



HAL
open science

The Color Constancy Problem in Multispectral Remote Sensing - On the Impact of Surface Orientation on Spectral Signatures

Rafael Wiemker

► **To cite this version:**

Rafael Wiemker. The Color Constancy Problem in Multispectral Remote Sensing - On the Impact of Surface Orientation on Spectral Signatures. Human-Computer Interaction [cs.HC]. Universität Hamburg, 1997. English. NNT: . tel-00010773

HAL Id: tel-00010773

<https://theses.hal.science/tel-00010773>

Submitted on 26 Oct 2005

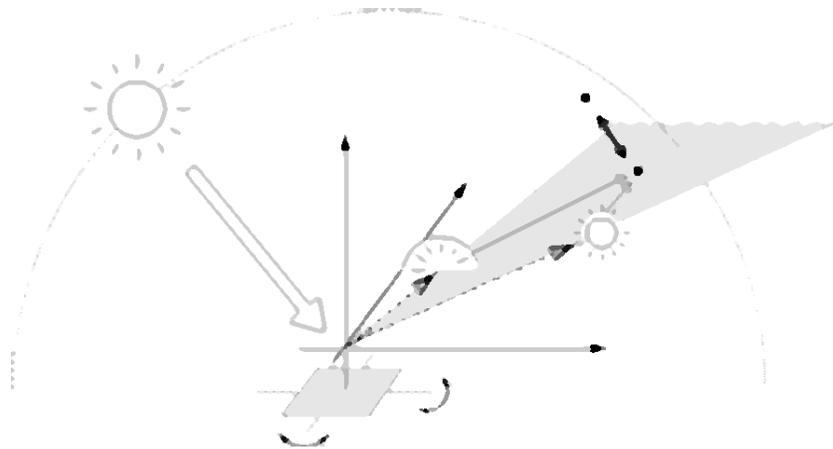
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Color Constancy Problem in Multispectral Remote Sensing

On the Impact of Surface Orientation on Spectral Signatures

Rafael Wiemker



Dissertation zur Erlangung des Doktorgrades
des Fachbereichs Physik der Universität Hamburg

Gutachterin / Gutachter der Dissertation: Prof. Dr. Hartwig Spitzer
Prof. Dr. Leonie Dreschler-Fischer (FB Informatik)
Prof. Dr. Bernd Jähne (IWR, Univ. Heidelberg)

Gutachter der Disputation: Prof. Dr. Hartwig Spitzer
Prof. Dr. Erich Lohrmann

Datum der Disputation: 17. Dezember 1997

Sprecher des Fachbereichs Physik und
Vorsitzender des Promotionsausschusses: Prof. Dr. Bernhard Kramer

Abstract

We suggest a novel approach to the surface orientation related color constancy problem for multispectral imagery. The basic problem addressed in this thesis is just how the observed spectral signature of a Lambertian target surface varies when its surface orientation changes.

Our approach is based on a dichromatic illumination model which we have verified by several thousands *in situ* measured spectra of a dozen samples of different surface materials. The two principal components of daylight illumination are direct sun light and diffuse sky light, which show distinctively different spectral characteristics. The observed spectrum of a given surface with a specific Lambertian reflectance spectrum varies with its surface orientation, since each orientation leads to different contributions of direct solar and indirect diffuse illumination. The ambiguity about the actual illumination of a surface, causing uncertainty about its spectral reflectance, has been recognized as the color constancy problem. In multispectral remote sensing, this leads to erroneous classification and segmentation as well as spurious results of change detection.

We introduce a transformation which is invariant against surface orientation. The suggested invariant is a linear mapping in the logarithmic feature space and filters out all spectral information which can possibly stem from an illumination change rather than from the reflectance of a given surface. Instead of recovering the reflectance signal, the suggested mapping produces a new only surface reflectance-dependent descriptor which is invariant against varying illumination. Sole input is the relative direct to diffuse illumination spectrum. No assumptions about the possible reflectance spectra are made. Error propagation through the transform is well understood. The mapping is a purely pixel-based, one-pass matrix operation and can preprocess multispectral images in order to segment them into regions of homogeneous reflectance, unperturbed by varying illumination conditions.

Apart from simulated and *in situ* measured data, the suggested transform has been successfully applied to experimental multispectral imagery. The quantitative results and example clippings from the imagery show significant improvements in the multispectral classification of target surfaces under varying surface orientation. Although the transformed data may not completely supersede the original spectral data, the suggested transformation is shown to be a powerful early processing step, allowing subsequent orientation invariant classification, edge detection and segmentation.

Das Farbkonstanzproblem in der multispektralen Fernerkundung

Über den Einfluß der Oberflächenneigung auf spektrale Signaturen

Rafael Wiemker

Kurzfassung

Die multispektrale Fernerkundung liefert Bilder der Erdoberfläche in verschiedenen Wellenlängenbereichen (typisch $N = 5 - 100$ Kanäle), so daß jedem Bildelement ein Spektrum zugeordnet werden kann. Diese *spektrale Signatur* dient der automatisierten Klassifikation und Änderungsdetektion. Die vorliegende Arbeit untersucht den Einfluß, den die Neigung einer Oberfläche auf ihre spektrale Signatur ausübt.

Ein Großteil der multispektralen Auswertungstechniken geht von *Lambert'schen* Oberflächen aus. Das sind solche, deren spektrale Signaturen nicht von der Beobachtungsrichtung des Sensors abhängen. Durch Spektrometermessungen an ausgewählte Oberflächen zeigt diese Arbeit, daß auch Lambert'sch reflektierende Flächen durch Änderung ihrer Oberflächenorientierung eine deutliche Änderung ihrer spektralen Signatur erfahren. Diese Änderung betrifft sowohl die Helligkeit (Höhe) als auch die Farbe (Verlauf) des Spektrums. Diese Orientierungsabhängigkeit führt dann in der Auswertung von fernerkundeten Multispektralbildern zu fehlerhaften Ergebnissen.

In der Analyse der Spektrometermessungen kann gezeigt werden, daß die Variabilität der spektralen Signaturen durch zwei Hauptkomponenten dargestellt werden kann, die auf die Reflexion von direktem Sonnenlicht einerseits und indirektem Himmelslicht andererseits zurückgeführt werden können. Die Anteile dieser beiden Komponenten können im allgemeinen nicht bestimmt werden. Diese unbekanntes spektrale Zusammensetzung der Beleuchtung ist als *Farbkonstanzproblem* bekannt.

Mit Hinblick auf die automatisierte Bildauswertung wird hier eine mathematische Abbildung entwickelt, die sich den *logarithmischen Spektralraum* zunutze macht, um ein beliebiges beobachtetes Spektrum auf einen spektralen Deskriptor zu projizieren, der invariant gegen den spektralen Einfluß der Oberflächenneigung sowie Eigen- und Schlagschatten ist. Diese *invariante spektrale Signatur* kann durch eine rechenzeitgünstige pixelweise Matrixmultiplikation errechnet werden ($\mathcal{O}(N^2)$). Die Vorteile der Invarianten werden anhand von simulierten Spektren und echten fernerkundeten Multispektralbildern gezeigt. Die Empfindlichkeit der Projektion gegenüber zufälligen und systematischen Fehlern wird im Rahmen der Fehlerfortpflanzung ausführlich untersucht.

Acknowledgements

I would like to thank my advisors Hartwig Spitzer (II. Institute for Experimental Physics, University of Hamburg) and Johann Bienlein (Desy, Hamburg) for the chance to join the remote sensing group of CENSIS (Center for Science and International Security, University of Hamburg). The CENSIS group provided an interdisciplinary and unconventional environment and a good atmosphere. I would like to thank my coworkers Ramon Franck, Thomas Hepp, Thomas Kollwe, Daniel Kulbach, Gerhard Meister, Borris Neumann, André Rothkirch and Anja Speck for help, inspiring discussions, and being nice guys, and particularly Martin Kollwe for encouragement. I appreciate our nice corridor companionship with the people of KOGS (Kognitive Systeme, Computer Science Department, University of Hamburg) and LKI (Labor für künstliche Intelligenz).

Very much I enjoyed the cooperation with Leonie Dreschler-Fischer (Computer Science Department, University of Hamburg) and the OSCAR group (Open Skies for Conventional Arms Reduction). Christian Drewniok, Harald Lange, Carsten Schröder and Rainer Sprengel introduced me to the niceties of the subject as well as to the pitfalls of the machinery.

Work on this thesis could not even have been started without the quick support I experienced from the staff of KOGS: Ingeborg Heer-Mück, Dieter Jessen, Jörn Tellkamp, and Ursula Wechsung ('Mission Control Center').

Fortunately, I had the opportunity to use the OVID spectrometer, and enjoyed the guidance of Barbara Bartsch and Margareta Betancor (Meteorological Institute, University of Hamburg).

The DAEDALUS image flights were conducted in collaboration with the DLR (German Aerospace Research Establishment, Oberpfaffenhofen), particularly with the help of Volker Amann, Peter Hausknecht and Rudolf Richter.

The Cusanus-Werk, Bonn, sponsored a workshop in the Mathematical Institute Oberwolfach which turned out to be very inspiring for this thesis.

Jiri Matas gave a very useful hint during our discussion in Prague.

Also, I would like to thank Ulf Borgeest and Dieter Reimers who encouraged me to start right away and go ahead.

My work at CENSIS was made possible by graduate scholarships by the University of Hamburg and the Volkswagen-Foundation, Hannover, whom I thank for their uncomplicated support.

*... bringt Licht in unsere Dunkelheit,
und dieses Welt- und Himmelslicht
weicht hunderttausend Sonnen nicht.*

(Kaspar Friedrich Nachtenhöfer, Kirchenlied, 1684)

The device also functioned as an ordinary calculator, but only to a limited degree. It could handle any calculation which returned an answer of anything up to 4. $1 + 1$ it could manage (2) and $1 + 2$ (3) and $2 + 2$ (4) or $\tan 74$ (3.487414), but anything above 4 it represented merely as "A Suffusion of Yellow." It was not certain if this was a programming error or an insight beyond his ability to fathom [...]

(Douglas Adams, *The Long Dark Tea-Time of the Soul*, London 1988)

Contents

1	Introduction	1
1.1	Introduction and Motivation	1
1.2	Color Constancy – Nature and Scope of the Problem	3
1.3	Brief Review of Previous Approaches to the Color Constancy Problem	7
1.4	Structure of this Thesis	11
2	Basic Concepts of Multispectral Imagery and its Classification	13
2.1	The Spectral Signal Formation Process	13
2.2	Multispectral Imagery, Spectral Vectors and the Feature Space	19
2.3	Spectral Distances and Classification in the Feature Space	23
2.4	Clusters and Separability Measures	25
2.5	Supervised Classification	29
2.6	Unsupervised Classification / Clustering	29
2.7	Principal Component Transformation	32
3	Color Constant Classification in the Multispectral Feature Space	35
3.1	Dichromatic Illumination Model	35
3.2	Planar Spectral Classes	39
3.3	Color Constant Enhancement of the Mahalanobis Distance	41
3.4	Some Derivatives of the Concept of Class Planes	44
4	Illumination Invariance in the Logarithmic Feature Space	47
4.1	The Logarithmic Feature Space	48
4.1.1	Cluster Separability in the Logarithmic Space	48
4.2	Brightness Constancy	51
4.2.1	Brightness Normalization	51
4.2.2	Logarithmic Brightness Filtering	52
4.3	The Dichromatic Illumination Model in the Logarithmic Space	59
4.4	The Spectral Filtering	60
4.4.1	Some Properties of the Algebraic Projector \mathbf{P}	63
4.4.2	Relation to Oppenheim’s Homomorphic Filtering	68
4.4.3	Correspondence to the Mahalanobis Distance	69

4.4.4	Processing Costs and Some Programming Considerations	70
4.4.5	Processing of the Spectra After Prior Filtering	71
4.5	Results on Simulated Data	72
4.5.1	Simulated Noisefree Data	72
4.5.2	Simulated Data With Random Noise	73
4.6	Analysis of Random Noise	80
4.6.1	Propagation of Multiplicative Noise	81
4.6.2	Propagation of Additive Noise	82
4.6.3	Expected Relative Accuracy of the Transformed Signals	82
4.7	Analysis of Systematic Errors	83
4.7.1	Propagation of Systematic Multiplicative Errors	84
4.7.2	Propagation of Systematic Additive Errors	85
4.8	Expected Performance of the Transform	85
4.8.1	Performance for a Given Pair of Reflectance Spectra	85
4.8.2	Performance for a Cluster of Different Reflectance Spectra	87
4.9	Decision on Reflectance vs. Illumination Change for a Given Pair of Spectra	89
4.9.1	Error Weighted Spectral Distance	89
4.9.2	Results on Simulated Data	90
4.10	Relaxation of the Lambertian Assumption	93
5	Experimental Findings on <i>in situ</i> Measured Spectral Data	95
5.1	The Sensor	95
5.2	Experimental Setup of the <i>in situ</i> Measurements	96
5.3	Some Examples in the Wavelength- and Feature-Space-Representation	98
5.4	Evaluating the Goodness of Invariance: Residual Variance after Filtering	105
5.5	The Impact of Surface Orientation: Magnitude of Brightness and Color Shifts	109
5.6	Principal Component Analysis of the Logarithmic Spectra	111
6	Experimental Findings on Multispectral Imagery	115
6.1	The Sensor	116
6.2	Calibration of the Image Data and Atmospheric Correction	117
6.3	Analysis of Selected Spectral Samples	122
6.3.1	<i>In scene</i> Example of the Color Shift	122
6.3.2	Comparison of Spectral Sample Pairs in Various Feature Spaces	124
6.3.3	Discussion of Results	129
6.3.4	Bitemporal Comparison of Spectral Samples	130
6.4	Evaluation by Multispectral Image Classification	135
6.4.1	Qualitative Results of the Classification	139
6.5	Unsupervised Classification / Clustering	140
6.6	Analysis of the Transform in the Feature Space	144
6.7	Statistical Evaluation of Spectral Sample Sets	148
6.7.1	Evaluation of Spectral Distances	149
6.7.2	Comparison in Different Feature Spaces	151
6.7.3	Discussion of Results	155
6.7.4	Comparison with respect to the Filter Rank	156
6.8	Sensor Specific Noise Analysis	158

6.8.1	Quantization Error	160
6.8.2	Magnitude of Color Shift and Estimation of Required Dynamic Range	161
6.8.3	Error Weighted Spectral Distances of the Spectral Sample Pairs	165
7	Conclusions	167
7.1	How Important is the Color Constancy Problem for Multispectral Remote Sensing?	167
7.2	Which Improvements have been Achieved with this Work?	168
7.3	Original versus Logarithmic Feature Space	171
7.4	Lessons Learned	172
7.5	Outlook	173
	Appendix	177
A	Analysis of the Spectral Variability due to Surface Orientation	177
A.1	The Pseudo-Inverse	177
A.2	Orthogonal Projections	179
A.3	Retrieval of the Reflectance Spectrum from the Class Plane	180
A.4	Minimal Distance Between Two Class Planes	182
A.5	Retrieval of the Path Radiance Spectrum from Two Class Planes	183
B	Spectra from <i>in situ</i> Measurements	185
B.1	Illumination Variability Eigenvectors	187
B.2	Eigenvectors of Residual Illumination Variability	199
C	Multispectral Imagery	203
C.1	Sample Pairs from Scene ‘Gewerbegebiet 1992’	204
C.2	Statistical Evaluation of Samples from Scene ‘Gewerbegebiet 1992’	214
C.3	Sample Pairs from Scene ‘Ziegelstein 1991’	217
C.4	Multispectral Classification of Samples from Scene ‘Ziegelstein 1991’	236
C.5	Bitemporal Sample Pairs from Scene ‘Gewerbegebiet’	242
	Bibliography	253

List of Figures

1.1	Effect of surface orientation	5
1.2	Approaches to the color constancy problem	10
2.1	Sensor target viewing geometry	15
2.2	Lambert reflector	16
2.3	Radiation paths: sun – target – sensor	18
2.4	Spectral vectors and signatures	20
2.5	Multispectral image cube	21
2.6	The multispectral feature space	22
2.7	Class ellipsoid in the feature space	24
2.8	Two cluster in the feature space	28
2.9	Clusters with orthogonal and parallel variance	28
3.1	Diffuse illumination	37
3.2	MODTRAN simulation	38
3.3	Class plane spanned by sun and sky light	41
3.4	Intersection of the class planes	45
4.1	Distance in original versus log space	50
4.2	Typical brightness scaled clusters in original and log space	58
4.3	Illustration of the filtering process.	62
4.4	Simulated relative illumination spectra	72
4.5	Mapping of simulated spectra	75
4.6	Mapping of simulated noisy spectra	76
4.7	Transformed feature space	77
4.8	Transformed covariance matrices	79
4.9	Transform efficiency w.r.t. noise	87
5.1	Experimental setup of the <i>in situ</i> measurements	97
5.2	Observed spectra	99
5.3	Illumination eigenvectors	102
5.4	Spectral clusters in the feature space	104
5.5	Illumination eigenvectors	112
5.6	PC Analysis of the residual variance	114

6.1	Scene ‘Ziegelstein 1991’	119
6.2	Scene ‘Gewerbegebiet 1992’	120
6.3	Scene ‘Gewerbegebiet 1994’	121
6.4	Clip from roof top	123
6.5	<i>In scene</i> measured color shift	123
6.6	Spectral sample pair from scene ‘Gewerbegebiet 1992’	125
6.7	Spectral sample pair from scene ‘Gewerbegebiet 1992’ (cont’d)	126
6.8	Spectral sample pair from scene ‘Ziegelstein 1991’	127
6.9	Spectral sample pair from scene ‘Gewerbegebiet 1992’ (cont’d)	128
6.10	Bitemporal spectral sample pair (scene ‘Gewerbegebiet’ 1992 vs. 1994)	132
6.11	Bitemporal spectral sample pair (cont’d)	133
6.12	Bitemporal spectral sample pair (cont’d)	134
6.13	Multispectral classification of samples	137
6.14	Multispectral classification of samples (cont’d)	138
6.15	Unsupervised multispectral classification	142
6.16	Unsupervised multispectral classification (cont’d)	143
6.17	Effect of filtering in the feature space	146
6.18	Effect of color filtering in the feature space	147
6.19	Spectral distance relative to the cluster orientation	150
6.20	Selected samples from scene ‘Ziegelstein 1991’	153
6.21	Surface plots of the projector matrices \mathbf{U} and \mathbf{V}	159
6.22	Relative quantization error	161
6.23	Color shift in shadow	162
6.24	Color shift in shadow, tilted surface	163
7.1	Alternative loss function	174
A.1	Illustration of an orthogonal projection in \mathbb{R}^3	179
B.1	Illumination variability eigenvectors of <i>in situ</i> spectra	187
B.2	Principal component analysis of the residual variance	199
C.1	Spectral sample pairs from scene ‘Gewerbegebiet 1992’	204
C.2	Set of selected spectral sample pairs (scene ‘Gewerbegebiet 1992’)	214
C.3	Spectral sample pairs from scene ‘Ziegelstein 1991’	217
C.4	Multispectral classification of sample pairs (scene ‘Ziegelstein 1991’)	236
C.5	Bitemporal spectral sample pairs (scene ‘Gewerbegebiet’ 1992 vs. 1994)	242

List of Tables

4.1	Mutual cluster separations	78
4.2	Error weighted spectral distance	92
5.1	Cluster separability	104
5.2	Variance reduction by filtering wrt. filtering rank K	107
5.3	Variance reduction by filtering wrt. atmospheric parameter γ	108
5.4	Magnitudes of brightness and color shifts.	110
6.1	Spectral bands of the DAEDALUS AADS 1268 line scanner	117
6.2	Cluster separation in different feature spaces	145
6.3	Investigated feature spaces.	149
6.4	Spectral distances wrt. feature space (scene ‘Ziegelstein 1991’)	154
6.5	Spectral distances wrt. filter rank (scene ‘Ziegelstein 1991’)	157
6.6	Assumed diffuse to global illumination ratio $\hat{\mathbf{m}}$	158
6.7	Magnitude of brightness and color shift	164
6.8	Error weighted spectral distance ratios	166
7.1	Additive and multiplicative mechanisms	172
C.1	Spectral distances wrt. feature space (scene ‘Gewerbegebiet 1992’)	215
C.2	Spectral distances wrt. filter rank (scene ‘Gewerbegebiet 1992’)	216

List of Symbols and Abbreviations

Physical Magnitudes:

λ	wavelength [$\mu\text{m}, \text{nm}$]
r	Lambert reflectance factor [dimensionless]
E	incident irradiance [$\text{W}/(\text{m}^2 \mu\text{m})$]
E^r	reflected irradiance [$\text{W}/(\text{m}^2 \mu\text{m})$]
E_{diff}	diffuse irradiance [$\text{W}/(\text{m}^2 \mu\text{m})$]
E_{glob}	global irradiance [$\text{W}/(\text{m}^2 \mu\text{m})$]
L	radiance [$\text{W}/(\text{m}^2 \text{sr} \mu\text{m})$]
L^r	radiance reflected from point source [$\text{W}/(\text{m}^2 \text{sr} \mu\text{m})$]
L^s	radiance measured at sensor [$\text{W}/(\text{m}^2 \text{sr} \mu\text{m})$]
L_o	path radiance [$\text{W}/(\text{m}^2 \text{sr} \mu\text{m})$]
T_i	transmittance in spectral band i [dimensionless]

Scalars:

γ	exponential constant for relative diffuse illumination
ϵ	small noise level
μ	overall brightness factor
μ	contribution of diffuse illumination
ν	contribution of direct illumination
$\eta = \nu/\mu$	relative contribution of direct illumination
ϕ	azimuth angle
$\sigma^2(x_i)$	variance, mean square scatter of x_i
θ	nadir / zenith angle
$\xi = x'/x''$	ratio of cluster centers
a	spectral class index
$d_i = x_i - x'_i$	signed spectral distance in band i
d_I	spectral distance normalized by Σ^{-1}
d_T	spectral distance normalized by $\text{tr} \Sigma$
d^2	squared spectral distance

d_E^2	squared Euclidean distance
d_M^2	squared Mahalanobis distance
d_w^2	squared error weighted distance
e	illumination component, or Euler number
$f(d_i)$	loss function
i	spectral band index
j	number of measurements
k	number of classes
m	diffuse illumination
\hat{m}	relative diffuse to global illumination
n	direct illumination
\hat{n}	relative direct to global illumination
r	reflectance
r^*	pseudo-reflectance
x	dimensionless value of observed pseudo-radiance
$s(x) = \sigma(x)/\sqrt{j}$	estimated error of x
t	invariant descriptor: transformed of x
$H = \text{rank } \mathbf{T}$	rank of transformation matrix \mathbf{T}
$J = \sqrt{\text{tr } \bar{\Sigma}}$	cluster radius, root mean squared errors
$K = N - \text{rank } \mathbf{P}$	number of filtered components
N	number of spectral bands
Vectors:		
$\lambda = (\lambda_1 \dots \lambda_N)^T$	wavelength vector
\mathbf{a}^T	the transposed of a column vector \mathbf{a}
\mathbf{d}	spectral distance vector
\mathbf{e}	eigenvector, or illumination vector
\mathbf{m}_a	center of spectral class a (mean vector)
\mathbf{m}	vector of diffuse illumination
$\hat{\mathbf{m}}$	vector of relative diffuse illumination
\mathbf{n}	vector of direct illumination
$\hat{\mathbf{n}}$	vector of relative direct illumination
\mathbf{r}	reflectance vector
$\hat{\mathbf{r}} = \mathbf{r}/\ \mathbf{r}\ $	normalized reflectance vector
$\mathbf{t} = (t_1 \dots t_N)^T$	transformed feature vector
$\ln \mathbf{t} = (\ln t_1 \dots \ln t_N)^T$	logarithmic transformed feature vector
$\mathbf{u} = \mathbf{1} = (1 \dots 1)^T$	unity vector
$\hat{\mathbf{u}} = \mathbf{u}/\ \mathbf{u}\ $	normalized unity vector, brightness shift
$\mathbf{v} = \mathbf{U}\mathbf{m}$	orthogonalized diffuse vector
$\hat{\mathbf{v}} = \mathbf{v}/\ \mathbf{v}\ $	ortho-normalized diffuse vector, color shift
$\mathbf{x} = (x_1 \dots x_N)^T$	feature vector
$\ln \mathbf{x} = (\ln x_1 \dots \ln x_N)^T$	logarithmic feature vector
$\hat{\mathbf{x}} = \mathbf{x}/\ \mathbf{x}\ $	normalized spectral vector

Matrices:

$\Sigma(\mathbf{x})$	covariance matrix of vectors \mathbf{x}
Σ_a	covariance matrix/ellipsoid of spectral class a
\mathbf{A}^T	the transposed of a matrix \mathbf{A}
\mathbf{A}^{-1}	the inverse of a matrix \mathbf{A}
\mathbf{A}^+	the pseudo-inverse of a matrix \mathbf{A}
\mathbf{A}	plane spanning matrix of class a
\mathbf{D}	diagonal matrix
\mathbf{E}	illumination matrix, or eigenvector matrix
\mathbf{I}	identity matrix
$\mathbf{P} = \mathbf{UV}$	color constant projector matrix in log space
\mathbf{Q}	powers of $\hat{\mathbf{n}}$ projector matrix in log space
\mathbf{R}	color constant projector matrix in orig space
\mathbf{S}	error covariance matrix of vectors \mathbf{x}
\mathbf{T}	principal component transformation matrix
$\mathbf{U} = \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T$	brightness shift filter
$\mathbf{V} = \mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^T$	color shift filter
\mathbf{X}	set of vectors \mathbf{x} , matrix with column vectors

Abbreviations:

ld	<i>logarithmus dualis</i> to base 2
ln	<i>logarithmus naturalis</i> to base e
log	logarithmic
rk	rank of matrix
rms	root mean square (deviation/error)
tr	trace of matrix
wrt.	with respect to
BIL	line-wise band-interleaved stored image data
BIP	pixel-wise band-interleaved stored image data
BRDF	bidirectional reflectance distribution function
BSQ	band sequential stored image data
CCD	charge coupled device
CET	Central European Time
CIE	Commission Internationale de l'Eclairage
DC	digital counts (sensor raw data)
DEM	digital elevation model
Eq.	equation
Fig.	figure
FA	factor analysis
IFOV	instantaneous field of view

MED	minimum Euclidean distance classification
ML	maximum likelihood classification
MMD	minimum Mahalanobis distance classification
MNF	maximum noise fraction transformation
MSE	mean squared error
OVID	optical visible and infra-red detector
PCA	principal component analysis
PCT	principal component transformation
PDF	probability density function
RGB	red-green-blue
RTC	radiative transfer code
SNR	signal to noise ratio
SSE	sum of squared error
SVD	singular value decomposition
VLSI	very large scale integration

Introduction

1.1 Introduction and Motivation

Multispectral remote sensing is a relative young field. Based on advances of solid state physics, opto-electronic sensors became available in the 1960ies, which produce digital image data in a substantial range of the electromagnetic spectrum (Rees 1990). State-of-the-art detectors take image data in selected wavelength bands from the visible blue light ($\lambda = 0.4 \mu\text{m}$) to the thermal infrared radiation ($\lambda = 13 \mu\text{m}$).¹ The resulting wealth of spectral image data yields much enhanced information about ground scenes, as compared to panchromatic (black and white) photography.

The processing and analysis of remotely sensed multispectral images² therefore developed into an important technique for a multitude of fields. It supports research and monitoring in virtually all geosciences, as well as in cadastre, land use management and urban planning. It is widely applied in military reconnaissance, and more recently its benefits are investigated for international confidence building by cooperative verification of arms control within the framework of the Open Skies Treaty (Ryan et al. 1996, Wiemker & Spitzer 1996, Spitzer 1997a, Spitzer 1997b, Spitzer et al. 1997).

Use of remotely sensed multispectral imagery, as taken from aircraft or satellite, is commonly made when the measurement of certain variables is required to cover a larger area, from hectares to thousands of square kilometers. Then comprehensive *in situ* measurements are too costly. As a cross reference, ground sample points are selected and some *in situ* measurements are conducted, synchronously if possible. This *ground reference* (or *ground truth*) is used for calibration of the remotely sensed image data. Hence the variables in question can be derived for the whole covered area. Obviously this line of argument only holds when other external parameters such as atmospheric

¹ Image data in the wavelength range $\lambda = 1 - 100 \text{ cm}$ ($0.3 - 30 \text{ GHz}$) can be obtained through the synthetic aperture radar method (SAR). Radar, however, is *active* remote sensing, which produces the necessary electromagnetic illumination itself. In contrast, the usual mode of multispectral remote sensing is *passive*, with either the necessary illumination provided by the sun as the primary and the atmosphere as the secondary source, or the thermal radiation of the target itself. Another available technique is also passive microwave remote sensing in the wavelength range $3 - 60 \text{ mm}$ ($5 - 100 \text{ GHz}$).

² An introduction into the concept of multispectral imagery is given in Chapter 2.

transmittance, surface orientation, ground humidity etc. do *not* change over the image scene.

There are two basic approaches to analyze image data. One is to search for salient statistical features within the data, to perform clustering and classification based on statistical models. Only in the last step are the statistically found classes identified with conceptual classes and linked to the ground reference.

The other way is to start from physical concepts of the observed objects, and to model the complete image formation process: The incident radiation being reflected by the object and transmitted to the sensor, and finally mapped to digital gray values. If this physical model of the image formation process is invertible, then the parameters of the ground object can be determined directly from the observations. Otherwise, the method of *inverse modeling* can be applied. It varies the physical parameters of the ground object and repeats the simulation of the image formation until a best fit between simulated results and the actual observations is reached.

Physical modeling has been important to the processing of remotely sensed multispectral images for the last two or three decades, and gained growing attention for computer vision more recently. Most commonly, a hybrid approach between statistical and physical theories is chosen, which connects some physical modeling with results of image statistics.

A large subfield of multispectral image analysis deals with the production of pixel-wise classified land cover maps, for mapping, planning or verification purposes in public, commercial and military applications. The underlying notion for classification is that of a *spectral signature*, *i.e.*, a specific reflectance spectrum which is typical for each class of surface cover. While the spectral reflectance signatures are assumed to be invariant, the spectral illumination and thus the recorded spectral radiances are certainly not so. The apparent change of reflectance where only the illumination has changed, has been labeled the *color constancy problem* in physiology and later in computer vision.

The color constancy problem has several aspects and is far from being solved. Previous work was mainly concerned with color constancy in the visible wavelength range ($0.4 - 0.7 \mu\text{m}$). In this thesis, we expand the scope by including also the region of the near and mid infrared radiation ($0.7 - 2.5 \mu\text{m}$). We discuss the color constancy problem with respect to the surface orientation of the reflecting objects, and the implications for the analysis of remotely sensed multispectral images.

Multispectral image analysis can be seen as a natural generalization of RGB color image analysis, and classification as a subdivision of the general *image segmentation* problem. In the conclusion of their comprehensive survey on recent results in color image segmentation, Skarbek & Koschan (1994) state that the main problem for color based methods is the need of color constancy. They further observe that physics based approaches, which try to overcome the problem of misclassifying specularly reflections as well as shadowed and shaded areas, are far from operational as to now.

A number of researchers have measured the spectral properties of outdoor illumination and its variability with respect to time, so that change of illumination effects can be assessed. A closer look reveals, however, that for remotely sensed images it is not sufficient to assume a global illumination which is valid for all imaged surface patches. We rather need to allow for a local illumination which can change rapidly from pixel to pixel due to changes in surface orientation. For a complete grid of surface orientation,

the local illumination can be modeled individually from surface orientation, horizon line and assumed sky intensity distribution. Such an orientation map with sufficient spatial resolution, however, will not be available in general.

In this thesis, we choose a different approach. First, we establish the spectral variability of a *Lambertian reflector* with respect to surface orientation.³ Secondly, instead of costly individual modeling of local illumination from geometrical considerations, we then simply filter this variability in order to produce a spectral descriptor which is invariant against the surface orientation induced change of illumination. Thus, we do not make any assumptions about the orientation of the surface patches or the sky's spectral distribution, but recover an invariant which can be used for further image processing. The search for invariants as robust features in image processing has since long been a primary goal in computer vision.

For analysis of the color constancy problem with respect to surface orientation, we introduce a logarithmic feature space which reduces the multiplicative nature of the physical reflection process into an additive one. The analysis and the invariant spectral descriptor are checked against two sets of experimental data: *in situ* measured spectra with high spectral and dynamical resolution, and remotely sensed multispectral imagery from an airborne scanner.

1.2 Color Constancy – Nature and Scope of the Problem

The color constancy problem arises whenever one wants to distinguish and classify spectra received by a multispectral sensor and does not have sufficient knowledge about the respective illuminating irradiance onto the reflecting surface patch. The significant reflectance spectrum can only be retrieved from the measured radiance spectrum if the proper illuminating spectral irradiance is known.

Lambertian reflection is a wide spread assumption in remote sensing image processing. It assumes that the spectral reflectance is specific to the reflecting surface and independent of the angle under which the sensor views the surface. In contrast, without the Lambertian assumption, also the source-reflector-sensor geometry of the recording setup has to be considered. Then we cannot measure a unique reflectance, but rather a single point of the more general bidirectional reflectance distribution function (BRDF, Nicodemus (1970), Kriebel et al. (1975)).⁴

This thesis does not consider BRDF effects, but shows that changes in the orientation of the reflecting surface causes brightness and color shifts in the measured radiance spectrum even for Lambertian reflectors.

³ Per definition, the spectrum of a fixed, Lambertian reflecting surface appears equal under all observation angles. However, if its surface orientation changes, the observed spectrum changes as well, because the illumination incident onto the surface depends on the geometry of the various light sources relative to the reflecting surface.

⁴ The BRDF (in units of $[\text{sterad}^{-1}]$) is a spectrally dependent function of the incident and scattered directions. Limiting cases are the totally diffuse *Lambertian scatterer* on the one hand, and the *specular scatterer* which reflects into the opposite direction of the incidence angle only. Even without specular reflection the BRDF effects can be as large as 100% for rough surfaces such as vegetation canopies (*hot spot effect*).

It is well known that the Lambertian reflection model works only poorly for many surface materials. Therefore, in computer vision many approaches have been made to refine the reflection models beyond Lambert's law (Oren & Nayar 1994, Oren & Nayar 1995, Meister et al. 1996, Wolff 1996). However, there are convincing reasons why analysis in this thesis starts out from Lambert's law:

- ▶ Lambertian reflection still is the most wide spread assumption in remote sensing image processing for land applications.
- ▶ Often the reflection is modeled as a superposition of specular reflection and Lambertian (totally diffuse) reflection, and our considerations are thus valid for a major component of such composite reflection models.
- ▶ Specular reflection is usually not of high interest for remote sensing image processing. Image pixels with specular reflection mostly exhibit values which are above the dynamic range of today's sensor detectors and are thus inapt for classification. Moreover, the number of pixels showing specular reflection is usually low in remotely sensed imagery, since the aircraft flight paths are planned such as to avoid just this effect.
- ▶ A number of BRDF-models have been investigated for use with remotely sensed imagery. However, as of today it remains unclear if there exists a general BRDF-model which could be used for a wide range of surface materials while not being too costly to, firstly, measure from samples, and secondly, to employ in multispectral image classification.

The color constancy problem as a complex contains quite different aspects:

- **Change of global illumination:** The illumination spectrum varies during the course of the day and year, with geographical position, cloud coverage, atmospheric composition and so on for outdoor images, and with the illuminating light sources for indoor applications of computer vision.
- **Change of local illumination:** *e.g.* due to
 - **surface orientation** (see the sketch in Fig. 1.1 on the facing page: apparent brightness relations can even be reversed in sign),
 - **shadowing** by other objects which block out the sun and part of the sky (individual horizon line),
 - **spatially varying atmospheric transmittance** (*e.g.* cloud coverage of certain areas),
 - **mutual illumination**, when the surface patch is not only illuminated by sun and sky but moreover by other neighboring reflecting surfaces,
 - and for indoor applications: **close, finite light sources**, *i.e.* non-parallel incident radiation.

For remote sensing image processing, the change of global illumination is less of a problem than for common computer vision applications. Remote sensing recording flights

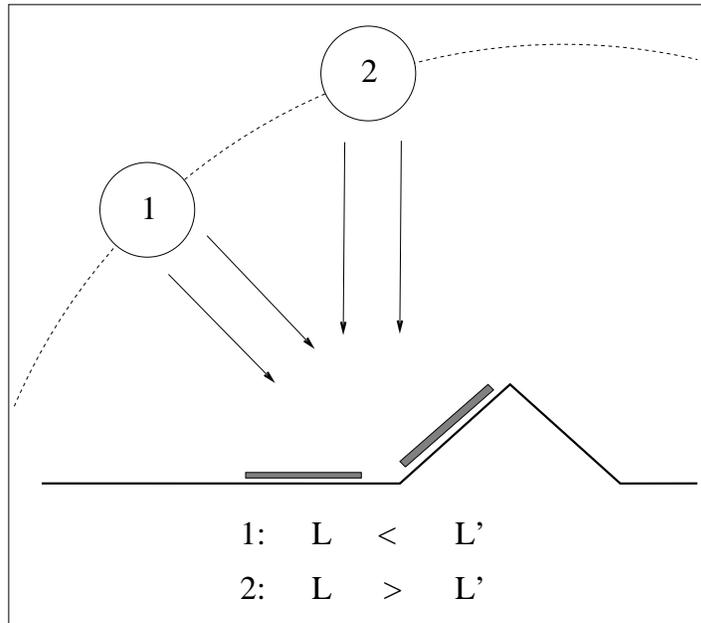


Figure 1.1: An example of the effect of surface orientation on the radiances reflected from surface patches of equal reflectance but different inclination: For sun position (1) the tilted surface seems brighter than the horizontal one, for sun position (2) vice versa.

take place at specified times and locations, and the particular illumination conditions can be assessed by radiation transfer codes (RTCs) and standard reflectance targets. Often meteorological data is available, and the sensor is radiometrically calibrated. The change of local illumination, however, remains an open problem.

In this thesis we concentrate on the impact of surface orientation on the Lambertian spectral signature. Note that we are not concerned with a BRDF dependent change of reflection properties, but rather with an apparent change of the received spectrum which is due to the fact that a different surface orientation causes a different illumination of the given (assumed Lambertian) surface patch.

Let us clarify the difference between BRDF-effects and surface orientation related color inconstancy: We consider a target surface of fixed orientation at a given sun position. When the sensor moves and the observed spectral signature changes, then these changes do necessarily stem from BRDF-effects because the illumination of the target surface has not changed. In contrast, when the sensor positions stays fixed, and the target surface's orientation changes, any change in the observed spectrum can stem from both the changed illumination (due to the new position of the surface relative to the sun and sky) as well as from BRDF-effects of the new geometry, such as *e.g.* specular reflection. The same is true for a fixed surface and a fixed sensor position with a varying sun position: the illumination will have changed because of the new sun position (Fig. 1.1) and BRDF-effects can occur as well.

A wide spread assumption in close range computer vision is that spatial changes in reflectance can occur quickly throughout the image scene, whereas the illumination varies spatially slowly (Land & McCann 1971, Wandell 1987). Moreover the changes in surface orientation are assumed to be spatially slow also, since the reflecting objects

are considered smooth and large in extent. It is claimed that these assumptions are built-in *a priori* knowledge of the human visual perception. These provisions then lead to the *homomorphic filtering* of Oppenheim & Schaffer (1975) where the large spatial frequencies of a given image are filtered out.

For remotely sensed multispectral imagery, however, this assumption of spatially slowly varying surface orientation does not seem to be adequate. Most man-made but also natural objects have image dimensions of only few pixels, and are not smooth in shape. Surface orientation, local illumination and land cover may all change abruptly, *i.e.*, on the scale of 1 pixel. Also most multispectral classification methods are purely pixel based without spatial context analysis. Therefore we here aim at an invariant which is computed from the spectrum of each pixel early in the image processing. No reasoning about the geometry of the spatial surrounding shall be tried. Our novel approach does not recover the true reflectance, but obtains an invariant as the result of spectral rather than spatial filtering.

If there was only one light source present, the spectra could be made invariant by reducing them to *spectral band ratios* (Sabins 1978, Lillesand & Kiefer 1987). Instead of the N signals x_i of spectral band i , one uses the $N - 1$ spectral band ratios x_i/x_{i+1} . Then a change of illumination, *i.e.*, a multiplication of all signals x_i with a common factor, cancels out and has no effect on the band ratios. However, with diffuse irradiance from the sky contributing significantly to the overall illumination, the assumption of a single light source no longer suffices. But with two light sources and respective contributions which can change *independently*, the spectral band ratios will not any longer be invariant.

The uncertainty about the actual irradiance incident onto a specific surface element is twofold:

1. In general, **the surface orientation for a specific surface element is not known**, because Digital Elevation Models have coarse resolution and also do not properly represent artificial objects like houses etc.
2. Moreover, **the skylight cannot be sufficiently modeled as isotropic**. Its directional distribution is strongly dependent on sun position, atmospheric aerosol content, etc. (Valko 1977, Häckel 1985). So even if the surface orientation was known we could not reliably estimate the diffuse contribution of the illuminating sky.

Facing these uncertainties, the aim of this thesis is to formulate a spectral descriptor which is invariant to change of surface orientation by virtue of being invariant to varying contributions from direct solar and diffuse sky light. Just as the spectral band ratio spectra are invariant to the sun light as an assumed only source of illumination, we need a more sophisticated invariant expression in order to account for two light sources (sun and sky) which exhibit different geometric dependencies and are spectrally different.

In this thesis we will not consider specular reflection. Specular reflection does not carry surface specific information (with the exception that it may put constraints on the surface orientation when the position of the illumination source is known), since it reflects the solar radiance independent of wavelength. Often the radiances resulting from specular reflection are so high in value that they saturate the CCD or the amplifier and thus cause overflows in sensors of lower dynamic range (8 bit) which cannot be properly evaluated. Nevertheless, for remotely sensed images in contrast to close range computer

vision applications, specular reflection over land (in contrast to oceanic) surfaces are relatively few and of lower interest.

1.3 Brief Review of Previous Approaches to the Color Constancy Problem

We want to relate the efforts of this thesis to previous approaches to the color constancy problem. The problem has not yet been treated thoroughly in the context of multispectral remote sensing. We clarify why most of the approaches developed for RGB based computer vision seem inappropriate for remote sensing image processing purposes, and point out the crucial differences to our approach.

Hermann Helmholtz (1821–1894) established psycho-physics, introduced physiological coordinates of human color perception (hue, saturation, intensity, Helmholtz (1896)), and commented also on color constancy. Most of the approaches to color constancy are based on three spectral bands, motivated by the three types of chromatic receptors of the human eye (cones), or the three channels of the standard red-green-blue (RGB) video systems.

The analysis of the color constancy problem naturally starts with an inspection of the range of occurring illumination spectra on the one hand, and reflectance spectra on the other.

As far as illumination by artificial light sources is concerned, the CIE (*Commission Internationale de l'Eclairage*, founded in 1931) has defined a number of standard sources with published spectra. The ‘colors’ of the spectra are defined by various three dimensional color space coordinate systems and a color temperature (Judd & Wyszecki 1975).

For outdoor scenes the daylight spectrum and its variability have been measured. The data is usually analyzed by principal component analysis. It is common practise to publish this spectral data as a mean spectral vector⁵ plus the most significant eigenvectors of the spectral covariance matrix and the respective variance eigenvalues. The number of given eigenvectors then is the degree of freedom which is estimated for the variability. The most cited measurements have been conducted by Judd et al. (1964), Judd & Wyszecki (1975) and Dixon (1978). Appropriate spectral data for various air masses can also be found in Accetta & Shumaker (1993). The CIE has proposed various standard daylight spectra with two and three degrees of freedom.

To assess the range of possible reflectance spectra, many researchers rely on spectral measurements of the Munsell color chips. The Munsell colors are welcome as a standard, but it seems doubtful whether they are representative for the complete range of the natural as well as man made surfaces which can occur in remotely sensed multispectral images. Moreover, their gamut has been designed for eye appraisal and thus for the visible wavelength range only.

At the heart of most color constancy algorithms lies a model of finite-dimensional vector spaces for either illumination or reflectance or both, *i.e.*, illumination and/or reflectance are described as linear combinations of a certain number principal vectors.⁶

⁵ Basic terms and concepts are introduced in Chapter 2.

⁶ A RGB ‘spectrum’ contains only 3 signals: red, green and blue. The vector space is thus limited to three dimensions. But also spectroscopic measurements with N spectral bands can be confined

Multiplication of all principal illumination spectra with all principal reflectance spectra then yields a system of principal vectors which span the space of all possibly observable spectra.

The algorithm of Ho et al. (1990) tries to recover illumination and reflection spectrum by finding an minimum root mean square deviation to any combination of the principal illumination and reflectance spectra. The approach needs a well sampled observed spectrum, not only RGB bands. The range of possible reflectance spectra is based on the Munsell colors.

Most RGB based algorithms, however, need not only one given observed spectrum but the spectra of all pixels of the image in question or at least extended subregions thereof. Each spectrum forms a point in the *color space* or *feature space* (see Chapter 2). The observed points within the feature space are fitted to either a planar subspace (Maloney & Wandell 1986, Wandell 1987), an ellipsoid (Petrov & Kontsevich 1994, Drew & Kontsevich 1994) or a convex polyhedron (Forsyth 1990) or polygon (Matas et al. 1994). From the parameters of this fitted manifold the illumination spectra can be recovered, using certain assumptions.⁷ Finally, by division of the observed spectra by the recovered illumination, the reflectance spectra are found. In principle these formalisms could be enhanced for multispectral images with N spectral bands which are a natural extrapolation of the three-band RGB images. The calculus only deviates for some cases where properties of the three dimensional vector space are explicitly used by the previously discussed approaches, *e.g.* unique normals to two dimensional planes etc.

Also the approach of Freeman & Brainard (1995) aims at recovering the illuminant which fits best to the observed spectra. In contrast to the afore mentioned subspace models they use a different loss function instead of the common squared error.

The approach of Ohta & Hayashi (1994) not only needs the complete RGB image, but rather several images of the same scene to recover the illuminant and thus the reflectances.

Some of the techniques intertwine the spectral analysis with spatial filters. They start from assumptions how spectral and spatial changes are related within a certain image area and are thus using spatial context information (Oppenheim & Schafer 1975, Gershon et al. 1986).

Another way to recover the illuminant spectrum is to identify highlights, *i.e.*, specular reflection (Klinker et al. 1987, Klinker et al. 1988, Klinker et al. 1990). In the feature space, all clusters formed by points of specular reflection must converge to the illuminating spectrum, since specular reflection is ‘white’, *i.e.*, the reflection does not change the color of the incident light. An even more special approach is the detection of mutual illumination of several objects which also allows recovery of the illuminant (Funt et al. 1991).

A number of approaches rely on the assumption of and are tested on images of a so-called *Mondrian world* (after the Dutch painter Piet Mondrian 1872–1944), *i.e.*, large rectangular homogeneous color patches on a flat surface. The treatment of the color

to a, say, $5 < N$ dimensional subspace, by means of principal component or factor analysis. The assumption is that all N dimensional spectra can be built by linear combination of, say, five principal components, and can thus be fully represented by the five combination coefficients.

⁷ *E.g.*, if the same scene is observed under different global illumination (*i.e.*, constant within each scene respectively), then all points appear uniformly shifted between the two respective feature spaces, and from this shift the spectral change in illumination can be determined.

constancy problem with respect to varying surface orientation is hence directly opposite to the assumption of a constantly horizontal Mondrian world.

Nagao & Grimson (1995) are probing beyond the Mondrian world and allow changes in surface orientation. They suggest the use of photometric invariants. However, their invariant is based on the assumption that the contribution of the diffuse (also called ambient) illumination is equal for all surface patches not regarding orientation.

Also Barnard et al. (1996) consider scenes with spatially varying illumination. Their approach is based on Forsyth (1990) and also needs the input of the whole scene. The algorithm is aimed at recovering the illuminating spectra from the convex set spanned by the observed spectra in the RGB feature space (color space). The approach is thus necessarily depending on the image scene contents.

Finlayson et al. (1996) suggest a way to perform color constant mapping of whole RGB-images to single predefined objects (object recognition). Instead of the common histogram based classification, they essentially use the spectral correlations of the RGB-images. They exploit the fact that in contrast to variances and covariances, the correlations are invariant against bandwise multiplication or addition.

A number of works treat the *shape from shading* problem (*e.g.* Horn (1986), Jähne (1993b)). *Shading* is the spatial brightness or color variation which is due to the local orientation of the reflecting surface. The basic assumption of shape from shading is that the local surface orientations of small planar surface patches can be recovered from their brightness or color. The orientations of all surface elements can then be integrated in order to recover a three-dimensional surface or terrain model. Practically all shape from shading algorithms assume either point light sources or completely isotropic illumination and use the Lambertian reflection model. An exception is Jähne (1993b) who suggests *shape from reflection* for sea applications and assumes specular reflection under an extended light source. Shape from shading on gray-value images is ambiguous. The ambiguity can be overcome *e.g.* by using multiple images with different light source positions (*e.g.* Heipke (1992), Heipke & Piechullek (1994)), or by using color imagery obtained under several point light sources of different illumination spectra (Schlüns 1992, Drew & Kontsevich 1994, Kontsevich et al. 1994). These approaches are related to the subject of this thesis as they analyze how the spectral appearance of a surface patch changes with orientation. However, the named approaches are constrained to laboratory conditions, whereas this thesis analyzes the problem with respect to the outdoor situation encountered in remote sensing. For remote sensing applications, shape from shading approaches could work only for spectral bands with wavelength $\lambda > 2\mu\text{m}$, because only then the diffuse illumination is essentially vanishing and the direct solar irradiance the sole illumination (see Fig. 3.2 on page 38). To account for diffuse illumination in shorter wavelength bands must fail since the diffuse skylight cannot satisfactorily be modeled as isotropic (Valko 1977). Also, shape from color approaches cannot be tried, as the sun is the only available point source.

Some approaches are motivated by a physiological interest in visual perception and strongly based on findings about the human visual neural system. The influential retinex theory was introduced by Land & McCann (1971). Also neural network approaches to the color constancy problem have been made (Moore et al. 1991). This thesis, however, is not concerned with modeling human visual perception but motivated from a physics perspective. Hence we are not so much interested in neural classifiers, which

are parameter-free but need extensive training. Here we rather start from a physical model of the spectral signal formation, check our analysis against experimental data, and construct an appropriate linear algebraic approach in the logarithmic domain.

The differences which distinguish most previous approaches from the one followed in this thesis can be briefly summarized, in that here we use:

- ▶ a physical parametric model,
- ▶ natural outdoor illumination,
- ▶ dichromatic illumination (direct plus diffuse, both with varying contributions)
- ▶ no Mondrian world assumption but varying surface orientation,
- ▶ pixel-wise processing of single spectra without spatial context assumptions,
- ▶ an invariant descriptor instead of costly recovery trial,
- ▶ reflectance spectra not constrained to Munsell colors or other subspace.

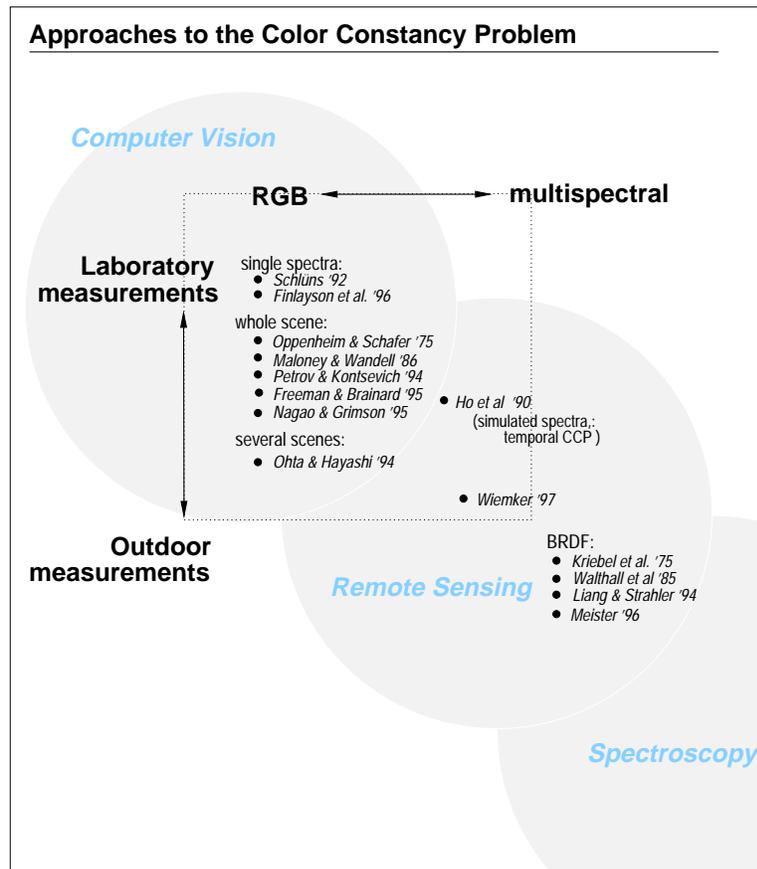


Figure 1.2: Tentative sketch of a color constancy approach topography. This work is located halfway between full scale BRDF measurements and computer vision approaches.

1.4 Structure of this Thesis

Following the formulation of the problem in the beginning of this chapter we have discussed previous approaches to the color constancy problem. It has been clarified to what extent these approaches are helpful to remote sensing image processing purposes, and why they are not appropriate for our objective of deriving a spectral invariant.

In Chapter 2 we will give a brief introduction and definition of those basic concepts of multispectral imagery and classification which will be used in later chapters.

Chapter 3 describes the spectral characteristics of direct sun and diffuse sky illumination. We analyze the effect of changing direct and diffuse contributions in the conventionally used feature space. This N -dimensional space is spanned by the axes of the N spectral bands i . An appropriate spectral distance and multispectral analysis methods are developed with respect to the color constancy problem.

In Chapter 4 we introduce the logarithmic feature space which is spanned by the logarithms $\ln x_i$ of the spectral signals x_i . This space is not commonly used in processing of digital images, and we investigate its properties with respect to multispectral classification. Then the color constancy problem is analyzed in this logarithmic space. An algebraic projector is introduced which is shown to map the spectra to an invariant spectral descriptor. The capabilities of the operator are demonstrated on simulated data, and an analysis of noise sensitivity and error propagation is carried out.

In Chapter 5 we present spectra which were recorded during *in situ* experiments with the radiometer OVID. The data is analyzed in a high dimensional logarithmic feature space, and the results show the validity of the considerations undertaken in Chapter 4.

In Chapter 6 the suggested projector is applied to the second set of experimental data, *i.e.*, to remotely sensed multispectral imagery recorded with the airborne line scanner DAEDALUS. Picking spectral samples of interest from the image data, we analyze the performance of the suggested transform in a variety of possible feature spaces. Moreover, analysis of the feature space densities illustrates just how the transform works on the experimental spectral image data. Both the original and the transformed spectral data subjected to unsupervised classification.

We finally draw conclusions in Chapter 7. A synopsis is tried between the conventional and the logarithmic feature space. Advantages and drawbacks of the suggested invariant mapping are discussed, and the most useful version of the transform is recommended. Last not least the progress made in this thesis is critically reviewed.

Basic Concepts of Multispectral Imagery and its Classification

Outline of this chapter – We introduce basic concepts of multispectral imagery and its classification. In particular, we define and briefly explain those notions and techniques which will be used throughout this thesis. We start with a physical model of the spectral image formation process. Later, we describe the multispectral feature space and common statistical techniques used for multispectral image classification.

Thorough introductions into scattering and reflection processes, radiometric measurement and radiometric magnitudes can be found e.g. in Gerthsen *et al.* (1977), Kraus & Schneider (1988), Rees (1990), Albertz (1991), and Hapke (1993), whereas the emphasis is more on image data processing and classification of imagery e.g. in Duda & Hart (1973), Ballard & Brown (1982), Horn (1986), Jähne (1993a), and Richards (1993).

2.1 The Spectral Signal Formation Process

■ Measurement

We consider multispectral image data recorded by an opto-electronic sensor. The sensor has a fixed relatively small aperture and records radiation in N spectral bands with index i . Each spectral band i has a spectral sensitivity distribution function with a centroid wavelength λ_i . The output voltage of the detector elements is amplified and converted to digital counts [DC] by an AD-converter. The calibration functions of the sensor – as established in the laboratory – allow to convert the recorded raw digital counts into the physical magnitude of *spectral radiances* L in units of $[\text{W m}^{-2} \mu\text{m}^{-1} \text{sterad}^{-1}]$. The radiance L^s describes the radiation power per unit of area, wavelength and solid angle which is received by the sensor from a particular direction. Radiances L are defined as the radiant flux Φ incident on an area element dA from a particular direction and from an infinitesimally small solid angle $d\Omega$. *Spectral radiances* L_λ are

radiances in an infinitesimally small wavelength interval:

$$\text{radiance:} \quad L = \frac{d\Phi}{dA d\Omega} \quad (2.1)$$

$$\text{spectral radiance:} \quad L_\lambda = \frac{d\Phi}{dA d\Omega d\lambda} \quad , \quad (2.2)$$

and, if the incident beam hits the sensor under an angle θ^s :

$$L_\lambda^s(\theta^s) = \frac{d\Phi^s}{dA^s d\Omega^s d\lambda \cos \theta^s} \quad . \quad (2.3)$$

Our sensor, of course, has a finite aperture and a finite wavelength interval. But using the calibration functions and the simplification that the radiance is constant within the small aperture and narrow wavelength band i of the sensor, the radiation measured by the sensor can be converted to N radiances L_i^s .

■ Reflection

Throughout this thesis we use the common and widespread assumption of Lambertian reflection, which is also called totally diffuse or ideally matte reflection. Its most prominent property is that when looking with a small aperture sensor at a Lambert reflecting surface, the measured radiance will be the same under all observation angles θ . In other words, the Lambertian surface has a uniform bidirectional reflectance distribution. Therefore we do not need to consider the geometry of the recording setup relative to the reflecting surface. This idealization gives rise to the notion of a stable spectral signature, which allows to identify different types of surface cover from their spectral appearance.

Let us consider a calibrated sensor with a small aperture optic which observes an extended planar horizontal Lambertian surface. The reflecting surface is illuminated by the incident irradiance E^\downarrow (direct plus diffuse), which is integrated from all incident radiances L^\downarrow over the complete hemisphere of the sky dome:

$$\text{irradiance:} \quad E^\downarrow = \int_{2\pi} L^\downarrow(\theta') \cos \theta' d\Omega \quad (2.4)$$

with the zenith angle θ' . The Lambert law states that the radiance L^r reflected from an infinitesimally small point source is proportional to the incident irradiance E^\downarrow multiplied by the Lambert reflectance factor r and an angular cosine dependence (Fig. 2.2 on page 16):

$$L^r = \frac{1}{\pi} r E^\downarrow \cos \theta \quad (2.5)$$

with the observation angle θ . There is a subtle difference, however, between the reflected radiance L^r and the radiance measured by the sensor L^s (Fig. 2.1 on the facing page). They are related as

$$L^s = L^r / \cos \theta \quad (2.6)$$

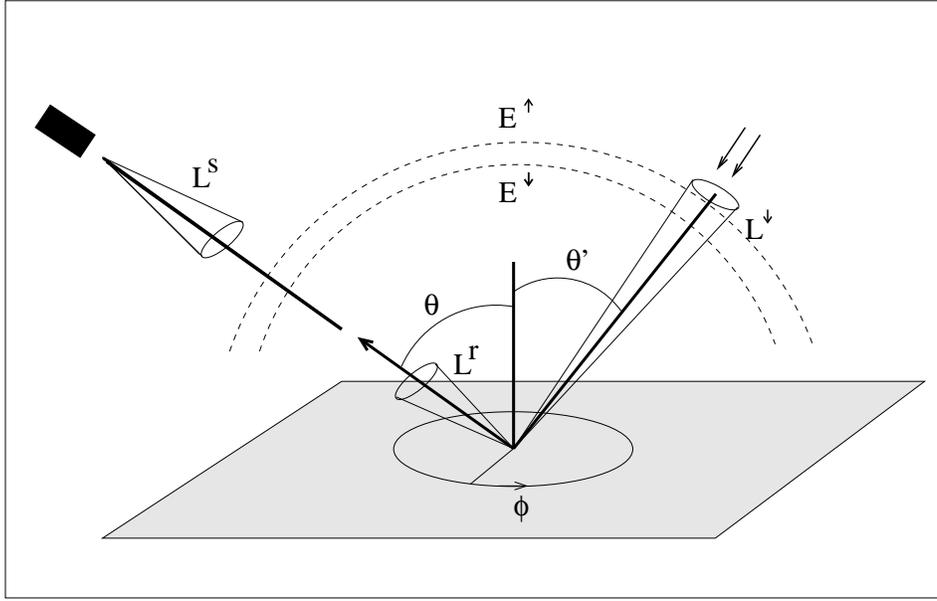


Figure 2.1: Sensor target viewing geometry, and incident and reflected radiances L^s , L^r , L^\downarrow and irradiance E^\downarrow .

because the flux Φ^s on the detector element dA^s is inverse proportional to the squared distance D between sensor and reflector, and proportional to the sensor's instantaneous field of view (IFOV) dA on the reflecting surface:

$$\frac{d\Phi^s}{dA^s} = \frac{L^r dA}{D^2} \quad (2.7)$$

where the IFOV is given by $dA = D^2 d\Omega^s / \cos \theta$ (Fig. 2.2 on the next page). Hence

$$\frac{d\Phi^s}{dA^s} = \frac{L^r D^2 d\Omega^s}{D^2 \cos \theta} \quad (2.8)$$

so that the distance D cancels out, and the radiance L^s received by the sensor is

$$L^s = \frac{d\Phi^s}{dA^s d\Omega^s} = L^r / \cos \theta \quad . \quad (2.9)$$

Substitution of L^r from Eq. 2.5 on the facing page in Eq. 2.6 on the preceding page then yields

$$L^s = \frac{1}{\pi} r E^\downarrow \quad . \quad (2.10)$$

Thus the cosine dependence cancels out and we see the Lambertian properties: the observed radiance L^s is invariant against the viewing angle θ and the sensor-target-distance D .

Integration of the reflected radiance L^r over the upper hemisphere yields the total reflected *exitance* E^\uparrow (Rees 1990), the equivalent to the incoming irradiance E^\downarrow . The

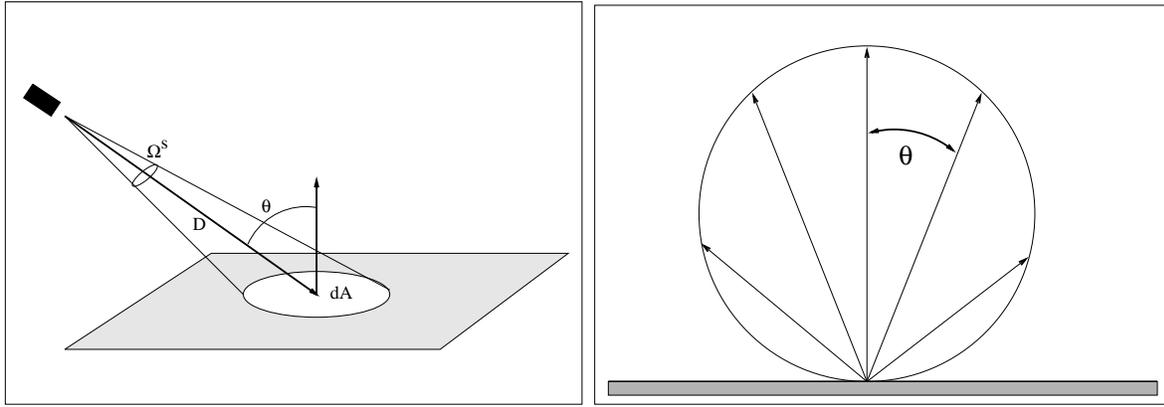


Figure 2.2: Left: Reflector area viewed by the sensor (IFOV).

Right: Cosine dependence of the radiance reflected from a single particle of a Lambertian reflecting surface.

reflected exitance E^\uparrow equals the incident irradiance E^\downarrow multiplied with the Lambert reflectance r :

$$\text{exitance:} \quad E^\uparrow = \int L^r(\theta, \phi) d\Omega = \int \frac{1}{\pi} r E^\downarrow \cos \theta d\Omega \quad (2.11)$$

$$= \frac{1}{\pi} r E^\downarrow \int_0^{\pi/2} \cos \theta \sin \theta d\theta \int_0^{2\pi} d\phi \quad (2.12)$$

$$= \frac{1}{\pi} r E^\downarrow \cdot \frac{1}{2} \cdot 2\pi \quad (2.13)$$

$$= r E^\downarrow \quad (2.14)$$

with θ the zenith and ϕ the azimuth angle of the reflected radiation in the angular system of the surface (see Fig. 2.2).

For most surface materials the Lambert law certainly is only an approximation. A more refined representation of the reflection properties is given by the bidirectional reflectance distribution function (BRDF). However, the cost of measuring this function for a sufficient number of illumination and reflection angles is quite high (Kriebel 1978, Meister 1995), and also for its use in subsequent image processing. Therefore most remote sensing digital image processing starts from the Lambertian approximation.

All these magnitudes can be spectrally dependent: $L_\lambda, E_\lambda, r_\lambda$. For our considerations of a multispectral sensor, we denote them with the index i for the spectral bands i with the centroid wavelength λ_i : L_i, E_i, r_i . A stricter formulation of the Lambert law even demands a vanishing wavelength dependence $r_i = r, \forall i$, *i.e.*, a ‘white’ reflector.

■ Dependence of Illumination on Surface Orientation

When illuminated by a point source, the irradiance E^\downarrow incident onto a given surface patch is proportional to the cosine of the angle θ' between the surface normal and the direction of the point source (inner product of the two normal vectors):

$$E_{\text{dir}}^\downarrow \propto \cos \theta' \quad (2.15)$$

The sun is commonly considered as a point source, and for a horizontal surface θ' then equals the solar zenith angle θ_\odot .

If the sky light is considered at all, then it is commonly assumed to be a diffuse source which is distributed isotropically over the hemisphere of the sky. This diffuse irradiance onto a surface of inclination θ' can be shown (Horn 1986) to yield the angular dependence:

$$E_{\text{diff}}^\downarrow \propto \frac{1}{2}(1 + \cos \theta') = \cos^2(\theta'/2) \quad . \quad (2.16)$$

However, as pointed out earlier, the sky's intensity distribution is usually not isotropic (Häckel 1985, Minnaert 1993). Moreover, the horizon line is at most times occluded by other objects and thus not the whole hemisphere will be visible to the reflecting surface.

■ Atmospheric Influence

Scattering and absorption are the mechanisms which cause an attenuation of the reflected radiance by the atmosphere. This attenuation is described by the *transmittance* $T \leq 100\%$. For solar irradiance from an solar azimuth angle θ_\odot , the transmittance is related to the optical depth τ by $T = \exp(-\tau \cos^{-1} \theta_\odot)$. For spectral band i with the centroid wavelength λ_i , a horizontal reflecting surface experiences an illuminating irradiance

$$E_i^\downarrow = E_{\text{dir},i}^\downarrow T_i(\theta_\odot) \cos \theta_\odot + E_{\text{diff},i}^\downarrow \quad (2.17)$$

where the direct irradiance $E_{\text{dir}}^\downarrow$ is extra-atmospherical and attenuated by the atmosphere, whereas the diffuse irradiance $E_{\text{diff}}^\downarrow$ is generated just there by scattering (see Fig. 2.3 on the next page). A part r/π of the irradiance E_i^\downarrow is reflected in the direction θ and again attenuated by the atmosphere. So we introduce another transmittance $T'(\theta)$ for the sensor recording under an observation angle θ (Fig. 2.3).

$$L_i^s = \frac{1}{\pi} r_i T'_i(\theta) E_i^\downarrow + L_{o,i} \cos^{-1} \theta \quad (2.18)$$

$$= \frac{1}{\pi} r_i T'_i(\theta) \left[E_{\text{dir}}^\downarrow T_i(\theta_\odot) \cos \theta_\odot + E_{\text{diff}}^\downarrow \right] + L_{o,i} \cos^{-1} \theta \quad (2.19)$$

Here L_o is the *path radiance*. It originates via diffuse scattering of the incoming solar radiation in the path of the view line between sensor and target surface. The path radiance stems from solar irradiance which is actually scattered towards the sensor by the atmospheric constituents before it reaches the ground (Fig. 2.3). We set L_o for the path radiance received at nadir viewing angle ($\theta = 0$), and assume that the path radiance is proportional to the length of the view path through the atmosphere to the target: $L_o \cos^{-1} \theta$. Also the transmittance $T'(\theta)$ between target and sensor can be approximated as $T' \cos^{-1} \theta$. If this angular dependence in Eq. 2.19 is neglected (*e.g.* for small swath angles), or $T'(\theta) \approx 100\%$ for low recording flight altitudes, we actually get a linear relation between reflectance r and measured radiance L^s for all pixels of a

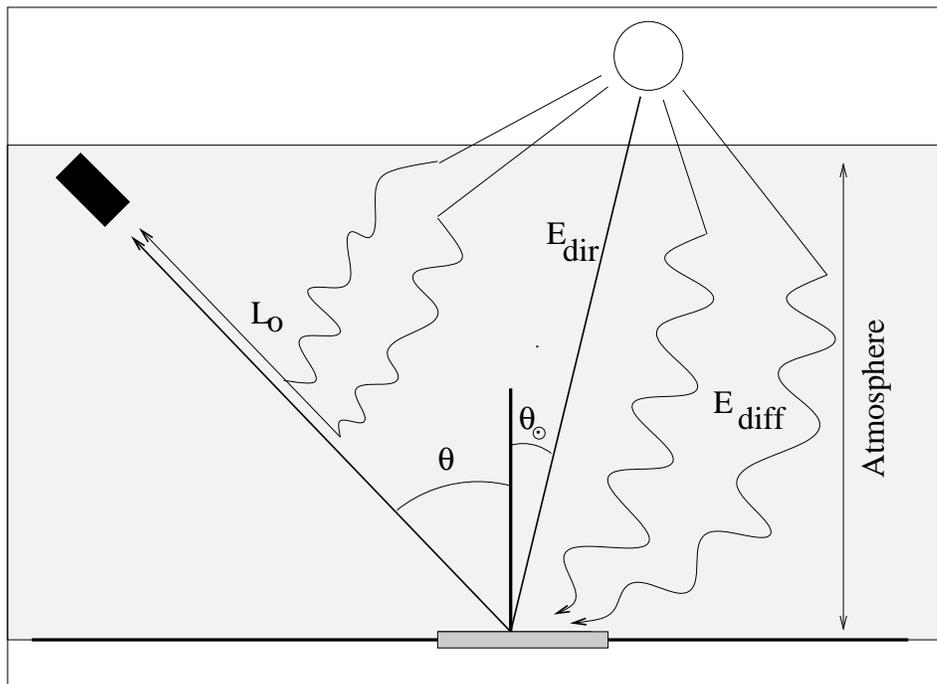


Figure 2.3: Various radiation paths: sun – target – sensor.

given image band i :

$$L_i^s = c_i r_i + L_{o,i} \quad (2.20)$$

$$\text{where } c_i = \frac{1}{\pi} T'_i \left[E_{\text{dir}}^\downarrow T_i(\theta_\odot) \cos \theta_\odot + E_{\text{diff}}^\downarrow \right] \quad (2.21)$$

$$(2.22)$$

Experiences have shown that this linear relation is a good first order approximation (Richter 1992, Hepp 1994, Kollewe 1995). A unique linear approximation of the atmospheric influence over the complete image band is computationally uncstly and has been used *e.g.* by Schott et al. (1988), Moran et al. (1990), Wegener (1990), Hall et al. (1991), Jahn & Grassl (1991).

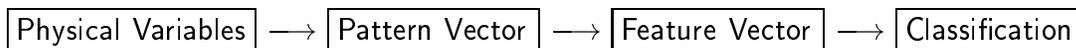
■ Atmospheric Correction

Following a common atmospheric correction scheme, the magnitudes $T, T', E_{\text{dir}}, E_{\text{diff}}, L_o$ of Eq. 2.19 can be computed for a horizontal surface patch using a radiative transfer code (RTC), *e.g.* SENSAT (Richter 1990, Richter 1992). With the measured L_i^s , the equation then can be solved for the sought reflectance r_i . However, we prefer to call these so recovered magnitudes *pseudo-reflectances* r_i^* , because their determination neglects their respective surface orientations. The surface orientation can differ for each pixel and is in general unknown.

2.2 Multispectral Imagery, Spectral Vectors and the Feature Space

For image processing purposes, these pseudo-reflectances r_i^* are coded into dimensionless numerical values $x_i \hat{=} r_i^*$ and stored as a data block. In its band sequential storage (BSQ), a *multispectral image* then consists of N image layers, each of which holds the gray values of one spectral band (alternatively the data can also be stored pixel-wise or line-wise band interleaved, BIP and BIL). The BSQ format is a three dimensional data array (x - y -pixel coordinate in the k th layer), which is often thought of as an image cube (Fig. 2.5 on page 21). A spectrum then consists of the entry at the same coordinate in each of the N layers. We call $\mathbf{x} = (x_1 \dots x_N)^T$ the *spectral vector*, which is a column vector with N entries. The spectrum can then be plotted from the N entries x_i versus the proper wavelengths λ_i (see Fig. 2.4 on the following page).

The set of spectral vectors \mathbf{x} can be represented as points in an N -dimensional vector space. In the field of pattern recognition, the terminology is such (Kohonen 1995) that the physical variables are registered by detectors and transduced to a *pattern vector*. By preprocessing, application specific *features* can be derived from these pattern vectors, and stored into a *feature vector*, which is used for the final classification (labeling) or identification.



In multispectral imagery each pixel has a vector of values from the N spectral bands of the respective sensor. The values can be either raw detector values (Digital Counts [DC]), calibrated radiances [$W/(m^2 \text{ sr } \mu\text{m})$], or reflectances [dimensionless], or other processed quantities. In the field of multispectral remote sensing it is common in that all these vector spaces are simply called *spectral space* or *feature space*, without differentiation between various levels of processing. Also throughout this thesis we will use the term *feature space* for the vector space as spanned by the N reflectance values¹ in the N spectral bands of the sensor, and the term *transformed feature space* after the transformations – which are suggested by this thesis – are applied.

Again, in the N -dimensional feature space, each pixel of a particular image is presented by a point with coordinates given by the value of the pixel in each spectral band. Then the basic idea of classification is that the pixels form clusters in the feature space corresponding to various ground cover types (see Fig. 2.6 on page 22). Each cluster is supposed to represent a *spectral class*. A spectral class is not necessarily identical to a *conceptual class*, such as ‘water’, ‘streets’, ‘houses’, ‘forest’ or ‘meadow’, but it is expected that relations between the spectral and the conceptual classes can be established.

The shape of the clusters in the feature space depends on the intrinsic variability of the cover type’s spectral signature, on noise, moisture content, BRDF effects, topographic effects such as surface orientation, etc. In practise, the clusters will not be as well separable as depicted in Fig. 2.6, but rather be only more or less pronounced aggregations embedded in a continuum of points in the feature space.

¹ The reflectances r_i (of spectral band i) are computed by calibration of the detector raw values and atmospheric correction under the assumption of Lambertian reflection (see Section 2.1 on page 13). We choose to call them *pseudo-reflectances* r_i^* if they are computed under the assumption of horizontal surface orientation.

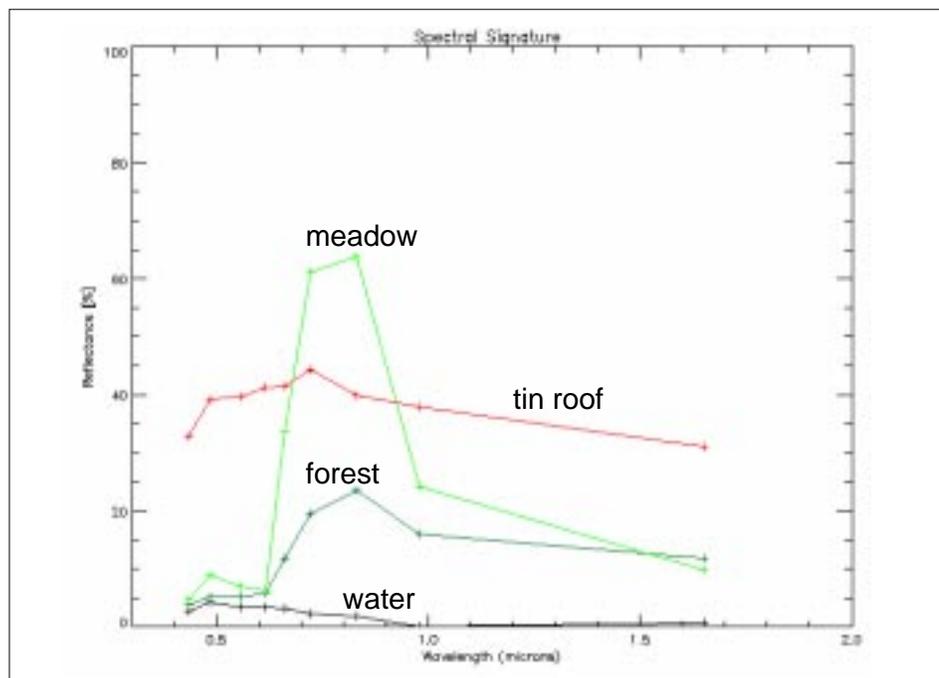
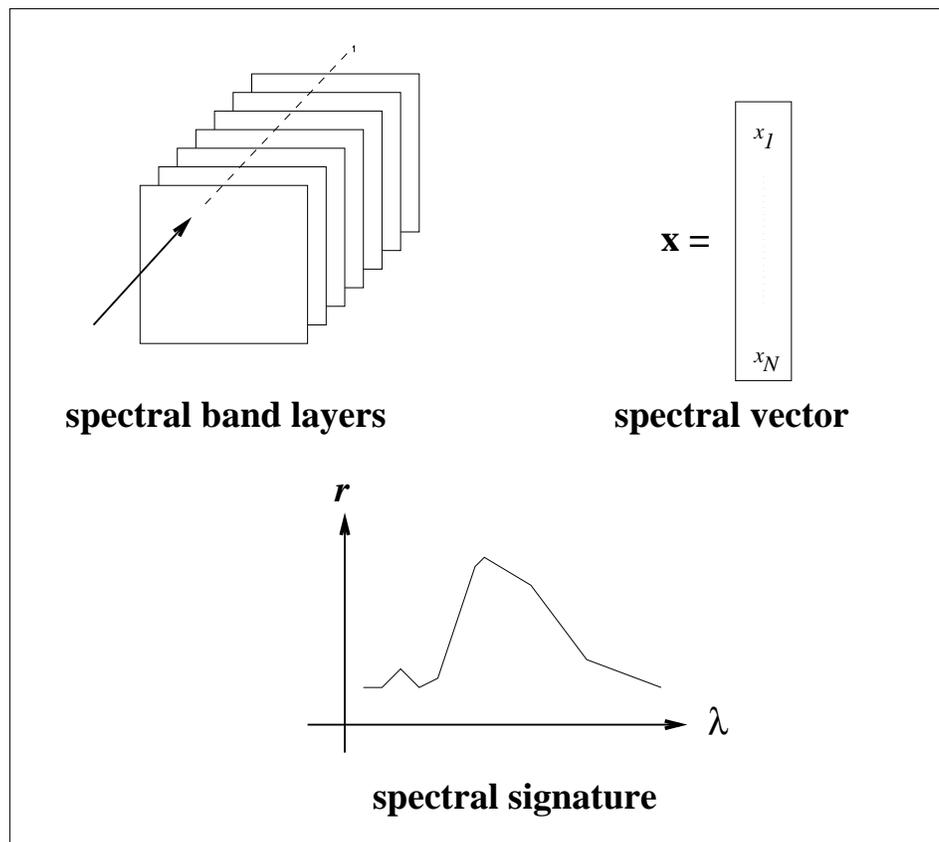


Figure 2.4: Top: Concept of multispectral imagery, spectral vectors and signatures. Bottom: Some exemplary spectra from multispectral imagery ($N = 9$ spectral bands, DAEDALUS sensor).

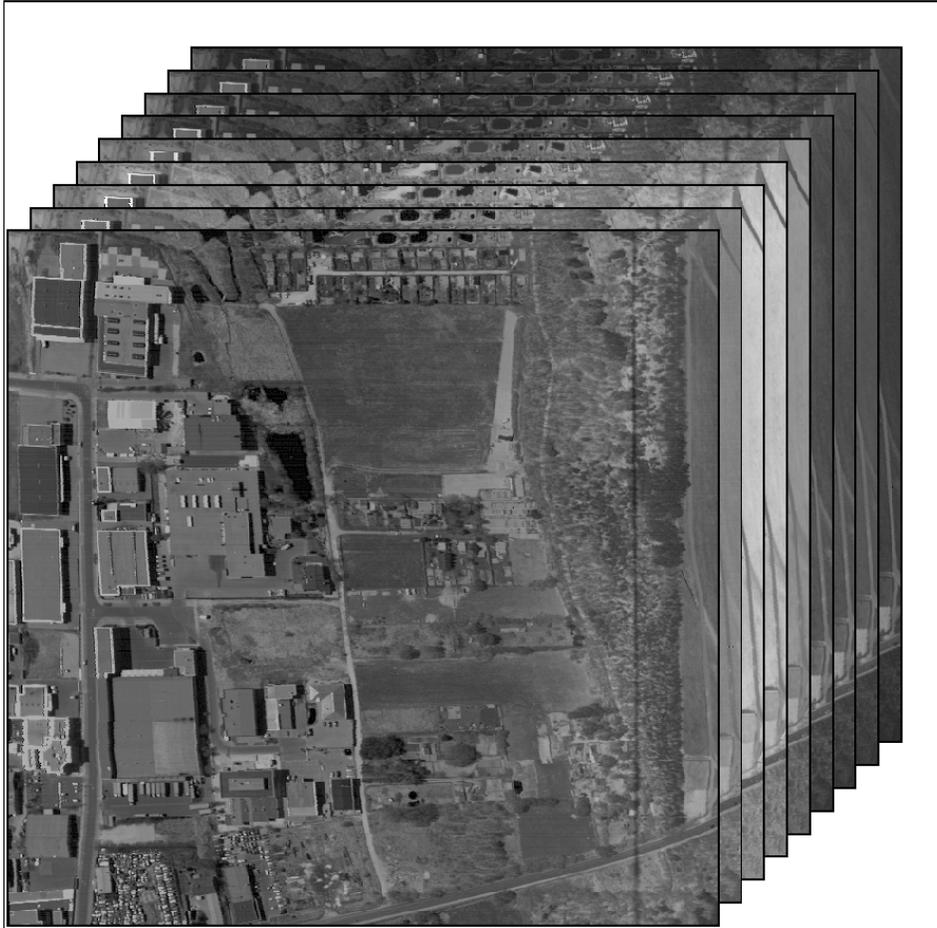


Figure 2.5: *Multispectral image cube with $N = 9$ spectral bands.*

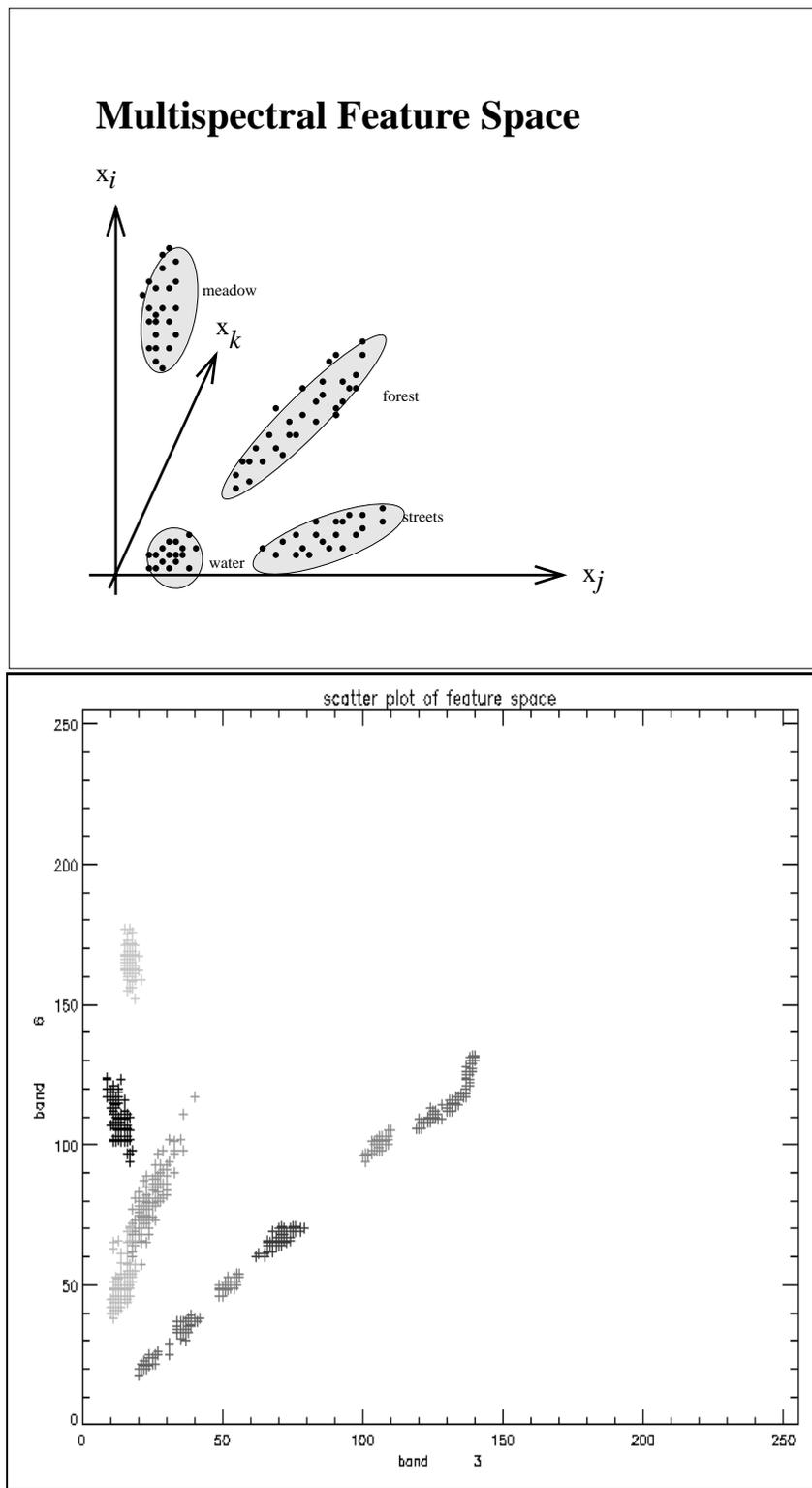


Figure 2.6: *The multispectral feature space.*

Top: Concept sketch.

Bottom: Exemplary scatter plot of two spectral bands (615 vs 720 nm) of selected regions of the remotely sensed multispectral image shown in Fig. 2.5.

2.3 Spectral Distances and Classification in the Feature Space

Most multispectral classification concepts are based on a specific model of the *spectral distance* of an observed spectrum \mathbf{x} to a certain spectral class denoted by an index a . Then the probability for the observed spectrum \mathbf{x} to belong to class a increases with decreasing spectral distance. Following the maximum likelihood principle derived from Bayesian decision theory (Duda & Hart 1973, Richards 1993), the spectrum \mathbf{x} is assigned to the spectral class a for which the spectral distance is minimal. So the definition of the spectral class and the spectral distance is crucial for the multispectral classification process.

Spectral classes are commonly described as points belonging to class a centered around a *cluster center* \mathbf{m}_a in the N dimensional feature space. Let \mathbf{X}_a be the set of all n_a vectors \mathbf{x} belonging to class a . Then class a has a cluster center \mathbf{m}_a ,

$$\mathbf{m}_a = \frac{1}{n_a} \sum_{\mathbf{x} \in \mathbf{X}_a} \mathbf{x} \quad (2.23)$$

and the spectral distance between a feature vector \mathbf{x} and the class center \mathbf{m}_a can *e.g.* be the squared *Euclidean distance*:

Euclidean distance:

$$d_E^2(\mathbf{x}, \mathbf{m}_a) = \|\mathbf{x} - \mathbf{m}_a\|^2 \quad (2.24)$$

$$= (\mathbf{x} - \mathbf{m}_a)^T (\mathbf{x} - \mathbf{m}_a) \quad . \quad (2.25)$$

The Euclidean distance can be modified by a metric \mathbf{G} , where the symmetric $N \times N$ matrix $\mathbf{G} = \mathbf{G}^T$ is injected into the inner product and yields the more general quadratic form

Metric induced distance:

$$d_G^2 = (\mathbf{x} - \mathbf{m}_a)^T \mathbf{G} (\mathbf{x} - \mathbf{m}_a) \quad . \quad (2.26)$$

The Euclidean distance (Eq. 2.25) implies that points of equal distance lie on hyperspheres centered around \mathbf{m}_a . This means that we expect no correlations between the features and the variance of each feature is equal. If the scatter of class points in the feature space around the class center is not spherically distributed, we need another spectral distance. The distribution of the points around the cluster center is *multivariate* as it varies in N different variables x_i . The representation of a multivariate Gaussian normal distribution for class a is known as an ellipsoid (see Fig. 2.7 on the next page) given by the *covariance matrix* Σ_a which is estimated as

$$\Sigma_a = \frac{1}{n_a} \sum_{\mathbf{x} \in \mathbf{X}_a} (\mathbf{x} - \mathbf{m}_a)(\mathbf{x} - \mathbf{m}_a)^T \quad . \quad (2.27)$$

For an assumed Gaussian normal distribution this leads to the well known *Mahalanobis distance* d_M^2 (Duda & Hart 1973, Richards 1993), where points of equal distance lie on

ellipsoids. For the appropriate spectral distance the inverse of the estimated covariance Σ_a is taken as a metric \mathbf{G} and substituted in Eq. 2.26 on the preceding page:

Mahalanobis distance:

$$d_M^2(\mathbf{x}, \mathbf{m}_a, \Sigma_a) = (\mathbf{x} - \mathbf{m}_a)^T \Sigma_a^{-1} (\mathbf{x} - \mathbf{m}_a) \quad . \quad (2.28)$$

If the covariance ellipsoids are identical for all classes, $\Sigma_a = \Sigma, \forall a$, then all vectors \mathbf{x} can be mapped to $\mathbf{x} \mapsto \mathbf{x}' = \Sigma_a^{-1/2} \mathbf{x}$, and the quadratic Mahalanobis classifier reduces to a linear classifier (Eq. 2.25) which may be computationally more efficient. We will show in Chapter 4 that this is just the case for the illumination covariance matrix in the logarithmic spectral feature space.

