segmentation of eye features. Ngo et al. (2009) utilize a 12-channel system to study how wavelengths affect iris recognition performance in detail. Where these studies are focused on frontal gaze, Crihalmeanu and Ross (2012) propose a solution for capturing a non-frontal gaze by combining multispectral sclera-texture and vasculature-pattern information, thus enhancing recognition. In a recently published book, Zhang et al. (2015) showcase a collection of studies in which the optimal bands for clustering, multispectral iris recognition, and image fusion techniques are investigated. Such research indicates the strong interest in the scientific community in incorporating spectral systems into studies related to the eye.

The majority of the challenges faced in the aforementioned studies relate to switching between physical filters or illumination bands in order to scan the electromagnetic spectrum for capturing a spectral image (see Table 3.3). Such procedures make the acquisition process slow, and speed is crucial when imaging a fast-moving object such as the eye. In addition, every time switching or scanning is needed recording of a video becomes impossible. Some of the systems involve switching the filter or light manually, some involve manual focusing; some use solutions to overcome these issues, such as electronically controlled aids. However, in all cases, the systems seem to be quite complicated. As such, the need to capture a fast-moving object using a much simpler and faster system remains. Table 3.3 indicates the numbers of spectral channels and the wavelength ranges that the existing systems cover. Most of them operate dominantly in the VIS range and capture a low number of channels. Some attempts have been made to combine VIS and NIR, and to increase the number of channels. The resulting systems, however, tend to be relatively slow and bulky. Therefore, smaller and faster systems that capture more channels are required.

Another field in which the human eye is central is that of eye tracking. Movements of the eye is recorded to facilitate the exploitation of gaze direction for various needs, such as interactions with computers and the understanding of cognitive processes (Hansen and Ji, 2010). These recordings are made in the form of image or video and are used to create databases that serve as learning tools for generating gaze-direction estimation algorithms. Learning by Labeling (LL) and Learning by Synthesis (LS) are just two of many recent database examples (Tonsen et al., 2016; Wood et al., 2016). Usually the images and video data in these datasets are recorded with cameras that are sensitive only to the VIS or NIR parts of the electromagnetic spectrum (Wood et al., 2016; Tonsen et al., 2016; Proenca et al., 2010). The data typically consists of

standard three-channel RGB or single-channel monochrome grayscale images and videos (see Table 3.3). This means that, due to the construction of existing eye-tracking hardware, a large amount of wavelength-dependent information gets lost and remains unexploited.

All of the information above serves as motivation to apply the available small, fast, seven-channel FD system in tracking the eye. In preparing for this study, a collaboration was initiated between two fields and the first steps for applying spectral data analysis in eye tracking were made. The key hypothesis is that there is a benefit to exploiting multiple wavelengths for the purpose of eye tracking. To the author's knowledge, the potential of spectral eye tracking has not yet been investigated; the database provided here is the first to serve such a purpose. The mission was to create a publicly available combined spectral image/spectral video database to introduce the possibility of using spectral video data to track eye features (i.e., the skin, the iris, the pupil, blood vessels, the sclera, and hair). The result is SPEED. Unlike other databases, SPEED incorporates spectral acquisition at video rates with eye-tracking elements, such as observational tasks and glint. This allows the spectral signatures of eye features to be explored while, at the same time, the features are tracked and the gaze is interpreted. The database provides a platform for creating new spectral-data-based enhanced eye-tracking devices, software, and algorithms. Moreover, the spectral datasets could be of use in other eye-related research areas, such as medicine and eye/vision research.

In **Publication III,** the SPEED data-collection method is presented, alongside example spectral images and videos.

Table 3.3 Summary of the state-of-the-art spectral eye research

	Dataset	Reference	Wavelength Range	Number of Channels	Switching
SPECTRAL RESEARCH	Vilaseca et al.	Vilaseca et al., 2008	VIS	6-7	YES motorized filter wheel
	Medina et al.	Medina et al., 2011	VIS	20	YES LCTF
BIOMETRICS	BATH	Proenca et al., 2010	NIR	1	NA
	CASIA	Proenca et al., 2010	NIR	1	NA
	ICE2005	Proenca et al., 2010	NIR	1	NA
	ICE2006	Proenca et al., 2010	NIR	1	NA
	MMU	Proenca et al., 2010	NIR	1	NA
	WVU	Proenca et al., 2010	NIR	1	NA
	UPOL	Proenca et al., 2010	VIS	3	NA
	UBIRIS	Proenca et al., 2010	VIS	3	NA
	UTIRIS	Hossein et al, 2010	VIS and NIR	4	YES manual 2 cameras
	Boyce et al.	Boyce et al., 2006	VIS and NIR	4	NO Snapshot
	Ngo et al.	Ngo et al., 2009	VIS and NIR	12	YES illumination
	Crihalmeanu and Ross	Crihalmeanu and Ross, 2012	VIS and NIR	4	NO Snapshot
	Zhang et al.	Zhang et al., 2015		12	YES illumination
EYE-TRACKING	LL	Wood et al., 2016	NIR	1	NA
	LS	Tonsen et al., 2016	VIS	3	NA
	SPEED	Publication II	VIS and NIR	7	NO Snapshot
	SPEED	Publication II	VIS and NIR	51	YES LCTF

4 SPECTRAL VIDEO: A SUMMARY OF CONTRIBUTIONS

In the previous chapters, a summary of existing spectral systems and state-of-the-art spectral eye research has been presented. Additionally, a novel seven-channel spectral video system has been introduced. The information provided serves as the basis for the spectral-video-based experiments conducted for this thesis. In this chapter, the experiments and the main contributions of the thesis will be delineated via a summary of the three publications that have been born of this project. These publications focused on analyzing the spectral video system used here and applying the system practically in spectral eye-video acquisition.

Upon interacting with the novel spectral video system for the first time, there was a need to gain an understanding of its working principles and its accuracy. It was found that the specifications provided by the manufacturer do not offer sufficient information about system-calibration procedures. In addition, it was taken into account that differences might occur from device to device as a result of variations caused in the production line. In order to avoid inaccurate result, which might be achieved if one relied only on the system properties published by the manufacturer, the spectral video system was analyzed and characterized using several procedures. Those procedures were selected to serve as relatively simple tools one can use for profiling a spectral video system, almost like a recipe. For **Publication I** a set of characterization data was collected for a single image output while in **Publication II** the characterization data was extended for a video output. As these publications compile many methods and to avoid repetition, here the discussion is focused on three aspects of the characterization procedure that are considered to be the most relevant for the later application of the device. Those are: spectral sensitivities as they determine the range of application for the system, spectral reflectance estimation as it allows for estimating reflectance spectra of higher spectral resolution and calculations of color, and noise as it, if present, corrupts the measured signal.

After achieving a better understanding of the system, the next step was to test its performance in real applications. There are a number of circumstances in which there is a need to conduct monitoring via spectral video – i.e., industrial inspections and medical procedures (Hirvonen et al., 2014; Hasnat et al., 2016). The author's main motivation for selecting the human eye as an application was a personal interest in the human anatomy and physiology, as well as all its related applications (e.g., eye tracking, medicine, surgery, diagnostics, and biometrics). In addition, the fact that the eye is a moving object that has multiple features with distinct spectral signatures has led the organ to be of keen interest to the scientific community, as demonstrated by the great number of relevant publications in diverse fields. The involvement of spectral still-image technologies in these fields is a hot research topic. Finally, by

bringing a fascination with the human eye together with spectral imaging, the author saw the possibility of introducing spectral techniques to a novel field – i.e., eye tracking. Thus, the potential for developing spectral-data-based eye-tracking hardware and software was determined. This led to the development of the hypothesis explored in this work: the idea that there are benefits to exploiting multiple wavelengths for the purposes of recording, analyzing, and tracking the eye. To test this hypothesis, the author has compiled a novel combined spectral image/spectral video database named SPEED (SPectral Eye vidEo Database). The database is presented in **Publication III** and will be discussed further in Section 4.3.

4.1 COLOR AND IMAGE CHARACTERIZATION

For **Publication I**, we collected a set of methods through which to characterize a spectral still camera and applied those methods on the FD spectral video system, for its still image output. The main motivation was to understand the behavior of the system and identify its strengths and weaknesses. The compilation of color and image characterization methods consist of: spectral sensitivity, linearity, spatial uniformity, spatial sensor alignment and noise measurements. Those where chosen in order to understand the efficiency of light detection of each individual CCD in the FD system and to be able to identify weather there is any differences between them. In addition, spectral reflectance estimation was performed and evaluated for 400-700 nm range using empirical regression methods and the standard color charts (ColorChecker Classic [CC24] and Digital ColorChecker SG [DSG96]). Together all these provide a functional portfolio of the FD system.

The characterization data collected for **Publication I** can be used, in fact, in calibration of the FD system and in its preparation for practical application. From the findings in **Publication I** this thesis benefits the most from the following:

- The actual spectral sensitivities of the FD system have been determined, thus making them available for this project and for all future research (for example, see Bartczak et al., 2015).
- The knowledge gained about the location of the peak sensitivities and the Full Width Half Maximum (FWHM) of the seven channels play vital role in identifying possible fields of application for the FD system.
- The sensitivity measurements taken reveal weaknesses of the FD device (i.e., a sensitivity gap between 715 nm and 740 nm). The effect of the sensitivity gap on the estimation of the reflectance spectra is made evident in **Publication II**.
- The spectral-reflectance-estimation experiments conducted here suggest that the spectral and color accuracy achieved with the apparatus is reasonable for use with standard color charts, which indicates the suitability of the FD system for use in applications that require the acquisition of colorimetric and spectral data.

- Miss-alignment of the three CCD sensors was identified here and a procedure for correcting this problem was proposed. In this way it was ensured that the content of a frame is spatially aligned for all the channels and the spectral data for every pixel is spatially accurate.
- The need for flat-field correction was identified as small amount of spatial non-uniformity was determined as being present in all the CCDs.

