JYVÄSKYLÄ STUDIES IN COMPUTING
184

Ilkka Pölönen

Discovering Knowledge in Various Applications with a Novel Hyperspectral Imager

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN COMPUTING 184

Ilkka Pölönen

Discovering Knowledge in Various Applications with a Novel Hyperspectral Imager

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksen auditoriossa 3
joulukuun 14. päivänä 2013 kello 10.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in building Agora, Auditorium 3, on December 14, 2013 at 10 o'clock.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2013

Discovering Knowledge in Various Applications with a Novel Hyperspectral Imager

JYVÄSKYLÄ STUDIES IN COMPUTING 184

Ilkka Pölönen

Discovering Knowledge in Various Applications with a Novel Hyperspectral Imager

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2013

Editors
Timo Männikkö
Department of Mathematical Information Technology, University of Jyväskylä
Pekka Olsbo, Ville Korkiakangas
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-5538-0
ISBN 978-951-39-5538-0 (PDF)

ISBN 978-951-39-5537-3 (nid.)
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä 2013

ABSTRACT

Pölönen, Ilkka
Discovering knowledge in various applications with a novel hyperspectral imager
Jyväskylä: University of Jyväskylä, 2013, 48 p.(+included articles)
(Jyväskylä Studies in Computing
ISSN 1456-5390; **184**)
ISBN 978-951-39-5537-3 (nid.)
ISBN 978-951-39-5538-0 (PDF)
Finnish summary
Diss.

Hyperspectral imaging is a rapidly growing field of science that has a multidisciplinary nature. This thesis introduces the knowledge discovery process for a novel hyperspectral imager based on the Fabry-Perot interferometer (FPI-HSI). The study explains step by step how hyperspectral imaging data can be preprocessed and analyzed for various applications. The research includes various areas, ranging from unmanned aerial vehicle base remote sensing to medical applications and crime scene investigation. Hyperspectral imagers based on the Fabry-Perot interferometer will become more common in the near future. The mathematical tools and computational algorithms behind the preprocessing and data analysis are tailored to process FPI-HSI data. This study implies that these tools are applicable for the process chain of FPI-HSI data.

Keywords: Hyperspectral imaging, knowledge discovery, data analysis, Fabry-Perot interferometer

Author

Ilkka Pölönen
Department of Mathematical Information Technology
University of Jyväskylä
Finland

Supervisors

Professor Pekka Neittaanmäki
Department of Mathematical Information Technology
University of Jyväskylä
Finland

Professor Amir Averbuch
School of Computer Science
Tel Aviv University
Israel

Reviewers

Professor Yehuda Roditty
School of Computer Sciences
The Academic College of Tel-Aviv-Yaffo
Israel

Professor Maria Skopina
Faculty of Applied Mathematics and Control Processes
Saint-Petersburg State University
Russia

Opponent

Professor Jarmo Alander
Department of Electrical Engineering and Automation
University of Vaasa
Finland

ACKNOWLEDGEMENTS

I would like to express my gratitude my supervisors, Professors Pekka Neittaanmäki and Amir Averbuch; without their advice and support, this process would have been much harder. I want to thank my reviewer professors, Yehuda Roditty and Maria Skopina. All of their comments made this work much better and more understandable. I am grateful to professor Valery Zheludev who has advised me many times.

In addition I would like to thank my fellow authors and other participants in this multidiscliplinary project: Heikki Saari, Eija Honkavaara, Hannu-Heikki Puupponen, Jere Kaivosoja, Heikki Salo, Liisa Pesonen, Teemu Hakala, Jussi Mäkynen, Christer Holmlund, Paula Litkey, Noora Neittaanmäki-Perttu, Mari Grönroos, Taneli Tani, Annamari Ranki, Olli Saksela, Erna Snellman, Jaana Kuula, Tuomas Selander, Tapani Reinikainen, Tapani Kalenius, Paavo Nieminen, and Tuomo Sipola.

I want to express my gratitude to the following colleagues for their fruitful conversations: Tero Tuovinen, Auri Kaihlavirta, Erkki Kurkinen, Tytti Saksa, Moshe Salhov, Ville Tirronen, Antti-Jussi Lakanen, Tapani Ristaniemi, Tuomo Rossi, Tiina Parviainen, Sakari Tuominen, Ismo Pellikka, Osmo Schroderus, Onni Kyrö, Timo Männikkö, and many others.

My relatives have encouraged me during this journey. My parents have given me the strength to carry on. In particular I would like to thank Saku and Lea for being a role model and guiding me in academic life. I would also like to thank my wife, Kreeta. She has supported and cheered me up whenever needed it. Finally, I thank my three lovely children Martta, Sohvi, and Väinö.

LIST OF ABBREVIATIONS

F	Matrix of filter vectors
J	Unit matrix
λ_i	Wavelength of certain waveband i
$\mathbf{m}$	Abundance vector for endmembers
m_i	Abundance of certain endmember
M	Endmembers abundances for each detected spectra
$\mathbf{\Psi}$	Non-linear function
R_{NDVI}	Normalized Difference Vegetation Index
R_{MCARI}	Modified Chlorophyll Absorption Ratio Index
$\mathbf{s}$	Pure spectrum also known as endmember
s_i	Endmembers value for waveband i
S	Matrix of endmembers
$\mathbf{x}$	Observed spectrum
x_i	Observed spectrum's value for waveband i
X	All observed spectra
x_i^f	Feature vector i
X^f	Matrix of feature vectors
W	Gaussian kernel matrix
w	Data noise
BRDF	Bidirectional Reflectance Distribution Function
CIR	Color-infrared
CMOS	Complementary Metal Oxide Semiconductor
DN	Digital number
DSM	Digital surface model
FPI	The Fabry-Perot interferometer
FPI-HSI	The hyperspectral imager based on FPI
GPS	Global Positioning System
KDD	The knowledge discovery process from databases
KNN	K-nearest neighbour algorithm
MCARI	Modified Chlorophyll Absorption Ratio Index
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
PCA	Principal component analysis
SVD	Singular value decomposition
SWIR	Short wave infrared
VNIR	Visible and near infrared
XVGA	Extended Video Graphics Array
UAV	Unmanned aerial vehicle
UAS	Unmanned aerial system

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF INCLUDED ARTICLES

PI **Ilkka Pölönen**, Heikki Salo, Heikki Saari, Jere Kaivosoja, Liisa Pesonen and Eija Honkavaara. Hyperspectral imaging based biomass and nitrogen content estimations from light-weight UAV. *Proceedings of SPIE Vol. 8887, Remote Sensing for Agriculture, Ecosystems, and Hydrology XV, 88870J (October 16, 2013); doi:10.1117/12.2028624.*, (2013).

PII **Ilkka Pölönen**, Heikki Salo, Heikki Saari, Jere Kaivosoja, Liisa Pesonen and Eija Honkavaara. Biomass estimator for NIR-Image with few additional spectral band images taken from light UAV. *Proceedings of SPIE Vol. 8369, Sensing for Agriculture and Food Quality and Safety IV, 836905 (May 1, 2012); doi:10.1117/12.918551.*, (2012).

PIII Eija Honkavaara, Heikki Saari, Jere Kaivosoja, **Ilkka Pölönen**, Teemu Hakala, Paula Litkey, Jussi Mäkynen and Liisa Pesonen. Processing and assesment of spectrometric, stereoscopic, imagery collected by a light weight UAV spectral camera for precision agriculture. *Remote Sensing, Vol. 5, No. 10, p. 5006-5039, doi:10.3390/rs5105006*, (2013).

PIV Noora Neittaanmäki-Perttu, Mari Grönroos, Taneli Tani, **Ilkka Pölönen**, Annamari Ranki, Olli Saksela and Erna Snellman. Detecting field cancerization using hyperspectral imaging system. *Lasers in Surgery and Medicine, Vol. 45, No 7, p. 410-417, doi:10.1002/lsm.22160*, (2013).

PV Jaana Kuula, **Ilkka Pölönen**, Hannu-Heikki Puupponen, Tuomas Selander, Tapani Reinikainen, Tapani Kalenius and Heikki Saari. Using VIS/NIR and IR spectral cameras for detecting and separating crime scene details. *Proceedings of SPIE Vol. 8359, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense XI, 83590P (May 1, 2012); doi:10.1117/12.918555;*, (2012).

PVI Paavo Nieminen, **Ilkka Pölönen** and Tuomo Sipola. Research Literature mapping using diffusion maps. *Journal of Informetrics, Vol. 7, No. 4, P. 874-886, doi:10.1016/j.joi.2013.08.004.*, (2013).

Hyperspectral imaging has a highly multidisciplinary nature. Hence, this work also has a base in a multidisciplinary research projects. The author has been the project manager and principal data analyst in the following three research projects:

- UASI, Unmanned aerial system innovations, 2011-2013
- Diagnostic, New imaging techniques for early detection, monitoring, and guided treatment of cancer, 2011-2013
- SpeCSI, Crime Scene Investigations by spectral imaging, 2012

These projects were funded by the Finnish Funding Agency for Technology and Innovation (TEKES). In these projects, data analysis has been mainly the author's responsibility. The articles included, **PI-PV**, introduce the research results from these projects.

Articles **PI-PIII** are based on the UASI project. The author participated in planning and executing the research setup and, after that, the processing of the results. The UASI project aimed to solve problems between traditional hyperspectral imaging and UAVs.

The project developed and demonstrated a novel small hyperspectral imager for real-life applications. The research article introduces this hyperspectral imager in a precision agriculture application. Articles **PI** and **PII** concentrate on estimating biomass and nitrogen content from a Finnish wheat field. Here, the author is the main author and performed most of the data analysis work. Article **PIII** concerns the preprocessing of hyperspectral data. The author's contribution to the estimation part of the text is considerable.

Article **PIV** is the result of the diagnostic project. Here, the author also participated in planning and executing the research setup and delivered the analyzed data to physicians. The author also collected most of the research data with the clinicist. The article utilizes a hyperspectral imaging system to detect field cancerization. The author wrote the technical parts of the paper. Article **PIV** will be also part of the main author's, Noora Neittaanäki-Perttu, dissertation.

Article **PV** is based on the SpeCSI project. The author participated in planning and executing the research setup and interpreting the analyzed data. The article introduces the potential of hyperspectral imaging in crime scene investigations.

Article **PVI** reveals information about the popularity of certain types of data mining methods. It also introduces the knowledge discovery process for scientometrics and shows how generally usable this process is. The author contributed to the phrasing of the research questions, the design of the data mining algorithm and the actual data analysis, and the visualization and interpretation of the results.