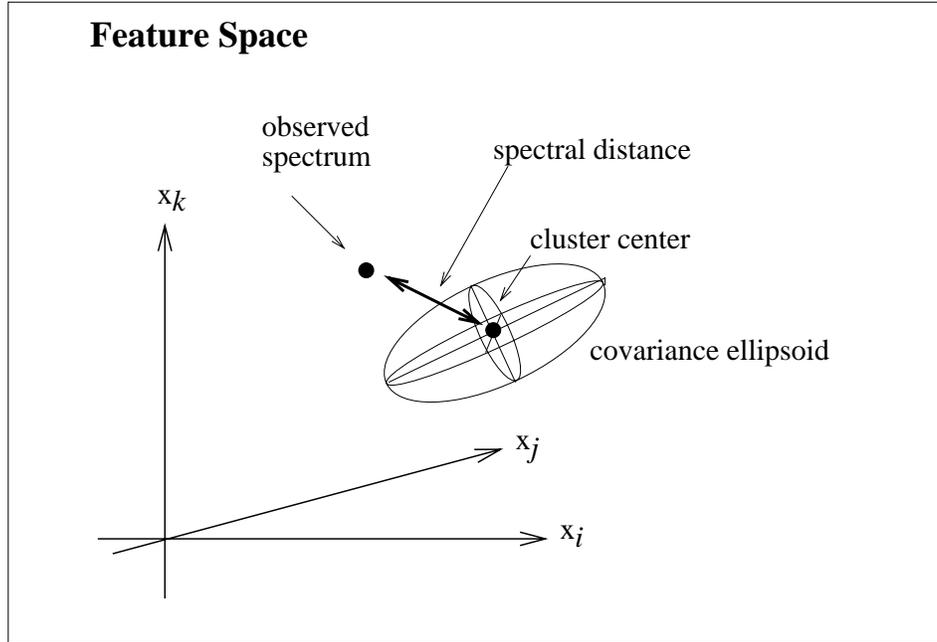


Figure 2.7: The feature space with the center \mathbf{m}_a of class a , the covariance ellipsoid Σ and the distance $\mathbf{d} = \mathbf{x} - \mathbf{m}_a$ between the class cluster and an observed spectrum \mathbf{x} .

For a given pixel with feature vector \mathbf{x} we have the Gaussian *probability density function* (PDF) with respect to class a :

$$p(\mathbf{x}|a) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_a|}} e^{-\frac{1}{2} d_M^2} = \frac{1}{\sqrt{(2\pi)^N |\Sigma_a|}} e^{-\frac{1}{2} (\mathbf{x} - \mathbf{m}_a)^T \Sigma_a^{-1} (\mathbf{x} - \mathbf{m}_a)} \quad (2.29)$$

where $|\Sigma_a|$ is the determinant of the covariance matrix Σ_a . The PDF is normalized so that $\int_{-\infty}^{\infty} p(\mathbf{x}|a) d\mathbf{x} = 1$. Let each class a have an *a priori* probability of $P(a)$. Then *Bayes Rule* changes this *a priori* probability into the *a posteriori* probability

$$P(a|\mathbf{x}) = \frac{p(\mathbf{x}|a) P(a)}{\sum_a p(\mathbf{x}|a) P(a)} \quad (2.30)$$

which describes the probability that the pixel with feature vector \mathbf{x} belongs to class a . All probabilities are normalized so that $\sum_a P(a|\mathbf{x}) = 1$. The *Bayes decision rule* states

that pixel should be assigned to the class for which $P(a|\mathbf{x})$ becomes maximal, because this minimizes the probability of error (Duda & Hart 1973).

In case the *a priori* probabilities $P(a)$ are equal for all classes a , Eq. 2.30 reduces to

$$P(a|\mathbf{x}) = \frac{p(\mathbf{x}|a)}{\sum_a p(\mathbf{x}|a)} \quad (2.31)$$

and yields a *likelihood* decision instead of a probability decision, because it is based entirely on the likelihood $p(\mathbf{x}|a)$ (Duda & Hart 1973). Again, we decide for the maximum likelihood. This principle is called the *Maximum Likelihood Classification* (ML).

When deciding for the class a with the highest probability $P(a|\mathbf{x})$ of Eq. 2.31, the computation of the denominator $\sum_a p(\mathbf{x}|a)$ is not really necessary, since it is the same for all classes. It suffices to decide for the class with the highest probability density $p(\mathbf{x}|a)$ (Eq. 2.29). This, in turn, is the same as deciding for highest logarithmic probability density:

$$\ln p(\mathbf{x}|a) = -\frac{1}{2}N \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_a| - \frac{1}{2}d_M^2 \quad (2.32)$$

$$= -\frac{1}{2}N \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_a| - \frac{1}{2}(\mathbf{x} - \mathbf{m}_a)^T \boldsymbol{\Sigma}_a^{-1} (\mathbf{x} - \mathbf{m}_a) \quad (2.33)$$

For reasons of analogy to the previously introduced spectral distances, we might want to look for a minimum distance principle instead of a maximum logarithmic probability density. This gives rise to the spectral distance as derived from Maximum Likelihood decision, which is just the negative logarithmic probability density:

distance derived from Maximum Likelihood decision:

$$d_{ML}^2 = \frac{1}{2}N \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{\Sigma}_a| + \frac{1}{2}(\mathbf{x} - \mathbf{m}_a)^T \boldsymbol{\Sigma}_a^{-1} (\mathbf{x} - \mathbf{m}_a) \quad (2.34)$$

Computation of the leading term is not even necessary as it is a normalization factor which is equal for all classes.

2.4 Clusters and Separability Measures

We want to look briefly into measuring separabilities of spectra and spectral clusters. Separability is the basic prerequisite for classification. Separability measures will be used in the statistical evaluation in Chapter 5 and Chapter 6.

A simple and widely used criterion for clustering is to minimize the mean-squared-error (MSE) within each cluster a :

$$J_a^2 = \frac{1}{n_a} \sum_{\mathbf{x} \in \mathbf{X}_a} \|\mathbf{x} - \mathbf{m}_a\|^2 \quad (2.35)$$

$$= \text{tr } \boldsymbol{\Sigma}_a \quad (2.36)$$

where \mathbf{X}_a is the set of all n_a vectors \mathbf{x} belonging to class a , and

$$\mathbf{m}_a = \frac{1}{n_a} \sum_{\mathbf{x} \in \mathbf{X}_a} \mathbf{x} \quad \text{and} \quad \boldsymbol{\Sigma}_a = \frac{1}{n_a} \sum_{\mathbf{x} \in \mathbf{X}_a} (\mathbf{x} - \mathbf{m}_a)(\mathbf{x} - \mathbf{m}_a)^T \quad (2.37)$$

are the mean vector and covariance matrix respectively. Actually it can be shown that the cluster center \mathbf{m}_a of Eq. 2.37 is just the one which minimizes J_a^2 (Duda & Hart 1973). J_a^2 measures the square of the mean scattering radius, since it is the mean sum of the variances in all coordinate directions. From linear algebra we know that the trace $\text{tr } \Sigma_a$ is invariant against coordinate rotations (Jänich 1993). By *rotation* of the coordinate system we mean a transformation $\mathbf{x} \mapsto \mathbf{G}\mathbf{x}$ with an orthogonal metric \mathbf{G} which has the properties $\mathbf{G}^T\mathbf{G} = \mathbf{I}$ and $|\mathbf{G}| = 1$. Then the trace remains unchanged:

$$\text{tr } \Sigma_a \mapsto \text{tr } (\mathbf{G}\Sigma_a\mathbf{G}^T) \quad (2.38)$$

$$= \text{tr } (\Sigma_a\mathbf{G}^T\mathbf{G}) \quad (2.39)$$

$$= \text{tr } \Sigma_a \quad (2.40)$$

because of the general property of the trace function: $\text{tr}(AB) = \text{tr}(BA)$.

Let \mathbf{m} and \mathbf{m}' be the cluster centers of two classes a and a' with covariances Σ and Σ' . Then the displacement vector between the two clusters is $\mathbf{d} = \mathbf{m} - \mathbf{m}'$ (see Fig. 2.8 on page 28). We want to have a measure which describes a scalar separability between the clusters which takes into account their extensions. To this aim we introduce the *between-cluster covariance matrix*

$$\Sigma_b = \frac{1}{k} \sum_{a=1}^k (\mathbf{m}_a - \bar{\mathbf{m}})(\mathbf{m}_a - \bar{\mathbf{m}})^T \quad (2.41)$$

$$\text{where } \bar{\mathbf{m}} = \frac{1}{k} \sum_{a=1}^k \mathbf{m}_a \quad (2.42)$$

for k classes, with $k = 2$ in our example of the unprimed and primed class a and a' (Fig. 2.8). Moreover, we need a *within-cluster covariance matrix*

$$\Sigma_w = \frac{1}{k} \sum_{a=1}^k \Sigma_a \quad (2.43)$$

Now a normalized separability measure between the k clusters can be defined as

$$d_S^2 = \text{tr } \Sigma_b / \text{tr } \Sigma_w \quad (2.44)$$

giving the mean cluster separation $\text{tr } \Sigma_b$ relative to the mean cluster extension $\text{tr } \Sigma_w$. As stated above, the measure is invariant against rotation of the coordinate system. However, the coordinate axes might also be rotated and *rescaled* by a transformation $\mathbf{x} \mapsto \mathbf{T}\mathbf{x}$ with a matrix \mathbf{T} , so that $\mathbf{T}^T\mathbf{T} = \mathbf{D}$ is diagonal but not equal to the identity matrix: $\mathbf{T}^T\mathbf{T} \neq \mathbf{I}$. Then the measure d_S^2 is no longer invariant, because $\text{tr}(\mathbf{T}\Sigma_a\mathbf{T}^T) \neq \text{tr } \Sigma_a$. Therefore we consider a second normalized separability measure between the k clusters:

$$d_S^{2*} = \text{tr}(\Sigma_b\Sigma_w^{-1}) \quad (2.45)$$

which is invariant against rotation and scaling of the axes of the coordinate system. The invariance of d_S^{2*} can be easily seen, since the distance d_S^{2*} in the transformed space can

be obtained from the original distance d_S^{2*} as

$$d_S^{2*' } = \text{tr} \left((\mathbf{T}\boldsymbol{\Sigma}_b\mathbf{T}^\text{T})(\mathbf{T}\boldsymbol{\Sigma}_w\mathbf{T}^\text{T})^{-1} \right) \quad (2.46)$$

$$= \text{tr} \left(\mathbf{T}\boldsymbol{\Sigma}_b\mathbf{T}^\text{T}(\mathbf{T}^\text{T})^{-1}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}^{-1} \right) \quad (2.47)$$

$$= \text{tr} \left(\mathbf{T}\boldsymbol{\Sigma}_b(\mathbf{T}^{-1}\mathbf{T})^\text{T}\boldsymbol{\Sigma}_w^{-1}\mathbf{T}^{-1} \right) \quad (2.48)$$

$$= \text{tr} \left(\mathbf{T}\boldsymbol{\Sigma}_b\boldsymbol{\Sigma}_w^{-1}\mathbf{T}^{-1} \right) \quad (2.49)$$

with the general trace property $\text{tr}(AB) = \text{tr}(BA)$

$$= \text{tr} \left(\boldsymbol{\Sigma}_b\boldsymbol{\Sigma}_w^{-1}\mathbf{T}^{-1}\mathbf{T} \right) \quad (2.50)$$

$$= \text{tr} \left(\boldsymbol{\Sigma}_b\boldsymbol{\Sigma}_w^{-1} \right) \quad (2.51)$$

$$= d_S^{2*} \quad (2.52)$$

and remains indeed unchanged.

We try to illustrate the difference between the separability measures d_S^2 (Eq. 2.44) and d_S^{2*} (Eq. 2.45) in Fig. 2.9 on the following page. We imagine two clusters with a distance \mathbf{d} and the covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}'$. In the first case we show mainly orthogonal variance, in the other case mainly parallel variance. The cluster separability measure d_S^2 (Eq. 2.44) is the same for both cases, whereas the other cluster separability measure d_S^{2*} (Eq. 2.45) is smaller for the parallel case (left), and thus indicates a worse separability for the parallel case.

We see that the separability measure d_S^{2*} is more sensitive and invariant against rotation and rescaling of the coordinate system axes. It is the separability measure which is relevant for classification performed with the Mahalanobis or maximum likelihood distances (Eq. 2.28 and Eq. 2.34). Separability measure d_S^2 is still worthwhile to consider, as it is the relevant one for most unsupervised clustering algorithms (see Section 2.6) which use the Euclidean distance (Eq. 2.25).

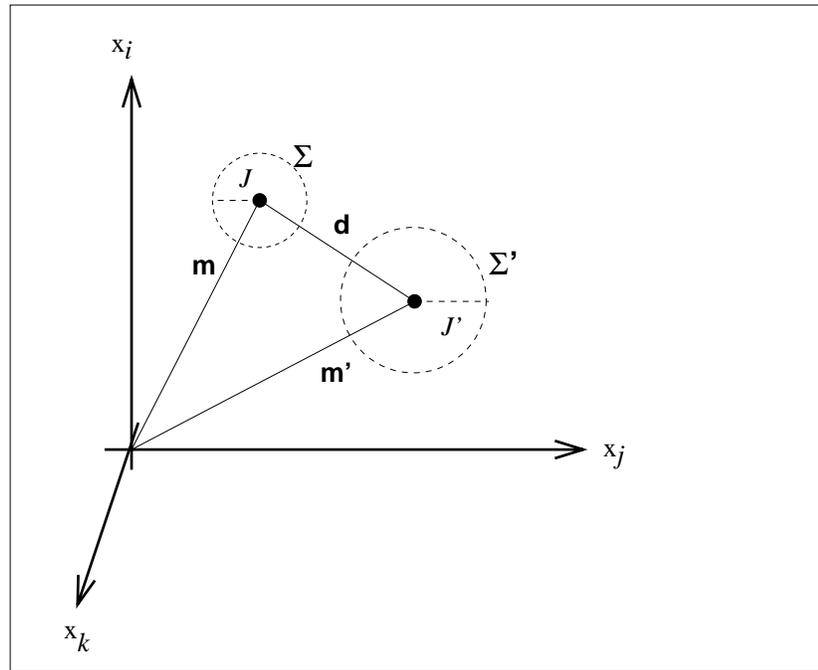


Figure 2.8: 3D projection of the feature space with two clusters, the distance between the cluster centers, and the cluster radii $J = \sqrt{\text{tr } \Sigma}$.

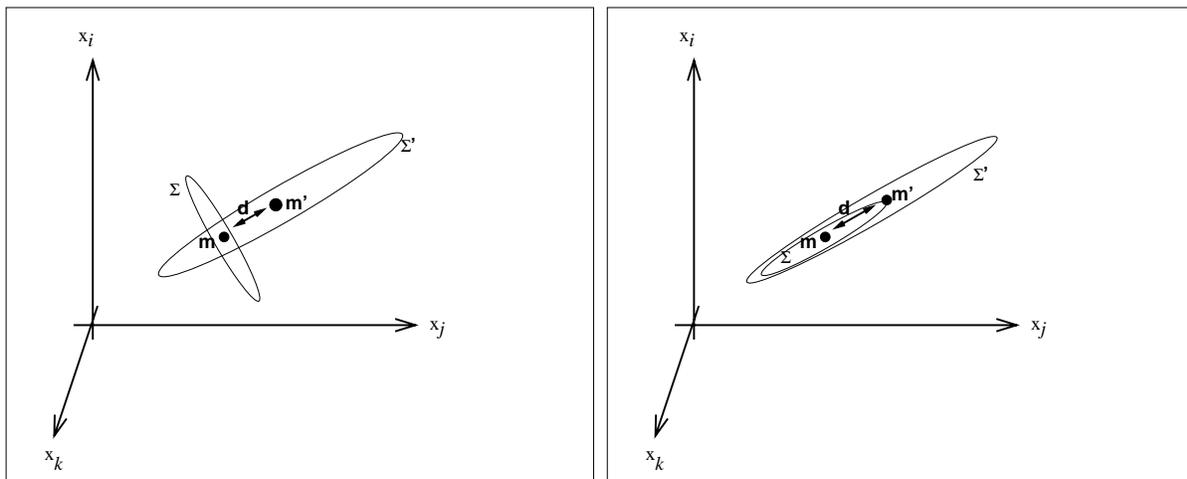


Figure 2.9: Clusters with mainly orthogonal and mainly parallel variance: the distance d and the covariance matrices Σ and Σ' are the same in both figures. The cluster separability measure d_S^2 (Eq. 2.44) is thus the same for both cases, whereas the other cluster separability measure d_S^{2*} (Eq. 2.45) is not and indicates better separability for the mainly orthogonal variance (left).

2.5 Supervised Classification

The *classification* or *labeling* of a multispectral image means that a class identifier a (label) is assigned to each pixel. Belonging to a certain class is usually derived by analysis of either the *spatial* context of the respective pixel (texture²), or the spectral context (spectral signature), or both. Throughout this thesis we concentrate on the pixel-wise, purely spectral classification as the most wide-spread method in multispectral remote sensing (see Chapter 6).

The underlying idea of *parametric classification* is that a number of k class prototypes are represented by a certain set of parameters (*e.g.* cluster center (mean spectrum) and covariance matrix). The feature vector of each pixel is compared to all class prototypes by means of a well defined distance, in our case the spectral distance d^2 (see Section 2.3 on page 23). Finally the pixel is assigned (labeled) to the class where the distance becomes minimal.

Supervised classification then means that the set of classes (represented by their respective parameters) is provided by the analyst, usually based either on eye appraisal of the image or on ground reference. The analyst will determine the necessary class parameters from *training areas* which are characteristic samples for the conceptual classes the operator wishes to use. Having the classes predefined, the classification then is a one-pass process, its computational effort depending only linearly on the number of pixels, the number of spectral bands N and the number of classes k .³

In the simplest form all spectra are compared to each spectral class. This however can be modified into the concept of hierarchical classification where a decision tree is worked down for each spectrum. This will decouple the linear dependence of the classification cost on the number of classes. Even for the simple non-hierarchical classification the run time can be optimized by ordering the spectral comparisons and using partial sum logic (Venkateswarlu & Singh 1995).

2.6 Unsupervised Classification / Clustering

In contrast to supervised classification, where the class parameters are trained on areas which are specified by the analyst, the *unsupervised* classification tries to estimate the class parameters automatically from the overall input data itself. After the parameter estimation, a final classification takes place as described above. Unsupervised classification will be applied in Chapter 5 and Chapter 6.

The task of the unsupervised classification algorithms is to discover structures which are inherent in the input data, more specifically, to identify clusters in the feature space. The underlying idea is that each feature vector (spectrum) is a point in the feature space, and that the points of related feature vectors form separable clusters in the feature space. The process of *clustering* then means to find these clouds of points in the feature space, in particular to find the cluster centers and to quantify the scatter around them. The result of clustering is then a set of statistically meaningful classes which can be attributed to conceptual classes only *a posteriori* by the analyst. Statistical

² The *texture* of a pixel's local neighborhood in a gray-value image is characterized by features such as 'smoothness', 'coarseness', 'graininess', directional 'co-occurrences' etc. (Gonzalez & Wintz 1987).

³ Provided that the used distance is an inner product induced norm such as described in Section 2.3.

classes may comprise quite different spectral sets than conceptual classes. E.g. for the here considered case of multispectral image classification, clusters with spectrally similar pixels do not necessarily coincide with functionally similar pixels derived from concepts such as street, house, vegetation or water.

In the context of data compression, unsupervised classification is also called *vector quantization*, as the large number of feature vectors is coded into a small number of prototypes contained in a *code book* (the set of class centers).

Main methods for grouping points in the feature space into clusters are (Duda & Hart 1973, Hartigan 1975, Bezdek 1981):

► **Hierarchical splitting / merging:**

The data points are iteratively split into subsets (top down), or the most similar points and then subsets are merged (bottom up), or both (hybrid).

► **Feature space density estimation:**

The feature space density can be estimated by virtue of Parzen windows. The feature space is divided into bins and a multivariate histogram is computed from all feature vectors. Then the bins which are local maxima in the feature space are designated as class centers.

► **Geometrical clustering:**

After tentative classification, the geometric centroids of the clusters in the feature space are computed in order to minimize the scatter around each cluster center.

Practically all algorithms work iteratively. In this thesis we deploy the widely used iterative geometrical clustering of multispectral data (Duda & Hart 1973, Hartigan 1975, Richards 1993). The number of classes (clusters) k is predefined by the analyst, and the clustering procedure then iteratively finds the coordinates of the cluster centers which are distributed throughout the feature space. The ‘best’ partition (classification) of the data points into the k classes is defined as the partition which minimizes a certain functional. A most commonly chosen functional is the overall mean squared errors (MSE)

$$J^2 = \frac{1}{n} \sum_a n_a J_a^2 \quad (2.53)$$

with J_a^2 being the scatter matrix of a class a as defined in Eq. 2.35 on page 25, n_a the number of data points belonging to class a , and the total number of points $n = \sum_a n_a$. So J^2 is the average combined scatter of all data points around their respective class centers. To find the partition which yields the global minimum of the functional J^2 is computationally very demanding, since the number of possible partitions of $n = \sum_a n_a$ data points rises as fast as $k^n/k!$.

■ **Hard k -means:**

Among the sub-optimal but computationally feasible algorithms (which guaranty convergence only to a local minimum), the family of ‘ k -means’ algorithms (MacQueen 1967) is the most wide-spread. They are also called ‘centroid clustering’, ‘migrating means’, or ‘basic ISODATA’ (Ball & Hall 1967). An initial classification must be provided, if

possible ‘close’ to the assumed global solution, or, if an educated estimate is unavailable, at random. Then the iteration process starts. In the j th iteration all points \mathbf{x} are assigned to the class a with the least distance $d^2(\mathbf{x}, \mathbf{m}_a)$ between the point \mathbf{x} and the current cluster center \mathbf{m}_a . Then the k cluster centers \mathbf{m}_a are recomputed as the centroid (*i.e.*, the mean value in each spectral band, Eq. 2.37 on page 25) of all vectors currently assigned to it. The justification for this lies in the fact that the mean \mathbf{m}_a is indeed the very point which minimizes the within-class scatter $n_a J_a^2$ (Eq. 2.35).

The procedure is repeated with the new means (cluster centers) and so forth. The means are expected to converge eventually at the true cluster centers, although only convergence to a local minimum can be proven. The iteration is stopped when the change in classification or in cluster center coordinates is below a fixed threshold.

When the clusters are characterized only by their means \mathbf{m}_a , this implicitly assumes that the class members are scattered concentrically around the class center, *i.e.*, k -means clustering will ‘seek’ for hyperspheres. Also an adaptive k -means algorithm has been investigated (Gustafson & Kessel 1979) where at each iteration a covariance matrix Σ_a is computed from the currently assigned vectors and used in a Mahalanobis distance (Eq. 2.28 on page 24) to the cluster center. This procedure is of course much more time consuming, and convergence can be shown only under the boundary condition of a fixed volume of the class’s hyperellipsoid (which is proportional to the determinant of the covariance matrix $|\Sigma_a|$).

■ Fuzzy k -means:

A fuzzy generalization of k -means, the *fuzzy k -means*, has been introduced by Bezdek (1973). Here each vector is not assigned entirely to a single class, but a fuzzy membership $0 < w_{a\mathbf{x}} < 1$ gives the degree of membership of a point \mathbf{x} to a class a . The memberships are normalized as $\sum_a w_{a\mathbf{x}} = 1$ for a given point \mathbf{x} . The membership is computed as a function of the distances $d_{a\mathbf{x}}$ in the feature space between the point \mathbf{x} and the cluster centers \mathbf{m}_a

$$w_{a\mathbf{x}} = \left[\sum_j \left(\frac{d_{a\mathbf{x}}}{d_{j\mathbf{x}}} \right)^{2/(z-1)} \right]^{-1} \quad (2.54)$$

where z controls the fuzzyness. The parameter $z > 1$ is often set to $z = 2$. In the limiting case of $z \rightarrow 1$, the fuzzy k -means degenerates to the previously discussed hard k -means. A column vector $\mathbf{w}_{\mathbf{x}} = [w_{1\mathbf{x}} \dots w_{k\mathbf{x}}]^T$ contains the membership $w_{a\mathbf{x}}$ of the point \mathbf{x} to all k classes.

For the final classification a *defuzzification* is necessary. The vector is assigned to the class where membership is highest, possibly requiring a minimum membership threshold (α -cutoff). This corresponds to the simple minimum distance classification as described above.

The computational effort of the fuzzy k -means is higher than for the hard k -means, because the membership matrix $\mathbf{W} = [.. \mathbf{w}_{\mathbf{x}} ..]$ which contains the membership vectors $\mathbf{w}_{\mathbf{x}}$ for all observed vectors \mathbf{x} , has to be updated throughout the iteration process. The main advantage of the fuzzy k -means lies in the fact that via fuzzy membership the whole iterative process of minimizing the objective function becomes continuous differentiable. Convergence to a local minimum can be proven (Bezdek 1981). Points do not ‘flip’ from class to class, but rather change their membership weights gradually, ensuring a

smooth convergence. Reliability of the classification can be estimated from the absolute height of the maximum membership $w_{a\mathbf{x}}$, since the membership is a normalized value. Separability of classes can be estimated from the off-diagonal elements of a similarity matrix ($\mathbf{W}\mathbf{W}^T$).

Fuzzy k -means clustering has been applied to remotely sensed multispectral imagery (LANDSAT) by (Trivedi & Bezdek 1986).

■ Cluster Validation:

A serious problem with all clustering algorithms is to set the number of classes k which is usually provided by the analyst. Bezdek (1981) points out that this so called *cluster validation* is a hard problem and depends, after all, crucially on the definition of what a ‘good’ cluster is. A promising way to determine a sensible number of classes k , resting on a maximum entropy approach, is presented *e.g.* by Buhmann & Kühnel (1993).

2.7 Principal Component Transformation

The *principal component transformation* (PCT, in computer vision also known as the Karhunen-Loève transform) is a common technique in order to obtain overall uncorrelated feature vectors. We will investigate PC-transformed feature spaces in Chapter 6.

With multispectral image data the correlation between the spectral bands can be quite strong, so that the spectral values are essentially redundant to a certain degree. Significant correlation between the spectral bands is indicated by large entries in the off-diagonal of the overall covariance matrix Σ :

$$\Sigma = \frac{1}{n} \sum_{\mathbf{x} \in \mathbf{X}} (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \quad (2.55)$$

$$\text{with } \mathbf{m} = \frac{1}{n} \sum_{\mathbf{x} \in \mathbf{X}} \mathbf{x} \quad (2.56)$$

where \mathbf{X} is the set of all n vectors \mathbf{x} in the image. The basic idea of the PCT is that the high dimensional vectors \mathbf{x} can be represented by linear combinations of only few *principal components*.

The foundation of the PCT is that the symmetric covariance matrix Σ of all observed data vectors \mathbf{x} can be decomposed⁴ into $\Sigma = \mathbf{E}\mathbf{D}\mathbf{E}^T$, where \mathbf{E} is an orthogonal matrix: $\mathbf{E}^T\mathbf{E} = \mathbf{I}$. The matrix \mathbf{E} contains the eigenvectors of Σ as columns, and \mathbf{D} is a diagonal matrix containing the corresponding eigenvalues σ_{ii}^2 in order of descending magnitude. The eigenvalues represent the variances which the data has after rotation into an uncorrelated coordinate system $\mathbf{x} \mapsto \mathbf{E}^T\mathbf{x}$.

Here we want to define the PCT as follows: First, the mean vector \mathbf{m} is subtracted of all observed vectors \mathbf{x} , so that the cluster becomes centered around the origin. Then

⁴ Singular value decomposition (SVD) decomposes a given matrix Σ into $\Sigma = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where \mathbf{D} is a quadratic diagonal matrix containing the singular values of Σ in order of decreasing magnitude. For a real and symmetric matrix Σ (as any covariance matrix is by definition) we know that $\mathbf{U} = \mathbf{V} = \mathbf{E}$, where \mathbf{E} is a quadratic matrix that contains the eigenvectors of Σ as columns, and \mathbf{D} contains the corresponding eigenvalues of Σ (Hämmerlin & Hoffmann 1994). The SVD is computed by virtue of Householder reflections and provided as a standard matrix operation by most mathematical software packages.

the vectors are projected onto the ortho-normal eigenvectors by multiplication with \mathbf{E}^T . After this rotation of the coordinate system the new vector entries are uncorrelated, in other words, the covariance matrix of the transformed vectors has only entries in the diagonal. Finally, the variance along each of the new coordinate axes is forced to unity by dividing each component by the square root of the respective eigenvalues:

$$\mathbf{x} \mapsto \mathbf{x}' = \mathbf{D}^{-\frac{1}{2}} \mathbf{E}^T (\mathbf{x} - \mathbf{m}) \quad (2.57)$$

$$= \mathbf{T} (\mathbf{x} - \mathbf{m}) \quad (2.58)$$

with the PCT-matrix $\mathbf{T} = \mathbf{D}^{-\frac{1}{2}} \mathbf{E}^T$. Note that \mathbf{T} is not an orthogonal matrix, $\mathbf{T}^T \mathbf{T} \neq \mathbf{I}$, because it presents a rotation and a rescaling of the coordinate system.

Generally, with N entries for each vector \mathbf{x} , the dimensionality of the data is N . But if some eigenvalues σ_{ii}^2 of $\mathbf{\Sigma}$ are vanishingly small, then the observed vectors are essentially confined to a subspace which is of a lower dimension $H < N$. Practising *lossy data compression*, the corresponding dimensions of the data space may be neglected. Then $N \times N$ matrix \mathbf{T} may be reduced to a $H \times N$ matrix by leaving out the eigenvectors with small corresponding eigenvalues. Then the matrix \mathbf{T} is only of rank $\text{rank}(\mathbf{T}) = H < N$, and also the set of transformed vectors $\mathbf{x}' = \mathbf{T}\mathbf{x}$ will then only be of dimension H .

It is as well possible to use the correlation matrix instead of the covariance matrix to determine the eigenvectors and -values and thus the principle components. The correlation matrix entries C_{ij} are obtained from the covariance matrix entries Σ_{ij} as: $C_{ij} = \Sigma_{ij} / \sqrt{\sigma_{ii}^2 \sigma_{jj}^2}$. Mather (1987) and Singh (1993) have pointed out that there is no simple relationship between these two options (which are also called *unstandardized* and *standardized principal components analysis*, (Fung & LeDrew 1987, Eklundt & Singh 1993)).

Other comparable techniques of compressing vector data are Factor Analysis (FA) (Harris 1975), and Maximum Noise Fractions (MNF) transformation (Nielsen & Larsen 1994, Hurcom et al. 1994).

Color Constant Classification in the Multispectral Feature Space

Outline of this chapter – We define a dichromatic illumination model which accounts for direct sun light and diffuse skylight. Combining the dichromatic illumination with Lambertian reflection we develop the concept of classes as two dimensional planes in the feature space. An adequate color constant distance measure is derived, using a linear algebraic projector matrix. We show that this projector matrix can be seen as similar to a change of metric, and that it can be combined with the covariance matrix into the common Mahalanobis distance at equal computational cost.

Then we investigate three questions which arise as consequences from the concept of planar classes. We show how to determine the proximity of planes and thus the probability of misclassification. We show that all class planes must intersect at the origin of the feature space, and how we can recover residual path radiances in case they have not been subtracted correctly. Finally, we show that it is possible to recover the normalized Lambertian reflectance spectrum from the plane which is spanned in the feature space by the set of observed spectra from a specific surface under varying illumination.

3.1 Dichromatic Illumination Model

For multispectral remote sensing applications as well as for outdoor scenes in computer vision it is common to model the global illumination onto a horizontal surface $E_{\text{glob}} = E_{\text{dir}} + E_{\text{diff}}$ as a *dichromatic illumination* with two basic sources:

1. **direct illumination** E_{dir} (sunlight) from the sun approximated as a point source,
2. **diffuse illumination** E_{diff} (skylight) from the whole remaining upper hemisphere (distributed source).

The diffuse illumination E_{diff} is generated by sunlight scattered on air molecules (diameter < 50 nm, Rayleigh-scattering) and aerosols (diameter ≈ 500 nm, Mie-scattering).

Direct and diffuse illumination have distinct spectra, the ratio of which is known from atmospheric physics and can be well approximated as:

$$\frac{E_{\text{diff}}}{E_{\text{glob}}} \hat{=} \frac{m}{n+m} = \hat{m} \propto \lambda^{-\gamma} \quad (3.1)$$

with a typical exponent $\gamma = [0.5 \dots 4]$, where $\gamma = 4$ corresponds to Rayleigh scattering, $\gamma = 1$ to Mie scattering, and $\gamma = 0$ to a scatter function of geometric optics, *i.e.*, wavelength independent optics (particle diameter $> 5\mu\text{m}$, Gerthsen et al. (1977), Minnaert (1993)). Various diffuse to global ratio spectra are plotted in Fig. 3.1. This exponential wavelength dependence of the diffuse sky irradiance can be found *e.g.* in Sabins (1978), Mather (1987), Kraus & Schneider (1988), Chavez (1989), Warnecke (1991), and Richards (1993). Schott (1993) cites Piech & Walker (1974) for a method to measure the relative diffuse illumination from the image data itself, namely from transitions of sunlit to shadowed areas of the same surface. This has been used on multispectral image data in order to determine the exponential constant γ by Wiemker & Hepp (1994). An exponential fit to the *in scene* measured spectra yielded $\gamma = 1.7 \pm 0.3$ for that particular recording day, and simulations with the atmospheric correction package SENSAT (Richter 1992) yielded $\gamma = 1.4$. Another simulation was run with the radiative transfer code MODTRAN (Anderson et al. 1995) for various visibilities (Fig. 3.2; simulation with ‘mid-latitude sommer atmosphere’, ‘rural aerosol model’, for a solar zenith angle of $\theta_{\odot} = 40^{\circ}$ (*i.e.*, *e.g.* September 1, 12h⁰⁰ CET, $\approx 50^{\circ}$ northern latitude)). For a given atmospheric parameter γ , the magnitude of the diffuse contribution depends on the visibility.

Obviously the value of γ for a particular recording situation depends on the actual aerosol density and on the density of bigger particles (dust, soot). For common atmospheric situations the coefficient is $\gamma \in [0.7..2.0]$ (Kraus & Schneider 1988). Warnecke (1991) gives $\gamma = 1.3$ as a typical exponent.

For tilted surfaces the contributions of direct and diffuse illumination change with surface orientation. We find different angular dependencies for point and distributed sources respectively. As many small scale and man made objects are too small to be considered in a digital elevation model (DEM, also digital terrain model, DTM), the surface orientation of a given surface patch often is not available. Consequently we cannot determine the mixture of the two illumination contributions to which a given surface patch is exposed. Also, for man made objects in aerial imagery we cannot assume smooth and continuous surfaces.

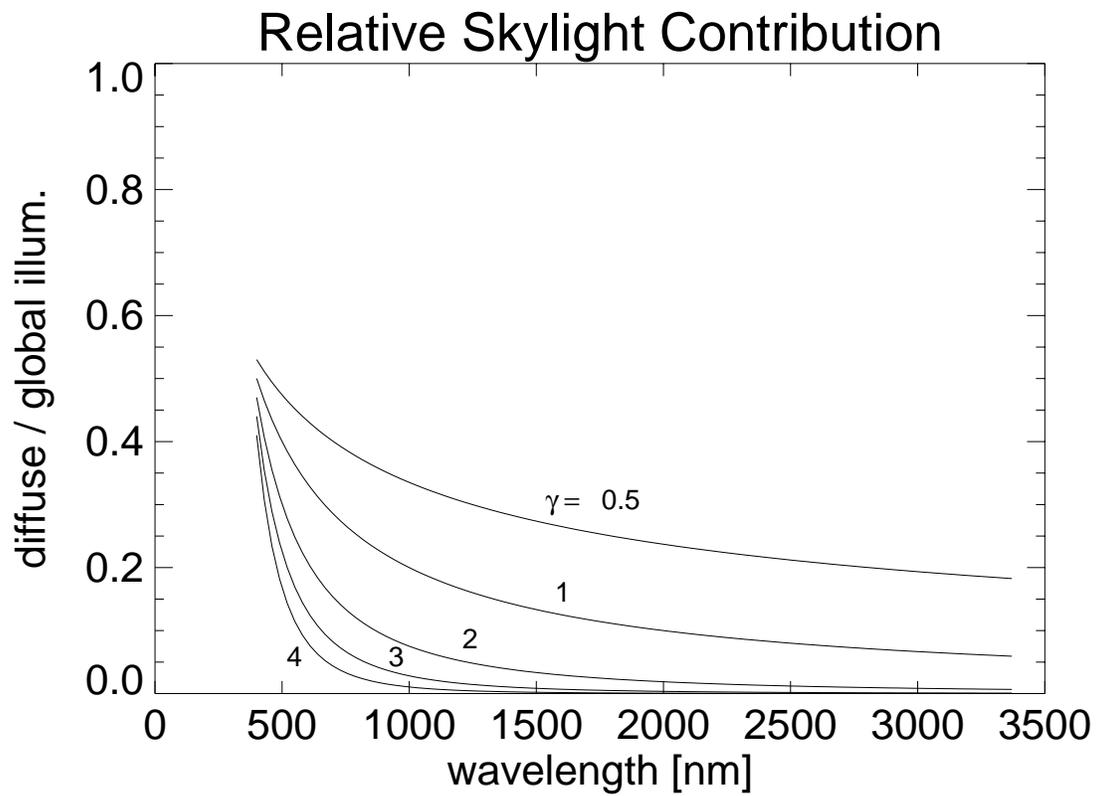


Figure 3.1: *The ratio of diffuse to global illumination modeled as an exponential function of wavelength $\hat{m} \propto \lambda^{-\gamma}$, plotted with $\gamma = [0.5, 1, 2, 3, 4]$.*

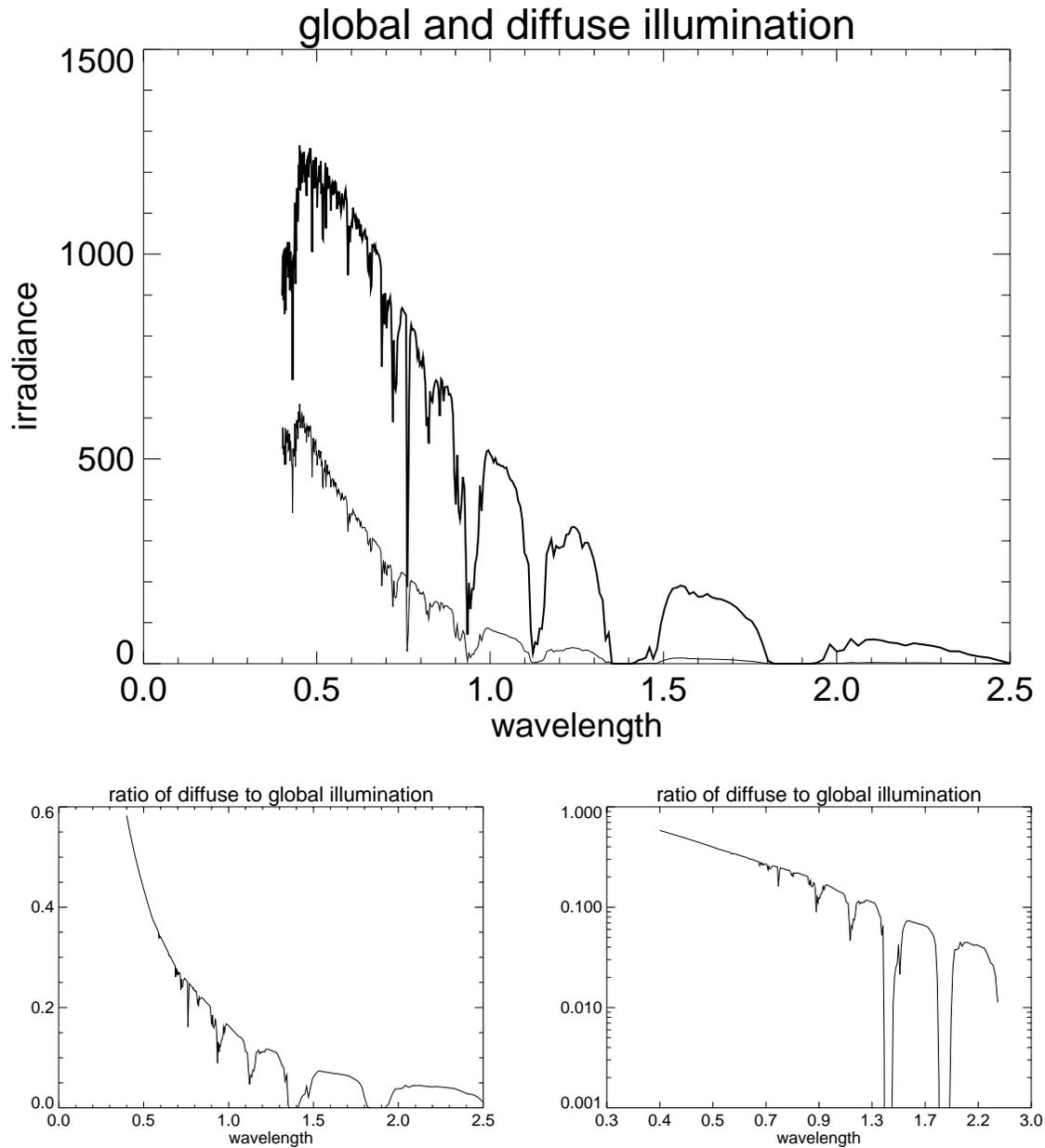


Figure 3.2: Top: Global irradiance (= direct + diffuse, higher curve) and diffuse irradiance (lower curve) obtained from MODTRAN simulation (wavelength in μm , irradiance in $\text{W}/(\text{m}^2\mu\text{m})$).

Bottom: The ratio of diffuse to global illumination. Right: Logarithmic representation. The atmospheric parameter for the chosen atmospheric model is $\gamma = 1.5$, whereas the magnitude of the diffuse contribution depends on the atmospheric visibility:

CONTRIBUTION OF DIFFUSE ILLUMINATION				
visibility [km]	5	10	15	20
diffuse to global ratio at $\lambda = 0.5\mu\text{m}$	80%	62%	51 %	44%

3.2 Planar Spectral Classes

We consider a sensor with N spectral bands which observes a homogeneous surface patch. The surface can have varying surface orientations (tilts). We now ask for the illumination e_i which is incident on the surface in each spectral band i . Neglecting atmospheric attenuation and Lambertian reflection provided, the sensor will receive spectral radiances $x_i = r_i e_i$ which are simply proportional to the incident illumination e_i , where r_i is the Lambertian reflectance. So if the incident illumination e_i changes due to varying surface orientation of the surface patch, the radiances received by the sensor will change as well. Let all possible reflected spectra x_i from this particular surface form a *spectral class* a . We will determine the possible spectra a_i in this class.

For reasons of simplicity, we will use the dimensionless constants n_i and m_i for the direct and diffuse illumination $E_{\text{dir}}(\lambda_i)$ and $E_{\text{diff}}(\lambda_i)$ for spectral band i . For the N spectral bands this combines into the spectral vectors \mathbf{n} and \mathbf{m} . With $\mathbf{n} = [\dots n_i \dots]^T$ the spectrum of the sunlight and $\mathbf{m} = [\dots m_i \dots]^T$ the spectrum of the skylight, the dichromatic illumination model states that the illumination of a flat surface is given – apart from a dimension factor $[W/(m^2 \mu m)]$ – by

$$\mathbf{e} = \nu \mathbf{n} + \mu \mathbf{m} \quad , \quad \nu, \mu \in \mathbb{R}_0^+ \quad (3.2)$$

$$\begin{bmatrix} \vdots \\ e_i \\ \vdots \end{bmatrix} = \nu \begin{bmatrix} \vdots \\ n_i \\ \vdots \end{bmatrix} + \mu \begin{bmatrix} \vdots \\ m_i \\ \vdots \end{bmatrix} \quad (3.3)$$

$$= \begin{bmatrix} \vdots & \vdots \\ n_i & m_i \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \nu \\ \mu \end{bmatrix} \quad (3.4)$$

$$\mathbf{e} = \mathbf{E} \mathbf{c} \quad (3.5)$$

with \mathbf{E} containing \mathbf{n} and \mathbf{m} as column vectors and the contributions $\mathbf{c} = (\nu, \mu)^T$. The factors ν and μ are dimensionless contribution factors from the positive real numbers including zero \mathbb{R}_0^+ .

Let the surface have a Lambertian reflectance spectrum $\mathbf{r} = [\dots r_i \dots]^T$. As we want to concentrate on the ‘form’ of the spectral signature and neglect its brightness (*i.e.*, the absolute magnitude of the reflectance spectrum), only the *normalized reflectance* $\hat{\mathbf{r}} = \mathbf{r}/\|\mathbf{r}\|$ is meaningful for our purpose. Then for a spectral class a all possible spectra \mathbf{a} reflected from a given surface under varying orientation are

$$a_i = \hat{r}_i e_i \quad (3.6)$$

$$= \hat{r}_i (\nu n_i + \mu m_i) \quad (3.7)$$

$$= \nu \hat{r}_i n_i + \mu \hat{r}_i m_i \quad (3.8)$$

or in vector/matrix notation:

$$\begin{bmatrix} \vdots \\ a_i \\ \vdots \end{bmatrix} = \nu \begin{bmatrix} \vdots \\ \hat{r}_i n_i \\ \vdots \end{bmatrix} + \mu \begin{bmatrix} \vdots \\ \hat{r}_i m_i \\ \vdots \end{bmatrix} \quad (3.9)$$

$$\mathbf{a} = \mathbf{A} \mathbf{c} \quad (3.10)$$

where we have stacked the reflected direct and diffuse components column-wise into a $N \times 2$ matrix \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} \vdots & \vdots \\ \hat{r}_i n_i & \hat{r}_i m_i \\ \vdots & \vdots \end{bmatrix} \quad (3.11)$$

with the contributions $\mathbf{c} = (\nu, \mu)^\top$.

All vectors $\mathbf{a} = \mathbf{A}\mathbf{c}$ lie in a two-plane, *i.e.*, a two dimensional linear subspace of the feature space \mathbb{R}^N . The subspace is spanned by the two column vectors in \mathbf{A} . The two degrees of freedom reflect the arbitrary contributions ν, μ of the two light sources (direct and diffuse, for illustration see Fig. 3.3 on the next page). This concept has been investigated for the three dimensional case of RGB-colors (Shafer 1985). For two point sources in RGB-space, Drew & Kontsevich (1994) cite Petrov to have shown that all possibly observed spectra \mathbf{x} from Lambertian reflection lie on an ellipse. However, for the distributed skylight the diffuse contribution μ is not simply connected to the inner product of surface normal and direction of incidence as for a point source, and the two-plane in the multispectral \mathbb{R}^N feature space is constrained only by the condition of $\nu, \mu \geq 0$.

Consequently, in the presence of direct sunlight and diffuse skylight, the spectral class formed by a homogeneous ground surface patch under varying surface orientation must not be represented by a cluster center as a single point (or direction, in case of spectral angle classification) in feature space, but rather by the respective two-plane spanned by reflected direct and diffuse illumination.

For several distinct surface types, the respective Lambertian reflectance spectrum $\hat{\mathbf{r}}$ will be different, whereas the illumination spectra \mathbf{n} and \mathbf{m} remain the same, of course. Each surface type will span a distinct class plane in the feature space, determined by the specific reflectance $\hat{\mathbf{r}}$ and the illumination spectra \mathbf{n} and \mathbf{m} (Eq. 3.10 on the preceding page).

In general, several such two-planes might but do not necessarily intersect in \mathbb{R}^N . In the spectral feature space as discussed here, however, all class planes intersect at the origin. This happens when no irradiance is incident on the reflecting surface patch whatsoever and thus no reflected radiances are measured by the sensor (provided correct subtraction of path radiances).

From the concept of class planes it follows that a meaningful, relevant distance of an observed spectrum \mathbf{x} to a spectral class ($\hat{\mathbf{r}}, \mathbf{n}, \mathbf{m}$) is the distance between \mathbf{x} and its *proximum* on the class plane (Fig. 3.3).

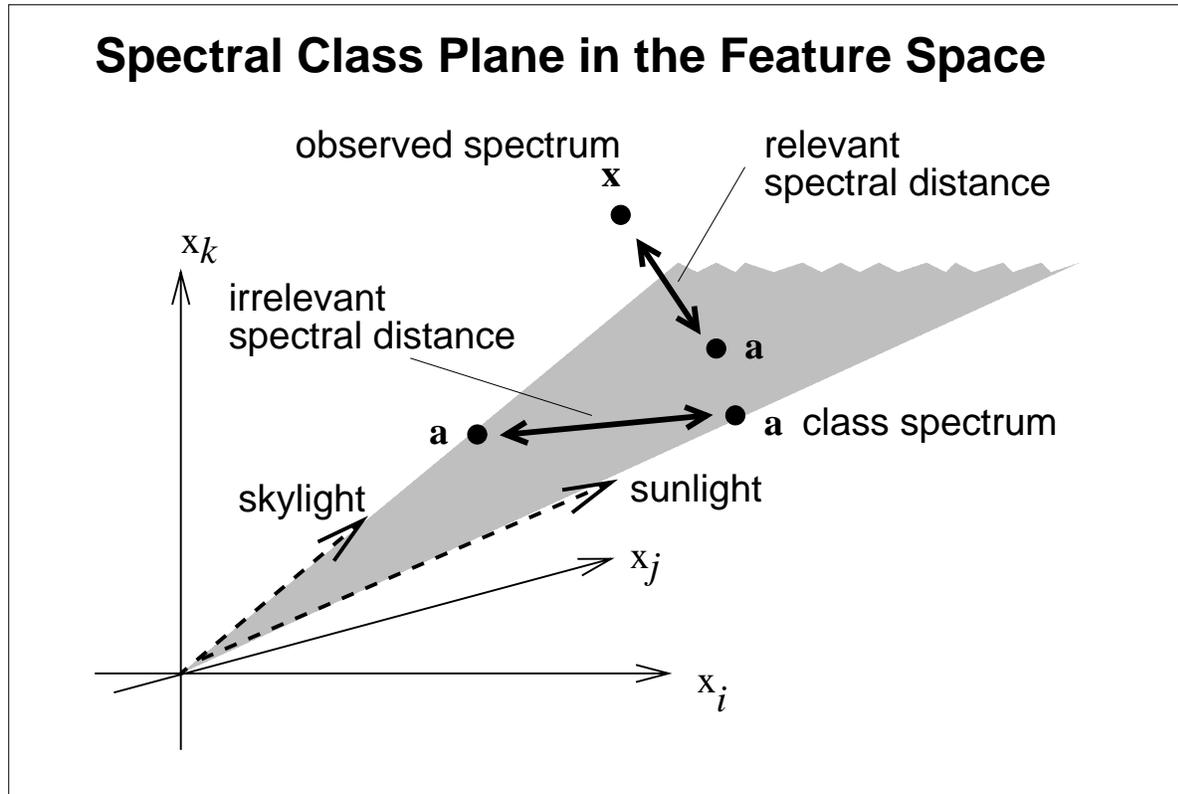


Figure 3.3: The class plane spanned by the reflectance multiplied with sunlight and skylight respectively. Distances parallel to the class plane are irrelevant and filtered out.

3.3 Color Constant Enhancement of the Mahalanobis Distance

In order to determine the distance between an observed spectrum \mathbf{x} and a given class plane \mathbf{A} (determined by $[\hat{\mathbf{r}}, \mathbf{n}, \mathbf{m}]$,¹ Eq. 3.10 on page 39), we need to find the point on plane \mathbf{A} which is closest to \mathbf{x} . Obviously this shortest distance vector between plane \mathbf{A} and point \mathbf{x} will be orthogonal to the plane. We thus consider the orthogonal projection of \mathbf{x} onto the plane:

The plane is formed by all vectors $\mathbf{a} = \mathbf{A}\mathbf{c}$ which can be formed by an arbitrary coefficient vector \mathbf{c} and a fixed matrix \mathbf{A} . Now we want to find the point $\mathbf{a}_0 = \mathbf{A}\mathbf{c}_0$ on the class plane closest to an observed spectrum \mathbf{x} :

$$\mathbf{x} \approx \mathbf{A}\mathbf{c}_0 \quad (3.12)$$

where the ‘approximately’ sign (\approx) means that the sum of squared deviations between \mathbf{x} and $\mathbf{A}\mathbf{c}_0$ is minimal:

$$\|\mathbf{x} - \mathbf{A}\mathbf{c}_0\| = \min \quad (3.13)$$

¹ In practice, the class plane may be determined by ground based reflectance measurements or by selection of spectral training areas in the imagery.

The least square solution for \mathbf{c} then is

$$\mathbf{c}_o = \mathbf{A}^+ \mathbf{x} \quad (3.14)$$

where the $N \times N$ matrix \mathbf{A}^+ is the pseudo-inverse (see Appendix A.1 on page 177). Thus the proximum to \mathbf{x} on the class plane is given by $\mathbf{A}\mathbf{c}_o = \mathbf{A}\mathbf{A}^+ \mathbf{x}$ which is also called its *orthogonal projection* (Appendix A.1 on page 177).

Then the distance vector between \mathbf{x} and its proximum on the plane $\mathbf{A}\mathbf{A}^+ \mathbf{x}$ becomes

$$\mathbf{d} = \mathbf{x} - \mathbf{A}\mathbf{A}^+ \mathbf{x} \quad (3.15)$$

$$= (\mathbf{I} - \mathbf{A}\mathbf{A}^+) \mathbf{x} \quad (3.16)$$

$$= \mathbf{R}\mathbf{x} \quad (3.17)$$

with \mathbf{I} the identity matrix and $\mathbf{R} = \mathbf{I} - \mathbf{A}\mathbf{A}^+$. The squared magnitude of the distance becomes

$$d^2 = \|\mathbf{R}\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{R}^T \mathbf{R} \mathbf{x} \quad (3.18)$$

$$= \mathbf{x}^T \mathbf{R} \mathbf{x} \quad (3.19)$$

We note that the matrix \mathbf{R} is symmetric, $\mathbf{R}^T = \mathbf{R}$:

$$\mathbf{R}^T = (\mathbf{I} - \mathbf{A}\mathbf{A}^+)^T \quad (3.20)$$

$$= \mathbf{I}^T - (\mathbf{A}\mathbf{A}^+)^T \quad (3.21)$$

$$= \mathbf{I} - \mathbf{A}\mathbf{A}^+ \quad \text{because } \mathbf{I} \text{ and } \mathbf{A} \text{ are symmetric matrices} \quad (3.22)$$

$$= \mathbf{R} \quad (3.23)$$

We also note that the matrix \mathbf{R} has the defining projector property, $\mathbf{R}\mathbf{R} = \mathbf{R}$:

$$\mathbf{R}\mathbf{R} = (\mathbf{I} - \mathbf{A}\mathbf{A}^+) (\mathbf{I} - \mathbf{A}\mathbf{A}^+) \quad (3.24)$$

$$= \mathbf{I} - 2\mathbf{A}\mathbf{A}^+ + \mathbf{A}\mathbf{A}^+ \mathbf{A}\mathbf{A}^+ \quad \text{with } \mathbf{A}\mathbf{A}^+ \mathbf{A} = \mathbf{A} \text{ (Appendix A.1)} \quad (3.25)$$

$$= \mathbf{I} - 2\mathbf{A}\mathbf{A}^+ + \mathbf{A}\mathbf{A}^+ \quad (3.26)$$

$$= \mathbf{I} - \mathbf{A}\mathbf{A}^+ \quad (3.27)$$

$$= \mathbf{R} \quad (3.28)$$

With the properties $\mathbf{R}^T = \mathbf{R}$ and $\mathbf{R}\mathbf{R} = \mathbf{R}$, the $N \times N$ matrix \mathbf{R} is another orthogonal projector. \mathbf{R} effectively filters out all those components of a vector \mathbf{x} which are contained as columns in \mathbf{A} .

The filter matrix \mathbf{R} can also be applied to any spectral distance \mathbf{d} . **The underlying idea is that \mathbf{R} filters out all components from the distance \mathbf{d} which are irrelevant because they can be explained by varying illumination conditions without a necessary change in reflectance (Fig. 3.3 on the preceding page). Here we mean in particular varying illumination due to a tilt of the observed ground surface patch.**

For a given class a we will designate the projector as \mathbf{R}_a . We may want to call \mathbf{R}_a a class specific color constancy projector, since it solves the color constancy problem in

the sense that after filtering observed spectra \mathbf{x}_a with \mathbf{R}_a they are invariant against the actual surface orientation of the observed surface patch.

In particular, \mathbf{R} can be applied to the Euclidean distance (Eq. 2.25 on page 23) $d^2 = \|\mathbf{d}\|^2 = \|\mathbf{x} - \mathbf{m}_a\|^2$ between an observed vector \mathbf{x} and a class center \mathbf{m}_a (mean vector of class a , Eq. 2.37 on page 25):

$$\mathbf{R}\mathbf{d} = \mathbf{R}(\mathbf{x} - \mathbf{m}_a) \quad (3.29)$$

$$= \mathbf{R}\mathbf{x} \quad (3.30)$$

where the projected distance of the mean vector \mathbf{m}_a of class a vanishes: $\mathbf{R}\mathbf{m}_a = 0$, obviously because \mathbf{m}_a is lying on the plane spanned by \mathbf{A} . So the distance becomes

$$d^2 = \mathbf{x}^T \mathbf{R}\mathbf{x} \quad . \quad (3.31)$$

We now want to apply \mathbf{R} also to the Mahalanobis distance (Eq. 2.28 on page 24)

$$d^2 = \mathbf{d}^T \boldsymbol{\Sigma}_a^{-1} \mathbf{d} \quad (3.32)$$

which uses the covariance $\boldsymbol{\Sigma}_a$ of a class a (Eq. 2.37).

Injecting a class specific projector \mathbf{R}_a this becomes

$$d^2 = \mathbf{d}^T \mathbf{R}_a^T \boldsymbol{\Sigma}_a^{-1} \mathbf{R}_a \mathbf{d} \quad (3.33)$$

$$= (\mathbf{x} - \mathbf{m}_a)^T \mathbf{R}_a^T \boldsymbol{\Sigma}_a^{-1} \mathbf{R}_a (\mathbf{x} - \mathbf{m}_a) \quad ; \quad \mathbf{R}_a \mathbf{m}_a = 0 \quad (3.34)$$

$$= \mathbf{x}^T \mathbf{R}_a^T \boldsymbol{\Sigma}_a^{-1} \mathbf{R}_a \mathbf{x} \quad (3.35)$$

which degrades to Eq. 3.19 for $\boldsymbol{\Sigma}_a = \mathbf{I}$ because of $\mathbf{R}^T \mathbf{R} = \mathbf{R}\mathbf{R} = \mathbf{R}$. Again, in Eq. 3.33 the matrix \mathbf{R} filters out all components from the distance \mathbf{d} which are irrelevant because they can be explained by varying illumination conditions without a necessary change in reflectance. After passing the filter, the remaining distance is then weighted with the inverse covariance as usual.

Note that the computational cost of the Mahalanobis distance is not increased by the color constant projector, since the metric $\mathbf{R}_a^T \boldsymbol{\Sigma}_a^{-1} \mathbf{R}_a$ needs to be determined only once for each class a and is equal for all tested spectra \mathbf{x} . Indeed, the computational cost is even slightly reduced because we do not need to compute the difference $\mathbf{d} = (\mathbf{x} - \mathbf{m}_a)$, as $\mathbf{R}\mathbf{m}_a$ vanishes anyway.

3.4 Some Derivatives of the Concept of Class Planes

Three questions arise from the above developed concept of class planes which we want to address briefly in this section and refer to the appendix for details.

■ Retrieval of the Reflectance From The Class Plane

Given a set of observations \mathbf{x} of the same surface under varying illuminations, these points \mathbf{x} will form a two-plane in the feature space. The question arises if and how we can retrieve the reflectance \mathbf{r} from the set of observed spectra \mathbf{x} . It turns out that the reflectance \mathbf{r} can indeed be retrieved up to an overall scaling factor, so that we get at least a normalized reflectance $\hat{\mathbf{r}} = \mathbf{r}/\|\mathbf{r}\|$, provided that the illumination spectra \mathbf{n} and \mathbf{m} are known or can be extracted from a white Lambertian reference target. The necessary calculations in order to retrieve the reflectance $\hat{\mathbf{r}}$ are set forth in Appendix A.3 on page 180. The computational cost is negligible since it is an analytic solution. Note that linear algebraic solutions of this kind are known to be sensitive to noise. The feasibility of the approach set out here has been demonstrated on experimental spectral data by Wiemker (1995c).

■ Minimal Distance Between Two Class Planes

In order to comment on the separability of two class planes spanned by \mathbf{A} and \mathbf{A}' , say, we want a measure for the minimal distance between those two class planes. This measure is important, firstly, to decide if two class planes of two brightness normalized spectral signatures $\hat{\mathbf{r}}$ and $\hat{\mathbf{r}}'$ are separable at all, and secondly, to assess how large the chance of misclassification between these two class planes is. Obviously the distance between all possible class planes vanishes at the origin of the feature space, when no illumination is available at all and all spectral values are zero (complete shadow). But a meaningful distance measure between the class planes spanned by \mathbf{A} and \mathbf{A}' can be given by means of the *condition* of the system of column vectors in \mathbf{A} and \mathbf{A}' (for a derivation we refer to Appendix A.4 on page 182). This condition can be easily computed from the *singular values* of the matrix $[\mathbf{A}, \mathbf{A}']$. Singular value decomposition is a standard matrix operation.

■ Retrieval of the Path Radiance From Two Class Planes

We have stated that all class planes of different surface types intersect at the origin, *i.e.*, at vanishing illumination. This is true provided the path radiance was properly subtracted beforehand by atmospheric correction (Section 2.1 on page 13). If, vice versa, the planes formed by different surface types under varying surface orientation do *not* intersect at the origin, we can assess any residual path radiance from the point of intersection of at least two given class planes (see Fig. 3.4). The planes must intersect when no light at all is incident onto the surfaces, and any remaining spectral radiance must then be due to path radiance. The path radiance is the same for all surface types as it depends only on the line of view of the sensor through the atmosphere.

As two-planes need not necessarily intersect in the n -dimensional space, we define the intersection as the center between the respective proxima of each plane to the other. The coordinates of this point must then be the spectral radiances due to path radiance. The proper calculations are developed in Appendix A.5 on page 183. In summary, this formalism provides a theoretical possibility to assess the path radiance for atmospherically uncorrected data, or to assess any path radiance residuals from an insufficient atmospheric correction. The computational cost is again negligible, once the training data sets, *i.e.*, the spectra observed from at least two different surface reflectance types under varying surface orientation, are selected.

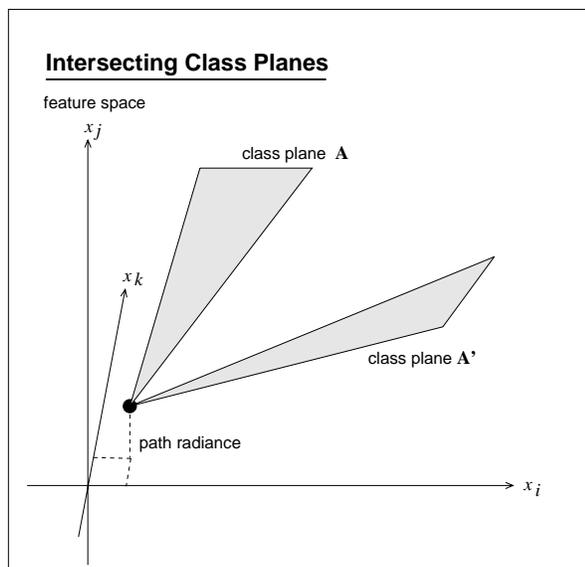


Figure 3.4: Intersection of the class planes in the spectral feature space.

Illumination Invariance in the Logarithmic Feature Space

Outline of this chapter – We introduce a feature space which is novel for multispectral classification. It is based on the logarithm of the spectral reflectances and accounts for the multiplicative nature of the reflection process. The comparison of the relative size of one cluster in the logarithmic in contrast to the original space shows no general preference for either one. However, investigation shows that the separability of two clusters is better in the log space if the within-cluster-variance is caused by changes of illumination with fixed reflectances.

We derive a logarithmic brightness normalization and discuss its advantages over conventional vector normalization in the original space, particularly with respect to error propagation.

The dichromatic illumination model is analyzed in the logarithmic space, in which the spectral illumination variance can be expanded into a Taylor series. The crucial finding is that in the logarithmic space the variability of the observed spectra – as caused by varying illumination – is independent of the reflectance.

We derive the principal modes of this spectral variance and develop a linear algebraic filter to remove the principal variance components. In this way we define an illumination – and thus surface orientation – invariant spectral descriptor. The properties of the suggested mapping are determined and illustrated on simulated data. The mapping operator can be split into a brightness and a color normalization.

Then the error propagation properties of the suggested mapping are derived for additive and multiplicative systematic errors and random noise. The performance of the transform under the presence of noise is assessed, and it is shown that there is always a noise level below which the transformed data will outperform the original data with respect to cluster separability. Moreover, the transform has an overall smoothing property on multiplicative random noise.

We investigate how to decide whether the difference of a given pair of observed spectra can be explained by change in illumination or necessarily indicates a change in reflectance. The distance of the transformed spectra must be compared to its expected standard deviation as caused by noise. As an option, the transformed spectral distance can be error weighted according to the earlier derived error propagation properties of

the transform. Simulations show that e.g. the quantization noise of the sensor causes errors in the spectral distances which can be significantly mitigated by error weighting.

Finally the Lambertian assumption, which was used for derivation, is revisited, and it is shown that the here suggested invariant is supposed to work also under a considerably relaxed reflectance assumption.

4.1 The Logarithmic Feature Space

In 1834 the psycho-physicist Ernst Heinrich Weber (1795–1878) suggested that the human senses distinguish intensities of stimuli in ratios rather than in differences. This finding was reformulated by Gustav Theodor Fechner (1801–1887) into the Weber-Fechner-Law (Fechner 1860), stating that the human senses perceive stimuli as the logarithm of their physical intensities.

When working with radiance images (or pseudo-reflectance images, see page 18) instead of reflectance images, it seems natural to change into the logarithmic domain, since the radiance image formation is a *multiplicative* process. The multiplication of illumination and reflectance then becomes an *additive* process in the logarithmic space, and thus a simple superposition which can be tackled with linear algebraic methods. In the context of panchromatic images, Ballard & Brown (1982) have pointed out as another advantage that – while the radiance signal is always positive – the logarithmic signal is unrestricted in sign. Therefore we can do without inequations which constrain linear algebraic solutions and complicate the calculus.

Oppenheim & Schafer (1975) have used the logarithmic domain for panchromatic images and spatial filtering in the frequency domain. For color image processing the logarithmic approach has been used only in physiologically/neurally motivated efforts, namely in the context of the retinex theory (Land & McCann 1971) and in a VLSI realization (electronic hardware) by Moore et al. (1991).

Despite its apparent usefulness, the logarithmic feature space has only rarely been used in RGB-based computer vision, and we do not know about any applications in multispectral image processing earlier than Wiemker (1995a) and Wiemker (1995b). Therefore we start with an investigation of its properties in comparison to the original feature space.

4.1.1 Cluster Separability in the Logarithmic Space

We want to compare the separability of clusters in the original versus the logarithmic feature space. For simplicity, we drop the spectral band index i and work out a one-dimensional example. Let us consider two cluster with means x' and x'' and variances $\sigma^2(x')$ and $\sigma^2(x'')$, assuming a Gaussian normal distributions for the clusters. The means are the cluster centers, and the variances are the cluster sizes. In the logarithmic space, the mean values x' and x'' become $\ln x'$ and $\ln x''$, and the variances $\sigma^2(\ln x')$ and $\sigma^2(\ln x'')$.

Then the squared relative distance between the two clusters in original (left-hand)

versus in logarithmic space (right-hand) is

$$\frac{(x' - x'')^2}{\sigma^2(x') + \sigma^2(x'')} \longleftrightarrow \frac{(\ln x' - \ln x'')^2}{\sigma^2(\ln x') + \sigma^2(\ln x'')} \quad (4.1)$$

where we have the absolute distances between the cluster centers in the numerators, weighted with the cluster sizes in the denominators.

Error propagation yields the relation between the variance in the original and the resulting variance in the logarithmic space:

$$\sigma^2(\ln x) = \left(\frac{\partial \ln x}{\partial x} \right)^2 \sigma^2(x) \quad (4.2)$$

$$= \frac{1}{x^2} \sigma^2(x) \quad . \quad (4.3)$$

■ Equal Cluster Extensions in the Original Space

Let the two clusters have equal variance in the original space: $\sigma^2(x') = \sigma^2(x'')$. Using Eq. 4.3 this transforms into the logarithmic space as $\sigma^2(\ln x') = x'^{-2} \sigma^2(x')$ and $\sigma^2(\ln x'') = x''^{-2} \sigma^2(x'') = x''^{-2} \sigma^2(x')$. By substitution in Eq. 4.1 we get

$$\frac{(x' - x'')^2}{\sigma^2(x') + \sigma^2(x'')} \longleftrightarrow \frac{(\ln x' - \ln x'')^2}{x'^{-2} \sigma^2(x') + x''^{-2} \sigma^2(x')} \quad (4.4)$$

for the original (left-hand) and the log space (right-hand). In order to compare these expressions we multiply by $\sigma^2(x')$ and introduce the ratio $\xi = x'/x'' > 0$

$$\frac{(\xi - 1)^2}{2} \longleftrightarrow \frac{(\ln \xi)^2}{\xi^{-2} + 1} \quad . \quad (4.5)$$

These two expressions are equally zero for $\xi = 1$ but the left-hand expression is larger elsewhere (for illustration see Fig. 4.1 on page 50, left)

$$\frac{(\xi - 1)^2}{2} \geq \frac{(\ln \xi)^2}{\xi^{-2} + 1} \quad , \quad \forall \xi > 0 \quad (4.6)$$

meaning, that the relative distance between the clusters is larger and thus the separation is better in the original space (left-hand expression), provided that the cluster variances are equal in the original space.

■ Equal Cluster Extensions in the Logarithmic Space

In a second step we compare two clusters with equal variance in the logarithmic space: $\sigma^2(\ln x') = \sigma^2(\ln x'')$. Then Eq. 4.1 becomes

$$\frac{(x' - x'')^2}{x'^2 \sigma^2(\ln x') + x''^2 \sigma^2(\ln x')} \longleftrightarrow \frac{(\ln x' - \ln x'')^2}{\sigma^2(\ln x') + \sigma^2(\ln x'')} \quad (4.7)$$

and using the same transformations as above yields

$$\frac{(\xi - 1)^2}{\xi^2 + 1} \longleftrightarrow \frac{(\ln \xi)^2}{2} \quad (4.8)$$

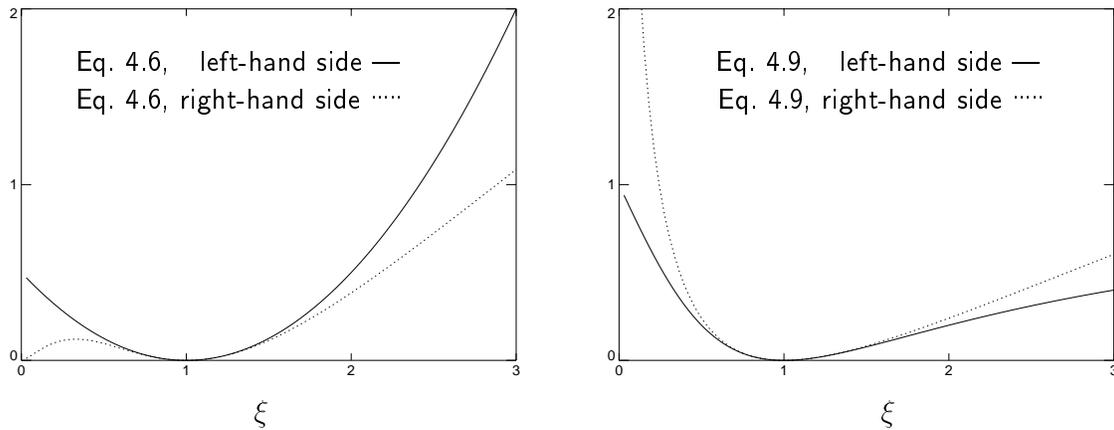


Figure 4.1: Illustration of inequation Eq. 4.6 on the preceding page and Eq. 4.9. The normalized distance between two clusters in original (solid) versus in log space (dotted) as a function of the ratio of the cluster centers $\xi = x'/x''$.

which again is equal for $\xi = 1$, but here the right-hand expression is larger elsewhere (for illustration see Fig. 4.1 on page 50, right)

$$\frac{(\xi - 1)^2}{\xi^2 + 1} \leq \frac{(\ln \xi)^2}{2}, \quad \forall \xi > 0 \quad (4.9)$$

meaning, that in this case the separation is better in the logarithmic space (right-hand expression).

Hence we can draw the conclusion that

- ▶ if the natural variance of the clusters is caused by an additive process and thus of equal extension in the original space, then the cluster separability (as measured by the normalized distance) is better in the original space, whereas
- ▶ if the natural variance of the clusters is caused by a multiplicative process and thus of equal extension in the log space, then the separability is better in the log space.

We have reason to believe that the logarithmic space is more appropriate since the reflection of illuminating light on a surface is an inherently multiplicative process (following Lambert's law). In particular, when two surface patches of different reflectance signatures are observed with changing surface orientations and thus changing illumination, then this will result in two clusters which are of equal extension in the logarithmic feature space. In contrast, in the original feature space the cluster extension will be proportional to the respective reflectance value. Hence, the separation of the two clusters will be better in the logarithmic space.

4.2 Brightness Constancy

4.2.1 Brightness Normalization

The foremost effect of varying surface orientation is the changing of the brightness of the observed spectrum, *i.e.*, an overall scaling of all spectral values by a common factor. We now discuss how such a brightness variability is conventionally dealt with, and introduce a novel brightness filtering method in the logarithmic space.

Previous approaches to normalize the effect of arbitrary brightness scalings suggest to decompose the observed spectrum \mathbf{x} into a ‘brightness’ $\|\mathbf{x}\|$ (the vector magnitude of \mathbf{x}), and a ‘chromatic vector’ $\hat{\mathbf{x}}$ which is mapped from \mathbf{x} as

$$\mathbf{x} \mapsto \hat{\mathbf{x}} = \frac{1}{\|\mathbf{x}\|} \mathbf{x} \quad , \quad (4.10)$$

in other words, to normalize the feature vector \mathbf{x} (Duda & Hart 1973). This has been applied to multispectral image data *e.g.* by Baraldi & Parmiggiani (1995). The set of observed vectors \mathbf{X} is by normalization mapped to $\hat{\mathbf{X}}$. Then the distance between two normalized vectors is

$$d^2 = \|\hat{\mathbf{x}} - \hat{\mathbf{x}}'\|^2 \quad (4.11)$$

Another approach (Duda & Hart 1973) is to consider the angular distance, or ‘spectral angle’ α between the pair of spectra \mathbf{x} and \mathbf{x}'

$$\cos \alpha(\mathbf{x}, \mathbf{x}') = \frac{\mathbf{x}^T \mathbf{x}'}{\|\mathbf{x}\| \|\mathbf{x}'\|} \quad (4.12)$$

The spectral angle has been used in remote sensing image processing *e.g.* by Kruse et al. (1993) and Ben-Dor et al. (1994).

We can easily show that these two approaches – the distance between normalized vectors and the spectral angle – are equivalent, *i.e.*, correlated linearly with reverse signs:

$$d^2 = \|\hat{\mathbf{x}} - \hat{\mathbf{x}}'\|^2 \quad (4.13)$$

$$= (\hat{\mathbf{x}} - \hat{\mathbf{x}}')^T (\hat{\mathbf{x}} - \hat{\mathbf{x}}') \quad (4.14)$$

$$= \hat{\mathbf{x}}^T \hat{\mathbf{x}} + \hat{\mathbf{x}}'^T \hat{\mathbf{x}}' - 2\hat{\mathbf{x}}^T \hat{\mathbf{x}}' \quad (4.15)$$

$$= 2(1 - \hat{\mathbf{x}}^T \hat{\mathbf{x}}') \quad \text{with } \|\hat{\mathbf{x}}\| = \|\hat{\mathbf{x}}'\| = 1 \quad (4.16)$$

$$= 2(1 - \cos \alpha(\mathbf{x}, \mathbf{x}')) \quad (4.17)$$

■ Noise Analysis:

We now study the influence of random noise on the ‘brightness normalization’-mapping. The mapping of the original spectral vector \mathbf{x} to the normalized spectral vector $\hat{\mathbf{x}}$ is computed for each component as

$$x_i \mapsto \hat{x}_i = \frac{x_i}{\|\mathbf{x}\|} = \frac{x_i}{\sqrt{\sum_j x_j^2}} \quad (4.18)$$

and thus the partial derivatives are:

$$\frac{\partial \hat{x}_i}{\partial x_j} = -\frac{x_j}{\|\mathbf{x}\|^3}, \quad i \neq j \quad (4.19)$$

and

$$\frac{\partial \hat{x}_i}{\partial x_i} = \frac{1}{\sqrt{\sum_j x_j^2}} - \frac{\frac{1}{2}x_i \cdot 2x_i}{\sqrt{\sum_j x_j^2}^3} = \frac{\|\mathbf{x}\|^2 - x_i^2}{\|\mathbf{x}\|^3} \quad (4.20)$$

Then errors dx_i in \mathbf{x} propagate into errors $d\hat{x}_i$ in $\hat{\mathbf{x}}$ as:

$$\Rightarrow d\hat{x}_i = \frac{\partial \hat{x}_i}{\partial x_i} dx_i + \sum_{j \neq i} \frac{\partial \hat{x}_i}{\partial x_j} dx_j \quad (4.21)$$

$$= \frac{1}{\|\mathbf{x}\|^3} \left[(\|\mathbf{x}\|^2 - x_i^2) dx_i - \sum_{j \neq i} x_j dx_j \right] \quad (4.22)$$

We observe that errors propagate non-linearly, *i.e.*, the propagated errors depend on the input signal \mathbf{x} . Moreover the spectral bands i are not treated on equal footing. In particular, equal noise in all spectral bands i will not transform as such:

$$\left[\sigma^2(x_i) = \text{const } \forall i \right] \not\Rightarrow \left[\sigma^2(\hat{x}_i) = \text{const } \forall i \right] \quad (4.23)$$

4.2.2 Logarithmic Brightness Filtering

In this section we introduce a novel brightness filtering in the logarithmic space.

Let us assume that the true reflectance vector \mathbf{r} is given by the observed pseudo-reflectance vector \mathbf{x} up to an unknown scaling factor μ , which indicates changing illumination brightness:

$$\mathbf{x} = \mu \mathbf{r} \quad (4.24)$$

which in the logarithmic domain is

$$\begin{bmatrix} \vdots \\ \ln x_i \\ \vdots \end{bmatrix} = \ln \mu \begin{bmatrix} \vdots \\ 1 \\ \vdots \end{bmatrix} + \begin{bmatrix} \vdots \\ \ln r_i \\ \vdots \end{bmatrix} \quad (4.25)$$

$$\ln \mathbf{x} = \ln \mu \mathbf{u} + \ln \mathbf{r}, \quad \text{where } \mathbf{u} = [1 \dots 1]^T \quad (4.26)$$

$$= \mu' \hat{\mathbf{u}} + \ln \mathbf{r}, \quad \text{where } \mu' = \ln \mu \quad (4.27)$$

$$= \mu' \hat{\mathbf{u}} + \ln \mathbf{r}, \quad \text{where } \hat{\mathbf{u}} = \frac{1}{\|\mathbf{u}\|} \mathbf{u} \text{ and thus } \hat{u}_i = \frac{1}{\sqrt{N}} \quad (4.28)$$

We now want to filter the variable brightness component $\hat{\mathbf{u}}$ from the logarithmic observed pseudo-reflectance $\ln \mathbf{x}$ by means of a mapping

$$\ln \mathbf{x} \mapsto \ln \mathbf{x} - \hat{\mathbf{u}} \hat{\mathbf{u}}^T \ln \mathbf{x} \quad (4.29)$$

which can be expressed as a multiplication:

$$= (\mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T) \ln \mathbf{x} \quad (4.30)$$

$$= \mathbf{U} \ln \mathbf{x} \quad (4.31)$$

$$\text{where } \mathbf{U} = \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T \text{ and thus } U_{ij} = \delta_{ij} - \frac{1}{N} \quad . \quad (4.32)$$

It can be easily seen that the matrix \mathbf{U} indeed filters all components $\hat{\mathbf{u}}$:

$$\mathbf{U}(\mu'\hat{\mathbf{u}}) = \mu'(\mathbf{U}\hat{\mathbf{u}}) \quad (4.33)$$

$$= \mu'(\mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T)\hat{\mathbf{u}} \quad (4.34)$$

$$= \mu'(\hat{\mathbf{u}} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T\hat{\mathbf{u}}) \quad (4.35)$$

$$= \mu'(\hat{\mathbf{u}} - \hat{\mathbf{u}}) \quad (4.36)$$

$$= 0 \quad (4.37)$$

For the result of the mapping, we introduce a transformed spectrum \mathbf{t} :

$$\ln \mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{U} \ln \mathbf{x} \quad (4.38)$$

$$\text{and } \mathbf{x} \mapsto \mathbf{t} = e^{\mathbf{U} \ln \mathbf{x}} \quad , \quad (4.39)$$

where the transformed $\ln \mathbf{t}$ is brought back into the original domain by exponentiation. We can show that the transformed spectrum $\ln \mathbf{t}$ depends not on the brightness of \mathbf{x} but only on the true reflectance spectrum \mathbf{r} :

$$\ln \mathbf{x} = \ln(\mu\mathbf{r}) \mapsto \mathbf{U} \ln(\mu\mathbf{r}) \quad (4.40)$$

$$= \mathbf{U}(\mu'\hat{\mathbf{u}} + \ln \mathbf{r}) \quad (4.41)$$

$$= \mathbf{U}(\mu'\hat{\mathbf{u}}) + \mathbf{U}(\ln \mathbf{r}) \quad (4.42)$$

$$= 0 + \mathbf{U}(\ln \mathbf{r}) \quad (4.43)$$

because the brightness scaling factor $\mu' = \ln \mu$ vanishes (Eq. 4.37).

In other words, since the projector \mathbf{U} filters any component of the brightness shift vector $\hat{\mathbf{u}}$ from an observed logarithmic spectrum $\ln \mathbf{x}$, the mapped spectra $\mathbf{U} \ln \mathbf{x}$ are invariant against any brightness variability and depend only on the components of the logarithmic reflectance $\ln \mathbf{r}$ spectrum which are orthogonal to $\hat{\mathbf{u}}$.

We note that \mathbf{U} is an orthogonal projector (see Appendix A.2 on page 179): firstly, it is symmetric by construction

$$\mathbf{U}^T = \mathbf{U} \quad (4.44)$$

and, secondly, it has the defining projector property

$$\mathbf{U}\mathbf{U} = \mathbf{U} \quad (4.45)$$

$$\text{because } \mathbf{U}\mathbf{U} = (\mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T)^2 \quad (4.46)$$

$$= \mathbf{I} - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^T + \hat{\mathbf{u}}\hat{\mathbf{u}}^T\hat{\mathbf{u}}\hat{\mathbf{u}}^T \quad (4.47)$$

$$= \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T \quad , \quad \text{because } \hat{\mathbf{u}}^T\hat{\mathbf{u}} = 1 \quad (4.48)$$

$$= \mathbf{U} \quad . \quad (4.49)$$

The mapping $\ln \mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{U} \ln \mathbf{x} = \mathbf{U} \ln \mathbf{r}$ yields an invariant against an overall scaling factor μ . We see the following advantages of the here presented logarithmic brightness filtering (Eq. 4.39) over the above discussed normalization (Eq. 4.10 on page 51):

- ▶ The mapping accounts properly for the underlying model of illumination and reflection being a multiplicative process in the original domain \mathbf{X} and thus a superposition in the logarithmic domain $\ln \mathbf{X}$.
- ▶ The resulting brightness filtered spectrum $\ln \mathbf{t}$ is ‘centered’ around 0, *i.e.*, the values t_i add up to zero, and so the mean value is $\langle \ln t_i \rangle = 0$ (similarly, the exponentiated transform \mathbf{t} is centered around 1). The *filtered* spectrum is thus independent of the number N of spectral bands i , whereas the *normalization* of \mathbf{x} is dependent on N (see Eq. 4.18 on page 51).
- ▶ Moreover this mapping can conveniently be extended to account for a dichromatic illumination model as we will show later.

■ Noise Analysis:

In the logarithmic domain we consider the mapping

$$\ln \mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{U} \ln \mathbf{x} = \mathbf{U} \ln \mathbf{x} \quad (4.50)$$

which is computed in each spectral band i as

$$\ln t_i = \sum_j U_{ij} \ln x_j \quad (4.51)$$

so that errors $d(\ln x_i)$ in $\ln \mathbf{x}$ propagate as

$$\Rightarrow d \ln t_i = \frac{\partial \ln t_i}{\partial \ln x_i} d \ln x_i \quad (4.52)$$

$$= \sum_j U_{ij} d \ln x_j \quad (4.53)$$

and using $d(\ln t_i) = dt_i/t_i$ and $d(\ln x_i) = dx_i/x_i$ yields

$$\frac{dt_i}{t_i} = \sum_j U_{ij} \frac{dx_j}{x_j} \quad (4.54)$$

for the original domain. Then the variances transform as:

$$\sigma^2(\ln t_i) = \sum_j \left(\frac{\partial \ln t_i}{\partial \ln x_j} \right)^2 \sigma^2(\ln x_j) \quad (\text{no covariances } \sigma^2(\ln x_i, \ln x_j) \text{ assumed}) \quad (4.55)$$

$$= \sum_j U_{ij}^2 \sigma^2(\ln x_j) \quad . \quad (4.56)$$

Now let us assume uncorrelated noise σ of equal level ϵ in all spectral bands i in the logarithmic feature space with the spectra $\ln \mathbf{x}$:

$\sigma^2(\ln x_i) = \epsilon^2$, and $\Sigma(\ln \mathbf{x}) = \epsilon^2 \mathbf{I}$ ('spherical' noise), where $\Sigma(\ln \mathbf{x})$ is the covariance matrix of the logarithmic spectra $\ln \mathbf{x}$. Then

$$\sigma^2(\ln t_i) = \epsilon^2 \sum_j U_{ij}^2 \quad (4.57)$$

which is equal for all i , due to the symmetry properties of \mathbf{U} . The overall noise of the original logarithmic spectra \mathbf{x} is

$$\text{tr } \Sigma(\ln \mathbf{x}) = \epsilon^2 N \quad . \quad (4.58)$$

After the filtering, the original covariance matrix $\Sigma(\ln \mathbf{x})$ of the logarithmic spectra $\ln \mathbf{x}$ is mapped to $\Sigma \mapsto \mathbf{U}\Sigma\mathbf{U}^T = \mathbf{U}\Sigma\mathbf{U}$ (common behavior for matrix induced linear mappings (Brandt 1992)). So the trace of the transformed covariance and thus the overall noise for the transformed logarithmic spectra $\ln \mathbf{t}$ is

$$\text{tr } \Sigma(\ln \mathbf{t}) = \text{tr } (\mathbf{U}\Sigma(\ln \mathbf{x})\mathbf{U}) = \sum_i \sigma^2(\ln t_i) \quad (4.59)$$

substituting Eq. 4.57 yields

$$= \epsilon^2 \sum_{ij} U_{ij}^2 \quad (4.60)$$

$$= \epsilon^2 \|\mathbf{U}\|_F \quad (4.61)$$

which is the squared *Frobenius norm* (Hämmerlin & Hoffmann 1994) of the matrix \mathbf{U} and can be easily evaluated by virtue of Eq. 4.32 on page 53 to

$$= \epsilon^2 (N - 1) \quad . \quad (4.62)$$

In matrix notation, a 'spherical' covariance matrix in the logarithmic space transforms as:

$$\left[\Sigma(\ln \mathbf{x}) = \epsilon^2 \mathbf{I} \right] \mapsto \left[\epsilon^2 \mathbf{U}\mathbf{I}\mathbf{U} = \epsilon^2 \mathbf{U} \right] \quad (4.63)$$

because $\mathbf{U}\mathbf{U} = \mathbf{U}$.

In conclusion, we see the following advantages of the logarithmic brightness filtering:

- ▶ $\ln \mathbf{t} = \mathbf{U} \ln \mathbf{x}$ is a linear transform from $\ln \mathbf{x}$ to $\ln \mathbf{t}$, and thus the error $\sigma^2(\ln x_i)$ propagates independently of the signal x_i and allows error estimation for the complete set of spectra $\ln \mathbf{X}$.
- ▶ Due to the symmetry properties of \mathbf{U} , 'spherical error' ($\Sigma(\ln \mathbf{x}) = \epsilon^2 \mathbf{I}$) in the logarithmic domain remains as such:

$$\left[\sigma^2(\ln x_i) = \text{const } \forall i \right] \quad \Rightarrow \quad \left[\sigma^2(\ln t_i) = \text{const } \forall i \right] \quad .$$

- We can show that the overall error is always mitigated by the mapping: We form the ratio of the noise in the transformed space (Eq. 4.62) over the noise in the original space (Eq. 4.58) and confirm this ratio to be smaller than one:

$$\frac{\text{tr}(\mathbf{U}\Sigma(\ln \mathbf{x})\mathbf{U})}{\text{tr} \Sigma(\ln \mathbf{x})} = \left(\frac{N-1}{N}\right) < 1 \quad .$$

- The logarithmic brightness filtering is a mapping of the original N dimensional spectral space to a $N-1$ dimensional linear subspace. This means that the mapped spectra are confined to a *hyperplane* in the N dimensional spectral feature space. In contrast, the brightness normalization (Eq. 4.10 on page 51) maps the original space onto a *hypersphere*, which is a curvilinear manifold in the N dimensional spectral feature space and much harder to treat as far as classification algorithms are concerned.

