Techniques and Applications of Hyperspectral Image Analysis

Hans F. Grahn and Paul Geladi

BICENTENNIAL 1807 WILEY 2007

John Wiley & Sons, Ltd

Techniques and Applications of Hyperspectral Image Analysis

Techniques and Applications of Hyperspectral Image Analysis

Hans F. Grahn and Paul Geladi

1807 WILEY 2007
BICENTENNIAL

John Wiley & Sons, Ltd

Copyright © 2007 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging in Publication Data

Techniques and applications of hyperspectral image analysis / [edited by] Hans Grahn and Paul Geladi.
p. cm.
Includes bibliographical references.
ISBN 978-0-470-01086-0 (cloth)
1. Image processing—Statistical methods. 2. Multivariate analysis. 3. Multispectral photography.
I. Grahn, Hans. II. Geladi, Paul.
TA1637.T42 2007
621.36′7—dc22 2007021097

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-470-01086-0

Typeset in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by TJ International, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

This book is about multivariate and hyperspectral imaging, not only on how to make the images but on how to clean, transform, analyze and present them. The emphasis is on visualization of images, models and statistical diagnostics, but some useful numbers and equations are given where needed. The idea to write this book originated at an Image Analysis Session at the Eastern Analytical Symposium (Somerset, NJ) in November 2002. At this session, the lectures were so inspiring that it was felt necessary to have something on paper for those not present.

An earlier book, also published by John Wiley & Sons, Ltd, came out in 1996. It was called *Multivariate Image Analysis* by Geladi and Grahn. This book contains a lot of the basic theory. The examples in this book are not very advanced because in the early 1990s it was not so easy to get 10 or more wavelength bands in an image. There has also been an evolution in theory and algorithms, requiring additions to the 1996 book, but a major difference is that image files are much larger in size and sizes are expected to keep on growing. This is a challenge to the data analysis methods and algorithms, but it is also an opportunity to get more detailed results with higher precision and accuracy.

It would have been possible to make a revised second edition of Geladi and Grahn, but it was considered more useful to include extra authors, thus creating a multi-authored book with chapters on hyperspectral imaging. The chapters would be written by groups or persons whom we felt would be able to contribute something meaningful. The book can roughly be divided into two parts. The earlier chapters are about definitions, nomenclature and data analytical and visualization aspects. The later chapters present examples from different fields of science, including extra data analytical aspects. The subdivision in theory and application parts is not ideal. Many attempts of putting the chapters in the correct order were tried and the final result is only one of them.

Chapter 1 is about the definition of multivariate and hyperspectral images and introduces nomenclature. It contains basic information that is applicable to all subsequent chapters. The basic ideas of multivariate interactive image analysis are explained with a simple color (three channels) photograph.

Chapters 2–5 give a good insight into factor and component modeling used on the spectral information in the images. This is called multivariate image analysis (MIA). Chapter 2 introduces interactive exploration of multivariate images in the scene and variable space in more detail using an eight channel optical image taken from an airplane. The role of visualization in this work is extremely important; something that Chapter 2 succeeds in highlighting. Chapter 3 gives a good overview of classification in optical images of agricultural products. The special topics of fuzzy clustering and clustering aided by spatial information are explained. In Chapter 4, the SIMPLISMA technique and its use on images are explained. This technique is an important alternative to those explained in Chapters 2 and 3. SIMPLISMA is not just exploratory, but tries to find deterministic pure component spectra by using spectroscopic constraints on the model. The examples are from Fourier transform infrared (FTIR) and time-of-flight - secondary ion mass spectrometry (TOF-SIMS) imaging. Chapter 5 is about even more factor analysis methods that can be applied to hyperspectral images. The special case of unsymmetrical noise distributions is emphasized.