4.1.1 The spectral sensitivity of FluxData 1665-MS7

In general, knowing the sensitivities of a device can help one to describe the relative efficiency of its light detection. As the FD system has many optical elements, it is important to know the sensitivities with all the optical components included in the optical path. For this reason, sensitivity measurements were performed in the range of 380 nm to 910 nm with a 5nm wavelength interval. The resulting spectral shapes are presented in Figure 4.1. Details of the measurement procedure can be found in **Publication I Section 3.1**.

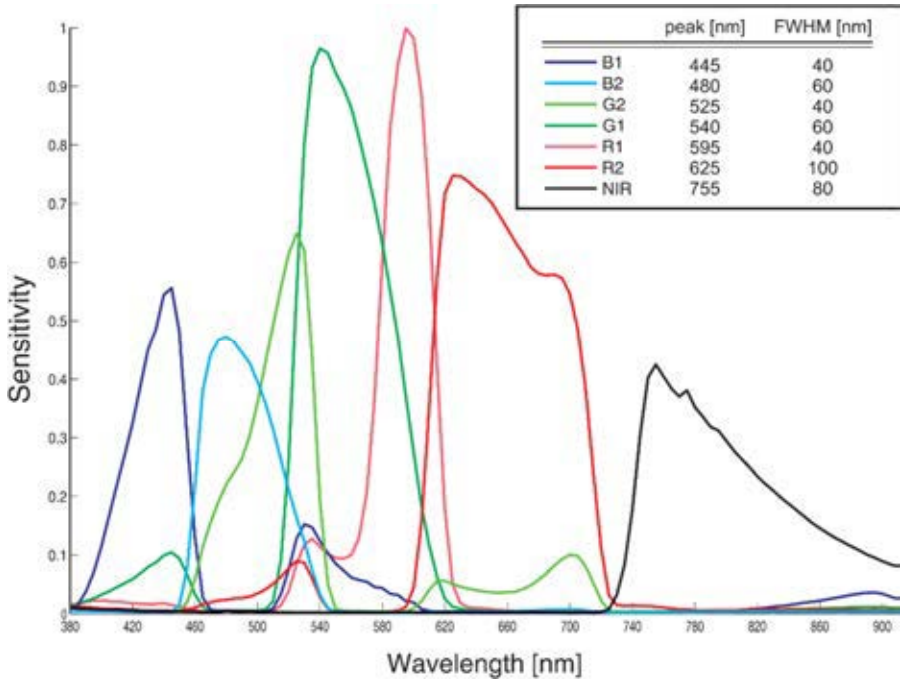

Figure 4.1. Measured sensitivities of FD in 380 – 910 nm range with 5 nm wavelength interval. Legend shows the sensitivity peaks and their Full Width Half Maximum (FWHM). The measurements were performed using GigaHertz Optik Integrating Sphere (diameter of 500 mm), monochromator with a halogen light source (placed at 45°), Hamamatsu PMA-11 optical fibre spectrometer (measuring the radiant flux) and the FD system (placed at a 0° angle). The FD sensitivities were obtained by dividing the camera response (spatial average of a 100 x 100 pixel area) for each channel with the measured radiant flux of the light.

Besides reporting on the actual sensitivities these measurements provided valuable information for selecting a suitable range of applications for the FD system. One example application, the human eye, was therefore explored here. Looking at Figure 4.1, we can see how the sensitivity bands for the FD system coincide with some of the main chromophores and elements discussed earlier in Section 3.3. The absorption peaks of oxyhemoglobin coincide with the sensitivities of channels G1 and R1, while the absorption peaks of deoxyhemoglobin coincide with the sensitivities of channels B1, G1, and NIR. In the existing literature, it was shown that the density of blood capillaries is reflected around 550 nm (in FD system G1) and veins are enhanced in the 620 nm (in FD system R2) or longer wavelength (Yamaguchi et al., 2006b). This data suggests that the FD channels mentioned can be used to detect blood vessels in the eye – especially when it comes to segmenting them from the sclera.

Absorption of both melanins decreases in the NIR region, especially after 750 nm, which coincides with the sensitivity of the NIR channel. This channel could be used, therefore, to differentiate between the pupil and the iris in a way similar to that used in traditional eye-tracking hardware. Research also indicates that melanin quantities can be observed at around 470 nm (Yamaguchi et al., 2006b). Channel B2 could provide more information, therefore, about melanin patterns in the iris.

The first absorption peaks of water appear in the NIR region; they also coincide with the sensitivity of the NIR channel. It should be noted, therefore, that this channel could be used to differentiate between eye features in terms of water content.

These findings suggest that use of the FD system is applicable to all fields in which samples containing these chromophores and elements are used – e.g., medicine and biometrics. The region of the human eye, as examined in this thesis, has features that contain these chromophores and elements (the skin, the iris, blood vessels, the pupil, the sclera, and hair). Moreover, as the FD system can be used to record a spectral video, it can be considered worthwhile to test capability of detecting, analyzing, and tracking these eye features with this model.

In addition to the elements mentioned in Section 3.3, it is clear that the FD system could be used for other applications. For instance, its green, red, and NIR channels coincide with the absorption peaks needed for vegetation and moisture discrimination (FluxData Inc., 2011).

4.1.2 Spectral reflectance estimation

The aim of reflectance estimation is to map a low dimensional device response v to a corresponding higher dimensional reflectance r. This computationally turns a low-spectral-resolution device into a high-spectral-resolution device. Improved spectral resolution is one way to improve the ability to discriminate between objects; it is therefore desirable in spectral imaging. Reflectance estimation has inspired many researchers to explore the idea of turning spectral systems into spectrophotometers

– including even consumer cameras. Such a method would render spectral acquisition more economical and widely available, not to mention making it faster. In **Publication I**, focusing on VIS range only, the possibility of increasing the low, six-channel response of the FD system is explored in relation to still-image output. In **Publication II**, this work is extended to video output and full spectral range of the system.

The methods used to perform and evaluate the reflectance estimation are presented in **Publication I, Section 3.2**. We estimated the spectral reflectance factors for the standard chart color samples (CC24 and DSG96) in 400–700 nm range using empirical kernel based regression models: Polynomial kernel (Heikkinen et al., 2008) and Martén kernel (Heikkinen et al., 2013). **Table 4** in **Publication I** summarizes the results of the reflectance estimation accuracy in terms of spectral (Root-Mean-Square Error [RMSE] and Pearson Distance [PD]) and color errors (ΔE_{00} for CIE D65, A and F11 illuminants). The figures for both spectral and color errors suggest that the reflectance estimation here is reasonably accurate for standard color charts with an average RMSE varying between 0.0183 and 0.0243, and an average ΔE_{00} color accuracy varying between 1.63 and 2.42 units. The results depend on the illuminant and estimation methods, suggesting that the best combination has to be determined for each specific application.

4.1.3 Comparison of FluxData 1665-MS7 to other spectral systems

In the present study, to extend on the work done in **Publication I**, an attempt has been made to compare the spectral accuracy and color accuracy of the FD system (when reflectance estimation is employed) with other spectral systems available at the Spectral Color Research laboratory at UEF. Even though a direct comparison of different devices' spectral and color accuracy might seem unfair due to variations in working principles and applications, such a procedure can give some indication of where, in terms of spectral and color quality, the FD system might be on a scale of spectral devices. Table 4.1 summarizes the results of this comparison.

In order to obtain the results reported in Table 4.1, the following devices were used: the PerkinElmer (PE) Lambda 1050 spectrophotometer (PerkinElmer, Inc., USA – used for the spectral ground truth), the Specim V10E line scan (LS) system (Specim, Ltd., Finland), the Nuance EX LCTF-based system (CRi, Inc., USA), the Nikon D800 (ND800) digital RGB-camera (Nikon, Inc., Japan) and the FD system. All the devices were employed using their most common measuring geometries and settings. To perform the comparison spectral and color difference metrics (Goodness-of-Fit Coefficient [GFC], RMSE and ΔE_{00}) were computed between each devices output spectra and the PE ground truth for the 24 patches of the CC24 chart. In the case of the FD system and the ND800, prior to the comparison being made, a spectral reflectance factor estimation was performed using an empirical linear regression model (Heikkinen et al., 2008) and DSG96 training data. To avoid repetition of the

training samples in the evaluation set, the 96 color patches of the DSG96 chart imaged by LS were used as training data. A comparison of the devices is then possible across three spectral ranges: 400–700 nm, 450–700 nm, and 450–900 nm. The evaluations we completed using the spectral metrics GFC and RMSE (explained in **Publication II**) and the color difference ΔE_{00} for the D65, A, and F11 illuminants.

Table 4.1 Spectral-reflectance-factor-estimation evaluation for the ranges 400–700 nm, 450–700 nm, and 450–950 nm, with 1nm sampling. The representative spectra from the imaged color chart samples was obtained by averaging a 5x5 pixel area from the middle of each patch. For additional noise removal temporal averaging was performed: in case of the FD system 20 images were averages; in case on ND800 10 images were averaged. The lowest error values are highlighted in yellow. Increases in errors with increased evaluation ranges are highlighted in white (mean / 95th percentile / maximum).