We will illustrate this topic by virtue of simulated data in the following paragraph.

■ Illustration on Simulated Data:

We want to illustrate the concept of the logarithmic space and logarithmic brightness filtering on some examples with simulated data.

A typical pattern in the feature space are ‘rays’ pointing towards the origin which are made up from spectra of the same true reflectance spectrum but varying illumination intensity (Fig. 4.2 on page 58, top left). Each spectral class makes up one ray, and all rays are pointing towards the origin but with class-specific, different directions in the N -dimensional spectral feature space.

The advantage of the logarithmic feature space is that here all such rays are running parallel (Fig. 4.2, top right) rather than radial as in the original space. The brightness scaled spectra (simulated data) form clusters of the form $x_j = cx_i$ which become $\ln x_j = \ln c + \ln x_i$ in the logarithmic space. Since all spectral clusters are represented by rays of identical inclination (namely unity) in the log space, it becomes easier to formulate effective decision surfaces between the ray-shaped clusters. A simple coordinate system rotation by 45° suffices to provide uncorrelated coordinates $[\ln x_i + \ln x_j]$ and $[\ln x_i - \ln x_j]$ (Fig. 4.2, center left). The bottom row of Fig. 4.2 shows the effect of log brightness filtering ($\mathbf{U} \ln \mathbf{x}$, left) and brightness normalization ($\mathbf{x}/\|\mathbf{x}\|$, right). Note that in contrast to the normalized cluster centers the log brightness filtered cluster centers lie on a *hyper-plane* rather than a *hyper-sphere*. The same is true for the principal direction of the noise variability. Multiplicative noise was superimposed on the original artificial data ($\text{SNR} = 3\% \hat{=} 30.5 \text{ dB}$). For the log brightness filtered points, the noise causes variability in only a single linear coordinate component. This noise component has the same directions for all three clusters and is thus easier to consider during image processing. In contrast, in the brightness normalized space, the noise varies in both coordinate directions, and is moreover cluster dependent and thus more difficult to deal with.

Since the variance of the log brightness filtered points in the feature space is on a linear subspace, for this $N = 2$ -dimensional example we can catch the whole information in a single histogram of the principal component ($[\mathbf{U} \ln \mathbf{x}]_j - [\mathbf{U} \ln \mathbf{x}]_i$, Fig. 4.2, center right). This would not be possible for the normalized spectra, because – as stated above – they vary in more than one principal component. It might be argued that a

transformation of the normalized spectra (Fig. 4.2, bottom right) into an angular system would also allow to comprehend the information in a single magnitude. However, the concept of angles becomes complicated for more than $N = 3$ dimensions. Moreover, this would require a non-linear transformation, whereas in the log brightness filtered a simple linear combination is sufficient. We recall that the usual classifier algorithms provide linear (minimum Euclidean distance, Eq. 2.25 on page 23) and quadratic (maximum likelihood, Eq. 2.34 on page 25) combinations of the spectral values. Hence, we expect better performance for these common classifiers in a log brightness filtered space than *e.g.* in a normalized or an angle-transformed space.

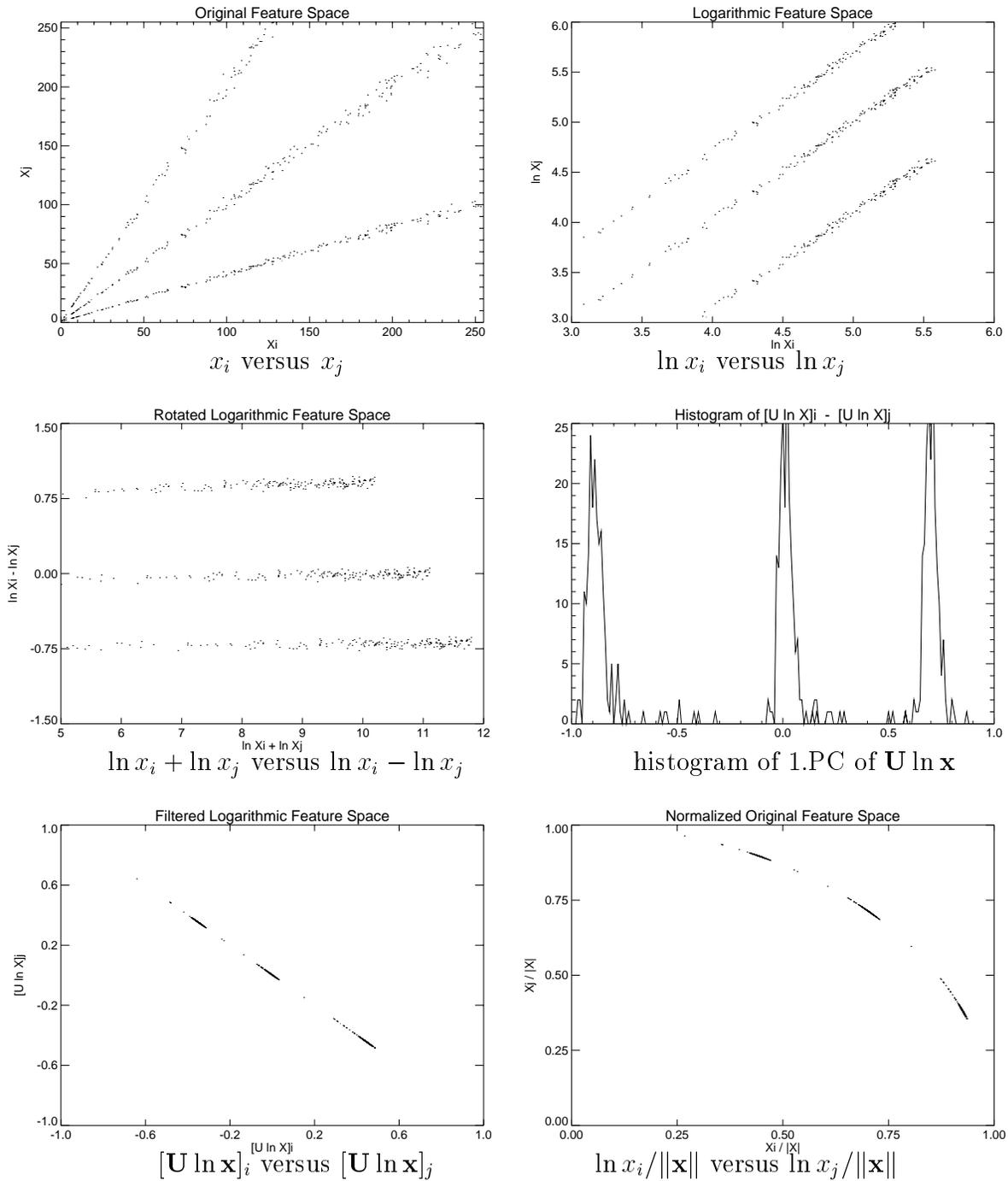


Figure 4.2: Typical brightness scaled clusters (simulated data, superposed with multiplicative noise of $\text{SNR} = 3\% \cong 30.5 \text{ dB}$), forming ‘rays’ out of the origin of the feature space (top left), transformed into equidistant ‘lines’ in the log space (top right). A simple coordinate system rotation by 45° then suffices to provide uncorrelated coordinates (center left). A histogram of the principle component of the log brightness filtered spectra is shown in the center row, right. The bottom row shows the effect of log brightness filtering ($\mathbf{U} \ln \mathbf{x}$, left) and brightness normalization ($\mathbf{x} / \|\mathbf{x}\|$, right).

4.3 The Dichromatic Illumination Model in the Logarithmic Space

After having analyzed the effect of pure brightness changes in both the original and the logarithmic space in the previous section, we now address the variability of spectral signatures as caused by changing surface orientations. This spectral variability is introduced by a dichromatic illumination with changing contributions. Let us recall the dichromatic illumination model with direct solar and diffuse sky light. The respective contributions are varying due to changing surface orientation of a given patch which thus receives more or less sun and sky illumination. The changing illumination then leads to an apparently changing spectral signature of the surface patch in question. We call this the *orientation-dependent spectral variability*.

Assuming Lambertian reflection, the observed reflected radiance spectra \mathbf{x} are given by the direct and diffuse illumination spectra $\mathbf{n} \hat{=} E_{\text{dir}}$ and $\mathbf{m} \hat{=} E_{\text{diff}}$ multiplied component-wise with the specific surface reflectance spectrum \mathbf{r} :

$$x_i = r_i(\mu m_i + \nu n_i) \quad , \quad \mu > 0, \nu \geq 0 \quad (4.64)$$

where μ and ν are the contributions of diffuse and direct illumination respectively. The direct contribution may vanish completely in shadow, whereas the diffuse will always be present to a certain degree. So we can introduce a relative contribution $\eta = \nu/\mu$ of direct illumination, take μ as an overall brightness factor, and reformulate

$$x_i = r_i \mu (m_i + \eta n_i) \quad , \quad \mu > 0, \eta \geq 0 \quad (4.65)$$

For $\eta = 1$ we get simply a multiple of the illumination incident onto a horizontal surface ($m_i + n_i$).

In order to separate illumination and reflectance we change into the logarithmic domain:

$$\ln x_i = \ln r_i + \ln \mu + \ln(m_i + \eta n_i) \quad (4.66)$$

or

$$\ln x_i - \ln r_i = \ln \mu + \ln(m_i + \eta n_i) \quad (4.67)$$

$$\ln(x_i/r_i) = \ln \mu + \ln(m_i + \eta n_i) \quad (4.68)$$

where the illumination is on the right-hand side.

Now we expand the direct/diffuse illumination mixture term $\ln(m_i + \eta n_i)$ into a Taylor series in η for $\eta \approx 1$ (*i.e.*, for a nearly horizontal surface patch):

$$\ln(m_i + \eta n_i) \quad (4.69)$$

$$= \left[\ln(m_i + \eta' n_i) + \frac{n_i}{(m_i + \eta' n_i)}(\eta - 1) - \frac{n_i^2}{(m_i + \eta' n_i)^2}(\eta - 1)^2 \dots \right]_{\eta'=1} \quad (4.70)$$

$$= \ln(m_i + n_i) - \sum_{k=1}^{\infty} (-1)^k (k-1)! (\eta - 1)^k \left(\frac{n_i}{n_i + m_i} \right)^k \quad (4.71)$$

Here we see that the leading term $\ln(m_i + n_i)$ does not contain the varying η and thus does not contribute to the variability at all. Limiting the Taylor series expansion to the first $(K - 1) \leq N$ terms, we can describe the variability of the mixture term $\ln(m_i + \eta n_i)$ as

$$\delta \ln(m_i + \eta n_i) \approx \sum_{k=1}^{K-1} c_k \hat{n}_i^k \quad (4.72)$$

where c_k are contribution factors, and $\hat{n}_i = \frac{n_i}{n_i + m_i}$ is the relative direct to global illumination. Likewise we want to introduce $\hat{m}_i = \frac{m_i}{n_i + m_i}$ as the relative diffuse to global illumination, and have $\hat{n}_i + \hat{m}_i = 1$, or $\hat{\mathbf{n}} + \hat{\mathbf{m}} = \mathbf{u}$.

Limiting the series expansion (Eq. 4.72) to $(K - 1)$ terms means that the expression is only an approximation when the mixture coefficient η is no longer close to $\eta = 1$. The accuracy of the approximation depends on the relative direct and diffuse to global illumination factors n_i and m_i , and thus on the sensor-specific positions λ_i of the spectral bands.¹

The overall scaling factor in the logarithmic illumination term (Eq. 4.66) can be written as $\ln \mu = c_0 \mathbf{1}$, allowing a convenient vector/matrix notation. Then the variability of the illumination $\ln \mathbf{x} - \ln \mathbf{r}$ can be approximated as

$$\delta (\ln \mathbf{x} - \ln \mathbf{r}) \approx c_0 \mathbf{u} + \sum_{k=1}^{K-1} c_k \hat{\mathbf{n}}^k \quad (4.73)$$

$$= \mathbf{Q} \mathbf{c} \quad (4.74)$$

where the coefficient vector $\mathbf{c} = [\dots c_k \dots]^T$ is multiplied with the $N \times K$ matrix $\mathbf{Q} = [\hat{\mathbf{n}}^0 \dots \hat{\mathbf{n}}^{K-1}]$, the columns of which are given by the powers of $\hat{\mathbf{n}}$.

$$Q_{ik} = \hat{n}_i^{k-1} \quad (4.75)$$

$$\mathbf{Q} = \begin{bmatrix} 1 & \hat{n}_1 & \hat{n}_1^2 & \dots & \hat{n}_1^{K-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \hat{n}_N & \hat{n}_N^2 & \dots & \hat{n}_N^{K-1} \end{bmatrix} \quad (4.76)$$

We note the fact that the variability of the spectral signature, or spectral variability, does not in the logarithmic space depend on the *absolute* direct and diffuse spectra \mathbf{n} and \mathbf{m} , but rather only on the *relative* direct or diffuse to global illumination ratios $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$. In contrast to the absolute spectra, the relative direct and diffuse to global illumination spectra are relatively smooth and simple function as pointed out in Section 3.1 on page 35.

4.4 The Spectral Filtering

We have identified the variability vectors in the logarithmic feature space which result from varying contributions of sunlight and skylight. These variability vectors are the

¹ For the specific case of the DAEDALUS sensor, examples were computed, with up to 15% approximation error for the extreme case of complete shadow ($\eta = 0$), and only 0.1% approximation error for surface tilts of $\theta = 45^\circ$ or 60° under a solar zenith angle $\theta_\odot = 45^\circ$ (Table 6.7 on page 164).

column vectors of the matrix \mathbf{Q} (Eq. 4.76). They are identified as the first K powers of the normalized relative direct illumination spectrum $\hat{\mathbf{n}} : \mathbf{Q} = [\hat{\mathbf{n}}^0 \dots \hat{\mathbf{n}}^{K-1}]$.

Now the aim is to filter out the variable components $\hat{\mathbf{n}}^k$ from an observed spectrum $\ln \mathbf{x}$ in order to produce an invariant spectral descriptor for a surface material, independent of its tilt. To this aim we determine the contributions c_k of the variability vectors $\hat{\mathbf{n}}^k$ which are present in the observed logarithmic spectrum $\ln \mathbf{x}$: $\ln \mathbf{x} \approx \mathbf{Q} \mathbf{c}$. This coefficient vector \mathbf{c} can be found as $\mathbf{c} = \mathbf{Q}^+ \ln \mathbf{x}$, where \mathbf{Q}^+ is the pseudo-inverse (Appendix A.1 on page 177) of \mathbf{Q} . Then the closest approximation to $\ln \mathbf{x}$ by linear combination of the variability vectors is $\mathbf{Q} \mathbf{c} = \mathbf{Q} \mathbf{Q}^+ \ln \mathbf{x}$. It are just these components of $\ln \mathbf{x}$ which may be affected by surface-orientation-caused illumination variability and shall therefore be removed by subtraction:

$$\ln \mathbf{x} \mapsto \ln \mathbf{x} - \mathbf{Q} \mathbf{Q}^+ \ln \mathbf{x} \quad (4.77)$$

$$= (\mathbf{I} - \mathbf{Q} \mathbf{Q}^+) \ln \mathbf{x} \quad (4.78)$$

$$= \mathbf{P} \ln \mathbf{x} \quad (4.79)$$

The subtraction can be expressed as a multiplication with the orthogonal projector $\mathbf{P} = \mathbf{I} - \mathbf{Q} \mathbf{Q}^+$, where $\mathbf{P} = \mathbf{P}^T$ is a symmetric $N \times N$ matrix and has the defining projector property $\mathbf{P} \mathbf{P} = \mathbf{P}$. It is thus an orthogonal projection $\mathbb{R}^N \mapsto \mathbb{R}^{N-K}$ (see Appendix A.2 on page 179).

Following the filtering in the logarithmic feature space, the exponential function takes us back into the original feature space:

$$\mathbf{x} \mapsto \exp(\mathbf{P} \ln \mathbf{x}) \quad (4.80)$$

The filtering process through the various domains is sketched in Fig. 4.3 on the next page.

It can be easily shown that the projector \mathbf{P} indeed filters all column vectors contained in \mathbf{Q} :

Let $\mathbf{c} = (c_1, c_2)^T$ be a coefficient vector which linearly combines the column vectors of \mathbf{Q} into a resulting N dimensional vector $\mathbf{Q} \mathbf{c}$. Any change in the two contribution factors in \mathbf{c} then leads to a logarithmic illumination shift vector of the form $\mathbf{Q} \mathbf{c}$ (Eq. 4.74). Then application of the projector \mathbf{P} filters this illumination shift as follows:

$$\mathbf{P} \mathbf{Q} \mathbf{c} = (\mathbf{I} - \mathbf{Q} \mathbf{Q}^+) \mathbf{Q} \mathbf{c} \quad (4.81)$$

$$= (\mathbf{Q} - \underbrace{\mathbf{Q} \mathbf{Q}^+ \mathbf{Q}}_{\mathbf{Q}}) \mathbf{c} \quad (4.82)$$

$$= (\mathbf{Q} - \mathbf{Q}) \mathbf{c} \quad (4.83)$$

$$= 0 \quad (4.84)$$

because of the defining property of the pseudo-inverse $\mathbf{Q} \mathbf{Q}^+ \mathbf{Q} = \mathbf{Q}$.

We can now show that the transformed spectrum $\ln \mathbf{t} = \mathbf{P} \ln \mathbf{x}$ depends not on possible

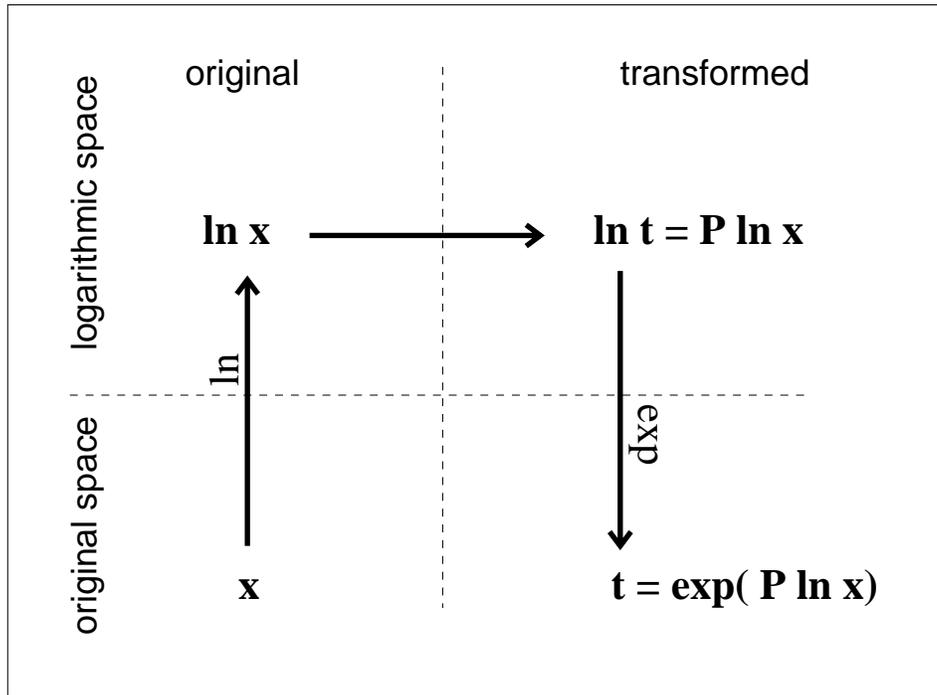


Figure 4.3: Illustration of the filtering process.

brightness or color shifts \mathbf{Qc} in an observed spectrum \mathbf{x} , but only on the true reflectance spectrum \mathbf{r} :

$$\ln \mathbf{x} = \ln \mathbf{r} + \mathbf{Qc} \mapsto \mathbf{P} \ln \mathbf{x} \quad (4.85)$$

$$= \mathbf{P}(\ln \mathbf{r} + \mathbf{Qc}) \quad (4.86)$$

$$= \mathbf{P} \ln \mathbf{r} + \mathbf{PQc} \quad (4.87)$$

$$= \mathbf{P} \ln \mathbf{r} + 0 \quad (\text{using Eq. 4.84}) \quad (4.88)$$

$$= \mathbf{P} \ln \mathbf{r} \quad (4.89)$$

In other words, since the projector \mathbf{P} filters any component of the brightness and color shift vectors contained in the matrix \mathbf{Q} from an observed logarithmic spectrum $\ln \mathbf{x}$, the mapped spectra $\ln \mathbf{t} = \mathbf{P} \ln \mathbf{x}$ are invariant against any brightness or color shift variability and depend only on the components of the logarithmic reflectance spectrum $\ln \mathbf{r}$ which are orthogonal to the column vectors in \mathbf{Q} .

Now we consider a given surface patch with reflectance spectrum $\ln \mathbf{r}$ forming a single point in the logarithmic feature space. The spectra $\ln \mathbf{x}$ observed from this surface under varying surface orientation and thus varying illumination will form a cluster in the logarithmic feature space. We have shown that this cluster can be described by linear combinations of the K variability vectors in the $N \times K$ matrix \mathbf{Q} which are simply added to the reflectance spectrum $\ln \mathbf{r}$. Then the illumination variability causes

a covariance matrix Σ_I of the cluster in the logarithmic feature space which is given by

$$\Sigma_I = \frac{1}{n} \sum_{j=1}^n (\mathbf{Q}\mathbf{c}_j - \langle \mathbf{Q}\mathbf{c} \rangle) (\mathbf{Q}\mathbf{c}_j - \langle \mathbf{Q}\mathbf{c} \rangle)^T \quad (4.90)$$

where we sum over all n occurring coefficient vectors \mathbf{c}_j , and $\langle \mathbf{Q}\mathbf{c} \rangle = \frac{1}{n} \sum_{j=1}^n \mathbf{Q}\mathbf{c}_j$ is the mean vector. The covariance Σ_I is independent of the specific reflectance $\ln \mathbf{r}$ because this is fixed for this cluster and does not vary. If we now apply the filter matrix \mathbf{P} , the covariance is transformed to

$$\Sigma_I \mapsto \mathbf{P}\Sigma_I\mathbf{P}^T = \frac{1}{n} \sum_{j=1}^n \mathbf{P}(\mathbf{Q}\mathbf{c}_j - \langle \mathbf{Q}\mathbf{c} \rangle) (\mathbf{Q}\mathbf{c}_j - \langle \mathbf{Q}\mathbf{c} \rangle)^T \mathbf{P}^T \quad (4.91)$$

$$= \frac{1}{n} \sum_{j=1}^n (\mathbf{P}\mathbf{Q}\mathbf{c}_j - \langle \mathbf{P}\mathbf{Q}\mathbf{c} \rangle) (\mathbf{P}\mathbf{Q}\mathbf{c}_j - \langle \mathbf{P}\mathbf{Q}\mathbf{c} \rangle)^T \quad (4.92)$$

$$= 0 \quad (4.93)$$

and vanishes because $\mathbf{P}\mathbf{Q}\mathbf{c} = 0 \forall \mathbf{c}$ (Eq. 4.84).

The filtering process will necessarily remove reflectance information as well as illumination information. The basic idea is, however, that the reflectance spectra will differ in features which cannot possibly be explained by varying illumination. We have shown that the illumination variability vectors are powers of the relative direct versus diffuse illumination spectra. From atmospheric physics we know that these relative illumination spectra $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ are relatively smooth and monotonous functions of the wavelength λ (see Fig. 3.1 on page 37), and thus the filtering will not remove ‘higher frequency’-features from the reflectance spectrum. As suggested by the experimental findings, we have achieved good results by filtering $K = 2$ components, *i.e.*, with a $N \times K$ matrix \mathbf{Q} of rank $K = 2$ (see Chapter 5).

4.4.1 Some Properties of the Algebraic Projector \mathbf{P}

If the matrix \mathbf{Q} were of full rank, $\text{rank } \mathbf{Q} = N$, the mapping would be invariant but trivial: $\mathbf{P} \ln \mathbf{x} = 0, \quad \forall \ln \mathbf{x}$ (because in this case $\mathbf{Q}\mathbf{Q}^+ = \mathbf{I}$ and $\mathbf{P} = \mathbf{I} - \mathbf{Q}\mathbf{Q}^+ = 0$). However, the illumination variability can be represented sufficiently by a limited number of terms K , *i.e.*, $(K - 1)$ of powers of \mathbf{n} , so that \mathbf{Q} is not of full rank: $\text{rank } \mathbf{Q} = K < N$. Thus some information of the reflectance spectrum \mathbf{r} is retained through the mapping, but the components which are most prone to illumination variability are removed. Essentially, this is possible whenever a reflectance spectrum \mathbf{r} contains higher spatial frequencies than the relative direct or diffuse illumination spectra $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$, which are smooth functions of low spatial frequency. The number of dimensions K and thus the rank of $\mathbf{Q}\mathbf{Q}^+$ can be chosen, corresponding to the number of terms of the Taylor expansion in Eq. 4.72 on page 60. This is basically a trade-off between less illumination variability on one hand and more preserved reflectance information on the other hand.

In the following we want to state some properties of \mathbf{P} and related expressions.

where the matrix \mathbf{E} contains the eigenvectors of \mathbf{P} . It has $(N - K)$ eigenvalues of 1, and K eigenvalues of 0. The zero-eigenvalues correspond to the eigenvectors which are filtered, whereas the non-zero-eigenvalues correspond to the eigenvectors which pass the filter matrix unchanged.

From the largest eigenvalue we see that the *spectral norm*⁴ is $\|\mathbf{P}\| = 1$, and that $\|\mathbf{P}\mathbf{d}\|_{\min} = 0$ and $\|\mathbf{P}\mathbf{d}\|_{\max} = \|\mathbf{d}\|$, where $\|\mathbf{P}\mathbf{d}\|_{\min}$ is the smallest vector magnitude of the $\mathbf{P}\mathbf{d}$ which can be produced by any vector \mathbf{d} , and $\|\mathbf{P}\mathbf{d}\|_{\max}$ the largest, respectively.

The *trace* of a matrix is equal to the sum of its diagonal elements, and particularly to the sum of its eigenvalues, and thus the trace of the here introduced projector matrix is $\text{tr } \mathbf{P} = N - K$.

The *Frobenius norm* (Jänich 1993) is $\|\mathbf{P}\|_F = \sqrt{\text{tr } \mathbf{P}^T \mathbf{P}} = \sqrt{\text{tr } \mathbf{P}} = \sqrt{N - K}$. Since \mathbf{P} is a symmetric projector matrix, the square of the Frobenius norm is equal to the trace, $\|\mathbf{P}\|_F^2 = \text{tr } \mathbf{P} = N - K$.

The mapping is not injective, *i.e.*, a mapped spectrum cannot be transformed back into the original spectral domain, since the matrix \mathbf{P} is singular. Thus, if – in the way of classification – we compare a number of observed reflectance signatures to a certain reference reflectance signature, we have to map the observed spectra as well as the reference signature into the illumination invariant domain before the comparison takes place.

■ Definition of Logarithmic Brightness and Color Filter \mathbf{U} and \mathbf{V} :

Now we want to set the number of filtered components to $K = 2$, meaning that we want to filter one brightness and one color variability eigenvector which we will call \mathbf{u} (Eq. 4.28 on page 52) and \mathbf{v} respectively. In analogy to the brightness filter \mathbf{U} (Eq. 4.32) we want to define a filter for the color shift. To achieve linear independence we use *Schmidt-Gram ortho-normalization* (Fischer 1984).

$$\mathbf{v} = \mathbf{U}\hat{\mathbf{u}} \quad (4.99)$$

$$\Rightarrow \mathbf{v}^T \hat{\mathbf{u}} = 0 \quad (4.100)$$

$$\hat{\mathbf{v}} = \mathbf{v} / \|\mathbf{v}\| \quad (4.101)$$

$$\Rightarrow \hat{\mathbf{v}}^T \hat{\mathbf{v}} = 1 \quad (4.102)$$

$$\mathbf{V} = \mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^T \quad (4.103)$$

so that a projector \mathbf{P} of rank $N - K = N - 2$ is equal to

$$\mathbf{P} = \mathbf{V}\mathbf{U} \quad (4.104)$$

$$= (\mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T)(\mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^T) \quad (4.105)$$

$$= \mathbf{I} + \hat{\mathbf{u}}\hat{\mathbf{u}}^T \hat{\mathbf{v}}\hat{\mathbf{v}}^T - \hat{\mathbf{u}}\hat{\mathbf{u}}^T - \hat{\mathbf{v}}\hat{\mathbf{v}}^T \quad (4.106)$$

$$= \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T - \hat{\mathbf{v}}\hat{\mathbf{v}}^T \quad \text{because of } \hat{\mathbf{v}}^T \hat{\mathbf{u}} = 0. \quad (4.107)$$

⁴ The *spectral radius* of a matrix \mathbf{A} is equal to the largest absolute eigenvalue; the *spectral norm* of a matrix \mathbf{A} is equal to the square root of the spectral radius of $\mathbf{A}^T \mathbf{A}$ (Hämmerlin & Hoffmann 1994). Note that in these mathematical terms the word ‘spectral’ is not physically motivated and not in any way related to wavelength dependency.

Note that instead of using the relative direct illumination $\hat{\mathbf{n}}$ we may as well construct \mathbf{V} from the relative diffuse illumination $\hat{\mathbf{m}}$, because of $\hat{\mathbf{n}} + \hat{\mathbf{m}} = \mathbf{u}$

$$\mathbf{v} = \mathbf{U}\hat{\mathbf{m}} \quad (4.108)$$

$$= \mathbf{U}(\mathbf{u} - \hat{\mathbf{n}}) \quad (4.109)$$

$$= -\mathbf{U}\hat{\mathbf{n}} \quad \text{because of } \mathbf{U}\mathbf{u} = \mathbf{0} \quad (4.110)$$

where only the sign is reversed in comparison to Eq. 4.99. So the employment of $\hat{\mathbf{m}}$ instead of $\hat{\mathbf{n}}$ yields the same filter $\mathbf{V} = \mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^T$. The vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ span the same subspace as the first two column vectors in the logarithmic illumination variability matrix \mathbf{Q} (Eq. 4.76 on page 60).

■ Definition of Logarithmic Brightness and Color Index:

As a complement to the filtered set of vectors \mathbf{t} which do not any longer contain any contributions of $\hat{\mathbf{u}}$ or $\hat{\mathbf{v}}$ (Eq. 4.28 on page 52 and Eq. 4.101 on the page before), we can also look at the filtered components themselves:

$$\hat{\mathbf{u}}^T \ln \mathbf{x} \quad (\text{log brightness index}) \quad (4.111)$$

$$\hat{\mathbf{v}}^T \ln \mathbf{x} \quad (\text{color index}) \quad (4.112)$$

which we want to call *logarithmic brightness index* and *logarithmic color index* respectively, since the former measures the average magnitude of the logarithmic spectrum \mathbf{x} , whereas the latter basically measures the difference between shorter and longer wavelengths, similar to the frequently used color index of astronomy (Harwit 1988).⁵ The scalars $(\hat{\mathbf{u}}^T \ln \mathbf{x})$ and $(\hat{\mathbf{v}}^T \ln \mathbf{x})$ can be represented as image layers, yielding a brightness and a color index image. These two properties are the ones which are filtered because both are affected by surface orientation as pointed out above.

■ Some More Properties of \mathbf{P} :

From $\mathbf{U}\mathbf{u} = \mathbf{0}$ (Eq. 4.37 on page 53) and $\mathbf{P} = \mathbf{V}\mathbf{U}$ we can conclude that also $\mathbf{P}\mathbf{u} = \mathbf{0}$ (where $\mathbf{u} = [\dots 1 \dots]^T$ is the unity vector). In particular, this means that for each line i of the matrix equation we have

$$[\mathbf{P}\mathbf{u}]_i = \sum_j P_{ij}u_j = \sum_j P_{ij} \cdot 1 = 0 \quad (4.113)$$

and thus the projector property

$$\sum_j P_{ij} = 0 \quad (4.114)$$

⁵ There actually is an intriguing coincidence with the Hertzsprung-Russell-Diagram (HRD) which is one of the principal representations of astrophysics (Harwit 1988). In the HRD the logarithmic brightness of stars is plotted against their color index (which is an observable related to the effective temperature). However, while in astrophysics these two parameters are useful for classification of stars, they are non-significant for surface orientation independent spectral classification in remote sensing as we have just pointed out.

i.e., all *line sums vanish*. (By *line sum* we denote the sum of all elements of the i th line of the matrix.) This enables us to determine the expectation value $\langle \ln t_i \rangle$ of the transformed logarithmic spectrum :

$$\langle \ln t_i \rangle = \left\langle \sum_j P_{ij} \ln x_j \right\rangle \quad (4.115)$$

$$= \sum_j \langle P_{ij} \ln x_j \rangle \quad . \quad (4.116)$$

We assume that the expectation value $\langle \ln x_j \rangle$ of the input signals x_j is approximately identical for all spectral bands j . This simplifies the expression to

$$\langle \ln t_i \rangle = \langle \ln x_j \rangle \sum_j P_{ij} \quad (4.117)$$

$$= 0 \quad (4.118)$$

and shows that the expectation value $\langle \ln t_i \rangle$ vanishes.

Now we also wish to determine the sum of the squares $\sum_j P_{ij}^2$ in each matrix line i , the *line square sum*. Starting from

$$\mathbf{P} = \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T - \hat{\mathbf{v}}\hat{\mathbf{v}}^T \quad (4.119)$$

$$P_{ij} = \delta_{ij} - \hat{u}_i\hat{u}_j - \hat{v}_i\hat{v}_j \quad (4.120)$$

the line square sum is

$$P_{ij}^2 = \delta_{ij} + \hat{u}_i^2\hat{u}_j^2 + \hat{v}_i^2\hat{v}_j^2 \quad (4.121)$$

$$- 2\delta_{ij}\hat{u}_i\hat{u}_j - 2\delta_{ij}\hat{v}_i\hat{v}_j \quad (4.122)$$

$$+ 2\hat{u}_i\hat{u}_j\hat{v}_i\hat{v}_j \quad (4.123)$$

then the line square sums of the projector matrix are

$$\sum_j P_{ij}^2 = 1 + \hat{u}_i^2 \sum_j \hat{u}_j^2 + \hat{v}_i^2 \sum_j \hat{v}_j^2 \quad (4.124)$$

$$- 2\hat{u}_i^2 - 2\hat{v}_i^2 \quad (4.125)$$

$$+ 2\hat{u}_i\hat{v}_i \sum_j \hat{u}_j\hat{v}_j \quad (4.126)$$

using $\sum_j \hat{u}_j^2 = \sum_j \hat{v}_j^2 = 1$ and $\sum_j \hat{u}_j\hat{v}_j = 0$

$$= 1 + \hat{u}_i^2 + \hat{v}_i^2 - 2\hat{u}_i^2 - 2\hat{v}_i^2 \quad (4.127)$$

$$= 1 - \hat{u}_i^2 - \hat{v}_i^2 \quad (4.128)$$

with $\hat{u}_i = \frac{1}{\sqrt{N}}$

$$= 1 - \frac{1}{N} - \hat{v}_i^2 \quad . \quad (4.129)$$

This result leads to the other projector property:

$$\Rightarrow 0 < \sum_j P_{ij}^2 < 1 \quad . \quad (4.130)$$

The total of all line square sums is

$$\sum_i \sum_j P_{ij}^2 = \sum_i \left(1 - \frac{1}{N} - \hat{v}_i^2\right) \quad (4.131)$$

$$= N - 1 - \sum_i \hat{v}_i^2 \quad (4.132)$$

$$= N - 2 \quad , \quad (4.133)$$

so that the mean line square sum is

$$\left\langle \sum_j P_{ij}^2 \right\rangle = \frac{N - 2}{N} \quad (4.134)$$

or more general

$$\left\langle \sum_j P_{ij}^2 \right\rangle = \frac{N - K}{N} \quad . \quad (4.135)$$

Then we want to determine the diagonal elements P_{ii} of \mathbf{P} . Setting $j = i$ in Eq. 4.120 on the page before yields

$$P_{ii} = 1 - \hat{u}_i^2 - \hat{v}_i^2 \quad (4.136)$$

which is equal to Eq. 4.128 on page 67. So the diagonal elements are equal to the line square sums $P_{ii} = \sum_j P_{ij}^2$ and finally

$$\text{tr } \mathbf{P} = \sum_i P_{ii} = \sum_{ij} P_{ij}^2 = \|\mathbf{P}\|_F^2 = N - K \quad . \quad (4.137)$$

4.4.2 Relation to Oppenheim's Homomorphic Filtering

Oppenheim & Schaffer (1975) discuss the processing of homomorphic systems. These are represented by algebraically linear transformations between input and output vector spaces. For monochromatic images Oppenheim & Schaffer (1975) suggest a homomorphic filtering which is based on the assumption that illumination generally does not vary rapidly across a scene, reflectance however does. They assume the spatial illumination distribution to be a low-frequency signal, and the reflectance on the other hand as a high-frequency signal. Consequently the – irrelevant – illumination signal can be filtered in the frequency domain after Fourier image transform.

For remotely sensed aerial imagery, however, we *do* expect rapid changes in illumination, such as shadows or inclinations of objects small compared to the resolution. Therefore we leave the *spatial* dimension unchanged, but apply the above argument to the *spectral* dimension of multispectral imagery. The assumption of a dichromatic illumination model and our experimental results show that the illumination variance *spectra* are of low frequency, and can thus be filtered. Here we suggest not to filter a complete frequency band but rather specific variance vectors which can be obtained from atmospheric physics.

4.4.3 Correspondence to the Mahalanobis Distance

Let the illumination covariance Σ_I in the logarithmic spectral space be described by $K < N$ significant eigenvectors as

$$\Sigma_I = \mathbf{E} \begin{bmatrix} \sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \sigma_K^2 & & \\ & & & \epsilon & \\ 0 & & & & \ddots \\ & & & & & \epsilon \end{bmatrix} \mathbf{E}^T \quad (4.138)$$

Then the Mahalanobis distance (Eq. 2.28 on page 24) would suggest a distance weighted with the inverse of Σ_I . However, for $\epsilon \rightarrow 0$ the covariance Σ_I becomes singular and thus its inverse Σ_I^{-1} undefined. One could be tempted to employ the pseudo-inverse Σ^+ (Appendix A.1 on page 177) instead, where the significant eigenvalues are inverted $\sigma_i^2 \mapsto \sigma_i^{-2}$ and the non-significant eigenvalues are set to zero $\epsilon \mapsto 0$. But the pseudo-inverse would yield a weighting scheme just opposite to the intended: The directions with vanishing variance and thus with most significance would not contribute to the distance at all.

Therefore we choose a weight matrix

$$\mathbf{P} = \mathbf{E} (\epsilon \Sigma_I^{-1}) \mathbf{E}^T \quad (4.139)$$

$$= \mathbf{E} \begin{bmatrix} \epsilon/\sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \epsilon/\sigma_K^2 & & \\ & & & \epsilon/\epsilon & \\ 0 & & & & \ddots \\ & & & & & \epsilon/\epsilon \end{bmatrix} \mathbf{E}^T = \mathbf{E} \begin{bmatrix} \epsilon/\sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \epsilon/\sigma_K^2 & & \\ & & & 1 & \\ 0 & & & & \ddots \\ & & & & & 1 \end{bmatrix} \mathbf{E}^T \quad (4.140)$$

where the directions with vanishing variance are weighted with unity and the remaining directions by a lesser factor according to their variance. This is equal to the above introduced projector (Eq. 4.98 on page 64) except the reverse order of eigenvectors.

For $\epsilon \rightarrow 0$ the weight matrix \mathbf{P} becomes

$$\mathbf{P} = \mathbf{E} \begin{bmatrix} 0 & & & & & 0 \\ & \ddots & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ 0 & & & & & 1 \end{bmatrix} \mathbf{E}^T \quad (4.141)$$

which has the advantage of being free of parameters. Instead of estimating a Mahalanobis-like weight using the parameters σ_i^2 and ϵ we can achieve invariance by setting $\epsilon \mapsto 0$.

4.4.4 Processing Costs and Some Programming Considerations

The above presented mapping is fast because apart from the logarithmic and exponential function it is a simple matrix multiplication between a fixed matrix \mathbf{P} and the spectra \mathbf{x} of all considered pixels. The matrix multiplication can be optimized by exploiting the symmetry property of \mathbf{P} . The cost to compute the projector \mathbf{P} is negligible since it has to be done only once and is independent of the number of spectra to be processed. So, the per-pixel processing costs are the same as *e.g.* for the weights matrix in the Mahalanobis distance (see Eq. 2.28 on page 24) used for Maximum Likelihood Classification.

For a particularly convenient notation, let all spectra \mathbf{x} of a given multispectral image be arranged as columns in a matrix \mathbf{X} , where X_{ij} denotes the value of the j th spectrum at the i th spectral band. Then in the logarithmic domain, the mapping is a simple multiplication of two matrices:

$$\mathbf{X} \mapsto \exp(\mathbf{P} \ln \mathbf{X}) \quad (4.142)$$

In this way the algorithm was implemented in the PVWAVE interpreter language (produced by ‘Visual Numerics, Inc.’) as well as in the IDL interpreter language (produced by ‘Research Systems, Inc.’), and made usable within the ENVI remote sensing image processing environment (produced by ‘Better Solutions Consulting’). The above notation is very useful since both interpreter languages allow arithmetic manipulations of whole arrays without explicit looping over single array indices.

For the precomputation of \mathbf{P} from a given relative direct illumination spectrum \mathbf{n} , the generalized inverse \mathbf{Q}^+ can be computed by Singular Value Decomposition of \mathbf{Q} (see Eq. 4.75 on page 60). If only one ‘brightness’ and one ‘color’ component is filtered, however, the computation of $\mathbf{P} = \mathbf{V}\mathbf{U}$ can be done less costly using Eq. 4.32 on page 53 and Eq. 4.103 on page 65.

Let us consider the case of a specialized routine written *e.g.* in C with pixelwise filtering. Then, instead of performing the multiplication $\mathbf{P} \ln \mathbf{x}$ which takes $N \times N$ multiplications, it is more efficient to compute the equivalent $\ln \mathbf{x} - \mathbf{Q}(\mathbf{Q}^+ \mathbf{x})$, because $(\mathbf{Q}^+ \mathbf{x})$ requires only $N \times K$ multiplications (with $K < N$ and usually $K = 2$), and then $\mathbf{Q}(\mathbf{Q}^+ \mathbf{x})$ the same, so that overall only $2K \times N$ multiplications are required. So the cost is decreased by a factor $N^2/2KN = N/2K$, *e.g.* a factor 2.5 for the case of $N = 10$ multispectral bands and filtering with $K = 2$ for one ‘brightness’ and one ‘color’ component.

Particularly a pure ‘brightness’ filtering, *i.e.*, the computation of $[\mathbf{U} \ln \mathbf{x}]_i$, is much more efficient by computing the equivalent $[\ln x_i - \frac{1}{N} \sum_{i=1}^N \ln x_i]$ which requires only one multiplication (note that $\mathbf{U} \ln \mathbf{x} = \mathbf{x} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T \ln \mathbf{x}$, and $\hat{\mathbf{u}} = \frac{1}{N}\mathbf{u}$, and $\mathbf{u}^T \mathbf{x} = \sum_{i=1}^N \ln x_i$, see Eq. 4.28 and Eq. 4.32 on page 53).

4.4.5 Processing of the Spectra After Prior Filtering

After filtering of the surface orientation related variability, the filtered spectra can be subjected to whatever image processing routine, particularly to spectral classification, which then should be unaffected by surface orientation.

One has to bear in mind, though, that the filtered spectral data is no longer of full dimension N , but rather only $N - K$, as set out above. If *e.g.* a maximum likelihood classification is carried out and spectral training samples for a class a are selected, the resulting sample covariance matrix Σ_a will be rank-deficient and thus cannot be inverted. The singularity of a sample covariance matrix is as well as the *Hughes phenomenon* (Shahshahani & Landgrebe 1994) the cause of which is a sample size too small to determine a large covariance matrix. In our case, however, the matrix singularity is not caused by insufficient sample size but by prior filtering. Therefore we are allowed to employ the pseudo inverse instead.

After filtering, the variance of the surface-orientation-caused variability vectors is zero, and thus the covariance matrix Σ of a set of filtered spectra $\exp(\mathbf{P} \ln \mathbf{x})$ is singular. The pseudo-inverse Σ^+ is formed by inverting all eigenvalues which are non-zero, and leaving those which are zero (Appendix A.1 on page 177). So when the Mahalanobis distance (Eq. 2.28 on page 24) is computed as $\mathbf{d}\Sigma^+\mathbf{d}$ from the spectral distance \mathbf{d} , the dimensions corresponding to the filtered surface orientation variance components are weighted with zero and thus not taken into account at all which is the desired effect. Thus we can conclude that inversion of a covariance matrix Σ computed from filtered spectral samples is impossible but that the pseudo inverse Σ^+ will work fine in the sense of a meaningful weighting and suppression of the filtered dimensions.

For Maximum Likelihood classification Eq. 2.34 on page 25 the expression $\ln |\Sigma|$ is essential. For a singular matrix Σ computed from filtered spectral samples, however, the determinant $|\Sigma|$ is zero and thus the logarithm undefined. Generally, the determinant of a $N \times N$ matrix can be computed as the product of its N singular values following a singular value decomposition (Hämmerlin & Hoffmann 1994). Now, for our case of the rank deficient matrix Σ , consequently the determinant must be computed as the product of only those $N - K$ singular values which are non-zero.

4.5 Results on Simulated Data

4.5.1 Simulated Noise-free Data

In order to demonstrate the illumination invariance of the mapping, we have selected four artificial spectral reflectance primitives (Fig. 4.5 on page 75, left panel):

1. the unity reflectance spectrum,
2. an ascending parabolic reflectance spectrum,
3. a step like reflectance spectrum as typical for vegetation at $\lambda \approx 0.7 \mu\text{m}$,
4. a sinusoidal reflectance spectrum as a primitive for higher frequency spectra.

All simulated spectra are sampled at $N = 40$ spectral bands. We have ‘exposed’ these reflectance spectra to various illuminations, *i.e.*, multiplied them with different contributions ν and μ of direct and diffuse illumination spectra according to Eq. 4.65 on page 59. The used relative illumination spectra \mathbf{n} and \mathbf{m} are shown in Fig. 4.4 and are physically realistic as pointed out in Section 3.1 on page 35.

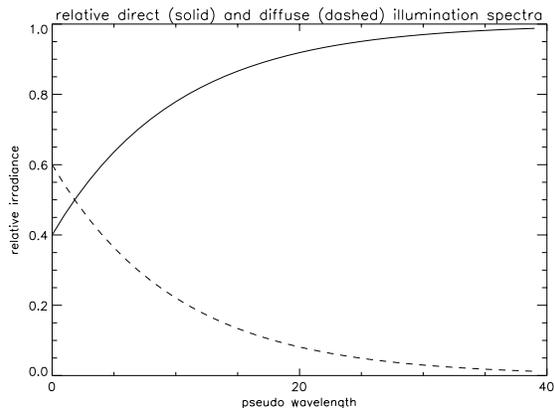


Figure 4.4: The assumed relative illumination spectra for relative direct (solid) and diffuse (dashed) illumination, sampled at $N = 40$ pseudo-wavelengths, typical for a wavelength range $\lambda \approx 0.4 - 2.5 \mu\text{m}$.

ance at different illuminations for a single reflectance spectrum, but the reflectance spectra 1 through 4 are still clearly distinguishable. In other words, if each spectral reflectance forms a spectral *class*, then the mapped spectra show vanishing in-class variance but remaining inter-class variance.

Each cluster in the spectral space formed by a given reflectance under varying illumination is contracted into nearly a singular point by the mapping. Hence a subsequent classification process can assign the reflectance spectra to different spectral classes independently of their unknown respective illumination. In order to illustrate the effect which the mapping has in the spectral feature space, we have visualized the $N = 40$ dimensional space by a projection into a three dimensional subspace. The optimal projection which captures most of the original variance in the data was established by a

The shown spectra are typical for a wavelength range $\lambda \approx 0.4 - 2.5 \mu\text{m}$. The such simulated spectra \mathbf{x} are shown in Fig. 4.5 (center panel). Only for reasons of display they have been normalized to $\mathbf{x} \mapsto \mathbf{x}/\|\mathbf{x}\|$ before plotting.

Then the discussed mapping (Eq. 4.80 on page 61) has been applied to the simulated spectra, with \mathbf{Q} made up from only the first four columns, *i.e.* $\mathbf{n}^0 \dots \mathbf{n}^3$, and the number of columns $K = 4$. Note that the sole input data to the mapping process is the relative direct illumination spectrum \mathbf{n} and the spectra \mathbf{x} to be mapped.

The results are shown in Fig. 4.5 (right panel) where *all* the mapped spectra of the center panel are plotted. Note that the mapped spectra show vanishing vari-

principal components analysis. The visualization in Fig. 4.7 on page 77 (top row) clearly shows that the simulated data of the four reflectance spectra is distributed on planes, and is contracted into points of vanishing in-cluster variance by the mapping.

4.5.2 Simulated Data With Random Noise

Critical noise sensitivity is a typical problem with color constancy algorithms and sometimes prohibits the use of linear algebraic methods. In order to test the robustness we have superimposed strong random noise on our simulated data (Fig. 4.6 on page 76, center panel) with a signal to noise ratio $\text{SNR} = x_i/\sigma(x_i) = 10$, *i.e.*, the standard deviation $\sigma(x_i)$ of the random noise is $\frac{1}{10}$ of the signal x_i . In logarithmic notation this corresponds to a noise level of: $10 \log(x_i^2/\sigma^2(x_i)) \text{ dB} = 20 \text{ dB}$.

Then the same mapping process was applied as described above. The results are shown in Fig. 4.6 (right panel). A principal components projection of the feature space into three dimensions is shown in Fig. 4.7 on page 77 in order to illustrate the mapping process also in the presence of noise.

In the presence of noise the mapping naturally cannot produce an invariant, but the performance can be evaluated by comparing the capability to discriminate different reflectance spectra under varying illumination before and after the mapping. The separability of two multivariate data clusters can be tested by the *Lawley-Hotelling trace criterion* (Press 1972, Duda & Hart 1973). It describes the distance between the clusters in the multidimensional feature space relative to the cluster sizes and directions of their respective covariance ellipsoids, assuming normal probability distributions. The trace criterion has been successfully used for optimal feature selection in remote sensing (Pyka & Steinnocher 1994). The squared relative distance T^2 between the clusters a and a' is computed as

$$d_S^{2*} = \text{tr}(\Sigma_b \Sigma_w^+) \quad (\text{see Eq. 2.45 on page 26}) \quad (4.143)$$

where the within-clusters covariance matrix $\Sigma_w = \frac{1}{2}(\Sigma_a + \Sigma_{a'})$ is the mean of the covariance matrices Σ_a and $\Sigma_{a'}$ of the two clusters, and the between-cluster covariance matrix $\Sigma_b = \frac{1}{2} \sum_{i=a,a'} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$ is the covariance matrix of the two cluster centers (mean spectra) \mathbf{m}_a and $\mathbf{m}_{a'}$, with the mean cluster center $\mathbf{m} = \frac{1}{2}(\mathbf{m}_a + \mathbf{m}_{a'})$ (compare Section 2.4 on page 25). In other words, the variance between the cluster centers is weighted by the inverse of the mean variance within the clusters, and then trace function takes the so called *sum of roots*, *i.e.*, the sum of the eigenvalues of $(\Sigma_b \Sigma_w^+)$,⁶ and this numerical value measures the ratio of between-cluster to within-cluster scatter in the directions of the eigenvectors (Duda & Hart 1973).

The mutual distances are summarized in Table 4.1 on page 78, tables 1 to 4. For the noise-free data before and after the mapping (Table 4.1, tables 1 and 2), the distances

⁶ The pseudo-inverse Σ_w^+ (see Appendix A.1) rather than the inverse Σ_w^{-1} has to be taken because the clusters are two-planes in \mathbb{R}^N (see Chapter 3). This means that the cluster covariance matrix Σ_a is only of rank $\text{rk } \Sigma_a = 2$, and thus a singular and not invertible matrix. As a matter of fact, the same reason prohibits the use of the better known *Bhattacharyya distance* (Richards 1993) for measuring the effective distance between clusters, as the latter works on the determinant of Σ_a which is vanishing for singular matrices such as formed by two-planes in \mathbb{R}^N .

The sometimes necessary employment of the pseudo-inverse Σ_w^+ instead of the inverse Σ_w^{-1} for parametric classification purposes is also mentioned by Kohonen (1995).

have increased almost infinitely (with respect to the computation accuracy of DOUBLE type variables) as the clusters have essentially been contracted to single points which represent the invariant transformed signals. For the random noise data before and after the mapping (Table 4.1, tables 3 and 4), the distances have still increased by one order of magnitude, except between reflectance spectra 1 and 2. For the last case we note that the differences in reflectance are of so low frequency and similar to the illumination spectra that they could obviously be explained by varying illumination and were removed as such.

Another way of evaluating the performance of the suggested mapping is to analyze the covariance matrices of the simulated clusters before and after the mapping is applied. In Fig. 4.8 on page 79 we see the entries of the $N \times N$ covariance matrices as heights in surface plots. The covariance matrix of the unmapped clusters (left panel) show strong correlations, *i.e.*, high values in the off diagonal entries. After the mapping is applied (right panel), all the variances have vanished for the noise-free case (top). For the random noise data we observe that the variances have been reduced by two orders of magnitude and are distributed only on the diagonal entries, *i.e.*, the correlations have been filtered out and only the uncorrelated random noise is left.

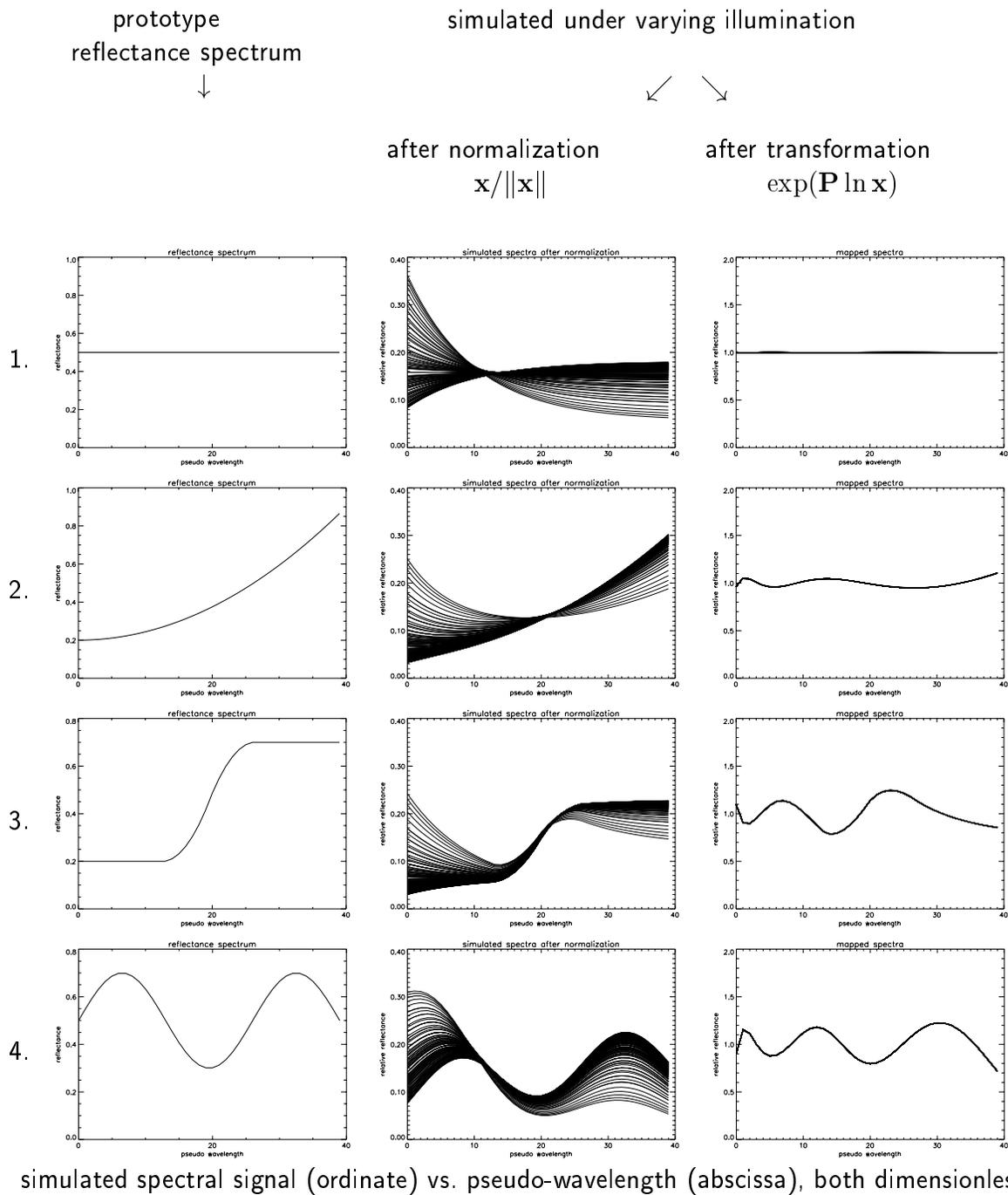


Figure 4.5: The simulated reflectance spectra 1 through 4 (left panel); the simulated noise-free ‘observed spectra’, i.e., the artificial reflectance spectra multiplied with various illumination spectra (center panel); and all these spectra after the mapping was applied (right panel).

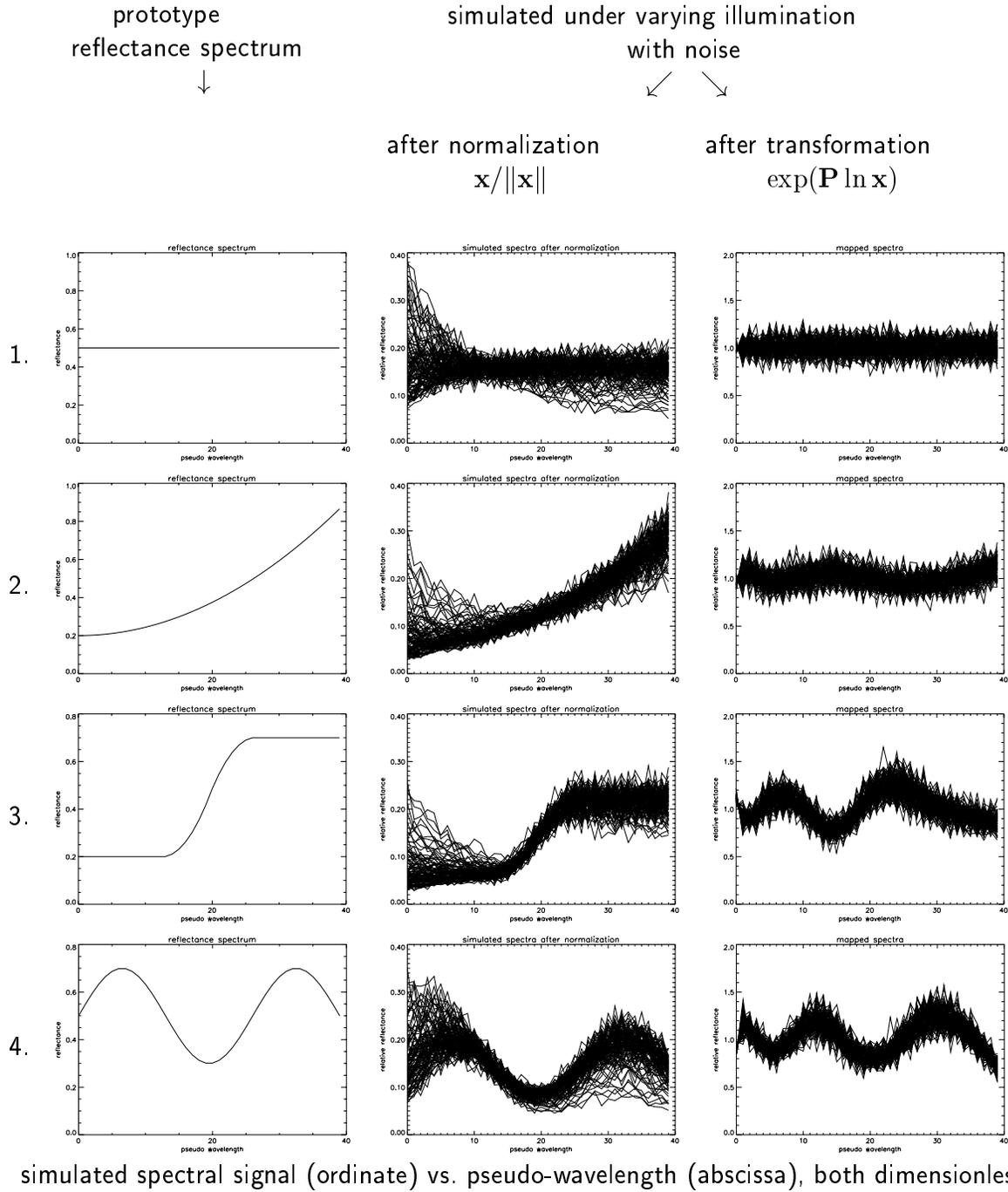


Figure 4.6: The simulated reflectance spectra 1 to 4 (left panel); the simulated noisy ‘observed spectra’, i.e., the artificial reflectance spectra multiplied with various illumination spectra and random noise ($\text{SNR} = 10 \hat{=} 20 \text{ dB}$) (center panel); and all these spectra after the mapping was applied (right panel).

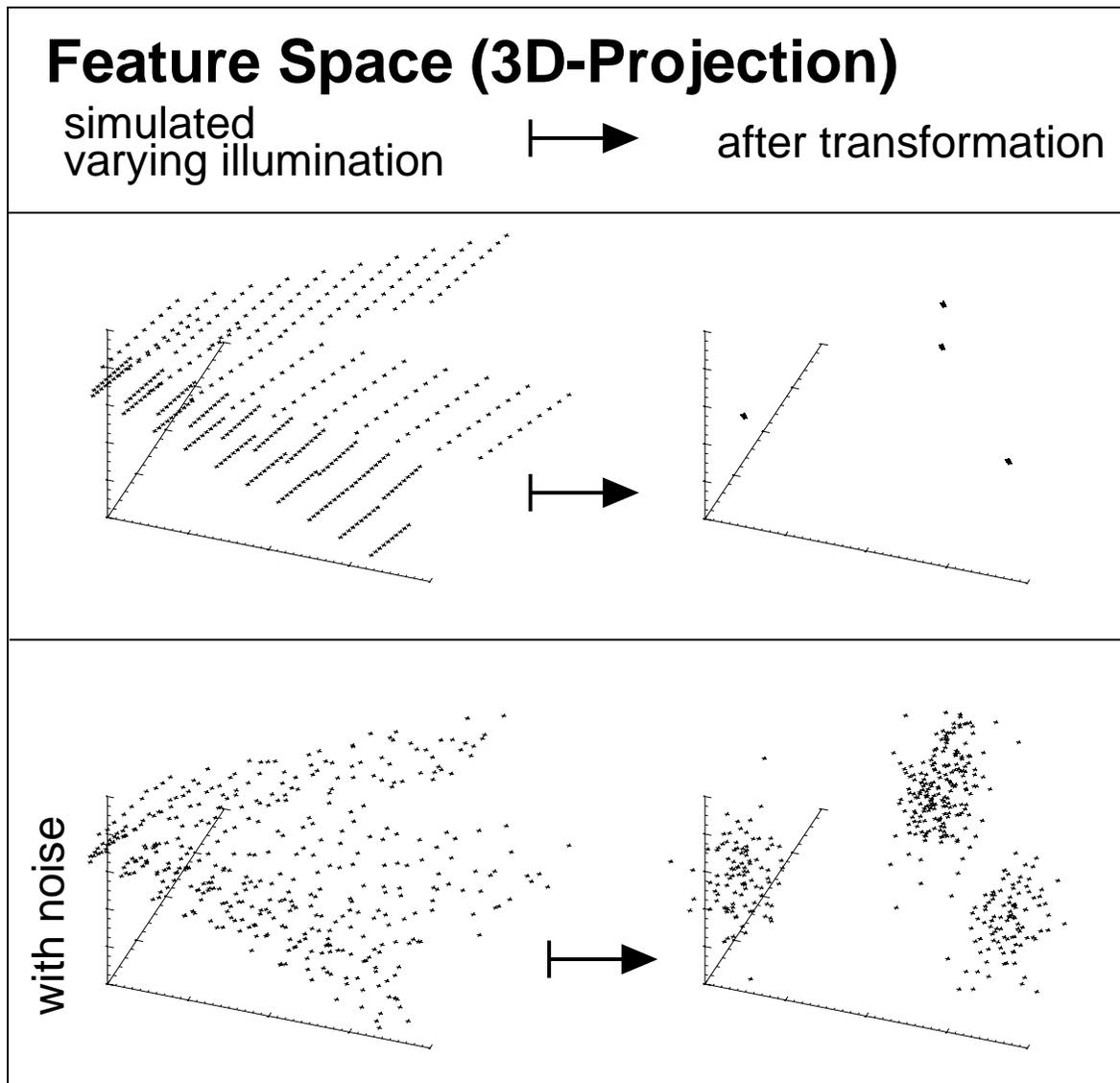


Figure 4.7: Feature space representation of the spectra in Fig. 4.5 and Fig. 4.6: The multivariate data clusters in the feature space which has been projected into a three dimensional subspace by principal component transformation.

The simulated data of all four reflectance spectra without noise (top left) distributed on four two-dimensional planes,

the respective mapped spectra (top right) with vanishing in-cluster variance,

the simulated data of all four reflectance spectra with noise $\text{SNR} = 10 \hat{=} 20 \text{ dB}$ (bottom left),

the respective mapped spectra (bottom right).

TABLE 1 LAWLEY-HOTELLING DISTANCES (WITHOUT NOISE)				
	C_1	C_2	C_3	C_4
C_1	0.0	$7.3 \cdot 10^4$	$7.2 \cdot 10^4$	$7.3 \cdot 10^4$
C_2		0.0	$7.2 \cdot 10^4$	$7.4 \cdot 10^4$
C_3			0.0	$7.4 \cdot 10^4$
C_4				0.0

TABLE 2 LAWLEY-HOTELLING DISTANCES (AFTER MAPPING)				
	C_1	C_2	C_3	C_4
C_1	0.0	$1.2 \cdot 10^{12}$	$1.2 \cdot 10^{14}$	$2.0 \cdot 10^{14}$
C_2		0.0	$1.1 \cdot 10^{14}$	$2.2 \cdot 10^{14}$
C_3			0.0	$5.8 \cdot 10^{14}$
C_4				0.0

TABLE 3 LAWLEY-HOTELLING DISTANCES (WITH NOISE)				
	C_1	C_2	C_3	C_4
C_1	0.0	$4.4 \cdot 10^4$	$5.1 \cdot 10^4$	$4.9 \cdot 10^4$
C_2		0.0	$4.6 \cdot 10^4$	$5.6 \cdot 10^4$
C_3			0.0	$5.8 \cdot 10^4$
C_4				0.0

TABLE 4 LAWLEY-HOTELLING DISTANCES (AFTER MAPPING)				
	C_1	C_2	C_3	C_4
C_1	0.0	$1.9 \cdot 10^4$	$2.2 \cdot 10^5$	$1.9 \cdot 10^5$
C_2		0.0	$4.0 \cdot 10^5$	$2.7 \cdot 10^5$
C_3			0.0	$5.2 \cdot 10^5$
C_4				0.0

Table 4.1: In presence of illumination variability, the four simulated spectral signatures form four clusters C_1 through C_4 (see Fig. 4.7 on the preceding page).

The mutual cluster distances (separations) before (left-hand side) and after the mapping (right-hand side), simulated without (top row) and with noise (bottom row) are given in the tables. The distances between the clusters are computed relative to the within-cluster scatter radii (Eq. 4.143 on page 73).

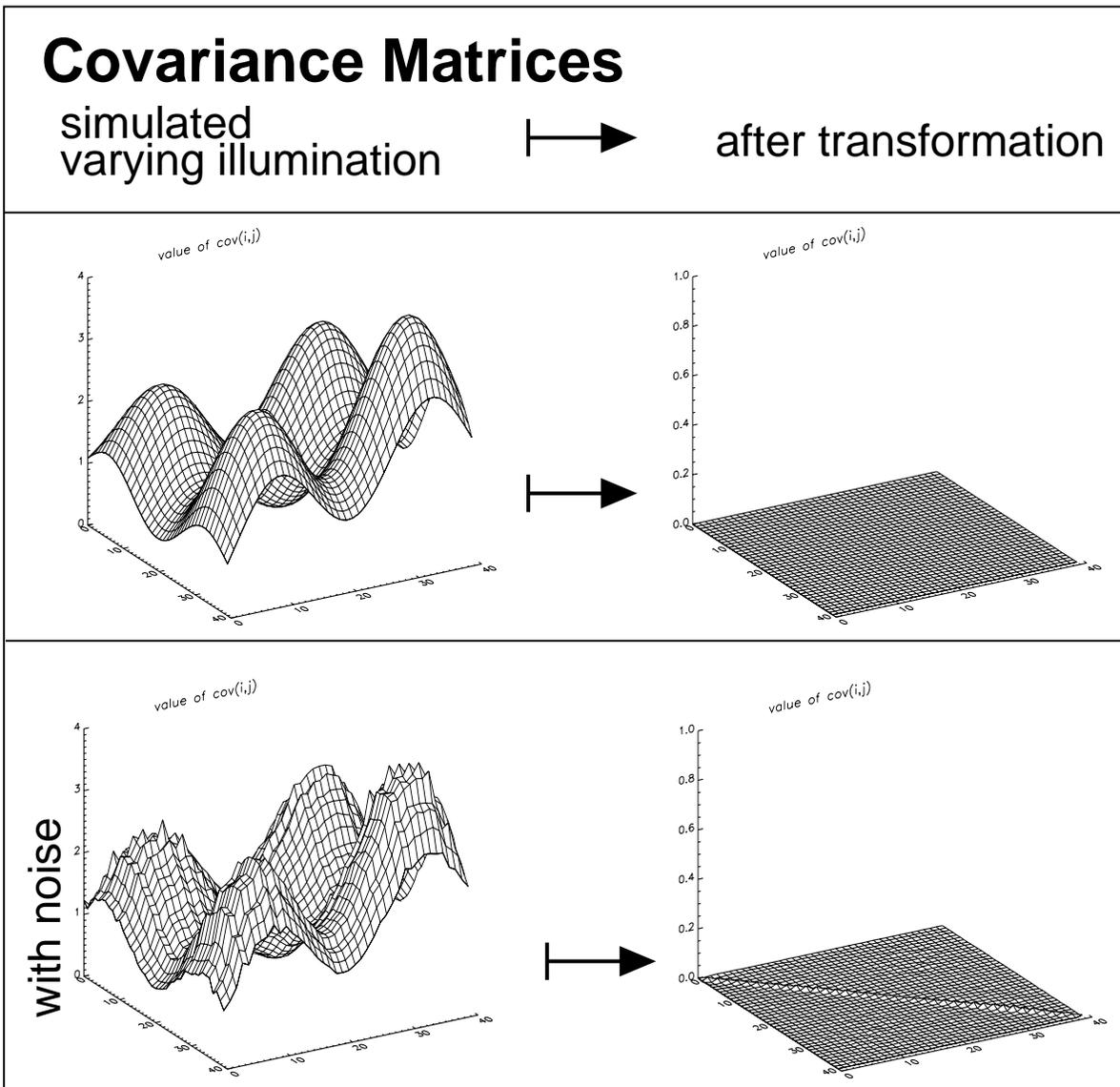


Figure 4.8: The covariance matrices (40×40) of the simulated spectra (Fig. 4.5 and Fig. 4.6) shown as surface plots:

covariances of simulated data for spectrum 4 without noise (top left),
 covariances of the respective mapped spectra (top right) which have vanished,
 covariances of simulated data for spectrum 4 with noise $SNR=10$ (bottom left),
 covariances of the respective mapped spectra (bottom right) which are reduced to uncorrelated variances on the matrix diagonal.

4.6 Analysis of Random Noise

In the previous sections of this chapter we have introduced the mapping $\mathbf{x} \mapsto \mathbf{t} = \exp(\mathbf{P} \ln \mathbf{x})$ and shown that it produces a spectral descriptor which is approximately⁷ invariant to varying contributions of direct and diffuse illumination. Now we investigate how the invariant is affected by errors in the spectral input signal \mathbf{x} . To this aim we follow the propagation of errors through the mapping. At first we consider random errors, and treat multiplicative and additive noise. We will address the effects of systematic errors in the following section.

Let the spectral input vector \mathbf{x} contain random error, and let $\mathbf{S}(\mathbf{x})$ be the covariance matrix of this error in \mathbf{x} . The vector \mathbf{x} is mapped to vector \mathbf{t} :

$$\mathbf{x} \mapsto \mathbf{t} = e^{\mathbf{P} \ln \mathbf{x}} \quad (4.144)$$

through three different steps. The first step is a non-linear transform and the error propagation is estimated by linearization:

$$\mathbf{x} \mapsto \ln \mathbf{x} \quad (4.145)$$

leads to a linearized error propagation as

$$\mathbf{S}(\mathbf{x}) \mapsto \mathbf{S}(\ln \mathbf{x}) = \mathbf{J}_1 \mathbf{S}(\mathbf{x}) \mathbf{J}_1 \quad (4.146)$$

where \mathbf{J}_1 is the Jacobi matrix of the partial derivatives

$$J_{1,ij} = \frac{\partial \ln x_i}{\partial x_j} = \frac{1}{x_i} \delta_{ij} \quad . \quad (4.147)$$

The next step is a linear transform:

$$\ln \mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{P} \ln \mathbf{x} \quad (4.148)$$

with corresponding error propagation as

$$\mathbf{J}_1 \mathbf{S} \mathbf{J}_1 \mapsto \mathbf{P} \mathbf{J}_1 \mathbf{S} \mathbf{J}_1 \mathbf{P} \quad . \quad (4.149)$$

Finally, the exponentiation is applied:

$$\mathbf{P} \ln \mathbf{x} \mapsto e^{\mathbf{P} \ln \mathbf{x}} \quad (4.150)$$

$$\text{or } \ln \mathbf{t} \mapsto e^{\ln \mathbf{t}} = \mathbf{t} \quad (4.151)$$

leading to a linearized error propagation as

$$\mathbf{P} \mathbf{J}_1 \mathbf{S} \mathbf{J}_1 \mathbf{P} \mapsto \mathbf{J}_e \mathbf{P} \mathbf{J}_1 \mathbf{S}(\mathbf{x}) \mathbf{J}_1 \mathbf{P} \mathbf{J}_e \quad (4.152)$$

where \mathbf{J}_e is the Jacobi matrix of the partial derivatives

$$J_{e,ij} = \frac{\partial e^{\ln t_i}}{\partial \ln t_j} = e^{\ln t_i} \delta_{ij} = t_i \delta_{ij} \quad . \quad (4.153)$$

⁷ The accuracy depends on the number of filtered components K (Eq. 4.72 on page 60).

Thus the mapping $\mathbf{x} \mapsto \mathbf{t}$ propagates the errors as

$$\mathbf{S}(\mathbf{x}) \mapsto \mathbf{S}(\mathbf{t}) = \mathbf{J}_e \mathbf{P} \mathbf{J}_1 \mathbf{S}(\mathbf{x}) \mathbf{J}_1 \mathbf{P} \mathbf{J}_e \quad (4.154)$$

where \mathbf{J}_1 , \mathbf{J}_e and \mathbf{S} are symmetric matrices depending on the signal \mathbf{x} . Note that a diagonal error matrix $\mathbf{S}(\mathbf{x})$, *i.e.*, with vanishing covariances, is not mapped into a diagonal error matrix $\mathbf{S}(\mathbf{t})$.

4.6.1 Propagation of Multiplicative Noise

We consider independent multiplicative noise with a variance proportional to the signal strength x_i :

$$S_{ij} = (\epsilon_i x_i)^2 \delta_{ij} \quad . \quad (4.155)$$

Then in the logarithmic domain the error matrix is still diagonal and becomes

$$[\mathbf{J}_1 \mathbf{S} \mathbf{J}_1]_{ii} = \frac{1}{x_i} (\epsilon_i x_i)^2 \frac{1}{x_i} = \epsilon_i^2 \quad (4.156)$$

and for identical variances $\epsilon_i = \epsilon$ this reduces to

$$\mathbf{S}(\ln \mathbf{x}) = \mathbf{J}_1 \mathbf{S}(\mathbf{x}) \mathbf{J}_1 = \epsilon^2 \mathbf{I} \quad . \quad (4.157)$$

In the logarithmic domain the error matrix of the transformed vector

$$\mathbf{S}(\ln \mathbf{t}) = \mathbf{P} \mathbf{J}_1 \mathbf{S}(\mathbf{x}) \mathbf{J}_1 \mathbf{P} \quad (4.158)$$

$$= \mathbf{P} \begin{bmatrix} \ddots & & \\ & \epsilon^2 & \\ & & \ddots \end{bmatrix} \mathbf{P} \quad (4.159)$$

for multiplicative error is independent of the signal \mathbf{x} and can thus be computed once for all vectors \mathbf{x} . For identical variances $\epsilon_i = \epsilon$ this reduces to $\mathbf{S}(\ln \mathbf{t}) = \epsilon^2 \mathbf{P} \mathbf{I} \mathbf{P} = \epsilon^2 \mathbf{P}$. For a particular component t_i of the transformed vector we have

$$\sigma^2(\ln t_i) = \frac{\sigma^2(t_i)}{t_i^2} = \sum_j P_{ij}^2 \frac{\sigma^2(x_i)}{x_i^2} \quad (4.160)$$

$$= \epsilon^2 \sum_j P_{ij}^2 < \epsilon^2 \quad (4.161)$$

because Eq. 4.130 on page 68 states that the line square sums $\sum_j P_{ij}^2$ are always smaller than unity and equal to $(N - K)/N$ on average. Thus, the relative error caused by multiplicative noise is reduced by the transform.

$$(4.162)$$

4.6.2 Propagation of Additive Noise

Here we consider additive noise with variance σ_i independent of the signal x_i and the other vector entries x_j . The error covariance matrix for \mathbf{x} is then diagonal

$$S_{ij} = \epsilon_i^2 \delta_{ij} \quad . \quad (4.163)$$

For identical variances $\epsilon_i = \epsilon$ this is simply

$$\mathbf{S}(\mathbf{x}) = \epsilon^2 \mathbf{I} \quad (4.164)$$

which in the logarithmic domain becomes

$$[\mathbf{J}_1 \mathbf{S}(\mathbf{x}) \mathbf{J}_1]_{ii} = \frac{\epsilon^2}{x_i^2} \quad . \quad (4.165)$$

This is dependant on the signal \mathbf{x} and must hence be computed individually for all vectors \mathbf{x} .

4.6.3 Expected Relative Accuracy of the Transformed Signals

We now compare the expected accuracy of the transformed signal in the transformed logarithmic space $\ln \mathbf{t} = \mathbf{P} \ln \mathbf{x}$ versus in the transformed original space $\mathbf{t} = e^{\ln \mathbf{P} \ln \mathbf{x}}$ in order to decide which may be the more favorable. To this aim we compare the relative errors $\sigma^2(t_i)/t_i^2$ and $\sigma^2(\ln t_i)/(\ln t_i)^2$ of both spaces.

The variance transforms through the logarithm/exponentiation according to error propagation as

$$\sigma^2(t_i) = t_i^2 \sigma^2(\ln t_i) \quad (4.166)$$

$$\frac{\sigma^2(t_i)}{t_i^2} = \sigma^2(\ln t_i) \quad (4.167)$$

the left-hand expression is the relative variance of the signal in the original space. Adding a denominator to the right-hand expression now yields two inequations:

$$\frac{\sigma^2(t_i)}{t_i^2} > \frac{\sigma^2(\ln t_i)}{(\ln t_i)^2} \quad \text{for } \ln t_i > 1, t_i > e \quad (4.168)$$

$$\frac{\sigma^2(t_i)}{t_i^2} < \frac{\sigma^2(\ln t_i)}{(\ln t_i)^2} \quad \text{for } \ln t_i < 1, t_i < e \quad . \quad (4.169)$$

These two inequations compare just the two relative errors we are interested in, and show that the inequality sign depends on the transformed signal t_i itself. However, we know that after applying the projector matrix \mathbf{U} to $\ln \mathbf{x}$ the expectation value of $\ln t_i = [\mathbf{U} \ln \mathbf{x}]_i$ vanishes: $\langle \ln t_i \rangle = 0 < 1$ (Eq. 4.118 on page 67).

Therefore we can expect the right-hand expression to be the larger one in most cases, and conclude that the transformed signal in the original space \mathbf{t} is the one with the relatively smaller error. **This suggests that the transformed original space $\mathbf{t} = e^{\mathbf{P} \ln \mathbf{x}}$ may be more favorable than the transformed logarithmic space $\ln \mathbf{t} = \mathbf{P} \ln \mathbf{x}$.**

4.7 Analysis of Systematic Errors

The calibration of multispectral sensors is performed individually for each spectral band i . It usually depends on an additive offset c_o (dark current plus minimum amplifier voltage) and a multiplicative factor c_1 (gain factor and voltage to radiance calibration). Both factors combine into a linear function between the measured signal and the calibrated radiance value.

The radiance is then converted into a Lambert reflectance value. This conversion requires measurement or estimation (*e.g.* by radiative transfer codes) of the irradiance from sun- and skylight, the atmospheric transmission, the path radiance, etc. The conversion in general is a non-linear function, due to consideration of multiple scattering and other code-specific computations.

However, as shown *e.g.* by Richter (1992), Hepp (1994), Kollwe (1995), for a number of cases it is a good first order approximation to assume a linear relation between radiance and reflectance value for a given situation of the data recording campaign. Then the path radiance (influenced by haze) determines the additive offset of the linear transform, whereas visibility and overall irradiance determine the multiplicative factor. From the simulation cases in Richter (1992) we can see that the dependence on the sensor observation angle is comparatively small on clear days (which chosen for imaging flight campaigns). Linear approximation has been applied in a large number of cases (*e.g.* Schott et al. (1988), Moran et al. (1990), Wegener (1990), Hall et al. (1991), Jahnen & Grassl (1991)), last not least due to its computational simplicity and numerical robustness.

Combining a linear calibration function with a linear relation between radiance and reflectance values again yields a linear function. This means that in the following discussion of systematic error propagation we can concentrate on two cases:

- ▶ a **systematic additive error** $dc_{o,i}$ added to the reflectance vector entry x_i (caused by uncertainties of the detector dark current, path radiance, and minimum detector voltage),
- ▶ a **systematic multiplicative error** $dc_{1,i}$ multiplied with the reflectance vector entry x_i (caused by uncertainties in the estimation of radiative transmission, irradiance or visibility),

with both constants for each spectral band i centered at wavelength λ_i .

Systematic errors dx_j propagate into the transformed vector entry error dt_i as follows:

The mapping $\mathbf{x} \mapsto \mathbf{t}$ with

$$t_i = e^{\sum_j P_{ij} \ln x_j} \quad (4.170)$$

has the total derivative

$$dt_i = \sum_j \frac{\partial t_i}{\partial x_j} dx_j \quad (4.171)$$

$$= t_i \sum_j P_{ij} \frac{1}{x_j} dx_j \quad (4.172)$$

and thus leads to a relative deviation of

$$\frac{dt_i}{t_i} = \sum_j P_{ij} \frac{1}{x_j} dx_j \quad (4.173)$$

whereas in the log space we have a total derivative of

$$d(\ln t_i) = \sum_j \frac{\partial \ln t_i}{\partial x_j} dx_j \quad (4.174)$$

$$= \sum_j P_{ij} \frac{1}{x_j} dx_j \quad . \quad (4.175)$$

4.7.1 Propagation of Systematic Multiplicative Errors

A systematic multiplicative error $dc_{1,i}$ in spectral band i yields an error in x_i of $dx_i = dc_{1,i} x_i$, or $dx_i/x_i = dc_{1,i}$. Using Eq. 4.172, this leads to

$$dt_i = t_i \sum_j P_{ij} dc_{1,j} \quad . \quad (4.176)$$

For the special case that the systematic multiplicative error $dc_{1,j} = dc_1$ is equal for all spectral bands j we get

$$dt_i = t_i dc_1 \sum_j P_{ij} \quad (4.177)$$

$$= 0 \quad (\text{because of Eq. 4.114 on page 66}) \quad (4.178)$$

and the systematic multiplicative error cancels out in the transformed signal \mathbf{t} !

We see that the relative deviation dt_i/t_i can be computed independently of the input signal \mathbf{x} . In the log space, however, even the absolute deviation is independent of \mathbf{x} .

$$d(\ln t_i) = \sum_j P_{ij} dc_{1,j} \quad (4.179)$$

Thus, a difference

$$d_i = \ln t_i - \ln t'_i \quad (4.180)$$

with systematic multiplicative error becomes

$$\implies (\ln t_i + d(\ln t_i)) - (\ln t'_i + d(\ln t'_i)) \quad (4.181)$$

$$= \ln t_i - \ln t'_i + \sum_j P_{ij} dc_{1,j} - \sum_j P_{ij} dc_{1,j} \quad (4.182)$$

$$= \ln t_i - \ln t'_i \quad . \quad (4.183)$$

This means that a difference $d_i = \ln t_i - \ln t'_i$ in the transformed logarithmic space $\ln \mathbf{t}$ is invariant against systematic multiplicative error $dc_{i,j}$ even if this error is different for each spectral band j !

4.7.2 Propagation of Systematic Additive Errors

For a systematic additive error we have $dx_i = dc_{o,i}$. Using Eq. 4.172 this leads to

$$dt_i = t_i \sum_j P_{ij} \frac{1}{x_j} dc_{o,j} \quad . \quad (4.184)$$

Here the deviation in t_i depends on the respective signal \mathbf{x} . In the log space we have

$$d(\ln t_i) = \frac{dt_i}{t_i} = \sum_j P_{ij} \frac{1}{x_j} dc_{o,j} \quad . \quad (4.185)$$

Obviously the impact of systematic additive error will be the stronger, the smaller the input signal x is. There is a special case when the ratio of additive systematic error and signal is roughly the same for all wavelengths: $\frac{dc_{o,j}}{x_j} \sim \xi$. Then

$$d(\ln t_i) = \xi \sum_j P_{ij} \quad (4.186)$$

$$= 0 \quad (4.187)$$

because the line sums of \mathbf{P} vanish (Eq. 4.114 on page 66).

$$(4.188)$$

4.8 Expected Performance of the Transform

4.8.1 Performance for a Given Pair of Reflectance Spectra

Let $\mathbf{d} = \mathbf{m}_a - \mathbf{m}_{a'}$ be the distance between two cluster centers \mathbf{m}_a and $\mathbf{m}_{a'}$ in the logarithmic spectral space, and let $\mathbf{\Sigma}$ be the within-cluster covariance matrix assumed to be equal for both clusters a and a' in the logarithmic spectral space. Thus $\text{tr } \mathbf{\Sigma}$ is the squared scattering radius of the two clusters (see section 2.4). Then the normalized spectral distance between the clusters is $\mathbf{d}^T \mathbf{d} / \text{tr } \mathbf{\Sigma}$. In other words, we evaluate the ratio of the squared distance between the two clusters over the squared scatter radius of the clusters. This should be as large as possible for unambiguous classification. In order to compare classification in the original versus the transformed space, we consider the normalized spectral distances in both spaces and will prefer the space with the larger one.

The transformation $\ln \mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{P} \ln \mathbf{x}$ in the logarithmic space will map the distance between the clusters as $\mathbf{d} \mapsto \mathbf{P} \mathbf{d}$, and the clusters' covariance matrix as $\mathbf{\Sigma} \mapsto \mathbf{P} \mathbf{\Sigma} \mathbf{P}^T = \mathbf{P} \mathbf{\Sigma} \mathbf{P}$. Consequently, the distances for the original clusters (left-hand) versus the transformed clusters (right-hand) are

$$\frac{\mathbf{d}^T \mathbf{d}}{\text{tr } \mathbf{\Sigma}} \longleftrightarrow \frac{\mathbf{d}^T \mathbf{P} \mathbf{d}}{\text{tr } \mathbf{P} \mathbf{\Sigma} \mathbf{P}} \quad (4.189)$$

We imagine the within-cluster scatter $\mathbf{\Sigma}$ to be the result of variation in the spectral signature \mathbf{r} with a proper covariance matrix $\mathbf{\Sigma}_R$, and of illumination variability with

a covariance matrix Σ_I . Since reflectance and illumination spectra are *multiplied*, we can model the overall within-cluster covariance in the logarithmic space as the *sum* of illumination variance and reflection variance: $\Sigma = \Sigma_I + \Sigma_R$. As shown previously, the illumination variance can be described by $K < N$ eigenvectors. Let these eigenvectors be column vectors in the matrix \mathbf{E} , then Σ_I can be written as

$$\Sigma_I = \mathbf{E} \begin{bmatrix} \sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \sigma_K^2 & & \\ & & & 0 & \\ 0 & & & & \ddots \\ & & & & & 0 \end{bmatrix} \mathbf{E}^T \quad (4.190)$$

and is filtered by the projector: $\mathbf{P}\Sigma_I\mathbf{P} = 0$ (Eq. 4.93 on page 63).