Chapters 6–9 introduce the concepts and models for regression modeling on hyperspectral images: multivariate image regression (MIR). Chapter 6 is about regression on image data. This is the situation where the spectrum in each pixel is able to predict the value of an external variable, be it another image, an average property or something in between like localized information. Emphasis is also given on cleaning and preprocessing a hyperspectral image to make the spectral information suitable for regression model building. Chapter 7 takes up the important aspect of validation in classification and regression on images. The example of Chapter 2 is reused by defining one of the channels as a dependent variable. Also, a new example for the calibration of fat content in sausages is introduced. The advantage of image data is that many pixels (=spectra) are available, making testing on subsets a much easier task. Chapter 8 describes classical, extended and general least squares models for Raman images of aspirin/polyethylene mixtures. The theory part is extensive. Chapter 9 is about the need for expressing hyperspectral data in the proper SI and IUPAC units and about standards for multivariate and hyperspectral imaging. In particular, diffuse reflection, the most practical technique of imaging used in the laboratory, is in need of such standardization. Without the standards, reproducible image data would not be available and spectral model building and interpretation would be hampered severely.

The applied chapters do not give a complete overview of all possible applications, but they give a reasonable catalog of things that can be done with hyperspectral images using different types of variables. Chapter 10 is about multivariate movies in different variables, mainly optical, infrared, Raman and nuclear magnetic resonance (NMR). Multivariate movies represent huge amounts of data and efficient data reduction is needed. The applications are in polymer and pharmaceutical tablet dissolution. Chapter 11 describes the DECRA technique as it can be used on phantoms and brain images in magnetic resonance imaging. Chapter 12 gives an overview of agricultural and biological applications of optical multivariate and hyperspectral imaging. Chapter 13 is about brain studies using positron emission tomography (PET). The PET images are extremely noisy and require special care. Chapter 14 is about chemical imaging using near infrared spectroscopy. Pharmaceutical granulate mixtures are the examples used.

When writing a book one should always have students in mind. Books are ideal as course material and there is not much material available yet for learning about nonremote sensing hyperspectral imaging. Recommendations for newcomers are to read Chapters 1–9 together with Geladi and Grahn (Geladi and Grahn, 1996) in order to get the basics. Chapters 2–5 form the factor analysis block and Chapters 6–9 form the regression/calibration block. More advanced readers may review the basics quickly and plunge directly into the applied chapters (10–15). An alternative choice of reading would be Bhargava and Levin (Bhargava and Levin, 2005). There are also some interesting books from the remote sensing field (Chang, 2003; Varhsney and Arora, 2004).

REFERENCES

Bhargava, R. and Levin, I. (Eds) (2005) *Spectrochemical Analysis Using Infrared Multichannel Detectors*, Blackwell, Oxford.

Chang, C. (2003) *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Kluwer Academic Publishers, Dordrecht.

Geladi, P. and Grahn, H. (1996) *Multivariate Image Analysis*, John Wiley & Sons, Ltd, Chichester.

Varshney, K. and Arora, M. (Eds) (2004) *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer, Berlin.

List of Contributors

Brian Antalek, Eastman Kodak Company, Research Laboratories B82, Rochester, NY 14650-2132, USA

Vincent Baeten, Département Qualité des Produits agricoles, Centre wallon de recherches agronomiques CRA-W, 24, Chaussée de Namur, B-5030 Gembloux, Belgium

Mats Bergström, Novartis Pharma AG, CH-4002, Basel, Switzerland

James E. Burger, Burger Metrics, Applied Hyperspectral Imaging for Automation and Research, Ladehammerveien 36, 7041 Trondheim, Norway

Pierre Dardenne, Département Qualité des Produits agricoles, Centre wallon de recherches agronomiques CRA-W, 24, Chaussée de Namur, B-5030 Gembloux, Belgium

Jane Dubois, Malvern Instruments, Analytical Imaging Systems, 3416 Olandwood Court #210, Olney, MD 20832, USA

Kim H. Esbensen, ACABS, Aalborg University Esbjerg (AAUE), Niels Bohrs Vej 8, DK-6700 Esbjerg, Denmark

Neal B. Gallagher, Eigenvector Research, Inc., 160 Gobblers Knob Lane, Manson, WA 98831, USA