1 INTRODUCTION

Hyperspectral imaging is based on the electromagnetic scattering of different substances at certain wavelengths. The imaging sensor captures the spectrum of the reflecting substances and creates a spatially and spectrally accurate datacube, which can be analyzed further. Each spatial pixel in the image forms a spectrum through a spectral dimension. Hyperspectral imaging itself is highly multidisciplinary: electronics, sensor technology, optics, data processing, data analysis, and, finally, some interpreted information about the imaged and analyzed objects are all needed.

The concept of the separate visible light spectrum was first introduced by English mathematician and physicist Sir Isaac Newton in *Opticks* (Newton, 1704).

> *This Image or Spectrum P T was coloured, being red at its least refracted end T, and violet at its most refracted end P, and yellow green and blew in the intermediate spaces.*

The method that Newton described is still topical. Most modern spectrometers and hyperspectral imagers are based on the idea of refracting light. The method can be utilized for visible light, but it is also possible to apply it elsewhere in the electromagnetic spectrum (Figure 1). However, infrared and ultraviolet optics are more expensive than visible light optics.

The basic principle in spectroscopy is that different substances reflect different wavebands of light differently. Therefore many substances have a unique spectral signature fingerprint. In spectral analysis, the goal is to analyze different substances based on their properties to reflect or absorb light. Hyperspectral imaging diverges from normal reflectance spectroscopy in that it produces a composite image of all the spectra. As Figure 2 illustrates, a hyperspectral image is composed of a stack of intensity images. Each of these images represents intensity distributed at different wavelength of light. Each pixel in the stack originates spatially from the same place. This means that if a vector of pixels is taken through this stack, it forms a spectrum from the imaged object.

Hyperspectral sensors used in spatial scanning can be divided into four different categories (Chang, 2007):

– Line scanners,

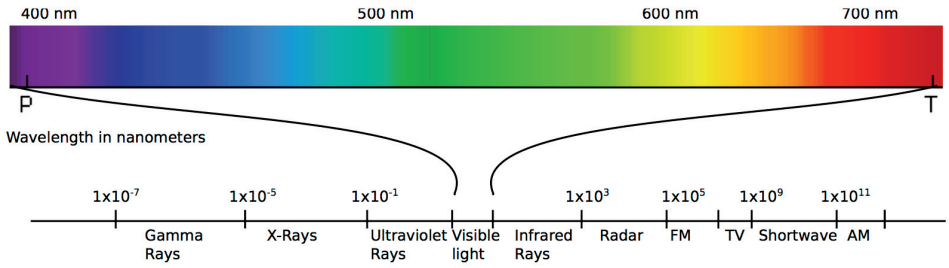

FIGURE 1 Electromagnetic spectrum and the spectrum of visible light.

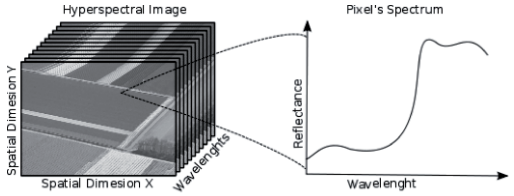

FIGURE 2 Illustration of a hyperspectral image. Each pixel in the image forms a spectrum through imaged wavebands. Figure is originally published in Pölönen et al. (2013).

- Whiskbroom scanners,
- Pushbroom scanners, and
- Framing cameras.

Line scanners were the earliest form of hyperspectral scanning systems. Line and whiskbroom scanners use rotating mirrors in spatial scanning, which makes them impractical, for example, in flight applications. The pushbroom scanner is a popular method to capture spatial data. It does not require moving mirrors because its focal plane consists of a linear array of detectors. A limited field of view is the disadvantage of the pushbroom scanners. Both pushroom and whiskbroom scanners have quite a high signal-to-noise ratio. Framing cameras take on two dimensional planes at once, so in their case, the spectral separation system must be fast or only few spectral bands will be recorded.

Spectral separation in these cases is based on various technologies, which can be divided into five different type (Chang, 2007):

- Prisms,
- Grating,
- Filter wheels,
- Interferometers and
- Acousto-optical tunable or liquid-crystal tunable filters.

The world's leading hyperspectral imager producer (Specim, 2013), Specim Ltd, uses, basically, prisms and pushbroom scanners to capture hyperspectral images.

Chapter 2 of this thesis introduces a novel hyperspectral imager developed by the VTT Technical Research Centre of Finland, the function of which is based on framing cameras and an interferometer.

The beginning of hyperspectral imaging goes back to the late 1970s. Its first applications were implemented in remote sensing in the early 1980s when NASA built the Airborne Imaging Spectrometer (AIS) (Goetz et al., 1985). In remote sensing, hyperspectral cameras were first used to search for minerals (Lillesand et al., 2004; Chang, 2007) and for vegetation research (Thenkabail et al., 2012; Jones and Vaughan, 2010). Due to the high price of hyperspectral imagers, they have not become very common. The main user groups are academics and the military. In addition, extractive industry utilizes hyperspectral imaging when they are looking for minerals or establishing new mines.

Hyperspectral data itself has a high-dimensional nature, from dozens to hundreds of dimensions. One image can contain information from millions of spatial pixels, and it is the possible that the obtained spectrum is actually a mixture (linear or nonlinear) of the spectra. Analysis of the hyperspectral images and their problems can be divided into six main categories (Bioucas-Dias et al., 2013):

- Data fusion (Segl et al., 2003; McKeown et al., 1999; Tsagaris et al., 2005),
- Unmixing (Keshava and Mustard, 2002; Winter, 1999; Nascimento and Dias, 2005; Plaza et al., 2011; Bioucas-Dias et al., 2012; Averbuch et al., 2012; Averbuch and Zheludev, 2012),
- Classification (Goel et al., 2003; Mountrakis et al., 2011; Camps-Valls and Bruzzone, 2005; Chang, 2003),
- Target detection (Chang, 2007; Manolakis, 2002; Averbuch et al., 2012; Averbuch and Zheludev, 2012),
- Physical parameter retrieval (Dorigo et al., 2007; Thenkabail et al., 2012) and
- Fast computing (Plaza et al., 2011).

Of these problems, this thesis will deal mainly with those of unmixing and classification.

Because of the high price of hyperspectral sensors and the nature of the computational complexity, these devices have not been widely used. In the 2010s, the increased computational power of home computers and GPUs have made it possible to handle hyperspectral data almost anywhere. Novel ideas in spectral separation make these devices more readily available to a broader selection of applications.

In this thesis, hyperspectral imaging has been applied to remote sensing from unmanned aerial vehicles, to precision agriculture, and as a medical *in vivo* imaging device in dermatology and in crime scene investigation. According to the United Nations Population Division, in 2050, there will be more than 9 billion people living on the earth. This means that food production must increase almost 50 %, which implies that resources must be used very efficiently in the near future. Precision agriculture aims to use resources as efficiently as possible. For example, there is the desire to bring fertilization to its peak efficiency. Precise fertilization requires accurate machinery and also information about where and

how much fertilization is needed. With hyperspectral imaging, it is possible to create estimation maps for the nitrogen content and biomass of crops.

Non-melanoma skin cancers (NMSC) represent the most common cancer type in Caucasians. The incidence of NMSC is constantly increasing worldwide now up to 2-3 million new NMSCs diagnosed each year (WHO). Skin field cancerization refers to the presence of NMSC, multiple precursors called actinic keratoses and clinically invisible dysplastic keratinocytes with the potential to transform into actinic keratoses on sun-exposed skin areas. Field cancerization increases the risk for NMSC and thus early therapy is recommended. In this study we used hyperspectral imager for early detection of the affected skin areas.

In addition, solving crimes is not as easy as it looks in television series. A crime scene can be very large, and there is a need for methods for screening of the scene. In screening, attempts are made to find, for example, biological stains and separate them from each other. This can speed up investigation processes and lead to the suspect's trail. With hyperspectral image processing, it is possible to detect small, even sub-pixel-level traces and identify them.

1.1 Research approach

This study focuses on analyzing hyperspectral imaging data for various applications. Its main focus is on hyperspectral imaging applications based on the Fabry-Perot interferometer (FPI). The FPI-based approach in spectral imaging offers potential in new application areas such as medical imaging, remote sensing from UAVs, and crime scene investigation. This novel hyperspectral imager technology is explained in detail in Chapter 2. This is the first doctoral-level study that, in quantitative experiments, addresses the whole data analysis chain with hyperspectral imagers based on the Fabry-Perot interferometer. The main research questions of this study are as follows:

1. Can hyperspectral imaging based on the Fabry-Perot interferometer be effectively utilized in UAV, medical, and crime scene investigation applications?
2. How, in these three novel application areas, can we preprocess and analyze hyperspectral imaging data based on the Fabry-Perot interferometer?

This work approaches these questions utilizing the process of knowledge discovery from databases introduced by Fayyad et al. (1996). This process allows us to perceive how raw data can be turned into intelligence. It describes the steps to be taken before the actual end user can utilize the data.

1.2 Structure of the work

First, a knowledge discovery process will be utilized for analyzing hyperspectral data in different applications. After that, data analysis results from the included original research articles will be discussed.

Chapter 2 explains the knowledge discovery process from databases in detail. The steps of this process will be tied to the included articles. Chapter 3 introduces various application areas where the developed novel hyperspectral imaging technology has been utilized. Then, the analysis results are presented. Finally, Chapter 4 summarizes the results.

1.3 Other publications and public presentations

The author has also contributed to several other studies that touch on the topic of this thesis. The results of these studies were published in the following articles:

- Hyperspectral imaging based delineation of malignant tumours (Zheludev et al., 2013).
- Miniaturized hyperspectral imager calibration and UAV flight campaigns (Saari et al., 2013).
- A case study of a precision fertilizer application task generation for wheat based on classified hyperspectral data from UAV combined with farm history data (Kaivosoja et al., 2013).
- Spectral imaging from UAVs under varying illumination conditions (Hakala et al., 2013).
- Methods for estimating forest stem volumes by tree species using digital surface model and CIR images taken from light UAS (Salo et al., 2012).
- Using hyperspectral imaging for detecting destructive subjects and materials (Kuula et al., 2012).

2 THEORETICAL FOUNDATION

In this chapter, the theory and research methods relevant for this thesis are introduced. First, the knowledge discovery process for databases is introduced, and after that, the executed research is introduced based on the steps of the process.