We assume the reflection covariance ellipsoid as ‘spherical’: $\Sigma_R = \epsilon^2\mathbf{I}$, *i.e.*, uncorrelated and of equal variance in all spectral bands. Then

$$\text{tr } \Sigma_R = \epsilon^2 \text{tr } \mathbf{I} = N\epsilon^2 \quad , \quad (4.191)$$

and the trace of the illumination covariance matrix is

$$\text{tr } \Sigma_I = \sigma_1^2 + \dots + \sigma_K^2 \quad . \quad (4.192)$$

So

$$\text{tr } \mathbf{P}\Sigma\mathbf{P} = \text{tr } \mathbf{P}(\Sigma_I + \Sigma_R)\mathbf{P} \quad (4.193)$$

$$= \text{tr } \mathbf{P}\Sigma_R\mathbf{P} \quad (4.194)$$

$$= \epsilon^2 \text{tr } \mathbf{P}\mathbf{I}\mathbf{P} \quad (4.195)$$

$$= \epsilon^2 \text{tr } \mathbf{P} \quad (4.196)$$

$$= \epsilon^2(N - K) \quad . \quad (4.197)$$

Thus Eq. 4.189 becomes

$$\frac{\mathbf{d}^T\mathbf{d}}{\text{tr } \Sigma_I + \epsilon^2N} \longleftrightarrow \frac{\mathbf{d}^T\mathbf{P}\mathbf{d}}{\epsilon^2(N - K)} \quad (4.198)$$

The performance can be seen as a trade off between shrinking distance versus vanishing variance:

$$\frac{\mathbf{d}^T\mathbf{P}\mathbf{d}}{\mathbf{d}^T\mathbf{d}} \leq 1 \text{ and } \frac{\epsilon^2(N - K)}{\text{tr } \Sigma_I + \epsilon^2N} \leq 1 \quad (4.199)$$

For the noise-free case, *i.e.*, $\epsilon \rightarrow 0$, the transform increases the effective cluster distance indefinitely (provided that the two clusters centers are still different after the transform, *i.e.*, $\mathbf{d}^T\mathbf{P}\mathbf{d} \neq 0$), while the original distance has an upper bound. With noise, we distinguish two cases (Fig. 4.9):

1. the relative distance of the transformed spectra is greater than for the original spectra for all noise levels $\epsilon \in \mathbb{R}^+$,
2. the relative distance of the transformed spectra is greater for all noise levels $\epsilon < \epsilon_o$ below a certain threshold ϵ_o .

So the transform will always improve the separability for low noise.

Note that ϵ_o depends strongly on $\frac{\mathbf{d}^T \mathbf{P} \mathbf{d}}{\mathbf{d}^T \mathbf{d}}$ and thus on the direction of the vector \mathbf{d} .

For the example in Fig. 4.9 we have used $N = 40, K = 2, \sigma_1 = 1.0, \sigma_2 = 0.1$, and the two compared reflectance spectra are (2),(3) (left) and (1),(2) (right) from Fig. 4.5 on page 75.

In conclusion, for any given pair of reflectance spectra and given reflectance error ϵ in logarithmic spectral space, we can determine whether the transform will improve the separability or not. The separability as defined in Eq. 4.189 on page 85 is relevant for Euclidean distance based classification such as *e.g.* most unsupervised clustering algorithms (Section 2.6 on page 29).

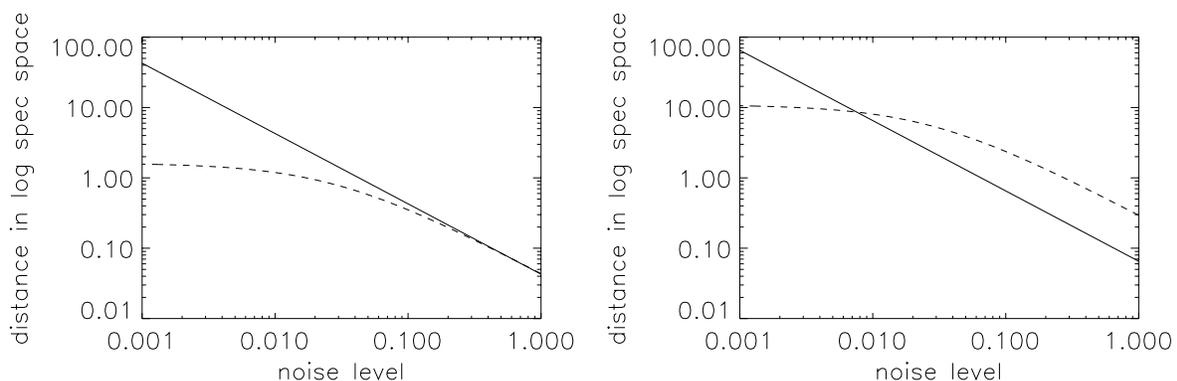


Figure 4.9: The effective distance in dependence of the noise level (SNR), for the original (dashed) and the transformed (solid) spectra.

4.8.2 Performance for a Cluster of Different Reflectance Spectra

We now consider a whole set of different reflectance spectra $\ln \mathbf{r}$ in the logarithmic spectral feature space which each form a particular cluster. Let Σ_b be the between-cluster covariance matrix, and Σ_w the mean within-cluster covariance matrix of all clusters.

As before, we want to measure the overall separability of the clusters in the logarithmic feature space as the ratio of the squared scatter radius of the reflectances $\text{tr } \Sigma_b$ over the mean squared scatter radius within the individual clusters Σ_w . So the separability before (left-hand) and after the transform (right-hand) is

$$\frac{\text{tr } \Sigma_b}{\text{tr } \Sigma_w} \longleftrightarrow \frac{\text{tr } \mathbf{P} \Sigma_b \mathbf{P}}{\text{tr } \mathbf{P} \Sigma_w \mathbf{P}} \quad (4.200)$$

and as before, we model the within-cluster variance Σ_w in the logarithmic feature space as the sum of the variance Σ_R of the reflectance $\ln \mathbf{r}$ and the illumination variance Σ_I : $\Sigma_w = \Sigma_I + \Sigma_R$. Then Eq. 4.200 becomes

$$\frac{\text{tr } \Sigma_b}{\text{tr } (\Sigma_I + \Sigma_R)} \longleftrightarrow \frac{\text{tr } \mathbf{P} \Sigma_b \mathbf{P}}{\text{tr } \mathbf{P} (\Sigma_I + \Sigma_R) \mathbf{P}} \quad (4.201)$$

$$\longleftrightarrow \frac{\text{tr } \mathbf{P} \Sigma_b \mathbf{P}}{\text{tr } \mathbf{P} \Sigma_R \mathbf{P}} \quad (4.202)$$

because the projector matrix filters the illumination variability: $\mathbf{P} \Sigma_I \mathbf{P} = 0$ (Eq. 4.93 on page 63). Then for symmetric and uncorrelated noise $\Sigma_R = \epsilon^2 \mathbf{I}$ we have $\text{tr } \Sigma_R = \epsilon^2 \text{tr } \mathbf{I} = \epsilon^2 N$ and $\text{tr } \mathbf{P} \Sigma_R \mathbf{P} = \epsilon^2 \text{tr } \mathbf{P} = \epsilon^2 (N - K)$ (Eq. 4.137) and arrive at

$$\frac{\text{tr } \Sigma_b}{\text{tr } \Sigma_I + N \epsilon^2} \longleftrightarrow \frac{\text{tr } \mathbf{P} \Sigma_b \mathbf{P}}{(N - K) \epsilon^2} \quad (4.203)$$

As in the previous section, **we can see that the right-hand expression and thus the separability for the transformed case will always become greater than for the untransformed data (left-hand) if only the noise level ϵ becomes sufficiently small: $\epsilon < \epsilon_o$.**

The threshold can be determined to $\epsilon_o = \sqrt{(\text{tr } \Sigma_I \text{tr } \mathbf{P} \Sigma_b \mathbf{P}) / ((N - K) \text{tr } \Sigma_b)}$.

If we moreover assume that also the set of different reflectance spectra \mathbf{r} be distributed spherically in the logarithmic spectral feature space, its covariance matrix becoming $\Sigma_b = R \mathbf{I}$ and $\text{tr } \Sigma_b = R \text{tr } \mathbf{I} = RN$, then the inequation is

$$\frac{RN}{\text{tr } \Sigma_I + N \epsilon^2} \longleftrightarrow \frac{R \text{tr } \mathbf{P}}{(N - K) \epsilon^2} \quad (4.204)$$

$$\frac{R}{(\text{tr } \Sigma_I) / N + \epsilon^2} \longleftrightarrow \frac{R(N - K)}{(N - K) \epsilon^2} \quad (4.205)$$

$$\frac{R}{(\text{tr } \Sigma_I) / N + \epsilon^2} \longleftrightarrow \frac{R}{\epsilon^2} \quad (4.206)$$

and obviously the left-hand expression is always smaller than the right-hand expression:

$$\frac{R}{(\text{tr } \Sigma_I) / N + \epsilon^2} < \frac{R}{\epsilon^2} \quad (4.207)$$

So for this case of spherical distribution of the reflectance spectra in logarithmic spectral space, the separability of the clusters from the different reflectances is better in the transformed feature space than in the original feature space at any noise level ϵ , due to the filtered illumination variability of magnitude $(\text{tr } \Sigma_I) / N$.

4.9 Decision on Reflectance vs. Illumination Change for a Given Pair of Spectra

We consider the case of two given spectra, \mathbf{x} and \mathbf{x}' , for which a decision is sought whether these two spectra are differing necessarily due to a change in the reflecting material or rather due to a mere change in surface orientation, *i.e.*, illumination. The spectra may be provided interactively by an analyst or by any algorithm for *e.g.* classification, change or edge detection purposes.

Ideally, the spectra \mathbf{x} and \mathbf{x}' will be considered to stem possibly from equal reflectances if their spectral distance vanishes after transformation to the invariant descriptors \mathbf{t} and \mathbf{t}' :

$$d^2(\mathbf{t}, \mathbf{t}') = (\mathbf{t} - \mathbf{t}')^T(\mathbf{t} - \mathbf{t}') = (\mathbf{x} - \mathbf{x}')^T \mathbf{P} (\mathbf{x} - \mathbf{x}') \quad (4.208)$$

$$\stackrel{!}{=} 0 \quad .$$

For real data, however, the spectral distance will seldomly yield exactly $d^2(\mathbf{t}, \mathbf{t}') = 0$, due to measurement and quantization noise. So the possibility of the spectra stemming from an equal reflectance spectrum will be sustained if the transformed spectral distance is small, *i.e.*, below a certain threshold. Such a threshold will certainly depend on certain image properties such as the noise level, systematic calibration errors etc. If we consider an example with equal multiplicative noise variance ϵ^2 for all bands i , and *e.g.* a transformed spectrum $\mathbf{t} \approx \mathbf{t}' \approx \mathbf{u}$, then we the expected standard deviation in the distance $d^2(\mathbf{t}, \mathbf{t}')$ between the transformed spectra can be evaluated to

$$\sigma^2(d^2) = \sum_i \sigma_i^2(t_i - t'_i) \quad (4.209)$$

$$= \sum_i \sigma_i^2(t_i) + \sigma_i^2(t'_i) \quad ; \quad \text{using Eq. 4.160 on page 81} \quad (4.210)$$

$$= \sum_i (t_i^2 + t_i'^2) \sum_j P_{ij}^2 \epsilon^2 \quad ; \quad \text{let } t_i \approx t_i' \approx 1 \quad (4.211)$$

$$= 2\epsilon^2 \sum_i \sum_j P_{ij}^2 \quad ; \quad \text{using Eq. 4.137 on page 68} \quad (4.212)$$

$$= 2\epsilon^2 \text{tr } \mathbf{P} \quad (4.213)$$

$$= 2\epsilon^2(N - K) \quad (4.214)$$

For making a decision on given pair of spectra \mathbf{x} and \mathbf{x}' , the absolute transformed distance $d^2(\mathbf{t}, \mathbf{t}')$ must be compared to its standard deviation $\sigma(d^2)$. A proper absolute threshold value such as *e.g.* 1σ , 2σ or 3σ must be derived from the image context. When the transformed distance $d^2(\mathbf{t}, \mathbf{t}')$ is below this threshold, then we will consider the spectra as equal in the sense that they could stem from two surfaces of equal reflectance but differing surface orientation.

4.9.1 Error Weighted Spectral Distance

The spectral distance used for the distance ratio in Eq. 4.217 on the next page includes all spectral bands on equal footing. With a given error caused covariance matrix $\mathbf{S}(\mathbf{x})$,

however, the contributions of the bandwise differences $d_i = x_i - x'_i$ should contribute to the overall spectral distance with weights according to their estimated error. After having established the error propagation and the proper covariance error matrices $\mathbf{S}(\mathbf{x}) \mapsto \mathbf{S}(\mathbf{t})$ (Section 4.6 on page 80), we now employ the estimated errors for a weighting scheme of the spectral distance.

$$d_w^2(\mathbf{x}, \mathbf{x}') = \frac{(\mathbf{x} - \mathbf{x}')^T [\mathbf{S}(\mathbf{x}) + \mathbf{S}(\mathbf{x}')]^{-1} (\mathbf{x} - \mathbf{x}')}{\text{tr}([\mathbf{S}(\mathbf{x}) + \mathbf{S}(\mathbf{x}')]^{-1})/N} \quad (4.215)$$

$$d_w^2(\mathbf{t}, \mathbf{t}') = \frac{(\mathbf{t} - \mathbf{t}')^T [\mathbf{S}(\mathbf{t}) + \mathbf{S}(\mathbf{t}')]^{-1} (\mathbf{t} - \mathbf{t}')}{\text{tr}([\mathbf{S}(\mathbf{t}) + \mathbf{S}(\mathbf{t}')]^{-1})/N} \quad (4.216)$$

The numerator in this expression weights the bandwise differences $d_i = x_i - x'_i$ according to their variances and covariances, while the denominator normalizes the sum of all weights. Note that the trace in the denominator is divided by N in order to yield $\text{tr} \mathbf{I}/N = 1$ for the case of a spherical error matrix $\mathbf{S}(\mathbf{t}) = \mathbf{S}(\mathbf{t}') = \epsilon^2 \mathbf{I}$, and thus correspondence of the weighted and unweighted distance $d_w^2 = d^2$ for this specific case.

4.9.2 Results on Simulated Data

As stated above, the distance $d^2(\mathbf{t}, \mathbf{t}')$ between the transformed spectra will depend on the image specific noise level. Therefore, for the purpose of general analysis in the framework of this thesis, we rather consider the ratio of the transformed over the original spectral distance:

$$\frac{d^2(\mathbf{t}, \mathbf{t}')}{d^2(\mathbf{x}, \mathbf{x}')} = \frac{(\mathbf{t} - \mathbf{t}')^T (\mathbf{t} - \mathbf{t}')}{(\mathbf{x} - \mathbf{x}')^T (\mathbf{x} - \mathbf{x}')} \quad (4.217)$$

Here $d^2(\mathbf{x}, \mathbf{x}') = (\mathbf{x} - \mathbf{x}')^T (\mathbf{x} - \mathbf{x}')$ is the original spectral distance which possibly exists merely because of different surface orientation. So for the following simulations and later evaluation of real spectral image data (Section 6.8.3 on page 165), instead of the absolute distances we rather give the relative reduction of the distances achieved by the transformation to the invariant descriptor. Due to the above shown transformation properties (Eq. 4.98 on page 64) this ratio will be in the range $[0..1]$.

We use the same distance ratio also for the investigation of the performance of the error weighted distance d_w^2 :

$$\frac{d_w^2(\mathbf{t}, \mathbf{t}')}{d_w^2(\mathbf{x}, \mathbf{x}')} \quad (4.218)$$

For evaluation of the weighted versus the unweighted spectral distance we have simulated 1000 spectral pairs. The first spectrum \mathbf{x} of each spectral pair was taken at random out of $0 < x_i \in \mathbb{R} < 255$ at the center wavelength λ_i of the sensor specific spectral bands of the DAEDALUS AADS 1268 sensor (described in Section 6.1 on page 116). For simulation of the case of ‘equal reflectance’, the second spectrum of each pair \mathbf{x}' was computed as the first spectrum in shadow illumination $x'_i = x_i \cdot \hat{m}_i$ (see values in Table 6.6 on page 158). Only then the spectra \mathbf{x} and \mathbf{x}' were converted to integers, accounting for the 8 bit quantization range $[0..255] \in \mathbb{N}$ (integer numbers).

In contrast, for simulation of the case of ‘random reflectance’ also the second spectrum of each pair \mathbf{x}' was chosen at random before quantization of both. No further artificial noise was superimposed. The error covariance matrices \mathbf{S} were computed assuming an additive random error of $\epsilon = 1$ using the formulae of Section 4.6 on page 80.

Then the transformed to original spectral distance ratio was computed. The simulation was evaluated for three different mappings. Results are listed in Table 4.2 on the following page. The mean distance ratios are given in lines (1) to (4). For ‘equal reflectances’ the distance ratio should be low. The ratio of lines (1)/(2) shows that the error weighted spectral distance ratio performs two orders of magnitude better for the mapping $\mathbf{x} \mapsto \mathbf{t}$, whereas the average performance is equal for the pure linear mapping $\ln \mathbf{x} \mapsto \ln \mathbf{t}$. This indicates that the tracing of the error propagation is most necessary for the non-linear logarithm and exponentiation mappings.

Comparing lines (1) and (3), *i.e.*, the unweighted distance ratio for ‘equal’ and ‘random’ reflectances, shows that they are clearly distinguishable. The ratio (1)/(3) is one to two orders of magnitude lower for ‘equal’ than for ‘random’ reflectance, and a proper threshold can be set accordingly. The same holds for comparison of lines (2) and (4), *i.e.*, the weighted distance ratio for ‘equal’ and ‘random’ reflectances. Here the ratio for pairs of equal reflectance is 3000 ($= 0.00034^{-1}$) times smaller than for pairs of random reflectance.

However, the lowest entry among the (1)/(3) ratios is in the logarithmic domain $\ln \mathbf{x} \mapsto \ln \mathbf{t}$, whereas the lowest entry among the (2)/(4) ratios is in the original domain $\mathbf{x} \mapsto \mathbf{t}$. This means that when employing the unweighted distance, the transformed to original spectral distance ratio is most useful in the logarithmic domain, while the use of the error weighted distance gives the most valuable ratio in the original domain. Moreover, even with quantization as the only noise source, the error weighted distance ratio performs two orders of magnitude better in distinction between spectral pairs of random and equal reflectance than the unweighted spectral distance ratio.

Hence we can summarize that the error weighted spectral distance should be used whenever computation time allows this, and that the transformed to original spectral distance ratio can be thresholded in order to decide whether two spectra are possibly related by an equal reflectance under differing illumination.

ERROR WEIGHTED SPECTRAL DISTANCE RATIOS FOR EQUAL REFLECTANCE PAIRS			
ratio $\frac{\text{transformed}}{\text{original}}$ distance	$\mathbf{x} \mapsto \mathbf{t}$	$\mathbf{x} \mapsto \ln \mathbf{t}$	$\ln \mathbf{x} \mapsto \ln \mathbf{t}$
(1) [mean \pm dev] $\cdot 10^{-6}$ unweighted	3.161 ± 0.1	1.779 ± 0.04	5370.3 ± 76.7
(2) weighted	0.009 ± 0.0002	0.81 ± 0.02	4226.7 ± 66.9
ratio (1)/(2) unweighted/weighted	344	22	1.3

ERROR WEIGHTED SPECTRAL DISTANCE RATIOS FOR RANDOM REFLECTANCE PAIRS			
ratio $\frac{\text{transformed}}{\text{original}}$ distance	$\mathbf{x} \mapsto \mathbf{t}$	$\mathbf{x} \mapsto \ln \mathbf{t}$	$\ln \mathbf{x} \mapsto \ln \mathbf{t}$
(3) [mean \pm dev] $\cdot 10^{-6}$ unweighted	127.1 ± 2.0	120.0 ± 1.5	770066 ± 5644
(4) weighted	26.88 ± 0.5	60.75 ± 0.8	1289498 ± 10971
ratio (3)/(4) un-/weighted	4.7	2.0	0.6

COMPARISON: EQUAL VS RANDOM REFLECTANCE PAIRS			
ratio (1)/(3) equal/random reflectance unweighted	0.025	0.015	0.007
ratio (2)/(4) equal/random reflectance weighted	0.00034	0.0133	0.0032

Table 4.2: Mean unweighted and error weighted spectral distance for 1000 simulated spectral pairs at the spectral bands of the DAEDALUS AADS 1268. For ‘equal reflectance’, the first spectrum is chosen at random and the second multiplied by the relative diffuse illumination $\hat{\mathbf{m}}$, whereas for ‘random reflectance’ the second spectrum is chosen at random as well. No artificial noise is superimposed, however, the spectra are quantized into 8 bit.

4.10 Relaxation of the Lambertian Assumption

In the foregoing considerations we have assumed the reflecting surface to be Lambertian (see Section 2.1 on page 13). After having introduced the invariant transformation $\mathbf{x} \mapsto \mathbf{t}$, we want to revisit the reflection assumption and show that we can actually relax this assumption to a certain degree.

The Lambert reflection term is

$$L^r = \frac{1}{\pi} r E^\downarrow \cos \theta \quad (\text{Eq. 2.5 on page 14}) \quad (4.219)$$

where L^r is the reflected radiance, E^\downarrow the incident irradiance, r the Lambertian reflectance, and θ the observation angle. By factorization, we can express the Lambert equation as a wavelength dependent function $R(\lambda) = \frac{1}{\pi} r_\lambda$ and a geometry function $G(\theta) = \cos \theta$

$$L^r(\lambda, \theta) = R(\lambda) \cdot G(\theta) \cdot E^\downarrow \quad (4.220)$$

Now the geometry function G does not necessarily need to be the cosine law $G(\theta) = \cos \theta$. We can rather allow for $G(\theta, \phi)$ to be any function of the angles (θ, ϕ) as long as it is not wavelength dependent.

$$L^r(\lambda, \theta, \phi) = R(\lambda) \cdot G(\theta, \phi) \cdot E^\downarrow \quad (4.221)$$

In other words, the geometry dependence of the reflection can be arbitrary but will only scale the overall spectrum by a common factor, *i.e.*, increase the brightness but not change the color. Then the suggested mapping $\mathbf{x} \mapsto \mathbf{t}$ will be invariant, since overall scaling factors are filtered by the first component of the projector $\mathbf{P} = \mathbf{U}\mathbf{V}$ ($\mathbf{U} = \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^T$, Eq. 4.32 on page 53). In fact, all color constancy algorithms (see Section 1.3 on page 7) agree in that an overall brightness factor, *i.e.*, the absolute height of a reflectance spectrum, cannot be recovered.

This requirement for the reflection is weaker than the Lambertian law. Some researchers even claim that the above factorization is an accurate description of the diffuse reflection⁸ in the visible range (Healey 1986), and some color constancy approaches in computer vision explicitly build on that (Matas et al. 1994).

However, from a physics based point of view this seems to be a simplification rather than a physical law. Reflection properties are influenced by surface roughness (Torrance & Sparrow 1967). The degree of roughness, however, will necessary be measured in relation to the wavelength of the incident light. Therefore it is unlikely that the BRDF properties of a surface should be independent of wavelength in general. This is confirmed by experimental findings of Meister (1995).

Nevertheless, this reflection assumption is considerably weaker than the Lambertian law and will thus be met better by real world materials. Hence the applicability of the invariant as introduced here is high.

⁸ a specular component is often added

Experimental Findings on *in situ* Measured Spectral Data

Outline of this chapter – *Here we confirm the previously developed dichromatic illumination model and its analysis in the logarithmic space by real spectral data. We analyze in situ measured spectra from selected surfaces under natural outdoor illumination. The variability of the measured radiance spectra from a specific surface is interpreted as the effect of the varying surface orientation. We identify the expected brightness and color shift in the principal components of this variability. At first we show some examples in the wavelength representation and in the feature space representation. Then we apply three different methods of quantitative evaluation to all the datasets: Principal component analysis in the logarithmic feature space, logarithmic filtering of the variance in order to assess the achieved invariance, and assessing the overall magnitude of the impact of surface orientation on the spectral signature. Analysis of spectral data sets from several surfaces shows that the variability is indeed independent of the specific target surface reflectance in the logarithmic space.*

The goodness of the desired invariance is evaluated by virtue of the reduction of variance which is achieved by the suggested filtering. The reduction of variance is investigated with respect to the number of filtered components K , and with respect to the estimated atmospheric parameter γ . It is shown that the filtering of two components is sufficient and can reduce the variance by $\approx 99.8\%$. Moreover, for the given wavelength range the invariant descriptor is nearly independent of the choice of the atmospheric exponential parameter γ .

Finally, the magnitudes of brightness and color shifts as caused by arbitrary change in surface orientation are assessed, with the result that the color shift is about an order of magnitude smaller than the brightness changes and in the 5% range in the measurement scale. There is no systematic correlation between brightness and color shift.

5.1 The Sensor

The instrument OVID (Optical Visible and Infrared Detector) is a fast recording, spectrally high resolving multichannel (non-imaging) spectrometer. It was designed and is

operated by Meteorologisches Institut der Universität Hamburg and Max Planck Institut für Meteorologie Hamburg for ground based as well as airborne measurements such as described by Bartsch et al. (1994). A thorough description of the sensor is given by Bartsch (1996).

For our experiment we used OVID's VIS module (visible wavelength range) which is based on a CCD array of 1024×256 pixels, of which the 1024 pixels are used to cover a wavelength range of 550 – 1100 nm (0.5686 nm per pixel), and the 256 pixels to integrate the beam incident from the slit via the grating. With a grating of 300 lines/mm the spectral resolution is $\Delta\lambda = 1.7 \text{ nm} \hat{=} 3 \text{ pixels}$. The spectrometer is fed via a fiber optic by a telescope with an 8 cm spherical mirror, yielding a spatial resolution of $0.27^\circ \hat{=} 4.7 \text{ mrad}$ (aperture angle). Typical integration times for a spectrum are 5 – 100 ms, with a maximum recording rate of 25 Hz. The intensities are recorded with a dynamic range of 14 bit, corresponding to an intensity resolution of 1/16384.

The wavelength calibration is performed before and after recordings using an Argon light source and accurate to ca. 4 pixels $\hat{=} 2.3 \text{ nm}$. Radiance calibration is also done before and after recordings by means of an Ulbricht integrating sphere, the window of which offers an approximately Lambertian radiating area. Absolute radiances are recovered by cross-calibration with a standard light source to $\pm 5\%$.

5.2 Experimental Setup of the *in situ* Measurements

We have recorded spectra of selected surfaces under arbitrary orientations (see Fig. 5.1 on the facing page). Vegetation canopies exhibit quite complex reflection properties and are a typical object of BRDF-measurements. This thesis was embedded in the context of image processing with respect to the Open Skies Treaty (Wiemker & Spitzer 1996). Therefore targets were chosen from man-made surfaces. Note that we were not interested to learn about specific properties of the selected surfaces. Rather, our interest here is to analyze the spectral effects of varying surface orientation which are common to all surfaces (Lambertian reflection assumed). So the emphasis in selecting the surface samples was to arrive at a wide range – rather than a representative range – of artificial outdoor materials. The target surfaces and the number of recorded spectra can be seen in Table 5.2 on page 107.

The recordings took place outdoors under natural illumination of sun and skylight, at different times of day during several summer days in 1994, some of which were clear and some slightly hazy or cloudy. The position and attitude of the spectrometer telescope was varying between recording days, but fixed within each recording session. Each single data set was recorded with one of the surface samples. The orientation of this target was changed slowly but constantly, while the OVID spectrometer continuously recorded spectra at a rate of 25 Hz. The orientation changes were conducted at random but in a fashion as to cover a wide range of possible illumination and observation angles. The described recording scheme was chosen in order to acquire as many spectra as possible and thus to build a broad statistical basis. The angles under which the single spectra were measured, were not determined, since in this thesis we are not concerned with angular dependent BRDF effects, but rather with the overall spectral variability as caused by arbitrary changes in surface orientation. Therefore it was of prominent importance to measure as many spectra as possible, and the specific recording geometry

had to be neglected.

The target surfaces were all larger than $15 \times 15 \text{ cm}^2$ and positioned in $0.5 - 1.5 \text{ m}$ distance from the spectrometer telescope. We chose short integration times of mostly 5 ms and can thus be certain that the recording time for each spectrum is shorter than the typical time scale of atmospheric variability of $\approx 0.1 \text{ s}$.

The recorded spectra were dark current corrected, wavelength and radiance calibrated, and cleaned from overflows. In order to avoid low SNR regions at the margins of the wavelength range we used only the range of $613 - 883 \text{ nm}$. Before further computation the spectra were subsampled from then $N = 475$ down to $N = 200$ spectral bands.

A set of spectra was recorded as described above from a near-Lambertian reflectance reference panel (SPECTRALON of Labsphere Inc., Boulder, Colorado). The properties of this particular panel were measured and described in Meister (1995) and Meister et al. (1996). The recorded spectra are plotted in Fig. 5.2 on page 99. Note that the reflectance of the reference panel is wavelength independent, $r_\lambda = 0.50 \pm 0.005$, and that all spectral features stem from the solar spectrum (in particular the Fraunhofer lines at $\lambda = 656.3 \text{ nm}$ (H) and $\lambda = 849.8 \text{ nm}$, 854.2 nm , 866.2 nm (Ca II)), and the atmospheric absorption bands (in particular the O_2 A-band centered at $\lambda = 760 \text{ nm}$ and the O_2 B-band centered at $\lambda = 690 \text{ nm}$).

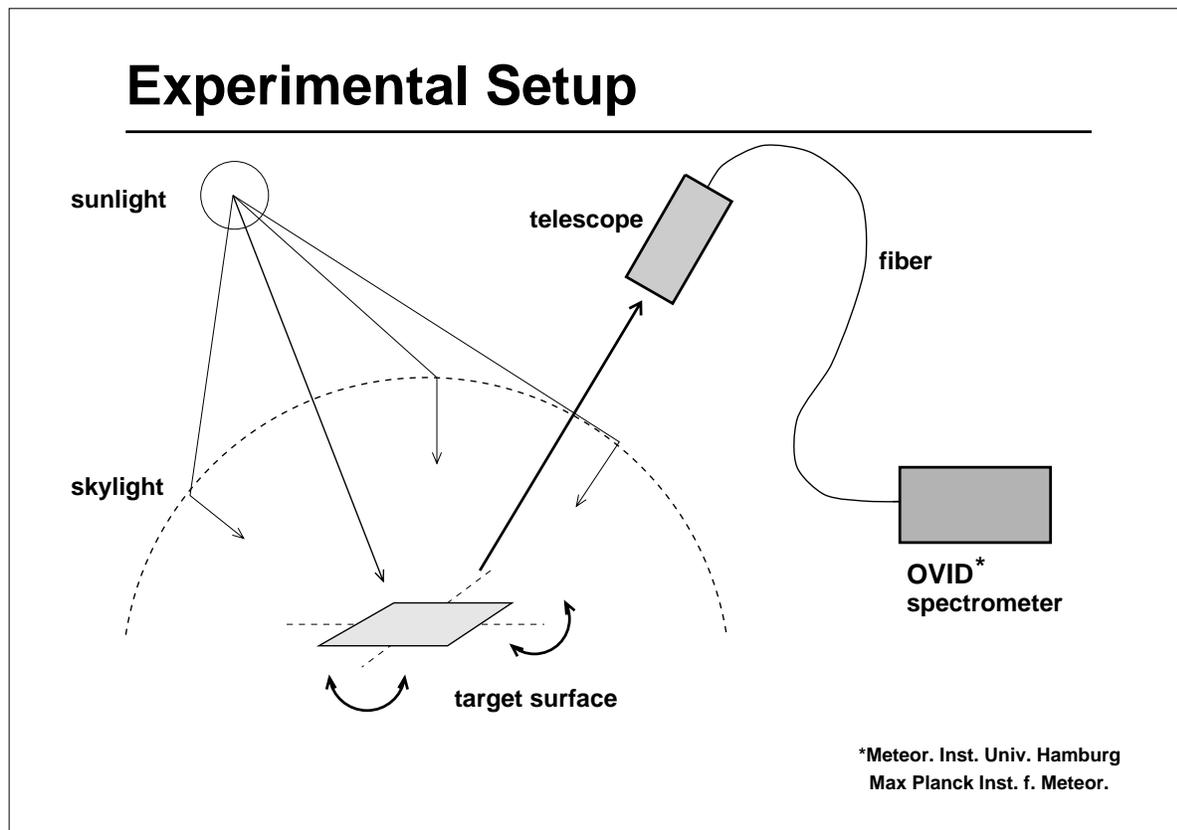


Figure 5.1: The spectra from selected target surfaces, observed under arbitrary surface orientation angles. The target surface is tilted around both axes while the spectrometer telescope remains fix.

5.3 Some Examples in the Wavelength- and Feature-Space-Representation

Before we start with a comprehensive evaluation of all experimentally measured data sets, we choose three exemplary sets in order to illustrate the impact of surface orientation on the spectral signature. In Fig. 5.2 on the next page (top) we show a number of spectra observed under varying surface orientation angles from a near-Lambertian reference surface. All measured spectra are plotted in one graph. Note again that the annotated absorption band features stem from the illuminating sun and sky light, whereas the reflectance of the Lambertian reference surface is wavelength independent $r = 0.5$, in other words, featureless ‘flat white’. By virtue of this property, via measuring the reflected spectra we actually observe the illumination spectra but up to an overall scaling factor. Due to the approximately Lambertian reflexion of the reference panel, we can neglect BRDF related effects, but rather observe pure illumination changes.

In order to illustrate the invariance achieved by the suggested logarithmic brightness and color filtering, all spectra are again plotted into one graph after brightness and color filtering respectively (Fig. 5.2). The spectra are brightness filtered as

$$\mathbf{x} \mapsto \exp(\mathbf{U} \ln \mathbf{x})$$

and brightness and color filtered as

$$\mathbf{x} \mapsto \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$$

where the projectors \mathbf{U} and \mathbf{V} are computed as specified in Eq. 4.32 on page 53 and Eq. 4.103 on page 65.

We observe that after brightness filtering (bottom left) with \mathbf{U} (Eq. 4.32) there is clearly a remaining systematic shift which increases the radiances in the shorter wavelengths and decreases the ones in the longer wavelengths, and vice versa. We like to call this shift the *color shift*. Its nature was analyzed in Chapter 4.

We further observe, that color filtering (bottom right) with a projector \mathbf{V} (Eq. 4.103) can remove effectively the remaining variance and produces orientation invariant spectral descriptors. The filter \mathbf{V} deployed here reduces the rank of the data by $K = 1$ only, and was computed from an assumed atmospheric parameter $\gamma = 1$.

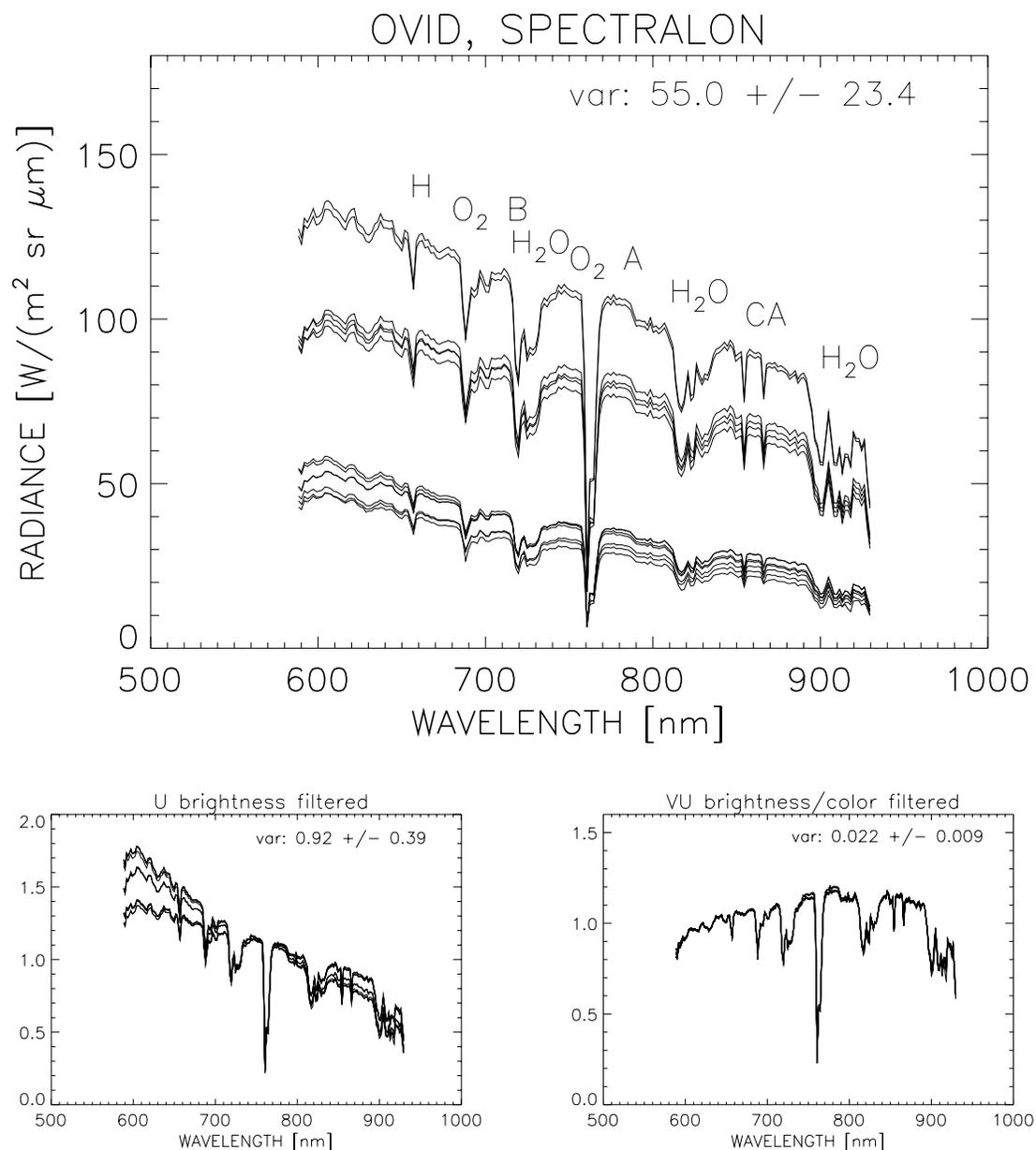


Figure 5.2: Spectra observed under varying surface orientation angles from a near-Lambertian reference surface (top).

bottom left: after logarithmic brightness filtering by projector \mathbf{U} ,

bottom right: after logarithmic brightness and color filtering by projector $\mathbf{P} = \mathbf{VU}$ with atmospheric parameter $\gamma = 1$.

The total variance with error margin is given in the top right corner of the plots.

■ Reduction of Within-Cluster Variability:

Since all of the spectra in Fig. 5.2 are reflected from the same surface, they can be considered as points in the feature space forming a single cluster denoted by a . We consider the variability within the cluster to be caused entirely by the varying surface orientation and the resulting changes in illumination. The variability or scatter within the cluster is given by $J_a^2 = \text{tr } \Sigma_a$ (Eq. 2.35 on page 25). (In this example we deal with one cluster only; thus we will conveniently drop the subscript a .) We prefer to determine the covariance matrix Σ from the logarithmic spectra $\ln \mathbf{x}$ rather than from the originally measured spectra \mathbf{x} . The advantage of the logarithmic space is its independence against scaling:¹ Let all observed spectral radiances x_i in spectral band i be multiplied by a constant factor c , $x_i \mapsto c x_i$, *e.g.* to account for an overall brighter solar illumination. Then the variance also changes with the factor c :

$$\text{var}\langle c x_i \rangle = c \text{var}\langle x_i \rangle \quad . \quad (5.1)$$

In contrast, the variance of the logarithmic spectra remains untouched by the scaling:

$$\text{var}\langle \ln(c x_i) \rangle = \text{var}\langle \ln c + \ln x_i \rangle \quad (5.2)$$

$$= \text{var}\langle \ln c \rangle + \text{var}\langle \ln x_i \rangle \quad (5.3)$$

$$= 0 + \text{var}\langle \ln x_i \rangle \quad (5.4)$$

$$= \text{var}\langle \ln x_i \rangle \quad . \quad (5.5)$$

The variances σ_{ii}^2 are the diagonal elements of the covariance matrix Σ and are estimated as described in Eq. 2.37 on page 25. The error which is inherent in these estimates of the variances σ_{ii}^2 depends on the number of samples n , *i.e.*, the number of spectra in the recording set. The variance of the variance σ_{ii}^2 as estimated from n samples is (Brandt 1992):

$$\text{var}\langle \sigma_{ii}^2 \rangle = \frac{2\sigma^4}{n-1} \quad . \quad (5.6)$$

Then the variance values are given with the error as the square root of the variance:

$$\sigma_{ii}^2 \pm \sigma_{ii}^2 \sqrt{\frac{2}{n-1}} \quad . \quad (5.7)$$

Thus also the trace $\text{tr } \Sigma = \sum_i \sigma_{ii}^2$ has this error:

$$\text{tr } \Sigma \pm \text{tr } \Sigma \sqrt{\frac{2}{n-1}} \quad . \quad (5.8)$$

¹ There is another advantage of using the logarithmic spectra: For the measured radiances x_i we certainly cannot assume a Gaussian probability distribution, since the values of x_i are positive definite $x_i \geq 0$, whereas the logarithmic values $\ln x_i$ are unbounded. Hence the *log normal distribution* may be a more appropriate approximation to the true – unknown – probability density function.

For more accuracy, the positive and negative deviation can be calculated separately at a given confidence level. The error interval is then given as

$$\left[\frac{n-1}{c_2}, \frac{n-1}{c_1} \right] \sigma_{ii}^2 \quad (5.9)$$

where the values of c_1 and c_2 are taken from integral tables of the χ^2 distribution function (Bronstein & Semendjajew 1991). For a confidence level of 99% and $n = 12$ this yields

$$[0.33, 3.1] \text{tr } \Sigma \quad . \quad (5.10)$$

Now we can quantify the reduction of the within-cluster scatter which has been achieved by the transforms. The squared scatter radius $J^2 = \text{tr } \Sigma$ is computed for the three sets plotted in Fig. 5.2:

$$\text{tr } \Sigma(\ln \mathbf{x}) = 55.0 \pm 23.4 \quad [18.4, 168.1] \quad (5.11)$$

$$\text{tr } \Sigma(\mathbf{U} \ln \mathbf{x}) = 0.92 \pm 0.39 \quad [0.31, 2.8] \quad (5.12)$$

$$\text{tr } \Sigma(\mathbf{V}\mathbf{U} \ln \mathbf{x}) = 0.022 \pm 0.009 \quad [0.007, 0.067] \quad (5.13)$$

Here we give the error of the estimated variance as stated in Eq. 5.8 and the error interval of Eq. 5.10.

The errors in the variance estimation are quite large in this example as only $n = 12$ spectra were used. Nevertheless, **the reduction in overall variance $\text{tr } \Sigma$ is clearly significant at the 99% confidence level. This indicates that the within-cluster scatter has been reduced and hence the cluster's aptitude for classification is improved. In particular, the color shift filtering by \mathbf{V} has reduced the overall variance by a factor of ≈ 40 compared to brightness filtering with \mathbf{U} only.**

■ Principal Component Analysis:

For the spectra of the near-Lambertian reference surface (SPECTRALON, see the spectra of Fig. 5.2) we carry out a principal component analysis (see Section 2.7 on page 32) in the logarithmic space, since in the log space the variance is not affected by the reflectance spectrum (as set out in Chapter 4). The two most significant eigenvectors of the log covariance matrix $\Sigma(\ln \mathbf{x})$ are plotted in Fig. 5.3 on the next page (top row).

We now will show that the first two eigenvectors can be explained by a wavelength independent brightness shift vector $\hat{\mathbf{u}}$, and a monotonous wavelength dependent 'color' shift vector $\hat{\mathbf{v}}$ (Eq. 4.28 on page 52 and Eq. 4.101 on page 65). To this aim we filter the covariance matrix with the brightness vector $\hat{\mathbf{u}}$ (see Fig. 5.3, bottom left). Then the resulting covariance $\mathbf{U}\Sigma\mathbf{U}^T$ is again submitted to a principal component analysis, and the most significant eigenvector is plotted in Fig. 5.3 (bottom right). Overlaid is the color vector $\hat{\mathbf{v}}$ as computed with an atmospheric exponent $\gamma = 1.0$. We observe that the measured and the expected color shift vectors are in very good agreement. There are slight distortions by the absorption features, particularly at $\lambda = 760$ nm due to the atmospheric variability in the O_2 A-band.

The first two eigenvalues of the covariance matrix $\Sigma(\ln \mathbf{x})$ in the logarithmic space are e_1^2 and e_2^2 and denote the variances of the first two eigenvectors. The total variance is $\text{tr} \Sigma = \sum_{i=1}^N e_i^2$. The relative variances of the first two eigenvectors, $e_1^2 / \text{tr} \Sigma$ and $e_2^2 / \text{tr} \Sigma$, are annotated in the plots of Fig. 5.3. **Together they make up for 99.97% of the total variance $\text{tr} \Sigma$. Hence we are confident that essentially the complete variance of the recording set can be represented by the postulated brightness and color shift vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$.**

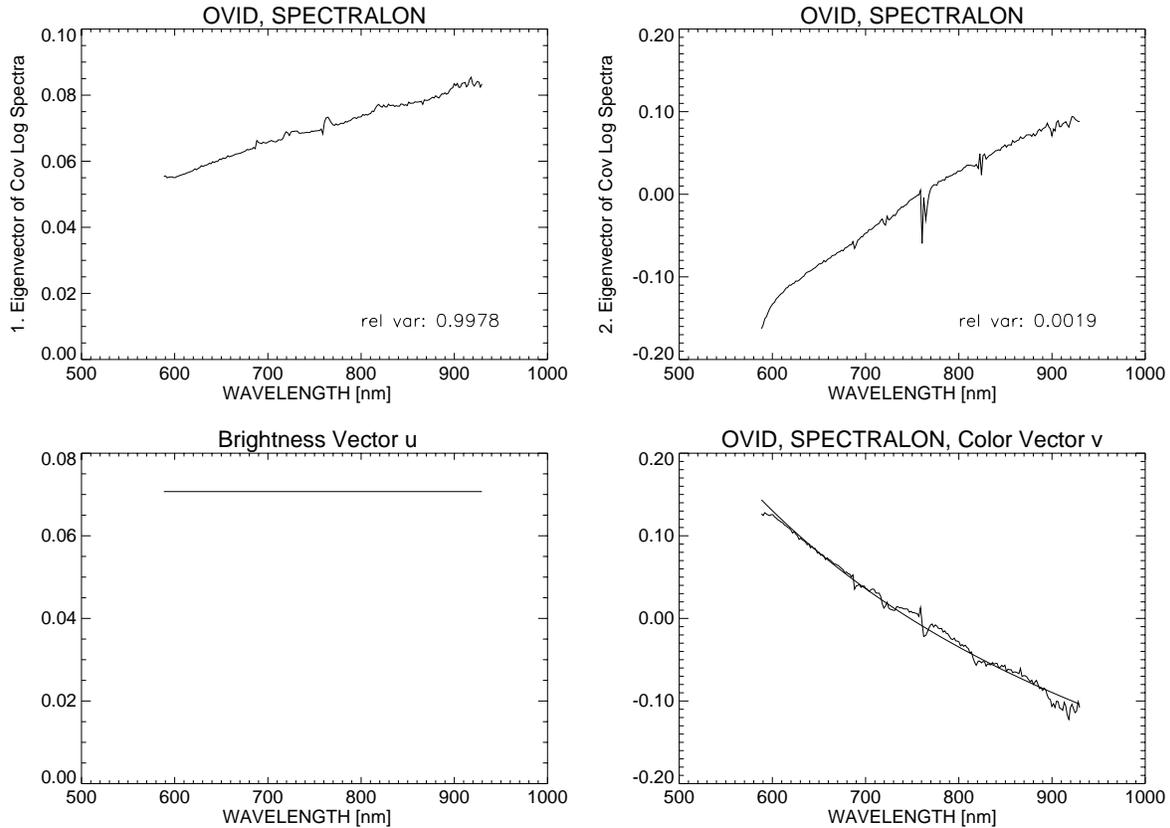


Figure 5.3: *Illumination eigenvectors (top row): The two most significant variance eigenvectors of the logarithmic spectra observed under arbitrary angles of the Lambertian reference target.*

Bottom row, left: The wavelength independent brightness vector $\hat{\mathbf{u}}$.

Bottom row, right: Plotted is firstly the most significant eigenvector of $\mathbf{U}\Sigma\mathbf{U}^T$, i.e., after filtering of the component $\hat{\mathbf{u}}$, and secondly, for comparison, the color shift vector $\hat{\mathbf{v}}$ as simulated with an atmospheric exponent $\gamma = 1.0$ (smooth curve).

■ Illustration in the Feature Space:

We finally want to illustrate the effects of surface orientation and the suggested filtering in the feature space, where the measured spectra are represented by scattered points. As there is no way of visualizing an N -dimensional feature space, we have to restrict ourselves to a two-dimensional projection of the feature space. Principal component analysis effectively tells us which components preserve the maximum variance present in the feature space, and thereby, which projection gives the most expressive view of the

feature space.

We selected recording sets of two different surface materials: a cork tile and a white ceramic tile, with 355 and 335 measured spectra respectively. The spectra were sub-sampled to $N = 95$ spectral bands.

In Fig. 5.4 on the following page (top left) we show the scatter plot of the experimentally measured original spectra from the two data sets. The elongated cluster shapes are caused by the strong brightness variations caused by the changing surface orientations. The feature vectors of the two materials are clearly distinguishable, however, not so by minimal Euclidean distance classification which can only discriminate spherical clusters. Here we would need to employ a metric induced distance measure which is computationally far more expensive and thus seldom used for unsupervised clustering methods.

The result of the brightness filtering in the logarithmic feature space is shown in Fig. 5.4 in the subfigure bottom left. We observe that separability of the two data sets has improved, but that even after logarithmic brightness filtering there is a systematic variance remaining.

The subfigure bottom right shows the result after additional logarithmic color filtering. On the one hand, the distance between cluster centers has decreased. On the other hand, the within-class scatter of the two sets is nearly spherical, *i.e.*, more similar to uncorrelated noise. So the overall separability for Minimum Euclidean distance classifiers is improved, as we will see below. We want to quantify the separability of the two clusters by virtue of the separability measures $d_S^2 = \text{tr } \Sigma_b / \text{tr } \Sigma_w$ and $d_S^{2*} = \text{tr}(\Sigma_b \Sigma_w^{-1})$ (Eq. 2.44 and Eq. 2.45 on page 26). The results for the original feature space \mathbf{x} , the brightness shift filtered space $\exp(\mathbf{U} \ln \mathbf{x})$, and the brightness and color shift filtered space $\exp(\mathbf{VU} \ln \mathbf{x})$ are given in Table 5.1 on the following page. Our example is apt to demonstrate some of the delicate pitfalls of cluster separability measures. The results in Table 5.1 indicate that for both measures the separability of the two clusters is increased by the brightness shift filtering. However, also for both measures, the separability decreases after the additional color shift filtering is applied. We note also that even after the color shift filtering the separability is still significantly better than between the cluster in the original space.

Finally, unsupervised classification of the spectra into $k = 2$ classes is performed, using the fuzzy k -means clustering algorithm (Section 2.6 on page 29). We set the number of classes to $k = 2$, but do not provide any ‘training’ spectra for the classes, nor cluster centers etc. The resulting classification accuracy is determined by comparing the unsupervised classification assignment to the true origin of each spectrum. The relative number of correctly labeled pixels is given in percent in the last row of Table 5.1. Only with the additional color shift filtering perfect classification is achieved. So the unsupervised classification works better after color shift filtering, even though both cluster separability measures are more promising for brightness shift filtering only. The reason becomes clear by appraisal of the feature space projections depicted in Fig. 5.4: Only in the color shift filtered space (bottom right) it is possible to draw a linear decision surface between the two classes which is perpendicular to the difference vector between the two cluster centers. This kind of decision surface is just the one implicitly used by Minimum Euclidean Distance classification (Section 2.3 on page 23).

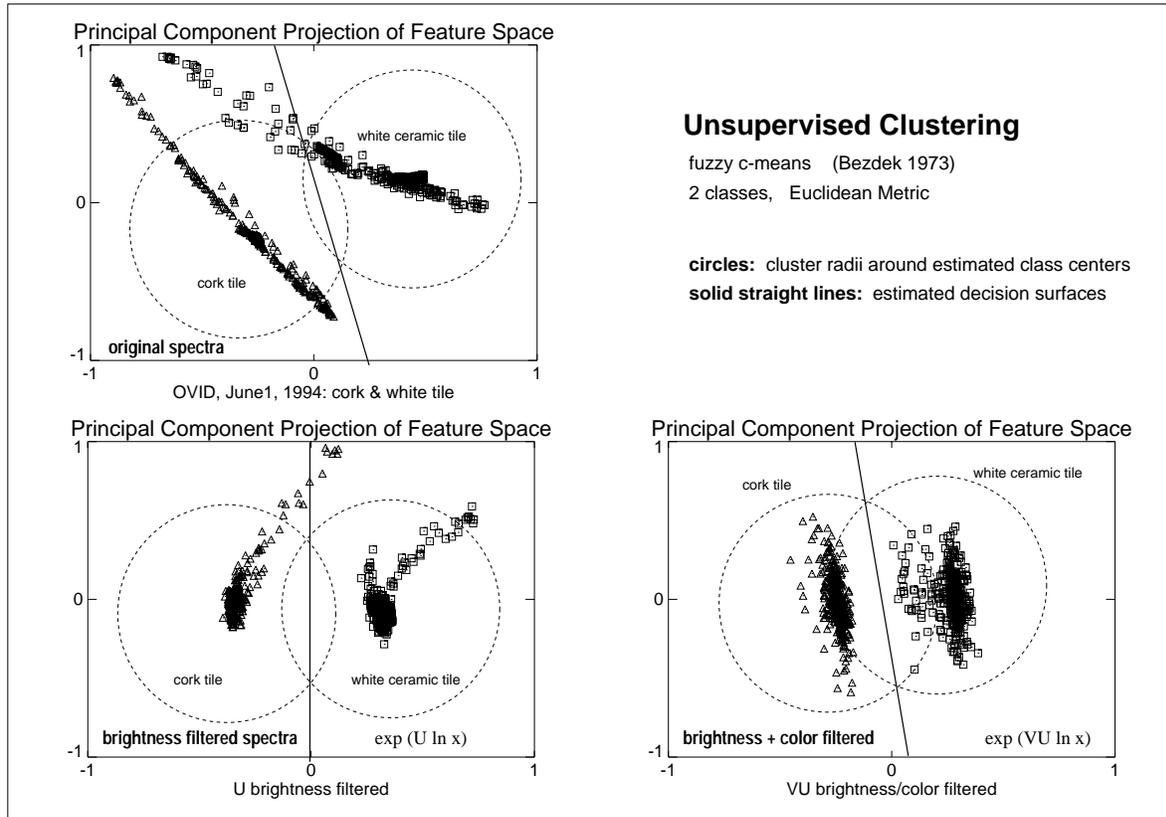


Figure 5.4: For eye appraisal the $N = 200$ dimensional feature space was projected down onto the subspace of the two most significant eigenvectors (individual PCT for each case).

top left: the 690 measured spectra (355 of cork tile Δ , 335 of white ceramic tile \square),

bottom left: after logarithmic brightness filtering by projector U ,

bottom right: after logarithmic brightness and color filtering by projector $P = VU$ with atmospheric parameter $\gamma = 0.5$.

CLUSTER SEPARABILITY IN DIFFERENT FEATURE SPACES			
Space	\mathbf{x}	$e^{U \ln \mathbf{x}}$	$e^{VU \ln \mathbf{x}}$
$\sqrt{\text{tr } \Sigma_b / \text{tr } \Sigma_w}$	27.7	104.9	55.3
$\sqrt{\text{tr } (\Sigma_b \Sigma_w^{-1})}$	205.4	498.0	295.8
unsupervised fuzzy k -means classification accuracy	89.6%	99.0%	100.0%

Table 5.1: Cluster separability, corresponding to Fig. 5.4 (top).

5.4 Evaluating the Goodness of Invariance: Residual Variance after Filtering

■ Quantitative Analysis:

With real data a true *invariance* can certainly not be achieved. For evaluation purposes we rather consider the *reduction* of variance achieved by the suggested filtering. In order to test which number $K = N - \text{rank } \mathbf{P}$ of filtered components is sufficient, we have applied filtering (with an atmospheric exponent $\gamma = 1$) to the recorded data sets with an increasing number of filtered components K . Let $\Sigma = \Sigma(\ln \mathbf{x})$ be the covariance matrix of the observed logarithmic spectra. Then $\text{tr } \Sigma$ is the total variance, and $\text{tr}(\mathbf{P}_K \Sigma \mathbf{P}_K)$ is the residual variance after filtering of K components.² In Table 5.2 on page 107 we give the variance reduction in percent achieved by sequential filtering with an increasing number K for each recorded data set:

$$\frac{\text{tr}(\mathbf{P}_K \Sigma \mathbf{P}_K) - \text{tr}(\mathbf{P}_{K+1} \Sigma \mathbf{P}_{K+1})}{\text{tr}(\mathbf{P}_K \Sigma \mathbf{P}_K)} \times 100\% \quad (5.14)$$

For a typical sample size of $n = 300$ we get an error for the estimated variance of $\pm 8.2\%$ (Eq. 5.8 on page 100). Demanding a 99% confidence level, we thus recognize a significant reduction of variance when it is larger than $3\sigma \hat{=} 25\%$.

From the tabulated results it becomes clear that brightness and color filtering with $K = 2$ always yields a substantial reduction of variance in comparison to pure brightness filtering with $K = 1$. The reduction is essentially always above the 25% margin required for the 99% confidence level. Results do not necessarily improve with filtering more components. A projector with $K = 3$ will only in some cases diminish the residual variance, and $K = 4$ finally is only of minor effect.

The figures indicate that it is a robust approach to filter with $K = 2$, *i.e.*, with the brightness and color shifts \mathbf{u} and \mathbf{v} . The corresponding filter matrix is $\mathbf{P}_{K=2} = \mathbf{V}\mathbf{U}$.

■ Dependence on the Atmospheric Parameter γ :

We also want to determine the dependence of the color shift filtering on the atmospheric parameter γ , which is needed to estimate the diffuse to global radiation ratio \mathbf{m} and to finally compute the color filtering projector matrix \mathbf{V} (see Eq. 4.103 on page 65). Note that the brightness filter matrix \mathbf{U} (see Eq. 4.32 on page 53) is parameter-free.

A comparison of filters with different values $\gamma \in [0.5, 1.0, 2.0, 3.0]$ is shown in Table 5.3 on page 108. The fourth column of Table 5.3 gives the relative reduction of the total variance achieved by filtering the brightness shifts using the projector \mathbf{U} :

$$\frac{\text{tr } \Sigma - \text{tr}(\mathbf{U}\Sigma\mathbf{U})}{\text{tr } \Sigma} \times 100\% \quad (5.15)$$

The fifth to eighth column of Table 5.3 gives the relative reduction of the total variance achieved by filtering the color shifts, where the projector $\mathbf{V} = \mathbf{V}_\gamma$ is computed with

² Recall that \mathbf{P} is symmetric: $\mathbf{P}^T = \mathbf{P}$. Thus $\mathbf{P}_K \Sigma \mathbf{P}_K^T = \mathbf{P}_K \Sigma \mathbf{P}_K$, and we can conveniently drop the superscript T ('transposed').

varying atmospheric parameter γ :

$$\frac{\text{tr}(\mathbf{U}\Sigma\mathbf{U}) - \text{tr}(\mathbf{V}_\gamma\mathbf{U}\Sigma\mathbf{U}\mathbf{V}_\gamma)}{\text{tr}(\mathbf{U}\Sigma\mathbf{U})} \times 100\% \quad (5.16)$$

The results indicate that the maximum variance reduction occurs at different parameter values of γ without revealing a clear scheme. We can hardly determine a certain value of γ for a recording day or even only a series of data sets.

On the other hand, the tabulated data suggests that the choice of the atmospheric parameter γ is only of minor importance for the reduction of variance. **This means that the color shift filtering discussed here is not critically sensitive to a correct estimation of γ , but rather robust and thus virtually parameter-free.** Following meteorological literature (see Section 3.1 on page 35) we will henceforth employ $\gamma = 1$ as a typical atmospheric exponent.

VARIANCE REDUCTION BY SEQUENTIAL FILTERING OF K COMPONENTS							
data set	material	no. of spectra	total variance	variance reduction [%]			
				$K = 1$	$K = 2$	$K = 3$	$K = 4$
june1G	white plastic	411	38.1	99.7	39.5	24.5	4.3
june1H	wood	284	61.3	99.2	87.8	38.7	6.5
june1I	white cardboard	599	10.8	99.3	79.2	8.4	1.2
june1K	cork tile	360	19.6	99.3	85.2	4.0	1.8
june1L	white tile	340	16.5	99.2	85.6	11.2	1.3
june1N	plywood	410	42.0	99.1	68.8	30.9	2.5
june1O	white plastic	517	25.6	99.4	86.3	11.9	1.6
june21R	wood	301	17.4	99.4	44.0	4.0	0.5
june21V	white cardboard	248	24.2	98.2	10.1	0.4	0.3
june23G	red roof tile	193	41.8	99.4	80.1	31.4	8.1
june23H	white brick	193	58.1	99.3	79.9	48.1	0.9
june23I	red brick	255	71.6	99.6	82.2	20.6	9.2
june23J	white tile	216	84.3	99.2	51.8	39.9	2.2
june23K	white tile	345	39.4	99.4	78.4	43.4	4.5
june23L	red tile	272	63.2	99.7	81.9	18.7	8.1
june23M	wood	300	52.8	99.3	76.0	54.2	0.5
june23N	mossy roof tile	286	37.3	99.0	61.5	10.3	17.6
june23O	white cardboard	334	67.6	99.5	80.1	37.0	2.0
june23P	roofing felt	337	21.3	92.6	12.6	1.8	1.6
june23Q	cork tile	308	146.3	99.7	84.3	14.4	17.9
june23U	pluster	245	103.7	99.5	89.6	18.9	13.8
july13G	cork tile	398	118.8	99.7	86.5	3.4	5.2
july13K	white tile	398	144.5	99.7	84.3	19.6	2.4
july13L	white brick	345	76.5	99.8	55.5	56.0	1.6
july13M	red roof tile	261	51.7	99.9	31.8	6.5	1.0
july13N	white cardboard	260	73.5	99.8	65.4	28.7	4.3
july13O	red brick	286	80.5	99.9	19.2	4.3	2.0
july13Q	wood	249	85.2	99.6	83.3	49.5	3.1

Table 5.2: Variance reduction by sequential filtering with an increasing number K of filtered components (atmospheric parameter $\gamma = 1$).

VARIANCE REDUCTION W.R.T VARYING γ PARAMETER							
data set	material	no. of spectra	variance reduction [%]				
			U filter	$V_{\gamma=0.5}$	$V_{\gamma=1}$	$V_{\gamma=2}$	$V_{\gamma=3}$
june1G	white plastic	411	99.7	38.9	39.5	40.6	41.5
june1H	wood	284	99.2	86.9	87.8	89.4	90.5
june1I	white cardboard	599	99.3	79.5	79.2	78.5	77.5
june1K	cork tile	360	99.3	85.4	85.2	84.6	83.6
june1L	white tile	340	99.2	86.0	85.6	84.6	83.3
june1N	plywood	410	99.1	68.0	68.8	70.2	71.3
june1O	white plastic	517	99.4	86.3	86.3	85.9	85.3
june21R	wood	301	99.4	43.6	44.0	44.6	45.0
june21V	white cardboard	248	98.2	10.1	10.1	10.2	10.2
june23G	red roof tile	193	99.4	79.0	80.1	82.0	83.4
june23H	white brick	193	99.3	78.6	79.9	82.2	84.2
june23I	red brick	255	99.6	81.3	82.2	83.5	84.5
june23J	white tile	216	99.2	50.8	51.8	53.8	55.6
june23K	white tile	345	99.4	77.1	78.4	80.6	82.5
june23L	red tile	272	99.7	80.3	81.0	82.1	82.8
june23M	wood	300	99.3	74.5	76.0	78.7	81.1
june23N	mossy roof tile	286	99.0	60.8	61.5	62.6	63.4
june23O	white cardboard	334	99.5	79.2	80.1	81.7	83.0
june23P	roofing felt	337	92.6	12.4	12.6	12.9	13.2
june23Q	cork tile	308	99.7	83.9	84.3	84.8	84.9
june23U	pluster	245	99.5	88.9	89.6	90.5	90.9
july13G	cork tile	398	99.7	86.6	86.5	86.1	85.4
july13K	white tile	398	99.7	84.7	84.3	83.3	82.0
july13L	white brick	345	99.8	54.6	55.5	57.2	58.8
july13M	red roof tile	261	99.9	31.4	31.8	32.5	33.0
july13N	white cardboard	260	99.8	64.3	65.4	67.4	69.1
july13O	red brick	286	99.9	19.0	19.2	19.4	19.5
july13Q	wood	249	99.6	82.1	83.3	85.6	87.4

Table 5.3: Variance reduction by filtering brightness shift and color shift with varying atmospheric parameter γ .

5.5 Assessing the Impact of Surface Orientation: Magnitude of Brightness and Color Shifts

Analysis of the experimental data enables us to assess the magnitude of the brightness and color shifts. In Eq. 4.112 on page 66 we have introduced a brightness index $\hat{\mathbf{u}}^T \ln \mathbf{x}$ and a color index $\hat{\mathbf{v}}^T \ln \mathbf{x}$, which tell us just how much of the variable brightness and color components is present in an observed log spectrum $\ln \mathbf{x}$. Both are determined for each spectrum of the experimental data sets. Then the respective variances are computed for each set:

$$\text{brightness variance: } \sigma_{\hat{\mathbf{u}}}^2 = \sigma^2(\hat{\mathbf{u}}^T \ln \mathbf{x}) \quad (5.17)$$

$$\text{color variance: } \sigma_{\hat{\mathbf{v}}}^2 = \sigma^2(\hat{\mathbf{v}}^T \ln \mathbf{x}) \quad (5.18)$$

$$\text{and the correlation: } \frac{\sigma_{\hat{\mathbf{u}}\hat{\mathbf{v}}}^2}{\sigma_{\hat{\mathbf{u}}} \cdot \sigma_{\hat{\mathbf{v}}}} \quad (5.19)$$

Leaving the log space, we are also interested to see how large these effects are in the original measurements space. Therefore we multiply the log standard deviation, *i.e.*, the square root of the above computed variance, with the magnitude of the respective component $\hat{\mathbf{u}}$ or $\hat{\mathbf{v}}$ in spectral band $i \in [1..N]$, and exponentiate the product. Then we can express the relative standard deviation in percent by the following formulae:

$$\text{positive deviations: } + (e^{\sigma_{\hat{\mathbf{u}}}\hat{u}_i} - 1) \times 100\% \quad (5.20)$$

$$+ (e^{\sigma_{\hat{\mathbf{v}}}\hat{v}_i} - 1) \times 100\% \quad (5.21)$$

$$\text{negative deviations: } - (1 - e^{-\sigma_{\hat{\mathbf{u}}}\hat{u}_i}) \times 100\% \quad (5.22)$$

$$- (1 - e^{-\sigma_{\hat{\mathbf{v}}}\hat{v}_i}) \times 100\% \quad (5.23)$$

The brightness shift is of equal magnitude in all spectral bands i . In contrast, the color shift varies with wavelength, and its magnitude is extreme at both ends of the wavelength range, and we thus take $i = 1$ corresponding to λ_{\min} for evaluation.

The results are given in Table 5.4 on the following page. **We observe that no systematic correlation between brightness and color shift is present. This means that the independent filtering of both shifts is really necessary and cannot be achieved by filtering of a single linear combination of both.**

The tabulated data shows that in our experiments the color shift caused relative deviations in the measurement space which are one order of magnitude smaller than those of the brightness shift, but still in the 5%-range. Note that these are standard deviations, and that the extreme deviations (*e.g.* in pure shadow) are significantly larger.

MAGNITUDE OF BRIGHTNESS AND COLOR SHIFTS								
data set	material	no. of spectra	variance in log space				rel. variance [%] in measurement space	
			total tr Σ	b.ness $\sigma_{\mathbf{u}}^2$	color $\sigma_{\mathbf{v}}^2$	correl.	brightness	color
june1G	white plastic	411	38.09	37.98	0.04	0.58	+54.6/-35.3	+2.9/-2.9
june1H	wood	284	61.27	60.81	0.40	0.73	+73.5/-42.3	+9.2/-8.4
june1I	white cardboard	599	10.79	10.71	0.06	-0.50	+26.0/-20.6	+3.5/-3.4
june1K	cork tile	360	19.56	19.42	0.11	-0.81	+36.5/-26.7	+4.8/-4.5
june1L	white tile	340	16.51	16.38	0.11	-0.71	+33.1/-24.8	+4.7/-4.5
june1N	plywood	410	42.00	41.61	0.26	-0.07	+57.8/-36.6	+7.4/-6.9
june1O	white plastic	517	25.60	25.45	0.13	-0.43	+42.8/-30.0	+5.1/-4.9
june21R	wood	301	17.37	17.26	0.04	0.80	+34.1/-25.4	+3.1/-3.0
june21V	white cardboard	248	24.15	23.72	0.04	0.80	+41.1/-29.1	+2.9/-2.8
june23G	red roof tile	193	41.79	41.56	0.18	0.88	+57.7/-36.6	+6.1/-5.8
june23H	white brick	193	58.12	57.74	0.30	0.97	+71.1/-41.5	+7.9/-7.3
june23I	red brick	255	71.55	71.27	0.22	0.96	+81.6/-44.9	+6.8/-6.3
june23J	white tile	216	84.33	83.64	0.35	0.74	+90.9/-47.6	+8.6/-7.9
june23K	white tile	345	39.41	39.19	0.17	0.90	+55.6/-35.7	+6.0/-5.6
june23L	red tile	272	63.23	63.03	0.15	0.90	+75.3/-42.9	+5.6/-5.3
june23M	wood	300	52.78	52.39	0.30	0.84	+66.8/-40.0	+7.8/-7.3
june23N	mossy roof tile	286	37.26	36.90	0.21	0.60	+53.6/-34.9	+6.6/-6.2
june23O	white cardboard	334	67.55	67.24	0.25	0.95	+78.5/-44.0	+7.2/-6.7
june23P	roofing felt	337	21.33	19.75	0.19	0.43	+36.9/-26.9	+6.3/-5.9
june23Q	cork tile	308	146.26	145.87	0.32	0.19	+134.9/-57.4	+8.2/-7.6
june23U	pluster	245	103.66	103.18	0.43	0.60	+105.0/-51.2	+9.5/-8.7
july13G	cork tile	398	118.82	118.46	0.31	-0.78	+115.9/-53.7	+8.0/-7.4
july13K	white tile	398	144.50	144.13	0.31	-0.45	+133.7/-57.2	+8.1/-7.5
july13L	white brick	345	76.54	76.42	0.06	0.33	+85.5/-46.1	+3.6/-3.4
july13N	white cardboard	260	73.46	73.34	0.07	0.91	+83.2/-45.4	+3.9/-3.8
july13O	red brick	286	80.47	80.37	0.01	0.47	+88.4/-46.9	+1.9/-1.9
july13Q	wood	249	85.20	84.85	0.29	0.76	+91.8/-47.8	+7.7/-7.2

Table 5.4: Magnitudes of brightness and color shifts.

5.6 Principal Component Analysis of the Logarithmic Spectra

In Section 4.3 on page 59 we have shown that the spectral variability caused by surface orientation changes may be conveniently analyzed in the logarithmic space. Let us recall that the variance in the logarithmic space is independent of the reflectance:

$$\text{var}\langle \ln x_i \rangle = \text{var}\langle \ln(r_i(\mu m_i + \nu n_i)) \rangle \quad (5.24)$$

$$= \text{var}\langle \ln r_i \rangle + \text{var}\langle \ln(\mu m_i + \nu n_i) \rangle \quad (5.25)$$

$$= 0 + \text{var}\langle \ln(\mu m_i + \nu n_i) \rangle \quad (5.26)$$

$$= \text{var}\langle \ln(\mu m_i + \nu n_i) \rangle \quad (5.27)$$

where $(\mu m_i + \nu n_i)$ is the spectral irradiance with direct and diffuse components n_i and m_i with contribution factors ν and μ , and r_i is the constant spectral reflectance. Hence we can state that **the covariance matrix $\Sigma(\ln \mathbf{x})$ in the logarithmic space is independent of the reflectance spectrum \mathbf{r} .**

We have applied the principal component analysis to the log spectra of all recorded data sets. Two examples are shown in Fig. 5.5, the other sets are shown in the appendix (see Fig. B.1 on page 187 ff.).

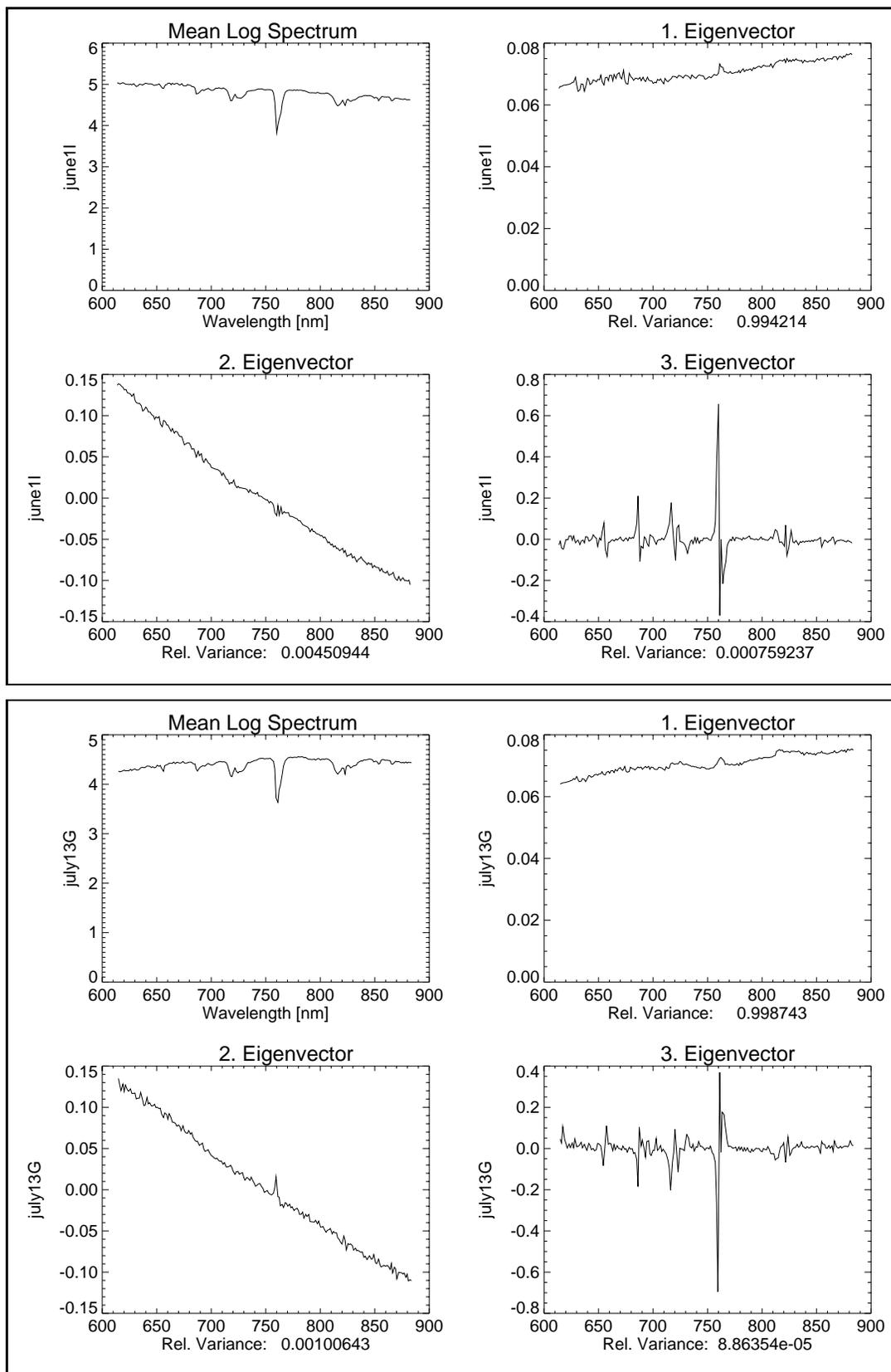


Figure 5.5: Mean logarithmic spectrum $\langle \ln \mathbf{x} \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra of two materials (top: white cardboard, bottom: reddish cork tile) observed under arbitrary angles on two different days.

The principal component analysis gives a good qualitative picture of the basic modes of the spectral variability, and shows it to be in good agreement with our analysis in Chapter 4:

- ▶ The difference of the surface material reflectances essentially shows up in the mean log spectra only, not in the log eigenvectors. (In Fig. 5.5 we see a white cardboard specimen versus a reddish cork tile.)
- ▶ The variance is always captured by the first few eigenvectors. We have plotted the first three eigenvectors for all data sets and annotated their respective relative variance (eigenvalues over total variance).
- ▶ Typically, the first eigenvector of the log covariance matrix $\Sigma(\ln \mathbf{x})$ is a rather wavelength independent brightness shift and represents a scaling of the overall spectrum.
- ▶ In contrast, the second eigenvector typically is a monotonous wavelength dependent ‘color’ shift, which is caused by the relative diffuse to global illumination ratio spectrum $\hat{\mathbf{m}}$.
- ▶ The typical third eigenvector is caused by variability in the absorption bands, and by slight drifts in the wavelength calibration: shifts of ca. 2 CCD-pixels $\hat{=} 1.2$ nm cause the up-down-spikes at the absorption band locations.
- ▶ Furthermore we also observe impurities in the 600–700 nm range which are probably caused by specular reflection (*e.g.* Fig. B.1 on page 194, top).

We note, however, that the variance eigenvectors are not identical and do not show pure brightness and color shift vectors (such as depicted in Fig. 5.3 on page 102, bottom row). Rather, we observe linear combinations of the brightness and color shift vectors. Moreover, these are also linear combined with measurement impurities such as wavelength instability (typical up-down spikes), and physical impurities such as specular reflection and atmospheric absorption band variability (for the absorption band positions compare with Fig. 5.2 on page 99) which are not related to the problem under investigation here. Note that we do not expect the variance eigenvectors to be identical for the various samples, because the sample sets were recorded from arbitrarily changed surface orientations and under varying illumination conditions. Again, in this thesis we are more interested in learning something about the scope and nature of the spectral variance, rather than about the results for specific illumination/recording geometries.

■ Principal Components of the Residual Variability:

In the first part of this section we have conducted a principal component analysis in order to verify that the variance of reflected spectra as caused by surface orientation changes can be represented by two eigenvectors which are similar to the proposed brightness and color shift vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. For completeness, we want to check this hypothesis the other way round: We analyze the residual variability after the filtering with the projector $\mathbf{P}_{(K=2)} = \mathbf{V}\mathbf{U}$ was applied. The aim is to falsify the possibility that there are other systematic effects hidden in the spectral variability.

The residual covariance matrix $\mathbf{P}\Sigma\mathbf{P}$ is again submitted to principal component analysis of each data set. Then the resulting most significant residual eigenvectors are identified. An example is shown in Fig. 5.6, the other sets are shown in the appendix (Fig. B.2 on page 199 through Fig. B.2).

The results indicate that the residual variability is made up by random noise, absorption band variability, small instabilities in the wavelength calibration (the up-down-spikes at the absorption bands), and specular reflection, but not by any systematic effect which we possibly would have overlooked so far. After appraisal of the residual variance eigenvectors we therefore conclude that we have successfully filtered all systematic variability which is due to surface orientation change, Lambertian reflection provided.

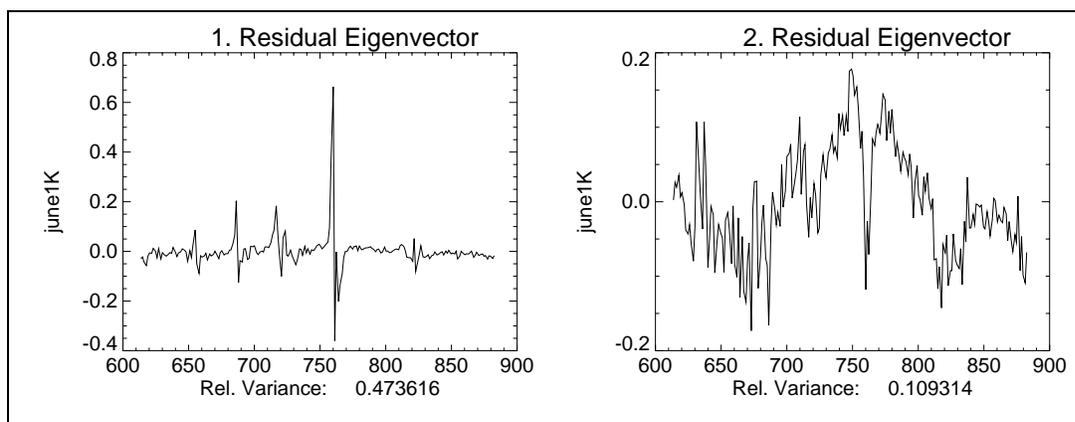


Figure 5.6: *Principal Component Analysis of the residual variance after filtering with $K = 2$.*

Experimental Findings on Multispectral Imagery

Outline of this chapter – After having established a theory of surface orientation related change of illumination in Chapter 3 and 4, and subsequent verification of our analysis and the suggested invariant descriptor in Chapter 5, we finally apply the discussed transform to real multispectral image data from an airborne sensor.

At first, we give a brief description of the airborne sensor, the data acquisition and calibration. Then we show an example of the color shift as measured from the multispectral imagery.

From the imagery we extract a number of sample pairs of spectra from the same surface under different orientation. For each pair, both spectra are compared in the original, the normalized, and the brightness and color filtered space. Both effects of surface orientation, the brightness and color shift, can be observed qualitatively in the plotted spectra. The same effects can be observed in a bitemporal comparison of tilted surface patches. The suggested logarithmic color filtering can substantially reduce the relative spectral distance between the two spectra of each sample pair.

Spectral distances are the basic key to pixel-wise multispectral image classification, such as Minimum Euclidean Distance (MED) or Minimum Mahalanobis Distance/Maximum Likelihood (MMD/ML) classification schemes. For qualitative evaluation of the suggested invariant, a simple classification into binary class images is performed on image clips around the sample pairs, using different classification schemes. We observe clear improvement of classification in the suggested transformed feature space, and diminished sensitivity to threshold parameters.

The effect of the suggested transition into the brightness and color filtered feature space is visualized by projection of the N -dimensional feature spaces onto a two-dimensional subspace spanned by the two most significant principal components. These feature space ‘slices’ are represented as contour plots of bivariate histograms of the two principle components. The suggested transform can be seen to filter a systematic shift inherent in the spectral image data.

Then, for statistical validation of the suggested invariant, we analyze sets of spectral sample pairs. The quantitative advantage achieved by introducing the invariant descriptor is measured as the spectral distance between the two spectra relative to the

spectral scatter radius of the overall image spectra. Ideally, the spectral distance should vanish, since the spectra originate from reflection off the same surface. We evaluate all sample pairs of a set, and the spectral distances and their variances are tabulated for the investigated feature spaces, also with principal component transformations (PCTs) of various ranks. Based on these statistics we can show that the suggested invariant reduces considerably the relative spectral distance between equal surfaces of different orientation.

Last not least, we pursue error considerations again, this time considering the specific case of the multispectral imagery used throughout this chapter. We show that –even after brightness normalization– the color shift can cause deviations up to 70% in the observed pseudo-reflectances as an upper bound. Analysis of the color constancy problem in the imagery used here is compromised as the quantization error is considerable for dimly illuminated surfaces (such as in shadowed areas), due to the restricted dynamic range of the sensor. Therefore we revisit the previously introduced error weighted spectral distance. Application to the sample spectra shows their usefulness, although the computation cost may be too high for common processing of large images.

6.1 The Sensor

The sensor DAEDALUS AADS 1268 is an airborne multispectral line scanner manufactured by Daedalus Enterprises, Inc., Ann Arbor, Michigan, USA. The instrument used for the recording of the subsequent imagery is owned and operated by the German Aerospace Research Establishment (DLR), and usually carried onboard a Dornier DO 228 aircraft.

The sensor has 11 spectral bands, ten in the wavelength range of reflected visible and infrared light, and one in the thermal infrared (see Table 6.1 on the next page). The spectral sensitivity functions (apparatus functions) can be found *e.g.* in (Kollewe 1995).

The instantaneous field of view (IFOV) is $2.5 \text{ mrad} = 0.14^\circ$, yielding a nadir pixel ground resolution of 0.7 m from an altitude of 300 m, and 4.2 m from 1800 m altitude. Only one ground pixel is observed at a time, and the incident beam is split and guided into the specific detectors of the spectral bands. The line scanning is achieved by means of a rotating mirror which drags the one-pixel-IFOV across the scan line perpendicular to the flight track. The next scan line is then offset by the flight velocity divided by the scan rate.

The DAEDALUS has various scan rate modes: 16.7, 33.3 and 100 Hz, *i.e.*, scan lines per second. The flight velocity and altitude has to be adjusted such that subsequent scan lines are indeed adjacent and do not over- or undersample the ground area being imaged. With the Do 228 aircraft, a scan rate of 100 Hz is used for 300 m altitude, 33.3 Hz for 900 m, and 16.7 Hz for 1800 m.

The scan angle θ ranges from -42.96° to $+42.96^\circ$, and this *swath* of 85.92° is divided into 716 pixels of equal angular step width $\Delta\theta = 85.92^\circ/716 = 0.12^\circ$. This causes a panoramic image distortion, *i.e.*, the pixel ground resolution varies with $\cos^{-2}\theta$. *E.g.*, the pixels on the left and right image boundary have a ground resolution twice as large as the nadir pixels. The proper rectification has been described by *e.g.* Zhang et al. (1994) or Wiemker (1996).

The sensor actually records 125 pixels more on each side, corresponding to an extra

15°. A giro subunit mounted on the sensor establishes possible deviations of the aircraft from horizontal flight (rolling). A 716 pixel window out of the overall 966 (= 716 + 2 × 125) scanned pixels is read out for data logging. According to the reading of the giro, the read-out-pixel-window is determined for each scan line such that any deviation from horizontal flight is compensated for (up to ±15°). The image data together with various flight parameters is logged to magnetic tape. For geocoding, the image data may be registered to a map subsequently. Registration techniques particularly suitable for airborne scanner imagery are described *e.g.* by Zhang et al. (1994), Ehlers (1994), Wiemker (1996) and Wiemker et al. (1996).

The analogue signals of the detectors are amplified twice, then digitized by a A/D-converter to a dynamic range of [0..255] and stored in byte types as digital counts [DC]. The gain control of the amplifier can be set in powers of 2 [0.5, 1, 2, 4, 8] in order to exploit the 8 bit dynamic range optimally for the various land or sea imaging applications and depending on weather conditions.

Due to the two amplifiers involved, one has to consider carefully that overflows can occur not only at digital counts of 255 DC, but also at lower DC numbers in case the pre-amplifier is already saturated.

Due to strong noise in the first spectral band $i = 1$, this band is omitted in the subsequent quantitative analysis of the surface orientation related color constancy problem in the multispectral image data.

MULTISPECTRAL LINE SCANNER DAEDALUS AADS 1268											
band number i	1	2	3	4	5	6	7	8	9	10	11
spectral band [μm]	0.420–0.450	0.450–0.520	0.520–0.600	0.605–0.625	0.630–0.690	0.695–0.750	0.760–0.900	0.910–1.050	1.550–1.750	2.080–2.350	8.500–13.000
center wavelength [μm]	0.435	0.485	0.560	0.615	0.660	0.723	0.830	0.980	1.650	2.215	10.750
detector material	Si							InSb		HgCdTe	

Table 6.1: Spectral bands of the multispectral line scanner DAEDALUS AADS 1268.

6.2 Calibration of the Image Data and Atmospheric Correction

The raw image as delivered in digital count [DC] gray-values, is converted to radiances by a linear calibration, taking into account also the gain setting of the recording. The additive and multiplicative calibration constants are determined on a DAEDALUS test bench in the laboratory after the sensor is unmounted from the aircraft platform.

The product of the calibration are radiance images which are dependent on the atmospheric conditions at the recording time. An atmospheric correction is applied in order to turn the measured radiance values into reflectances. The computation of the proper reflectances is done by the package SENSAT-5 (Richter 1990, Richter 1992) by piecewise linear reverse modelling. For this, four different reflectance values are assumed, and for each the way of the radiation coming from the sun, being scattered by the atmosphere, reflected from the assumed horizontal and Lambertian surface and re-transmitted through the atmosphere to the sensor, is simulated by the radiative transfer

code MODTRAN (Anderson et al. 1995) for all wavelength of the DAEDALUS spectral bands, provided a certain radiometric resolution. Thus, for each of the four considered reflectance values, a theoretical radiance value is computed which is expected to be measured by the airborne sensor. Now the actually measured radiances are converted into reflectances by piecewise linear interpolation between the simulated four radiance/reflectance pairs. The process of radiance calibration and reflectance computation of the imagery used here is described in detail by Hepp (1994), Kollewe (1995), and Kollewe et al. (1996).

Even following the assumption of Lambertian reflexion, we still prefer to call the values in the so obtained reflectance images only *pseudo-reflectances*, as the computation was based on the assumption of the reflecting surfaces being horizontal.

We want to point out the following effects which may compromise the analysis of color constancy problem if not properly corrected:

- **Path radiance:** Solar radiance which is scattered on air molecules and aerosols directly into the sensor without reflection from any ground surface. The path radiance is an additive offset constant for all reflexion signals.
- **Adjacency effect:** Radiance which was reflected in the neighborhood around the surface in the current field of view (IFOV pixel), and then scattered into the line of sight of the sensor.
- **Overflow/Underflow:** Very high and very low radiances cannot be properly coded when constrained by the 8 bit dynamic range.