Paul L. M. Geladi, NIRCE, The Unit of Biomass Technology and Chemistry SLU Röbäcksdalen, PO Box 4097, SE 90403 Umeå, Sweden

Hans F. Grahn, Division of Behavioral Neuroscience, Department of Neuroscience, Karolinska Institutet, S-17177, Stockholm, Sweden

Kenneth S. Haber, Malvern Instruments, Analytical Imaging Systems, 3416 Olandwood Court #210, Olney, MD 20832, USA

Joseph P. Hornak, Magnetic Resonance Laboratory, 54 Lomb Memorial Drive, Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623-5604, USA

Sergei G. Kazarian, Department of Chemical Engineering, ACE Building 208A/210, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Michael R. Keenan, Sandia National Laboratories, Albuquerque, NM 87185-0886, USA

Linda H. Kidder, Malvern Instruments, Analytical Imaging Systems, 3416 Olandwood Court #210, Olney, MD 20832, USA

E. Neil Lewis, Malvern Instruments, Analytical Imaging Systems, 3416 Olandwood Court #210, Olney, MD 20832, USA

Sharon Markel, Eastman Kodak Company, Rochester, NY 14650-2132, USA

Jacco C. Noordam, TNO Defence, Security and Safety, PO Box 96864, 2509 JG The Hague, The Netherlands

Juan Antonio Fernández Pierna, Département Qualité des Produits agricoles, Centre wallon de recherches agronomiques CRA-W, 24, Chaussée de Namur, B-5030 Gembloux, Belgium

Pasha Razifar, Computerized Image Analysis & PET Uppsala Applied Science Lab (UASL) GEMS PET Systems AB, Husbyborg, 752 28 Uppsala, Sweden

P. M. Thompson, Eastman Kodak Company, Rochester, NY 14650-2132, USA

Thorbjørn Tønnesen Lied, Kongsberg Maritime AS, R&D, Hydrography & Hydroacustic Division, Horten, Norway

Willie H. A. M. van den Broek, TNO Defence, Security and Safety, PO Box 96864, 2509 JG The Hague, The Netherlands

Jaap van der Weerd, Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Willem Windig, Eigenvector Research, Inc., 6 Olympia Drive, Rochester, NY 14615, USA

List of Abbreviations

2-D	Two-dimensional
3-D	Three-dimensional
ABES	Agricultural, biology and environmental sciences
ACD	Annihilation coincidence detection
AIS	Airborne imaging system
ALS	Alternating least squares
ANN	Artificial neural network
ARMA	Autoregressive moving average
AOTF	Acousto-optic tunable filter
API	Active pharmaceutical ingredient
ATR	Attenuated total reflection
AVHRR	Advanced very high resolution radiometer
AVIRIS	Airborne visible and infrared imaging spectrometer
CAS	Chemical Abstracts Service
CASI	Compact airborne spectrographic imager
CCD	Charge coupled device
CLS	Classical least squares
cFCM	Conditional Fuzzy C-Means
CSF	Cerebrospinal fluid
csiFCM	Cluster size insensitive Fuzzy C-Means
DECRA	Direct exponential curve resolution analysis
EDS	Energy dispersive spectrometer (mainly X-rays)
EMSC	Extended multiplicative scatter correction
ELS	Extended least squares
FA	Factor analysis
fALS	Factored alternating least squares
FCM	Fuzzy C-Means
FEMOS	Feedback Multivariate Model Selection