2.1 Knowledge discovery process for databases

The methodological background of data collection and processing pertains to all the research articles included in this thesis' knowledge discovery process from databases (KDD), which was introduced by Fayyad et al. (1996). The included research articles either utilize the whole discovery process or concentrate on some parts of it. The steps of the KDD process are introduced in Figure 3. These include data selection, preprocessing, transformation, data mining, and eventually interpretation/evaluation of the results.

Selection: With large databases and data sets, it is impractical to use all the available data. Thus, it is necessary to focus only on a subset of data or a subset of variables in data. Selection is determined by the goal of the KDD process itself about what is needed to be discovered.

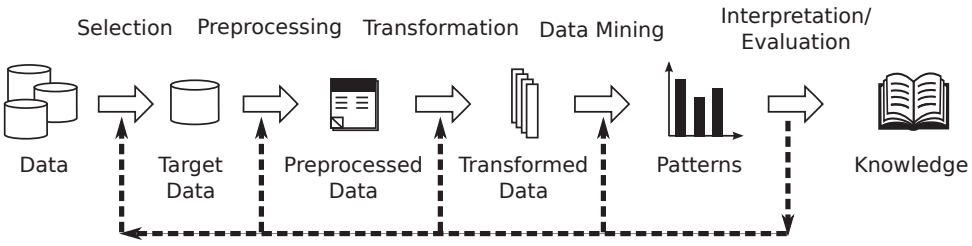

FIGURE 3 Illustration of the knowledge discovery process from databases. Figure is originally published in Nieminen et al. (2013).

Preprocessing: The selected data include noise, some parts of it may be missing, or there may be a need to normalize it somehow. These issues are corrected in preprocessing. For example, some noise removal can be done. Typically, preprocessing requires a large amount of signal processing work. Often workload of this step is underestimated when new applications are studied.

Transformation: Pure data from a data set may not give optimal results by itself. All feature extraction, feature generation, and dimensional reduction is done in this transformation step.

Data mining: In this step, patterns of interest in the data are searched for. The actual data mining process of this step may consist of, for example, clustering and classification patterns.

Interpretation: In the final step, the extracted information patterns are evaluated and some interpretations are carried out. After this step, the interpreted knowledge is ready for direct use.

It is worth noting that any step of the knowledge discovery process may involve loops between the steps and can have multiple iterations. This is also illustrated in Figure 3 with a dashed line and arrows. The KDD process is well suited for handling large and high-dimensional data sets.

The KDD process can be utilized in hyperspectral imaging data analysis. A hyperspectral image easily contains from hundreds of thousands to millions of data points, and each data point includes hundreds of parameters. The nice thing about hyperspectral data is that there are rarely any missing parameters. The data noise level is completely dependent on the instrument being used and on the imaging conditions.

2.2 Data acquisition and selection

All the hyperspectral data used in this thesis were collected during the research projects. The main instrument used for that purpose was a small and lightweight hyperspectral imager developed by VTT. Spectral separation in these devices is carried out through a piezo-actuated Fabry-Perot interferometer (FPI). The component is hermetically sealed in a metal can filled with nitrogen. Both the parallelism and the distance between the mirrors of the Fabry-Perot interferometer must be controlled with high accuracy. This is achieved with three closed-loop control channels at the edges of the mirror plates. Each channel has a piezoelectric actuator with a closely positioned capacitive measurement point, which is used to determine the mirror separation. Each channel is controlled with nanometer accuracy to obtain the desired parallelism and air gap between the mirrors.

Two of these imagers were designed for unmanned aerial vehicles, and one is a hand-held device custom-built for the University of Jyväskylä. Figure 4 shows the optical route of a hyperspectral imager based on the Fabry-Perot interferometer (FPI-HSI) and meant for UAVs. The university's hand-held spectral imager has a similar optical route: basically, just one folding mirror, which

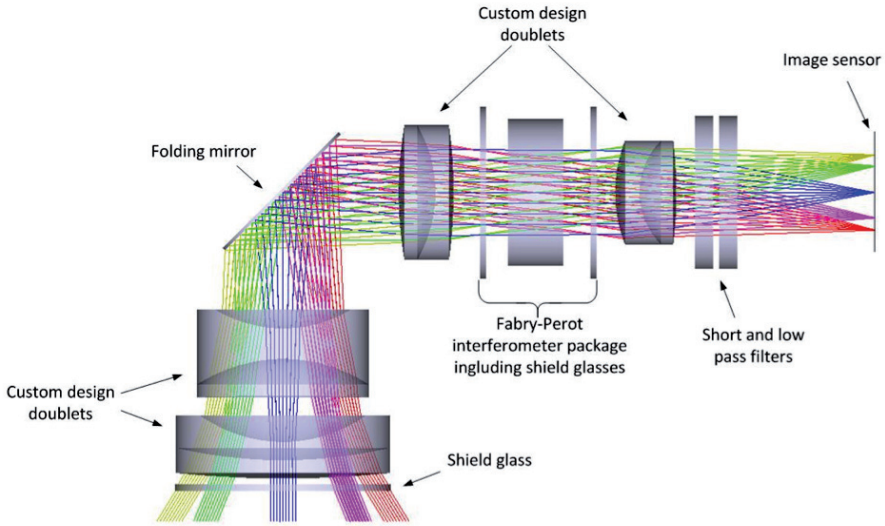

FIGURE 4 Optical path of the hyperspectral imager based on the Fabry-Perot interferometer. Figure is from Pölönen et al. (2013).

refracts the light route, is missing.

VTT's spectral imager uses the CMOS sensor. This and the piezo-actuated Fabry-Perot interferometer enable quite a high speed of image acquisition. Depending on the lighting conditions, the exposure time for each plane can vary from 0.1 to 1,000 milliseconds. The whole spectral cube can be recorded in approximately one second. This depends on how many wavebands we want to include in the measurement and which resolution we want to use. For example, the 2012 UAV imager uses the 4Mpix RGB CMOS CMV4000 image sensor from CMOSIS. The sensor has an on-chip analog-to-digital conversion (ADC) that provides 12-bit resolution. The default resolution used in the prototype is 2x2 binned XVGA (1024x648 pixels). In this mode, up to 24 raw images can be saved to the buffer memory in one image burst. The full 4Mpix mode can also be used, but the 32 MB buffer memory limits the number of images to three for each image burst. This mode can be used if only a small set of wavelength bands is required. A full detailed description of the FPI-HSI can be found in an article from Mäkynen et al. (2012)

FPI-HSI's sensor output consists of raw digital numbers (DN), that is, photons per pixel per second. Depending on what the final application is and how the lighting conditions vary, there might be a need to convert DNs to surface reflectance values. This is essential in remote sensing.

Table 1 presents the specifications for VTT's spectral imagers. The hand-held device and the UAV 2011 are, approximately, the same generation devices. Both were manufactured in 2011. The major difference between them is in the optics. For the hand-held device, it is possible to use various custom lenses, which

TABLE 1 Specifications of VTT's Fabry-Perot imagers: hand-held, UAV 2011, and UAV 2012

Parameter	Hand-held	UAV 2011	UAV 2012
Horizontal and vertical FOV (deg.)	>36,>26	>36,>26	>50,>37
Nominal focal length (mm)	9.3 ± 3 (Custom lenses)	9.3	10.9
Wavelength range (nm)	500-885	400-900	400-900
Spectral resolution at FWHM (nm)	9-40	9-45	10-40
Adjustable spectral resolution step	<1	<1	<1
f-number	<6.7	<7	<3
Maximum spectral image size (pixels)	2592×1944	2592×1944	2048×2048
Spectral image size with default binning (pixels)	320×240	640×480	1024×648
Camera dimensions (mm)	$62 \times 66 \times 219$	$65 \times 65 \times 130$	$80 \times 92 \times 150$
Weight (g)	<450	<420	<700

enables its use in different applications. The 2012 device contains the Fabry-Perot interferometer, which has a larger diameter than in the earlier imagers. This enables a better f-number, so this device's signal-to-noise ratio is much higher than that of the hand-held device or UAV 2012.

In articles PI-PIV, the main instruments are hyperspectral imagers based on the Fabry-Perot interferometer. In a precision agriculture application, UAV 2011 and UAV 2012 were utilized. The hand-held device was utilized in a medical application and in a crime scene investigation. Figure 5 shows a UAV 2012 hyperspectral imager. This device was on a small microcopter, which can carry a payload of approximately one kilogram. The microcopter flew along a route, which had been determined in advance, over a southern Finnish wheat field. Besides collecting hyperspectral data, the imager also gathered position information through its GPS receiver and information about changes in lighting conditions through its irradiance sensor. All the data was gathered on a CompactFlash memory card during the flight. The data from the memory card was transferred for further processing on the ground after the flight.

In article PIV, the hand-held hyperspectral imager was utilized for medical imaging. Medical imaging differs greatly from remote sensing: the object is not moving and, the lighting conditions are under human control. Figure 6 shows the sketch of a setup of an imaging system. A hyperspectral imager was mounted to an external (3D-printed) handle, shown in Figure 7. This handle also has a

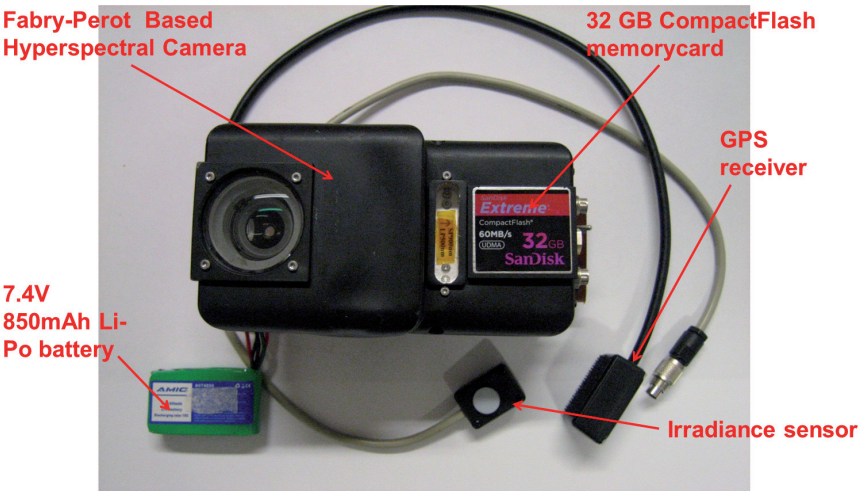

FIGURE 5 Fabry-Perot hyperspectral imaging system. Image is from Pölönen et al. (2012).

place for ring light, which were connected with a light fiber to a halogen light source. Hyperspectral cubes were connected by a USB cord to the laptop for further processing.