Interestingly enough, errors in the assessment of the solar irradiance or atmospheric transmissivity will *not* affect the analysis of the spectral impact of surface orientation, since we are always considering spectral ratios or logarithmic differences.

Three example images are given in Fig. 6.1 through Fig. 6.3. The prints show the gray-values of a single spectral band ($i = 7$, at $\lambda = 980$ nm) and are contrast enhanced by means of histogram equalization (Richards 1993). The imagery was recorded by the DAEDALUS AADS 1268 line scanner onboard a Dornier Do 228 during four campaigns from 1991 to 1995 in cooperation with the German Aerospace Research Establishment (DLR) at flight altitudes of 300 m and 1800 m (nadir ground resolution 70 cm and 4.2 m, respectively). For the spectral measurements in this thesis only the 300 m data was used because of the higher spatial resolution.



zieg91cor

Figure 6.1: Scene ‘Ziegelstein’, recorded 1991 at 300 m altitude over Nürnberg on August 21, 1991, 13h50 CET, solar zenith angle $\theta_{\odot} = 41^{\circ}$ (at $\lambda = 980$ nm, contrast enhanced).



gew92cor

Figure 6.2: Scene ‘Gewerbegebiet’ recorded from 300 m altitude over Nürnberg on April 25, 1992, 12h20 CET, solar zenith angle $\theta_{\odot} = 36^{\circ}$ (at $\lambda = 980$ nm, contrast enhanced).



gew94cor

Figure 6.3: Scene ‘Gewerbegebiet’ recorded from 300 m altitude over Nürnberg on October 18, 1994, 11h30 CET, solar zenith angle $\theta_{\odot} = 59^{\circ}$ (at $\lambda = 980$ nm, contrast enhanced).

6.3 Analysis of Selected Spectral Samples

6.3.1 *In scene* Example of the Color Shift

From a multispectral image we have selected a roof top and clipped a polygonal region where the same surface material is visible under different surface orientations (Fig. 6.4). The spectra contained in the selected area were extracted and treated in the logarithmic spectral space. First the brightness was normalized by filtering the unity vector: $\mathbf{x} \mapsto \mathbf{U} \ln \mathbf{x}$ (for definition of \mathbf{U} see Eq. 4.32 on page 53).

In order to find the next most significant eigenvector of the spectral variability which has remained after this logarithmic brightness filtering, a principal component analysis is conducted. The vector entries of the most significant eigenvector (which carries 80% of the total variance) are plotted in Fig. 6.5 (box marks). The spectral dependence of the eigenvector is compared to the diffuse to global illumination ratio spectrum \mathbf{m} with $m_i = c\lambda_i^{-\gamma}$ (Eq. 3.1 on page 36). In our example, the best fitting atmospheric parameter is $\gamma = 2.2$. For comparison with the variability eigenvector, the normalized and brightness filtered diffuse-to-global illumination ratio $\hat{\mathbf{v}} \propto \mathbf{U}\mathbf{m}$ (Eq. 4.101 on page 65) is plotted in Fig. 6.5 (solid line). We observe striking agreement.

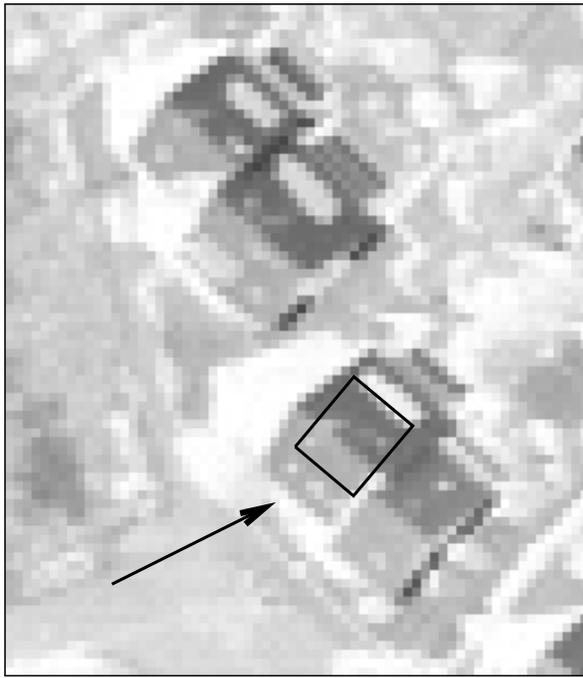


Figure 6.4: Region of analyzed spectra from a roof top (scene ‘Ziegelstein 1991’, Fig. 6.1 on page 119, bottom left near soccer field).

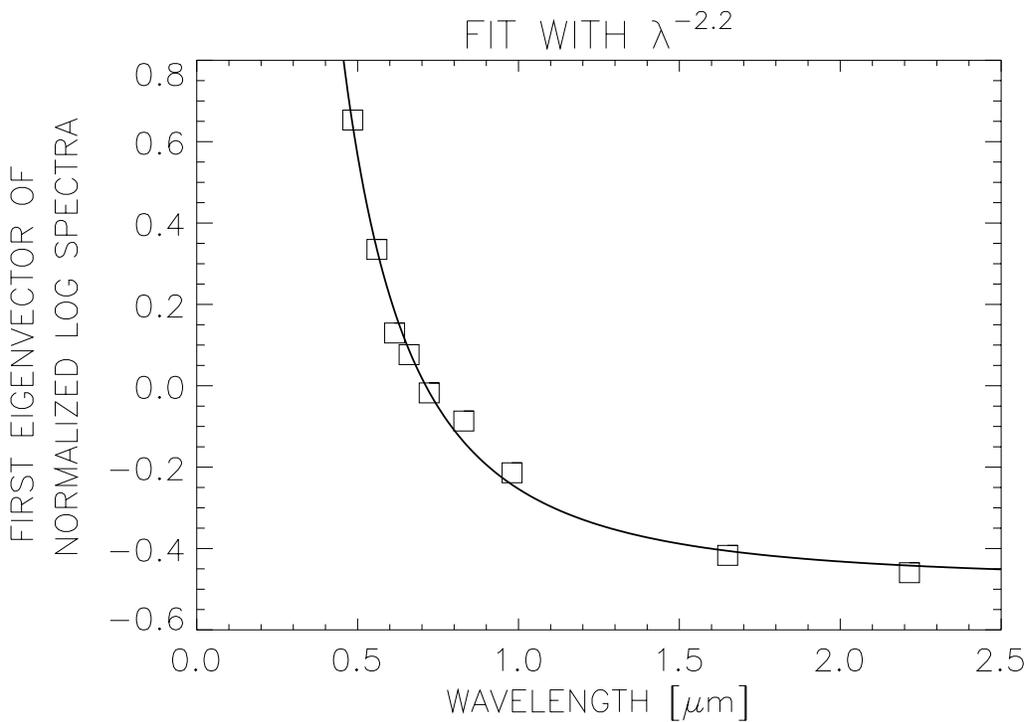


Figure 6.5: The entries of the most significant eigenvector (boxes) of the measured spectra, and the best fitting diffuse to global illumination ratio vector $\hat{\mathbf{v}} \propto \mathbf{U}\mathbf{m}$.

6.3.2 Comparison of Spectral Sample Pairs in Various Feature Spaces

Surveying several image scenes, we have located samples of spectral pairs which exhibit the aspect of the color constancy problem as discussed here, namely, which stem from equal surface materials under different surface orientation. The spectra are extracted from the imagery and we compare the two spectral signatures. Their pseudo-reflectance curves differ, and we want to test whether they become more similar in the transformed feature spaces.

A number of examples are given in Fig. 6.6 through Fig. 6.9 (more sample pairs are shown in the appendix, see Fig. C.1 on page 204 ff., and Fig. C.3 on page 217 ff.). On the top left of Fig. 6.6, a clip from the image scene shows the location of the selected pair of spectra. On top right, the two measured reflectance spectra \mathbf{x} and \mathbf{x}' are plotted (marked by boxes and triangles, respectively). The spectral values are extracted as the mean value from a 3×3 neighborhood. The spectra are plotted with error bars which indicate the root mean square scattering in this neighborhood. In all cases, we observe a clear difference in brightness between the two spectra.

Then the pseudo-reflectances \mathbf{x} are brightness normalized by conventional L_2 -norm vector normalization: $\mathbf{x} \mapsto \hat{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|$. The resulting normalized spectra $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}'}$ are plotted on the second row, left, in Fig. 6.6. The difference vector, *i.e.*, the vector containing the difference between the two normalized spectra for each spectral band i , is plotted on the right-hand side. (The error bars of the difference vector are computed according to error propagation as the square root of the added scattering squares.)

We observe a clear wavelength dependence of the difference vectors. Obviously, there is not only a brightness shift but also a color shift between the two spectral signatures. The color shift is due to varying contributions of skylight and sunlight as analyzed in Chapter 3 and 4. This color shift cannot be removed by the brightness normalization. We quantify the remaining residual between the two normalized spectra $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}'}$ as the mean relative deviation in all spectral bands i :

$$\left\langle \frac{\hat{x}_i - \hat{x}'_i}{\frac{1}{2}(\hat{x}_i + \hat{x}'_i)} \right\rangle \times 100\% \quad . \quad (6.1)$$

This residual is given in percent on top of plot of the the difference spectrum.

Then all spectra of the image clip are subjected to logarithmic brightness filtering: $\mathbf{x} \mapsto \mathbf{t} = \exp(\mathbf{U} \ln \mathbf{x})$. Again, two spectra are extracted as the mean value from a 3×3 neighborhood, and the root mean square scatter in this neighborhood is indicated by the error bars (Fig. 6.6, third row left). The residual difference spectrum is again plotted on the right-hand side, and the mean relative deviation given as:

$$\left\langle \frac{t_i - t'_i}{\frac{1}{2}(t_i + t'_i)} \right\rangle \times 100\% \quad . \quad (6.2)$$

Finally, the logarithmic brightness and color filtering is applied: $\mathbf{x} \mapsto \mathbf{t} = \exp(\mathbf{P} \ln \mathbf{x}) = \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$. The results are plotted in the bottom row of each figure.

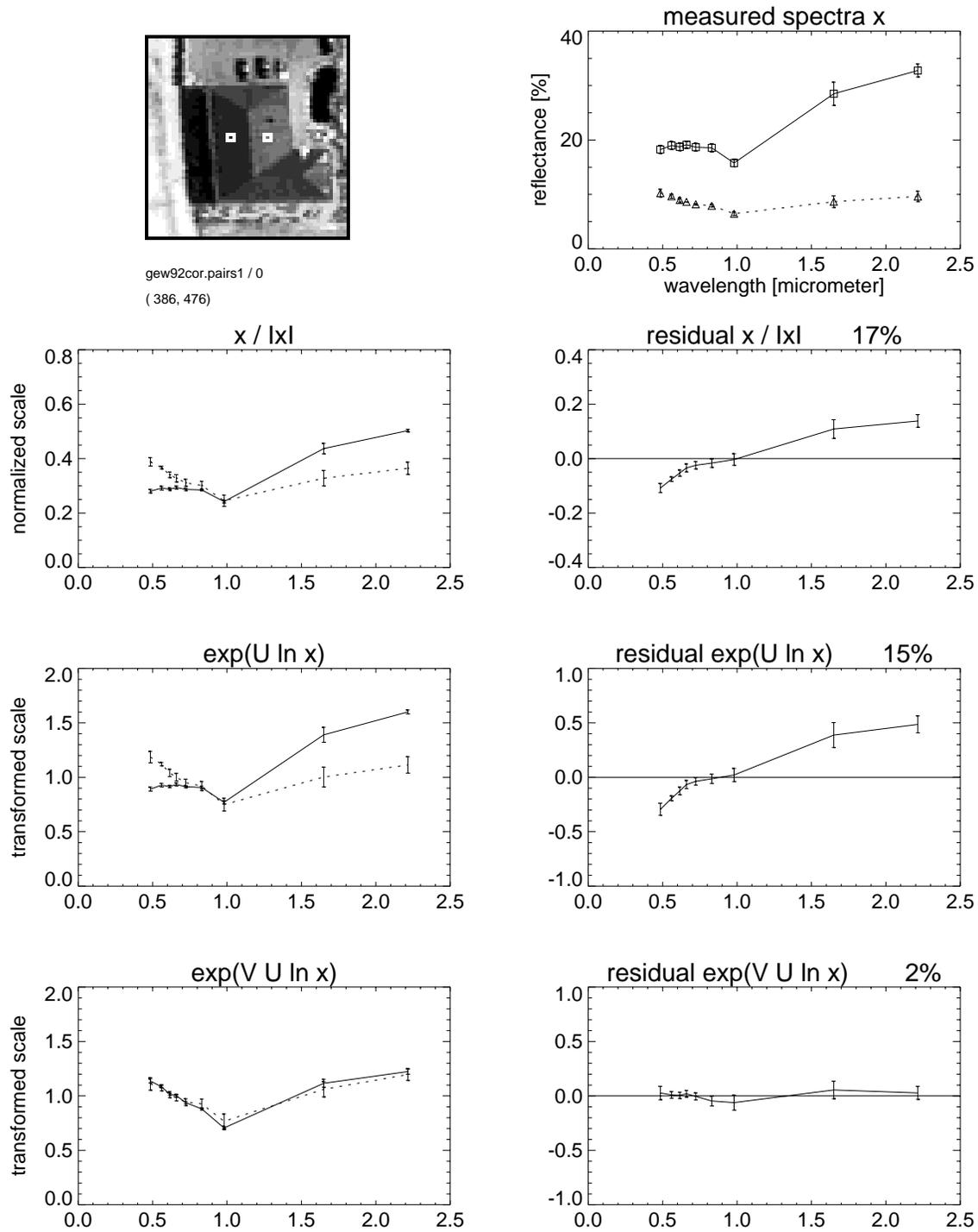


Figure 6.6: Spectral sample pair from scene ‘Gewerbegebiet 1992’ (Fig. 6.2 on page 120).

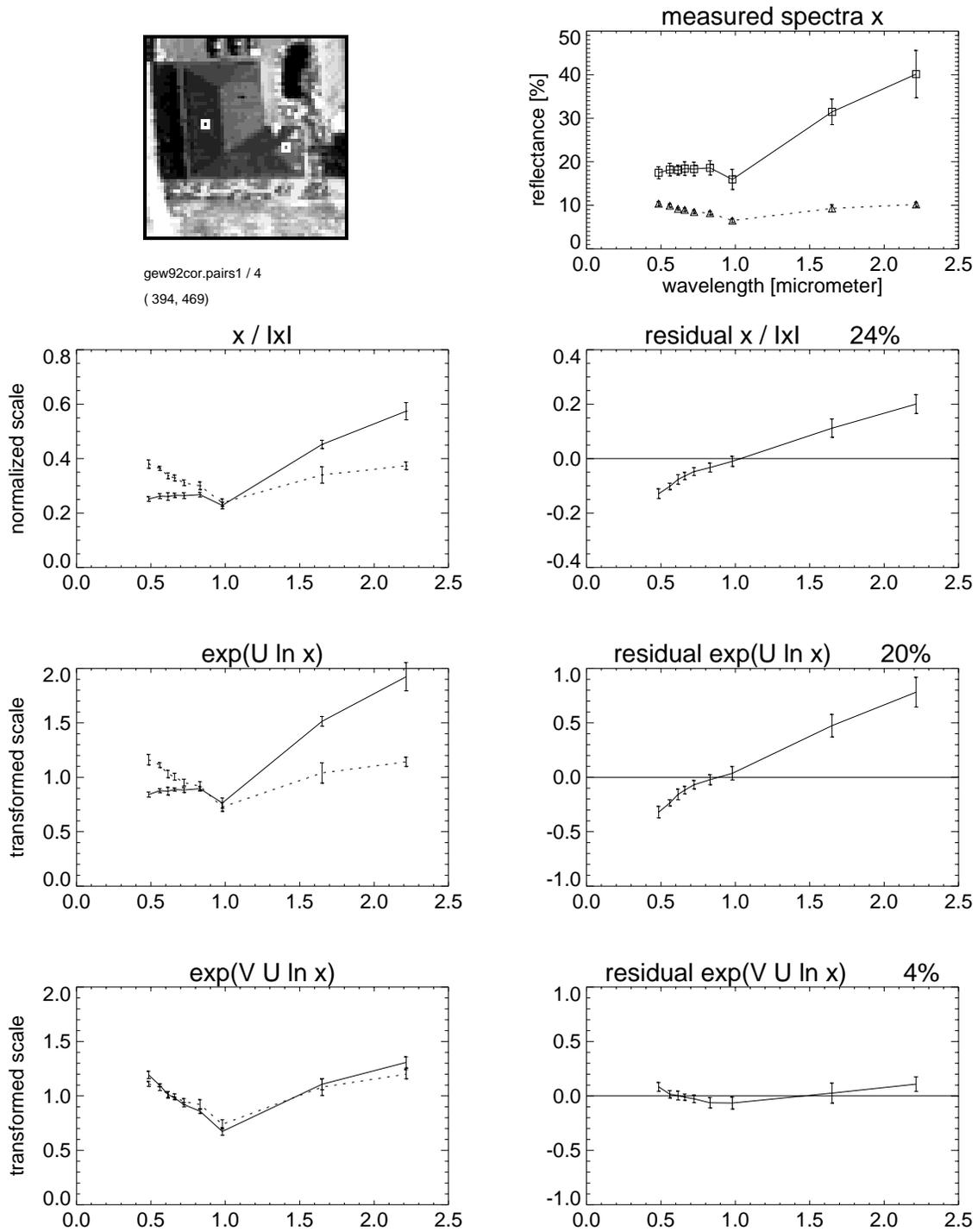


Figure 6.7: Spectral sample pair from scene ‘Gewerbegebiet 1992’ (Fig. 6.2 on page 120).

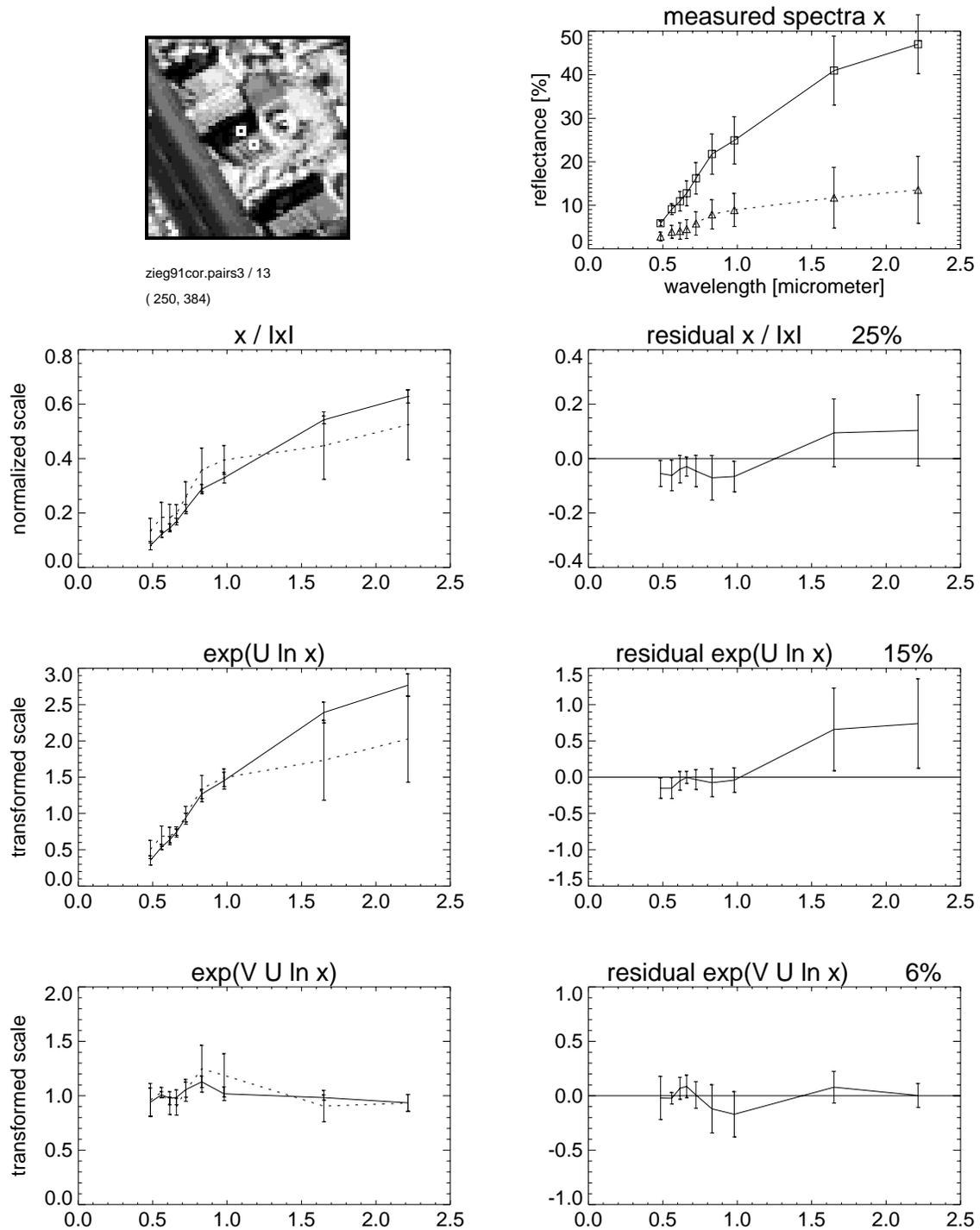


Figure 6.8: Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

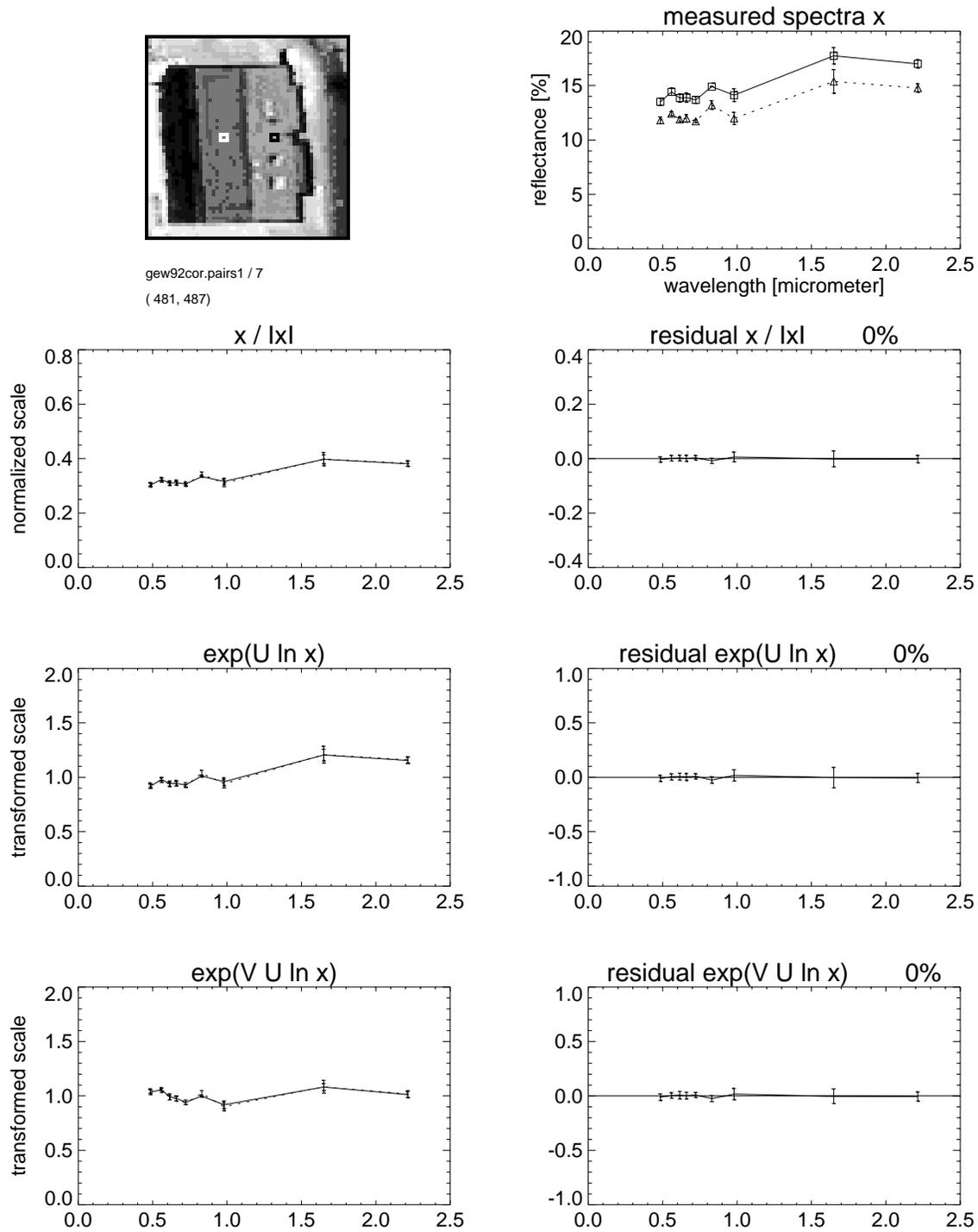


Figure 6.9: Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

6.3.3 Discussion of Results

We summarize our observations from the comparison of the spectral sample pairs as follows:

- ▶ Necessarily, the originally extracted spectral signatures disagree severely. One of them is always of much higher pseudo-reflectance than the other, even though they we are confident that they do stem from the same true reflectance spectrum.
- ▶ Some spectral pairs agree already perfectly after normalization or log brightness filtering only (see particularly Fig. 6.9).
- ▶ The reduction of the mean relative residual is similarly good for both the brightness normalization and the logarithmic brightness filtering, if somewhat more favourable for the logarithmic brightness filtering.
- ▶ In most of the spectral pairs, however, the residual deviation is not arbitrary, but we can rather clearly see a monotonous wavelength dependence (*e.g.* Fig. 6.6 and Fig. 6.7). This we interpret as the same color shift which we have observed in the spectral data discussed in Chapter 5.
- ▶ The additional logarithmic color filtering [$\mathbf{x} \mapsto \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$] then successfully filters this effect. The mean relative residual is reduced substantially.
- ▶ The remaining residuals seem quite obviously to stem from stray radiation from vegetated surfaces in the neighborhood. The characteristic rise at $\lambda \approx 700$ nm strongly suggests that the adjacency effect mixes vegetation spectra from nearby surfaces into the observed spectra (see particularly Fig. C.3 on page 220).

6.3.4 Bitemporal Comparison of Spectral Samples

Multitemporal *change detection* is a central task for all kinds of monitoring applications using remotely sensed imagery. It compares multitemporal imagery in order to detect changes in the land cover (Lillesand & Kiefer 1987, Richards 1993).

In general, remotely sensed multispectral imagery for monitoring purposes is recorded by overflights over the same land area at two times, T_1 and T_2 , say. An appropriate algorithm must then compare the two observed images of the same scene and assist the analyst by designating those areas where the ground cover has apparently changed. For specific applications, certain wavelength bands may be selected, whereas for general purpose monitoring, all spectral bands will be taken into account. Several strategies for change detection can be followed (Singh 1989):

- Register the two images and perform a pixel-wise comparison, or extract only spectra of interest from corresponding locations in both images. Then compare the spectral signature which the corresponding pixels offer in both observations (Wiemker & Hepp 1994, Wiemker & Spitzer 1996, Wiemker et al. 1996, Spitzer et al. 1997, Wiemker 1997a, Wiemker et al. 1997).
- Classify and segment both images independently, and then compare the results from corresponding locations.
- Start out from a knowledge base and a generic model of the scene, then compare the model-derived expectation with the image of the scene as received by the remote sensor (Dreschler-Fischer et al. 1993, Lange & Schröder 1994).

Here we will discuss only the direct comparison of the spectral signatures of the same object as it appears in two different images. This low level approach is the one which is most widely used.

We have selected spectral sample pairs with spectra of the same roof patch as observed in two different years: April 1992 and October 1994. As we have discussed in Chapter 1 and 2 we expect that the spectral signatures of tilted surfaces do not agree between two recording times with different solar angle, even if the spectral raw values have been atmospherically corrected into pseudo-reflectances (Section 2.1 on page 13).

Some examples are shown in Fig. 6.10 through Fig. 6.12 (more can be found in the appendix, Section C.5 on page 242). The corresponding image clips from the two recordings are shown on the top left of each figure. The regions from which the spectra were extracted are marked. The two spectral signatures are compared in the same way as described earlier in Section 6.3.2 on page 124. Comparing these spectra from patches of identical orientation but under varying solar angle, we observe the same results as before with the spectra taken from a single image but under different orientation (Section 6.3.3 on the preceding page):

- ▶ Sometimes the two spectra essentially differ only in brightness, *i.e.*, by an overall scaling factor (*e.g.* Fig. 6.12). After brightness normalization or log brightness filtering we find no more wavelength dependent difference.
- ▶ However, other spectral pairs differ severely even after brightness normalization / filtering (Fig. 6.10 and Fig. 6.11). Then the spectral difference exhibits the typical

wavelength dependence which can be explained by varying contributions of direct and diffuse illumination,

- The remaining spectral difference can be successfully filtered by the suggested transform. Then the spectral distance between the two spectra is reduced considerably (Fig. 6.10 and Fig. 6.11).

Thus, we can conclude that the surface orientation has a non-negligible impact on multitemporal comparison of spectral signatures even assuming Lambertian reflection, and that the suggested projector may successfully prevent erroneous change detection.

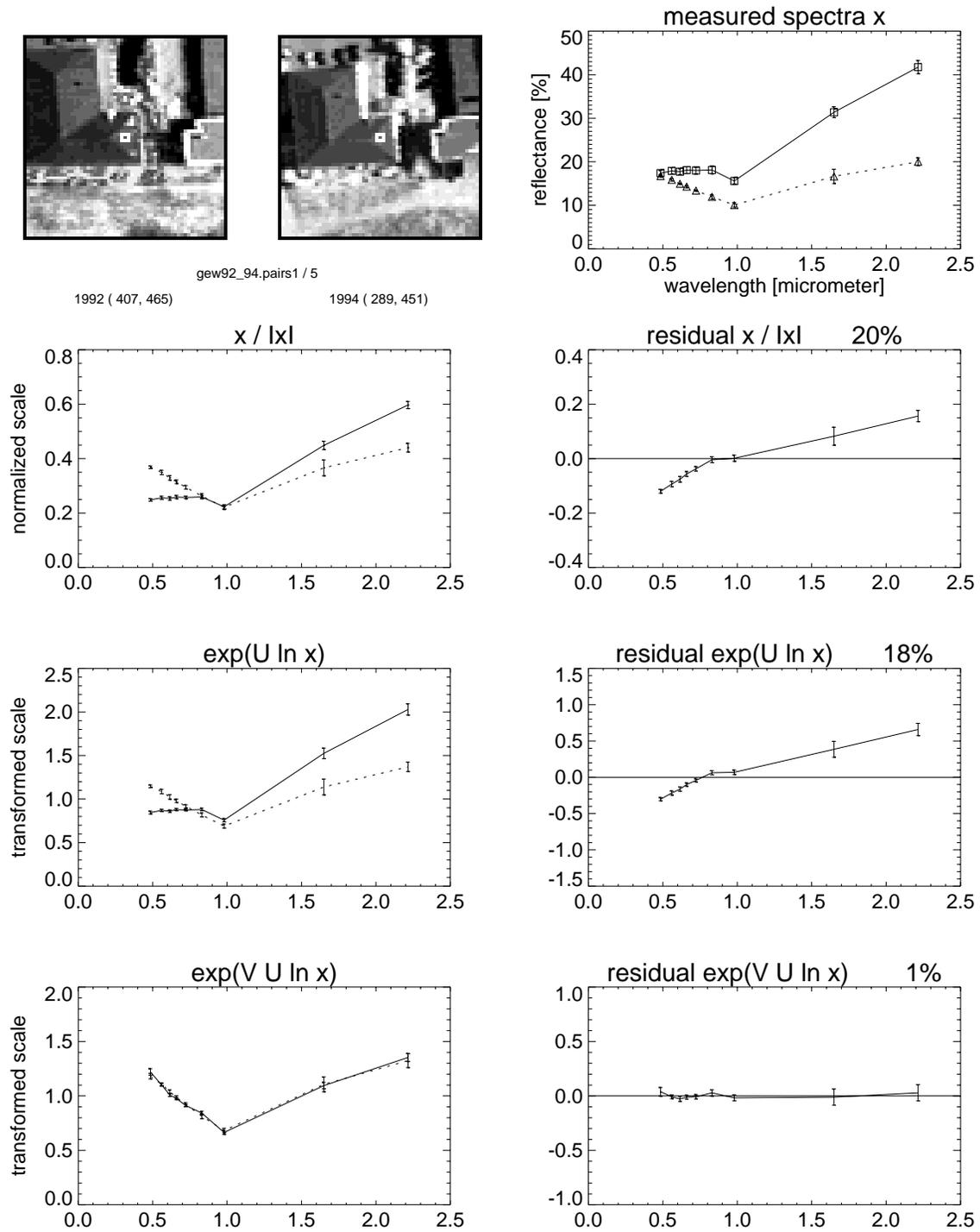


Figure 6.10: Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

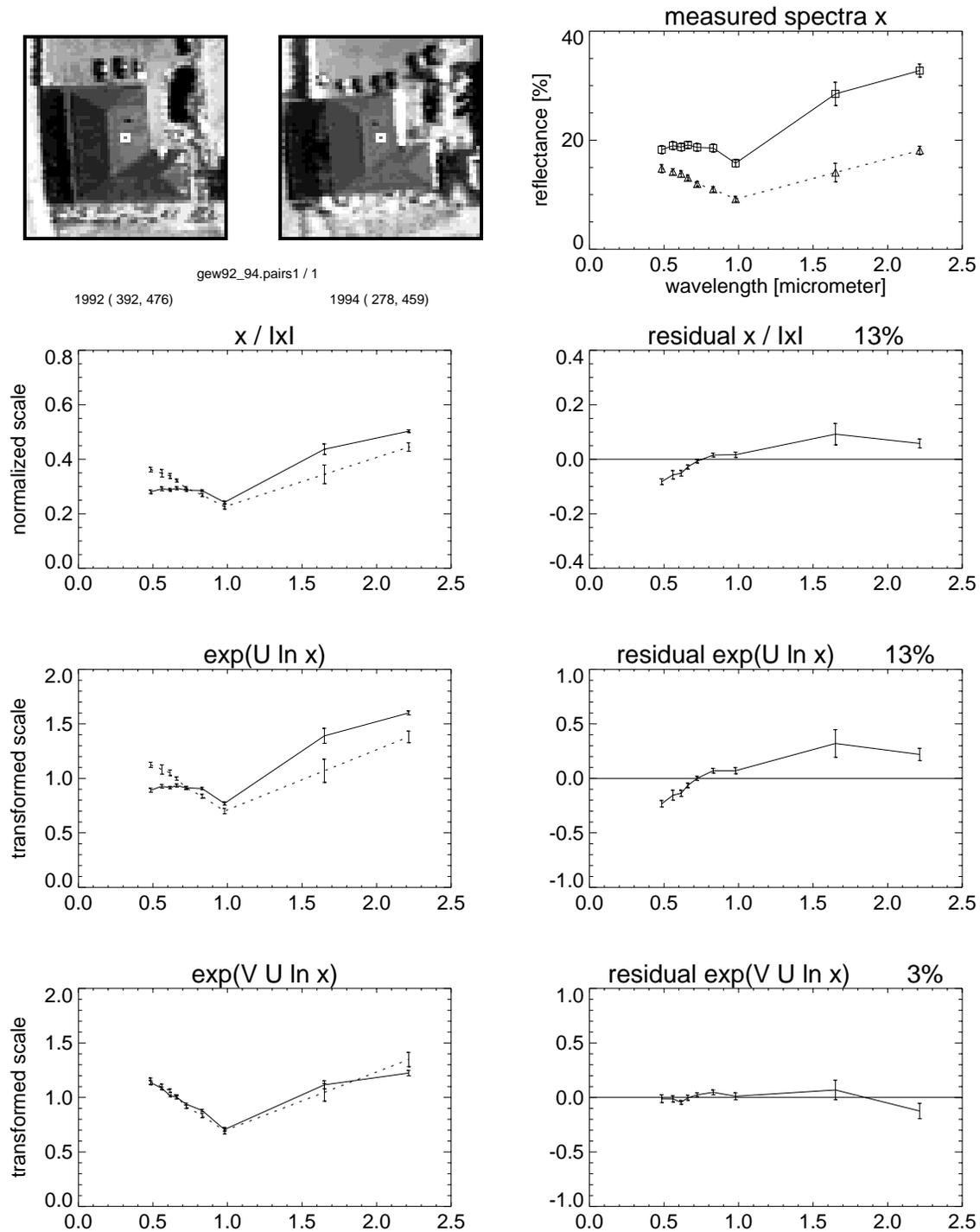


Figure 6.11: Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

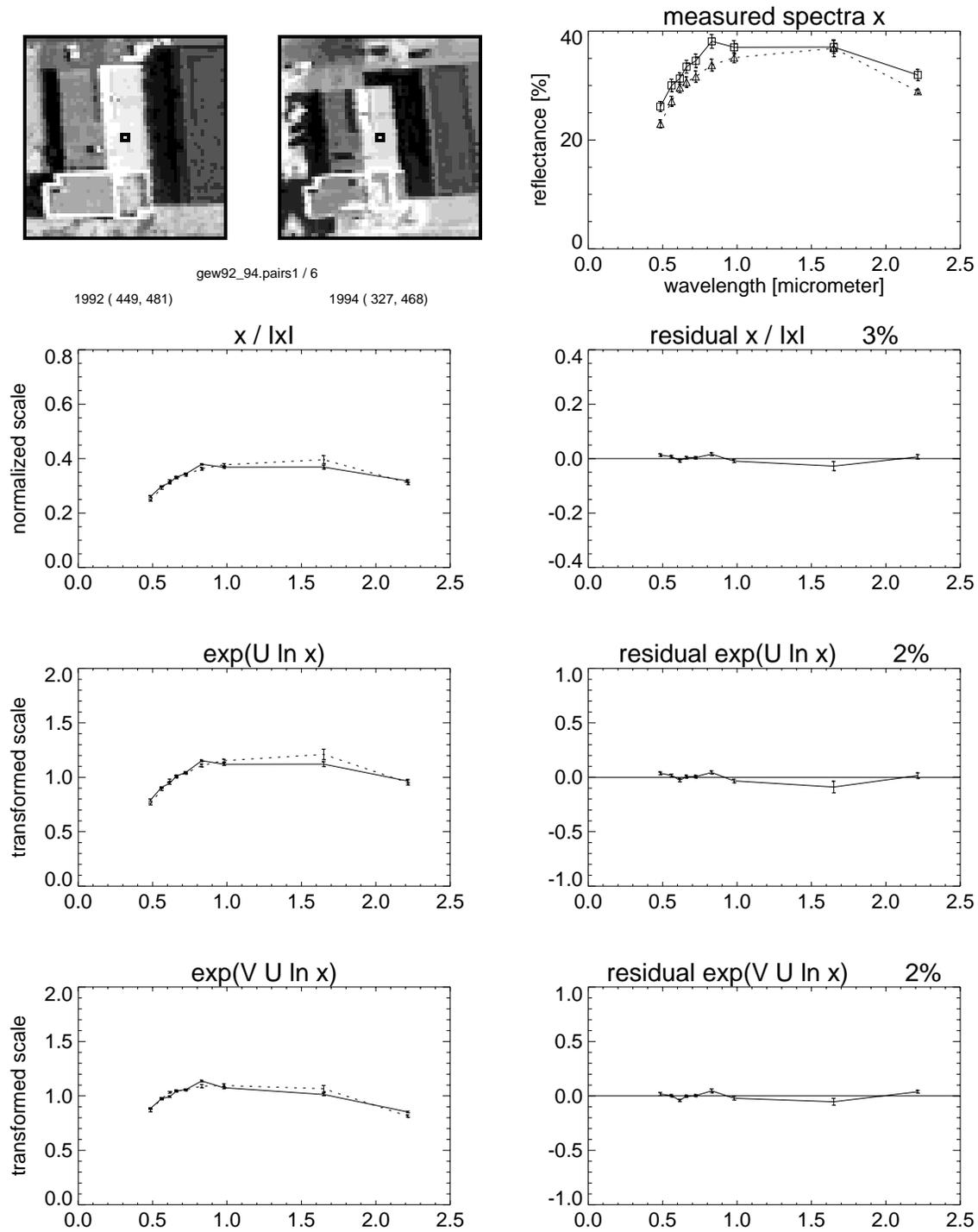


Figure 6.12: Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

6.4 Evaluation by Multispectral Image Classification

In the previous sections of this chapter we have investigated the quantitative effect the suggested transformation has on the spectral distance of the sample pairs. The ultimate aim of the orientation invariant descriptor is improvement of multispectral image processing. In this section we now want to show the qualitative effect of the transformation on multispectral image classification.

The results of any multispectral image classification depends very sensitively on the number of classes, on either the selection of training areas by hand or by application of an unsupervised clustering algorithm, on the employed spectral distance etc. In order to avoid this multitude of parameters, we have chosen the very simple example of classifying the image clips of the sample pairs into binary class images, all pixels of which either belong to the class or do not belong to the class. The underlying idea is, that an operator (or an image processing module) ‘points’ to an area of one of the roof tops chosen earlier as samples, and chooses a training area from the designated roof segment. Then a multispectral class is specified from this training area. The multispectral distance of each image pixel to this class is determined, and the distances are thresholded in order to decide which image pixels belong to the specified roof class and which do not. Ideally, all roof segments, regardless of their surface orientation, should fall into the specified class, while excluding all non-roof pixels. However, the brightness and color shifts induced by the varying surface orientation of the various roof segments will usually compromise the quality of such a classification. The scheme again in short:

-
1. For each feature space: Transform the spectral data (if necessary).
 2. For each sample pair: Train the class on a dark roof segment (*i.e.*, compute the mean spectrum, and the covariance matrix for ML only).
 3. Compute the spectral distance of all image pixels to the class, in the respective feature space.
 4. Set the threshold parameter such that the spectra on the brighter roof segment are just accepted, *i.e.*, on the maximum spectral distance value of all designated roof pixels.
 5. Optionally, multiply the threshold value by a tolerance factor 2.
 6. Classify the image into pixels accepted to belong to the class (distance lower than threshold, indicated as white), and pixels not accepted to belong to the class (distance larger than threshold, indicated as black).
 7. Check by eye appraisal how satisfactory the roof segments were classified as such and other areas of the image clip were not.
-

In practical terms, our test was conducted as follows: For each sample pair, an image clip was cut out of the overall image, centered around the sample pair (top left of Fig. C.4 on page 236). Then the training area for the class ‘roof’ was automatically designated as the 3×3 window around the darker one of the two spectra of each sample pair. We compare six different methods of classification and thresholding:

- ▶ Minimum Euclidean Distance (MED) in the original space.
- ▶ Maximum Likelihood (ML) classification in the original space¹.
- ▶ Minimum Euclidean Distance (MED) in the normalized space ($\mathbf{x}/\|\mathbf{x}\|$).
- ▶ Minimum Euclidean Distance (MED) in the brightness and color filtered space ($\exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$).
- ▶ The latter two classifications again, with the threshold parameter doubled by a tolerance factor 2.

At first, the simple Euclidean spectral distance between all image pixels and the mean spectral vector of the training area is computed (Eq. 2.25 on page 23). Then the threshold distance is fixed as the maximum distance of all pixels within the 3×3 window around of the darker and the brighter sample. In other words, the threshold is fixed as the minimum threshold which guarantees that all the selected 18 image pixels on the darker and brighter roof segment are just within the class. The pixels belonging to the class are shown as white, the non-belonging ones as black (top row, second column).

As a second classification method, we have applied maximum likelihood classification (Richards 1993). There, the class covariance matrix is determined from the spectra of the training area, and the multispectral distance to the mean vector is weighted with the inverse covariance matrix (Mahalanobis distance, Eq. 2.28, and Eq. 2.34 on page 25). Then the threshold is determined as above, such that the selected pixels of the darker and brighter roof segment are just falling into the class, and the image classified into roof and non-roof pixels (bottom row, second column).

For the next classification test, all spectral vectors are normalized. Within the normalized spectral space, the training (which in this case means only computation of the mean vector) is repeated, and the classification is performed with simple Euclidean distance and thresholding as above (top row, third column). In order to test the sensitivity of the classification against the threshold value, the classification is repeated with a doubled threshold value, *i.e.*, the multispectral distances are thresholded with the maximum distance of the selected pixels times a tolerance factor 2 (bottom row, third column).

Finally, the training and classification is performed after the transformation (log brightness and color filtering) of the spectra into the orientation invariant descriptor. Again, we show the results for the minimal threshold (top row, right) and for this threshold times a tolerance factor 2 (bottom row, right).

¹ Note that for this case of just a single class the Minimum Mahalanobis Distance classification is equivalent to Maximum Likelihood classification (Eq. 2.28, and Eq. 2.34 on page 25). The $N \times N = 9 \times 9$ covariance matrix is in our example estimated from the $3 \times 3 = 9$ pixel neighborhood which is numerically just sufficient although certainly underdetermined in practise (*Hughes phenomenon*, Shahshahani & Landgrebe (1994)).

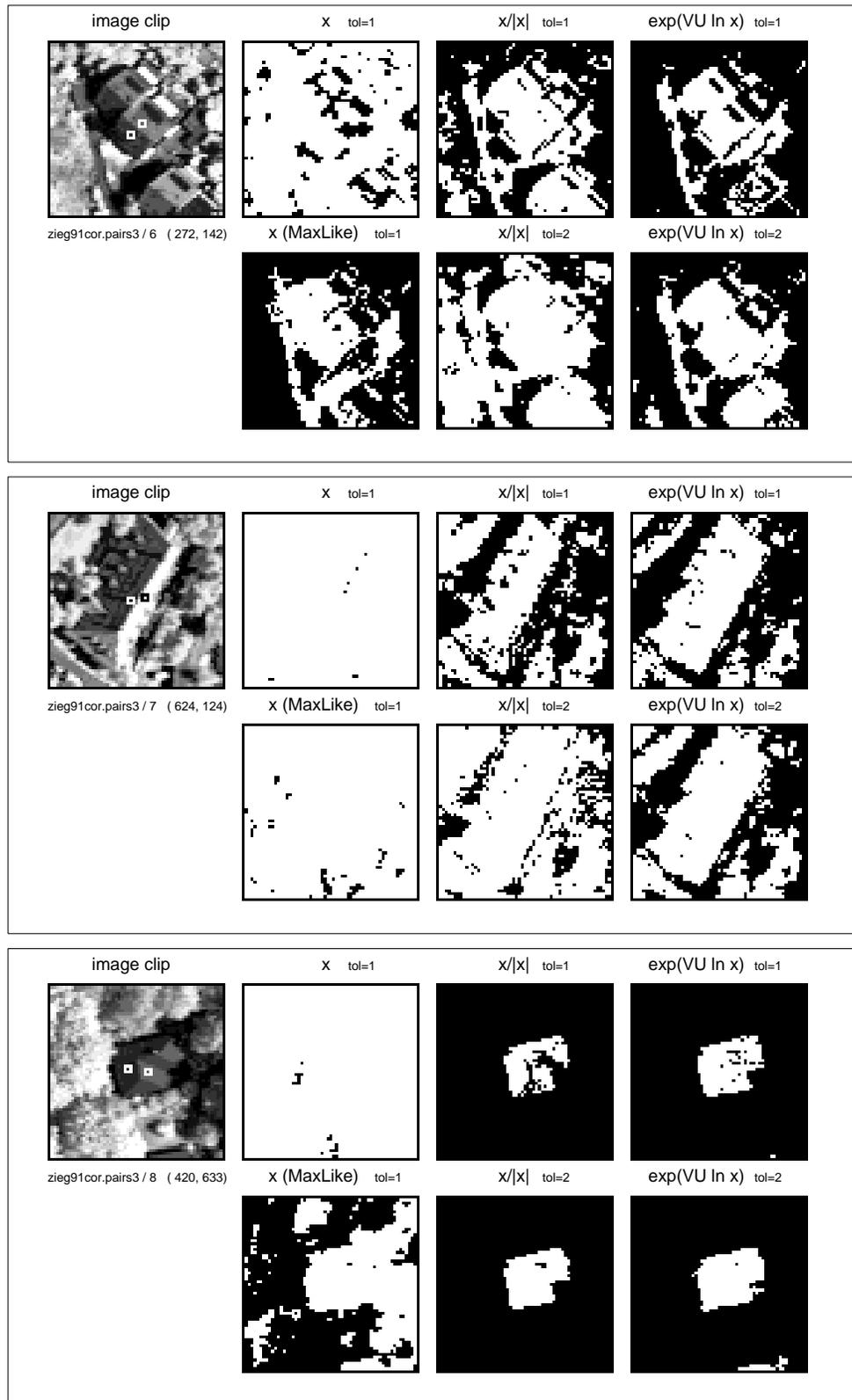


Figure 6.13: Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class.

White: belongs to class; black: belongs not to class.

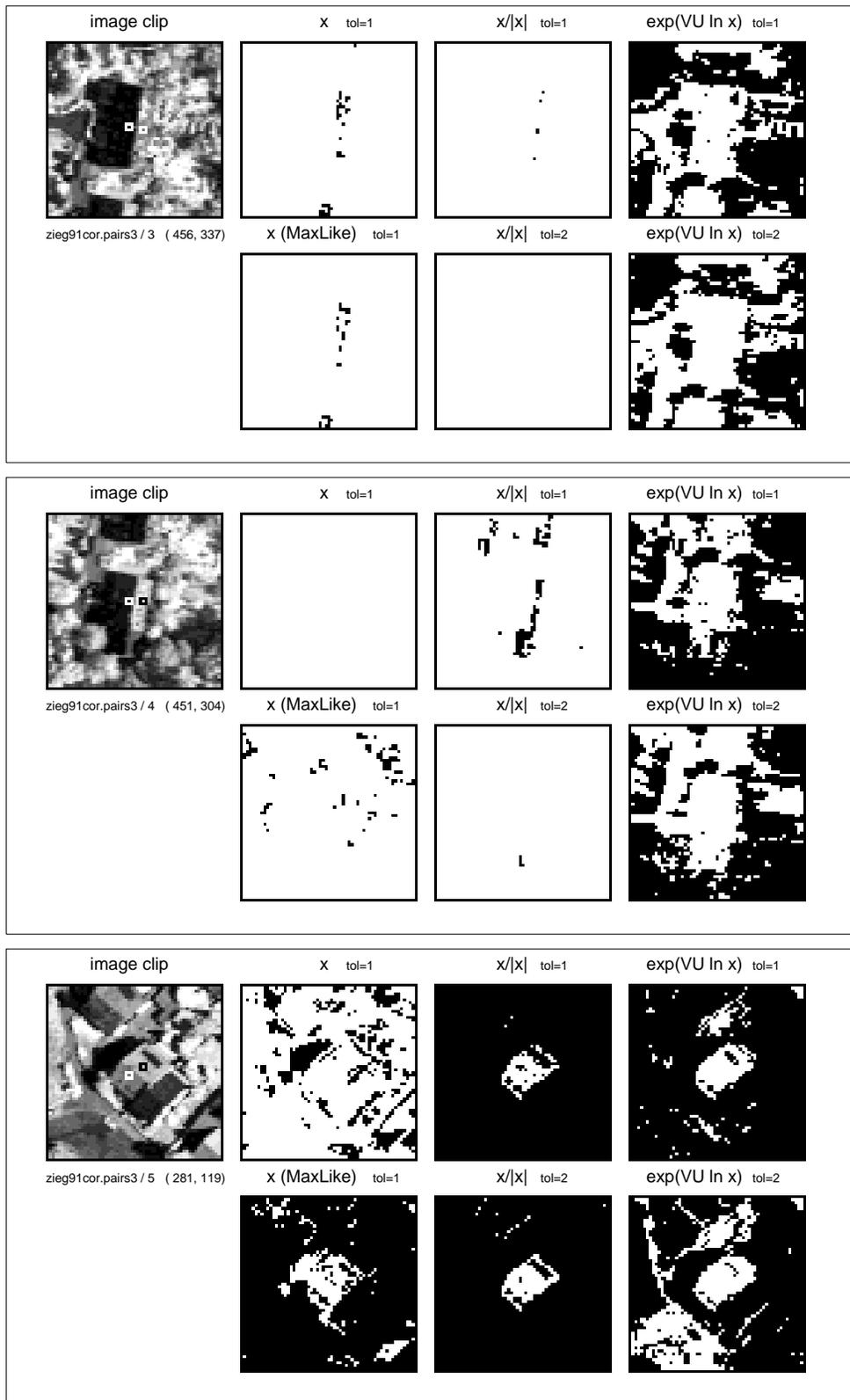


Figure 6.14: Multispectral classification into binary class images (continued).

6.4.1 Qualitative Results of the Classification

From the binary classifications shown in Fig. 6.13 and Fig. 6.14 (more in Section C.4 on page 236 in the appendix) we draw the following observations:

- ▶ The classification with simple Euclidean distance in the original space performs very poorly. The multispectral distance between the pixels of the darker and the brighter roof segment is very large. Demanding that the selected pixels of both roof segments are comprehended by one class requires nearly all other image pixels to fall into the class as well.
- ▶ The maximum likelihood classification, using the Mahalanobis distance determined with the covariance matrix of the darker roof segment pixels, performs often better than Euclidean distance classification. We want to recall that the computation costs are much higher, though, than for Euclidean distance.
- ▶ The Euclidean distance in the normalized space performs better than maximum likelihood classification in some cases. However, it shows to be very sensitive against the threshold value. In a number of cases, we observe poor results when the threshold value is relaxed by a tolerance factor 2.
- ▶ The Euclidean distance in the transformed space performs equally good or best in most cases in discriminating the roof pixels (plus pixels of similar spectra, of course) from others, regardless of surface orientation. Moreover, it is robust against variation in the threshold value, which can mostly be relaxed to double tolerance without severe effect. This indicates that the suggested transformation allows stable thresholding at salient gaps in the histogram of the spectral distances with respect to the specified class.

For the problem of identifying surface materials under varying orientation, we can conclude that after the suggested transformation into the orientation invariant feature space, we can employ a low-cost Euclidean distance classification approach which performs better than costly maximum likelihood classification or classification in the normalized vector space, and which is moreover less sensitive to threshold parameters.

6.5 Unsupervised Classification / Clustering

In the previous section we have performed *supervised* classification in the sense that the training areas were predefined. In this section we select an image clip which then is subjected as a whole to *unsupervised* classification / clustering (Section 2.6 on page 29). Note that this exercise can only be seen as an illustration, since the number of parameters which influence the outcome of unsupervised classification is just too large. In particular, the result of unsupervised classification depends strongly on the image scene contents (*i.e.*, the abundance of each ‘sort’ of spectra in the image), on the number k of spectral classes, on whether the spectral bands were PC-transformed or rescaled, etc.

Except to the original spectral feature space, The spectral data of the image clip shown on top of Fig. 6.15 was mapped to the normalized, the log brightness filtered, and the log brightness and color filtered space.

In order to reduce computation cost, the spectral data was subjected to principal component transformation (PCT, Section 2.7 on page 32) after the mapping to the respective feature space. By virtue of the PCT, the number of spectral bands was reduced from $N = 9$ to $H = 2$. The transformed bands have a vanishing mean value and are rescaled to unit variance.

Then the image clip shown on top of Fig. 6.15 was classified by unsupervised fuzzy k -means (Bezdek 1981) into $k = 5$ classes, as described in Section 2.6 on page 29. The results for the four different feature spaces can be seen in Fig. 6.15 and Fig. 6.16, left-hand side.

Most classification techniques as applied in multispectral remote sensing (Richards 1993) rely on purely *spectral* features and consider only one pixel at a time. More recently, a method for utilizing additional contextual information from neighboring pixels has been derived from Markov random field modeling (Besag 1986). This ‘ICM-algorithm’ has been shown to improve supervised classification results on multispectral imagery (Jhung & Swain 1996, Solberg et al. 1996). So far, this spectral-spatial labeling approach has been used in conjunction with supervised classification only, *i.e.*, the reference classes were established from training data by an analyst. Wiemker (1997b) describes the effects of incorporating spatial context information into *unsupervised* clustering techniques such as the hard and fuzzy k -means algorithms. It could be shown that not only the classification procedure is improved by accounting for local neighborhoods, but that also the clustering process yields different class centers than without using spatial features. The accuracy of the estimated class prototype spectra is significantly increased on simulated test imagery, and more satisfactory segmentation results could be achieved on real remotely sensed multispectral imagery.

The basic underlying idea of using additional context information is that around each pixel \mathbf{x} a local neighborhood $\mathcal{N}(\mathbf{x})$ is considered (here we used a 3×3 pixel window). If the surrounding pixels of a pixel \mathbf{x} do from their spectral appearance belong to the same class as \mathbf{x} , then this encourages the current classification of \mathbf{x} . Vice versa, if the surrounding pixels have low probabilities to belong to the same preliminary class as \mathbf{x} , then also \mathbf{x} ’s probability for this class decreases. This process is iterated until convergence. The results are shown in Fig. 6.15 and Fig. 6.16 on pages 142 and 143, right-hand side.

We discuss the ‘behaviour’ of some exemplary surface patches throughout the classifications in the different feature spaces (marked by numbered circles in Fig. 6.15):

- ▶ The roof parts **(4)** and **(5)** are classified (labeled) into two different classes using the original spectra $[\mathbf{x}]$, due to their opposite tilt. Brightness normalization $[\mathbf{x}/\|\mathbf{x}\|]$ or log brightness filtering $[\exp(\mathbf{U} \ln \mathbf{x})]$ does not help (Fig. 6.16). Only after logarithmic brightness and color filtering $[\exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})]$ is the complete roof labeled into one class.
In contrast, the roof parts **(9)** and **(10)** are always labeled as the same class in all feature spaces.
- ▶ The roof parts **(2)** and **(3)** are classified into one class using the original spectra $[\mathbf{x}]$, in spite of their opposite tilt. However, the roof parts are at least partially split after brightness normalization $[\mathbf{x}/\|\mathbf{x}\|]$ or log brightness filtering $[\exp(\mathbf{U} \ln \mathbf{x})]$. In the latter, it can be classified into one class again when using spatial context features. Without using context, only after log brightness and color filtering $[\exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})]$ is the complete roof labeled into one class again.
- ▶ The roof parts **(6)** and **(7)** are labeled into two different classes using the original spectra $[\mathbf{x}]$, due to their opposite tilt. After brightness normalization $[\mathbf{x}/\|\mathbf{x}\|]$ or log brightness filtering $[\exp(\mathbf{U} \ln \mathbf{x})]$ they resolve into one class, but only in the non-context clustering mode. After log brightness and color filtering $[\exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})]$ they are labeled as one class both with or without spatial context features.
- ▶ The shadowed street **(1)** is misclassified into the same class as the dark roof parts **(2)** and **(3)** using the original spectra $[\mathbf{x}]$. However, **(1)** is correctly distinguished from the adjacent roof top in all other spaces (Fig. 6.16).
In contrast, the shadowed yard **(8)** is never merged into one object with the neighboring roof tops **(6)** to **(10)**.

Using spatial context features yields overall more satisfactory results for thematic classification. It reduces speckle noise considerably. On the other hand, also real objects which are of smaller size than the used local neighborhood $\mathcal{N}(\mathbf{x})$ around a pixel \mathbf{x} may easily be suppressed and will remain undetected.

In summary, the examples of unsupervised classification show that the additional color filtering successfully removes color shifts which otherwise prevent identical surface patches of different orientation from being classified into the same cluster in the spectral feature space. On the other hand, the examples show also that the ‘behaviour’ of specific objects in various unsupervised clustering runs is highly complex, depending *e.g.* on the image scene content and the number of classes. The success can only be evaluated with respect to the usefulness for specific applications.

image clip from scene 'Gewerbegebiet 1992'
 (169 × 154 pixels, at $\lambda = 1650$ nm, contrast enhanced)



original spectra [x]

with context

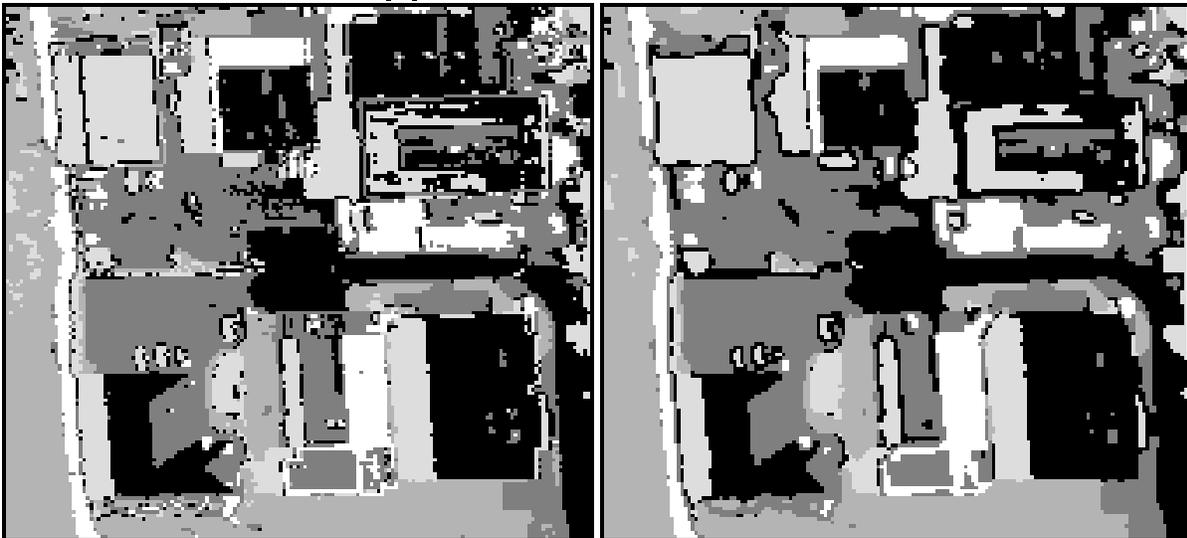


Figure 6.15: Unsupervised multispectral classification ($k = 5$ classes), using either simple pixel-wise classification (left-hand), or additional local neighborhood features (3×3 pixels, right-hand).

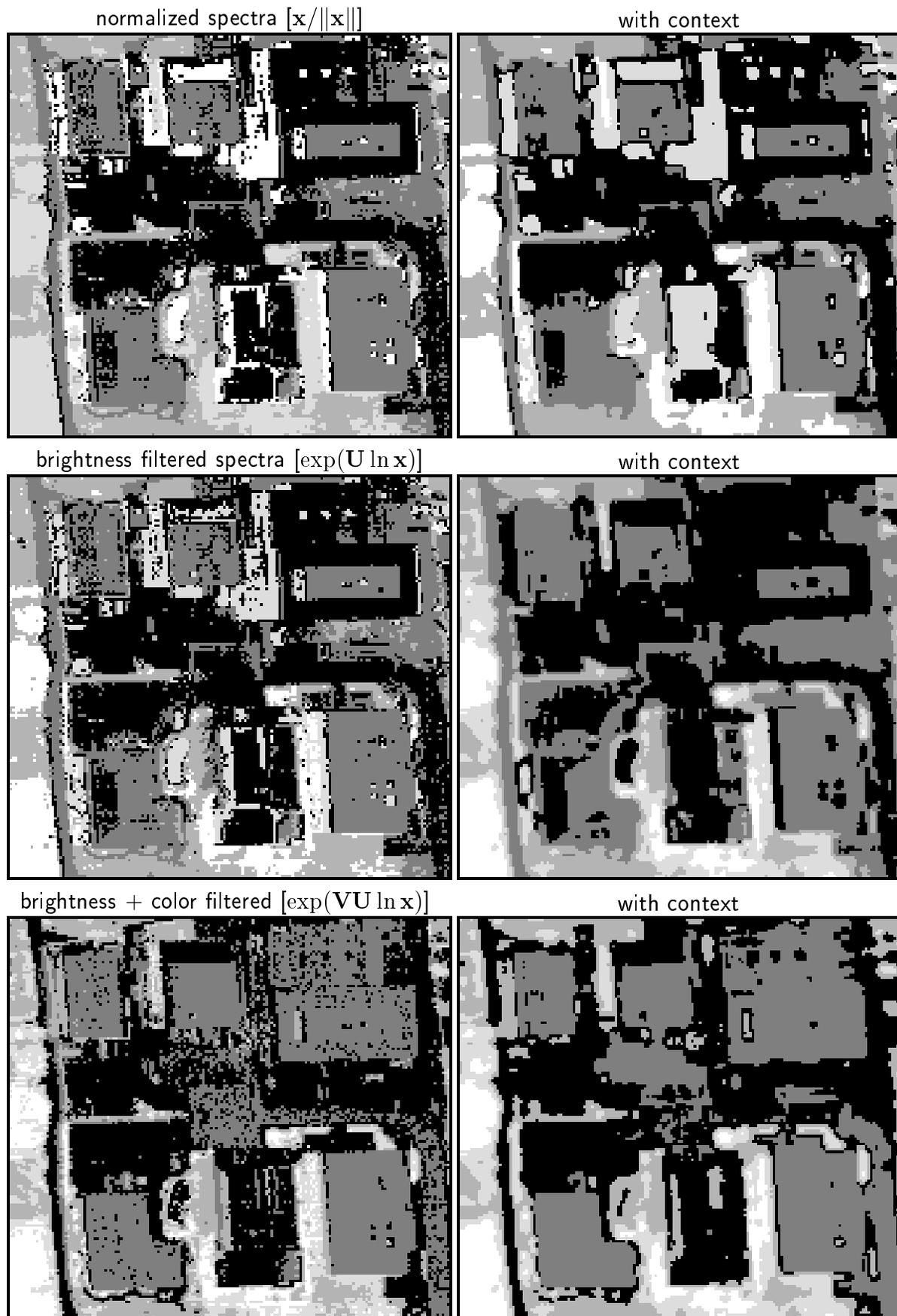


Figure 6.16: Unsupervised multispectral classification ($k = 5$ classes), using either simple pixel-wise classification (left-hand), or additional local neighborhood features (3×3 pixels, right-hand).

6.6 Analysis of the Transform in the Feature Space

In order to visualize the overall effect of the suggested transform, we appraise the distribution of the spectra in the feature space. We recall that each spectrum is a vector which consists of N entries and can be considered as a point in the N dimensional feature space. For inspection of the here $N = 9$ dimensional feature space we employ a principal component transform (see Section 2.7 on page 32): The mean vector and the covariance matrix of the data is determined, and its eigenvalues and eigenvectors are computed; then the mean spectrum is subtracted from all spectra, the spectra are projected onto the principal spectra, *i.e.*, the eigenvectors, and finally the variance of each principal component is scaled to unity by division of all principal components by the square root of the respective eigenvalue. The result is data with a mean vector of zero and the identity matrix as the covariance matrix.

For visualization of the feature space of scene ‘Ziegelstein 1991’ (Fig. 6.1 on page 119), all spectra are transformed to the two most significant principal components. In this example, two principal components capture between 93% and 97% of the total variance (depending on the feature space). The two vector entries of each such transformed spectrum are then entered into a bivariate histogram. Contour plots of these histograms are shown in Fig. 6.17 on page 146.

In the full $N = 9$ dimensional feature space we have run a fuzzy k -means unsupervised clustering algorithm (see Section 2.6 on page 29 and Bezdek (1981)) for $k = 2$ clusters. After convergence of the iterative clustering, we obtain the cluster centers \mathbf{m} and \mathbf{m}' for the two clusters and the respective covariance matrices Σ and Σ' . Then the cluster separation can be described as the length of the distance vector between the centers $\mathbf{d} = (\mathbf{m} - \mathbf{m}')$ relative to the root mean square extension of the clusters $\sqrt{(\text{tr } \Sigma + \text{tr } \Sigma')/2}$ (compare Section 2.4 on page 25):

$$\sqrt{\frac{(\mathbf{m} - \mathbf{m}')^T (\mathbf{m} - \mathbf{m}')}{(\text{tr } \Sigma + \text{tr } \Sigma')/2}} \quad (6.3)$$

This normalization of the separation does not take into account the shape of the cluster ellipsoids, since the basic k -means clustering algorithms only allow search for hyperspheres in the data. The separation for four different feature spaces is given in Table 6.2 on the facing page. In Fig. 6.17 on page 146 we have marked the fuzzy k -means-found cluster centers and plotted the cluster radii $\sqrt{\text{tr } \Sigma}$ in order to illustrate the compactness of the clusters. (Note that the cluster means, which minimize the root mean square within-cluster-deviation, are not necessarily identical to the histogram peaks.) The decision surfaces for Minimum Euclidean Distance classification are halfway between the cluster centers perpendicular to the separation vector.

In this example the two most prominent clusters are formed by vegetation and non-vegetation pixels which are of nearly equal abundance in the image. By eye appraisal (Fig. 6.17) we can clearly tell apart these two clusters in all four spaces, however, they are unequally easy to be discriminated by unsupervised clustering as the cluster ellipsoids in the original data are far from being spherical.

We observe that already the transition into the logarithmic feature space improves the aptitude for Minimum Euclidean Distance (MED) classification. The brightness filtered space is even better, but still exceeded by the brightness and color filtered

NORMALIZED SEPARATION IN DIFFERENT FEATURE SPACES

Space	\mathbf{x}	$\ln \mathbf{x}$	$e^{\mathbf{U} \ln \mathbf{x}}$	$e^{\mathbf{VU} \ln \mathbf{x}}$
$\sqrt{\frac{(\mathbf{m}-\mathbf{m}')^T(\mathbf{m}-\mathbf{m}')}{(\text{tr} \mathbf{\Sigma} + \text{tr} \mathbf{\Sigma}')/2}}$	1.781	2.338	2.506	2.859

Table 6.2: Cluster separation in the original, logarithmic, brightness filtered and brightness/color filtered feature spaces, normalized by the mean cluster extension.

space. The visual impression is supported by the relative cluster separation values in Table 6.2.

We want to look more closely at the final improvement achieved by the color filtering. To this aim we consider the principal component projection of the log feature space after brightness filtering ($\mathbf{U} \ln \mathbf{x}$, Fig. 6.18 on page 147). Within the bivariate histogram, all local maxima are found, *i.e.*, all bins with entries higher than all their 8 nearest neighbors. Then the highest 6 of these local maxima are marked (crosses). We notice that both in the vegetation (righthand) and the non-vegetation cluster (lefthand) the histogram peaks are arranged in a particular direction. For comparison, the brightness normalized vector $\mathbf{v} = \mathbf{U}\mathbf{m}$ (see Eq. 4.99 on page 65) with $\mathbf{m} = \boldsymbol{\lambda}^{-2}$ is plotted as the solid line. A color filtering with the projector \mathbf{V} will extract all variation along the plotted vector \mathbf{v} . The visualization of the feature space suggests that in this way within-cluster variance is reduced without diminishing the between-cluster separation, and thus explains the better cluster separability after color filtering.

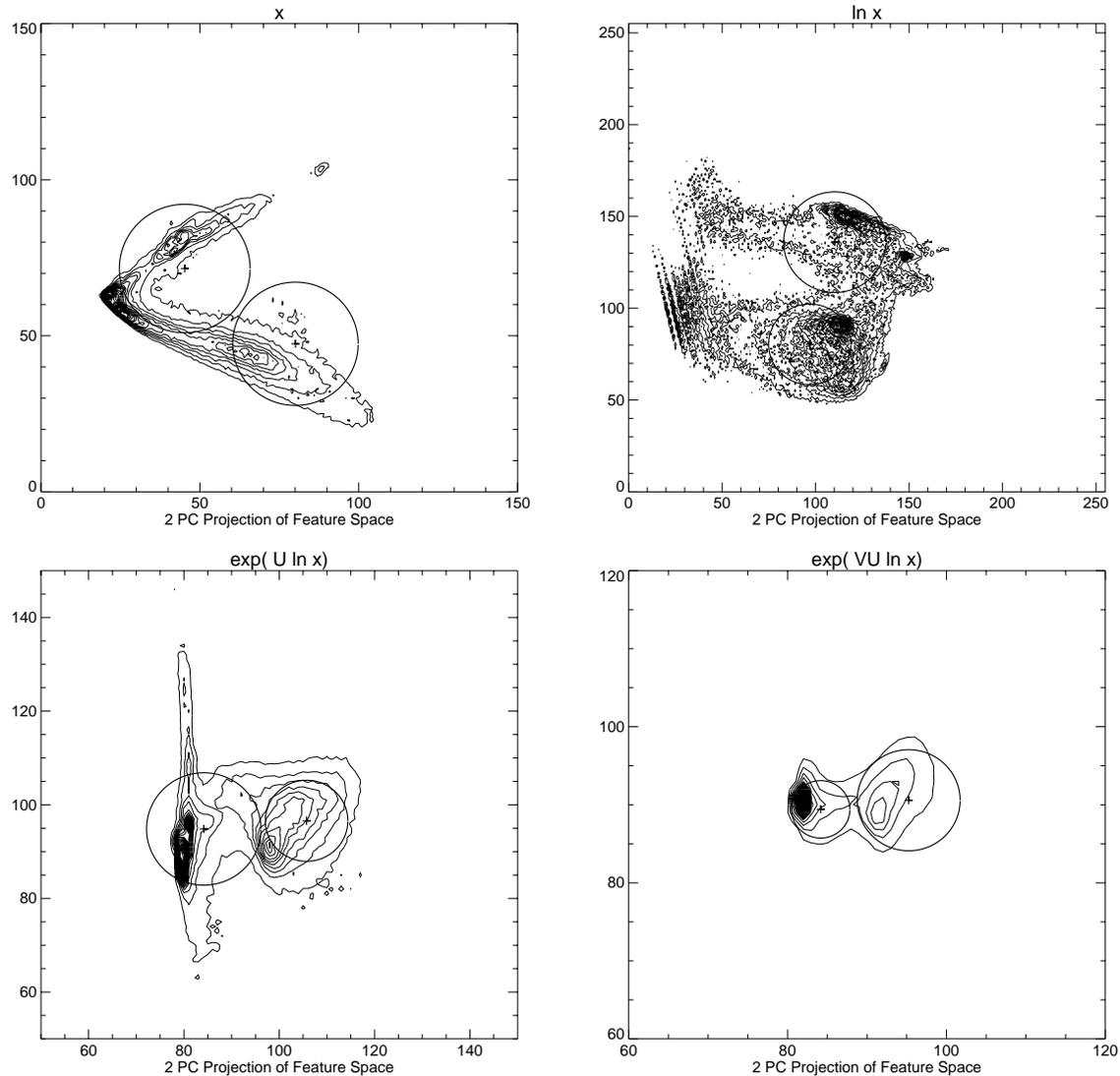


Figure 6.17: Iso-lines (contours) of the bivariate histogram of the principal component transformed feature space (Scene ‘Ziegelstein 1991’ of Fig. 6.1 on page 119).

top left: Original feature space (\mathbf{x}); right: Log feature space ($\ln \mathbf{x}$),
 bottom left: After brightness filtering ($\exp(\mathbf{U} \ln \mathbf{x})$); right: After brightness and color filtering ($\exp(\mathbf{VU} \ln \mathbf{x})$).

The crosses and circles are the cluster centers \mathbf{m} and scatter radii ($\text{tr } \Sigma$)^{1/2} of the clusters found by unsupervised fuzzy k -means clustering ($k = 2$). (The axis values are not significant, as the principal-component-projected data was scaled and translated for visualization.)

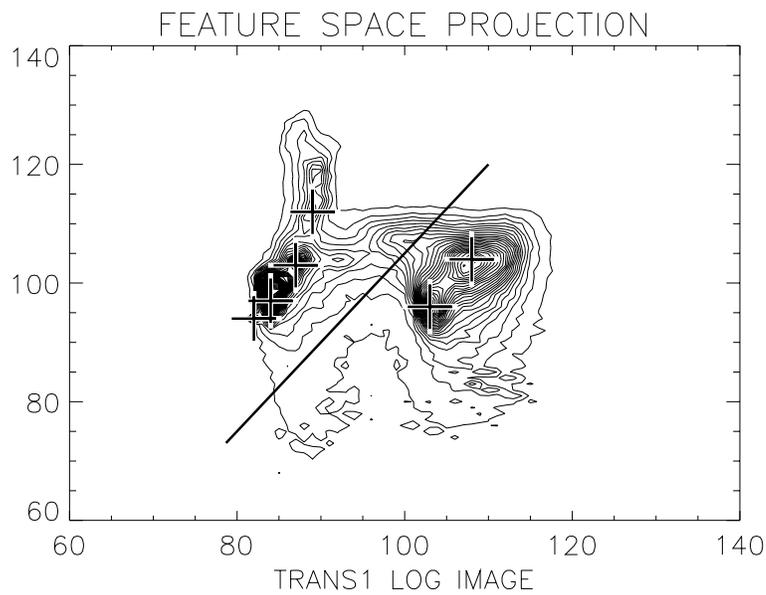


Figure 6.18: Iso-lines (contours) of the bivariate histogram of the principal component transformed feature space after log brightness filtering $\mathbf{U} \ln \mathbf{x}$ (Scene of Fig. 6.1 on page 119). The highest histogram peaks are marked by crosses, right hand is the vegetation cluster. Also plotted is the PC-transformed direction of the color shift \mathbf{v} , i.e., the direction of the variance which is filtered by the projector \mathbf{V} .

6.7 Statistical Evaluation of Spectral Sample Sets

In the previous sections of this chapter we have compared spectral signatures, either from corresponding surfaces of different orientation, or from the same inclined surfaces under different solar illumination angle. We have analyzed the difference between the observed spectra which can be separated into a large brightness shift and a small color shift.

Our aim is to find a transformed feature space in which the spectral distance between spectra which differ only in orientation but not in reflectance vanishes or, at least, becomes small. Inspection of our spectral samples has shown that the spectral distance between the two spectra of each sample pair can indeed be significantly reduced by application of the suggested logarithmic brightness and color filtering.

However, measuring the reduction of the spectral distance only is not sufficient. After all, we have to consider the possibility that *all* spectral distances are reduced by the brightness and color filtering.² Hence, a statistical evaluation has to show that the spectral distances have been reduced relative to the overall extension of all spectra in the feature space (see Section 2.4 on page 25).

Therefore, we have formed sets of spectral sample pairs which exhibit the aspect of the color constancy problem as discussed here, namely, which exhibit equal surface materials under varying surface orientation. We have selected a number of pairs of spectra where the reflecting material remains the same but surface orientation and thus the illumination is different (Fig. 6.20 on page 153). Another set also comprises additional spectral sample where we can be reasonably confident that both surface patches are horizontal and the change of appearance is caused the transition between sunlit and shadowed area (Fig. C.2 on page 214).

For our comparison, we consider the spectral information of surface pixels in seven basic feature spaces as depicted in Table 6.3 on the next page: The original pseudo-reflectance feature space (\mathbf{x}), as conventionally normalized under L_2 -norm ($\mathbf{x}/\|\mathbf{x}\|$), the logarithmic space ($\ln \mathbf{x}$), then filtered for brightness shift (\mathbf{U}) and color shift (\mathbf{V}), and exponentiated back into the non-logarithmic domain.

In order to decide which of the various possible feature spaces is most suitable for the here discussed problem, eye appraisal of the transformed multispectral imagery is certainly not a satisfactory option, as it cannot capture the multispectral nature of the data.

Instead, using the sets of spectral sample pairs, we will evaluate their separability, *i.e.* their relative spectral distances in the various feature spaces. Again, we aim at vanishing spectral distances. In other words, the separability between the two spectra of each sample pair should become low. The fact that each set comprises a number of samples allows to estimate the error of the mean separability. Thus we can clarify whether the separabilities in the different feature spaces differ substantially, *i.e.*, whether their differences are above a statistical significance threshold.

In the following sections we will first explain how we are going to measure and compare the separabilities in different feature spaces in a meaningful way, and then discuss results in Section 6.7.3 on page 155.

² Consider *e.g.* the trivial case of all spectra mapped to zero: $\mathbf{x} \mapsto \mathbf{0}$, which certainly reduces undesired spectral distances but is useless, of course.

FEATURE SPACES		
\mathbf{x}	$\ln \mathbf{x}$	–
$\mathbf{x}/\ \mathbf{x}\ $	$\mathbf{U} \ln \mathbf{x}$	$e^{\mathbf{U} \ln \mathbf{x}}$
–	$\mathbf{V}\mathbf{U} \ln \mathbf{x}$	$e^{\mathbf{V}\mathbf{U} \ln \mathbf{x}}$

Table 6.3: Investigated feature spaces.

6.7.1 Evaluation of Spectral Distances

■ Normalizing the Spectral Distances

The characteristic property of the spectral distances between the sampled pairs is that they are caused not by a change in the reflecting surface material but rather by a change in surface orientation and/or illumination. An invariant should thus reduce such distances to zero, and a transformed feature space should diminish such distances. On the other hand, a feature space transformation will also change the overall density distribution. Hence we aim to find the lowest normalized spectral distance, ideally a vanishing spectral distance between pairs but conserved finite extension of the overall image data cluster in the feature space.

As for the normalization of the spectral distances, we consider two normalized distance measures (compare Section 2.4 on page 25):

$$d_T = \sqrt{\frac{\mathbf{d}^T \mathbf{d}}{\text{tr } \Sigma}} = \sqrt{\frac{\text{tr}(\mathbf{d}\mathbf{d}^T)}{\text{tr } \Sigma}} \quad (6.4)$$

$$d_I = \sqrt{\frac{\mathbf{d}^T \Sigma^{-1} \mathbf{d}}{\text{rk } \Sigma}} = \sqrt{\frac{\text{tr}(\Sigma^{-1} \mathbf{d}\mathbf{d}^T)}{\text{rk } \Sigma}} \quad (6.5)$$

where $\text{rk } \Sigma$ denotes the rank of matrix Σ .

The first normalized distance d_T simply puts the Euclidean spectral distance $\sqrt{\mathbf{d}^T \mathbf{d}}$ in relation to the scatter radius $\sqrt{\text{tr } \Sigma}$ of the overall image cluster with the covariance ellipsoid Σ . This distance does not regard the orientation of the spectral distance relative to the overall cluster in the feature space (for illustration see Fig. 6.19 on the following page).

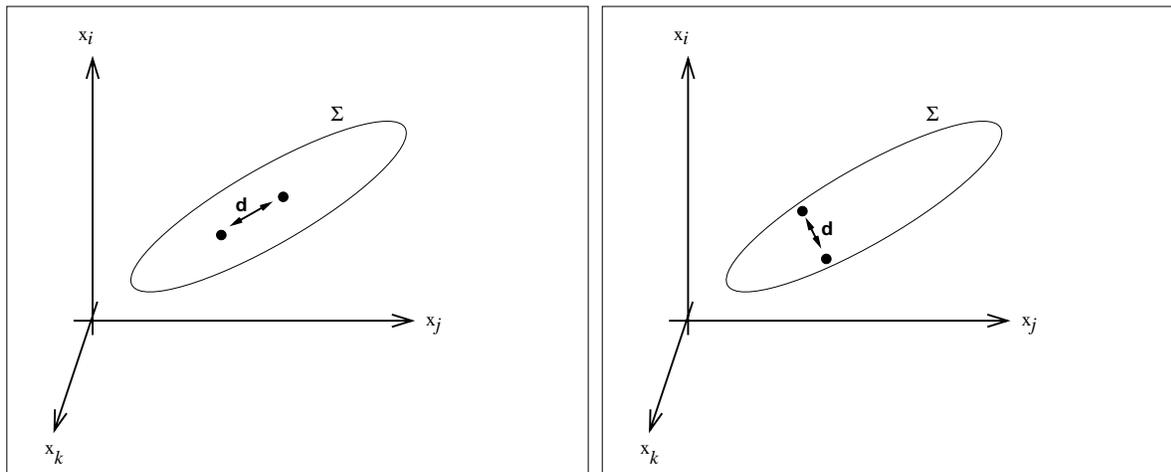


Figure 6.19: Illustration of the orientation of a spectral distance \mathbf{d} (between a pair of spectra) relative to the ellipsoid Σ of the overall image cluster in the feature space.

The second normalized distance d_I is sensitive to the relative orientations of spectral distance and overall cluster. By weighting with the inverse of the covariance Σ^{-1} , the distance \mathbf{d} is essentially transformed into a space of vanishing correlation and unit variance before evaluation of the vector magnitude.

In order to enable direct comparison between the two normalized distance measures d_T and d_I , the additional division by the rank $\text{rk } \Sigma$ of the overall covariance matrix Σ is necessary. We demand equality for the limiting case of a spherical cluster with $\Sigma = \epsilon^2 \mathbf{I}$. In this case, both expressions degenerate to

$$d_T = d_I = \sqrt{\mathbf{d}^T \mathbf{d} / (\epsilon^2 N)} \quad (6.6)$$

(because $\text{tr } \mathbf{I} = \text{rk } \mathbf{I} = N$).

If any filtering of K components with a projector \mathbf{P}_K is applied, this reduces the rank of the overall data from originally N down to $\text{rk } \Sigma = N - K$, and also $\text{tr}(\mathbf{P}_K \mathbf{I} \mathbf{P}_K^T) = N - K$, so for the spherical case $\Sigma = \epsilon^2 \mathbf{I}$ this yields equal distances

$$d_T = d_I = \sqrt{\mathbf{d}^T \mathbf{P}_K \mathbf{d} / (\epsilon^2 (N - K))} \quad . \quad (6.7)$$

The distance d_T is derived from the simple Euclidean distance $\sqrt{\mathbf{d}^T \mathbf{d}}$ which is the one naturally employed for most applications. For some pattern recognition applications, however, the computationally more expensive weighted distance $\sqrt{\mathbf{d}^T \Sigma^{-1} \mathbf{d}}$ yields useful results and was therefore considered here. Note that after filtering of components, the inverse Σ^{-1} can only be computed as the pseudo-inverse Σ^+ because Σ is no longer of full rank N and thus singular.

■ Dependence on Rescaling and Decorrelation

Any additional offsets in the data will not change spectral distances and cluster centers found by unsupervised clustering. However, band-wise scaling does change spectral distances and is well known to influence the outcome of all unsupervised clustering techniques which rest on inner-product-norms with fixed metrics (Hartigan 1975,

Bezdek 1981). Rescaling the input data to standard variance is a natural approach to account for grossly different variances in the spectral bands, or to eliminate influences of sensor gain settings or changing recording conditions such as solar irradiance and atmospheric transmissivity. Therefore we have investigated the spectral distances also in rescaled and transformed spaces. The transformation of the N -valued spectral data to $H \leq N$ uncorrelated principal components of unit variance is an often used tool which reduces the rank of the data to $\text{rk } \Sigma = H$.

6.7.2 Comparison in Different Feature Spaces

We now describe the computation of the relative separability between the two spectra of each spectral sample pair. The separability is measured as the mean value of the spectral distance d for each pair. The lowest separability, *i.e.*, the lowest mean spectral distance, designates the ‘most invariant’ feature space.

The results are shown in Table 6.4 on page 154. The spectra are taken from the pixel position as shown in Fig. 6.20 on page 153. In order to smooth noise, the spectral values are taken as the mean of the 3×3 -neighborhood around the center pixel. At first, all table entries were computed separately for each spectral pair. The final table then gives the mean distance value d of all j sample pairs for the respective entry. Below (denoted ‘ \pm ’), the proper deviation – needed for statistical comparison – is given: $s = \sigma/\sqrt{j}$, the standard deviation σ divided by the square root of the number of samples j .

In order to compare two table entries, *i.e.*, two sample means d and d' , Gosset’s so-called Student’s distribution (Brandt 1992, Lozán 1992) must be used (the ‘ t -test’):

$$t = \frac{\|d - d'\|}{\sqrt{\frac{\sigma^2 + \sigma'^2}{j}}} \quad (6.8)$$

$$= \frac{\|d - d'\|}{\sqrt{s^2 + s'^2}} \quad ; \quad \text{with } s = \frac{\sigma}{\sqrt{j}} \quad . \quad (6.9)$$

The difference of means $\|d - d'\|$ is considered significant when $t > t(\alpha, 2j - 1)$ which is to be taken from the Student distribution tables, where α is the required confidence level and $2j - 1$ the number of degrees of freedom, *e.g.*, $t(\alpha = 90\%, j = 20) = t(90\%, 41) = 1.30$. For this exemplary choice, a difference is considered significant, if and only if

$$\|d - d'\| > 1.3 \sqrt{\sigma^2 + \sigma'^2} \quad (6.10)$$

■ Separability with respect to the Basic Feature Spaces

For each column of Table 6.4 on page 154, the spectral image data was transformed into one of the seven investigated basic feature spaces (Fig. 6.3 on page 149). The first row gives the results for the distance d_I , the second row for the distance d_T .

The third row gives the results for the case that an additional band-wise rescaling to unit variance is applied after the transformation, *i.e.*, the diagonal of the covariance matrix becomes $\Sigma_{ii} = \sigma^2(x_i) = 1$.

The fourth row and the following rows give results for the case that an additional principal component transformation (PCT) was applied after the transformation. We varied the rank H of the $H \times N$ PCT-matrix \mathbf{T} (see Section 2.7 on page 32), *i.e.*, the

number of vector entries and thus the dimension to which the original data is reduced. Each transformed band has a zero mean and unit variance. The PCT-matrix \mathbf{T} is derived from the eigenvectors of the covariance matrix of the complete image data only after transformation into the 7 basic feature spaces.