400-700nm	GFC	RMSE	ΔE_{00} D65	ΔE_{00} A	ΔE_{00} F11
LS	0.9984/0.9944/0.9900	0.0455/0.0934/0.1130	3.35/4.65/4.87	3.08/3.97/3.98	3.34/4.62/4.81
FD*	0.9964/0.9873/0.9770	0.0312/0.0617/0.0781	2.71/5.45/6.30	2.61/4.95/5.95	2.70/5.41/6.25
ND800**	0.9897/0.9654/0.9535	0.0427/0.0771/0.0807	1.98/3.80/4.21	1.96/3.05/3.60	1.95/3.68/4.13

450-700nm	GFC	RMSE	ΔE_{00} D65	ΔE_{00} A	ΔE_{00} F11
LS	0.9997/0.9873/0.9770	0.0408/0.0853/0.1035	2.93/3.69/3.87	2.85/3.58/3.63	2.92/3.68/3.84
LCTF	0.9966/0.9882/0.9850	0.0355/0.0884/0.1030	4.16/8.42/9.22	3.96/6.79/7.31	4.16/8.33/9.10
FD*	0.9964/0.9848/0.9717	0.0308/0.0620/0.0734	2.62/4.77/5.58	2.58/4.71/5.60	2.61/4.77/5.77
ND800**	0.9890/0.9578/0.9420	0.0442/0.0766/0.0872	2.16/4.82/5.55	2.43/3.89/4.12	2.16/4.60/5.33

450-950nm	GFC	RMSE
LS	0.9993/0.9988/0.9880	0.0423/0.0808/0.0941
LCTF	0.9939/0.9639/0.9603	0.0535/0.1489/0.2017
FD*	0.9899/0.9510/0.9454	0.0517/0.1347/0.2379

* 20image average; ** 10image average

The results suggest that FD could be suitable for a number applications in which the acquisition of colorimetric and spectral data in still-image form is required. In terms of spectral accuracy, the FD system can achieve the accuracy of the LCTF system, as shown via the applied estimation model (see Table 4.1, highlighted in yellow). It is interesting to note that the colorimetrical accuracy is highest for the ND800. One would assume that a higher number of acquisition channels would lead to better spectral and color accuracy. In the case of the FD system and when compared to ND800, however, this seems to be true for spectral accuracy only.

The effects of the illuminant use and the wavelength range are shown to affect the evaluation. When simulating an A light source, color errors are reduced than when simulation D65 and F11. In addition, the evaluation range seems to have an effect on accuracy. Extending the range by 50 nm in the VIS (evaluation of 400 – 700 nm) increases errors in the FD system. Similar behavior is shown for the LS, while for the ND800, this move reduces such errors. Extending the range to include NIR is shown here to increase spectral errors for the FD system, which is most likely to be due to

the sensitivity gap at 730 nm and the fact that there is only one very broad NIR channel. This matter is discussed in more detail in **Publication II** and in Section 4.2.1.

The experimental results in **Publication I** suggest that the FD system could be suitable for several applications in which the acquisition of colorimetric and spectral data in still-image form is necessary. Moreover, the fast imaging, portability, and spatial properties of the FD system render it an attractive, practical spectral imaging system. The device comparison showed that, even when the simplest estimation model is used, FD can reach up to the accuracy of an LCTF. However, these conclusions hold for standard color chart cases and when both spatial and temporal averaging can be performed to remove noise. At this point, however, the video aspects of the FD system still need to examined. In the next step, the extension of still-image characterization to video characterization is discussed.

4.2 VIDEO CHARACTERIZATION

As the ultimate goals are to use the FD system in practical dynamic applications and to acquire spectral videos for utilization in the creation of novel research, it is important to understand the behavior and stability of the system during video acquisition. **Publication II** thus focuses on analyzing the FD systems output as a function of time. For this paper, the author compiled and implemented a set of methods through which to gain a temporal understanding of four aspects of the spectral video system: noise, mutuality, spectral performance, and color performance. The state-of-the-art of characterization and evaluation in spectral color research is based on still images and consequently still objects. However, in spectral video the scene content might change from frame to frame. Therefore, there is a need to extend the characterization procedures to the temporal domain. The four aspects listed above were selected for the following reasons: in video it is important to know whether movement comes due to noise or motion; mutuality facilitates prediction of time series; spectral information is valuable when systems are used for reflectance-based analysis and color is of interest for applications concerning the human observer.

The main experiments performed were classified into two main groups: monochromatic analysis and spectral analysis (see Figure 1 in **Publication III**). Within each group, intra-frame (one frame at a time) and inter-frame (between different frames) analysis were undertaken for the following spectral video data: a 30x30cm Spectralon 99% diffuse white reference plate, CC24 and DSG96 color charts and a dark sample, which was obtained by closing the lens with its lens cap. In addition, spectral reflectance estimation was performed using empirical linear regression (Heikkinen et al., 2008) and evaluated as a function of time for 400-700 nm and 400–1000 nm range and DSG96 and CC24 color charts. The data-gathering setup is presented in **Publication II, Section III B** and the computational methods used are described in **Publication II, Section II**.

The characterization data collected facilitated expansion of the knowledge about the FD system collected in **Publication I**, as well as provided a deeper understanding of the complexities of spectral video recording. There are multiple elements that vary during video acquisition. Here two aspects of the characterization procedure will be discussed in more detail: spectral reflectance estimation as a function of time (discussed in **Section 4.2.1**) and noise in video (discussed in **Section 4.2.2**). The data provides an insight into the FD system's accuracy in estimating reflectance for a given illumination and estimation model combination. Also, video noise is more complex than image noise as it may come from multiple sources – illumination changes over time, sensor changes over time, motion, and post-processing (e.g., demosaicing).

4.2.1 Spectral reflectance estimation in spectral video

For **Publication II,** in contrast with the method used in **Publication I**, the estimation model was applied individually to every frame, without averaging. The spectral dimensions were increased from seven to 601 (see **Publication II Section D** for a description of the reflectance estimation model utilized).

When inter-frame RMSE and GFC values are computed both before estimation and after, we can see that the spectral fit between the frames is reduced. The inter-frame RMSE increases on average from 0.0063 to 0.0102, and GFC decreases from 0.9994 to 0.9989 (see Figure 6 a, b, e, and d in **Publication II**). The reason behind this might involve the estimation model itself and/or the amplification of temporal noise when the estimation is applied. As such, estimation models and noise-reduction algorithms have to be selected with caution, bearing in mind the needs and tolerances of the final application.

When evaluating the estimates in an intra-frame manner, with respect to spectrophotometer output (as in Section 4.1.3 a PE spectrophotometer groud truth was used), the spectral errors were found to be larger for spectral estimation in the 400–1000 nm wavelength range than in the 400–700 nm wavelength range (see Figure 6 c and d in **Publication II**). With this particular device, this finding is probably the result of the relatively broad wavelength band of the NIR channel and the sensitivity gap at 730 nm (see device sensitivities in Figure 4.1), which affect the performance of the estimation model.

The results given above can be applied to cases where the possibility of taking a spatial average to represent a noise-reduced spectrum exists. When there is no possibility of performing spatial averaging, however, the mean errors increase significantly. When pixel-wise errors are computed (see Figure 7 in **Publication II**), the average RMSE, even without estimation, becomes almost three times higher and the GFC decreases to 0.9893. This suggests that if the FD system was used in applications where pixel-wise evaluation is important, there would be a need for additional noise reduction.

Colorimetrical evaluation is also important, especially for applications concerning a human observer. Inter-frame color-difference values indicate the perception of neighboring frames of the video with relative similarity in terms of color appearance (see Figure 8b in **Publication II**, ΔE_{00D65}~ 2.04). We should note that in some samples (8 and 13 on Figure 8b in **Publication II**), the color coordinates seem to overlap, possibly suggesting that it would be difficult for this device to differentiate between those samples in certain frames of the video.

These tests provided a valuable insight into the accuracy and precision of FD when used in video mode. The level of accuracy might be affected by estimation models used, temporal noise and the evaluation range. The precision in repeating the measurements on the other hand was shown to be high (for example, standard deviation for 400-700 nm range and intra-frame evaluation over the frames: ΔE_{00}=0.29; $RMSE$=0.0005; GFC=0.0006). These results apply when there is a possibility to remove noise with spatial averaging. When pixel-wise computations are needed the effect of spatiotemporal noise are evident and need to be studied in more detail in future work.

4.2.2 Noise in spectral video output

The characterization data gathered for **Publication II** provides a deeper understanding of the complexities of the spectral-video recording task, especially from the noise perspective. During motion, there are many elements that might change, leading to data being corrupted by noise. Illumination conditions, sensors, and the object are all subject to change. Post-processing (i.e., demosaicing) can also add to the noise effect.

The correlation values given in **Publication II Table I** indicate the fluctuations that could appear due to the sensor or the illumination conditions, such as light-source flicker. Even though the SNR (see Figure 5 f in **Publication II**) indicates that values are improved through averaging, there is no possibility of averaging frames in order to remove noise at video rates. In addition, pixel-wise errors (see Figure 8, **Publication II**) suggest that some noise effects are not completely removed via the commonly used flat-field-correction methods. For these reasons, sensors and illumination conditions must be controlled closely during acquisition, or else algorithms must be applied for noise removal in post-processing. Also, it would be beneficial to further investigate the effects of illumination angles on sensor noise and if certain areas on the sensors exhibit specific noise and sensor sensitivity related behavior.

Interestingly, it was found that post-processing is also a possible source of noise. If the light source flickers or changes spatially in any way during the acquisition process, the mutuality of spatial information between frames is affected. With these light changes, the interpolation employed by demosaicing can produce different pixel values for different frames, thus affecting frame mutuality further. The question

that must be investigated in more detail, therefore, is whether the use of a sensor-wise demosaicing technique is a good choice for the FD system or whether better multi-channel demosaicing options exist.

4.3 SPEED – SPECTRAL EYE VIDEO DATABASE

The next step in this project is to describe the practical implementation of the FD spectral video system. The most significant task here was the creation of a new, publicly available combined spectral image/spectral video database – i.e., SPEED (the SPectral Eye vidEo Database). The database features 180 seven-channel videos, 60 seven-channel images, and 30 fifty-one-channel images of the left eyes of 30 voluntary human subjects. It was created with a number of dynamic computer-vision applications in mind, such as spectral-reflectance information-based surveillance, reflectance-based object detection, illumination-independent object tracking, spectral enhancement, and temporal spectral analysis. All of these applications have the potential to be relevant to many eye-related research areas – e.g., medicine, eye tracking, biometrics, and in eye/vision research.