Article PV deals with the utilization of hyperspectral imaging in a crime scene investigation. Here, the instrument employed was VTT's hand-held hyperspectral imager, but two commercial available hyperspectral imagers from Specim Ltd were also used. One of the Specim imagers works in 400-1,000 nm and the another in 1,000-2,500 nm. These devices are often called VNIR (visible, near-infrared) and SWIR (short-wave infrared) cameras. They are based on the traditional pushbroom scanning technology. Compared to the small hyperspectral imagers based on the Fabry-Perot interferometer, these devices have a higher signal-to-noise ratio because their optical throughput is an order of magnitude higher. On the other hand, as scanning devices, these are not flexible to use. They need either a rotation stand or a moving platform to capture whole images.

The number of wavebands that are needed depends on the final application. There are situations where only a few wavebands are needed. Then, it would be a waste of time and memory to collect all the possible wavebands. FPI-HSI has a major advantage here compared to pushbroom scanners. Because its filter is tunable, it is possible to select only the needed wavebands to be captured. This shortens the imaging time.

2.3 Preprocessing of hyperspectral data

The preprocessing required depends on the final application. If there is a need to compare different images with each other or to use spectral libraries in a situation

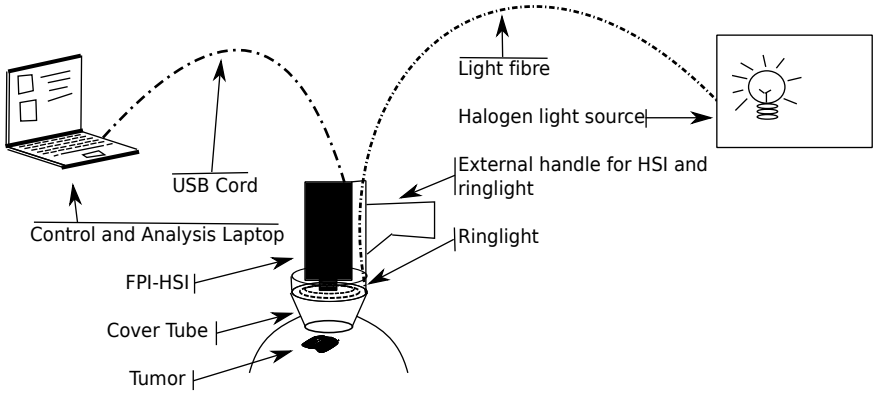

FIGURE 6 Imaging system setup used in a medical application.

FIGURE 7 Developed hand-held hyperspectral imaging system in use. Image is from Neittaanmäki-Perttu et al. (2013).

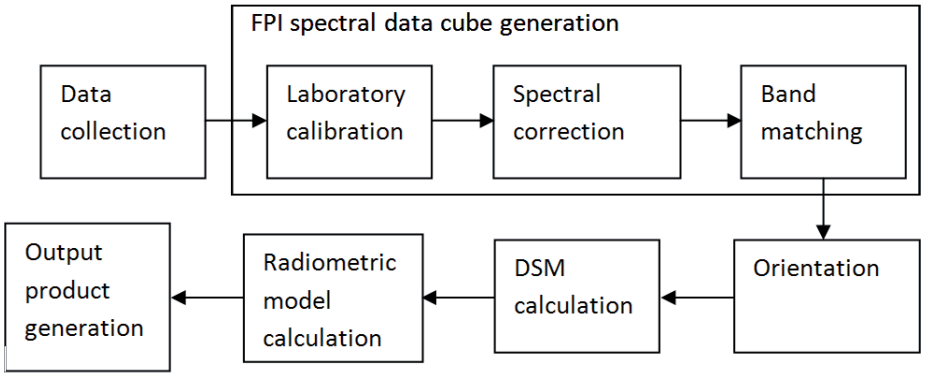

FIGURE 8 Workflow for the preprocessing of gathered hyperspectral data in remote sensing applications. Process chain is from Honkavaara et al. (2013).

where the lighting conditions vary among hyperspectral images, it is necessary to normalize the data and minimize the changes in the data. This situation is common in remote sensing applications. If the hyperspectral imager is in a satellite or in a plane flying high, complex atmospheric models and corrections are needed to reduce any distortion effects caused by the earth's atmosphere.

One widely-used way to reduce changes in the sun's irradiance received is to use bidirectional reflectance distribution functions (BRDF). The FPI-HSI system has an external irradiance sensor, which collects information about the changes in irradiance during flight operations. This information can be utilized for the radiometric model calculations. It is also possible to use a totally external irradiance sensor, which is kept on the ground during the flight. Different methods of performing these corrections and how they affect the data are explained in detail in article PIII.

Because each FPI-HSI takes images in bursts of three image planes, it takes approximately one second to take the whole cube. This is not a problem if the imaged object does not move, as in many medical application and forensics applications. However, for remote sensing this brings one extra challenge. There is a need to match different wavebands because a moving plane causes some drift in the hyperspectral datacube. Figure 8 presents the workflow of preprocessing for a hyperspectral datacube from UAV. The objective is to process all hyperspectral images taken into one coherent hyperspectral datacube. This datacube is basically a mosaic from the hyperspectral images taken. It is notable that FPI-HSI also delivers the digital surface model (DSM), which can be utilized in the analysis of images. A traditional pushbroom imager does not do this.

In articles PI-PIII, preprocessing from spectral images to a whole mosaic is described briefly in the conference paper of Honkavaara et al. (2013) with the following steps:

1. System corrections of the images using laboratory calibration, spectral cor-

rection, and dark signal correction. These parameters and algorithms are provided by VTT (Mäkynen et al., 2011).
2. Image quality assessment and signal-to-noise ratio calculation (SNR).
3. Pixel transformation from 32-bit float format to 16-bit unsigned integer format.
4. Registration of layers to form spectral data cubes of individual images.
5. Determination of the image orientations of the reference layers using a self-calibrating bundle block adjustment (Honkavaara et al., 2012a; Rosnell and Honkavaara, 2012).
6. Optionally, a DSM can also be calculated (Rosnell and Honkavaara, 2012).
7. Determination of the radiometric imaging model to compensate for atmospheric and illumination influences, and a non-uniformity related to a view (or illumination), as well as reflectance transformation. A radiometric block adjustment method is being developed to determine optimal parameters by utilizing overlapping images (Honkavaara et al., 2012a,b).
8. Calculation of output products, which include 3D spectral point clouds, digital surface models, spectrometric image mosaics, and object bidirectional reflectance data (Honkavaara et al., 2012b).

Because the lighting conditions are constant, and the imaged object does not move in the medical application and the crime scene investigation, there is not a need for as heavy preprocessing as in remote sensing. If there were a need to use external spectral libraries or to compare different hyperspectral images to each other when the lighting conditions change greatly, there would be a need to transform DN into reflectance. In these cases, transformation would be beneficial, but it would also affect the data's dynamic range. If the data must be converted into reflectance, the exposure of the hyperspectral image must be adjusted to the level of white reference. This can affect the maximum dynamic range.

2.4 Transformation of hyperspectral data

Hyperspectral data can be used directly in the data mining step. This basically depends on how characteristic the features of the imaged object are and what the goal of the data processing is. On the other hand, transformed hyperspectral data can provide information that is self explatonary. Spectral feature extraction can be divided into some major categories: spectral indices, spectral unmixing, signal processing methods (wavelets, etc.), and spatial feature extraction.

Vegetation indices are one example of physical parameter retrieval and they are also the easiest and computationally the cheapest way to extract features from spectral data. Basically, a vegetation index can be one spectral band divided by another. These are called as the Simple Ratio (SR) or Ratio Vegetation Index (RVI). For example, the classical Normalized Difference Vegetation Index (NDVI)

(Rouse et al., 1973) is calculated as follows:

$$R_{NDVI} = \frac{\lambda_{nir} - \lambda_{red}}{\lambda_{nir} + \lambda_{red}}.$$

NDVI contrasts the density with green vegetation density. Here λ refers to to certain wavebands. The wavebands for $\lambda_{nir} = 816$ nm and $\lambda_{red} = 690$ nm are used. For chlorophyll, there is a vegetation index called MCARI (Daughtry et al., 2000)

$$R_{MCARI} = ((\lambda_{700} - \lambda_{670}) - 0.2 * (\lambda_{700} - \lambda_{550}) * \frac{\lambda_{700}}{\lambda_{670}}.$$

One efficient way to extract feature information is spectral unmixing. This technique assumes that the pixel content in a hyperspectral image is a mixture of various substances. Unmixing can be divided into three different models: linear, bilinear, and non-linear. In this study, both the linear model (Keshava and Mustard, 2002) and the non-linear (Chen et al., 2011) model have been utilized for feature extraction.

Let $\mathbf{x} = (x_1, x_2, \dots x_L)^T$ be the observed spectrum with L wavebands. Let us assume that there are N pure spectra, often referred to as endmembers, in a hyperspectral image. Each endmember is vector $\mathbf{s} = (s_1, s_2 \dots s_L)$. Endmembers form an $L \times N$ matrix $S = (\mathbf{s_1}, \mathbf{s_2} \dots \mathbf{s_N})$. Now $\mathbf{m} = (m_1, m_2 \dots m_N)^T$ is the $(1 \times N)$ abundance vector for endmembers S. These abundance vectors depend on the observed spectrum $\mathbf{x}$.

As given in (Keshava and Mustard, 2002), the linear mixing model can be described for each waveband λ_n as

$$\mathbf{x}[\lambda_n] = \sum_{i=1}^{N} m_i \mathbf{s_i}[\lambda_n] + w[\lambda_n]$$

where N is the number of endmembers, w is a noise term, and n denotes a certain waveband. Expanding the linear mixing model to all k observed pixel spectra, we will have a matrix form for the model

$$X = MS + W,$$

where $X = (\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_k})^T$, $M = (\mathbf{m_1}, \mathbf{m_2}, \dots, \mathbf{m_k})$, and W is pixel-wise and waveband-wise noise.

A general non-linear model can be formulated as follows:

$$\mathbf{x} = \mathbf{\Psi}(\mathbf{S}) + w,$$

where $\mathbf{\Psi}$ is a non-linear function that describes the connection between endmembers and the obtained spectrum. Thus, in the linear model unknown parameter is M, and in the non-linear model, the unknown function is Ψ. It is quite clear that the linear unmixing model can be solved in multiple ways. A classical solution for the problem is given by the least squares method. One fast way of solving the

the problem is to use a filter vector algorithm (FVA) (Bowles et al., 1995), which is a linear way to solve inverse problem. As Pesses (1999) conclude that the least square inversion's computational cost is $O(n^2)$ higher than FVA's.