For the distance d_I we have not tested band-wise rescaling and principal component transformation. The rescaling with a transformation matrix \mathbf{T} (Section 2.7 on page 32) can certainly not change the distance d_I , because the distance $d_I^{2'}$ in the transformed system can be obtained from the original distance d_I^2 (Eq. 6.5) as:

$$d_I^{2'} = (\mathbf{T}\mathbf{d})^T(\mathbf{T}\Sigma\mathbf{T}^T)^{-1}\mathbf{T}\mathbf{d}/\text{rk } \Sigma \quad (6.11)$$

$$= \mathbf{d}^T\mathbf{T}^T(\mathbf{T}^T)^{-1}\Sigma^{-1}\mathbf{T}^{-1}\mathbf{T}\mathbf{d}/\text{rk } \Sigma \quad (6.12)$$

$$= \mathbf{d}^T(\mathbf{T}^{-1}\mathbf{T})^T\Sigma^{-1}\mathbf{T}^{-1}\mathbf{T}\mathbf{d}/\text{rk } \Sigma \quad (6.13)$$

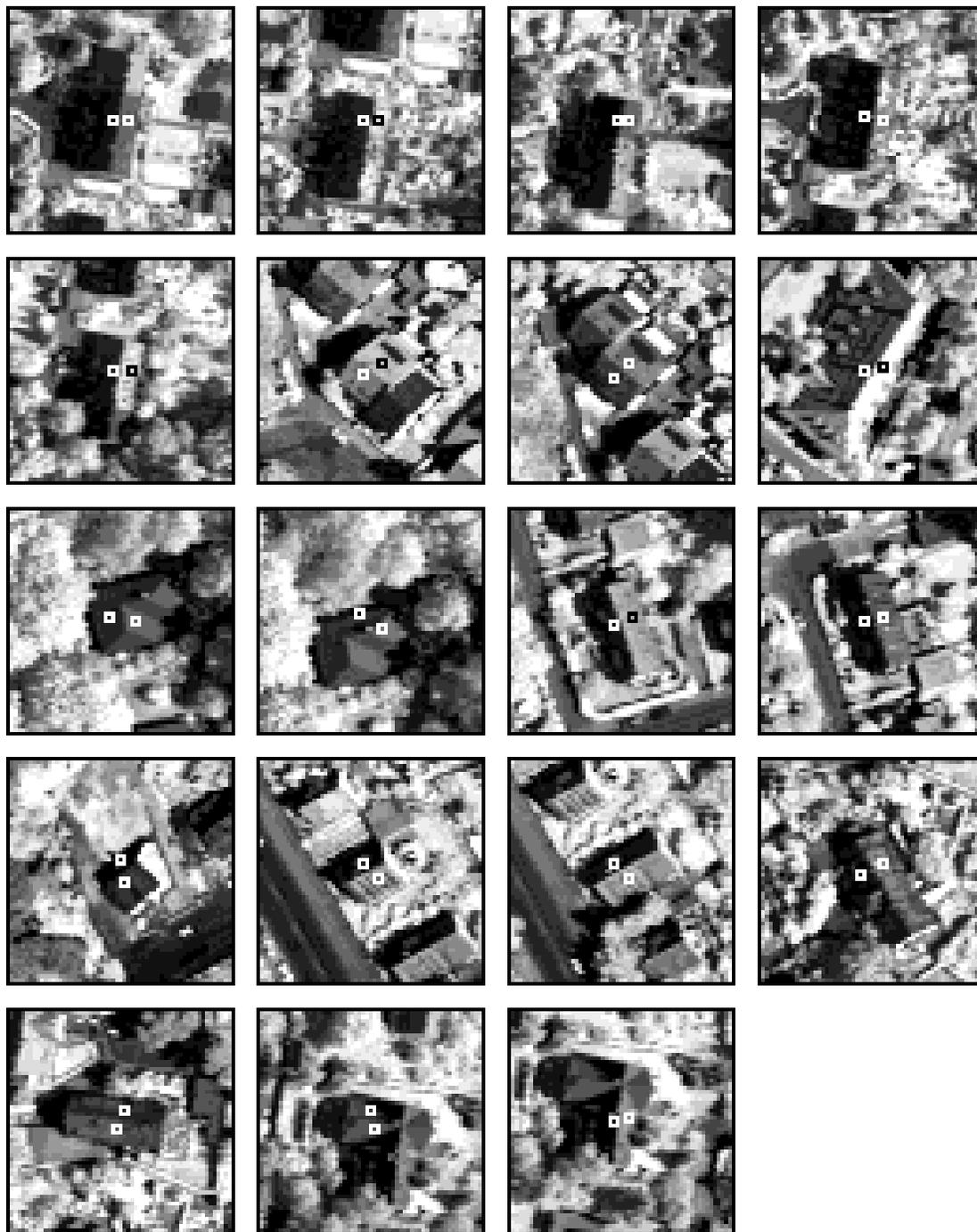
with $\mathbf{T}^{-1}\mathbf{T} = \mathbf{T}\mathbf{T}^{-1} = \mathbf{I}$

$$= \mathbf{d}^T\Sigma^{-1}\mathbf{d}/\text{rk } \Sigma \quad (6.14)$$

$$= d_I^2 \quad (6.15)$$

and thus remains unchanged.

Also, we do not need to consider PCT for the distance measure d_I , because by the weighting with the inverse of the overall cluster covariance Σ , this distance is independent of correlation and variance anyway. As a consequence, the distance d_I must yield equal results as a full rank PCT to $H = N$ principal components with distance measure d_T (compare first and last rows in Table 6.4). There can be no more meaningful principal component bands than the rank of the data to be transformed. After filtering with the orthogonal projector \mathbf{P}_K the dimension will be reduced to $\text{rk } \Sigma = N - K$ and thus the number of meaningful principal components H is limited.



zieg91cor.pairs3

Figure 6.20: Selected samples of surface pairs with varying orientation (set 1) from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119, 300 m altitude).

SEPARABILITY IN DIFFERENT FEATURE SPACES							
	\mathbf{x}	$\mathbf{x}/\ \mathbf{x}\ $	$\ln \mathbf{x}$	$\mathbf{U} \ln \mathbf{x}$	$\mathbf{VU} \ln \mathbf{x}$	$e^{\mathbf{U} \ln \mathbf{x}}$	$e^{\mathbf{VU} \ln \mathbf{x}}$
$\sqrt{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} / \text{rk } \boldsymbol{\Sigma}}$	1.709 ± 0.137	0.972 ± 0.132	0.737 ± 0.074	0.666 ± 0.081	0.627 ± 0.075	0.847 ± 0.108	0.575 ± 0.074
$\sqrt{\mathbf{d}^T \mathbf{d} / \text{tr } \boldsymbol{\Sigma}}$	2.282 ± 0.176	0.751 ± 0.145	1.252 ± 0.132	0.587 ± 0.097	0.356 ± 0.077	0.620 ± 0.092	0.221 ± 0.054
rescaled	2.076 ± 0.175	0.777 ± 0.138	1.240 ± 0.129	0.579 ± 0.093	0.385 ± 0.073	0.865 ± 0.124	0.465 ± 0.080
PCT, rank $H = 1$	2.100 ± 0.194	0.489 ± 0.126	1.384 ± 0.145	0.191 ± 0.069	0.297 ± 0.084	0.184 ± 0.058	0.179 ± 0.055
2	2.402 ± 0.177	0.896 ± 0.178	1.023 ± 0.109	0.710 ± 0.124	0.353 ± 0.078	1.221 ± 0.177	0.253 ± 0.057
3	2.410 ± 0.167	0.826 ± 0.160	0.927 ± 0.102	0.623 ± 0.101	0.490 ± 0.067	1.104 ± 0.160	0.283 ± 0.052
4	2.131 ± 0.150	0.957 ± 0.163	0.838 ± 0.093	0.638 ± 0.092	0.492 ± 0.069	1.083 ± 0.156	0.535 ± 0.093
5	2.062 ± 0.154	1.007 ± 0.161	0.822 ± 0.085	0.637 ± 0.088	0.466 ± 0.063	1.023 ± 0.144	0.567 ± 0.088
6	2.053 ± 0.167	1.059 ± 0.154	0.801 ± 0.082	0.606 ± 0.082	0.448 ± 0.057	0.972 ± 0.130	0.607 ± 0.088
7	1.911 ± 0.155	1.070 ± 0.148	0.756 ± 0.078	0.578 ± 0.075	0.627 ± 0.075	0.941 ± 0.122	0.600 ± 0.084
8	1.807 ± 0.146	1.018 ± 0.140	0.715 ± 0.072	0.666 ± 0.081	0.627 ± 0.075	0.892 ± 0.115	0.579 ± 0.077
9	1.709 ± 0.137	0.972 ± 0.132	0.737 ± 0.074	0.666 ± 0.081	0.627 ± 0.075	0.847 ± 0.108	0.575 ± 0.074

zieg91.cor.pairs3

Table 6.4: Spectral distances (mean \pm deviation) of the sample pairs (from scene ‘Ziegelstein 1991’) in 7 basic feature spaces, and after additional rescaling and principal component transformation (PCT).

6.7.3 Discussion of Results

Altogether, the tables contain 80 different tested feature spaces, compared by virtue of the spectral distances d_T and d_I (Eq. 6.4 and Eq. 6.5 on page 149). Concerning Table 6.4 on the facing page, we make the following observations:

- ▶ Comparison of the basic feature spaces with the distance measure d_I (first row in Table 6.4) shows that the transformed space ($\mathbf{t} = \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$) yields the best results. The improvement is of a factor ≈ 3 in comparison to the original space (\mathbf{x}), and still of a factor ≈ 2 compared to the normalized space ($\mathbf{x}/\|\mathbf{x}\|$).
- ▶ Comparison of the basic feature spaces with the distance measure d_T (second row in Table 6.4) shows also that the transformed space ($\mathbf{t} = \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$) yields the best results. The improvement is of a factor ≈ 10 in comparison to the original space (\mathbf{x}), and still of a factor ≈ 3 compared to the normalized space ($\mathbf{x}/\|\mathbf{x}\|$). Performance is slightly better in the non-log domain (\mathbf{t}) than in the log domain ($\ln \mathbf{t} = \mathbf{V}\mathbf{U} \ln \mathbf{x}$).
- ▶ The additional color filtering with $\mathbf{V}\mathbf{U}$ always improves the results compared to pure brightness filtering \mathbf{U} or brightness normalization ($\mathbf{x}/\|\mathbf{x}\|$).
- ▶ The filtering improvements are larger for the distance measure d_T than for d_I (Eq. 6.4 / Eq. 6.5). The absolute values are better for d_I in the original space (\mathbf{x}), and better for d_T in the transformed space ($\mathbf{t} = \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$).
- ▶ The logarithmic feature space ($\ln \mathbf{x}$) performs always better than the original space (\mathbf{x}).
- ▶ The logarithmic brightness filtering ($\mathbf{U} \ln \mathbf{x}$) performs always better than the brightness normalization ($\mathbf{x}/\|\mathbf{x}\|$). There is no clear trend in comparison of the spaces $\exp(\mathbf{U} \ln \mathbf{x})$ versus $\mathbf{x}/\|\mathbf{x}\|$.
- ▶ The band-wise rescaling to unit variance (third row in Table 6.4) worsens the results in practically all spaces, and yields its best result for the log transformed space ($\ln \mathbf{t} = \mathbf{V}\mathbf{U} \ln \mathbf{x}$).
- ▶ The usefulness of the principal components transformations (PCTs) can be evaluated in a meaningful way only by virtue of the distance d_T (because of Eq. 6.15 on page 152). Looking at the various PCT spaces, there is no clear preference for either $\mathbf{t} = \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x})$ or $\ln \mathbf{t} = \mathbf{V}\mathbf{U} \ln \mathbf{x}$.
- ▶ Wrt. the distance d_T (Eq. 6.4), the PCT can improve the results only for the case of a single component ($H = 1$). Then, however, only little of the original image information is retained, so that this approach does not seem viable for practical purposes. We can explain this result as follows: After brightness filtering or normalization of the example image scene, the next most significant principal component is the ‘greenness’, *i.e.*, the similarity to a vegetation spectrum. Because all the artificial surfaces exhibit very low greenness, their mutual spectral distances are essentially orthogonal to the greenness component. Thus the transformed distances between artificial surfaces almost vanish for the case of $H = 1$.

- We cannot make a general statement about the efficiency of the PCT with regard to our problem. For a given space, it is hard to tell in general what choice for the number of components H will yield best results. The outcome of a PCT of course depends very much on the content of the image scene. We can state, however, that with a rising number of principal components H , there is a continuous transition of the result values between the two normalized distance measures d_T and d_I , as can be expected because after PCT we have a spherical overall covariance $\Sigma = \mathbf{I}$ (see Eq. 6.7 on page 150).

In summary, we may conclude that the effectiveness of the suggested logarithmic filtering has been clearly shown on real multispectral image data. Improvement by the introduced color filtering for the normally used Euclidean spectral distance is of a factor ≈ 10 in comparison to the original space, and of a factor ≈ 3 in comparison to the brightness normalized space.

6.7.4 Comparison with respect to the Filter Rank

In another evaluation series (see Table 6.5) we compared results with respect to varying filter rank $K = N - \text{rk } \mathbf{P}$. Here we investigated only two of the seven basic feature spaces ($\ln \mathbf{x}$ and $\ln \mathbf{t} = \mathbf{P} \ln \mathbf{x}$), but for each column we filtered with a projector \mathbf{P}_K of increasing filter rank K up to $K = 5$. In particular, we have the filters

$$\mathbf{P}_{K=0} = \mathbf{I} \quad , \quad \mathbf{P}_{K=1} = \mathbf{U} \quad , \quad \mathbf{P}_{K=2} = \mathbf{V}\mathbf{U} \quad ,$$

so the values for the filter ranks $K = \{0, 1, 2\}$ in Table 6.5 are the same as the corresponding entries in Table 6.4.

Concerning the tests with different filter ranks K (Table 6.5 on the next page), we observe that:

- For the distance d_I (Eq. 6.5 on page 149, first row in Table 6.5) we get best results for $K = 3$. However, the improvements are not strictly significant in the sense of the t -test (Eq. 6.10 on page 151).
- For the distance d_T (Eq. 6.4 on page 149, second row in Table 6.5) we also get best results for $K = 3$. However, more precisely, the improvement between $K = 1$ (\mathbf{U}) and $K = 2$ ($\mathbf{V}\mathbf{U}$) is statistically significant, whereas the further improvement between $K = 2$ and $K = 3$ is not, since the difference is smaller than the error margins.

We summarize that filtering with $K = 3$ might even be better than with $K = 2$; however, we have to keep in mind that each filtered dimension also reduces the information content of the data. So, filtering with $K = 2$ ($\mathbf{V}\mathbf{U}$) may be a good compromise.

SEPARABILITY W.R.T. FILTER RANK						
	$\ln \mathbf{x}$	$\mathbf{P}_K \ln \mathbf{x}, K =$				
	0	1	2	3	4	5
$\sqrt{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} / \text{rk } \boldsymbol{\Sigma}}$	0.737 ± 0.074	0.666 ± 0.081	0.627 ± 0.075	0.617 ± 0.076	0.652 ± 0.083	0.670 ± 0.091
$\sqrt{\mathbf{d}^T \mathbf{d} / \text{tr } \boldsymbol{\Sigma}}$	1.252 ± 0.132	0.587 ± 0.097	0.356 ± 0.077	0.309 ± 0.065	0.351 ± 0.062	0.329 ± 0.063

zieg91cor.pairs3

Table 6.5: Spectral distances (mean \pm deviation) of the sample pairs (from scene ‘Ziegelstein 1991’) in 6 feature spaces of varying rank $\text{rk } \boldsymbol{\Sigma} = N - K$: The log space $\ln \mathbf{x}$ (for $\mathbf{P}_0 = \mathbf{I}$, i.e., no filtering), and the filtered spaces $\mathbf{P}_K \ln \mathbf{x}$, particularly $\mathbf{P}_1 = \mathbf{U}$ and $\mathbf{P}_2 = \mathbf{VU}$.

6.8 Sensor Specific Noise Analysis

Let us recall the analysis of error propagation as set out for the general case in Section 4.6 and 4.7 (page 80). For the specific case of the DAEDALUS sensor, we can fill in the respective parameters and evaluate some of the general error propagation formulae, in order to get an idea of how significant the errors and their propagation are in the here used multispectral image data.

The error propagation as well as the required quantization accuracy depend on the matrices \mathbf{P} , \mathbf{U} , \mathbf{V} (see Eq. 4.32 on page 53, and Eq. 4.103 on page 65) and thus on the assumed spectral diffuse to global illumination ratio $\hat{\mathbf{m}}$, and finally on the sensor specific positions of the spectral bands within the electromagnetic spectrum.

Assuming $\hat{\mathbf{m}} \sim \lambda^{-1}$ yields the diffuse to global illumination ratio values as given in Table 6.6. From $\hat{\mathbf{m}}$ we construct the proper projector matrices \mathbf{U} and \mathbf{V} (for illustration see Fig. 6.21 on the facing page). For all projectors the line sums cancel out: $\sum_j P_{ij} = 0$; $\forall j$ (Eq. 4.114 on page 66). Two values of interest for this sensor specific projector $\mathbf{P} = \mathbf{V}\mathbf{U}$ are:

$$\sum_{ij} P_{ij}^2 = \sum_i P_{ii}^2 = \text{tr } \mathbf{P} = 7.0 \quad [= N - 2] \quad (6.16)$$

(compare Eq. 4.133 and 4.137 on page 68)

$$\langle \sum_j P_{ij}^2 \rangle_i = \langle P_{ii} \rangle_i = 0.777 \quad [= (N - 2)/N] \quad (6.17)$$

(compare Eq. 4.134 on page 68)

where $\langle \rangle_i$ denotes the average value over all spectral bands i , and the single values are given in Table 6.6.

■ Reduction of Multiplicative Noise

From Eq. 4.161 on page 81 we know that the relative level of multiplicative noise $\sigma^2(x_i)/x_i^2$ in the original signal \mathbf{x} becomes transformed to a lesser noise level $\sigma^2(t_i)/t_i^2$ in the transformed signal \mathbf{t} . The reduction of the relative level of multiplicative noise is $\sum_j P_{ij}^2 = P_{ii}$ for spectral band i . On average this amounts to $\langle P_{ii} \rangle = 0.78$ (Eq. 6.17) in our specific case (Table 6.6).

SENSOR SPECIFIC PROJECTOR									
band number i	2	3	4	5	6	7	8	9	10
Wavelength [μm]	0.485	0.560	0.615	0.660	0.723	0.830	0.980	1.650	2.215
diffuse/global $\hat{\mathbf{m}}$	0.400	0.346	0.315	0.294	0.269	0.234	0.200	0.118	0.088
$\sum_j P_{ij}^2 = P_{ii} =$	0.629	0.783	0.841	0.867	0.885	0.885	0.856	0.679	0.575

Table 6.6: Assumed diffuse to global illumination ratio $\hat{\mathbf{m}}$ and the proper projector matrix line sums $\sum_j P_{ij}$ for $\mathbf{P} = \mathbf{V}\mathbf{U}$. The first spectral band $i = 1$ ($\lambda_1 = 435$ nm) was omitted because of its low signal-to-noise level.

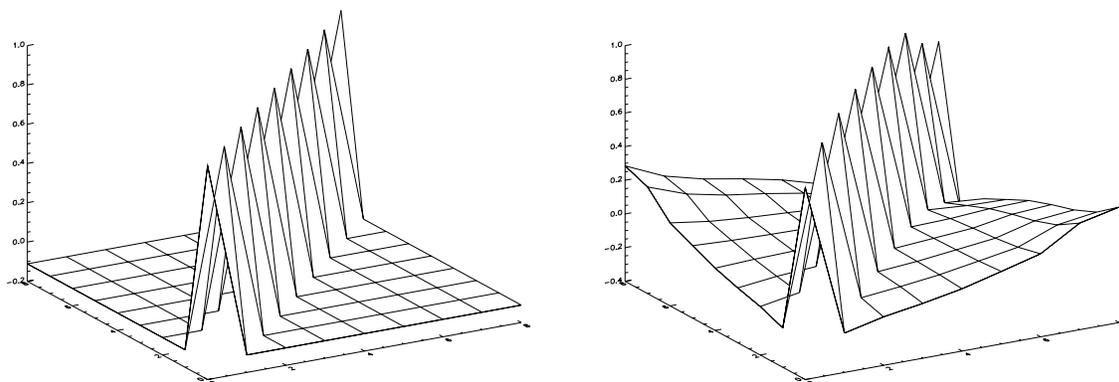


Figure 6.21: Surface plots of the projector matrices \mathbf{U} and \mathbf{V} .

■ Sensitivity to Systematic Errors

In Section 4.7.1 on page 84 we have shown the distances in the log space to be invariant against systematic multiplicative errors. In contrast, they are not invariant against systematic additive errors (Section 4.7.2). The characteristic dependence of the relative error on the signal value x_i is the same as with the relative quantization error plotted in Fig. 6.22 on page 161.

Substituting a sensor specific mean $\langle \sum_j P_{ij}^2 \rangle = 0.777$ from Eq. 6.17 into Eq. 4.184 on page 85 we get a relative error of

$$\frac{dt_i}{t_i} = d(\ln t_i) = 78\% \quad (6.18)$$

for the bad case of $dc_{o,j} \approx x_j$ where systematic additive error is of approximately the same magnitude as the input signal.

6.8.1 Quantization Error

We consider the quantization error as an additive random error independent of the signal strength. Thus it becomes relatively high for small signals and low for high signals.

The opto-electronic detector element of the sensor yields an amplified continuous voltage which is rounded into integer values at the analogue-digital-converter. The round-off to integer numbers, to the digital counts (DC), produces a maximum error of $\pm 1/2$. The root mean square deviation (rms) of the rounded DC from the original continuous value can be determined as follows: We consider an interval $[-1/2..1/2]$ and assume the continuous values to be equally distributed within this interval and all mapped to the 0 DC gray value. Then the rms deviation σ is the square root of the deviation of a continuous value x from the rounded value 0 integrated over the considered interval:

$$\sigma = \sqrt{\int_{-1/2}^{1/2} (x - 0)^2 dx} \quad (6.19)$$

$$= \sqrt{\left[\frac{1}{3} x^3 \right]_{-1/2}^{1/2}} \quad (6.20)$$

$$= \sqrt{\frac{1}{12}} \quad (6.21)$$

$$= 0.289 \quad (\text{in units of DC}) \quad (6.22)$$

$$\text{and } \sigma^2 = 0.083 \quad (6.23)$$

When the raw DC values are calibrated into radiance and reflectance values (see Section 2.1 on page 13) they undergo various multiplications. After these corrections, the resulting FLOAT-type values are commonly rounded to integer type values again, in order to save storage and computation costs. Thus the original quantization error of the AD-converter may well be amplified. Therefore we prefer a conservative estimate of the quantization error and assume $\sigma = 1$ instead of the lower bound of $\sigma \approx 0.3$ (Eq. 6.22).

For the here suggested operators \mathbf{U} , \mathbf{V} and \mathbf{P} in the logarithmic space, the relative variance σ/x_i is of interest. As stated above, it becomes high for small signals x_i and low for high signals. This behaviour is illustrated in Fig. 6.22.

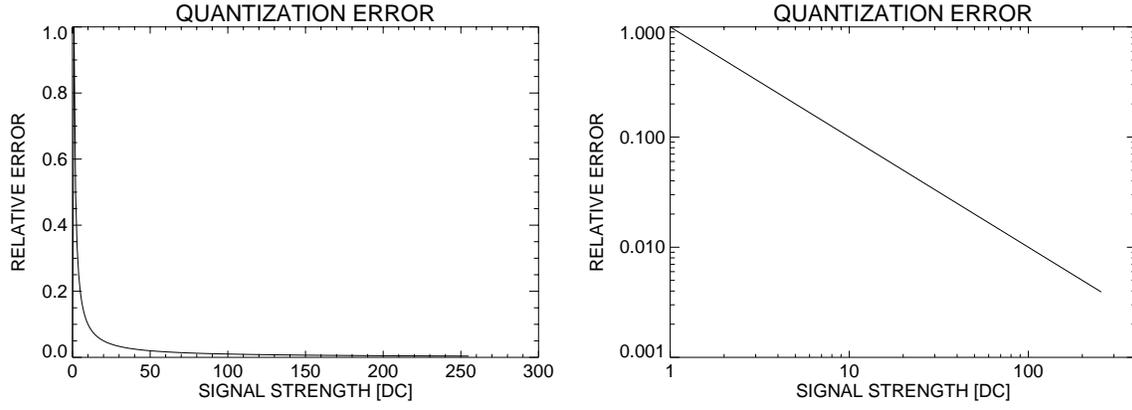


Figure 6.22: Relative error caused by 8 bit quantization, assuming an additive noise of 1 DC over a signal of [1..255] DC.

Right: log/log representation.

6.8.2 Magnitude of Color Shift and Estimation of Required Dynamic Range

Here we want to answer the question what the dynamic range of the sensor has to be to observe the color shift effects. To pose the question in a positive way, how coarse must the quantization be so that one is *not* bothered with the color constancy problem? This of course means also that spectral signatures from shadowed areas cannot be used due to their low signal-to-noise ratio.

The gain for an optimal recording will be set such that the mean signal gets half of the maximal possible number of Digital Counts (DC): $\langle x \rangle = \frac{1}{2} \text{DC}_{\max}$. (For convenience, the dimensionless unit of DC is dropped in the following equations.)

In order to assess the effect of color inconstancy we consider as an example a white reflectance under global illumination, a wavelength independent reflectance spectrum $\mathbf{x} = (c \dots c)^T$ with $c = \frac{1}{2}$ (Fig. 6.23 on the next page, upper solid line). When this reflectance spectrum is illuminated by the skylight spectrum $\hat{\mathbf{m}}$ only, we will observe a pseudo-reflectance spectrum \mathbf{x}' with $x' = \hat{m}_i x_i$ (lower solid line) which is color shifted and no longer 'white' (wavelength independent). Now we look for a wavelength independent spectrum $\mathbf{x}'' = (c'' \dots c'')^T$ which has the same log brightness index (see Eq. 4.112 on page 66) as \mathbf{x}' :

$$\hat{\mathbf{u}}^T \ln \mathbf{x}'' = \hat{\mathbf{u}}^T \ln \mathbf{x}' \quad (6.24)$$

$$\frac{1}{\sqrt{N}} \sum \ln c'' = \frac{1}{\sqrt{N}} \sum \ln x'_i \quad (6.25)$$

$$N \ln c'' = \sum \ln(x_i \hat{m}_i) \quad (6.26)$$

$$c'' = e^{\frac{1}{N} \sum_i \ln(x_i \hat{m}_i)} \quad (6.27)$$

$$= e^{\frac{1}{N} N \ln c} \cdot e^{\frac{1}{N} \sum_i \ln \hat{m}_i} \quad (6.28)$$

$$= c \cdot e^{\frac{1}{N} \sum_i \ln \hat{m}_i} \quad (6.29)$$

$$= 0.114 \quad (6.30)$$

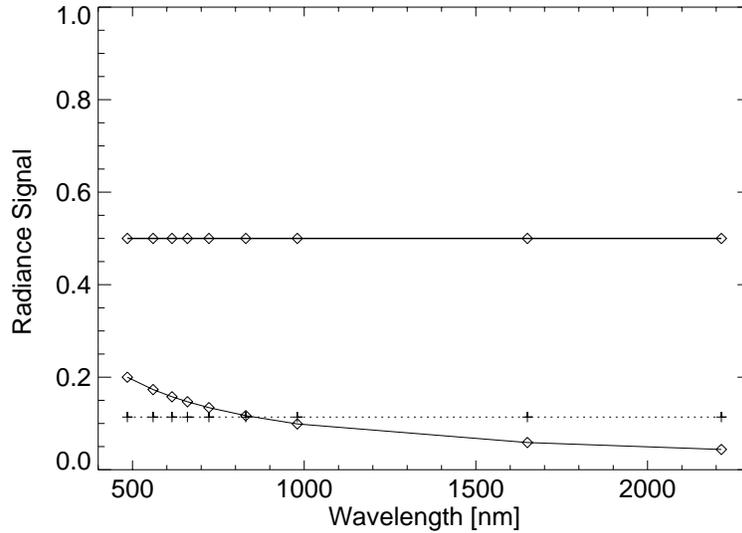


Figure 6.23: *Brightness and color shift experienced by the spectral signature of a white surface in the shadow.*

for the specific DAEDALUS wavelength positions and an assumed relative diffuse illumination $\hat{\mathbf{m}} \propto \lambda^{-1}$. We then consider the deviation between the even spectrum \mathbf{x}'' (dotted line in Fig. 6.23) and the actual one \mathbf{x}' as the effect of the color shift. For a specific band i we have a relative deviation to a white spectrum of

$$\frac{x'_i - x''_i}{x''_i} = \frac{m_i c - c''}{c''} \quad (6.31)$$

$$= \frac{\hat{m}_i}{e^{\frac{1}{N} \sum_i \ln \hat{m}_i}} - 1 \quad (6.32)$$

which is actually independent of the absolute intensity c for the reflectance spectrum \mathbf{x} which we assumed at the beginning. The deviations of \mathbf{x}' from the white spectrum \mathbf{x}'' are given in Table 6.7 on page 164 (see “surface cast in shadow, $\theta = 0^\circ$ ”), and have a mean absolute deviation of $\Delta = 0.043$. If we want to detect this deviation with a signal to noise ratio of $\text{SNR} = 10$ where the quantization noise is 1, then the number of coding bits necessary for this quantization accuracy is

$$-\text{ld}(\Delta) + \text{ld}(\text{SNR}) = \quad (6.33)$$

$$-\text{ld}(0.043/10) = 7.87 \quad (6.34)$$

$$\Rightarrow 8\text{bit}. \quad (6.35)$$

As the gain setting during recording usually cannot be optimal and pure diffuse illumination is the extreme case, we expect the color shifts to be clearly above quantization noise level only for the first and last spectral bands where the effect is most accentuated. The effect of the color shift can be sufficiently well measured and corrected in all spectral bands only for sensors with a dynamic range of more than 8 bit.

Three examples were computed to point out the magnitude of the color shift as experienced by the spectral signature of a tilted surface. For a solar zenith angle of

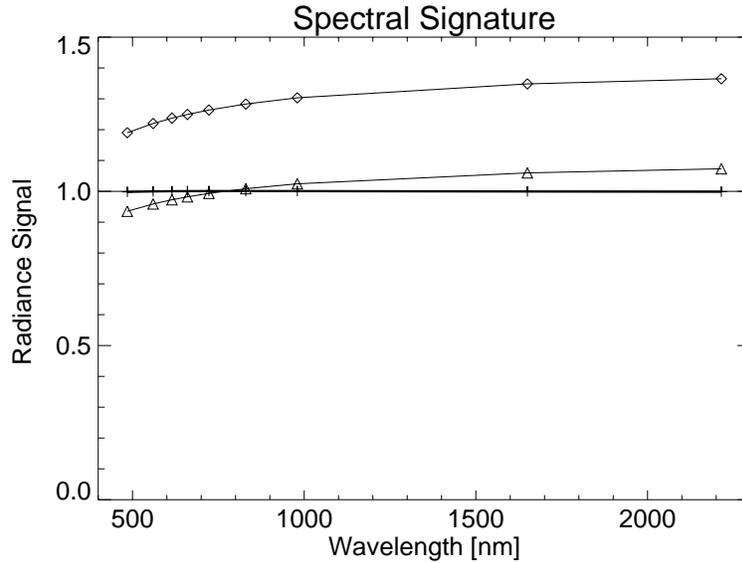


Figure 6.24: Brightness and color shift experienced by the spectral signature of a white surface with tilt $\theta = 45^\circ$ under a sun of solar zenith angle $\theta_\odot = 45^\circ$:

The original white reflectance spectrum (thin line —, $x_i = 1.0$), the pseudo-reflectance spectrum of the tilted surface, (thin line with \diamond symbol), then brightness filtered ($\exp(\mathbf{U} \ln \mathbf{x})$, thin line with \triangle symbol), and brightness and color filtered ($\exp(\mathbf{P} \ln \mathbf{x})$, thick line with $+$ symbol).

$\theta_\odot = 45^\circ$, we again consider a white reflecting surface with a spectral signature $x_i = 1.0$. The cosine between solar angle and surface normal, $\cos(\theta_\odot - \theta)$, is the contribution factor of the relative direct illumination \mathbf{n} , and $\cos^2(\theta/2)$ (Eq. 2.16 on page 17) the contribution factor of the relative diffuse illumination $\hat{\mathbf{m}}$. Given the sensor-specific positions of the spectral bands in the electromagnetic wavelength range, we can then compute the pseudo-reflectance signal \mathbf{x} as received by the sensor from the surface tilted with $\theta = -45^\circ, 45^\circ$ and 60° . The original and the filtered spectral signatures for $\theta = 45^\circ$ (surface is normal to the sun) are plotted in Fig. 6.24. The relative deviations of the original and the filtered signal are listed in Table 6.7). The spectral filtering cannot reduce the color shift completely because the filter $\mathbf{P}_{K=2} = \mathbf{V}\mathbf{U}$ represents only $K = 2$ terms of a Taylor series expansion (Eq. 4.72 on page 60). However, *e.g.* for the $\theta = 60^\circ$ case we expect a remaining deviation from the white reflectance of only 0.1%, whereas the original signature is expected to exhibit color and brightness shift effects of $\approx 70\%$, and the brightness filtered signature color shift effects of $\approx 7\%$.

MAGNITUDE OF COLOR SHIFT									
band number i	2	3	4	5	6	7	8	9	10
Wavelength [μm]	0.485	0.560	0.615	0.660	0.723	0.830	0.980	1.650	2.215
illumination ratio									
diffuse/global $\hat{\mathbf{m}}$	0.400	0.346	0.315	0.294	0.269	0.234	0.198	0.118	0.088
direct/global $\hat{\mathbf{n}}$	0.600	0.654	0.685	0.706	0.731	0.766	0.802	0.882	0.912

surface cast in shadow ($\theta = 0^\circ$)									
deviation from white spectrum [dimensionless pseudo-reflectance]:									
$x_i - \exp(\langle \ln x_i \rangle)$	0.086	0.059	0.044	0.033	0.020	0.003	-0.015	-0.055	-0.070

tilted surface in shadow ($\theta = -45^\circ$ under sun zenith $\theta_\odot = 45^\circ$)									
relative deviation from white spectrum [%]:									
\mathbf{x}	-66	-70	-73	-75	-77	-80	-83	-90	-93
$\exp(\mathbf{U} \ln \mathbf{x})$	76	52	39	29	18	3	-13	-48	-62
$\exp(\mathbf{VU} \ln \mathbf{x})$	-14	-4	2	5	9	12	12	-2	-15

tilted surface ($\theta = 45^\circ$ under sun zenith $\theta_\odot = 45^\circ$)									
relative deviation from white spectrum [%]:									
\mathbf{x}	19	22	24	25	26	28	30	35	37
$\exp(\mathbf{U} \ln \mathbf{x})$	-6	-4	-3	-2	-1	1	2	6	7
$\exp(\mathbf{VU} \ln \mathbf{x})$	-0.2	0.0	0.0	0.1	0.1	0.1	0.1	-0.1	-0.1

tilted surface ($\theta = 60^\circ$ under sun zenith $\theta_\odot = 45^\circ$)									
relative deviation from white spectrum [%]:									
\mathbf{x}	57	63	66	69	71	75	79	87	91
$\exp(\mathbf{U} \ln \mathbf{x})$	-9	-6	-4	-2	-1	1	3	8	10
$\exp(\mathbf{VU} \ln \mathbf{x})$	-0.3	0.0	0.1	0.1	0.2	0.2	0.1	-0.1	-0.2

Table 6.7: Assessment of the brightness and color shift magnitude:

Example of a tilted white reflecting surface under a solar zenith angle $\theta_\odot = 45^\circ$; in shadow ($\theta = -45^\circ$, pure diffuse illumination), with tilt $\theta = 45^\circ$ (normal to the sun), and with tilt $\theta = 60^\circ$.

The spectral values for the original and filtered signals are computed for the scanner specific wavelengths, with the relative colorshift given in percent. (See also the corresponding plots in Fig. 6.23 and 6.24.)

6.8.3 Error Weighted Spectral Distances of the Spectral Sample Pairs

In the previous section we have seen that the quantization error is considerable for small radiances. Thus, in particular the distances between spectra from dim or shaded areas are prone to error. We have discussed the proper error propagation in Section 4.6 on page 80. To account for the quantization error induced uncertainty, an error weighted spectral distance was developed in Section 4.9.1 on page 89.

As in Table 4.2 on page 92 for simulated data, we have computed the error weighted spectral distances for real image data (selected sample pairs, Fig. 6.20 on page 153) and compared them to the unweighted spectral distances. As before, we have tested three different feature space transformations:

$$\mathbf{x} \mapsto \mathbf{t} = \exp(\mathbf{V}\mathbf{U} \ln \mathbf{x}) \quad (\text{nonlinear transform}) \quad (6.36)$$

$$\mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{V}\mathbf{U} \ln \mathbf{x} \quad (\text{nonlinear transform}) \quad (6.37)$$

$$\ln \mathbf{x} \mapsto \ln \mathbf{t} = \mathbf{V}\mathbf{U} \ln \mathbf{x} \quad (\text{linear transform}) \quad (6.38)$$

For the error weighted distances computed according to Eq. 4.216 on page 90 we have assumed the quantization error as an additive error of ± 1 DC constant for all spectral bands. (It seems interesting to note that, provided the quantization error is constant for all spectral bands, the *value* of the assumed standard deviation does *not* influence the error weighted spectral distance because of the division by the trace in Eq. 4.216 on page 90.)

The results are listed in Table 6.8 on the next page. We observe that the error weighting improves considerably (by a factor ≈ 68) for the $\mathbf{x} \mapsto \mathbf{t}$ mapping, while making no difference for the other two mappings. Moreover we observe, that just for the error weighted $\mathbf{x} \mapsto \mathbf{t}$ mapping, the by far lowest spectral distance ratio is achieved. We can conclude that:

- ▶ The transformation $\mathbf{x} \mapsto \mathbf{t}$ is the preferable one.
- ▶ It is advisable to use error weighting for computation of the spectral distance.

However, the error weighting requires a full error covariance matrix to be stored for each image pixel when computing the transformed spectra. The error matrix cannot be restored from the transformed spectra only. The data required for the error weighting is thus $\approx N/2$ times greater than for the transformed image data itself.³ Therefore, the error weighting seems too costly with respect to both memory and computation time for the processing of large images which are common in remote sensing.

³ The transformed spectrum for one pixel consists of N spectral values. The error matrix for each pixel is an $N \times N$ matrix, with only $N(N + 1)/2$ significant values since the matrix is symmetric.

ERROR WEIGHTED SPECTRAL DISTANCE RATIOS			
ratio $\frac{\text{transformed}}{\text{original}}$ distance	$\mathbf{x} \mapsto \mathbf{t}$	$\mathbf{x} \mapsto \ln \mathbf{t}$	$\ln \mathbf{x} \mapsto \ln \mathbf{t}$
(1) $[\text{mean} \pm \text{dev}] \cdot 10^{-6}$ unweighted	23.1 ± 9.0	14.5 ± 4.1	19135.7 ± 4434.3
(2) weighted	0.34 ± 0.16	16.5 ± 4.3	20476.1 ± 5938.2
ratio (1)/(2) unweighted/weighted	67.8	0.88	0.93

zieg91cor.pairs3

Table 6.8: Error weighted spectral distance ratios compared to unweighted spectral distance ratios for spectral sample pairs.

Conclusions

7.1 How Important is the Color Constancy Problem for Multispectral Remote Sensing?

By far most procedures used in multispectral digital image processing, and particularly multispectral classification, are based on the explicit or tacit assumption of Lambertian reflectance, and are, moreover, assuming the same spectral signature of a certain surface material regardless of its surface orientation. Both assumptions are similar as they ignore geometry effects, but there is a subtle difference between these two aspects:

- ▶ The Lambertian assumption means that the orientation of the sensor's line of view relative to a fixed surface and fixed sun position can change without affecting the measured radiance spectrum. A more accurate model would have to take BRDF effects into account.
- ▶ A change in surface orientation or sun position with a fixed sensor position can exert a change in the received spectrum even if the reflecting surface satisfies Lambert's law, because the illumination of the target surface changes with its orientation.

When only the sensor position changes but target surface orientation and sun position are fixed, then any change in the received spectrum can only stem from non-Lambertian BRDF-effects. In contrast, when the surface orientation changes, and the sensor and the sun are fixed, then effects in the measured spectrum can stem both from BRDF effects as well as from illumination changes. This thesis is centered around the second aspect, the surface orientation related change in measured spectra, and assumes Lambertian reflection in a weak form (see Section 4.10 on page 93).

There is a widespread awareness of the compromising effect of the Lambertian simplification on multispectral image processing. A new stage of sophistication would clearly be to include BRDF modeling. In this work, however, we could show that even with the Lambertian assumption a varying surface orientation has severe effects such as brightness and color shifts, because of the direct and diffuse illumination. The effect can be up to 75% in single spectral bands for an exemplary sensor configuration (see Table 6.7 on

page 164). Although there are approaches where the illumination spectrum is estimated for each image pixel, these are rather insufficient as the surface orientation is known only to the accuracy of the terrain model (with a typical grid size of > 20 m) and the diffuse illumination is simplified to be isotropic over the sky's hemisphere.

We would like to call the spectral ambiguities introduced by varying surface orientation¹ the ‘Color Constancy Problem in Multispectral Remote Sensing’, since the underlying physical uncertainty is the unknown illumination spectrum. The problem of recovering a ‘color’ (reflectance spectrum) or an invariant from an ambiguous illumination spectrum, however, has a long history in RGB-based computer vision and theoretical biology as the ‘Color Constancy Problem’ (see Section 1.2 on page 3 and Section 1.3 on page 7).

This work shows that a certain aspect of the general color constancy problem, namely the surface orientation ambiguity, is of importance to multispectral image processing and has been neglected so far. Varying surface orientations can lead to brightness and color shifts in the spectral signature as observed under natural outdoor illumination. These brightness and color shifts may cause erroneous change detection and spurious classification results (see examples in Section 6.3 on page 122 through Section 6.5). The existence of the color shift has been clearly shown by this thesis. Compared with the magnitude of the well known brightness shift it is a second order effect.

With the sun/shadow illumination being just a sub-topic of the color constancy problem, the problem of the *hot spot effect* is also implicitly addressed by this work. The hot spot appears on rough surfaces such as vegetation and is induced by the dependence of the percentage of microshadows on the viewing angle relative to the sun (for a modeling see *e.g.* Borel et al. (1991)). It is thus just another effect of illumination ambiguity.

While specular reflection, the most severe BRDF related problem next to the hot spot effect, is usually confined to a finite number of pixels in a given image which can be identified and masked out, we consider the surface orientation effects on the measured spectra to be much more ubiquitous.

The importance of the surface orientation related color constancy problem rises with an increasing dynamic range of the radiance detectors. With 8 bit sensors, the analysis of comparatively brightly or poorly lit pixels fails due to overflow and insufficient signal to noise ratio. With a higher dynamic range, however, analysis of such pixels is possible, and it is just there that the effects of color inconstancy are the most pronounced.

7.2 Which Improvements have been Achieved with this Work?

We have developed a framework of a dichromatic illumination which describes the illumination as a linear combination of direct sun and diffuse sky light. We have shown that the spectra of a given surface reflectance lie on two dimensional class planes in the

¹ We include both shading and shadowing, where *shading* denotes the decrease of illumination due to surface orientation, whereas *shadowing* means that a surface patch is fully cast in shadow by a taller neighboring object.

feature space (Chapter 3). Furthermore, the principal modes of the surface orientation related variance can be found by a series expansion in the logarithmic feature space. There, these modes are equal for all Lambertian reflectance spectra (Chapter 4). Actually, the Lambertian assumption is a sufficient but not a necessary condition: **We can allow for BRDF-effects without endangering the abilities of the suggested filter, as long as these BRDF-effects are only geometry- but not wavelength-dependent.**²

Based on the analysis in the logarithmic feature space was the design of a mapping, a linear algebraic projector, which filters out the modes of surface orientation related spectral variance, namely the most significant brightness and color shift. The suggested transform is a mapping which maps each N -dimensional spectrum to a spectral invariant descriptor with N entries. The mapping operator could be separated into a logarithmic brightness and a color shift filter \mathbf{U} and \mathbf{V} . Due to the logarithmic brightness shift filter \mathbf{U} , we can actually relax the Lambertian assumption we started out from: The invariant can deal with non-Lambertian effects as long they are wavelength independent, in other words, if the BRDF can be factorized into a geometry dependent non-Lambertian and a wavelength dependent Lambertian function.

In the logarithmic feature space, the discussed transform is a linear mapping, and thus error propagation can be well understood. The transform has an overall smoothing effect on multiplicative random noise. How well the mapping improves the separability of two reflectance spectra which are broadened into clusters by variable surface orientation, depends on the noise level. The analysis of error propagation shows that the transform is increasingly useful once a certain signal to noise level is surpassed. The relative quantization accuracy is poor for radiances of only a few digital counts (DC). Including the accuracy estimation, individually for each spectrum and each spectral band respectively, leads to an error weighted spectral distance. The computation is significantly more costly than for the simple spectral distance. The results on simulated data justified the cost in that cluster separation improved with error weighting.

We have verified the dichromatic illumination by several thousands *in situ* measured spectra of a dozen samples of different surface materials (Chapter 5). Our analysis of the spectral variability with respect to surface orientation suggested that

- ▶ the brightness shifts are in the 100%-range (*i.e.*, factor 2),
- ▶ the color shifts are in the 5–10%-range,
- ▶ the brightness shift filtering reduces the variability by more than 99%,
- ▶ the color shift filtering reduces the variability by another 40–80%,
- ▶ for diffuse reflecting surface material no more than two dimensions need to be filtered,
- ▶ the choice of the atmospheric parameter γ needed for the construction of the filter projector is not critical (for the limited wavelength range of the sensor).

² Note however, that some forms of specular reflection are wavelength independent, in other words ‘white’, and thus *do* change the body reflectance of a non-white surface in a wavelength *dependent* fashion.

We then have applied the suggested transform to experimental multispectral imagery (Chapter 6). The transformed image data was subjected to three different techniques in order to assess the merits of the presented invariant: direct comparison of the spectral signatures, supervised and unsupervised multispectral classification, and statistical evaluation.

Sample pairs of surfaces of identical reflectance but different orientation were selected (roof top segments). Other bitemporal spectral sample pairs consist of spectra from the same surface with identical orientation but varying solar illumination angle. The respective mutual spectral distances between the two spectra of each pair were evaluated relative to the overall variance in the various feature spaces. Plots of the two spectra of each sample pair illustrate the effect of the filtering. The filtered feature space of the transformed spectra showed significantly lessened relative spectral distances. This means that it becomes easier to define a cluster in the feature space which accommodates all the spectra of a specific reflectance under varying orientation while excluding spectra of other reflectance spectra. Of the various tested feature spaces, best results were achieved for the transformed space ($\mathbf{x} \mapsto \mathbf{t} = \exp(\mathbf{P} \ln \mathbf{x})$). Moreover, the logarithmic brightness filtering ($\mathbf{x} \mapsto \exp(\mathbf{U} \ln \mathbf{x})$) proved to work better than the brightness normalization ($\mathbf{x} \mapsto \mathbf{x}/\|\mathbf{x}\|$).

Apart from the quantitative analysis, we employed the more intuitive test of multispectral classification of image clips surrounding the sample pairs into binary images. Ideally, all segments of a surface should be accepted without regard to its orientation, while rejecting as much as possible of the remaining image clip. The displayed classification results show that, firstly, the transformed data is often more suitable for this task, and that, secondly, the threshold criticality is lessened in the transformed feature space.

A third test of the suggested transform was conducted by unsupervised fuzzy clustering in the original and the transformed feature space. An exemplary image scene was classified into two classes. The mutual separation of the two automatically determined clusters in the feature space was measured relatively to their scatter radii and improved by about three times. At the same time, contour plots of the principal components of the feature space showed a much more satisfactory clustering result. Common unsupervised clustering searches for spherical clusters and is even more improved by the invariant transform than supervised classification.

Finally, the general noise and error propagation analysis was specified for the used multispectral sensor. Error weighting improved the results also for real imagery. However, it is quite costly for image-size spectral data.

The transformed spectral distance, and the error weighted transformed spectral distance (see Section 4.9 and Section 4.9.1 on page 89) between two designated spectra, provide a powerful interactive test for a screen analyst to decide if these spectra could possibly stem from an equal reflectance where just the surface orientation (or the illumination: sun/shadow) is different.

The invariant transform was implemented in PVWAVE and IDL interpreter languages. For portability and easy accessibility via a graphical user interface (GUI), the codes were incorporated in the remote sensing image processing environment ENVI (see Section 4.4.4 on page 70).

In summary, we have presented a novel spectral invariant which has little parameter sensitivity, which is scene-content independent, which is theoretically well founded and has a sound experimental validation. It can be computed pixel-wise at low cost³ (of order $\mathcal{O}(MN^2)$, *i.e.*, linear with the number of pixels M and quadratic to the number of spectral bands N).

We do not suggest that the transformed data should completely replace the original spectral data for all kinds of image processing applications. Too large is the information content lost by the filtering process. Rather, we see the suggested transformation as a useful early preprocessing step, enabling to perform subsequent classification, edge detection and segmentation in a way which is invariant against surface orientation. The transformed spectra may be used complementary to the original pseudo-reflectance spectra in order to decide if spectral discrepancies *could* be due to varying surface orientation.

We could supply a successful answer to a problem of interaction between observation geometry and spectral signature, which – just as BRDF-effects – has mostly been neglected so far in common multispectral image processing.

7.3 Original versus Logarithmic Feature Space

Common multispectral image analysis works in the original feature space which is spanned by the axes of the measured radiances x_i in the various spectral bands. As an alternative, *band ratios* are sometimes used: Instead of the value x_i of band i , the ratio x_i/x_{i+1} is used (Sabins 1978, Lillesand & Kiefer 1987). This essentially effects a brightness normalization.

For the color constancy problem with respect to surface orientation, we have found it useful to investigate the *logarithmic feature space* where the axes are denoting the logarithm of the measured radiance values. In the logarithmic feature space, all brightness shifts are along the unity vector \mathbf{u} .⁴ In Chapter 4 we could show that the logarithmic space is better suited for the analysis of multiplicative processes, whereas the original space is more suitable for additive mechanisms (Section 4.1 on page 48 through Section 4.3 on page 59). The nature of some important processes to be considered in remote sensing image formation is listed in Table 7.1 on the following page.

Unsupervised classification, such as hard or fuzzy k -means clustering, will not be affected by a systematic additive offset which merely exerts a uniform shift on all the clusters in the feature space. In contrast, a systematic multiplicative factor will change the form of the clusters depending on their position in the feature space. In the logarithmic feature space, the effects are vice versa: a simple unaffacting shift caused by a multiplicative factor, and change of cluster shapes by an additive offset.

The ideally unique spectral signature of a class is ‘broadened’ into a distributed cluster in the feature space by various non-systematic additive and multiplicative mechanisms. But the fundamental physical principle of passive remote sensing is the reflec-

³ For comparison, the widely used Maximum Likelihood classification has the same computation cost of order $\mathcal{O}(MN^2)$ for each spectral class.

⁴ A certain drawback with respect to the computational point of view may be that for the logarithmic space FLOAT values will be preferred instead of INTEGERS, causing higher memory and processing costs. Considering error propagation, the errors of the log values become high for low original radiance values (Fig. 6.22 on page 161).

additive:	multiplicative:
<ul style="list-style-type: none"> • path radiance • stray radiance from adjacent surfaces • specular reflection • AD-converter quantization accuracy • sensor dark current and amplifier offset 	<ul style="list-style-type: none"> • atmospheric transmittance • reflection • illumination changes • detector noise which is scaled up by the sensor amplifiers • sensor amplifier gain factor

Table 7.1: *Some additive and multiplicative processes which are important in multi-spectral remote sensing.*

tion of the illumination at the target surface, in other words, the spectral band-wise multiplication of the illumination spectrum with the reflectance spectrum. This multiplication makes up for the typical feature space appearance of ‘rays’ towards the origin which are caused by spectra of equal reflectance under varying illumination intensity. The advantage of the logarithmic feature space is that here all these rays are running parallel rather than radial as in the original space (see Fig. 4.2 on page 58). Also the cluster broadening caused by varying direct and diffuse illumination contributions is a multiplicative mechanism.

When accepting the postulate that cluster broadening stems mainly from changes in illumination and is thus a multiplicative process, then the logarithmic feature space clearly is the choice for performing multispectral analysis.

7.4 Lessons Learned

During this work we had to acknowledge a number of aspects which limited the performance of the suggested transform and its analysis. Even though some of the topics we name below are trivial, they pose some severe constraints which cannot be dismissed.

- ▶ Stray radiance from adjacent pixels (adjacency effect) and path radiance are superimposed on the true reflected signal. In contrast to atmospheric transmissivity or sensor gain calibration, these additive components compromise the analysis of the color constancy problem as related to surface orientation. Although the adjacent radiance and path radiance have been corrected for (see Section 6.2 on page 117), the analysis results can only be as good as the prior calibration and atmospheric correction.
- ▶ The 8 bit dynamic range of the multispectral image sensor used here is coarse for our purposes. In principle, our suggested invariant transform makes surfaces exposed to full sun light and surfaces cast in complete shadow equally classifiable. In practice, however, under- and overflows at either end of the [0..255] DC

range occur frequently, and the signal-to-noise ratio for shadowed surfaces at a quantization accuracy of 1 DC is approaching 1 (see Fig. 6.22 on page 161).

- ▶ For multispectral data of $N = 10$ spectral bands, say, the filtering of $K = 2$ components does not seem to present a severe loss of information. However, the dimension which is truly inherent in the multispectral image data is significantly smaller since the spectral data is highly correlated. E.g. in a typical image scene (Fig. 6.1 on page 119), the first two principal components account for as much as 93% of the total variance. Brightness filtering removes 76% of the overall variance, and color shift filtering another 8%.

By appraisal of the feature space plots of experimental imagery after brightness and color shift filtering (Fig. 6.17 on page 146), we actually observe only vegetation and non-vegetation as clearly perceivable clusters. When these are isolated as a first principal component, the remaining components apparently form a unimodal almost Gaussian distributed feature space density bare of salient clusters.

- ▶ The filtering approach which has been developed for the dichromatic illumination model (two light sources: sun and sky) can easily be generalized to an invariant against varying illumination contributions from K light sources ($K > 2$). However, as pointed out previously, with an increasing number of filtered components K the filtering will reduce the inherent dimensionality of the N -dimensional spectral data and eventually grate off its usefulness for classification at all.
- ▶ In our analysis of the usefulness of the suggested invariant for multispectral image classification we would have liked to employ unsupervised classification in order to establish a most impartial quality assessment not relying on the human analyst. However, there exists no general paradigm for image segmentation. We can show hand selected examples of what the invariant achieves and its limitations, but there is no general way of judging whether a segmentation of a given image is ‘better’ after application of the transform or not.

7.5 Outlook

Throughout this thesis we have investigated the properties of an invariant which is achieved by filtering out the variable components related to surface orientation. This spectral invariant can be achieved only in the logarithmic feature space, and only approximatively (to K^{th} order), as discussed in Chapter 4. If we drop for once the pursuit of a rigorous spectral invariant, there are other approaches which appear to be promising. Even though the investigation of these is beyond the scope of this thesis, we would like to sketch brief descriptions of two approaches which seem worthwhile to follow up on.

■ Alternative Spectral Distance with a Modified Loss Function

The invariance has been achieved by filtering out the variable components related to surface orientation. An obvious question is whether this approach might be too rigorous in that it might filter out too much spectral information as well. If we give up the objective of an invariant, we may instead develop a modified spectral distance. A more

gentle approach than the rigorous filtering would be to define a spectral distance measure which does not ignore distances in the direction of the surface orientation related variance completely but rather applies a different weighting to those in contrast to distances in other directions in the feature space. Generally, the spectral distance is connected to the notion of a ‘loss function’ $f(d_i)$ of the signed distance d_i in a certain band i .

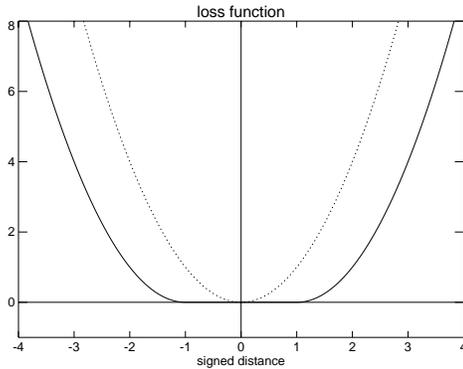


Figure 7.1: *Alternative loss function*

only after a certain threshold is surpassed (see solid line in Fig. 7.1). The underlying reasoning would be that small deviations in the particular directions of brightness and color shifts can be caused by variation in surface orientation rather than in reflectance, and that consequently they should not be taken into account for the spectral distance unless the deviation is so strong that it cannot possibly be explained by varying surface orientation. Physically meaningful thresholds can be derived from the brightness and color shift which is experienced by a surface cast in complete shadow, *i.e.*, with diffuse illumination only.

The computation of a such modified spectral distance could be realized by first computing the Euclidean distance in the invariant space

$$\sum_i f([\exp(\mathbf{P} \ln \mathbf{d})]_i) = \|\exp(\mathbf{P} \ln \mathbf{d})\|^2,$$

and then adding proper loss function values

$$f'(\hat{\mathbf{u}}^T \mathbf{d}) \quad \text{and} \quad f''(\hat{\mathbf{v}}^T \mathbf{d})$$

of the brightness and color indices $\hat{\mathbf{u}}^T \mathbf{d}$ and $\hat{\mathbf{v}}^T \mathbf{d}$ (see Eq. 4.28 on page 52 and Eq. 4.101 on page 65).

In this way one could conceive a spectral distance measure which is insensitive to changes in surface orientation but still sensitive to spectral deviations which are so large that a real change in surface reflectance has to be assumed.

■ Surface Orientation Invariant Classification in the Original Feature Space

Our analysis in Chapter 3 has shown that planar spectral classes are spanned in the original feature space by the dichromatic illumination. The disadvantage compared to the logarithmic space is that the plane spanning vectors are reflectance dependent and thus class specific. The advantage, however, is that the classes are true planes and not only local series expansions as in the logarithmic space. Moreover, in the original space

E.g. for the common Euclidean distance, each signed distance d_i in a spectral band i is ‘punished’ by the quadratic loss function $f(d_i) = d_i^2$ (see dashed line in Fig. 7.1). Then the overall spectral unsigned distance is summed up from all squares

$$\mathbf{d}^T \mathbf{d} = \sum_i d_i^2 = \sum_i f(d_i)$$

(see Eq. 2.25 on page 23). Alternatively, for distances in the directions related to surface orientation, one might think of a loss function f' which starts to ‘punish’

no memory cost increase with logarithmization (which requires FLOAT type storage) is encountered. As developed in Chapter 3, supervised classification similar to Maximum Likelihood classification, but surface orientation invariant, is possible. Moreover, with given class centers, meaningful thresholds to the extent of brightness and color shifts can be applied.

The theory has been applied to some experimental data by Wiemker (1995c). Test results on color constant classification of multispectral imagery in the original feature space remain to be worked out yet.

Appendix A

Analysis of the Spectral Variability due to Surface Orientation

A.1 The Pseudo-Inverse

The pseudo-inverse is also called Moore-Penrose Inverse or Generalized Inverse (Albert 1972, Hämmerlin & Hoffmann 1994). It essentially allows to invert matrices which are not quadratic or singular and thus cannot be inverted in the strict sense.

It can be shown that for each $m \times n$ matrix \mathbf{A} there exists a unique pseudo-inverse \mathbf{A}^+ which is an $n \times m$ matrix. Its properties are

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \tag{A.1}$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \tag{A.2}$$

$$\mathbf{A}\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^+)^T \tag{A.3}$$

$$\mathbf{A}^+\mathbf{A} = (\mathbf{A}^+\mathbf{A})^T \tag{A.4}$$

$$\mathbf{A}^+\mathbf{A} = \mathbf{I} \quad ; \quad \text{with } \mathbf{I} \text{ the } n \times n \text{ identity matrix.} \tag{A.5}$$

$\mathbf{A}\mathbf{A}^+$ is called the *orthogonal projection* (see Appendix A.2 on page 179) of the \mathbb{R}^m onto the *image* of \mathbf{A} , whereas $\mathbf{A}^+\mathbf{A}$ is the orthogonal projection of the \mathbb{R}^n onto the *orthogonal complement* of the *kernel* of \mathbf{A} . (The image of a matrix \mathbf{A} is the set of all vectors $\mathbf{A}\mathbf{x}$ produced by the mapping $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$; the kernel of a matrix \mathbf{A} is the set of all vectors \mathbf{x} which are mapped to $\mathbf{x} \mapsto \mathbf{A}\mathbf{x} = 0$ (Fischer 1984).)

In particular, an over-determined system of linear equations (adjustment problem)

$$\mathbf{A}\mathbf{x} \overset{!}{\approx} \mathbf{b} \tag{A.6}$$

with an $m \times n$ matrix \mathbf{A} and $m > n$, has the least square solution

$$\mathbf{x} = \mathbf{A}^+\mathbf{b} \tag{A.7}$$

in the sense that

$$\|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) = \min \quad . \quad (\text{A.8})$$

From the *singular value decomposition* $\mathbf{A} = \mathbf{UDV}^T$ with the $m \times m$ diagonal matrix \mathbf{D} , the pseudo-inverse of \mathbf{A} can be derived as

$$\mathbf{A}^+ = \mathbf{VD}^+\mathbf{U}^T \quad (\text{A.9})$$

where \mathbf{D}^+ can be derived from \mathbf{D} simply by inverting all diagonal elements which are non-zero and leaving all other matrix elements unchanged. Algorithmically, the singular value decomposition is computed by a series of Householder reflections. The operation is provided by most mathematical software packages (here PVWAVE was employed).

A.2 Orthogonal Projections

We follow the introduction of *orthogonal projections* as given in Kohonen (1977) and Kohonen (1995) :

If \mathcal{L} is a subspace of \mathbb{R}^N , then an arbitrary vector $\mathbf{x} \in \mathbb{R}^N$ can be uniquely decomposed into the sum of two vectors of which one, \mathbf{x}_{\parallel} , belongs to \mathcal{L} and the other, \mathbf{x}_{\perp} , is orthogonal to it:

$$\mathbf{x} = \mathbf{x}_{\parallel} + \mathbf{x}_{\perp} \quad , \quad \text{where } \mathbf{x}_{\parallel} \in \mathcal{L} \text{ and } \mathbf{x}_{\perp} \perp \mathcal{L} \quad . \quad (\text{A.10})$$

Furthermore, the *Projection Theorem* then states:

Of all decompositions of the form $\mathbf{x} = \mathbf{x}' + \mathbf{x}''$, where $\mathbf{x}' \in \mathcal{L}$, the one into orthogonal projections has the property that $\|\mathbf{x}''\|$ is minimum.

Orthogonal projections in three-dimensional space can be visualized as in Fig. A.1.

The decomposition of an arbitrary vector \mathbf{x} into its orthogonal projections $\mathbf{x}_{\parallel} \in \mathcal{L}$ and $\mathbf{x}_{\perp} \in \mathcal{L}^{\perp}$ can always be expressed in terms of linear transformations, whereby there always exists a symmetric projector matrix \mathbf{P} such that $\mathbf{x}_{\parallel} = \mathbf{P}\mathbf{x}$ and $\mathbf{x}_{\perp} = (\mathbf{I} - \mathbf{P})\mathbf{x}$.

The projection results into two orthogonal spaces \mathcal{L} and \mathcal{L}^{\perp} . Two vector spaces are called orthogonal, if every vector of the first space \mathcal{L} is orthogonal to every vector of the other space \mathcal{L}^{\perp} . With the orthogonal projection, the set of vectors \mathbf{x}_{\parallel} are elements of \mathcal{L} , and the set of vectors \mathbf{x}_{\perp} form the *orthogonal complement* \mathcal{L}^{\perp} of \mathcal{L} .

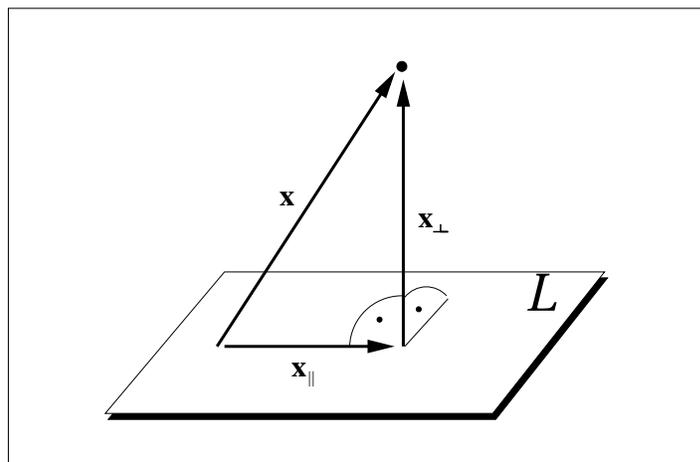


Figure A.1: Illustration of an orthogonal projection in \mathbb{R}^3

A.3 Retrieval of the Reflectance Spectrum from the Class Plane

Given a set \mathbf{X} of observations \mathbf{x} of the same surface under varying illuminations, we first compute the covariance matrix $\mathbf{\Sigma}(\mathbf{X})$. We then extract the two most significant eigenvectors \mathbf{a}_1 and \mathbf{a}_2 of $\mathbf{\Sigma}$ and stack them as column vectors in $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]$. The matrix \mathbf{A} characterizes the class plane (Eq. 3.11 on page 40). Furthermore we need the illumination plane $\mathbf{E} = [\mathbf{n}, \mathbf{m}]$ (Eq. 3.5 on page 39), formed by the two vectors \mathbf{n} and \mathbf{m} or linear combinations thereof (*e.g.* extracted as principal components of the observed class plane of a white reflectance standard).

Then it is possible to express the product of the unknown reflectance vector $\hat{\mathbf{r}}$ with the known illumination vectors \mathbf{n} and \mathbf{m} as linear combinations of the observed eigenvectors \mathbf{a}_1 and \mathbf{a}_2 :

$$\hat{r}_i n_i = c_1 a_{i1} + c_2 a_{i2} \quad (\text{A.11})$$

$$\hat{r}_i m_i = c'_1 a_{i1} + c'_2 a_{i2} \quad (\text{A.12})$$

This relation holds since the spectra resulting from reflection of direct and diffuse illumination should each be in the class plane \mathbf{A} .

After multiplication of Eq. A.11 and A.12 with m_i and n_i respectively we can demand equality of the two equations:

$$\hat{r}_i n_i m_i = (c_1 a_{i1} + c_2 a_{i2}) m_i \stackrel{!}{=} (c'_1 a_{i1} + c'_2 a_{i2}) n_i \quad . \quad (\text{A.13})$$

In practise, due to noise this will not be the case simultaneously for all N equations. Therefore, in matrix/vector notation, we rather need to look for the smallest possible deviation of both sides of the N equations:

$$\begin{bmatrix} \vdots & \vdots \\ a_{i1} m_i & a_{i2} m_i \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} \vdots & \vdots \\ a_{i1} n_i & a_{i2} n_i \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} c'_1 \\ c'_2 \end{bmatrix} \quad (\text{A.14})$$

$$\begin{bmatrix} \vdots & \vdots \\ a_{i1} m_i & a_{i2} m_i \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} \vdots & \vdots \\ a_{i1} n_i & a_{i2} n_i \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} c'_1 \\ c'_2 \end{bmatrix} \stackrel{!}{=} \min \quad (\text{A.15})$$

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i1} m_i & a_{i2} m_i & a_{i1} n_i & a_{i2} n_i \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ -c'_1 \\ -c'_2 \end{bmatrix} \stackrel{!}{=} \min \quad (\text{A.16})$$

where minimal deviation in the least squares sense is defined as

$$\|\mathbf{Bc}\|^2 \stackrel{!}{=} \min \quad (\text{A.17})$$

$$\text{with } \mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i1} m_i & a_{i2} m_i & a_{i1} n_i & a_{i2} n_i \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (\text{A.18})$$

and with a coefficient vector $\mathbf{c} = [c_1, c_2, c_3, c_4]^T \in \mathbb{R}^4$, and $c_3 = c'_1$ and $c_4 = c'_2$. Obviously this expression has a trivial solution $\mathbf{c} = \mathbf{0}$, and another non-trivial zero solution only if $\det \mathbf{B} = 0$, *i.e.*, the planes are undistinguishable. Following a common method of linear algebra (Harris 1975), we exclude the trivial solution by constraining \mathbf{c} to a non-vanishing magnitude $\|\mathbf{c}\|^2 = \mathbf{c}^T \mathbf{c} = 1$. In the formalism of Lagrange multipliers, the constraint is formulated as $(1 - \mathbf{c}^T \mathbf{c}) \stackrel{!}{=} 0$. Then the minimization problem is solved in presence of the constraint by virtue of the *constrained extremal condition*:

$$\mathbf{c}^T \mathbf{B}^T \mathbf{B} \mathbf{c} + \kappa(1 - \mathbf{c}^T \mathbf{c}) \stackrel{!}{=} \min \quad (\text{A.19})$$

with a Lagrange multiplier κ . Now we demand a vanishing partial derivation:

$$\frac{\partial}{\partial \mathbf{c}} \left(\mathbf{c}^T \mathbf{B}^T \mathbf{B} \mathbf{c} + \kappa(1 - \mathbf{c}^T \mathbf{c}) \right) \stackrel{!}{=} 0 \quad (\text{A.20})$$

$$\Rightarrow \quad \mathbf{B}^T \mathbf{B} \mathbf{c} - \kappa \mathbf{c} \stackrel{!}{=} 0 \quad (\text{A.21})$$

$$(\mathbf{B}^T \mathbf{B} - \kappa \mathbf{I}) \stackrel{!}{=} 0 \quad (\text{A.22})$$

which is the eigenvalue equation for the 4×4 matrix $\mathbf{B}^T \mathbf{B}$. Provided a perfect fit, the smallest eigenvalue is ideally 0. We take the eigenvector \mathbf{e}_{\min} corresponding to the smallest eigenvalue in order to get two estimations $\hat{\mathbf{r}} \approx \hat{\mathbf{r}}'$:

$$\hat{r}_i = \left([B_{i1} B_{i2}] \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right) / (n_i m_i) \quad \text{and} \quad \hat{r}'_i = \left([B_{i3} B_{i4}] \begin{bmatrix} c_3 \\ c_4 \end{bmatrix} \right) / (n_i m_i) \quad (\text{A.23})$$

The fitted reflectance is then found as the mean of the estimations $\bar{\mathbf{r}} = \frac{1}{2}(\hat{\mathbf{r}} + \hat{\mathbf{r}}')$. Moreover, the goodness of the fit can be assessed by the *condition* of the matrix $\mathbf{B}^T \mathbf{B}$. The condition of a matrix is defined as the ratio of the largest eigenvalue over to the smallest non-vanishing eigenvalue (Hämmerlin & Hoffmann 1994).

In principle, only two observations \mathbf{x} and \mathbf{x}' of the same surface under different orientation are sufficient to determine a two-plane in the feature space and thus to retrieve the normalized reflectance spectrum $\hat{\mathbf{r}}$. To this aim the two observations \mathbf{x} and \mathbf{x}' are substituted as the column vectors of the class plane spanning matrix \mathbf{A} . Note however, that not any two observations \mathbf{x} and \mathbf{x}' will allow a satisfying fit of \mathbf{r} ; in order to define a two-plane they must be linearly independent. The goodness of the linear independence can again be tested by the condition of matrix \mathbf{A} .

A.4 Minimal Distance Between Two Class Planes

Misclassification of an observed spectrum \mathbf{x} into class a instead of a' can happen because of noise. An erroneous classification is the more likely the smaller the distance between the class planes \mathbf{A} and \mathbf{A}' is. In order to evaluate the probability of correct or erroneous classification we want to know the minimal distance of any spectrum from class a (in plane \mathbf{A}) to a spectrum of another class a' (in plane \mathbf{A}'), and vice versa, in absence of noise. To this aim we stack the column vectors of \mathbf{A} and \mathbf{A}' into a $N \times 4$ matrix

$$\mathbf{D} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ A_{i1} & A_{i2} & A'_{i1} & A'_{i2} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} . \quad (\text{A.24})$$

Demanding the least possible distance between \mathbf{A} and \mathbf{A}' means no less but demanding that four coefficients $\mathbf{c} = [c_1, c_2, c_3, c_4]^T$ yield a linear combination of the four column vectors in Eq. A.24 which add up to a vector of minimal magnitude (compare Eq. A.14 through Eq. A.16 on page 180).

$$d^2 = \|\mathbf{D}\mathbf{c}\|^2 \stackrel{!}{=} \min \quad (\text{A.25})$$

$$\mathbf{c}^T \mathbf{D}^T \mathbf{D} \mathbf{c} \stackrel{!}{=} \min \quad (\text{A.26})$$

with a coefficient vector $\mathbf{c} = [c_1, c_2, c_3, c_4]^T$. Obviously this expression has a trivial solution $\mathbf{c} = \mathbf{0}$, and another non-trivial zero solution only if $\det \mathbf{D} = 0$, *i.e.*, the planes are undistinguishable. Following a common method of linear algebra (Harris 1975), we exclude the trivial solution by constraining \mathbf{c} to a non-vanishing magnitude $\|\mathbf{c}\|^2 = \mathbf{c}^T \mathbf{c} = 1$. Thus we formulate the constrained extremal condition

$$\mathbf{c}^T \mathbf{D}^T \mathbf{D} \mathbf{c} + \kappa(1 - \mathbf{c}^T \mathbf{c}) \stackrel{!}{=} \min \quad (\text{A.27})$$

where κ is a Lagrange multiplier. Then we demand a vanishing partial derivative:

$$\frac{\partial}{\partial \mathbf{c}} \Rightarrow \mathbf{D}^T \mathbf{D} \mathbf{c} - \kappa \mathbf{c} \stackrel{!}{=} \mathbf{0} \quad (\text{A.28})$$

$$(\mathbf{D}^T \mathbf{D} - \kappa \mathbf{I}) \mathbf{c} \stackrel{!}{=} \mathbf{0} \quad (\text{A.29})$$

which is the eigenvalue equation for the 4×4 matrix $\mathbf{D}^T \mathbf{D}$. If there are points which are common to both planes spanned by \mathbf{A} and \mathbf{A}' , then the four column vectors (Eq. A.24) cannot be linearly independent and the smallest eigenvalue of $\mathbf{D}^T \mathbf{D}$ will vanish. For very ‘similar’ class planes the smallest eigenvalue will not vanish but will be ‘relatively small’. This can be evaluated by virtue of the *condition* of the matrix $\mathbf{D}^T \mathbf{D}$ which is defined as the ratio of the largest eigenvalue over to the smallest non-vanishing eigenvalue (Hämmerlin & Hoffmann 1994). We can conclude that the smallest eigenvalue of $\mathbf{D}^T \mathbf{D}$ will vanish if the class planes are indistinguishable, and that the higher the condition of the symmetric matrix $\mathbf{D}^T \mathbf{D}$ is, the more the class planes are separated in the feature space and the less likely is a misclassification between the classes.

A.5 Retrieval of the Path Radiance Spectrum from Two Class Planes

All class planes intersect at the origin, *i.e.*, at vanishing illumination, provided the path radiance was properly subtracted beforehand. Vice versa, if the planes do not intersect at the origin, we can assess any residual path radiance from the point of intersection of at least two given class centers \mathbf{m}_a and $\mathbf{m}_{a'}$ and their respective plane spanning vectors in \mathbf{A} and \mathbf{A}' . Their point of intersection then is the one of highest proximity:

$$d^2 = \|\mathbf{D}\mathbf{c} + \mathbf{d}\|^2 = \min \quad (\text{A.30})$$

where \mathbf{D} and \mathbf{c} are as in (A.24) and $\mathbf{d} = \mathbf{m}_a - \mathbf{m}_{a'}$. This adjustment problem is according to Appendix A.1 on page 177 solved by

$$\mathbf{c}_{\min} = \mathbf{D}^+ \mathbf{d} \quad (\text{A.31})$$

where \mathbf{D}^+ is the pseudo-inverse of \mathbf{D} , and $\mathbf{c}_{\min} = (c_1, c_2, c_3, c_4)^T$. So we have two approximations to the true origin, *i.e.*, the vector of path radiances:

$$\mathbf{o} = \mathbf{m}_a + \mathbf{A} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad \text{and} \quad \mathbf{o}' = \mathbf{m}_{a'} - \mathbf{A}' \begin{bmatrix} c_3 \\ c_4 \end{bmatrix} \quad (\text{A.32})$$

and may take the centroid $\bar{\mathbf{o}} = \frac{1}{2}(\mathbf{o} - \mathbf{o}')$ between the two proxima as the estimated point of intersection. The coordinates of this intersection point then represents the path radiances in all spectral bands.

Appendix B

Spectra from *in situ* Measurements

B.1 Illumination Variability Eigenvectors

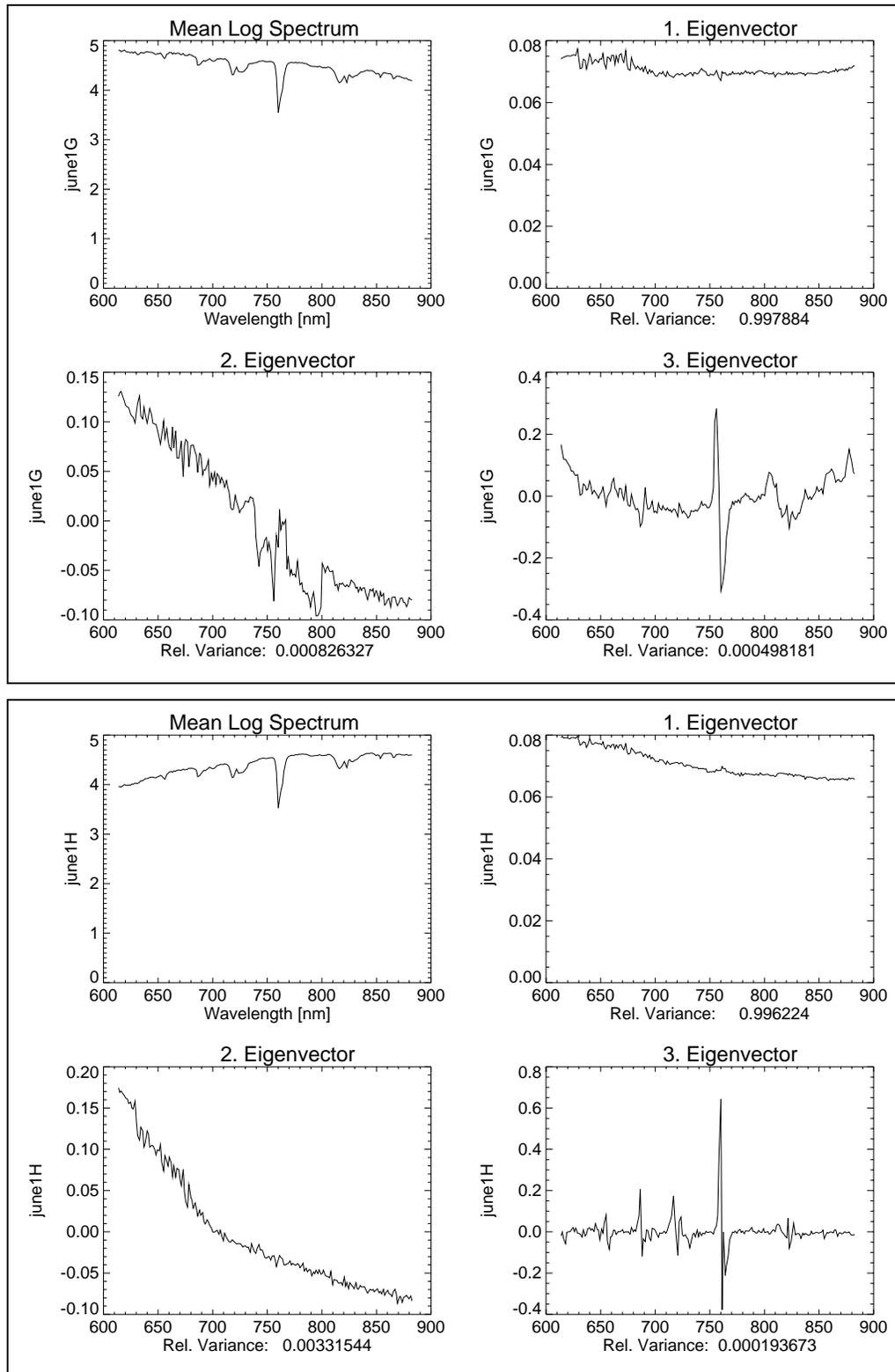


Figure B.1: Mean logarithmic spectrum $\langle \ln \mathbf{x} \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