In **Publication III**, the possibility of using spectral video to recording the eye while performing observational tasks was introduced to the eye-tracking research community. Unfavorable conditions, such as eyewear reflections, extreme angles, and make-up, can be considered problematic for eye tracking. Such issues have been included in the compilation of SPEED. As such, the database provides a platform for creating new methodologies for the imaging, training, analysis, and interpretation of eye-tracking data, especially where harsh conditions are present. The development of SPEED can be considered the first step towards the creation of new spectral-data-based enhanced eye-tracking devices, software, and algorithms. To the author's knowledge, the potential of spectral eye-tracking has not yet been investigated and the database compiled for this study is the first to serve such a purpose.

4.3.1 SPEED Database Description

In order to gather spectral data that would serve both spectral imaging and eye tracking, the standard spectral-imaging setup had to be modified. An NIR light-emitting diode (LED) (peak at 850nm), which functioned as a reference point that was reflected in the middle of the eye (glint), was added and also an observational task the subjects were performing during capture. This addition also provided additional benefits as it made it possible even to interpret the subject's gaze. One could, therefore, think of this experimental setup as constituting a remote spectral eye tracker. This unique setup is shown in **Publication III, Figure 4**. Figure 4.2 offers a detailed info-graph for SPEED that summarizes subject related basic statistics and shows generated RGB images as examples of conditions included in the database.

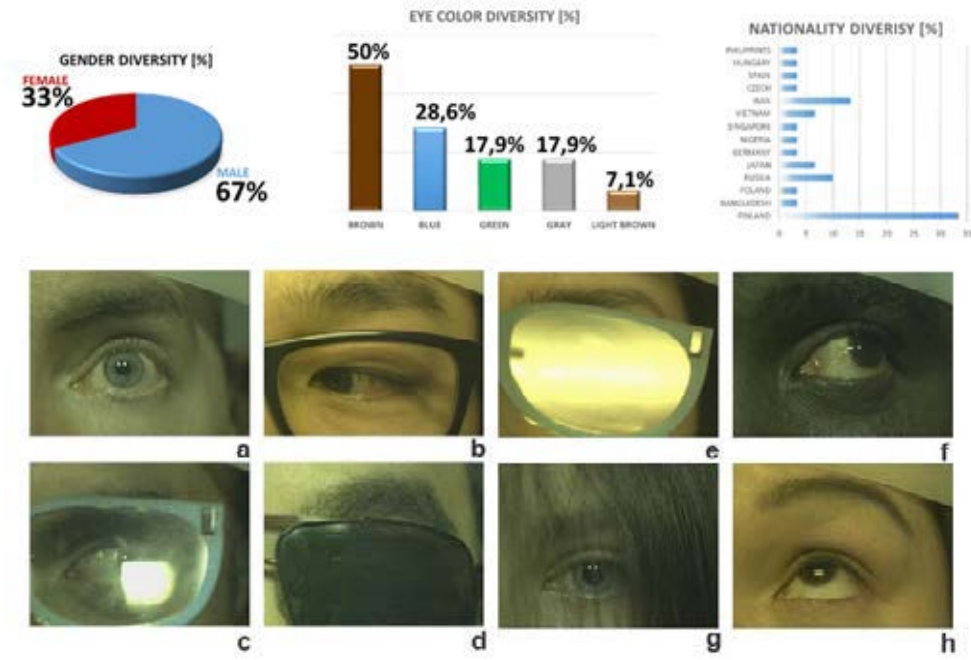

Figure 4.2. Info-graph for SPEED indicating the gender diversity, iris-color diversity, and nationality diversity found within the database. The example images demonstrate the different condition incorporated into the database: a) standard conditions, b) shadows, c) strong reflections from daylight, d) use of sunglasses, e) strong reflections from office light, f) extreme angles in daylight g) hair occlusion, h) extreme angles and make up in office light.

For SPEED Database collection two devices and two illumination conditions were used. The devices were LCTF and FD. As lighting condition two most common cases were simulated: sunny daylight with a 6500K color temperature D65 simulator and fluorescent office light simulator TL84.

The following data collection experiments were performed:

- fifty-one-channel spectral image collection (using LCTF): still spectral image collected to serve as spectral ground truth (see example Figure 3.8b and c)
- seven-channel spectral video collection (using FD): videos of eye movement recorded under two different light sources: standard illuminant simulator D65 (simulates sunny daylight of 6500K) and TL84 (simulates fluorescent office light) for the following cases:
    - standard – the most controlled case with a clear view of the eye (Figure 4.2a)
    - with eye glasses (own or glasses provided) – introduces shadows and reflections (Figure 4.2b)
    - with 3D-printed eye glasses – introduces extreme shadows and reflections (Figure 4.2c and d)

- seven-channel image collection (using FD): under D65 (Figure 4.2d) and TL84 while wearing sunglasses.

In total one hyperspectral image, 6 multispectral videos and 2 multispectral images per subject. All experiments were realized in the same setting and at a same occasion with duration of about 30min.

For all the video related experiments subjects were asked to perform an eye movement task. In this task subjects were asked to look at nine points marked by post-it notes and attached on fishing lines hanged from the ceiling. The location of points was known and fixed throughout the campaign. The main goal was to push the eye to its extreme positions is order to create data that could be used in eye tracking as well as an example of harsh conditions. At the same time this positions would reveal other features interesting for spectral tracking (blood vessles). We recorded instructions as a voice command and instructed subjects to perform the eye movement towards the number after hearing the number command. All subjects received the same instructions. In total one video is 5s long.

Additional details about the database can be found in **Publication III**, as follows: the acquisition setup used to gather the data is explained in **Section 2.2**, the devices utilized are discussed in **Section 2.1**, the methods applied are delineated in **Section 2.4,** and the observational task undertaken by the subjects is outlined in **Section 2.5**. Here the discussion about SPEED will be extended with more detailed analysis and examples of possible further exploration and use.

4.3.2 Spectral ground truth in SPEED database

To complement the low spectral resolution of the FD system, a 51-channel LCTF was used. The role of LCTF was to gather high-spectral-resolution and high-spatial-resolution ground-truth data for use in the analysis and identification of spectral signatures in the eye region, as well as in the evaluation of the accuracy of the FD video system. The data captured could also be used as training data for reflectance estimation algorithms. Figure 3.8c in Section 3.3 contains one example of the spectra sampled in an LCTF image of the main eye features discussed earlier (i.e., skin, the iris, the sclera, the pupil, blood vessels, and hair). Here, the spectral signatures discussed in Section 3.3 are recognizable and the different spectral shapes of the various eye features are evident. Those can be exploited to differentiate features and spectrally track them. To extend on the data reported in **Publication III**, here more detailed information about these features in the context of SPEED is provided in Figure 4.3.

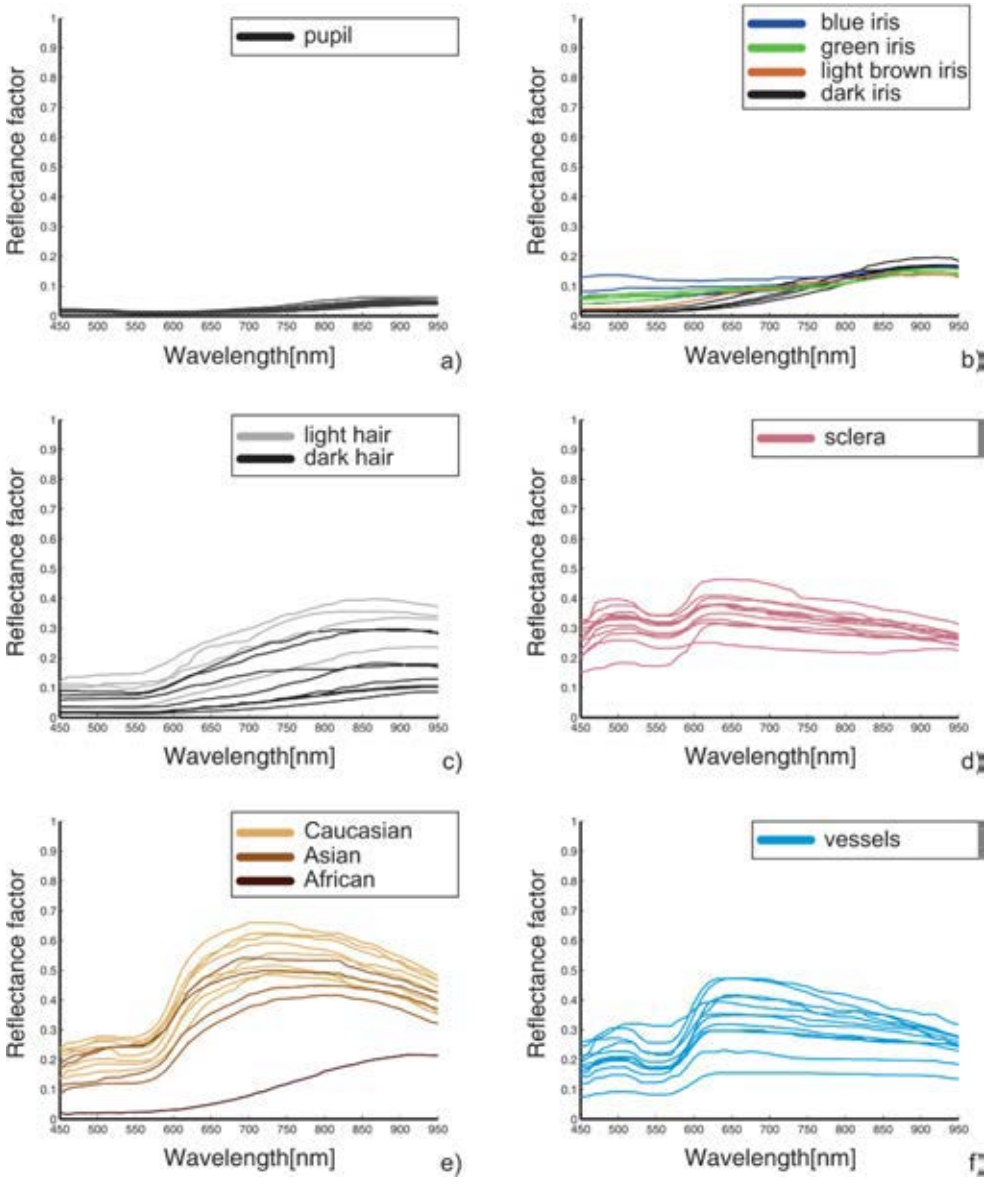

Figure 4.3 LCTF spectra in SPEED for a) the pupil, b) iris diversity, based on color; c) hair diversity, d) sclera, e) skin diversity within ethnicities, f) blood vessels.