FVA forms a set of filter vectors F, which are used to estimate the abundance coefficients M. The estimation is performed as follows:

$$M = FY,$$

where

$$F = (RS)^{-1}R$$

and

$$R = S^T - (\frac{J}{N}S)^T$$

where J is an $N \times N$ unit matrix. Now M reveals each endmember's abundances in the hyperspectral image.

While non-linear models have not been particularly attractive, linear models have been widely examined. The main reason for this is the complexity of non-linear problems. Actually, this is a quite common phenomenon, as indicated in article PVI, where the frequency of linear methods (125) in data mining literature is double that of non-linear methods (61). Recently, there has been more published research concentrating on non-linear unmixing. The K-Hype algorithm developed by Chen et al. (2011) was tested in article PI. K-Hype is a kernel-based approach to solving the non-linear unmixing problem. The abundance matrix M and the results of function $\mathbf{\Psi}$ can be used as features in the estimation process.

These unmixing models require knowledge about pure spectra. Spectra are known from the relevant libraries or perceived from hyperspectral data themselves. Algorithms such as N-Findr (Winter, 1999) or Vertex Component Analysis (VCA) (Nascimento and Dias, 2005), look for pure spectra from the datacube. Both algorithms are based on approach of data's geometry. There are quite many alternative approaches to this problem (Bioucas-Dias et al., 2012). VCA is widely used in hyperspectral remote sensing. The main assumption in VCA is that there is a pure spectra $\mathbf{s}$ presence in observed data set $X = (\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_d})$, where d is the number of spectra in the image. VCA basically determines the extreme points of data set X in the projected subspace, and these points are selected as endmembers. Figure 9 illustrates a projected subspace. The extreme points of this convex hull are marked with a star.

If we want to use all the extracted features without any feature selection, then a manifold learning approach can be utilized. The included articles PI, PII and PVI utilize a relatively new dimensional reduction method called diffusion maps (Coifman and Lafon, 2006a). For a chosen distance function, this method finds a low-dimensional representation. Thus, the high-dimensional data points become embedded in a lower-dimensional space. The dimensionality reduction yields a space where the Euclidean distance corresponds to the diffusion distance in the original space (Coifman and Lafon, 2006a; Nadler et al., 2008).

Let us consider a feature set that consists of feature vectors extracted from spectral data. Let $X^f = \left\{x_1^f, x_2^f, \ldots, x_n^f\right\}$ be the feature set and $x_i^f \in \mathbb{R}^p$ be a

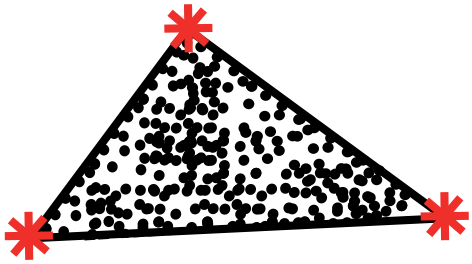

FIGURE 9 Illustration of the projection of a high-dimensional convex data set to a two-dimensional simplex. In this illustration VCA can find three endmembers (stars) from extreme points of convex hull. Data points, which are on line between two endmembers, are mixture of these two. Data points in convex hull are mixture of all three endmembers.

feature vector length of p. As described in Nieminen et al. (2013), diffusion map is a kernel method that utilizes the Gaussian kernel matrix W,

$$W_{ij} = \exp\left(-\frac{\text{dist}(x_i^f, x_j^f)}{\epsilon}\right),$$

where $\text{dist}(x_i^f, x_j^f)$ is the Euclidean distance. Matrix W is normalized a $P = D^{-1}W$, where $D_{ii} = \sum_{j=1}^{n} W_{ij}, i \in 1 \ldots n$. The matrix represents the transition probabilities between data points. To determinate the eigenvalues of P the conjugate matrix, $\tilde{P} = D^{\frac{1}{2}} P D^{-\frac{1}{2}}$ is calculated. From this, we get

$$\tilde{P} = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}.$$

This is a normalized graph Laplacian (Chung, 1997), and it maintains the eigenvalues (Nadler et al., 2008). Singular value decomposition (SVD) $\tilde{P} = U\Lambda U^*$ is used to find the eigenvalues $\Lambda = \text{diag}([\lambda_1, \lambda_2, \ldots, \lambda_n])$ and eigenvectors $U = [u_1, u_2, \ldots, u_n]$ for $\tilde{P}$. The eigenvalues for P are the same as for $\tilde{P}$. The eigenvectors for P are found with $V = D^{-\frac{1}{2}} U$ (Nadler et al., 2008). Low-dimensional coordinates Φ are created using $\Phi = V\Lambda$. It is notable that computational cost of diffusion maps is high. Thou, it is impractical to embed all data in feature extraction phase with diffusion maps. In manifold learning diffusion maps should be used only for training data. There are methods such as Nyström extension (Fowlkes et al., 2004) and geometric harmonics (Coifman and Lafon, 2006b) for embedding rest of the data to lower dimension .

As mentioned earlier, some extracted features can be utilized directly in interpretation. This is the case especially in unmixing. As a result of unmixing, abundance maps give information about each found substance's relative fraction in each pixel. For example, in the remote sensing of minerals, this information is highly profitable.

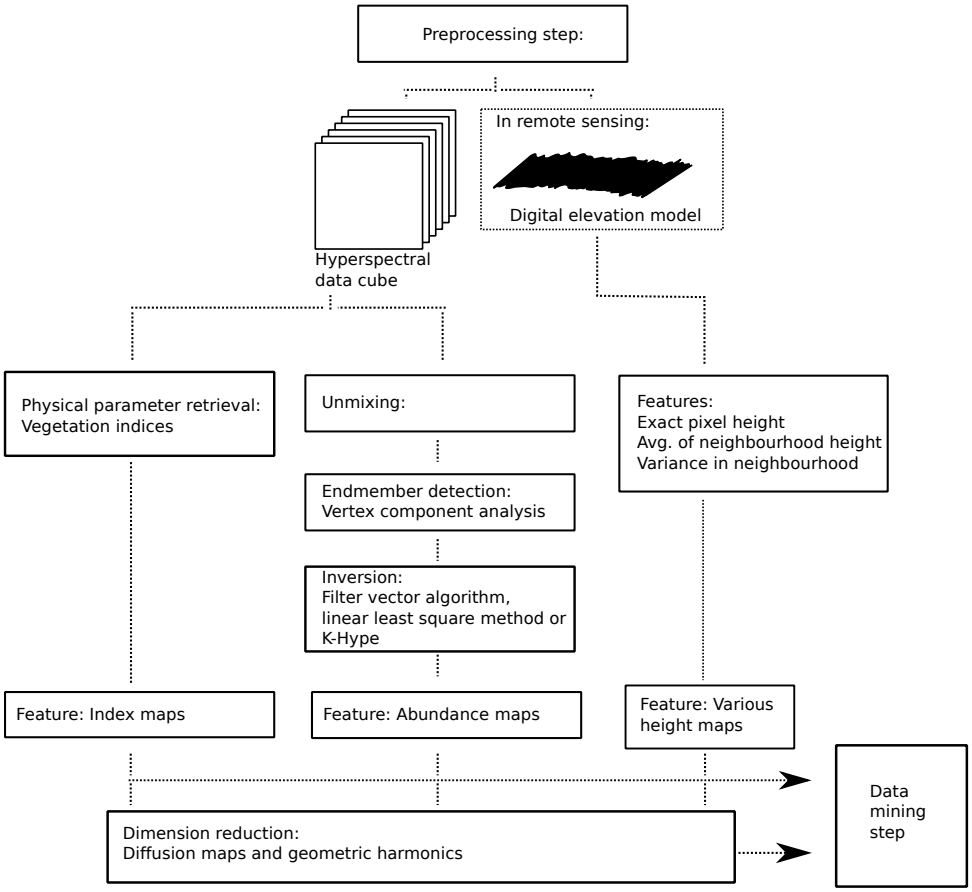

FIGURE 10 This drawing illustrates used feature extraction process. After preprocessing step there exist hyperspectral data cube and optionally digital elevation model. In included articles feature extraction schema follows two main lines vegatation indices and unmixing. Results of these lines are usable straight in data mining step, but it is possible also gather features together and use dimension reduction.

2.5 Data mining of hyperspectral data

Data mining itself is determined in multiple ways: one common definition introduced by Hand et al. (2001), is

> the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.

The data mining step can include for example, classification or clustering of the data. In this study, both methods are used. The included articles PI-PV utilize a supervised machine learning approach, which is illustrated in Figure 11. In these cases, training data was selected either from a single hyperspectral image, a hyperspectral image mosaic, or a collection of hyperspectral images. This training data received pixelwise labels. After the selection of the training data, features were extracted as explained earlier. Both the labels and the training data were directed to machine learning algorithms. Four different algorithms are utilized in this thesis:

- K-nearest neighbours (KNN) (Hand et al., 2001)
- Leave-one-out KNN for cross-validation (Hastie et al., 2001)
- Support vector machine (Hastie et al., 2001; Han and Kamber, 2006)
- Feedforward neural network with multilayer perceptron (Hastie et al., 2001; Hand et al., 2001; Han and Kamber, 2006)

After the training of the algorithm, the rest of the hyperspectral data can be analyzed. Similar feature extraction is then done. After that, the data is guided to a trained classifier. The way data is handled depends on the classifier. As a general principle, classifiers classify data so that the extracted feature vectors that are similar to each other are classified into the same group.

For example, KNN calculates the distance between points in a training data set and points e_n, which need to be estimated $(n = 1, ..., d)$. The algorithm selects k data points, the Euclidean distance of which to a point e_n is the shortest. With $k = 5$ selected, the five nearest neighbors in the Euclidean space are calculated from the trained data set. The estimate $\hat{e}$ is calculated from the label values l_i of these five neighbors with the following:

$$\hat{e} = \frac{\sum_{i=1}^{k} l_i}{k}.$$

This finally gives the prediction or estimation results for the whole hyperspectral image.

In the case of leave-one-out, KNN estimation results only in labeled training data being used. This method estimates each of the training data vectors so that, one by one, each vector is classified with the rest of the training data. In the end, the prediction results are formed from these classified points. This method indicates the reliability of the classifier and performs cross-validation for the model.