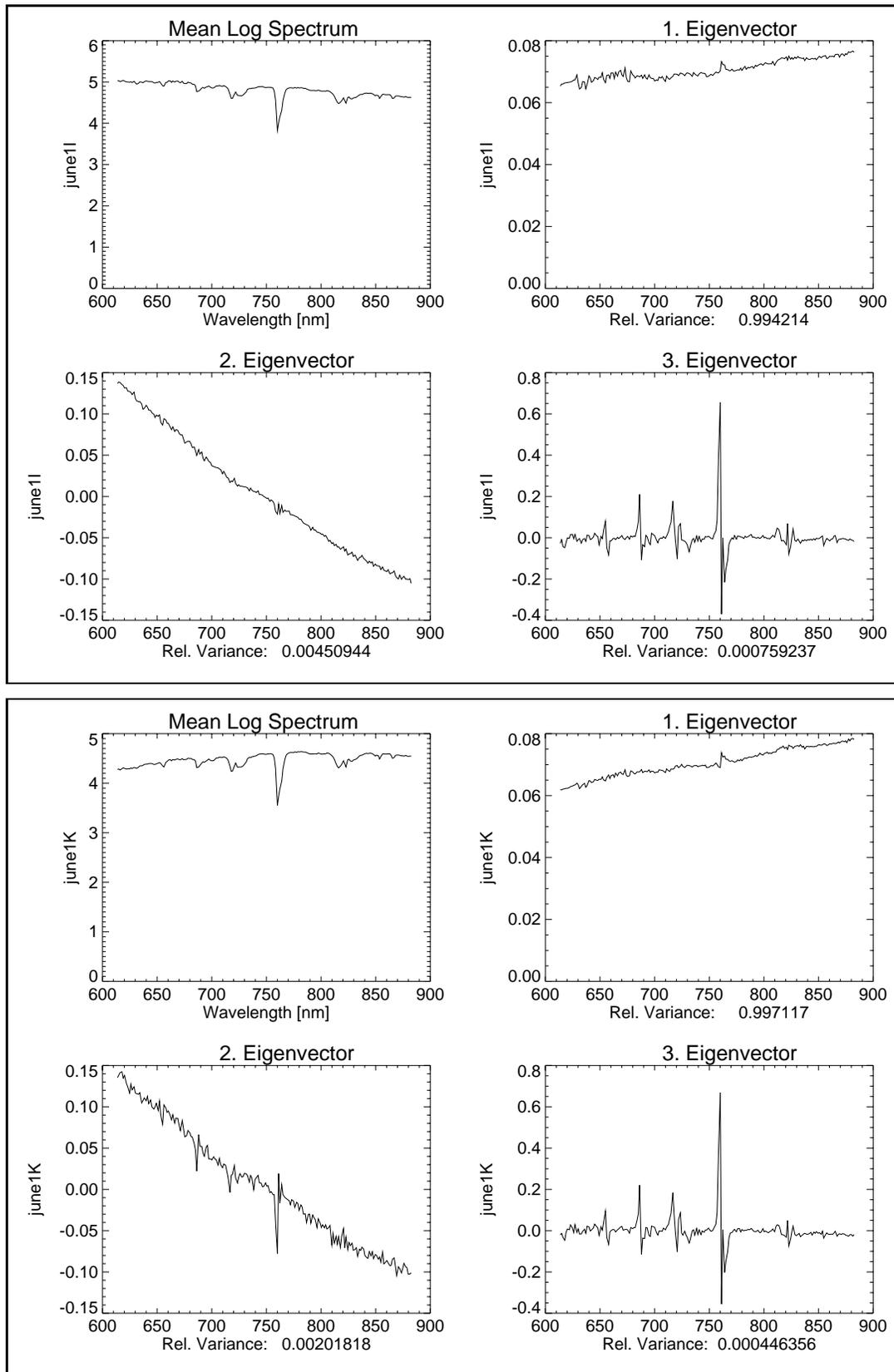


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

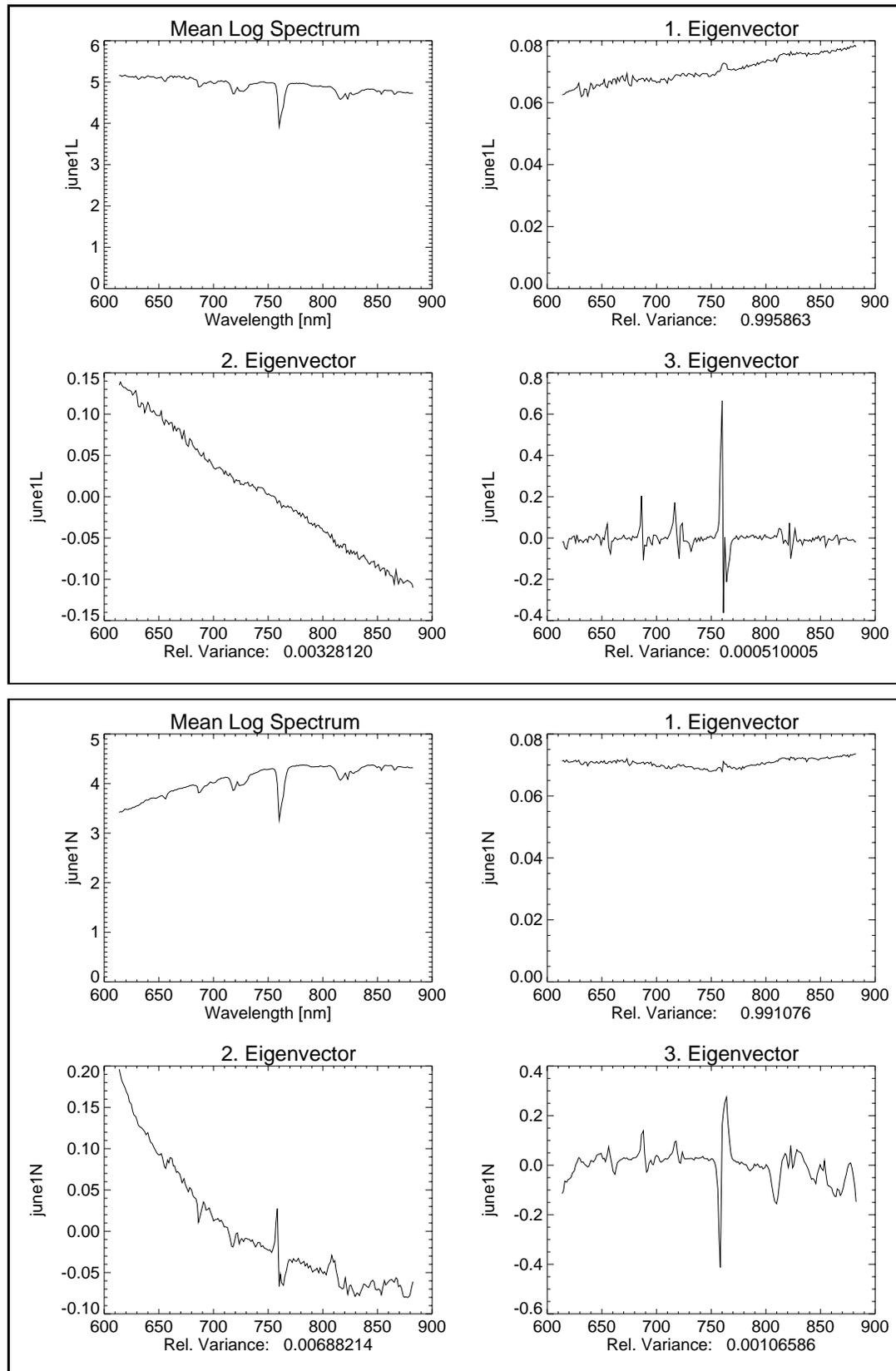


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

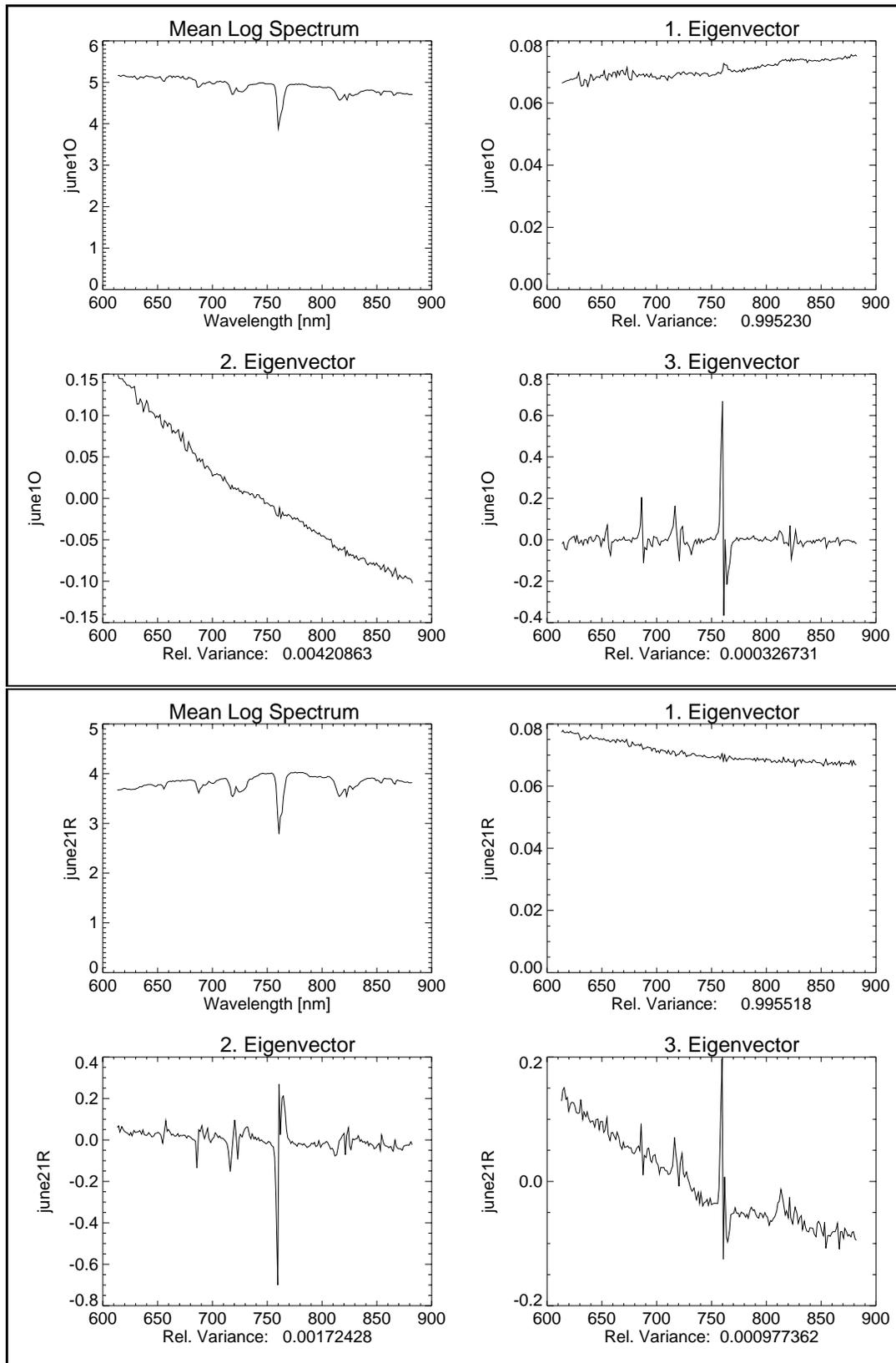


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

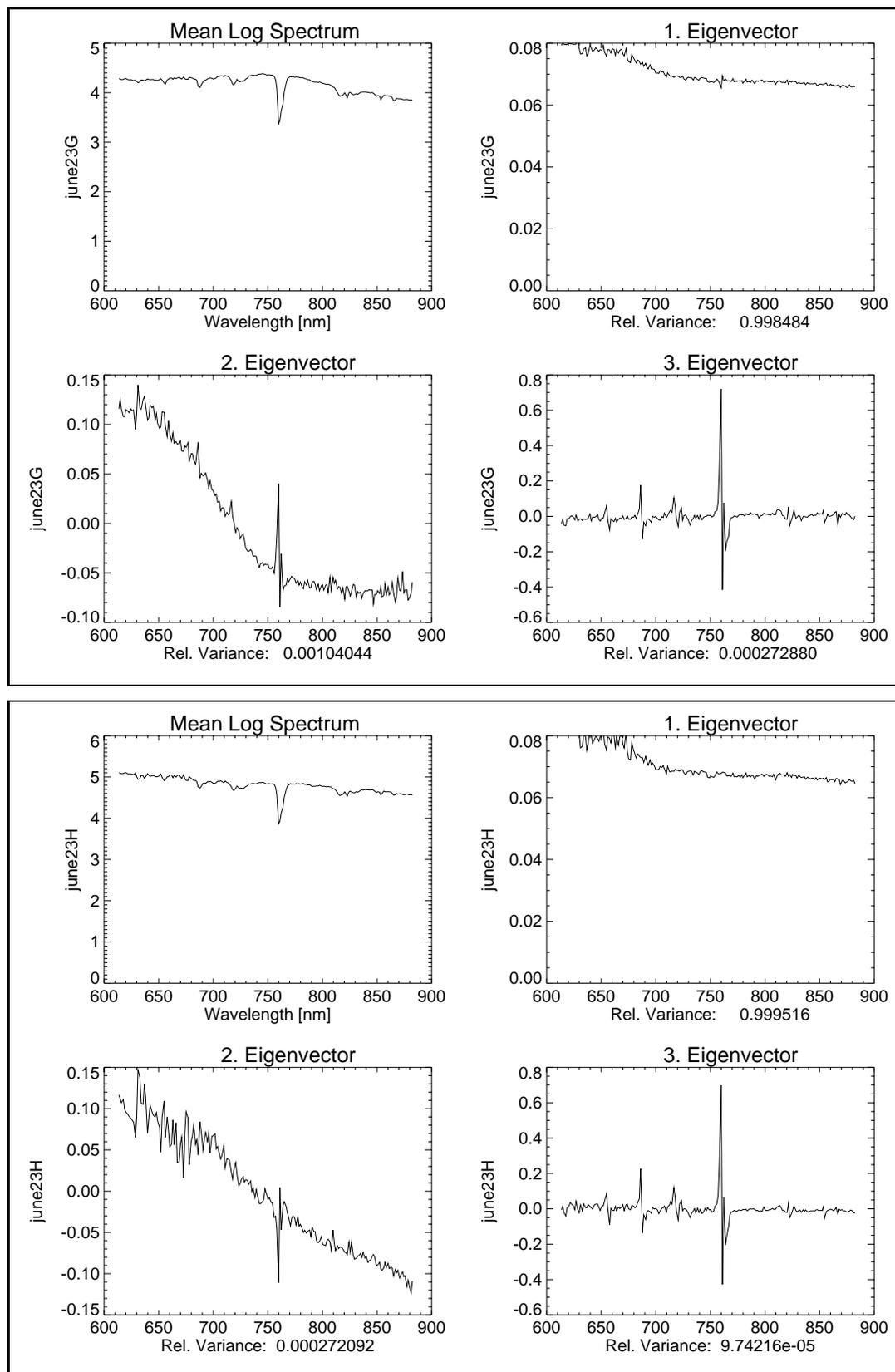


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

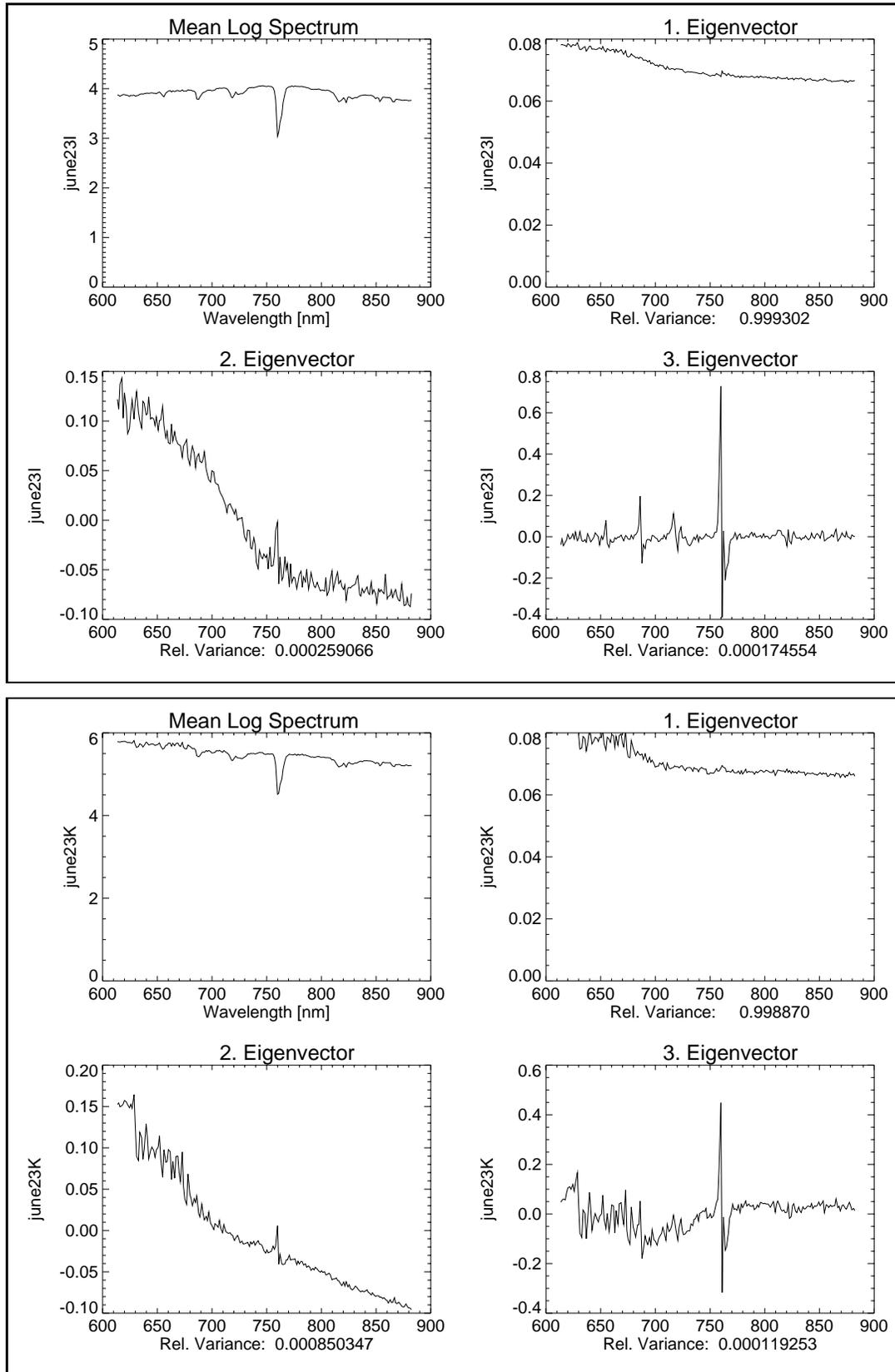


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

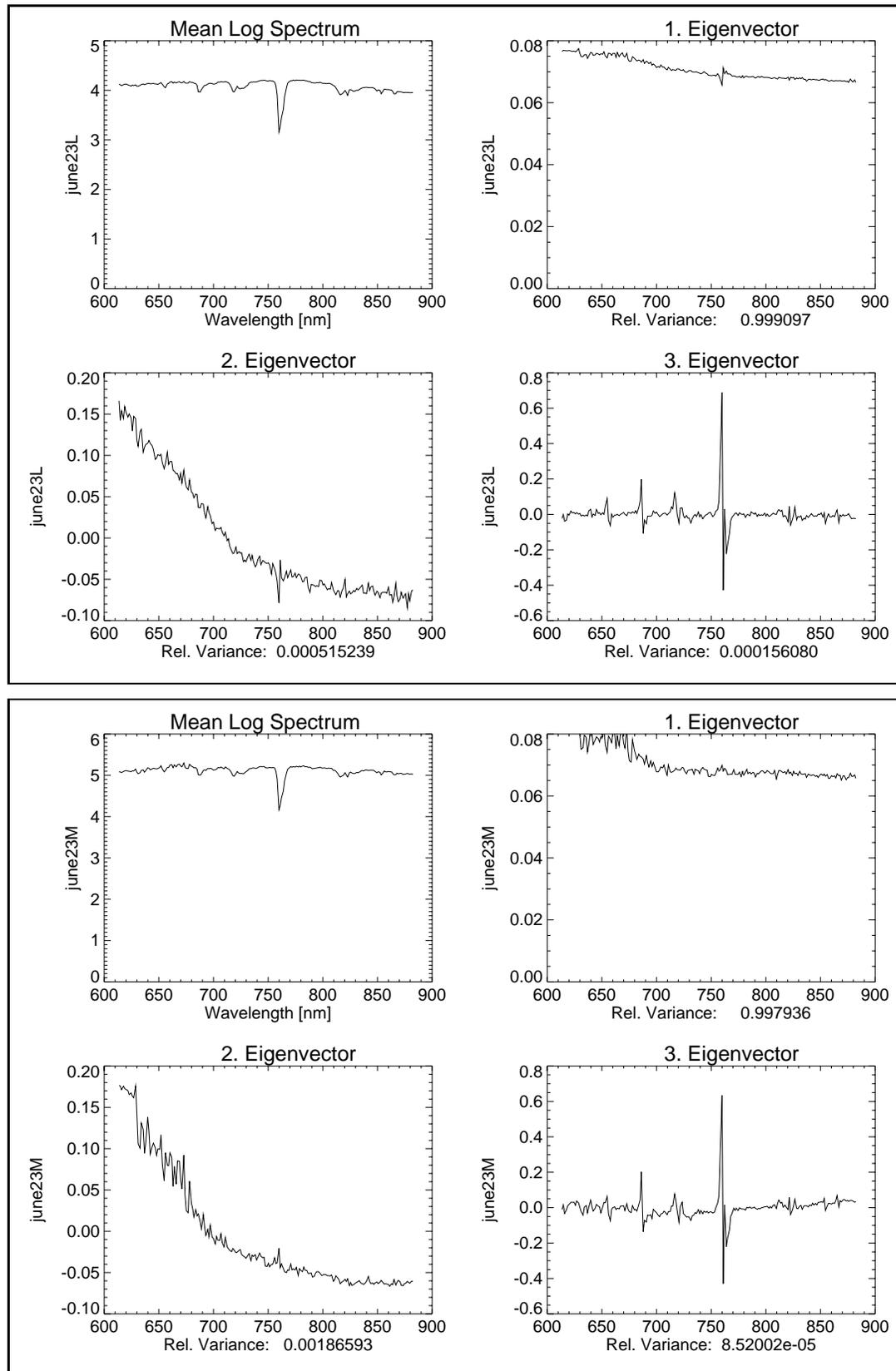


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

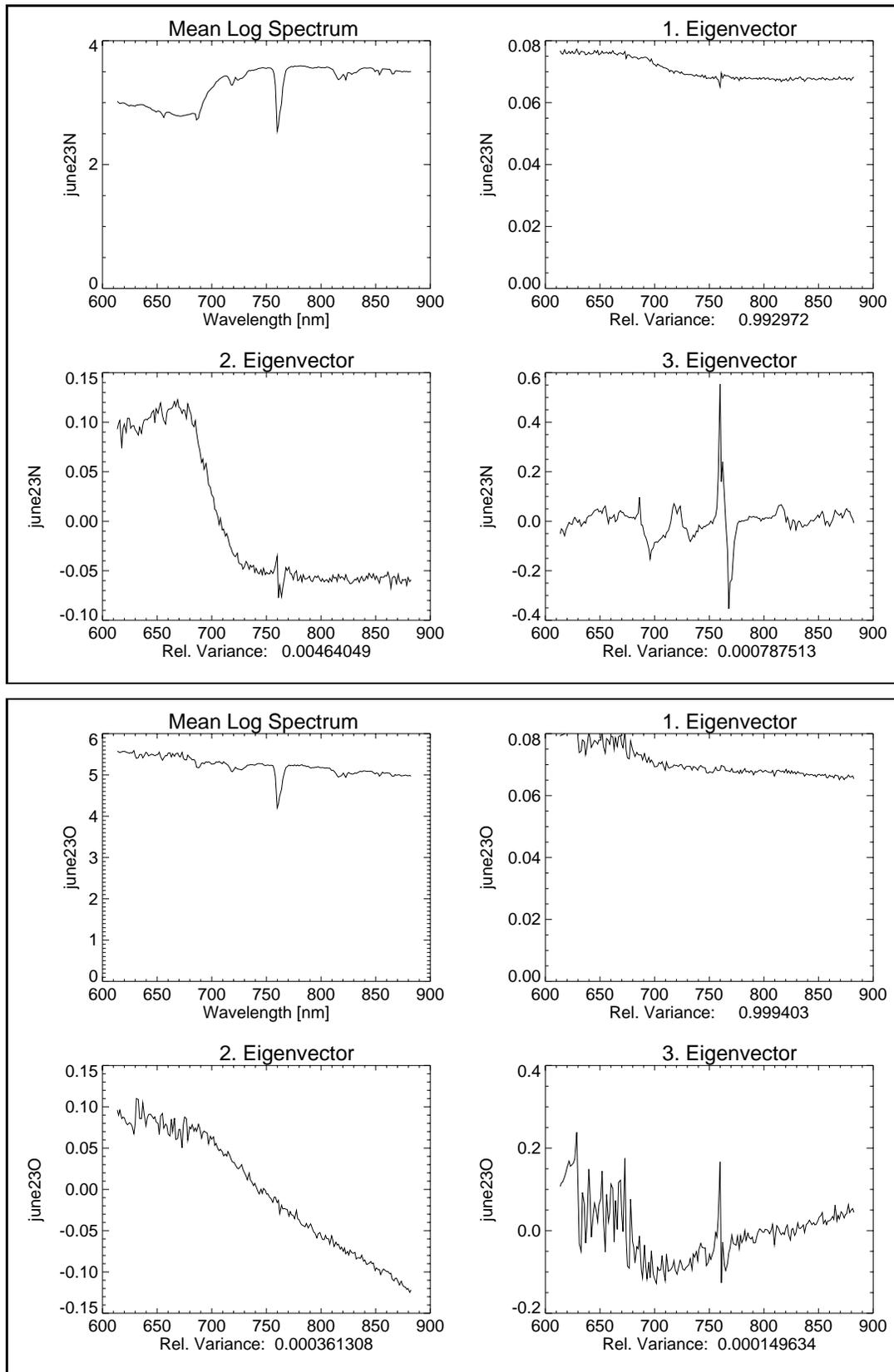


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

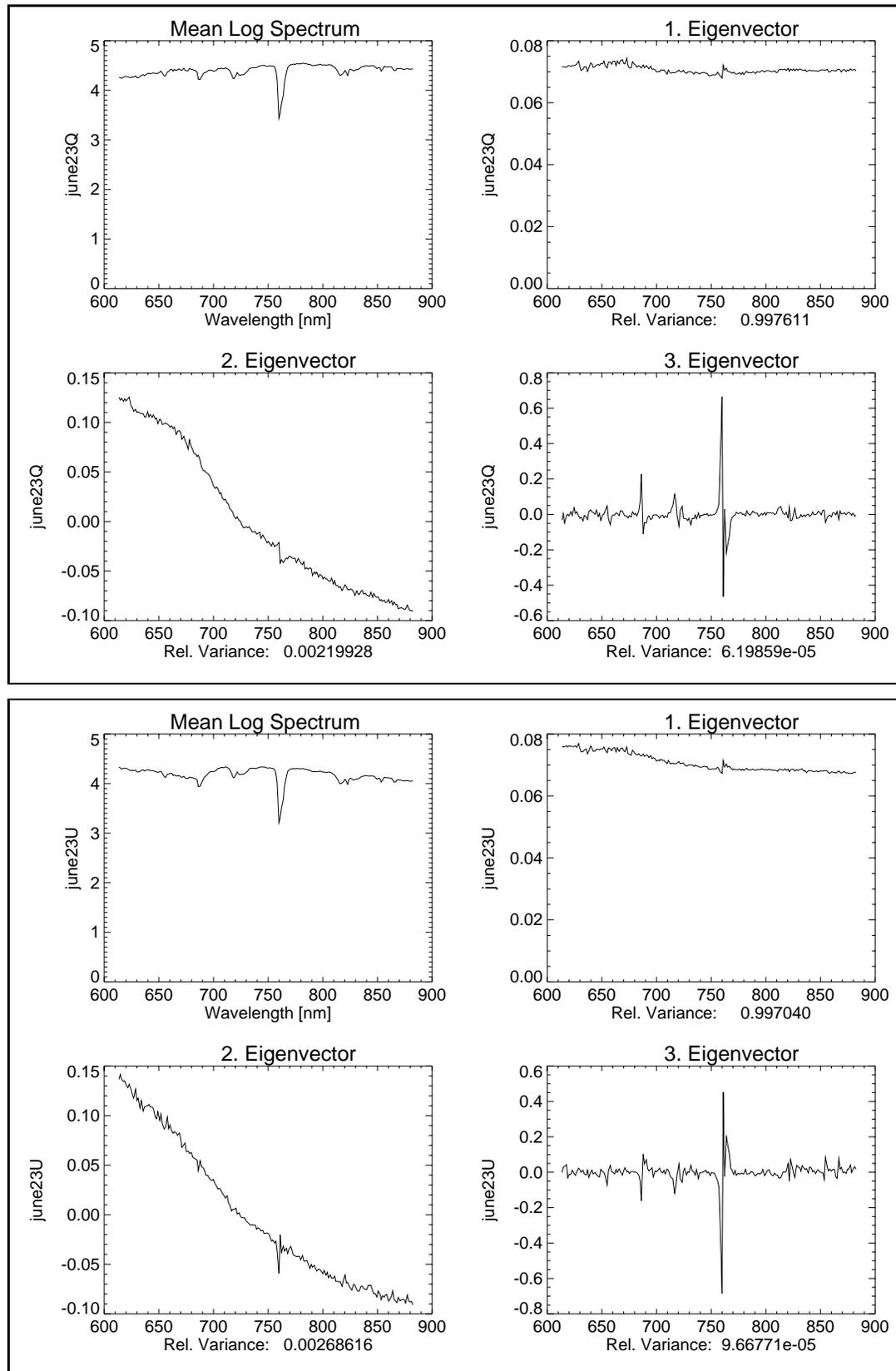


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

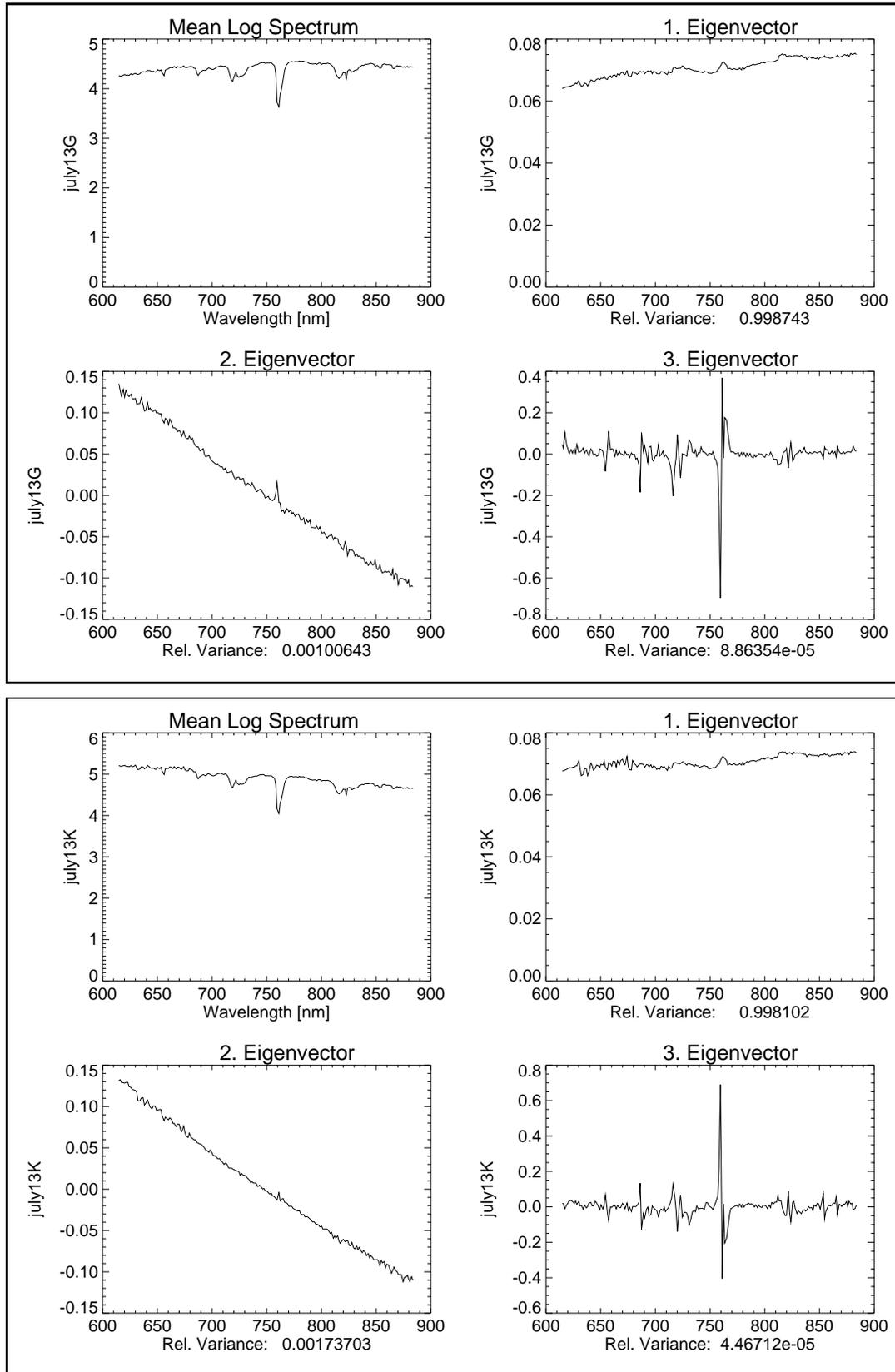


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

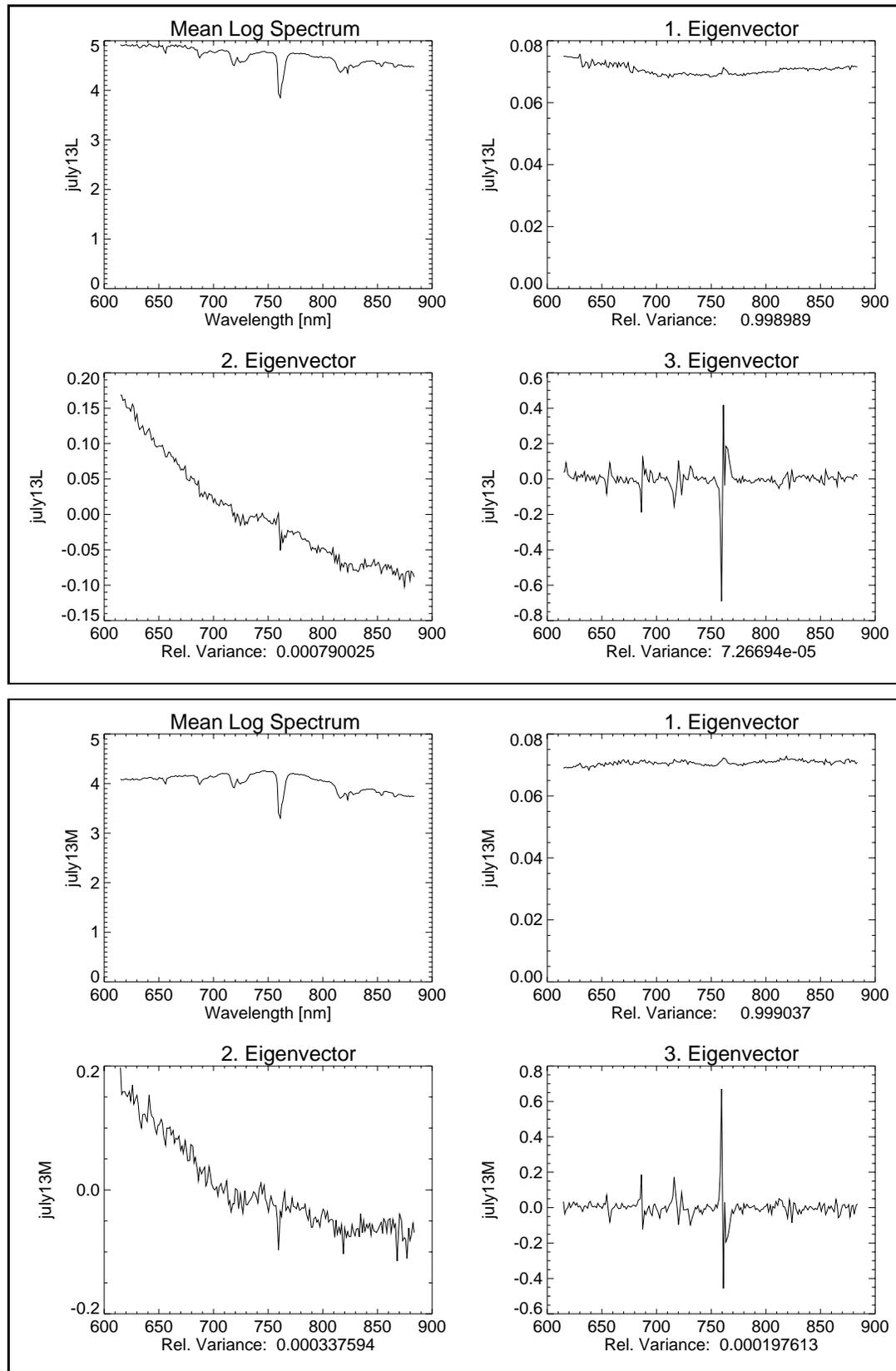


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

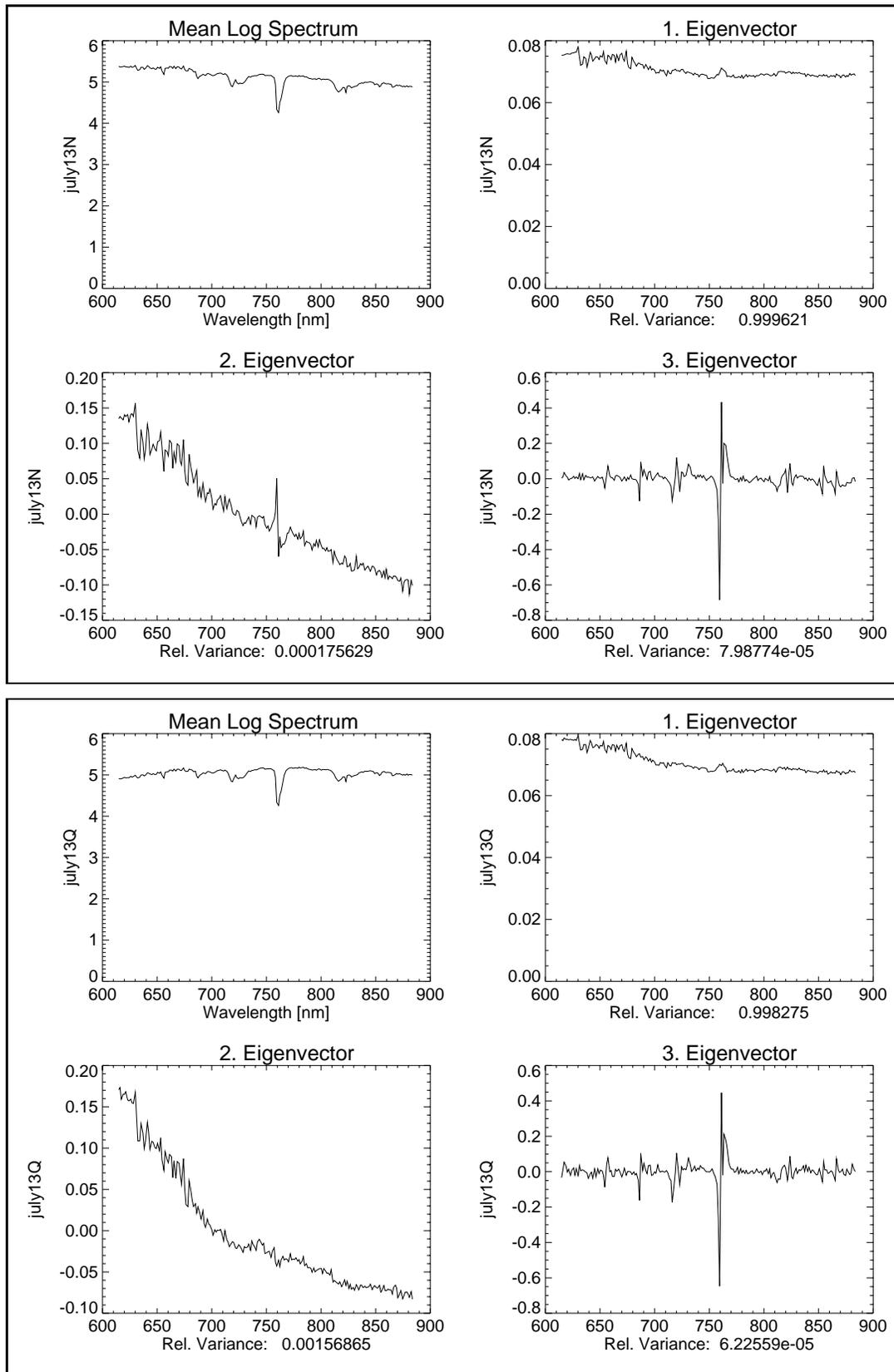


Figure B.1 (cont'd): Mean logarithmic spectrum $\langle \ln x \rangle$ and illumination eigenvectors of the covariance matrix of the logarithmic spectra observed under arbitrary angles. (The samples are denoted by the ordinate titles and specified in Table 5.2 on page 107.)

B.2 Eigenvectors of Residual Illumination Variability

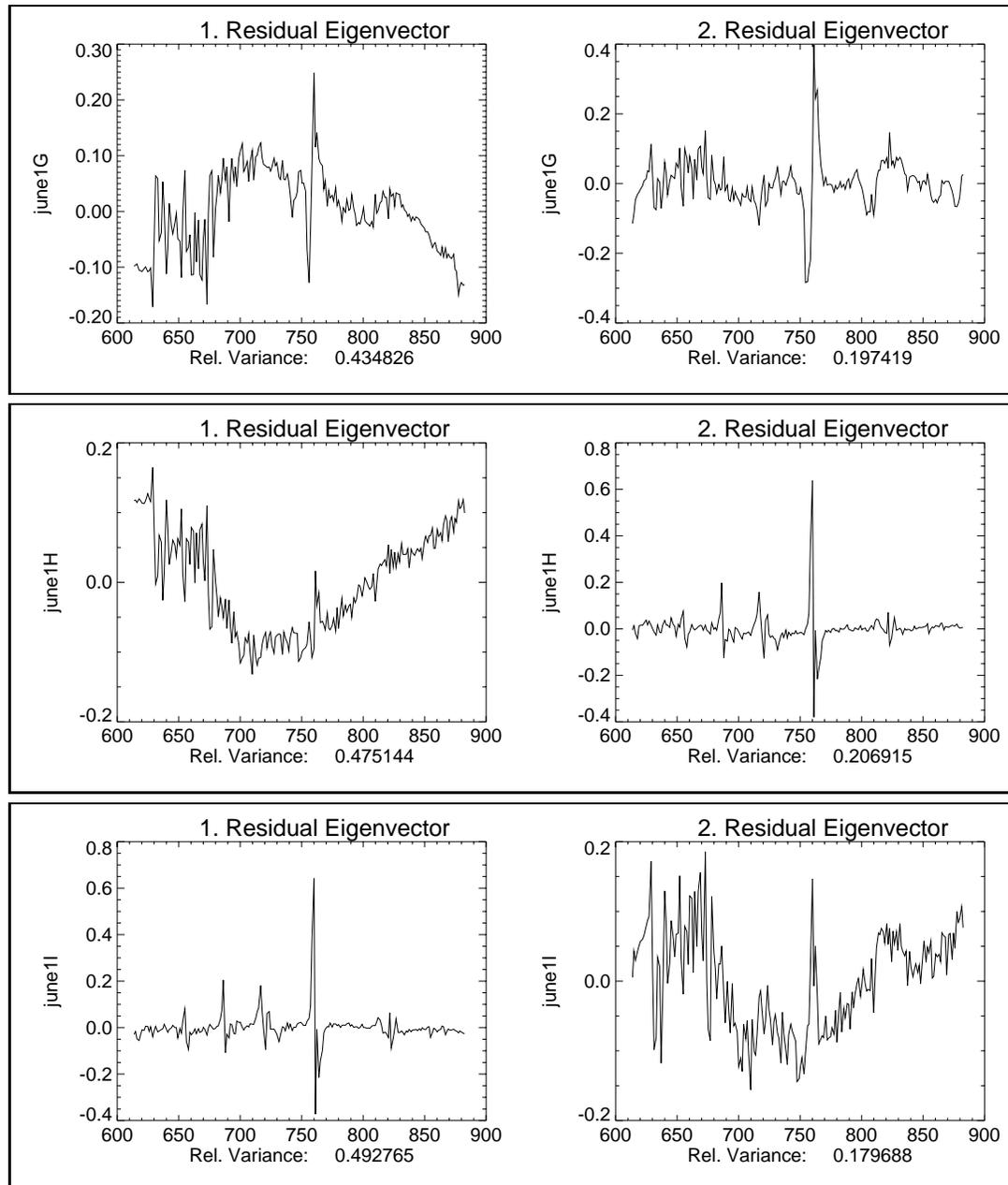


Figure B.2: Principal Component Analysis of the residual variance after brightness and color filtering (filter rank $K = 2$).

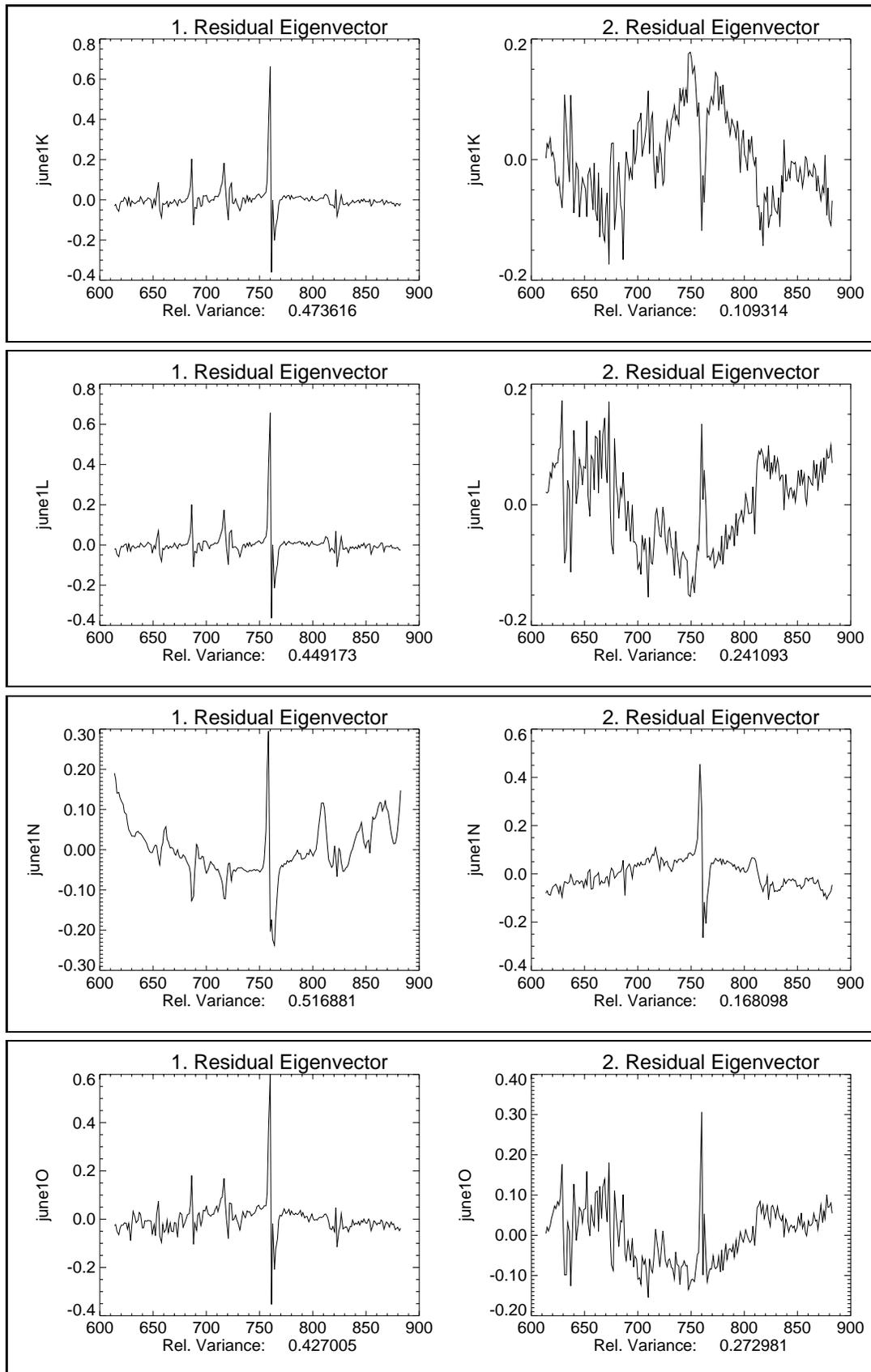


Figure B.2 (cont'd): Principal component analysis of the residual variance after brightness and color filtering (filter rank $K = 2$).

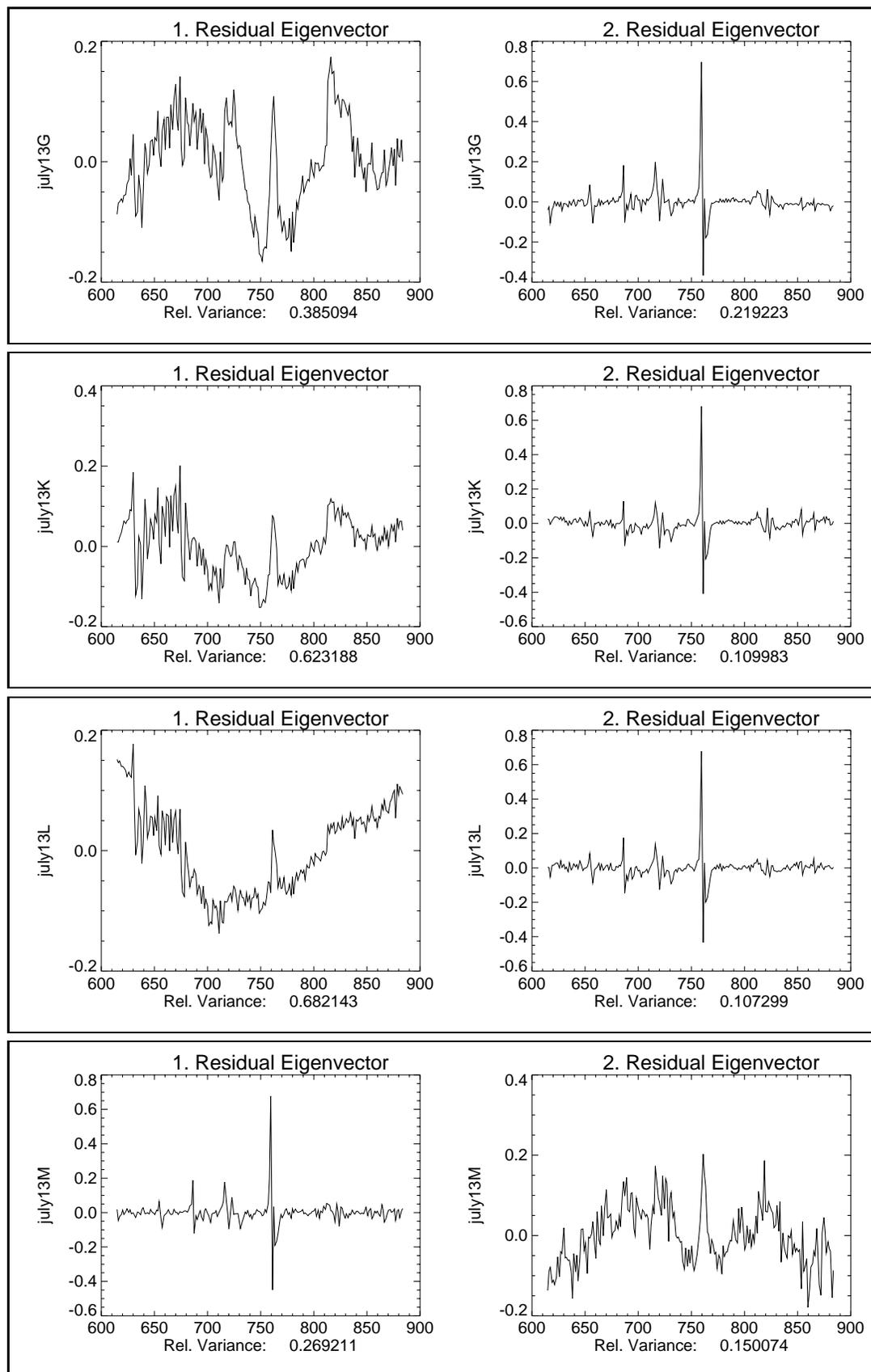


Figure B.2 (cont'd): Principal component analysis of the residual variance after brightness and color filtering (filter rank $K = 2$).

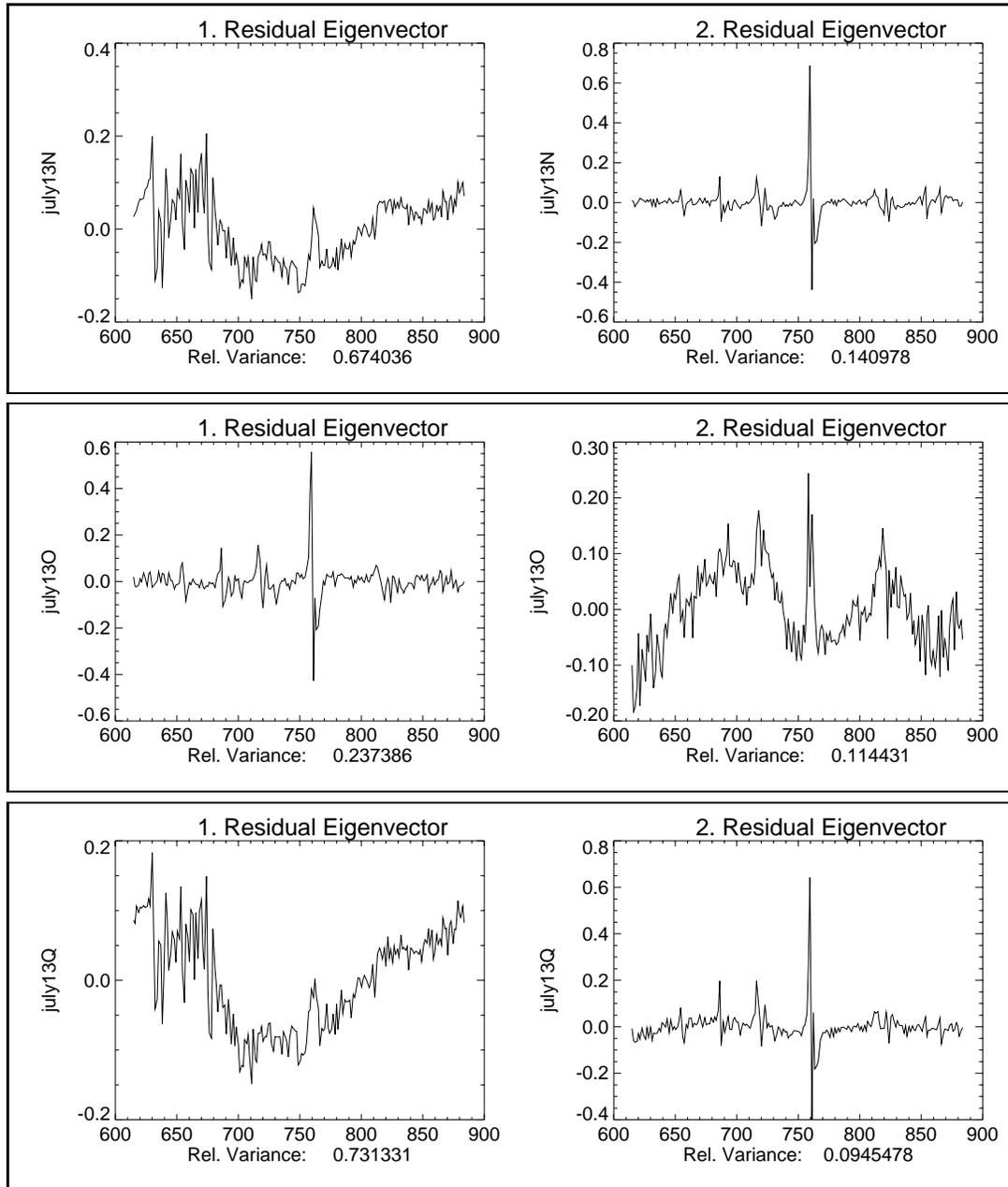


Figure B.2 (cont'd): Principal component analysis of the residual variance after brightness and color filtering (filter rank $K = 2$).

Appendix C

Multispectral Imagery

C.1 Sample Pairs from Scene ‘Gewerbegebiet 1992’

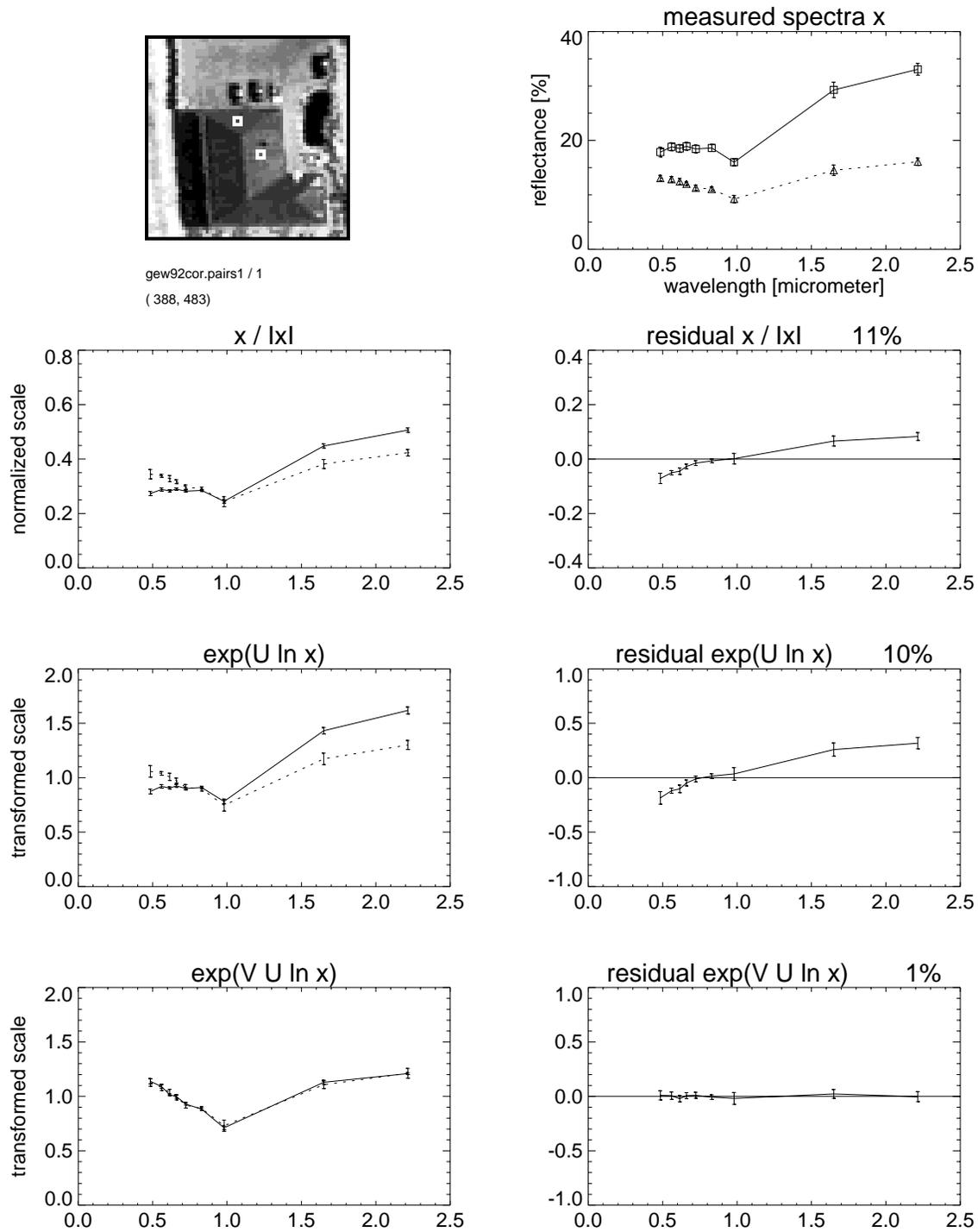


Figure C.1: Spectral sample pair from scene ‘Gewerbegebiet 1992’ (Fig. 6.2 on page 120).

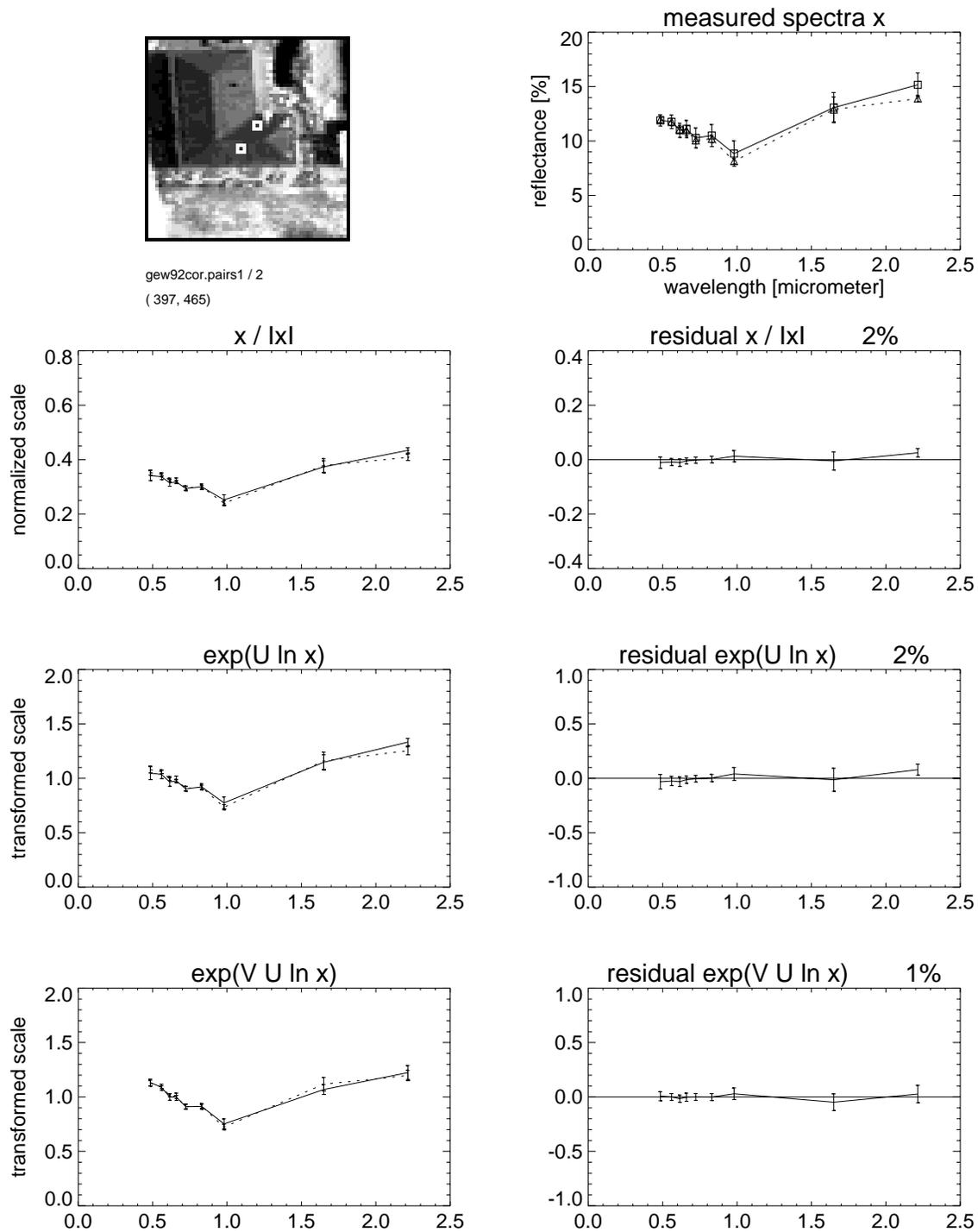


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

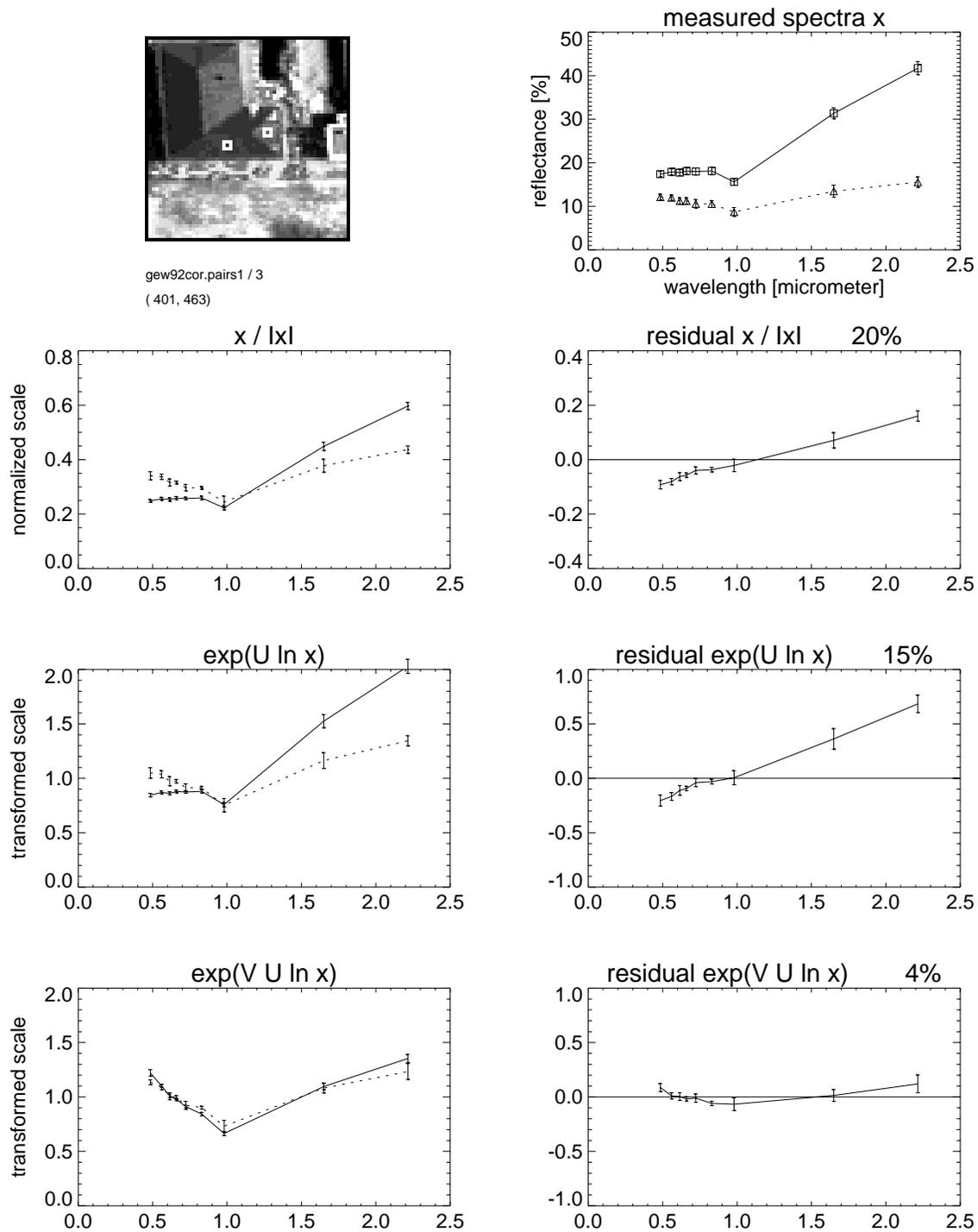


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

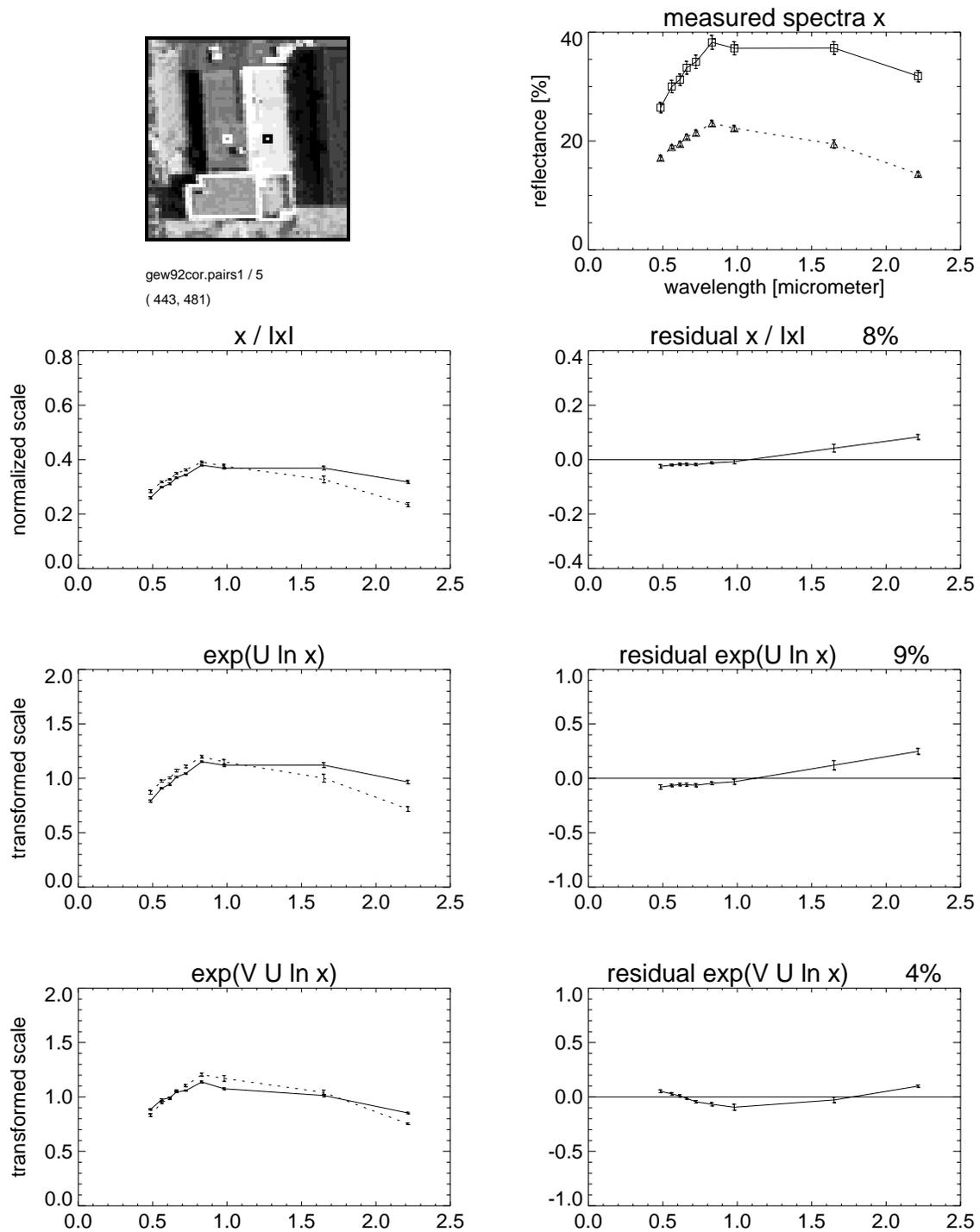


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

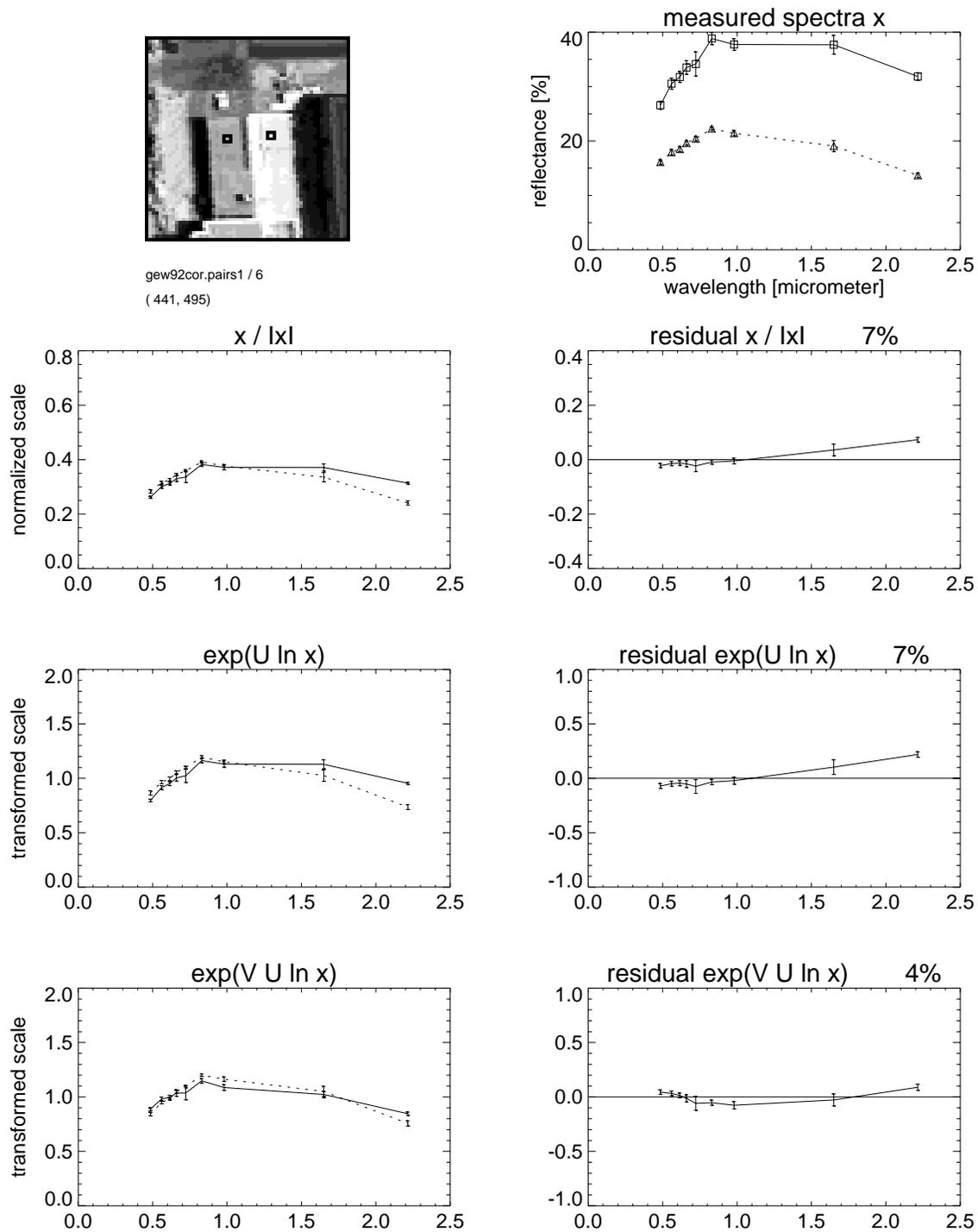


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

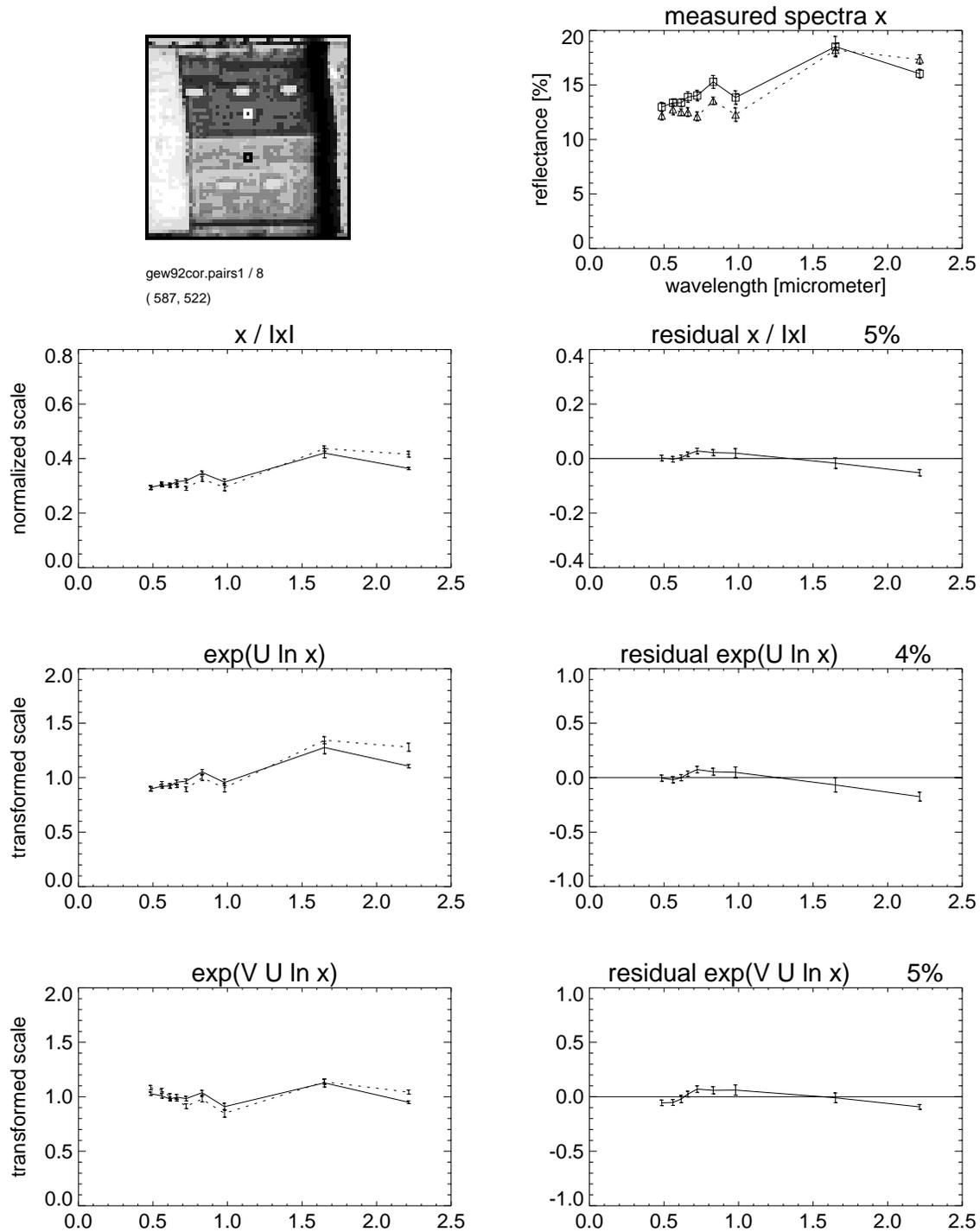


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

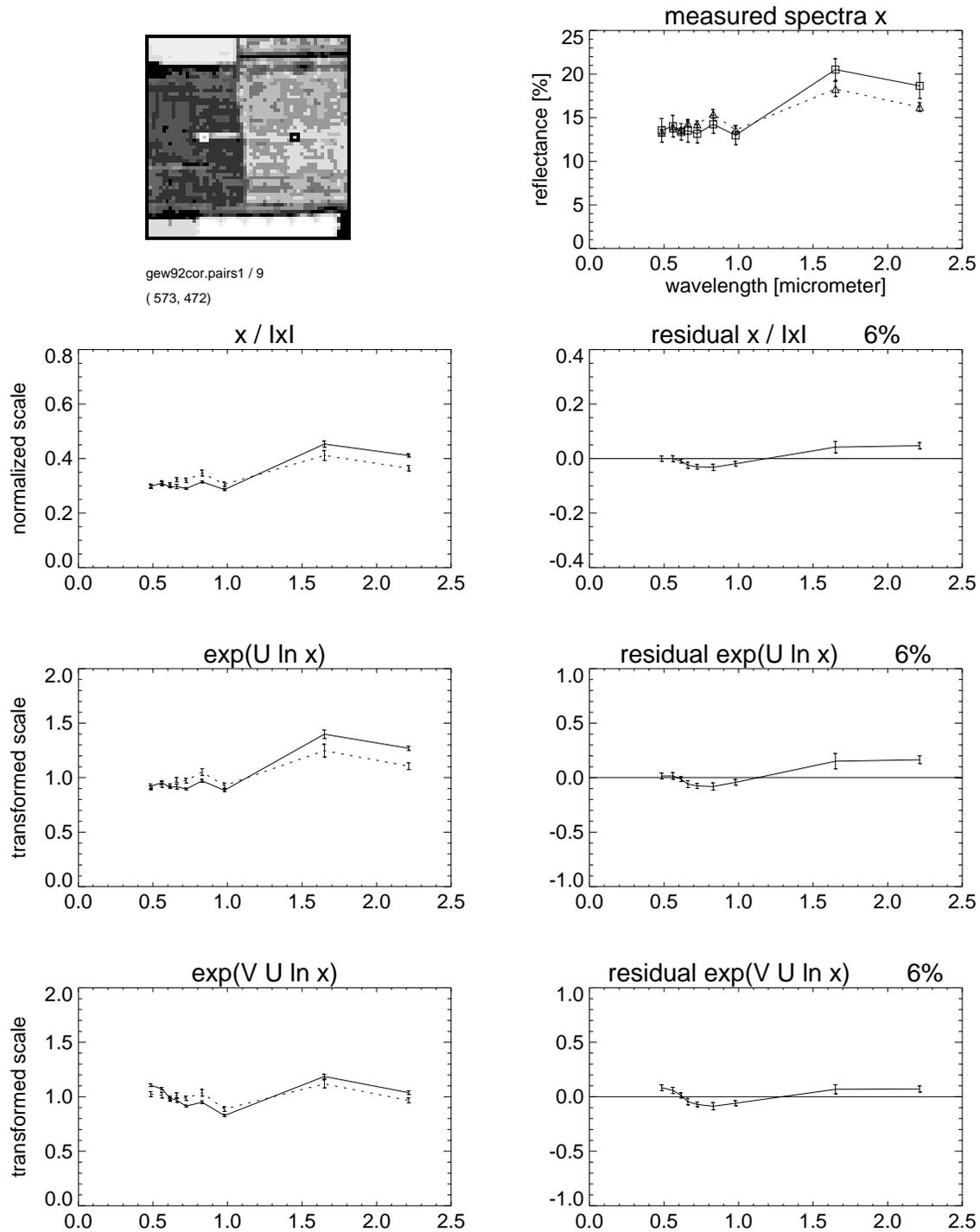


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

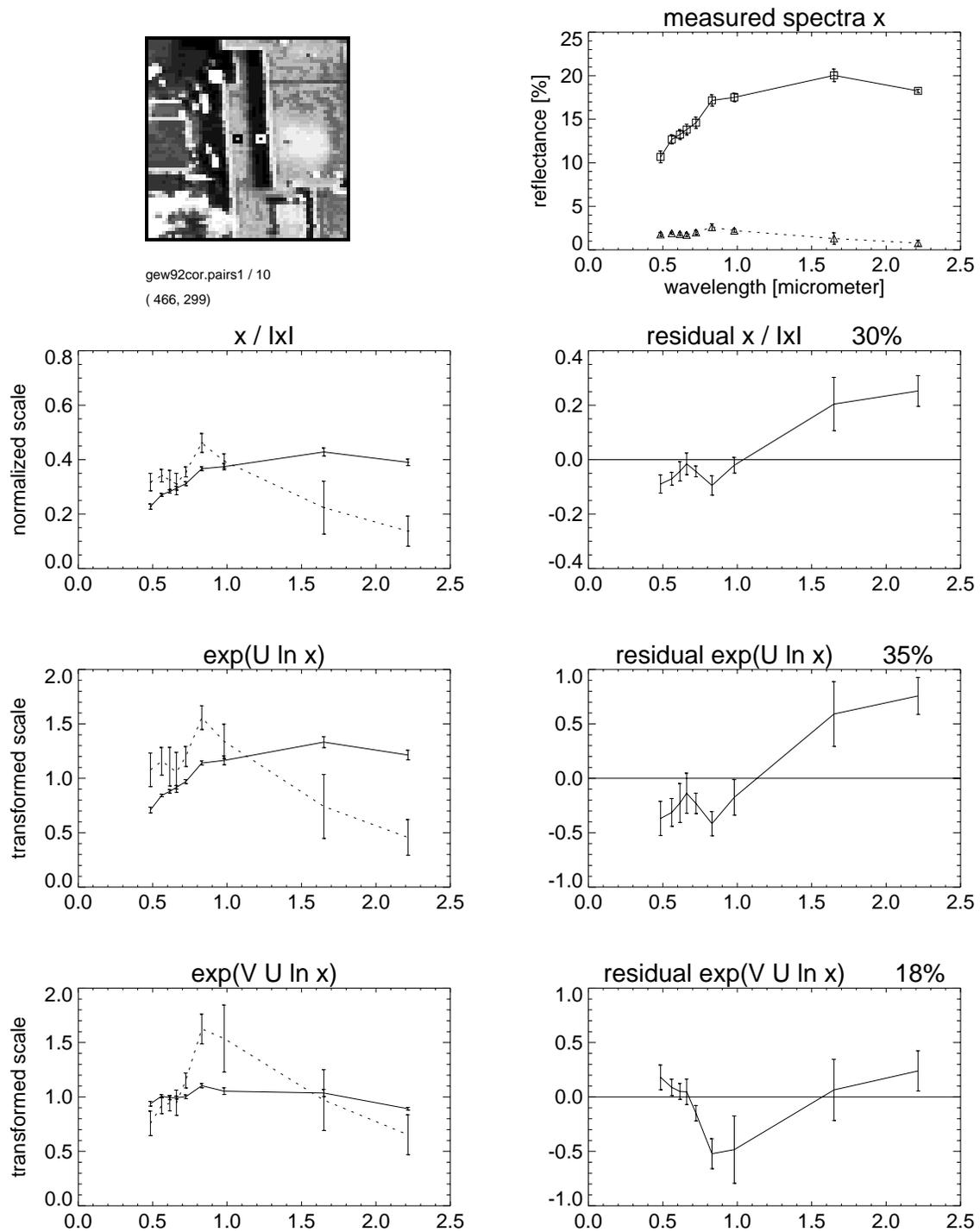


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

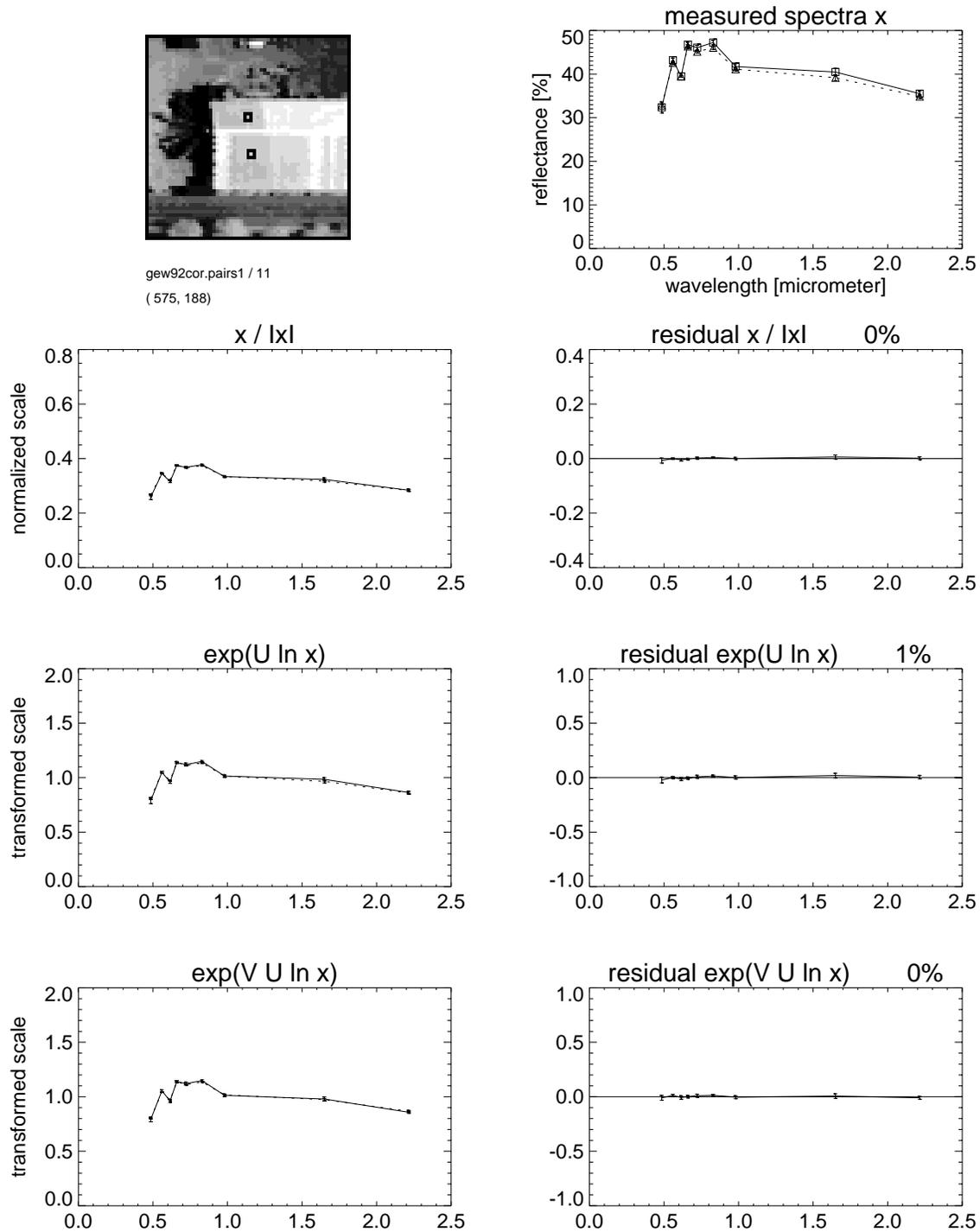


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

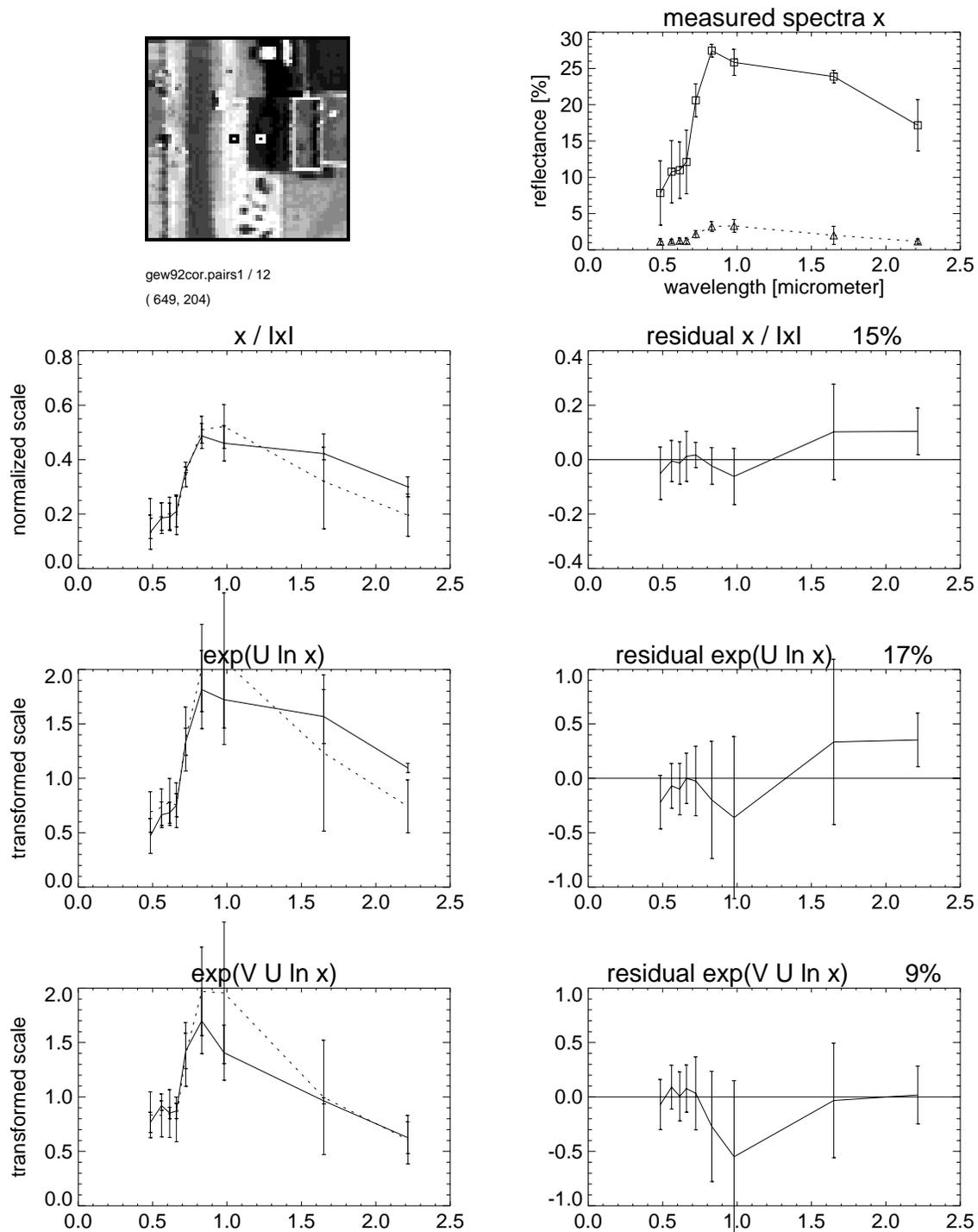
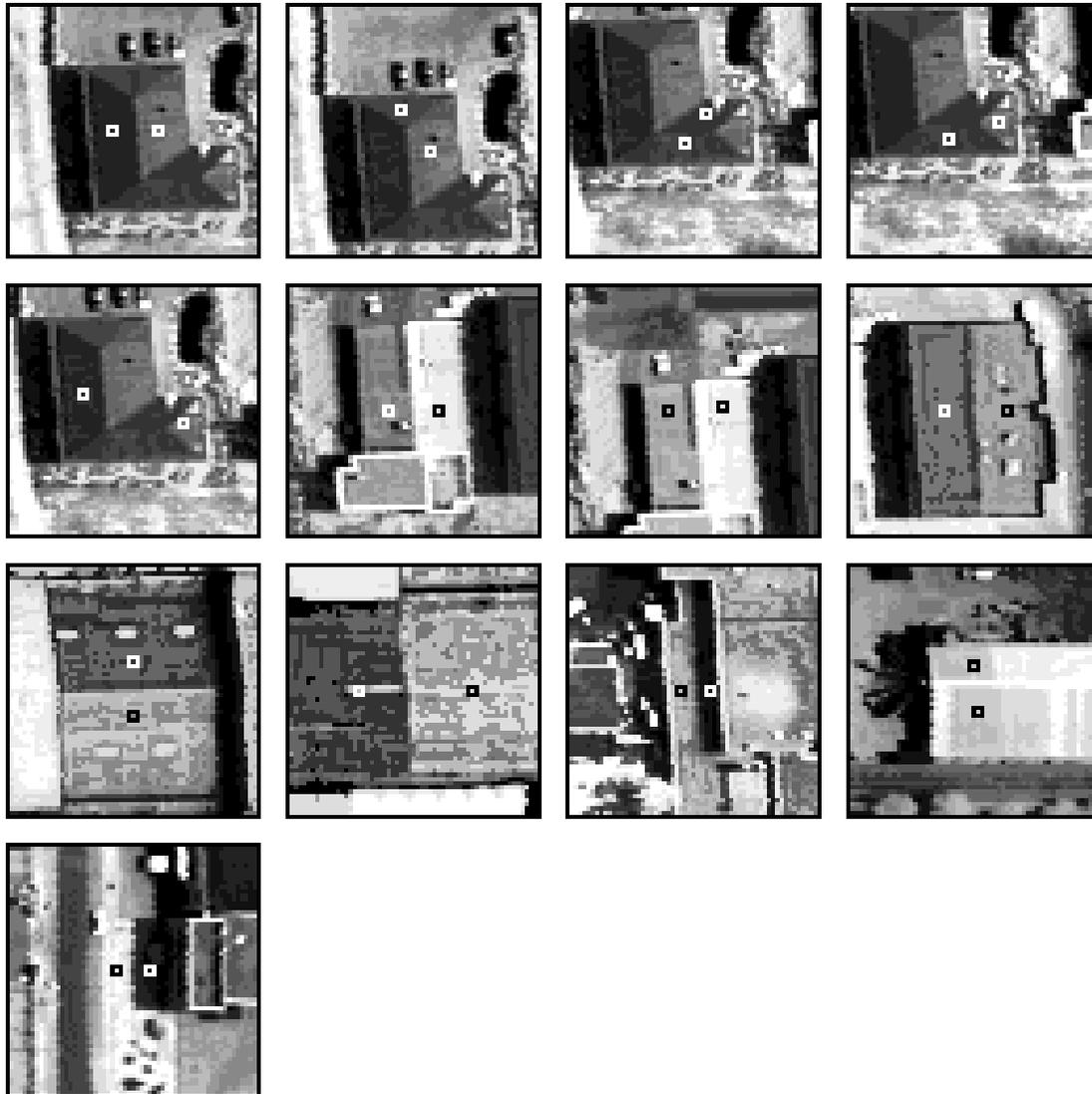


Figure C.1 (cont'd): Spectral sample pair from scene 'Gewerbegebiet 1992' (Fig. 6.2 on page 120).

C.2 Statistical Evaluation of Samples from Scene ‘Gewerbegebiet 1992’



gew92cor.pairs1

Figure C.2: Selected samples of surface pairs with varying orientation (from scene ‘Gewerbegebiet 1992’).