fNNMF	Factored non-negative matrix factorization
FBP	Filtered back projection
FOV	Field of view
FPA	Focal plane array
FT	Fourier transform
FTIR	Fourier transform infrared
FWHM	Full width at half maximum
GIS	Geographical information system
GLS	Generalized least squares
GOME	Global ozone monitoring experiment
GPS	Global positioning system
GRAM	Generalized rank annihilation method
HD	High density (for polymers)
HIA	Hyperspectral image analysis
HPLC	High pressure liquid chromatography
HPMC	Hydroxypropylmethylcellulose
ICV	Image cross-validation
ILS	Inverse least squares
ILS	Iterative partial least squares
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
IWLS	Iteratively weighted least squares
LCTF	Liquid crystal tunable filter
LD	Low density (for polymers)
LDA	Linear discriminant analysis
LOR	Line of response
LUT	Look up table
MBM	Meat and bone meal
MCR	Multivariate curve resolution
MEIS	Multispectral electro-optical imaging spectrometer
MI	Multivariate image
MIA	Multivariate image analysis
MIR	Multivariate image regression
MIR	Mid infrared
MLPCA	Maximum likelihood principal component analysis
MNF	Minimum noise fraction
MRI	Magnetic resonance imaging
MSC	Multiplicative scatter correction
MVWPCA	Masked volume wise principal component analysis
NIR	Near infrared
NIRCI	Near infrared chemical imaging

NMR	Nuclear magnetic resonance
NNMF	Non-negative matrix factorization
NSLS	National Synchrotron Light Source
OSEM	Ordered subset expectation maximization
PAL	Phase alternating line
PAA	Polyacrylamide
PAMS	Poly (α-methyl styrene)
PAS	Photo-acoustic spectroscopy
PBMA	Poly (butyl methacrylate)
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PE	Polyethylene
PET	Poly (ethylene terephthalate)
PET	Positron emission tomography
PETT	Positron emisison transaxial tomography
PFA	Principal factor analysis
PGP	Prism–grating–prism
PGSE	Pulse gradient spin echo
PLS	Partial least squares
PMSC	Piecewise multiplicative scatter correction
PNNMF	Poisson non-negative matrix factorization
PVA	Poly (vinyl alcohol)
PVC	Poly (vinyl chloride)
PW	Pixel wise
PRESS	Prediction residual error sum of squares
QA	Quality assurance
QC	Quality control
RGB	Red green blue
RMSEC	Root mean square error of calibration
RMSECV	Root mean square error of cross-validation
RMSEP	Root mean square error of prediction
ROI	Region of interest
SECAM	Séquentielle couleur a mémoire
SECV	Standard error of cross-validation
SEM	Scanning electron microscope
SIA	Self-modeling image analysis
sgFCM	Spatially guided Fuzzy C-Means
SI	Système international
SLDA	Stepwise linear discriminant analysis
SIMCA	Soft independent modeling of class analogy

SIMPLISMA	Simple-to-use interactive self-modeling mixture analysis
SIMS	Secondary ion mass spectrometry
SNR	Signal to noise ratio
SNV	Standard normal variate
SPOT	Satellite pour l'observation du terre
SRM	Standard reference material
SS	Sum of squares
SSE	Sum of squared errors
SVD	Singular value decomposition
SVM	Support Vector Machine
SWPCA	Slice wise principal component analysis
TAC	Time–activity curve
TOF-SIMS	Time-of-flight-secondary ion mass spectrometry
TOMS	Total ozone mapping spectrometer
USP	United States Pharmacopeia
UV	Ultraviolet
VWPCA	Volume wise principal component analysis
WPCA	Weighted principal component analysis
WTFA	Window target factor analysis

1

Multivariate Images, Hyperspectral Imaging: Background and Equipment

Paul L. M. Geladi, Hans F. Grahn and James E. Burger

1.1 INTRODUCTION

This chapter introduces the concepts of digital image, multivariate image and hyperspectral image and gives an overview of some of the image generation techniques for producing multivariate and hyperspectral images. The emphasis is on imaging in the laboratory or hospital on a scale going from macroscopic to microscopic. Images describing very large scenes are not mentioned. Therefore, the specialized research fields of satellite and airborne imaging and also astronomy are left out. A color image is used to introduce the multivariate interactive visualization principles that will play a major role in further chapters.

1.2 DIGITAL IMAGES, MULTIVARIATE IMAGES AND HYPERSPECTRAL IMAGES

All scientific activity aims at gathering information and turning this information into conclusions, decisions or new questions. The information may be qualitative, but is often and preferably quantitative. This

Techniques and Applications of Hyperspectral Image Analysis Edited by H. F. Grahn and P. Geladi

means that the information is a number or a set of numbers. Sometimes even a large set of numbers is not enough and an image is needed. Images have the dual property of both being large datasets and visually interpretable entities.