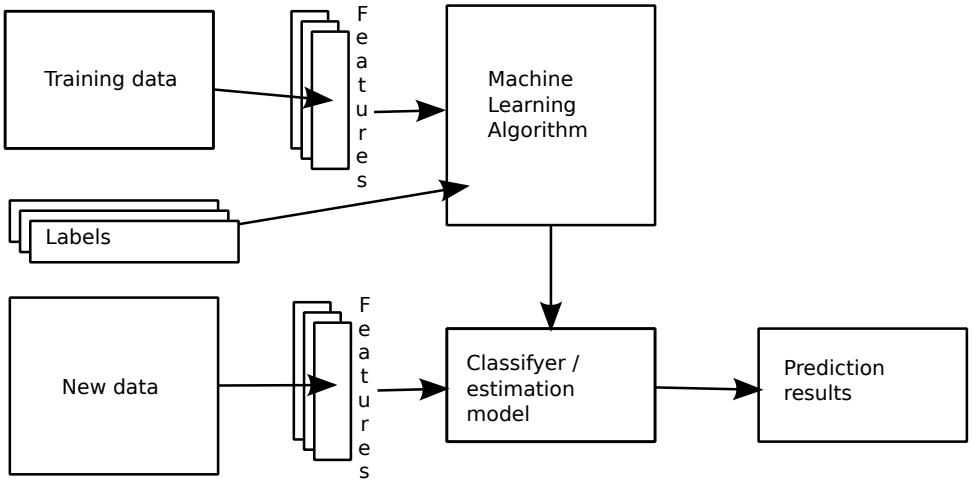

FIGURE 11 Supervised machine learning approach.

Actually, it does not really matter what kind of classifier is used. Most of them give similar answers; the differences are in their computational efficiency (Read et al., 2012). This underlines the importance of feature extraction.

The included research article PVI deals with the data mining step with clustering. A classical agglomerative clustering using the Ward method for cluster distances is applied there. The agglomerative hierarchical clustering scheme is discussed in Everitt et al. (2001) and Hastie et al. (2001). The number of clusters is determined using the silhouette measure, and the number yielding the highest average silhouette for a clustering is chosen, as recommended by Rousseeuw (1987). Clustering can also be utilized in hyperspectral image analysis, for example, to distinguish among different crops from the field.

2.6 Interpretation of the results

The application areas of this research are multiple. Interpretation of the results is highly related to the knowledge of experts in the application area. Interpretation can be easy if the labels of the training data are based on accurate measurements. Then it will be possible to see how correct the estimations are, and there will also be a possibility to calculate some correlations and prediction error values such as the root mean square error. This is the case, for example, in the precision agriculture application, where the label values from sample plots in a field are measured in the laboratory after collection.

On the other hand, for example, in medicine, the ground truth is based on a pathologist's opinion, which is finally based on biopsies. In the crime scene investigation paper, most of the samples were manufactured by actual crime scene

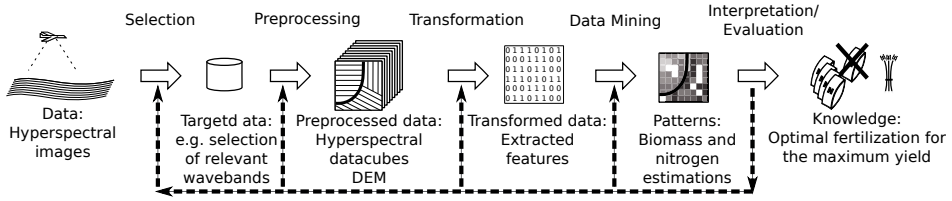

FIGURE 12 Illustration the KDD process applied to precision agriculture. Figure is from Pölönen et al. (2013).

investigators from the Finnish National Bureau of Investigation. In these cases, interpretation is not something measurable but more of an expert opinion about the relevance of the results.

2.7 Overview of analysis workflow in applications

Different applications uses processing and analysis workflows, which consist of the methods described above. The combination of methods depends on the final application. In precision agriculture, the goal is create more accurate fertilization maps, which could lead to the use of less fertilization and a richer harvest. In medicine the hyperspectral imager was used to help the dermatologists to define the area of subclinical precancerous skin lesions (i.e. subclinical actinic keratoses) for early treatment of the affected area. Crime scene investigator collect's huge amount of samples to further investigation. To speed up the laboratory process some priority information between the samples is needed. Figures from 12 to 14 illustrate how the KDD process is utilized in these application areas.

In Figure 12, the KDD process is applied to precision agriculture. In the included articles PI-PIII, data was first collected with a hyperspectral imager, which is mounted to a small microcopter. Images are gathered over a field. In pre-processing, spectral cubes are registrated, and then some atmospheric and irradiance corrections are done. After that, some features are extracted based on unmixing and vegetation indices. Then a supervised machine learning algorithm (KNN) is applied to estimate nitrogen content and biomass. Based on these estimations, it is possible to create fertilization plans.

In dermatology applications, patients are first imaged. The included article PIV uses two approaches; the first one is based on supervised machine learning based on neural networks, and the second exploits abundance maps, which are generated using VCA and FVA or least square inversion. In the preprocessing step, only light noise removal is conducted, which is based on avaraging each spatial pixel to its neighbourhood and, in the spectral dimension, each waveband to the closest neighbours. Figure 13 presents KDD process with an unmixing procedure for a dermatology application.

A challenge in crime scene investigation is the variation of different crime

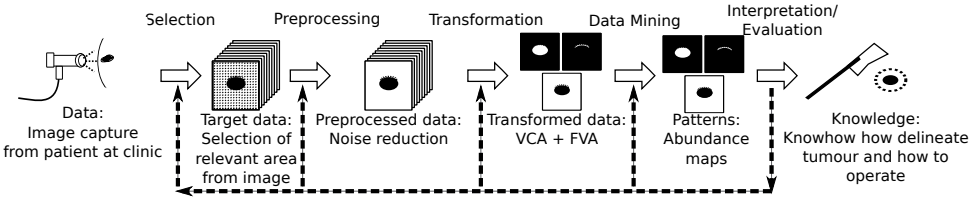

FIGURE 13 Illustration the KDD process in a medical application.

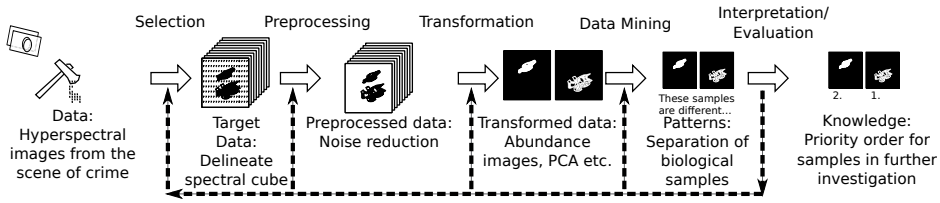

FIGURE 14 Illustration of the KDD process applied to crime scene investigation.

scenes and external conditions. Crime scenes can be in the darkest cellar or broad daylight. Crimes happen year-round, so temperature and moisture conditions also vary. If these are ignored and we concentrate only hyperspectral imagings proof of relevance in crime scene investigation, which has been done in controlled enviroments, we can apply the KDD process as suggested in Figure 14. In the included article PV, after image capture, noise removal was done as described in previously. Then some patterns are extracted based on unmixing. In addition some dimension reduction methods can used for separating different substances. For example, principal components analysis returns false color composition based on calculated eigenvectors of spectra. These can be used, for example, to compare different colourants in the same image.

3 RESULTS

In this chapter, the results of the included articles are presented. The first articles, PI-PIII, introduce FPI-HSI in the field of remote sensing and precision agriculture. After that, FPI-HSI is implemented in a medical device for dermatology and in a crime scene investigation in articles PIV and PV. Finally, article PVI shows how the knowledge discovery from database process can be applied to totally different topics, such as scientometrics.

3.1 Precision agriculture

The climate in northern Europe makes the utilization of hyperspectral imaging for precision agriculture challenging. Cloud conditions combined with short growing seasons make traditional satellite and airborne hyperspectral imaging almost useless. In recent years, the fast development of unmanned aerial vehicles (UAV) and hyperspectral imagers has created totally new aspects for this problem. Research papers PI-PIII introduce one solution to this problem. The solution relies on the hyperspectral imager based on the Fabry-Perot interferometer and UAVs.

Our test field at Vihti is located in Southern Finland. Article PI reports results based on flight campaigns from 2012 taken with the FPI-HSI UAV 2012 prototype camera. The flight and imaging campaign was executed in mid-summer 2012 just before tillering. In Finland, there is usually only one extra fertilization period, due to the short growing season. Seeding and fertilization had been done in May. In the field (Figure 15), two types of wheat and barley were seeded. In addition, different amounts of seed and nitrogen were administered. In total, 50 spots of one square meter each, were selected. The spots' accurate location was determinate with GPS. GPS receivers accuracy was 3 cm, which is below spectral images spatial resolution. All vegetation in those spots was collected and then measured in the laboratory. The measured values were dry weight, moisture, and nitrogen content.

FIGURE 15 The test field located in Southern Finland, Vihti. The hyperspectral images were taken in mid-summer 2012. Image is from Kaivosoja et al. (2013).

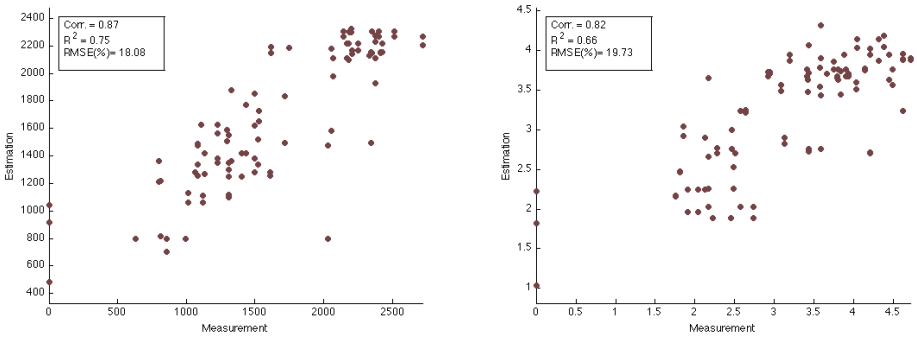

FIGURE 16 Scatter plot between estimations and measured values (left: biomass, right: nitrogen). These are originally published in Pölönen et al. (2013).

Article PII precedes article PI, and it basically gives information about how the whole system of UAV, FPI-HSI, preprocessing, and analysis could work. The actual estimation results were not outstanding, but they provided encouragement to continue the studies in the second year. Figure 16 illustrates the accuracy of the estimation results. The biomass results are well in line with the results obtained by airplanes with other sensors. The nitrogen content results indicate a major advance for precision agriculture.

Validation of the features between different years and locations must be done before the results can be confirmed. Basically, the results indicate that hyperspectral imaging from UAV is possible and that the analysis methods used provide added value for precision agriculture.