Figure 4.3e demonstrates differences in skin-reflectance spectra in relation to ethnicity. Figure 4.3c indicates differences in hair type and Figure 4.3b showcase differences in eye color. In all these features, the effects of melanin can be noticed—especially for dark features. In the VIS part of the spectrum, melanin-rich, dark features have low reflectance, while in the NIR part of the spectrum, all of the reflectance figures increase. Where the iris is concerned, we can note that

discrimination of types is possible in the VIS part of the spectrum only; in the NIR part, the spectral shapes strongly overlap. There are other chromophores and elements that affect the whole spectral shape, however; in VIS, for instance, we can also detect the effects of hemoglobin in blood vessels (Figure 4.3f), the sclera (Figure 4.3d), and skin (Figure 4.3e).

A feature that is important for eye tracking specifically is the pupil, as many algorithms rely on pupil detection. Figure 4.3a indicates the high spectral shape similarities between the pupils of all the subjects. This feature is dark and has a reflectance close to zero in the VIS part of the spectrum. For some subjects, certain other dark features (e.g. a dark iris or dark hair) are shown to be highly similar in spectral shape to the pupil in the VIS part of the spectrum. In such cases, discrimination between these features becomes difficult in VIS, which explains the importance of using NIR cameras in eye tracking. The author was surprised to find that there is no standardized LED bandwidth within eye tracking; the LEDs used in commercial eye-tracking hardware differ from manufacturer to manufacturer. Aside from confirming the relevance of using NIR here, the LCTF data can also provide information, about which of the LED bands is optimal for tracking. For instance, by computing the wavelength-wise RMSE maximal, pupil-iris and pupil-hair differences can be determined. The location of the average maximum pupil-iris difference is 920 nm, and the maximum pupil-hair difference is at 840 nm. Additionally, pupil-iris difference reaches a level of 90% at 850 nm, while pupil-hair difference reaches 90% at 740 nm.

Finally, observing the NIR part of the spectrum also enables the effect of water content to be established. For instance, it was found that the sclera (Figure 4.3d) absorbs NIR radiation far more than skin as its water content is higher.

The knowledge gained about the features spectra stresses the potential for multiple-feature detection based on the spectral signatures of the features in question.

4.3.3 Examples of spectral eye videos in SPEED

The LCTF data demonstrates that the variations between the spectral signatures of different eye features (as discussed in Section 3.3) are recognizable within the eye region imaged here. In Section 4.2.1, we saw how the sensitivities of the FD system coincide with some of the main peaks in the eye features' spectral signatures. The next move, then, was to exploit these signatures to differentiate between features and track them spectrally over time. For this purpose 180 seven-channel spectral videos were collected. To challenge the FD system in the same ways in which an eye tracker is challenged, unfavorable conditions (such as eyewear reflections, extreme angles, shadows and make-up) were incorporated into the acquisition process. It was hypothesized that the use of spectral videos in eye-tracking would be beneficial as it would allow the spectral signatures of different eye features to be exploited for the

enhancement of eye-tracking devices, software, and algorithms. In this section, example spectral videos are provided in order to open up a discussion about the potential of spectral videos.

Figures 4.4 and 4.7 give examples of three frames picked at certain moments from the eye spectral vides, as well as present the seven-channel spectral reflectances for the six eye features within each frame. Figure 4.4 depicts a standard-condition video, while Figure 4.7 offers an example of a spectral video with harsh conditions. It should be noted that, for viewing purposes, the reflectances are plotted sharply; the seven responses from the FD system were connected very simply, using a straight line and no shape estimation.

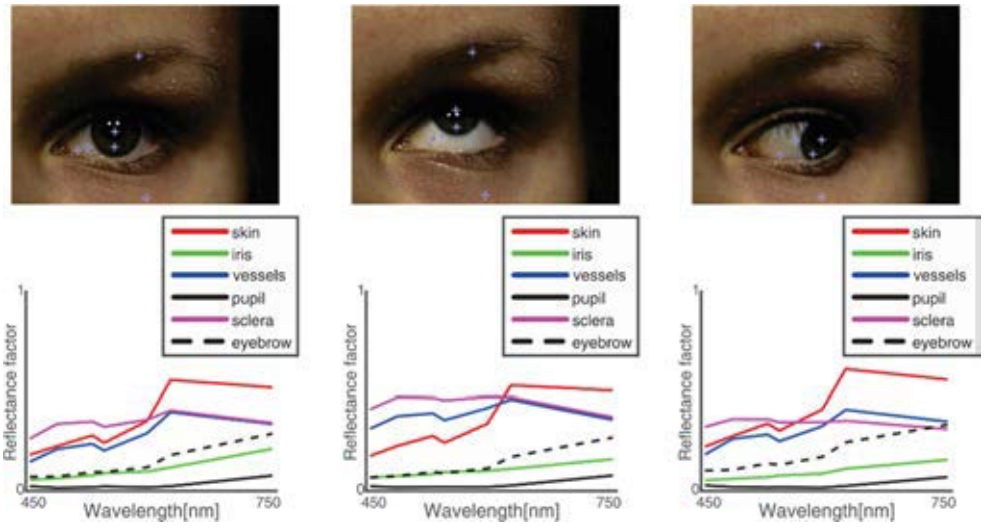

Figure 4.4 Example of standard video frames from SPEED and the spectra of six eye features (i.e., skin, the iris, the sclera, the pupil, blood vessels, and hair). Every spectrum is an average of a 10x10 pixel area marked on the RGB images. Note that these spectral shapes are rather superficial. These and the RGB images have been generated simply for viewing purposes.

Even with this very sparse spectral representation, it is possible to differentiate between eye features, based on the spectral signatures in each of the frames. The similar behavior of the features and the LCTF data can also be noted here. The pupil is the darkest feature. The effects of hemoglobin can be seen in the VIS region, while the effects of melanin and water are clear in both the VIS and NIR region. These findings offer inspiration for multiple future projects. One direction might be to analyze stability/change in eye features' spectral signatures over time and during observational tasks. Here, inter-frame RMSE and GFC could be used as measures of spectral similarity within a video. In the case shown in Figure 4.4, the lowest RMSE can be seen for the pupil, while the least stable element is the sclera.

Spectral data was also used for detecting and segmenting eye featrues. Many available tools are described in the literature for this purpose – e.g.., Principal Component Analysis (Rodarmel and Shan, 2002), Spectral Angle Mapping (SAM) (Rashmi et al., 2014), Derivatives (Tsai and Philpot, 2002), and Non-negative Matrix Factorization (Wang et al., 2013). Groundworks have been performed with First Derivative and SAM. Figure 4.5 indicates the possibility of undertaking pupil detection via First Derivative, while Figure 4.6 demonstrates the potential for eye-feature classification via SAM.

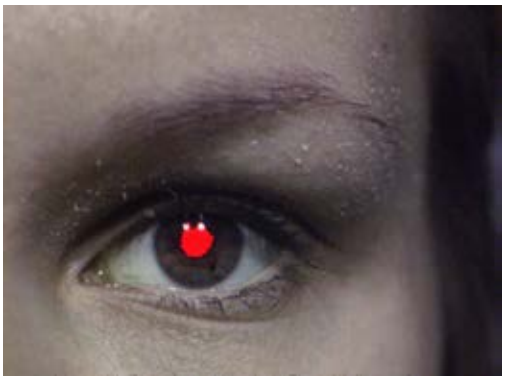

Figure 4.5 Pupil detection via 1st derivative. The algorithm for performing the detection consists of the following steps: 1) select representative spectra for each feature (10x10px area averaged spectra shown earlier), 2) calculate the 1st derivative for the representative spectra, 3) identify derivative values of importance for pupil and 4) compare with every pixel in the image to those values to classify them to pupil or no-pupil class

SAM is a method that compares spectra by computing the angles which describes their angular separation.

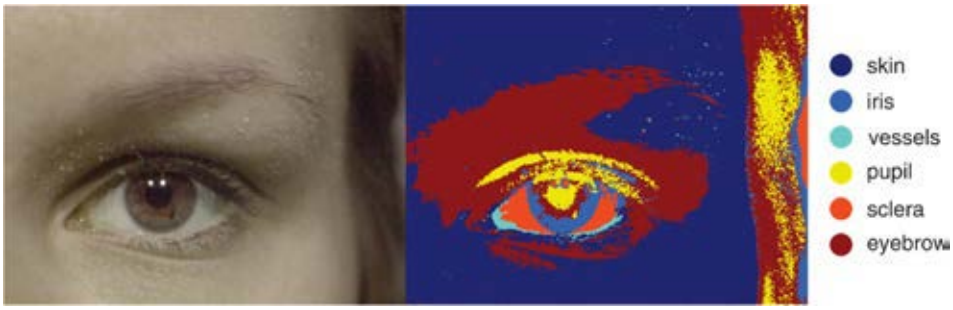

Figure 4.6 Eye-feature segmentation via SAM. The algorithm for performing the SAM segmentation consists of the following steps: 1) select representative spectra for each feature (10x10px area averaged spectra shown earlier), 2) calculate the angles between every pixel spectra and the representative spectra, 3) label the pixel spectra to a certain feature based on the smallest angle

The SAM algorithm has been shown to have good performance across homogenous regions (see skin, Figure 4.6); shadow regions do pose a problem, however. Thus, further analysis is needed to explore whether shadow detection could be performed prior to SAM, or whether a seventh class – the "shadow" class – could be incorporated into the classification. In future research, other methods for feature classification will be tested and their performance will be evaluated across all video frames.