Freehand drawing and photography have been used extensively in the sciences to convey information that would be too complicated to be expressed in a text or in a few numbers. From the middle of the 1900s the TV camera and electronic image digitization have become available and images can be saved in digital format as files (Geladi and Grahn, 2000). A digital image is an array of I rows and J columns made of I × J greyvalues or intensities, also called pixels. A pixel is a greyvalue with an associated coordinate in the image. The image is also a data matrix of size I × J with the greyvalues as entries. (Pratt, 1978; Rosenfeld and Kak, 1982; Gonzalez and Woods, 1992; Schotton, 1993) For three-dimensional images, the array has I rows, J columns and H depth slices. The pixel becomes a voxel. For color imaging in TV, video and on computer screens, three images are needed to contain the red, green and blue information needed to give the illusion of color to the human eye (Callet, 1998; Johnson and Fairchild, 2004) For photocopying and printing the three primary colors are yellow, cyan and magenta. One may say that the pixels (or voxels) are not greyvalues anymore, but triplets of numbers (Figure 1.1).

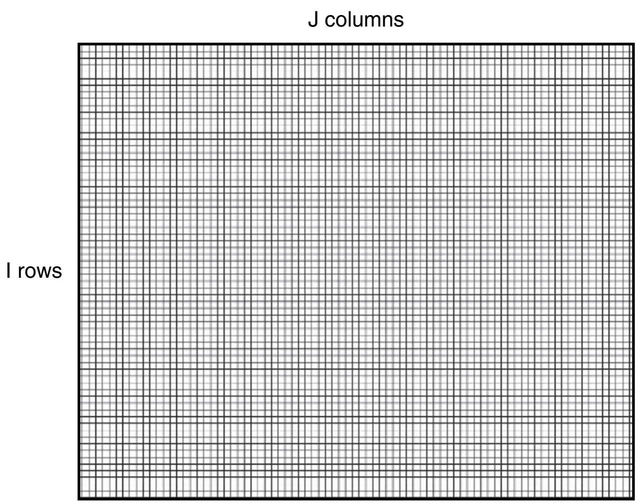

Figure 1.1 A digital image is an array of I rows and J columns. Each coordinate pair has a greyvalue and the small grey square (or rectangle) is called a pixel. For color images, the pixel becomes a red, green and blue triplet instead of a greyvalue

Because pixels are digitized greyvalues or intensities, they may be expressed as integers. Simple images may have a greyvalue range of 2^8 meaning that 0 is the blackest black and 255 is the whitest white. In more advanced systems, 2^{12} grey levels (0–4095), 2^{14} or 2^{16} greylevels are used. Some systems average images over a number of scans. In such a case, greylevels may have decimals and have to be expressed as double precision numbers.

The fact that a digitized image is a data matrix makes it easy to do calculations on it. The result of the calculations can be a number, a vector of numbers or a modified image. Some simple examples would be counting of particles (image to number), the calculation of intensity histograms (image to vector), image smoothing and edge enhancement (image to modified image). There are many books describing how this is done (Pratt, 1978; Rosenfeld and Kak, 1982; Low, 1991; Gonzalez and Woods, 1992).

Color images have three layers (or bands) that each have different information. It is possible to make even more layers by using smaller wavelength bands, say 20 nm wide between 400 nm and 800 nm. Then each pixel would be a spectrum of 21 wavelength bands. This is the multivariate image. The 21 wavelength bands in the example are called the image variables and in general there are K variables. An $I \times J$ image in K variables would form a three-way array of size $I \times J \times K$. (Figures 1.2–1.4).