The included article PIII strengthens the above conclusions. The article provides new knowledge about high-resolution, passive UAV remote sensing. The results show that all FPI technology-related processing steps (image preprocessing and spectral datacube generation) can be addressed in separate steps and that the rest of the processing can be carried out using regular photogrammetric and remote sensing software. The fact that images can be processed using regular

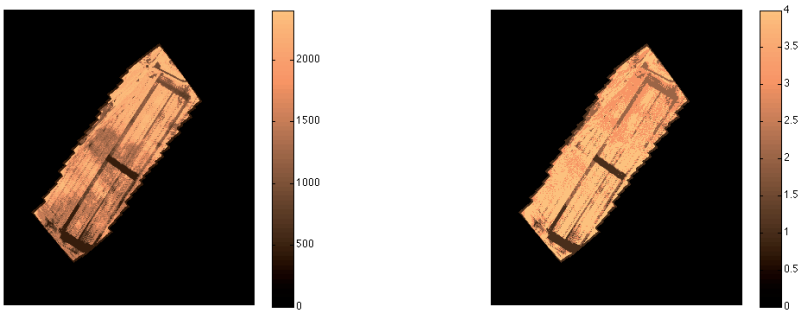

FIGURE 17 Estimation results of the whole field [(left: biomass (kg/ha), right: nitrogen (%)].These are originally published in Pölönen et al. (2013).

software is an important factor for users integrating an FPI spectral camera into their operational workflows. This was the first quatitative experiment with the FPI camera-type technology covering the entire remote sensing processing chain. The results confirmed the operability of the FPI camera in UAV remote sensing and the high potential of lightweight UAV remote sensing in general.

3.2 Dermatology

Actinic keratosis is a precursors for malignant non-melanoma skin cancer, squamous cell carcinoma. To prevent squamous cell carcinomas from developing it is crucial to treat the precursors early and effectively. Field cancerization denotes subclinical abnormalities in a tissue chronically exposed to UV radiation. These abnormalities can be found surrounding the clinically visible actinic keratoses. The aim of article PIV was to test the feasibility of a hyperspectral imaging system in the detection of multiple clinical and subclinical actinic keratosis for early treatment of the affected areas.

Altogether, 52 clinical actinic keratoses in 12 patients were included in this study. In 6 patients, digital photos were taken of the naive actinic keratosis, and again after a methylaminolevulinate (MAL)-fluorescence, diagnosis which was used to teach hyperspectral imaging system to find subclinical lesions. After 2-3 days, once the MAL had vanished, the hyperspectral images were taken. Biopsies were taken from clinical actinic keratosis, healthy-looking skin, and several suspected subclinical actinic keratosis. In the other six patients, digital and hyperspectral images were taken of the naive actinic keratosis followed by one biopsy per patient.

The hand-held FPI-HSI that was employed seemed to detect all, clinically visible actinic keratosis and numerous subclinical lesions. The histopathologies of the 33 biopsied lesions were concordant with the imaging results, showing either actinic keratosis or photo damage (Figure 18). Of the 28 histopathologically

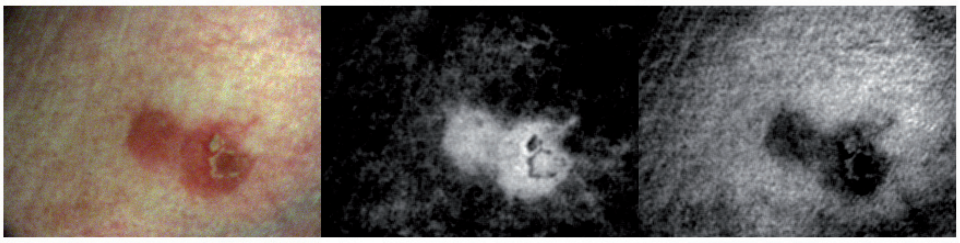

FIGURE 18 Left: False color composition of actinic keratosis. Middle: Abundance map of clinical and subclinical actinic keratosis. Right: Abundance map of healthy skin.

confirmed actinic keratoses, 16 were subclinical. A specific diffuse reflectance spectrum of an actinic keratosis and healthy skin was defined. The hyperspectral imaging system offers a new, non-invasive method for the early detection of field cancerization. Research in this area is ongoing with validation of these results and expansion of the research area to malignant tumors. For example, Zheludev et al. (2013) is concentrated on delineating malignant tumors with hyperspectral imaging and framelet processing of the signal.

3.3 Crime Scene Investigation

Detecting invisible details and separating mixed evidence is critical for forensic inspection. If this can be done reliably and fast at the crime scene, irrelevant objects do not require further examination in the laboratory. This will speed up the inspection process and release resources for other critical tasks. The included article PV reports on tests carried out at the University of Jyväskylä together with the Central Finland Police Department and the National Bureau of Investigation for detecting and separating forensic details with hyperspectral technology. Article itself does not contain information about mathematical tools and algorithms that were used. Used tools and algorithms are explained previously in chapter 2 and in section 2.7 there is an overview how these were used in this application.

In the tests, at an assumed violent burglary scene, evidence was sought with the help of VTT's hand-held FPI-HSI camera, Specim's VNIR camera, and Specim's SWIR camera. The tested details were dried blood on a ceramic plate, a stain of four types of mixed and absorbed blood, and blood that had been washed off a table. Other examined details included untreated latent fingerprints, gunshot residue, primer residue, and layered paint on small pieces of wood. All the cameras could detect visible details and separate mixed paint. From the SWIR data it was possible to separate four types of human and animal blood, which were mixed in the same stain and absorbed into a piece of fabric. None of the cameras, however, could detect primer residue, untreated latent fingerprints, or

blood that had been washed off. The results are encouraging and indicate the need for further studies. The results also emphasize the importance of creating optimal imaging conditions for crime scenes for each of subject and background.

3.4 Scientometrics

The author has with his co-authors utilized a knowledge discovery process from databases in the field of scientometrics. This research shows that the KDD process is a quite universal approach to perceive data analysis chains. Scientometrics itself is commonly defined as the quantitative study of science. Ivancheva (2008) provided a categorization of scientometrics methodology for research subjects, information types, and method classes. The subject of PVI can be seen as *science itself* because the authors in PVI tried to understand the structure of a field of science. The field is limited, focused, and concrete, so the information type of this research is *operational*. Finally, in the classification of Glänzel (2003), PVI positions itself in *structural scientometrics*, trying to map the research area.

The authors applied the knowledge discovery process to the mapping of current topics in a particular field of science. The main interest was in how articles form clusters and what the contents of the found clusters are. A framework involving web scraping, keyword extraction, dimensionality reduction, and clustering using the diffusion map algorithm were presented. Article PVI uses publicly available information about articles in high-impact journals. The presented method should be of use to practitioners or scientists who want to overview recent research in a field of science. As a case study, PVI maps the topics in data mining literature for 2011.

4 CONCLUSION

This thesis is the first research study that implements the whole knowledge discovery process from databases to hyperspectral imaging based on the Fabry-Perot interferometer. Both of the research questions introduced in Chapter 1 have been answered. As papers PI-PV show, there is no doubt that hyperspectral imaging based on the Fabry-Perot interferometer has usability in UAV, medicine, and crime scene investigation applications. The mathematical tools and computational algorithms behind the preprocessing and analysis are mostly known from the literature. This study implies that these tools are also applicable to the process chain of FPI-HSI data. The included articles show how FPI-HSI data can be preprocessed and analyzed.

Hyperspectral imagers based on the Fabry-Perot interferometer will become more common in the near future. Their production costs are much smaller than those of traditional pushbroom scanners. These FPI-HSI cameras are utilizable in applications where traditional hyperspectral imagers seem clumsy. In this sense, this research is valuable for anyone who wants to use or develop services based on FPI-HSI.

This research study also indicates some factors that require further development. The development needs, such as calibration, are related not only to FPI itself but also to completed application research. Basically, all the presented applications need reiterations and larger test series. For example, in the agriculture application, there is a need to repeat the test for several years and at different locations around the globe or at least in Europe.

The hyperspectral imager based on the Fabry-Perot interferometer also offers new opportunities for research. In the UAV application, a hyperspectral mosaic was created based on the nearest points directly below the UAV. This mosaic could also be based on the mean value of all the detected pixels in the same spatial location in different images. Then the points at the edges of the image could provide more information to the mosaic. At the edges, camera lens distortion overturns high objects such as trees so that the bole of the trees remains partly visible. The spectrum of the boles could be added to the hyperspectral mosaic, and their abundance could be unmixed with a proper unmixing model. In hy-

perspectral imaging of soft tissue such as human skin, more complicated models that take into account the absorption and transmittance of components in human skin could be used.

The results of this thesis provide new information for the utilization of hyperspectral imaging based on the Fabry-Perot interferometer. The nature of this research is highly multidisciplinary, and this is reflected in many places in this work. The research itself also points out that a novel sensor or a mathematical formula is not useful by itself. Society is affected when a novel technology and nice mathematics are tied with a real-life application. With the refinement of the information in this chain of technology, theory, and application, it is possible to improve our quality of life. The author's opinion is that great achievements in science arise, more and more, from multidisciplinary context.

YHTEENVETO (FINNISH SUMMARY)

Hyperspektrikuvantaminen on nopeasti kasvava monitieteinen ala. Tämä väitöskirja, jonka otsikko on "Tiedon havaitseminen uudenlaisen hyperspektrikameran sovelluksista", käsittelee nimensä mukaisesti Fabry-Perot-interferometriin pohjatuvan hyperspektrikameran sovelluksien tiedonhavaitsemisprosessia. Väitöskirjassa kerrotaan vaihe vaiheelta miten hyperspektrikuva-aineisto tulisi esikäsitellä ja analysoida. Tutkimuksessa uutta hyperspektrikameraa käytetään kolmessa eri sovelluskohteessa: pieniin miehittämättömiin lennokkeihin pohjautuvassa kaukokartoituksessa, lääketieteellisessä ihotautisovelluksessa ja rikospaikkatutkinnassa. Fabry-Perot interferometri pohjaiset spektrikamerat tulevat yleistymään lähivuosina. Väitöskirjassa käytetty matemaattiset työkalut ja algoritmit on sovitettu käsittelemään Fabry-Perot-interferometripohjaisella spektrikameralla tuotettua aineistoa. Tämä tutkimus osoittaa, että valitut työkalut soveltuvat hyvin tähän tehtävään.