Besides the standard conditions the contribution of the spectrum was examined for harsh conditions as well.

Figure 4.7 demonstrates how a specific harsh condition (i.e., the use of eyewear) affects the spectra of the eye features. Eyewear creates shadows, reduces the intensity of the light reaching the camera, and introduces reflection. When traditional pupil detection algorithm is applied to such an image the pupil detection fails (especially in the case enlarged on Figure 4.8). Therefore, the hypothesis is that spectral information can be exploited for facilitating the detection either by spectrally enhancing the pupil and/or allowing for spectrum-based feature detection and segmentation.

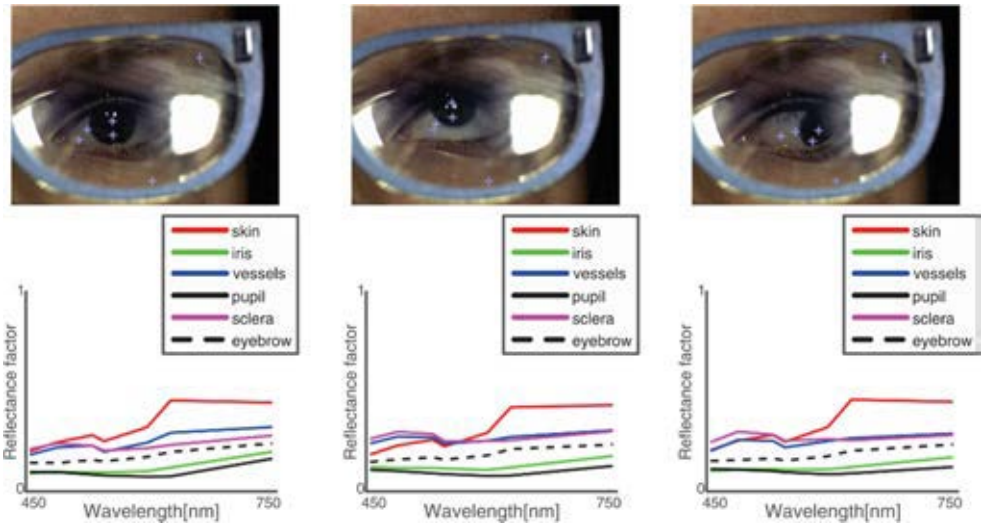

Figure 4.7 Example of harsh video from SPEED database. NOTE these spectral shapes are rather superficial for preview purpose, also RGB images are generated just for preview purpose

All of these aspects affect the final shape of the reflected spectra. Further research is needed to understand the degree of change. As a preliminary observation, when comparing the pupil spectra to the spectra in Figure 4.4, a rise in the levels can be seen, which is most likely to be due to reflections on the surface of the glasses. Despite

the harsh conditions in the figure above, the spectral differences between the six features remain, which indicates the potential for further exploitation of spectral data in eye tracking – especially under harsh conditions.

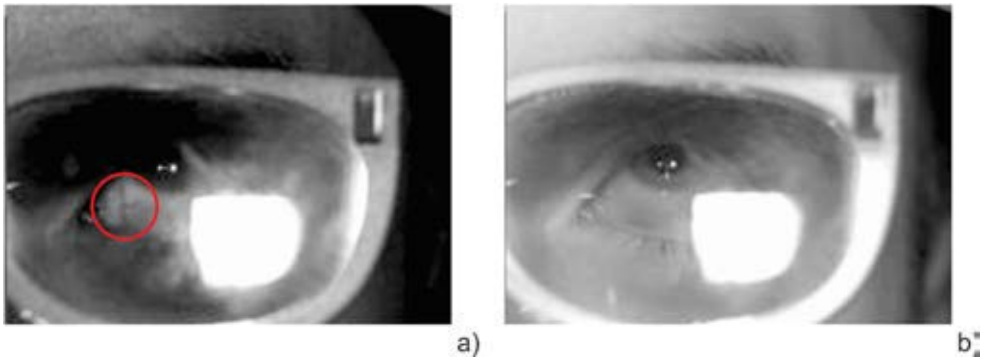

Figure 4.8 Potential of multi-channel capture for multi-feature tracking. a) Eye image of the channel G1 in an FD image. Highlighted in red is an additional feature (blood vessel) that becomes visible in this channel. b) Eye image of the NIR channel in an FD image. This channel can be considered to simulate well an image obtained traditional eye-tracking hardware. This type of image shows high contrast difference between pupil and iris however due to the difficult shape of pupil and strong reflections traditional pupil-detection algorithms fail here.

Looking at individual channels (Figure 4.8) another aspect can also be considered for future research into spectra in eye tracking. Traditionally, the NIR channel is used alone, leading to eye tracking failing under harsh conditions due to the occlusion of the pupil. With multi-channel tracking, such failures could be compensated by searching for other features that are invisible in NIR but are visible in other channels – for example, the blood vessels (Figure 4.8 B2 and G2).

The results presented in **Publication III** and examined in more detail here suggest the need for further exploration of spectral techniques in eye tracking. Spectral analysis, optimization, enhancement, annotation, and gaze interpretation are future areas of study to be researched on the basis of the hypothesis that knowledge about spectral signatures and their behavior can be used to help overcome the difficulties arising from unfavorable tracking conditions.

5 DISCUSSION

The main goals of this thesis were to characterize a novel spectral video system – the FluxData 1665-MS7 system – and apply it in practice. The characterization was achieved in two stages, using (1) single-image output and (2) video output. After the characterization stage, the system was applied in the development of a novel combined spectral image/spectral video database named SPEED (the SPectral Eye vidEo Database). SPEED consists of 30 fifty-one-channel spectral images, 60 seven-channel spectral images, and 180 seven-channel spectral videos. Harsh conditions such as eyewear reflections, extreme angles, and make-up were incorporated into the videos. SPEED was then introduced to the eye-tracking research community, and preliminary tests were performed for the detection, classification, and tracking of eye features during observational tasks. The database will continue to grow and is publicly available to all researchers at http://www.uef.fi/fi/web/spectral/speed. SPEED can be used non-commercially in teaching, for example, or for creating and testing new/improved methods of eye analysis via spectra.

For **Publications I** and **II**, a set of methods for characterizing a spectral video camera were compiled and implemented. As specifications provided by the manufactures can at time be found insufficient for system calibration procedures and differences in systems might as well appear from device to device in the production line, the methods compiled here can potentially serve as a recipe for researchers for profiling future spectral video systems. In this work those methods revealed that FD could be suitable for several applications that require acquisition of colorimetric and spectral data in a still image form. In video the spectral and color accuracy is challenged by noise which stresses the need for noise removal especially when used for a pixel-wise analysis. Once noise is removed, for example via spatial averaging, high precision for standards color charts can be noted. This encouraged further application of the examined system in practical application.

The data collected for these publications can be used in the calibration of the FD spectral video system and in its preparation for practical application. The data provides knowledge about spectral sensitivities, linearity, sensor misalignment, non-uniformity, noise, and color/spectral performance. Understanding the sensitivities related to the system helped us to identify the application range. More specifically, it is notable that the sensitivity bands for the FD system coincide with the main chromophores and elements present in the eye region. Compared to some of the state-of-the-art technologies the FD system trades a good spectral resolution for better spatial and temporal resolutions, but it is a very compact system that is easy to use, thus rendering it practical for many applications. In addition, a higher spectral resolution can be achieved via the use of reflectance estimation methods.

Other characterization procedures used here revealed the complexity of the spectral-video recording task, as multiple elements are subject to variance during acquisition. In video, noise is more complex than in an image as it may come from multiple sources: illumination changes over time, sensor changes over time, motion, and post-processing (e.g., demosaicing). Thus, illumination settings have to be selected with caution. Additionally, the fact that exposure times can be set sensor-wise, rather than channel-wise, leads to problems with exposure at the blue ends of the spectrum. It is also imperative that the stability of the light source is checked or compensated for if needed because light-source flickering can be captured by FD due to the system's high frame rate, thus corrupting the spectral data with noise. Demosaicing is one aspect of the system that must be researched further in future. The question remains of whether the use of standard sensor-wise Bayer demosaicing techniques is suitable, or whether multi-channel demosaicing techniques might be preferable. The strengths and weaknesses of both need to be established for this particular system. Color accuracy and spectral accuracy tests showed that the system could be suitable for several applications where the acquisition of colorimetric and spectral data is required. When used in video mode and for pixel-wise measurements, the level of accuracy is affected by noise. In such cases, additional noise removal is needed.

The practical application of the FD system resulted in a creation of the SPEED database. Compared to other state-of-the-art spectral video databases, SPEED contains video data of a scientifically relevant object that has been proven to interest many fields of science and industry, the human eye. Compared to other state-of-the-art eye databases, SPEED extends the available data in either spectral or temporal domain. Therefore, the database provides a platform for creating novel spectral-data-based methodologies to be used in the imaging, training, analysis, and interpretation of spectral video. Moreover, it provides valuable data for use in all eye-research-related fields when creating new spectral-data-based enhanced devices, software, and algorithms. In **Publication III**, the potential of exploiting spectral video to record the eye is introduced. Here, using a unique recording setup that could be thought of as a remote spectral eye tracker, 180 seven-channel spectral videos were recorded, some of which incorporate the harsh conditions that eye trackers face today. In addition, ground truth data for a 51-channel spectral image collection has been gathered and analyzed

The existence of the database provides the impetus for several possible areas of future work. Novel dynamic computer-vision applications such as spectral-reflectance information-based monitoring, reflectance-based object detection, illumination-independent object tracking, spectral enhancement, and temporal spectral analysis have the potential to be applied in many eye-related research areas, including medicine, eye tracking, biometrics, and eye/vision research. The groundworks reported in this thesis tests possible uses for SPEED in relation to enhancing eye-tracking hardware and software, and processing algorithms. Based

on the data recorded via both the LCTF and FD systems, we hypothesize that there would be benefits to exploring the use of multiple wavelengths further for the purposes of recording, analyzing, and tracking the motion of the eye. The examples of spectrum-based pupil detection, feature classification, and segmentation presented in this thesis also suggest possibilities for future applications.