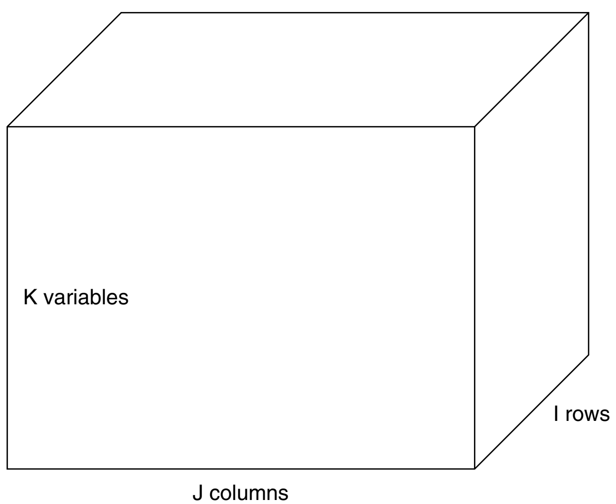

Figure 1.2 An $I \times J$ image in K variables is an $I \times J \times K$ array of data

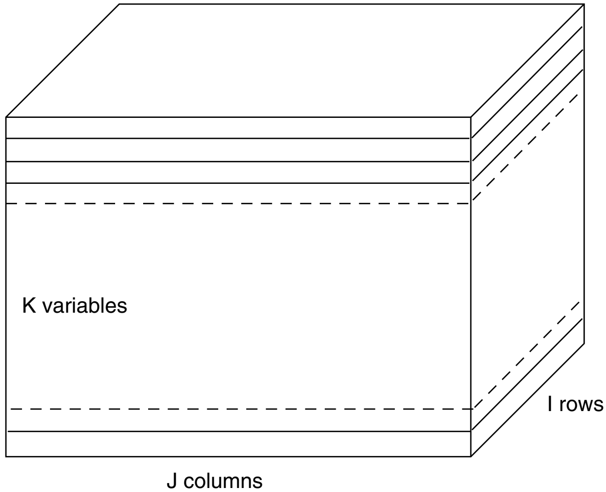

Figure 1.3 The I × J × K image can be presented as K slices where each slice is a greyvalue image

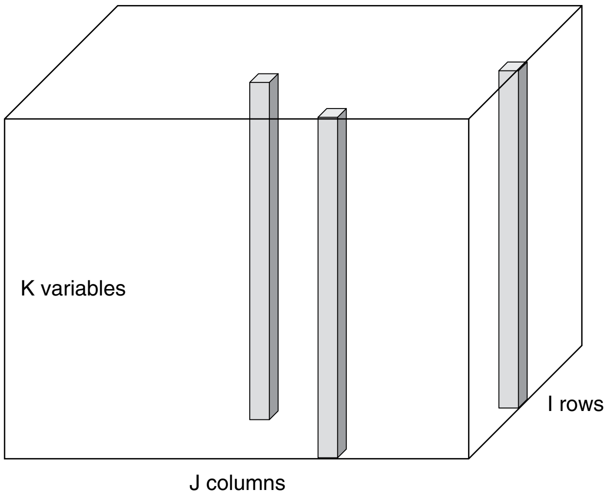

Figure 1.4 The I × J × K image can be presented as an image of vectors. In special cases. the vectors can be shown and interpreted as spectra

The human eye only needs the three wavelength bands red, green and blue in order to see color. With more than three wavelength bands, simple color representation is not possible, but some artificial color images may be made by combining any three bands. In that case the colors are not real and are called pseudocolors. This technique makes no sense when more than three bands are combined because of the limitations of the human visual system.

Many imaging techniques make it possible to make multivariate images and their number is constantly growing. Also, the number of variables available is constantly growing. From about 100 variables upwards the name hyperspectral images was coined in the field of satellite and airborne imaging (Vane, 1988; Goetz and Curtiss, 1996), but hyperspectral imaging is also available in laboratories and hospitals. The following sections will introduce some multivariate and hyperspectral images and the physical variables used to make them with literature references.

Images as in Figures 1.2–1.4 with $K = 2$ or more are multivariate images. Hyperspectral images are those where each pixel forms an almost continuous spectrum. Multivariate images can also be mixed mode, e.g. $K = 3$ for an UV wavelength image, a near infrared (NIR) image and a polarization image in white light. In this case, the vector of three variables is not really a spectrum.