REFERENCES

Averbuch, A., Zheludev, M. & Zheludev, V. 2012. Unmixing and target recognition in airborne hyper-spectral images. Earth Science Research 1 (2), 200-228.

Averbuch, A. & Zheludev, M. 2012. Two linear unmixing algorithms to recognize targets using supervised classification and orthogonal rotation in airborne hyperspectral images. Remote Sensing 4 (2), 532-560.

Bioucas-Dias, J., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P. & Chanussot, J. 2012. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (2), 354-379.

Bioucas-Dias, J., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. & Chanussot, J. 2013. Hyperspectral remote sensing data analysis and future challenges. Geoscience and Remote Sensing Magazine, IEEE 1 (2), 6-36.

Bowles, J., Palmadesso, P., Antoniades, J., Baumback, M. & Rickard, L. 1995. Use of filter vectors in hyperspectral data analysis. Proc. SPIE 2553, 148-157.

Camps-Valls, G. & Bruzzone, L. 2005. Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 43 (6), 1351-1362.

Chang, C. 2003. Hyperspectral Imaging: Techniques for Spectral Detection and Classification. Kluwer Academic/Plenum.

Chang, C. 2007. Hyperspectral Data Exploitation: Theory and Applications. Wiley.

Chen, J., Richard, C., Bermudez, J.-C. & Honeine, P. 2011. A modified non-negative LMS algorithm and its stochastic behavior analysis. ASILOMAR 11, 6-9 November.

Chung, F. R. K. 1997. Spectral Graph Theory. AMS Press.

Coifman, R. R. & Lafon, S. 2006a. Diffusion maps. Applied and Computational Harmonic Analysis 21 (1), 5–30.

Coifman, R. R. & Lafon, S. 2006b. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. Applied and Computational Harmonic Analysis 21 (1), 31 - 52.

Daughtry, C. S., Walthall, C. L., Kim, M. S., de Colstoun, E. B. & McMurtrey, J. E. 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of Environment 74 (2), 229-239.

Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R. & Schaepman, M. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation 9 (2), 165 - 193.

Everitt, B. S., Landau, S. & Leese, M. 2001. Cluster Analysis. Oxford University Press.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM 39 (11), 27–34.

Fowlkes, C., Belongie, S., Chung, F. & Malik, J. 2004. Spectral grouping using the Nyström method. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2), 214 -225.

Glänzel, W. 2003. Bibliometrics as a research field. (A course on theory and application of bibliometric indicators. Course Handouts.).

Goel, P., Prasher, S., Patel, R., Landry, J., Bonnell, R. & Viau, A. 2003. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. Computers and Electronics in Agriculture 39 (2), 67 - 93.

Goetz, A. F., Vane, G., Solomon, J. E. & Rock, B. N. 1985. Imaging spectrometry for earth remote sensing. Science 228 (4704), 1147-1153.

Hakala, T., Honkavaara, E., Saari, H., Mäkynen, J., Kaivosoja, J., Pesonen, L. & Pölönen, I. 2013. Spectral imaging from UAVs under varying illumination conditions. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2, 189–194.

Han, J. & Kamber, M. 2006. Data mining: Concepts and Techniques. Morgan Kaufmann.

Hand, D., Mannila, H. & Smyth, P. 2001. Principles of Data Mining. MIT Press. Adaptive computation and machine learning.

Hastie, T., Tibshirani, R. & Friedman, J. 2001. The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc. Springer Series in Statistics.

Honkavaara, E., Hakala, T., Kirjasniemi, J., Lindfors, A., Mäkynen, J., Nurminen, K., Ruokokoski, P., Saari, H. & Markelin, L. 2013. New light-weight stereosopic spectrometric airborne imaging technology for high-resolution environmental remote sensing - case studies in water quality mapping. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. SPRS Hannover Workshop 2013 XL-1/W1, 139-144.

Honkavaara, E., Hakala, T., Saari, H., Markelin, L., Mäkynen, J. & Rosnell, T. 2012a. A process for radiometric correction of UAV image blocks. Photogrammetrie, Fernerkundung, Geoinformation (PFG) 2012 (2), 115-127.

Honkavaara, E., Kaivosoja, J., Mäkynen, J., Pellikka, I., Pesonen, L., Saari, H., Salo, H., Hakala, T., Markelin, L. & Rosnell, T. 2012b. Hyperspectral reflectance signatures and point clouds for precision agriculture by light weight UAV imaging system. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I-7, 353–358.

Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., Mäkynen, J. & Pesonen, L. 2013. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. Remote Sensing 5 (10), 5006-5039.

Ivancheva, L. 2008. Scientometrics today: A methodological overview. In the fourth International Congerence on Webometrics, Informetrics, and Scientometrics & Ninth COLLNET Meeting.

Jones, H. & Vaughan, R. 2010. Remote Sensing of Vegetation: Principles, Techniques, and Applications. Oxford University Press.

Kaivosoja, J., Pesonen, L., Kleemola, J., Pölönen, I., Salo, H., Honkavaara, E., Saari, H., Mäkynen, J. & Rajala, A. 2013. A case study of a precision fertilizer application task generation for wheat based on classified hyperspectral data from UAV combined with farm history data. Proc. SPIE 8887.

Keshava, N. & Mustard, J. 2002. Spectral unmixing. IEEE Signal Processing Mag. 19 (1), 44-57.

Kuula, J., Pölönen, I. & Puupponen, H.-H. 2012. Using hyperspectral imaging for detecting destructive subjects and materials. Defence Forces Technical Research Centre Publications: 8th symposium on CBRNE threats - How does society cope? 26, 181-184.

Lillesand, T., Kiefer, R. & Chipman, J. 2004. Remote Sensing and Image Interpretation. Wiley.

Manolakis, D. G. 2002. Overview of algorithms for hyperspectral target detection: theory and practice. Proc. SPIE 4725, 202-215.

McKeown, D.M., J., Cochran, S., Ford, S., McGlone, J., Shufelt, J. & Yocum, D. 1999. Fusion of hydice hyperspectral data with panchromatic imagery for cartographic feature extraction. Geoscience and Remote Sensing, IEEE Transactions on 37 (3), 1261-1277.

Mountrakis, G., Im, J. & Ogole, C. 2011. Support vector machines in remote sensing: A review. {ISPRS} Journal of Photogrammetry and Remote Sensing 66 (3), 247 - 259.

Mäkynen, J., Holmlund, C., Saari, H., Ojala, K. & Antila, T. 2011. Unmanned aerial vehicle (uav) operated megapixel spectral camera. Proc. SPIE 8186.

Mäkynen, J., Holmlund, C., Saari, H., Ojala, K. & Antila, T. 2012. Multi- and hyperspectral UAV imaging system for forest and agriculture applications. Proc. SPIE 8374.

Nadler, B., Lafon, S., Coifman, R. & Kevrekidis, I. G. 2008. Diffusion maps – a probabilistic interpretation for spectral embedding and clustering algorithms. In Principal Manifolds for Data Visualization and Dimension Reduction, Vol. 58. Springer. Lecture Notes in Computational Science and Engineering, 238–260.

Nascimento, J. & Dias, J. 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 34 (4), 898-910.

Neittaanmäki-Perttu, N., Grönroos, M., Tani, T., Pölönen, I., Ranki, A., Saksela, O. & Snellman, E. 2013. Detecting field cancerization using a hyperspectral imaging system. Lasers in Surgery and Medicine 45 (7), 410–417.

Newton, I. 1704. Opticks: or, A treatise of the reflections, refractions, inflections & of light. Royal Society.

Nieminen, P., Pölönen, I. & Sipola, T. 2013. Research literature clustering using diffusion maps. Journal of Informetrics 7, 874–886.

Pesses, M. 1999. A least-squares-filter vector hybrid approach to hyperspectral subpixel demixing. IEEE Transactions on Geoscience and Remote Sensing 37 (2), 846-849.

Plaza, A., Plaza, J., Paz, A. & Sanchez, S. 2011. Parallel hyperspectral image and signal processing [applications corner]. Signal Processing Magazine, IEEE 28 (3), 119-126.

Pölönen, I., Saari, H., Kaivosoja, J., Honkavaara, E. & Pesonen, L. 2013. Hyperspectral imaging based biomass and nitrogen content estimations from lightweight UAV. Proc. SPIE 8887.

Pölönen, I., Salo, H., Saari, H., Kaivosoja, J., Pesonen, L. & Honkavaara, E. 2012. Biomass estimator for nir image with a few additional spectral band images taken from light UAS. Proc. SPIE 8369.

Read, J., Bifet, A., Pfahringer, B. & Holmes, G. 2012. Batch-incremental versus instance-incremental learning in dynamic and evolving data. Advances in Intelligent Data Analysis XI, Lecture Notes in Computer Science 7619, 313-323.

Rosnell, T. & Honkavaara, E. 2012. Point cloud generation from aerial image data acquired by a quadrocopter type micro unmanned aerial vehicle and a digital still camera. Sensors 12, 453-480.

Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351 1, 309-317.

Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20 (1), 53 - 65.

Saari, H., Pölönen, I., Salo, H., Honkavaara, E., Hakala, T., Holmlund, C., Mäkynen, J., Mannila, R., Antila, T. & Akujärvi, A. 2013. Miniaturized hyperspectral imager calibration and UAV flight campaigns. Proc. SPIE 8889.

Salo, H., Tirronen, V., Pölönen, I., Tuominen, S., Balazs, A., Heikkilä, J. & Saari, H. 2012. Methods for estimating forest stem volumes by tree species using digital surface model and CIR images taken from light UAS. Proc. SPIE 8390.

Segl, K., Roessner, S., Heiden, U. & Kaufmann, H. 2003. Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing 58 (1-2), 99-112.

Specim 2013. Specim Ltd's webpage. ⟨URL:http://www.specim.fi⟩.

Thenkabail, P., Lyon, J. & Huete, A. 2012. Hyperspectral Remote Sensing of Vegetation. CRC Press/Taylor & Francis.

Tsagaris, V., Anastassopoulos, V. & Lampropoulos, G. 2005. Fusion of hyperspectral data using segmented PCT for color representation and classification. Geoscience and Remote Sensing, IEEE Transactions on 43 (10), 2365-2375.

Winter, M. E. 1999. N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data. Proc. SPIE 3753, 266-275.

Zheludev, V., Pölönen, I., Averbuch, A., Neittaanmäki-Perttu, N., Grönroos, M., Neittaanmäki, P. & Saari, H. 2013. Hyperspectral imaging based delineation of malignant tumours. Reports of the Department of Mathematical Information Technology, Series B, Scientific Computing. 8/2013. University of Jyväskylä.