The main challenges identified in this study were the complexity of the spectral-video recording task and the amount of data required. For instance, changing illumination conditions introduce noise into the system, and very-fast object movement continues to challenge spectral-acquisition technology. In the case of the eye, some blurry frames are captured during motion, which indicates the need for higher frame rates to be achieved. One must also make trade-offs between high quality and data size. One seven-channel spectral image is about 20 MB in size. In SPEED, one video contains about 80 frames, giving a total of approximately 14,400 seven-channel spectral images. Each video is about five seconds long, resulting in a size of about 1.6 GB on average. For each subject, there is about 9.8 GB of data, which gives a total of about 293 GB for the whole database. We anticipate that spectral technology and large data management will evolve to meet these challenges.

Looking at the pros and cons of the spectral video technology used in this thesis and considering how they affect the video-recording task, there are a few challenges that must be overcome in order to use spectral video widely to track the eye. At the moment, given the size of the data and the device used here, the system can be used only as an offline solution. However, there is also a persistent tendency to make spectral devices smaller and capture times faster. To combat this issue, one might reduce the number of channels by using the SPEED data to identify the most significant ones only. The frame rate of the sensor utilized here was just sufficient for acquiring the data under controlled laboratory conditions. For more natural observational tasks, a higher frame rate is needed.

Perhaps the final challenge is to convince the eye-tracking community to explore spectra even more. This is a serious task, given that price, speed, size and practicality are crucial factors that need to be addressed in order for widespread use to be achieved. First steps towards providing evidence that there is benefits from spectral enhancement in pupil detection have already been made. At the moment there is ongoing research that has shown improvements in pupil center detection for both standard conditions and harsh conditions. These works involve application of spectral image enhancement methods on the seven-channel spectral video frames and comparison of the performance of two pupil detection algorithms (ExCuSe and SET) on both original and enhanced frames.

The future works can continue in multiple directions, both with the characterization procedures and SPEED. For characterization there is a need to test multiple noise removal algorithms and reflectance estimation algorithms. Demosaic is one aspect of the system that will be investigated in more detail. Ultimately, the

goal is to be able to use FD to capture full frame information and perform pixel-wise analysis.

Further analysis of the data would begin with analyzing the aforementioned spectral signatures as functions of time. Such an analysis should provide information about the stability or otherwise of spectral signatures during motion. Next, the classification and segmentation of features as functions of time should be conducted. Multiple methods must be tested for this purpose, as their performance might vary from feature to feature. Spectral-enhancement methods might also be applied in order to identify the optimal bands for, e.g., enhancing the contrast between different eye features. Finally, comparisons between existing pupil-detection algorithms for different spectral-band combinations should be made so that their weak points can be identified and enhancements can be proposed. Further research is also needed for deeper understanding the degree of change the harsh conditions introduce to the spectrum of the eye features. The aim would be to be able to provide a model of those changes. One interesting point for consideration is that one seeks to obtain the most accurate spectra possible in spectral imaging. In eye tracking, however, such an aim might not necessarily be of primary importance. This realization opens up the possibility of thinking about feature detection rather as an image processing task than spectral analisys task, in which additional spectral information can be used to enhance features of interest.

The work conducted for this thesis can be considered the first step necessary in the quest to create new eye-tracking technologies in the future. The characterization procedures highlighted device features one would need to keep in mind when practically applying such system and the application provided real world challenges for it. The results encourage further exploration and collaboration of spectral video acquisition and eye analysis and tracking. Aside from its use in eye tracking, SPEED also has the potential to be utilized in other eye-related research areas, such as medicine, biometrics, and eye and vision research.

BIBLIOGRAPHY

Anderson R.R., Parrish, J.A. 1981. The optics of human skin. *Journal of investigative dermatology*, 77(1).13-19.

Angelopoulo E., Molana R., Daniilidis, K. 2001. Multispectral skin color modeling. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference. Vol. 2, IEEE. II-II

Arnold T., De Biasio M., Leitner, R. 2013. Hyper-spectral video endoscopy system for intra-surgery tissue classification. *In Sensing Technology (ICST), 2013 Seventh International Conference. IEEE*. 145-150

Bartczak P., Gebejes A., Falt P., Parkkinen J., Silfstein, P. 2015.. LED-based spectrally tunable light source for camera characterization. *In Colour and Visual Computing Symposium (CVCS), IEEE*. 1-5.

Basler 2017a. http://www.baslerweb.com/en/products/cameras/area-scan-cameras/scout/sca1400-30gc - visited 6.3.2017

Basler 2017b. https://www.bnl.gov/atf/docs/scout-g_users_manual.pdf - visited 6.3.2017

Bodkin A., Sheinis A., Norton A., Daly J., Beaven S., Weinheimer, J. 2009, May. Snapshot hyperspectral imaging: the hyperpixel array camera. *In SPIE Defense, Security, and Sensing. International Society for Optics and Photonics*.

Boettner E.A., Wolter J.R., 1962. Transmission of the ocular media. *Investigative Ophthalmology & Visual Science*, 1(6). 776-783.

Bowen I.S., 1938. The Image-Slicer a Device for Reducing Loss of Light at Slit of Stellar Spectrograph. *The Astrophysical Journal*, 88, p.113.

Boyce C., Ross A., Monaco M., Hornak L., Li X. 2006, June. Multispectral iris analysis: A preliminary study51. *In Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference. IEEE.* 51-51.

Cao X., Du H., Tong X., Dai Q., Lin S., 2011b. A prism-mask system for multispectral video acquisition. *IEEE transactions on pattern analysis and machine intelligence*, 33(12), 2423-2435.

Cao X., Tong X., Dai Q., Lin S. 2011a. High resolution multispectral video capture with a hybrid camera system. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference. IEEE*. 297-304

Cesarini J.P. 1990. Hair melanin and hair color. Chapter 8 in Hair and hair diseases. *Springer Berlin Heidelberg*. 165-197

Crihalmeanu S., Ross A. 2012. Multispectral scleral patterns for ocular biometric recognition. *Pattern Recognition Letters*, 33(14). 1860-1869.

Descour M., Dereniak, E., 1995. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 34(22), 4817-4826.

Du H., Tong X., Cao X., Lin, S. 2009. A prism-based system for multispectral video acquisition. *In Computer Vision, 2009 IEEE 12th International Conference. IEEE*. 175-182

Edelhauser H.F., Ubels, J.L., 2003. The cornea and the sclera. Adler's Physiology of the Eye, Clinical Applications, edited by PL Kaufman and A. Alm, 10th ed., *Mosby, St. Louis, MO.*

Everitt J.H., Escobar D.E., Villarreal R., Noriega J.R., Davis, M.R., 1991. Airborne video systems for agricultural assessment. *Remote Sensing of Environment*, 35(2-3), 231-242.

Feng J., Fang X., Cao X., Ma C., Dai Q., Zhu H., Wang Y., 2014. Advanced hyperspectral video imaging system using Amici prism. *Optics express*, 22(16), 19348-19356.

FluxData Inc. 2011 http://www.fluxdata.com/news/fluxdata%E2%80%99s-multispectral-camera-heading-international-space-station – visited 9.3.2017

Ford B.K., Volin C.E., Rouse A.R., Sabharwal Y.S., Lynch R.M., Gmitro A.F., Bearman G.H., Descour, M.R. 1999, April. Video-rate spectral imaging system for fluorescence microscopy. *In BiOS'99 International Biomedical Optics Symposium. International Society for Optics and Photonics*. 216-226.

Fält P., Hiltunen J., Hauta-Kasari M., Sorri I., Kalesnykiene V., Uusitalo H., 2009. June. Extending diabetic retinopathy imaging from color to spectra. *In Scandinavian Conference on Image Analysis. Springer Berlin Heidelberg.* 149-158

Hagen N., Dereniak E.L. 2008. Analysis of computed tomographic imaging spectrometers. I. Spatial and spectral resolution. *Applied Optics*, 47(28), pp.F85-F95.

Hagen N., Kudenov M.W., 2013. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9). 090901-090901.

Haneishi H., Hasegawa T., Hosoi A., Yokoyam, Y., Tsumura N., Miyake, Y., 2000. System design for accurately estimating the spectral reflectance of art paintings. *Applied Optics*, 39(35), 6621-6632.

Hansen D.W., Ji Q. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence*, 32(3). 478-500.

Hardeberg J.Y., 2001. Acquisition and reproduction of color images: colorimetric and multispectral approaches. *Universal-Publishers*.

Hardeberg J.Y., Schmitt F., Brettel H., 2002. Multispectral color image capture using a liquid crystal tunable filter. *Optical engineering*, 41(10), 2532-2548.

Hasnat M., Parkkinen J. and Hauta-Kasari M., 2016. Spectral video construction from RGB video: Application to Image Guided Neurosurgery. *arXiv preprint arXiv:1612.04809*.

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

Heikkinen V., Mirhashemi A., Alho J, 2013. Link functions and Matérn kernel in the estimation of reflectance spectra from RGB responses. *J. Opt. Soc. Am. A* 30(11). 2444–2454

Hirvonen T., Orava J., Penttinen N., Luostarinen K., Hauta-Kasari M., Sorjonen M., Peiponen K.E., 2014. Spectral image database for observing the quality of Nordic sawn timbers. *Wood science and technology*, 48(5), 995-1003.