So what characterizes hyperspectral images? Two things:

- many wavelength or other variable bands, often more than 100;
- the possibility to express a pixel as a spectrum with spectral interpretation, spectral transformation, spectral data analysis, etc.

1.3 HYPERSPECTRAL IMAGE GENERATION

1.3.1 Introduction

Many principles from physics can be used to generate multivariate and hyperspectral images (Geladi and Grahn, 1996, 2000). Examples of making NIR optical images are used to illustrate some general principles.

A classical spectrophotometer consists of a light source, a monochromator or filter system to disperse the light into wavelength bands, a sample presentation unit and a detection system including both a detector and digitization/storage hardware and software (Siesler *et al.*, 2002). The most common sources for broad spectral NIR radiation are tungsten halogen or xenon gas plasma lamps. Light emitting diodes and tunable lasers may also be used for illumination with less broad wavelength bands. In this case, more diodes or more laser are needed to cover the whole NIR spectral range (780–2500 nm). For broad spectral sources, selection of wavelength bands can be based on specific bandpass filters based on simple interference filters, liquid crystal tunable filters (LCTFs), or acousto-optic tunable filters (AOTFs), or the spectral energy may be dispersed by a grating device or a prism–grating–prism (PGP) filter. Scanning interferometers can also be used to acquire NIR spectra from a single spot.

A spectrometer camera designed for hyperspectral imaging has the hardware components listed above for acquisition of spectral information plus additional hardware necessary for the acquisition of spatial information. The spatial information comes from measurement directly through the spectrometer optics or by controlled positioning of the sample, or by a combination of both. Three basic camera configurations are used based on the type of spatial information acquired; they are called point scan, line scan or plane scan.

1.3.2 Point Scanning Imaging

The point scanning camera configuration shown in Figure 1.5 can be used to measure a spectrum on a small spot. The sample is then repositioned before obtaining a new spectrum. By moving the sample

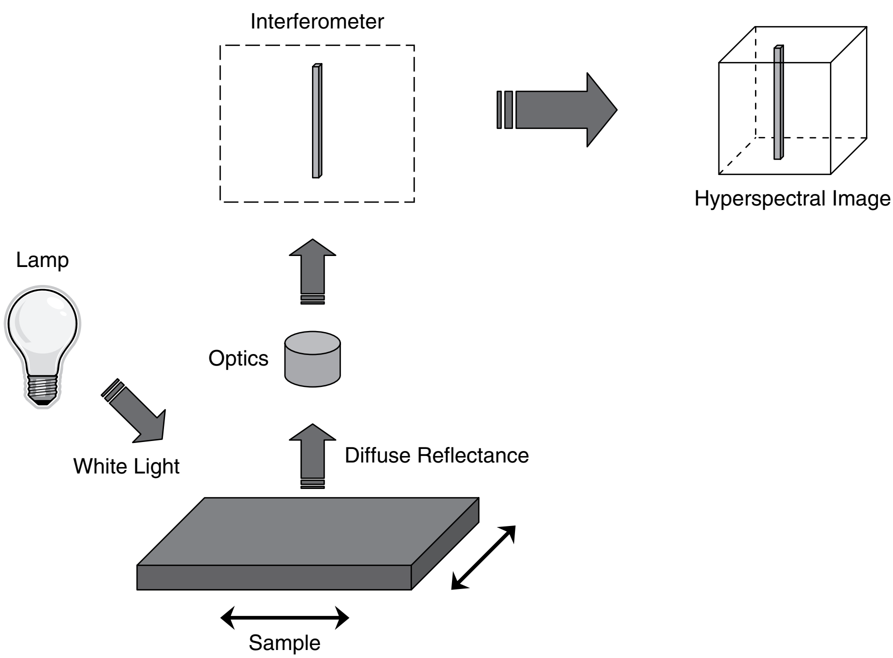

Figure 1.5 A scanning set-up measures a complete spectrum in many variables at a single small spot. An image is created by systematically scanning across the surface in two spatial dimensions