Hohmann M., Douplik A., Varadhachari J., Nasution A., Mudter J., Neurath M., Schmidt, M., 2011. Preliminary results for hyperspectral videoendoscopy diagnostics on the phantoms of normal and abnormal tissues: towards gastrointestinal diagnostics. *In European Conference on Biomedical Optics. Optical Society of America*. 80872N

Hohmann M., Kanawade R., Klämpfl F., Douplik A., Mudter J., Neurath M.F., Albrecht, H. 2016. In-vivo multispectral video endoscopy towards in-vivo hyperspectral video endoscopy. *Journal of biophotonics*.

Hosseini M.S., Araabi B.N., Soltanian-Zadeh, H. 2010. Pigment melanin: Pattern for iris recognition. *IEEE Transactions on Instrumentation and Measurement*, 59(4).792-804.

Iwama R., Mitsui M., Yamaguchi M., Haneisi H., Ohyama N. 2005. Real-time multispectral and multiprimary video system. *Proc. AIC Colour, 5*.

Jóźwik A., Siedlecki D., Zając M., 2014. Analysis of Purkinje images as an effective method for estimation of intraocular lens implant location in the eyeball. *Optik-International Journal for Light and Electron Optics*, 125(20). 6021-6025.

Kaiser P.K., Boynton R.M., 1996. Human color vision.

Kapany N. S. 2004. Fiber optics. *Concepts of Classical Optics, J. Strong, Ed.*, 553–579.

Kaplan M., 2015. The Secrets in Their Eyes: Transforming the Lives of People with Cognitive, Emotional, Learning, Or Movement Disorders Or Autism by Changing the Visual Software of the Brain. *Jessica Kingsley Publishers.* Chapter 1.

King D., Vlcek, J. 1990. Development of a multispectral video system and its application in forestry. *Canadian Journal of Remote Sensing*, 16(1), 15-22.

Koponen P., Kalviainen H., Parkkinen, J., 2000. Multispectral video. *In Pattern Recognition, 2000. Proceedings. 15th International Conference. IEEE.* (Vol. 3). 186-189.

Land M.F., Hayhoe M., 2001. In what ways do eye movements contribute to everyday activities? *Vision research*, 41(25). 3559-3565.

Lee D., Haynes R., Ren D., Allington-Smith, J. 2001. Characterization of lenslet arrays for astronomical spectroscopy. *Publications of the Astronomical Society of the Pacific*, 113(789), p.1406.

Leitner R., De Biasio M., Arnold T., Dinh C. V., Loog M., Duin, R. P. 2013. Multi-spectral video endoscopy system for the detection of cancerous tissue. *Pattern Recognition Letters*, 34(1), 85-93.

Liou H.L., Brennan N.A., 1997. Anatomically accurate, finite model eye for optical modeling. *JOSA A*, 14(8). 1684-1695.

Losson O., Macaire L., Yang, Y., 2010. Comparison of color demosaicing methods. *Advances in Imaging and Electron Physics*, 162, 173-265.

Lund-Andersen H., Sander B. 2003 "The vitreous," in Adler's physiology of the eye – 10th ed., P. L. Kaufman and A. Alm, eds. *Mosby, Inc*. 293–316.

Ma C., Cao X., Tong X., Dai Q., Lin, S. 2014. Acquisition of high spatial and spectral resolution video with a hybrid camera system. *International journal of computer vision*, 110(2), 141-155.

McDuff D., Gontarek S., Picard, R.W., 2014. Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering*, 61(10), pp.2593-2601.

Medina J.M., Pereira L.M., Correia H.T., Nascimento S.M., 2011. Hyperspectral optical imaging of human iris in vivo: characteristics of reflectance spectra. *Journal of biomedical optics*, 16(7).076001-076001.

Monno Y., Kiku D., Kikuchi S., Tanaka M., Okutomi M., 2014, October. Multispectral demosaicking with novel guide image generation and residual interpolation. *In Image Processing (ICIP), 2014 IEEE International Conference on. IEEE.* 645-649

Morovič P., Haneishi, H. 2006. Estimating reflectances from multi-spectral video responses. *In Color and Imaging Conference Vol. 2006, No. 1. Society for Imaging Science and Technology*. 131-137

Morris H.R., Hoyt C.C., Treado, P.J., 1994. Imaging spectrometers for fluorescence and Raman microscopy: acousto-optic and liquid crystal tunable filters. *Applied spectroscopy*, 48(7), 857-866.

Ngo, H.T., Ives, R.W., Matey, J.R., Dormo, J., Rhoads, M. and Choi, D., 2009, November. Design and implementation of a multispectral iris capture system. In Signals, Systems

and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on (pp. 380-384). IEEE.

Ohta N., Robertson A., 2006. Colorimetry: fundamentals and applications. *John Wiley & Sons*.

Park J.I., Lee M.H., Grossberg M.D., Nayar, S.K. 2007. Multispectral imaging using multiplexed illumination. *In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE*.1-8

Phinn S.R., Stow D.A., Zedler J.B. 1996. Monitoring wetland habitat restoration in southern California using airborne multi spectral video data. *Restoration Ecology*, 4(4), 412-422.

Photo Research, Inc. 1999. http://www.cs.unc.edu/Research/stc/FAQs/spectroradiometer/705spec.pdf - visited 10.3.2017

Proenca H., Filipe S., Santos R., Oliveira J., Alexandre L.A., 2010. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), p.1529.

Rashmi S., Swapna A., Venkat S., Ravikiran S. 2014. Spectral Angle Mapper Algorithm for Remote Sensing Image Classification. IJISET–International Journal of Innovative Science, Engineering & Technology, 50(4). 201-205.

Rodarmel C., Shan, J. 2002. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science*, 62(2). 115.

Sellar R.G., Boreman G.D., 2005. Comparison of relative signal-to-noise ratios of different classes of imaging spectrometer. *Applied Optics*, 44(9). 1614-1624.

Sharma G., eds., 2003. Digital color imaging handbook. *CRC press*.

Shinoda K., Ogawa S., Yanagi Y., Hasegawa M., Kato S., Ishikawa,M., Komagata H., Kobayashi, N., 2015. Multispectral filter array and demosaicking for pathological images. *In Signal and Information Processing Association Annual Summit and Conference (APSIPA), IEEE*. 697-703.

Specim, Spectral Imaging Ltd. http://www.specim.fi/downloads/PFD_Spectral_Camera-v1-16.pdf - visited 10.3.2017

Stoffels J., Bluekens A.A.J., Petrus, J.M.P. 1978. Color splitting prism assembly. US Philips Corporation, U.S. Patent 4,084,180.

Thody A.J., Higgins E.M., Wakamatsu K., Ito S., Burchill S.A., Marks J.M. 1991. Pheomelanin as well as eumelanin is present in human epidermis. *Journal of Investigative Dermatology*, 97(2).340-344.

Tonsen M., Zhang X., Sugano Y., Bulling A., 2016, March. Labelled pupils in the wild: a dataset for studying pupil detection in unconstrained environments. *In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. ACM*. 139-142.

Tsai F., Philpot W.D. 2002. A derivative-aided hyperspectral image analysis system for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 40(2), 416-425.

Uhlhorn S.R., Borja D., Manns F. and Parel J.M. 2008. Refractive index measurement of the isolated crystalline lens using optical coherence tomography. *Vision research*, 48(27). 2732-2738.

Vanderriest C. 1980 A fiber-optics dissector for spectroscopy of nebulosities arounds quasars and similar objects. *Publ. Astron. Soc. Pac*. 92(550), 858–862

Vecino E., Sharma S.C., 2011. Glaucoma: Basic and Clinical Concepts.

Vilaseca M., Mercadal R., Pujol J., Arjona M., de Lasarte M., Huertas R., Melgosa M., Imai F.H., 2008. Characterization of the human iris spectral reflectance with a multispectral imaging system. *Applied optics*, 47(30).5622-5630.

Wagadarikar A. A., Pitsianis N. P., Sun X., Brady D. J. 2009. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics express*, 17(8), 6368-6388.

Wang N., Du B., Zhang, L., 2013. An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2). 554-569.

Wood E., Baltrušaitis T., Morency L.P., Robinson P., Bulling, A. 2016. *Learning an appearance-based gaze estimator from one million synthesised images. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. ACM*. 131-138.

Wyszecki G., Stiles W.S., 1982. Color science (Vol. 8). *New York: Wiley*.

Yamaguchi M., Haneishi H., Fukuda H., Kishimoto J., Kanazawa H., Tsuchida M., Iwama R., Ohyama N. 2006a. High-fidelity video and still-image communication based on spectral information: Natural vision system and its applications. *In Electronic Imaging 2006 International Society for Optics and Photonics*. (pp. 60620G-60620G).

Yamaguchi M., Iwama R., Kanazawa H., Fujikawa N., Fukuda H., Haneishi H., Ohyama N., Wada H., Kambara T., Aihara M.. Yamakawa, Y. 2006b. Color reproducibility of skin lesions in multispectral video: Experimental evaluation. *In Color and Imaging Conference (Vol. 2006, No. 1). Society for Imaging Science and Technology*. 8-13

Zhang D., Guo Z., Gong, Y. 2015. Multispectral biometrics: systems and applications. *Springer*.

Zonios G., Bykowski J., Kollias, N. 2001. Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo using diffuse reflectance spectroscopy. *Journal of Investigative Dermatology*, 117(6). 1452-1457.

ANA GEBEJES

This thesis explores the potential of a novel spectral acquisition technique - spectral video acquisition. Protocols for characterization were developed to serve as future guidelines on spectral video systems evaluation in this emerging field. In addition, to boost research on spectral-data-based enhanced eye-tracking, a spectral video system was applied to capture a first of its kind combined spectral video/ spectral image database of the human eye: the SPectral Eye vidEo Database (SPEED).

UNIVERSITY OF
EASTERN FINLAND

uef.fi

PUBLICATIONS OF
THE UNIVERSITY OF EASTERN FINLAND
Dissertations in Forestry and Natural Sciences

ISBN 978-952-61-2550-3
ISSN 1798-5668