SEPARABILITY IN DIFFERENT FEATURE SPACES							
	\mathbf{x}	$\mathbf{x}/\ \mathbf{x}\ $	$\ln \mathbf{x}$	$\mathbf{U} \ln \mathbf{x}$	$\mathbf{VU} \ln \mathbf{x}$	$e^{\mathbf{U} \ln \mathbf{x}}$	$e^{\mathbf{VU} \ln \mathbf{x}}$
$\sqrt{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} / \text{rk } \boldsymbol{\Sigma}}$	0.894 ± 0.172	0.740 ± 0.124	0.556 ± 0.109	0.514 ± 0.096	0.378 ± 0.088	0.491 ± 0.093	0.296 ± 0.070
$\sqrt{\mathbf{d}^T \mathbf{d} / \text{tr } \boldsymbol{\Sigma}}$	0.901 ± 0.172	0.537 ± 0.101	0.860 ± 0.226	0.412 ± 0.081	0.215 ± 0.049	0.296 ± 0.058	0.144 ± 0.035
rescaled	1.011 ± 0.192	0.576 ± 0.110	0.852 ± 0.224	0.455 ± 0.090	0.250 ± 0.058	0.481 ± 0.097	0.241 ± 0.055
PCT, rank $H = 1$	0.894 ± 0.182	0.106 ± 0.019	0.974 ± 0.266	0.157 ± 0.035	0.145 ± 0.031	0.047 ± 0.008	0.096 ± 0.027
2	0.843 ± 0.164	0.808 ± 0.161	0.701 ± 0.187	0.640 ± 0.131	0.301 ± 0.124	0.575 ± 0.118	0.117 ± 0.039
3	1.017 ± 0.193	0.767 ± 0.154	0.707 ± 0.160	0.563 ± 0.125	0.293 ± 0.114	0.501 ± 0.102	0.178 ± 0.065
4	0.925 ± 0.170	0.851 ± 0.164	0.633 ± 0.147	0.559 ± 0.131	0.352 ± 0.103	0.505 ± 0.100	0.310 ± 0.089
5	0.911 ± 0.166	0.832 ± 0.141	0.611 ± 0.145	0.554 ± 0.117	0.385 ± 0.094	0.527 ± 0.102	0.328 ± 0.081
6	0.997 ± 0.203	0.791 ± 0.133	0.603 ± 0.130	0.548 ± 0.106	0.368 ± 0.085	0.509 ± 0.093	0.312 ± 0.075
7	0.954 ± 0.193	0.809 ± 0.142	0.592 ± 0.120	0.519 ± 0.097	0.378 ± 0.088	0.497 ± 0.090	0.301 ± 0.071
8	0.923 ± 0.186	0.765 ± 0.132	0.564 ± 0.111	0.514 ± 0.096	0.378 ± 0.088	0.509 ± 0.097	0.296 ± 0.068
9	0.894 ± 0.172	0.740 ± 0.124	0.556 ± 0.109	0.514 ± 0.096	0.378 ± 0.088	0.491 ± 0.093	0.296 ± 0.070

gew92cor.pairs1

Table C.1: Spectral distances (mean \pm deviation) of the sample pairs (from scene ‘Gewerbegebiet 1992’) in 7 basic feature spaces, and after additional rescaling and principal component transformation (PCT).

SEPARABILITY W.R.T. FILTER RANK						
	$\ln \mathbf{x}$	$\mathbf{P}_K \ln \mathbf{x}, K =$				
	0	1	2	3	4	5
$\sqrt{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} / \text{rk } \boldsymbol{\Sigma}}$	0.556 ± 0.109	0.514 ± 0.096	0.378 ± 0.088	0.321 ± 0.071	0.309 ± 0.077	0.275 ± 0.085
$\sqrt{\mathbf{d}^T \mathbf{d} / \text{tr } \boldsymbol{\Sigma}}$	0.860 ± 0.226	0.412 ± 0.081	0.215 ± 0.049	0.196 ± 0.052	0.196 ± 0.048	0.175 ± 0.053

gew92cor.pairs1

Table C.2: Spectral distances (mean \pm deviation) of the sample pairs (from scene ‘Gewerbegebiet 1992’) in 6 feature spaces of varying rank $\text{rk } \boldsymbol{\Sigma} = N - K$: The log space $\ln \mathbf{x}$ (for $\mathbf{P}_0 = \mathbf{I}$, i.e., no filtering), and the filtered spaces $\mathbf{P}_K \ln \mathbf{x}$, particularly $\mathbf{P}_1 = \mathbf{U}$ and $\mathbf{P}_2 = \mathbf{VU}$.

C.3 Sample Pairs from Scene 'Ziegelstein 1991'

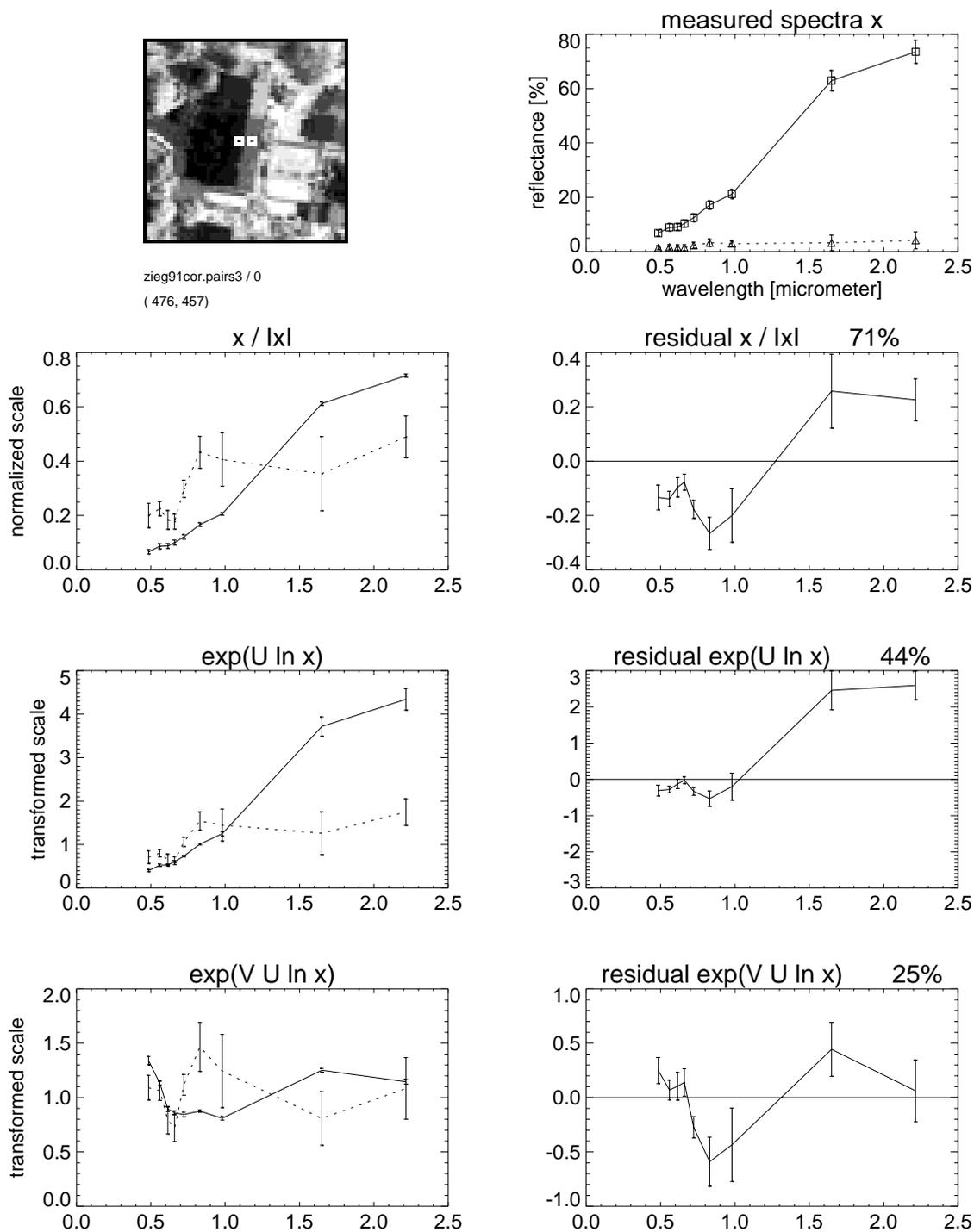


Figure C.3: Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

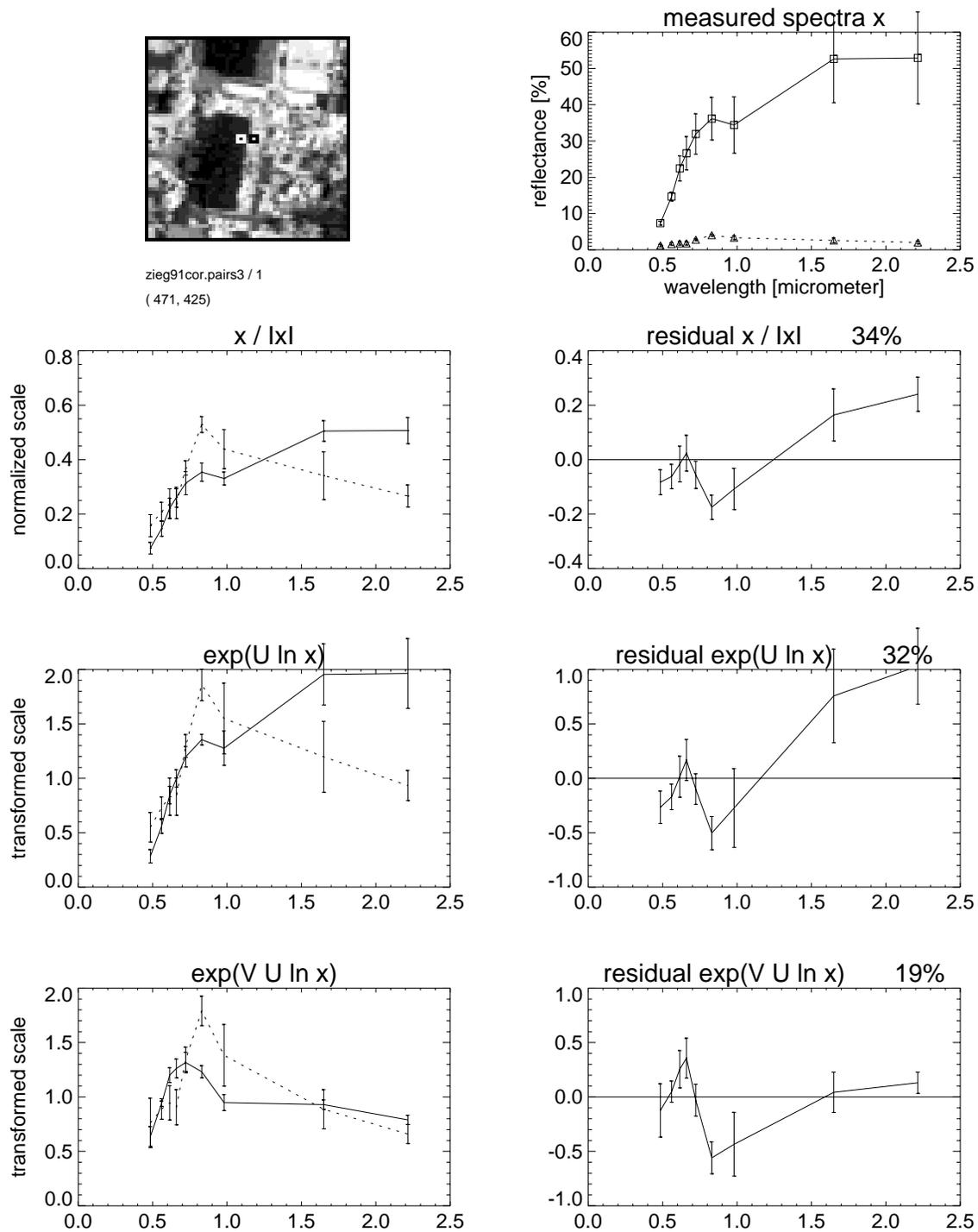


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

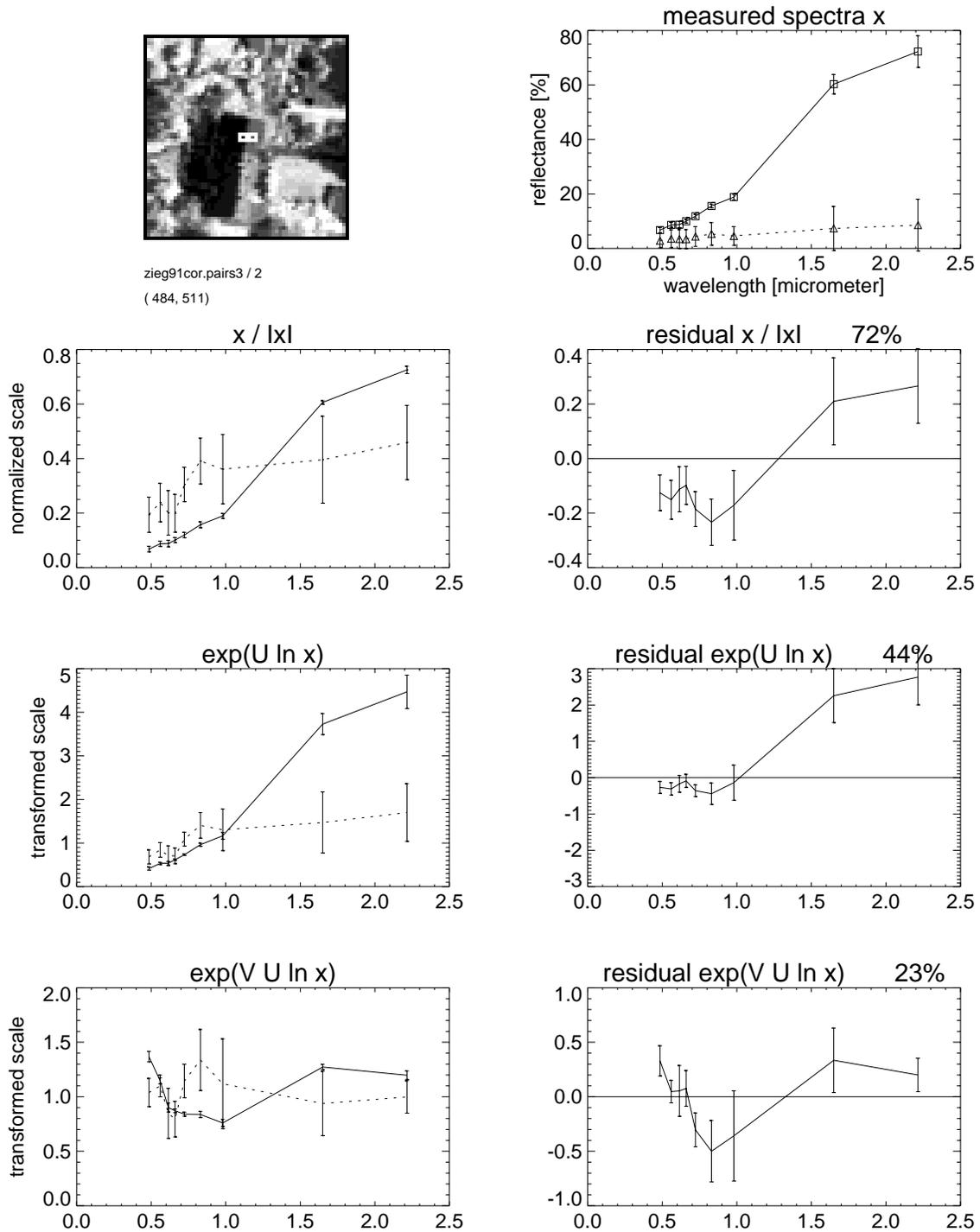


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

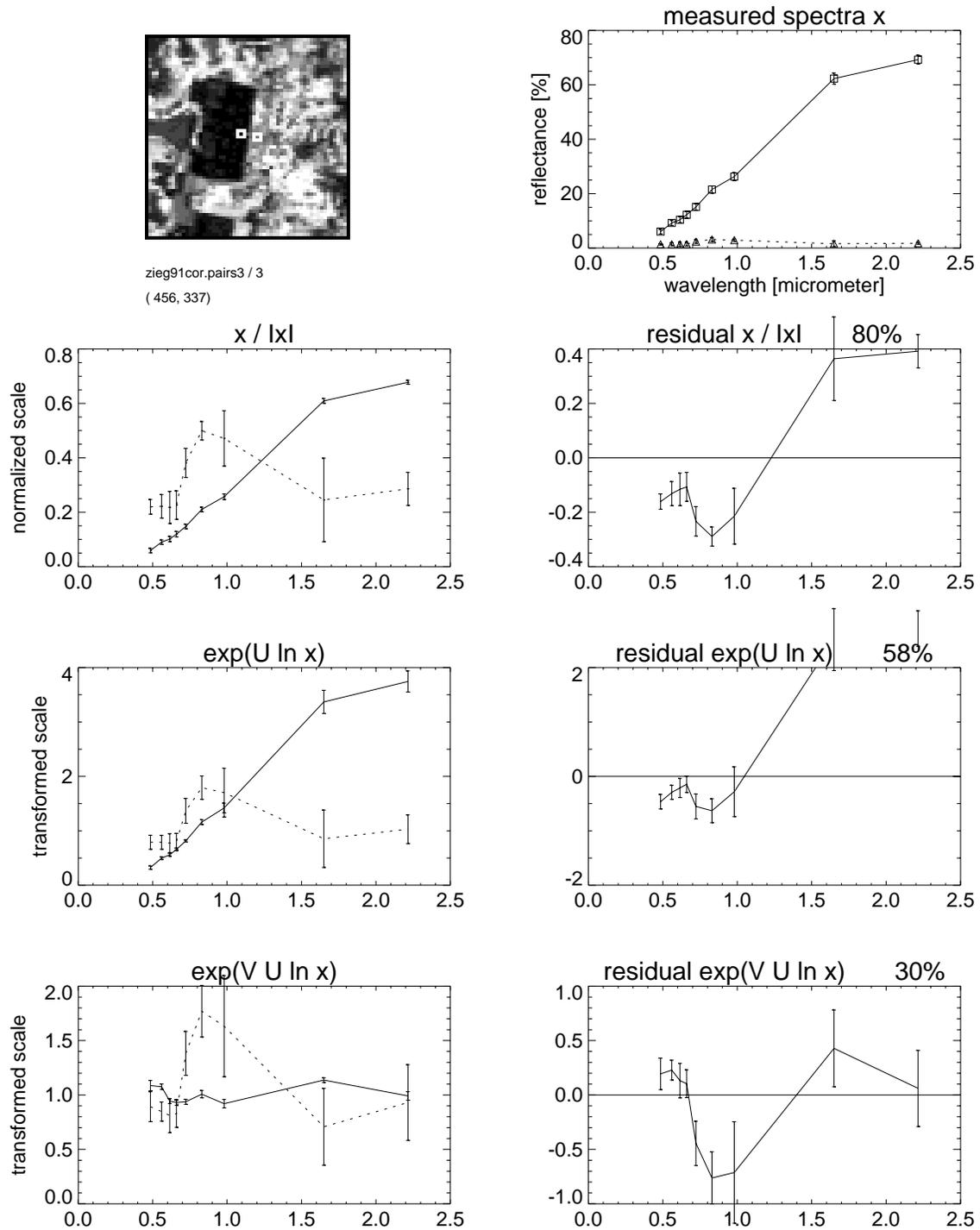


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

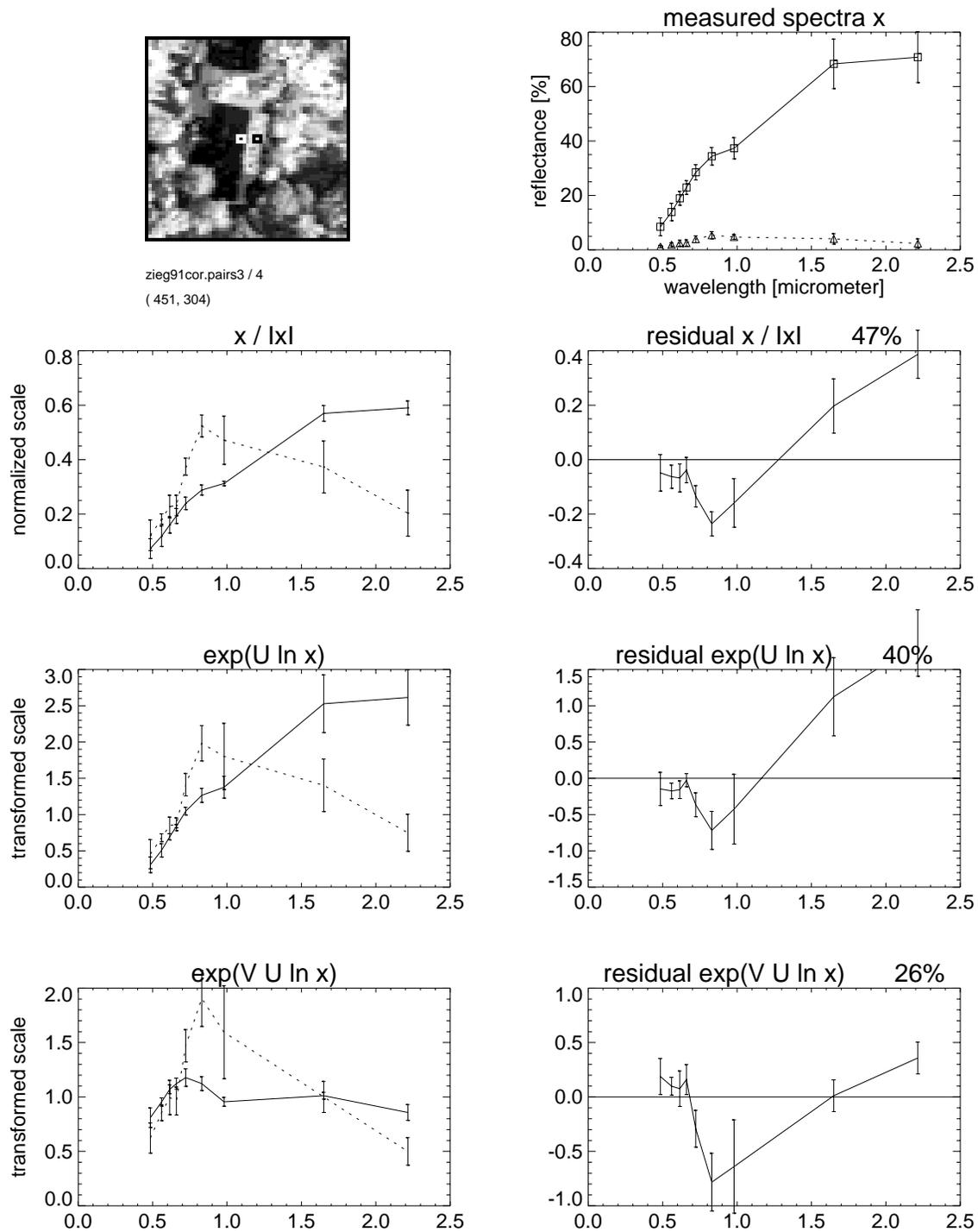


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

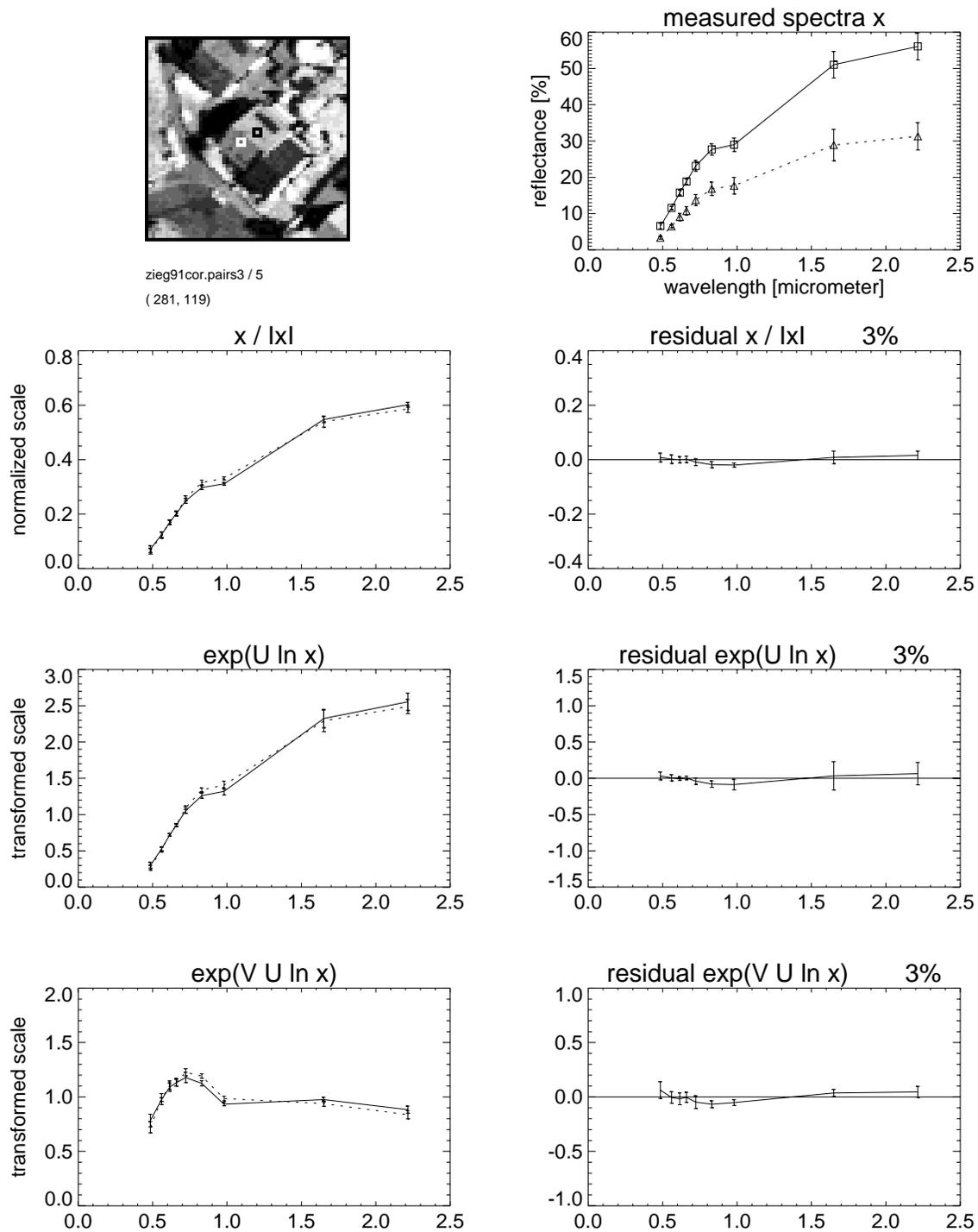


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

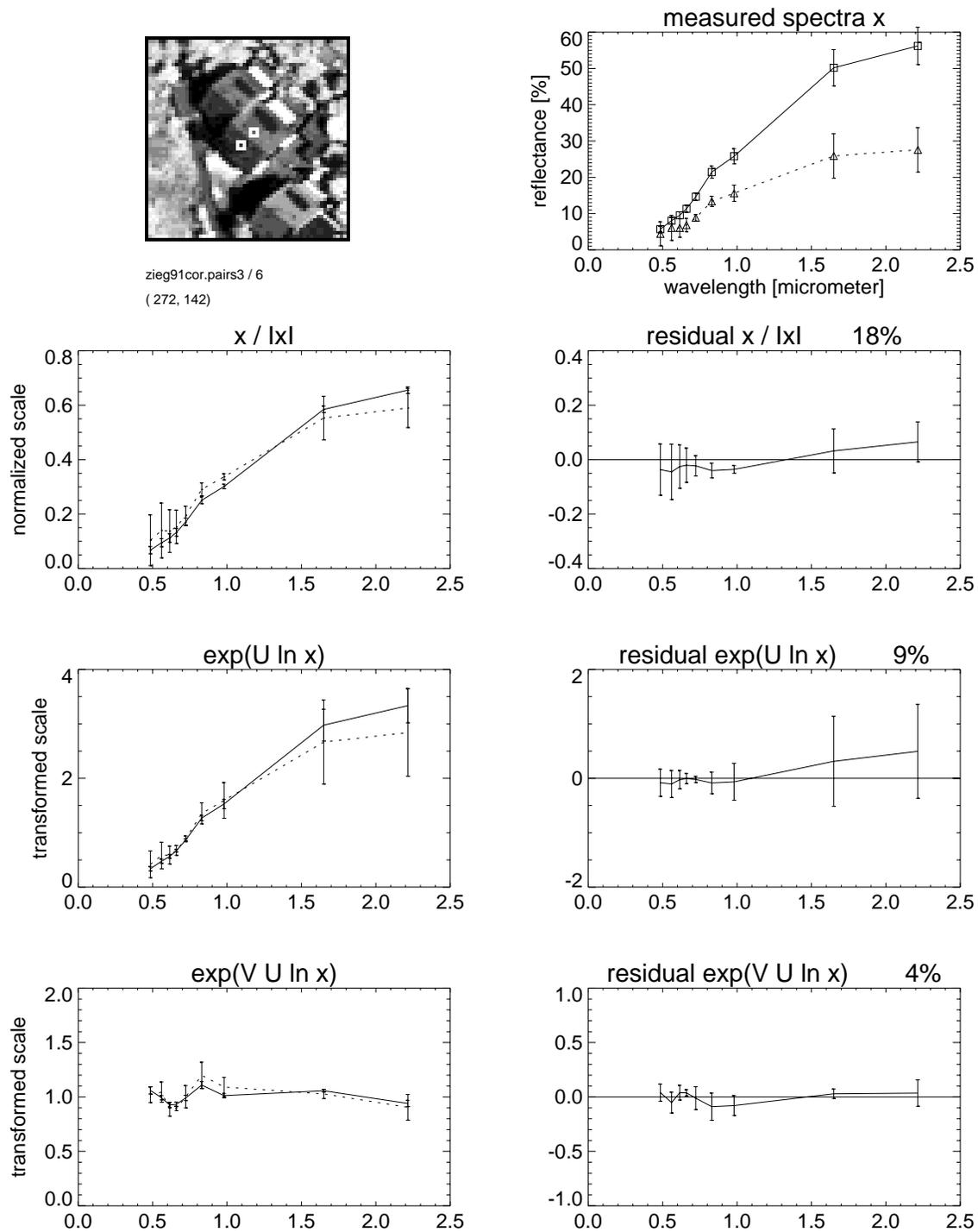


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

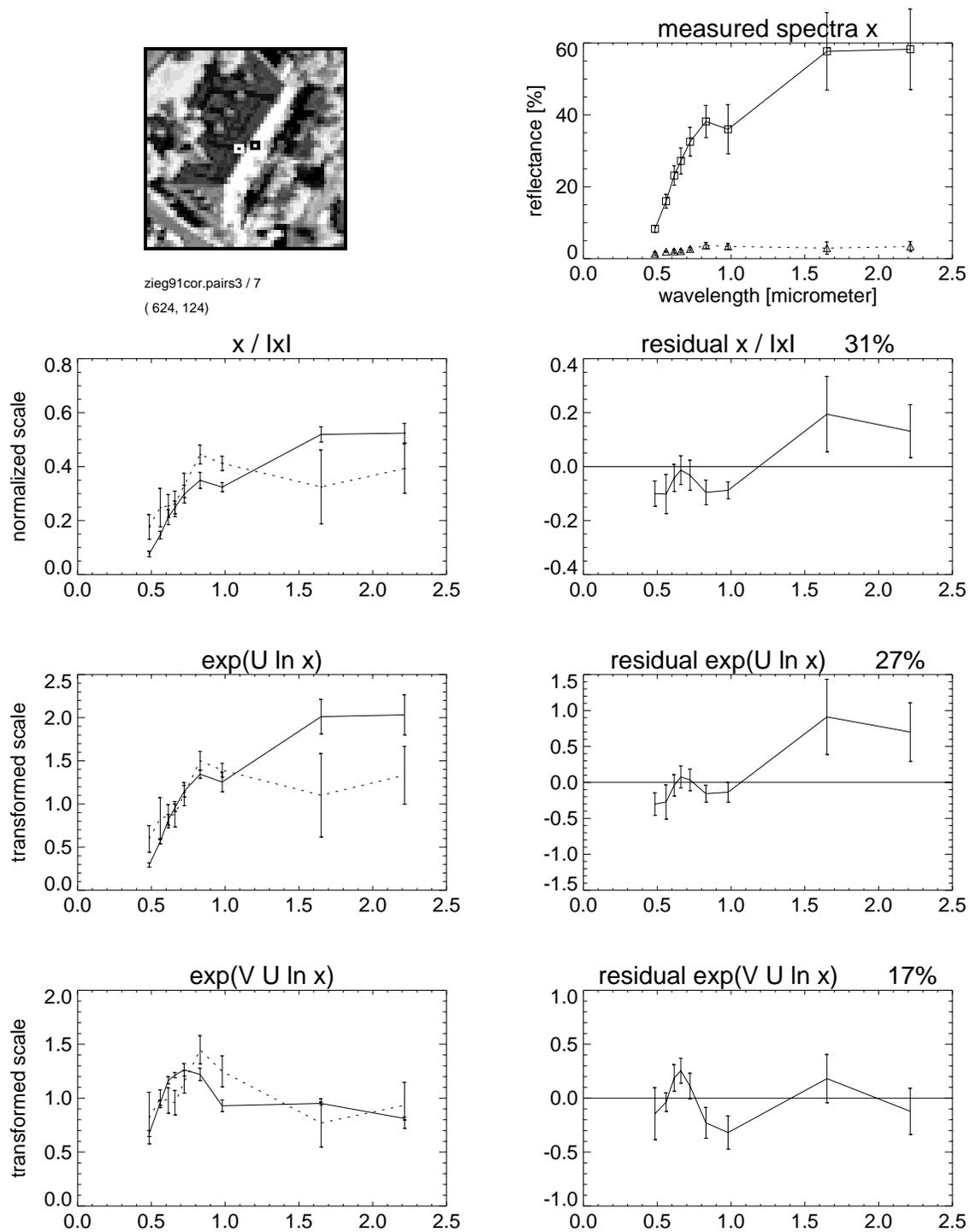


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

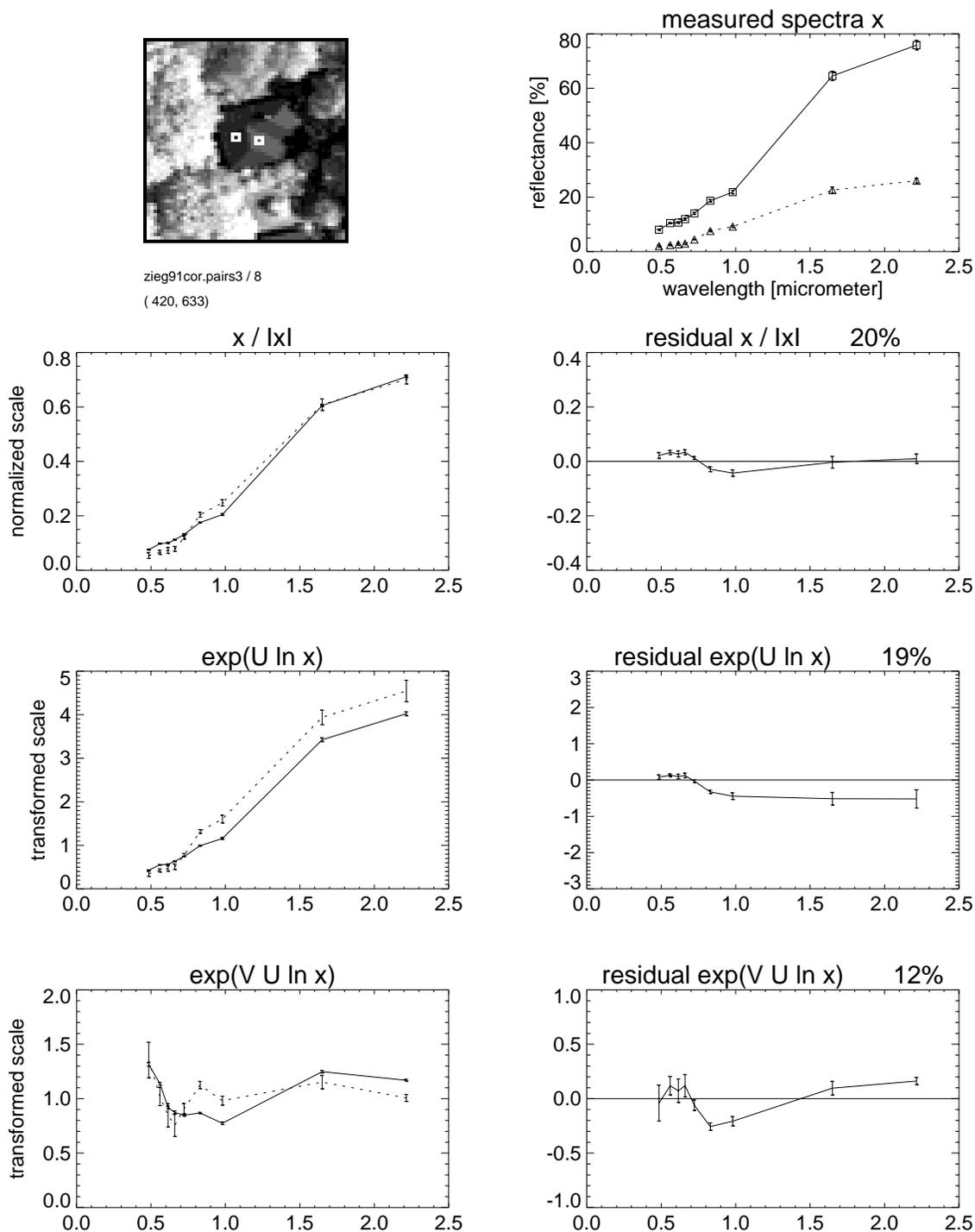


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

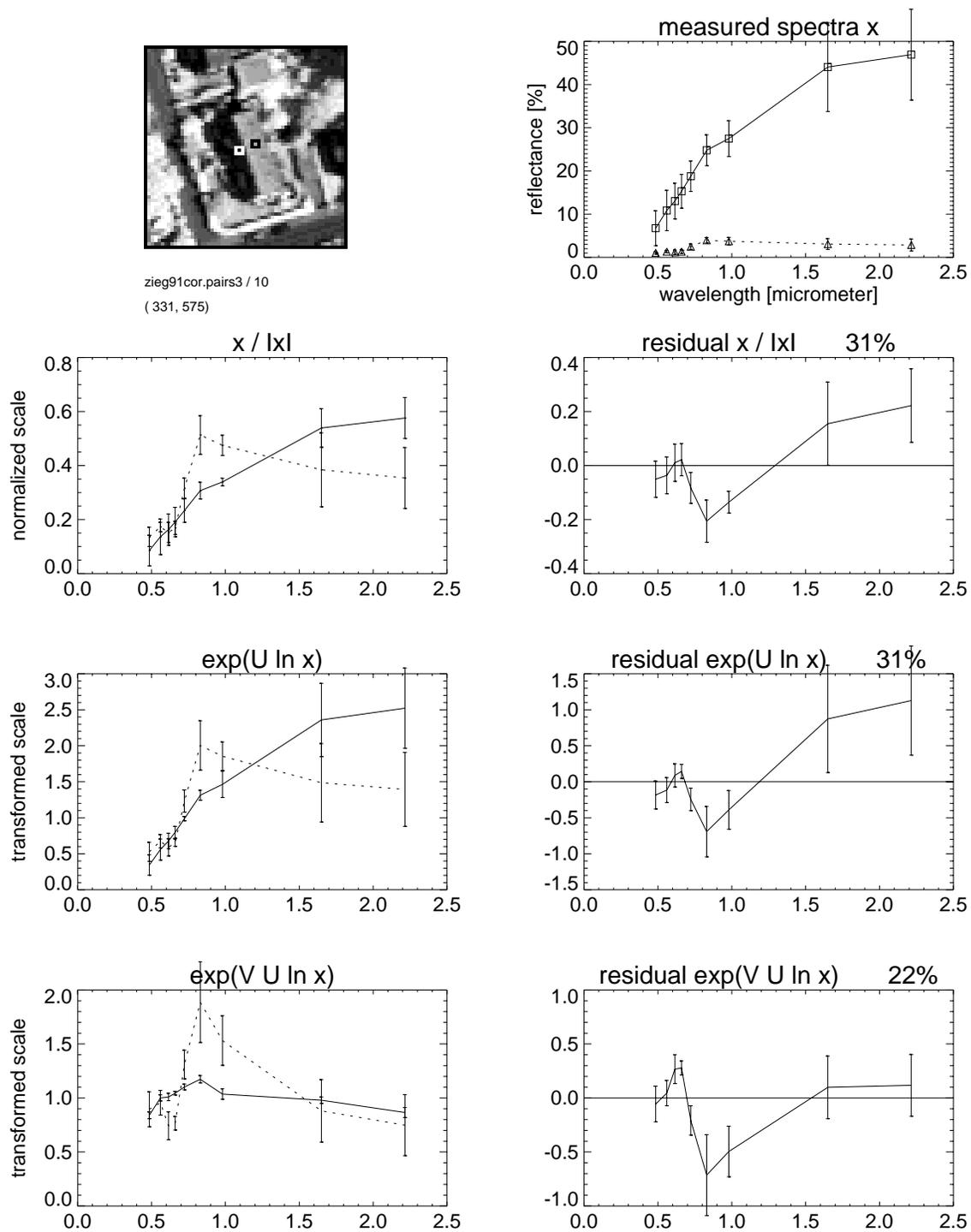


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

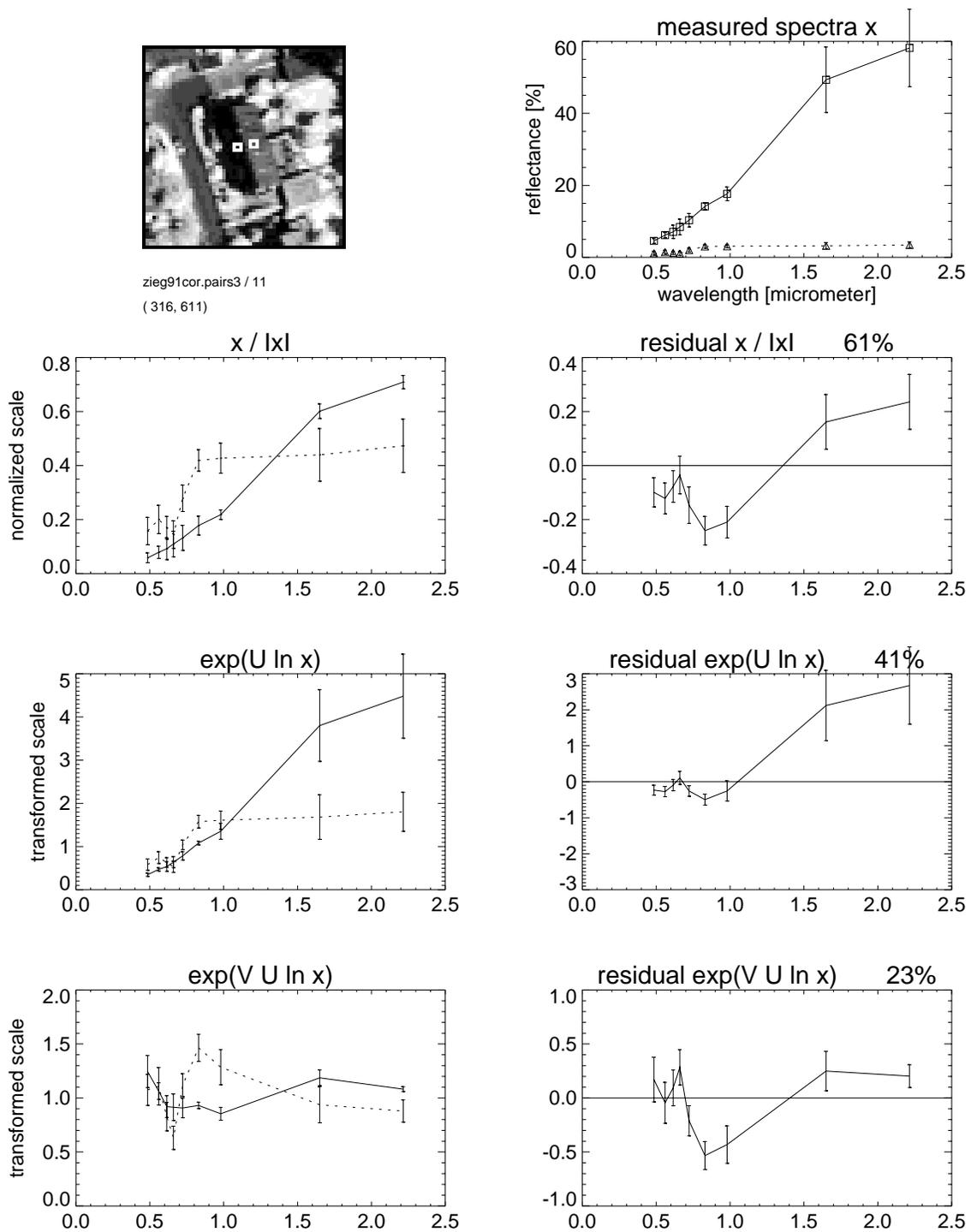


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

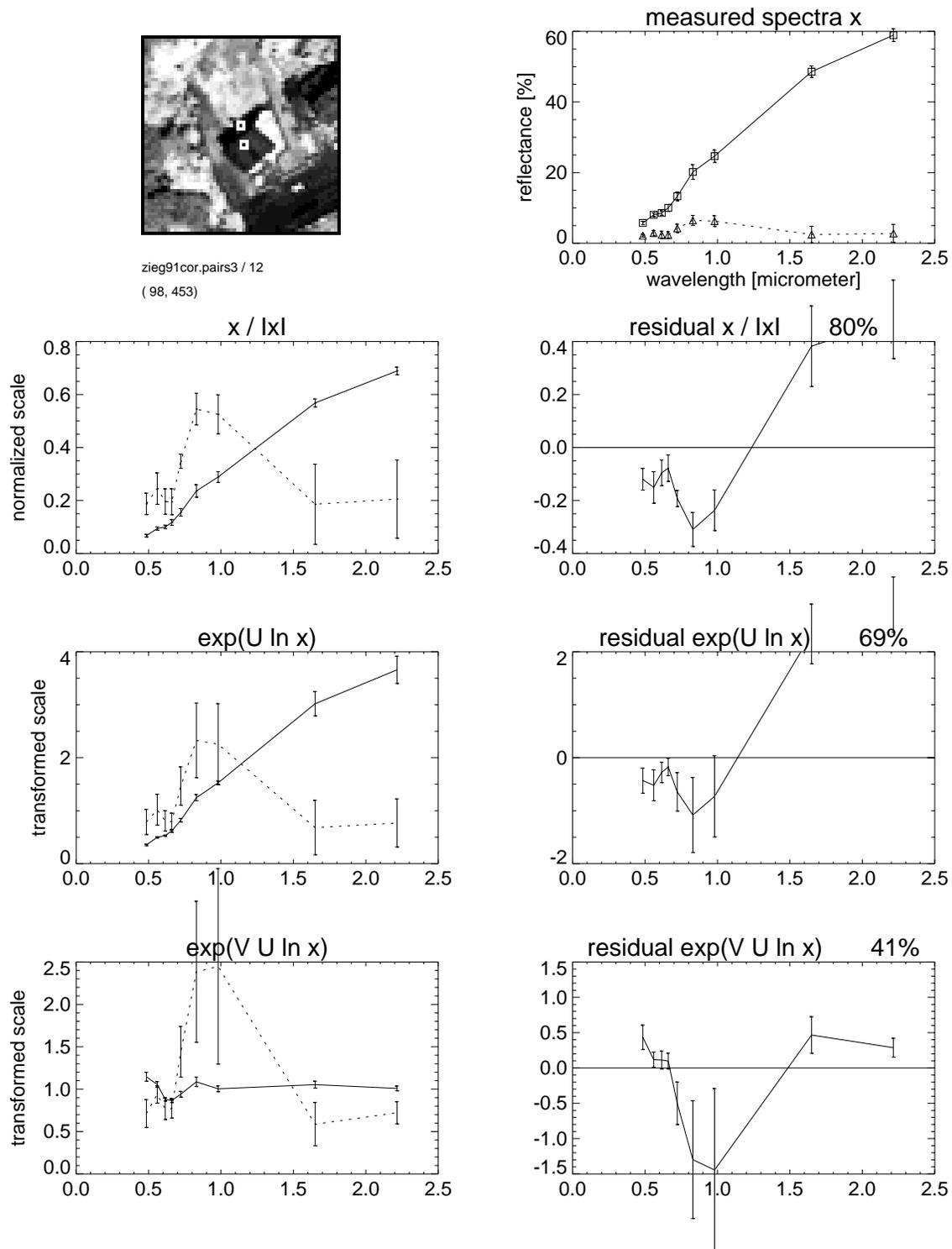


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

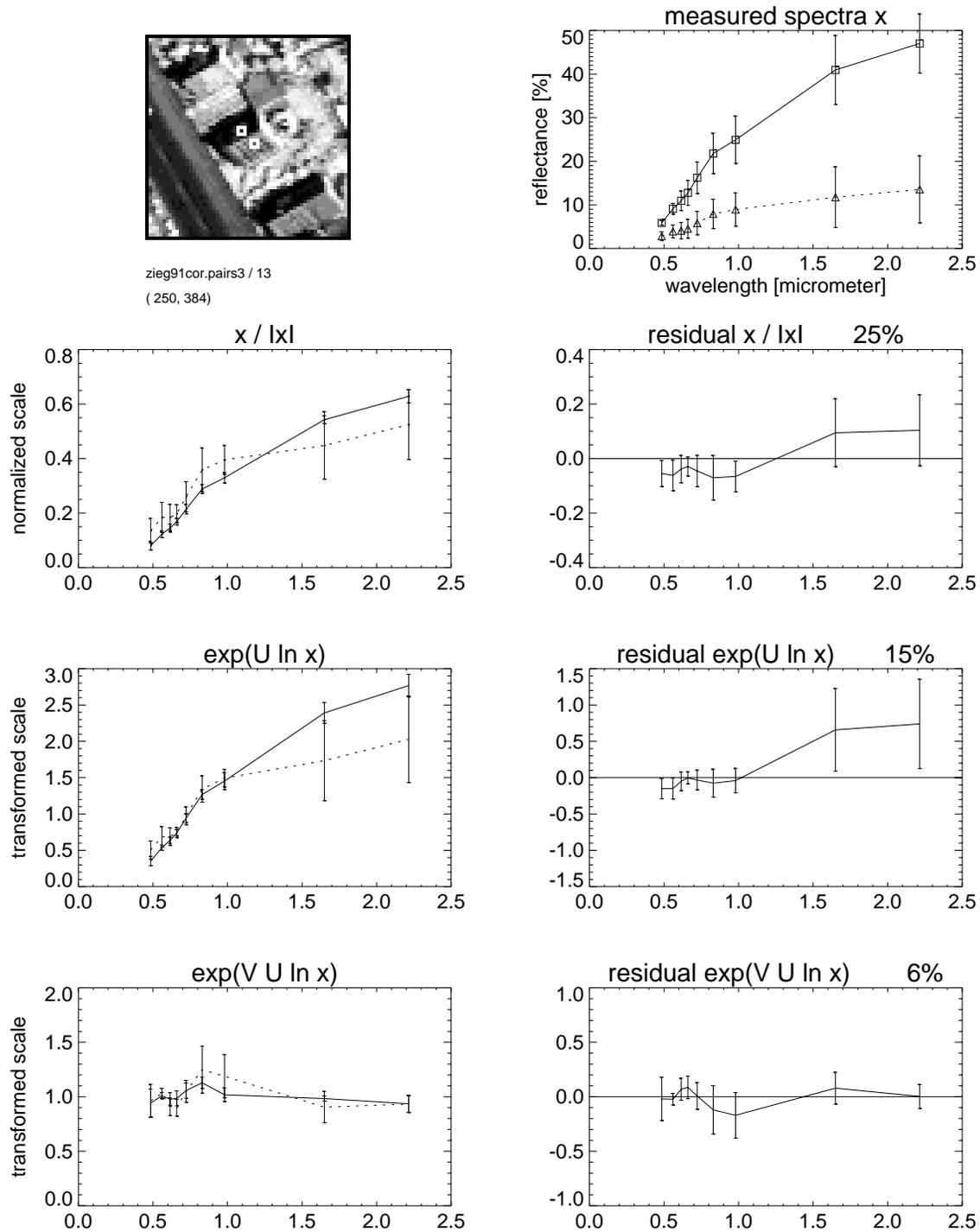


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

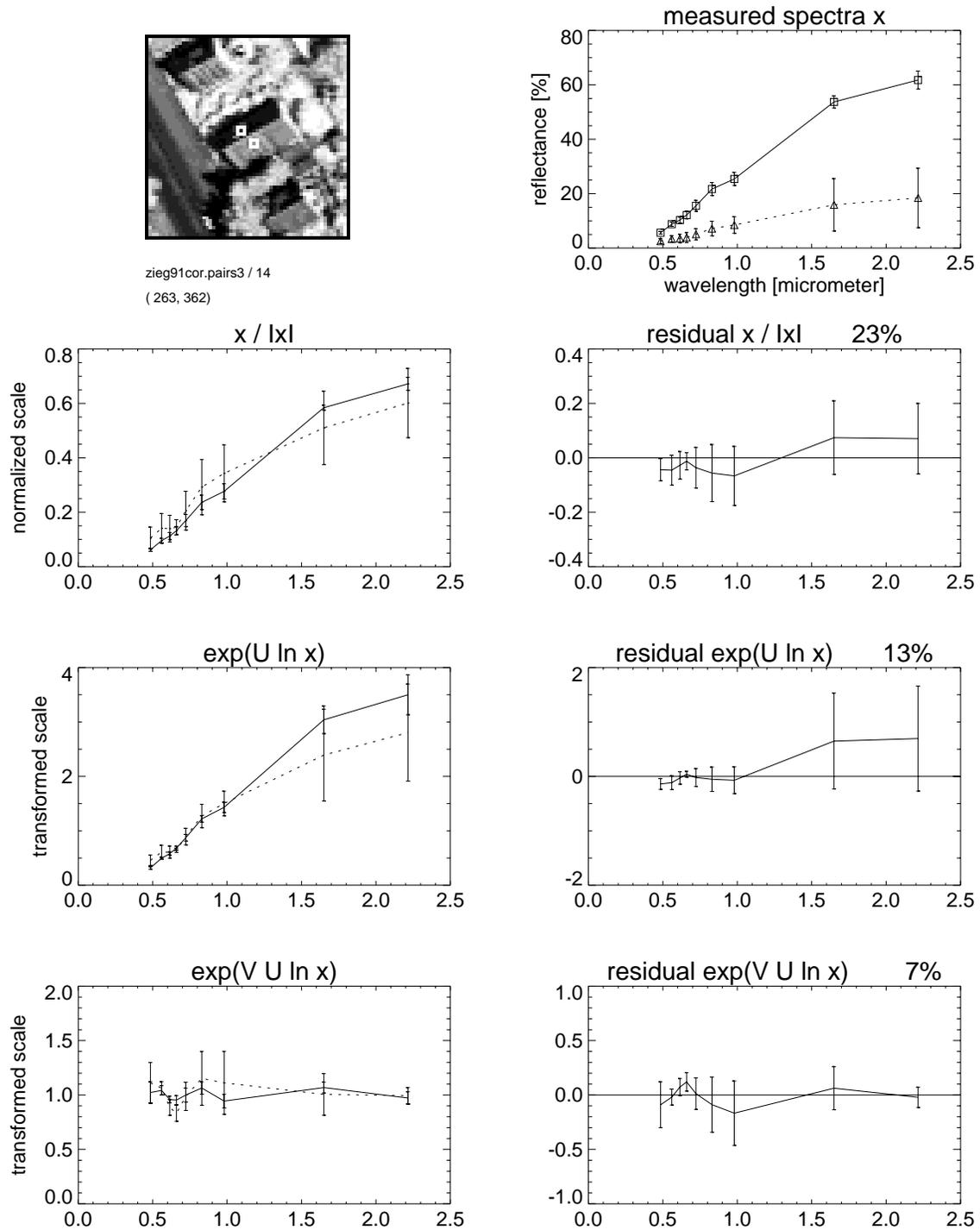


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

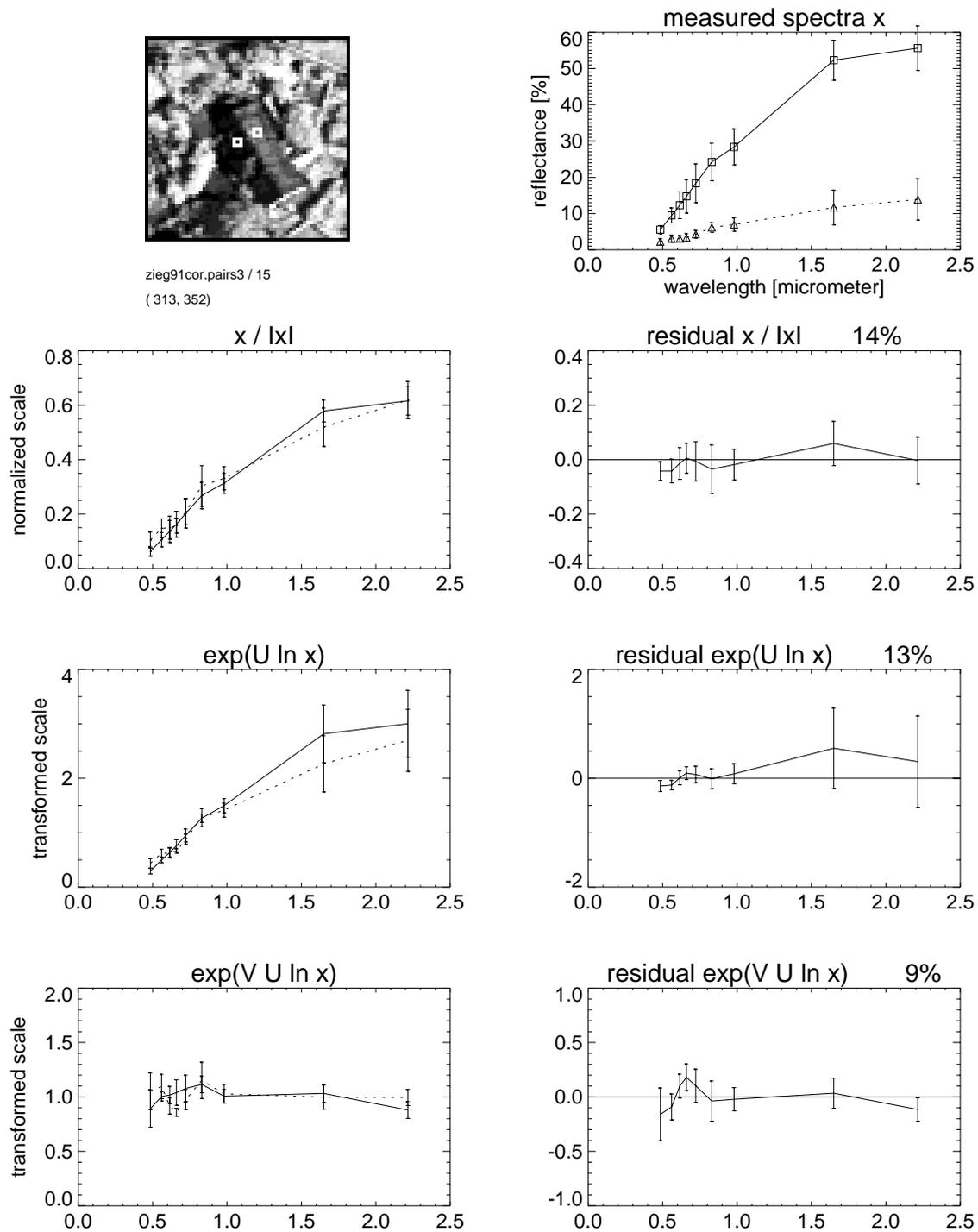


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

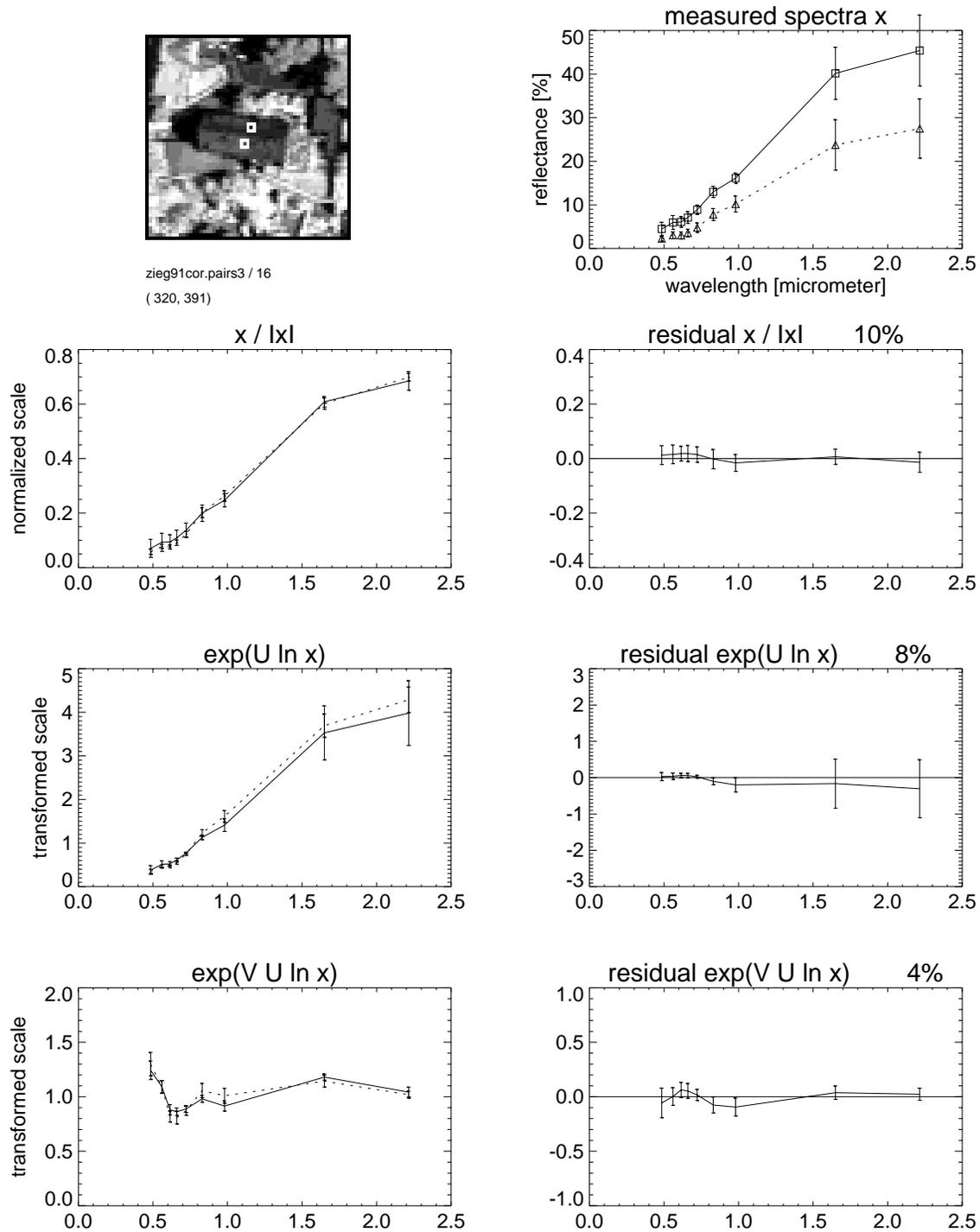


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

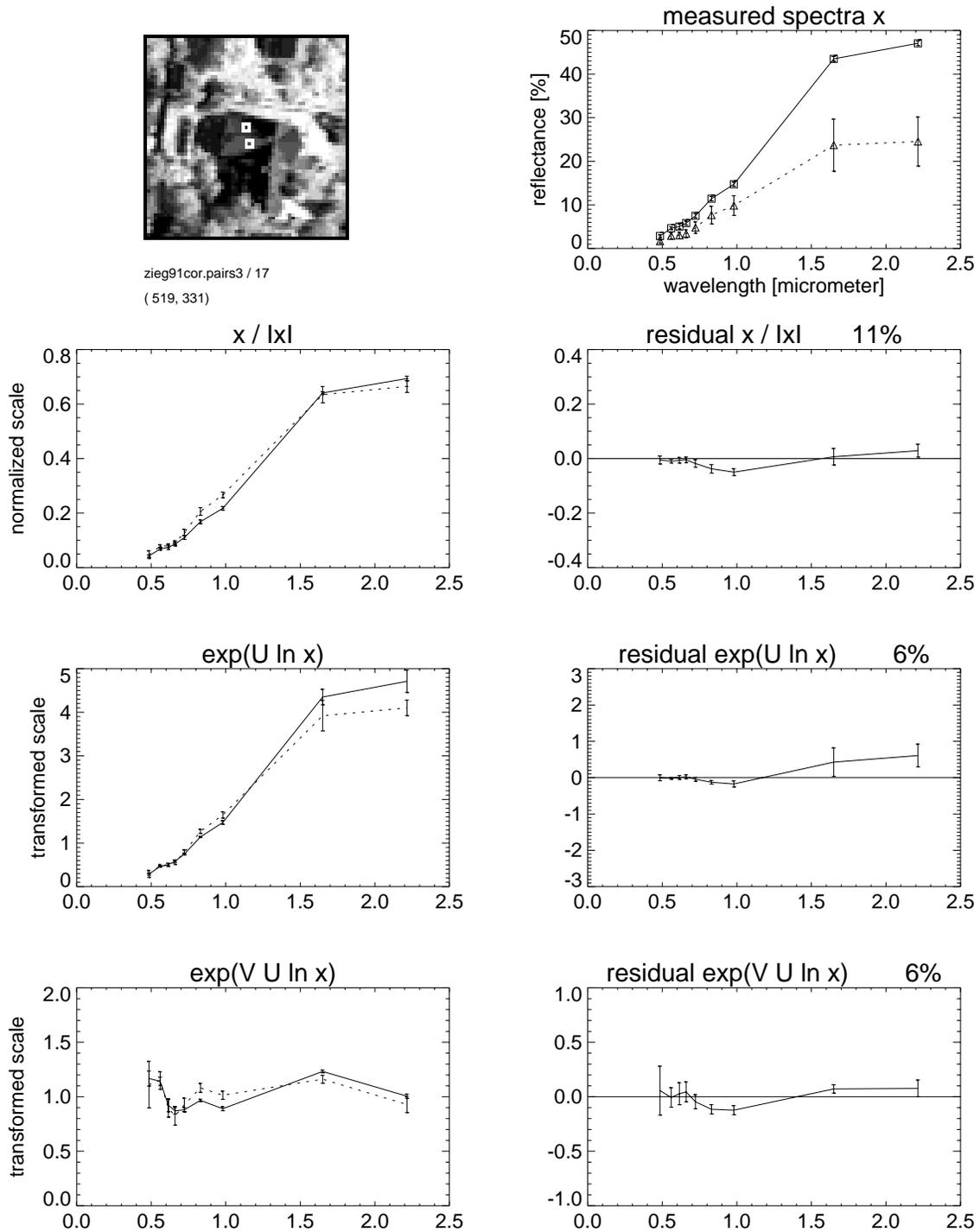


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

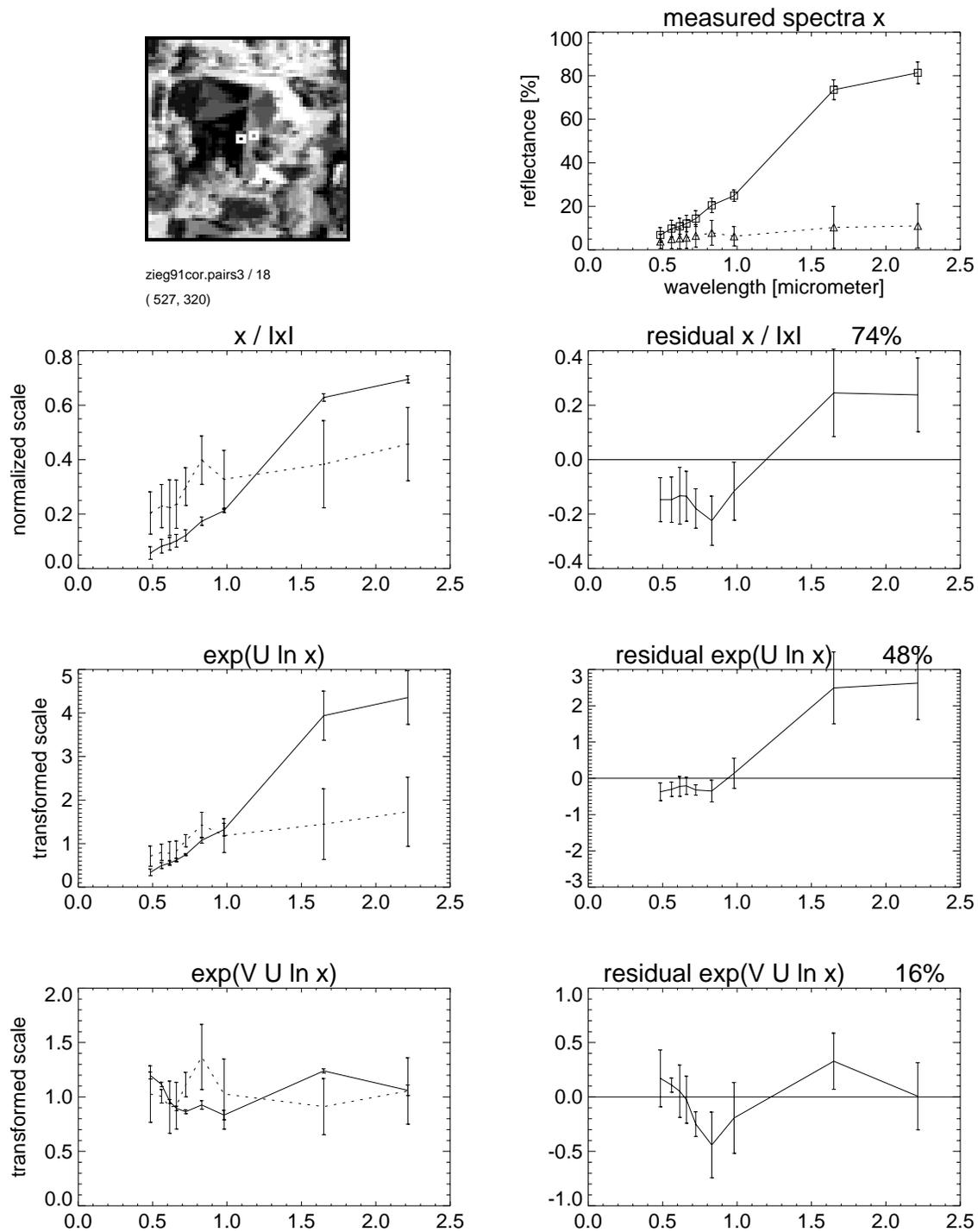


Figure C.3 (cont'd): Spectral sample pair from scene 'Ziegelstein 1991' (Fig. 6.1 on page 119).

C.4 Multispectral Classification of Samples

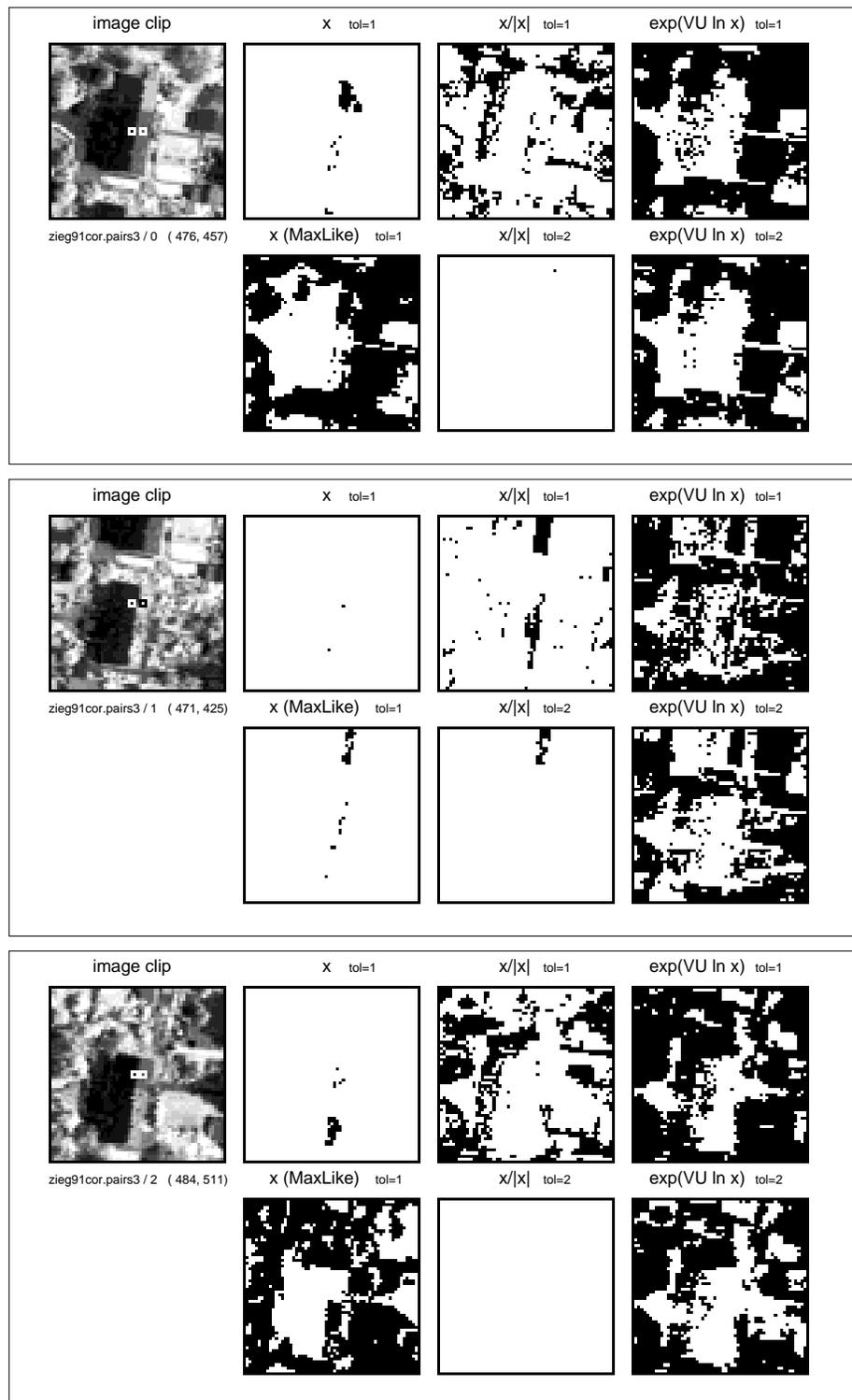


Figure C.4: Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class. (Scene ‘Ziegelstein 1991’)

White: belongs to class; black: belongs not to class.

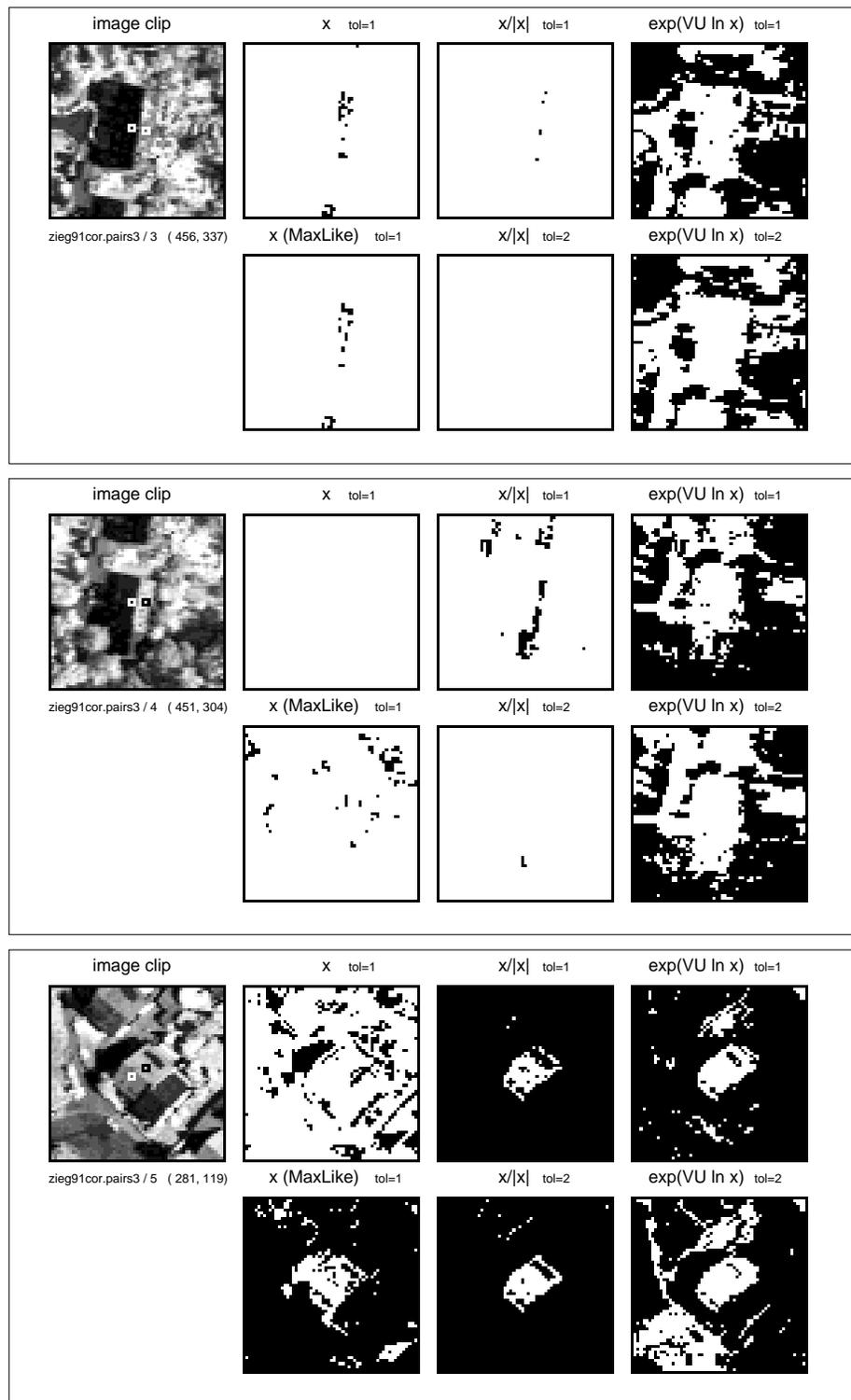


Figure C.4 (cont'd): Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class. (Scene 'Ziegelstein 1991')

White: belongs to class; black: belongs not to class.

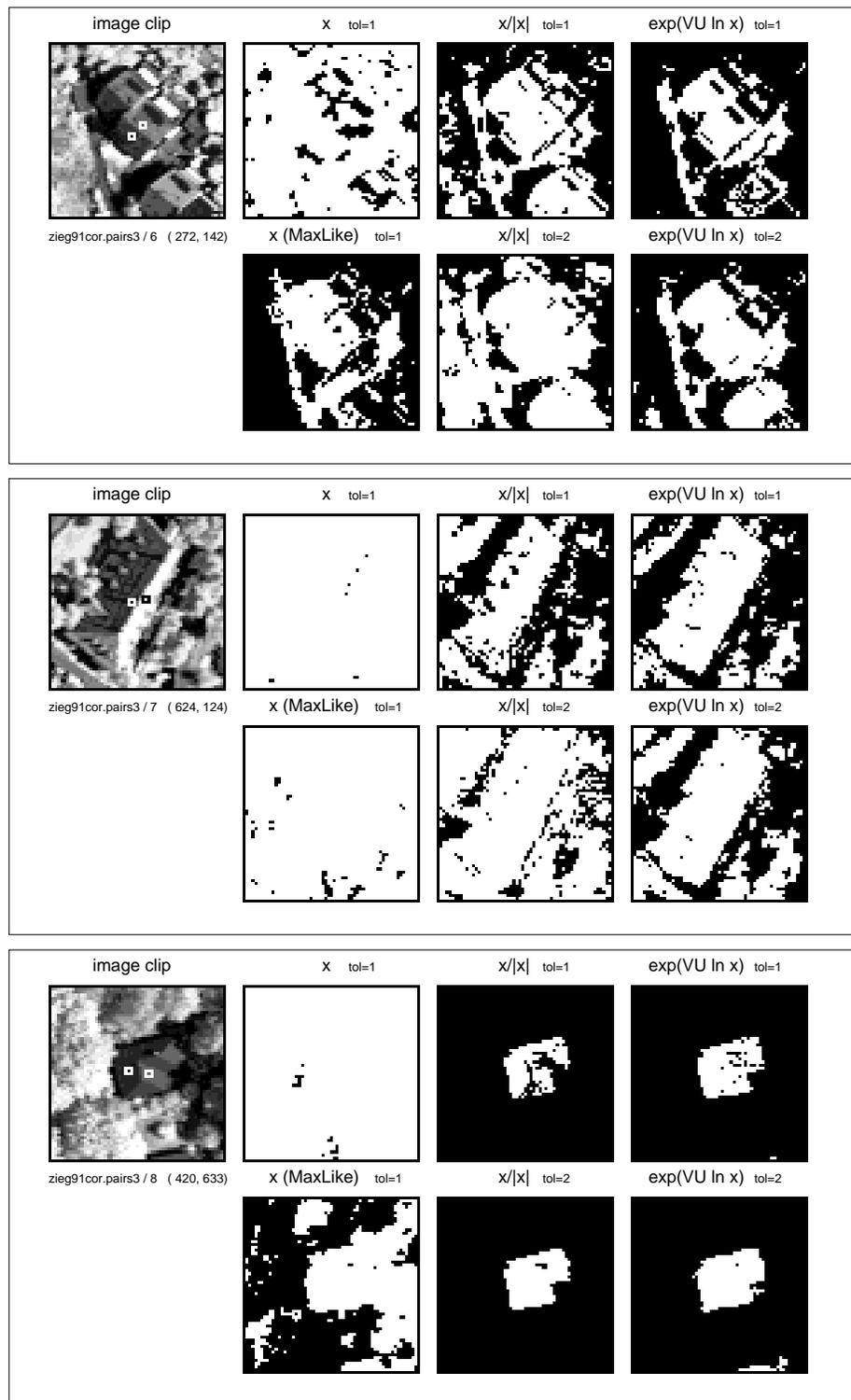


Figure C.4 (cont'd): Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class. (Scene 'Ziegelstein 1991')

White: belongs to class; black: belongs not to class.

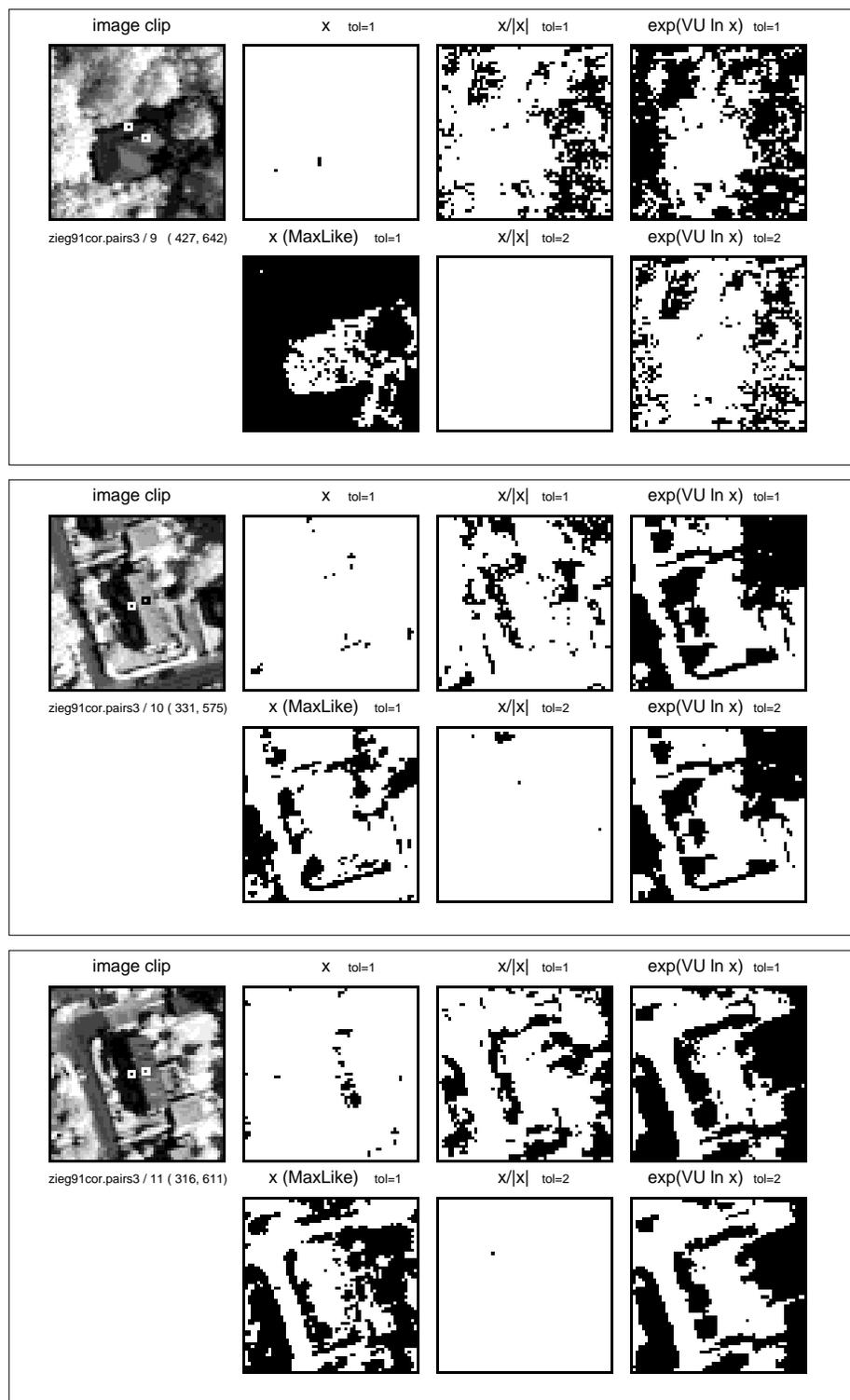


Figure C.4 (cont'd): Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class. (Scene 'Ziegelstein 1991')

White: belongs to class; black: belongs not to class.

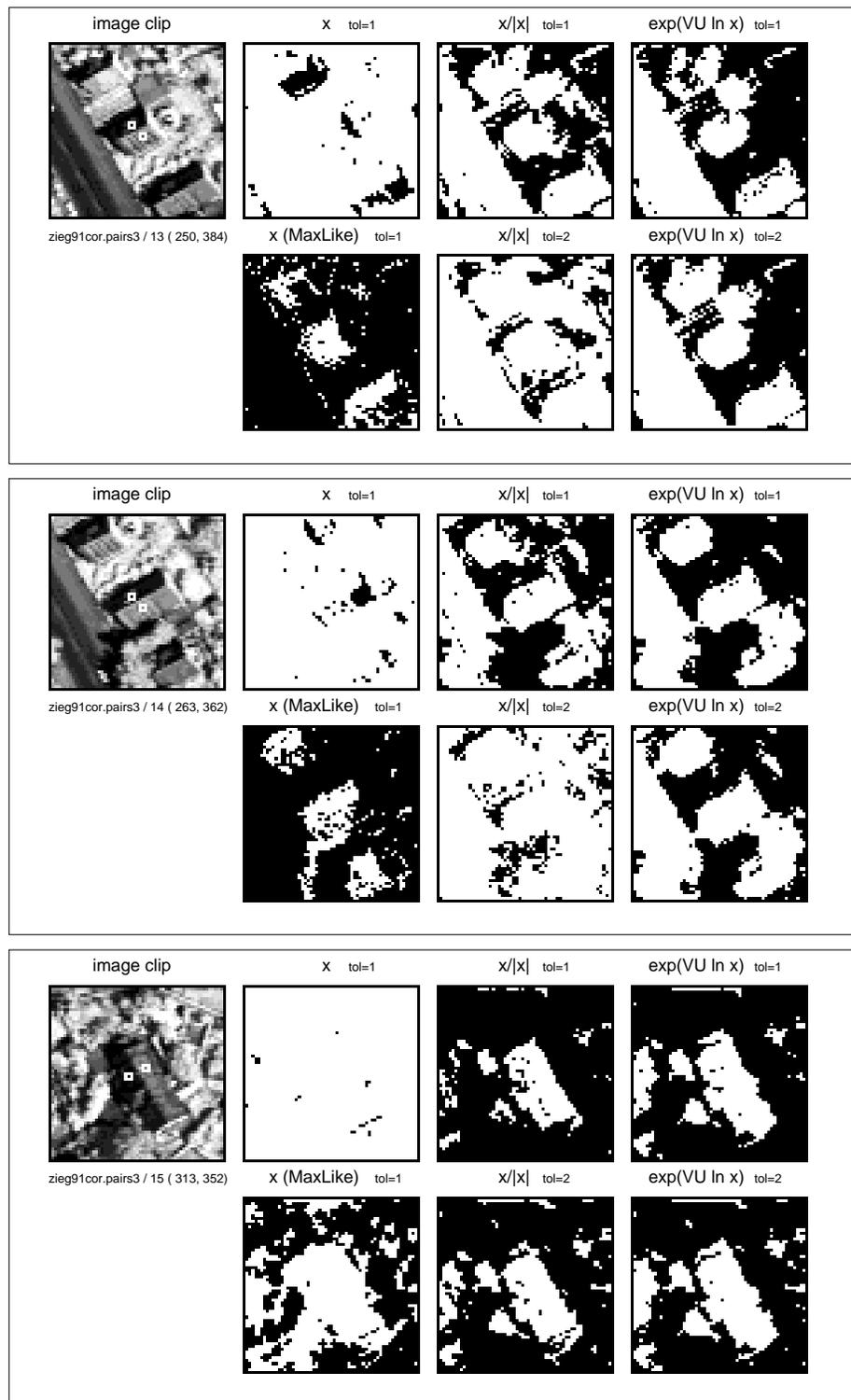


Figure C.4 (cont'd): Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class. (Scene 'Ziegelstein 1991')

White: belongs to class; black: belongs not to class.

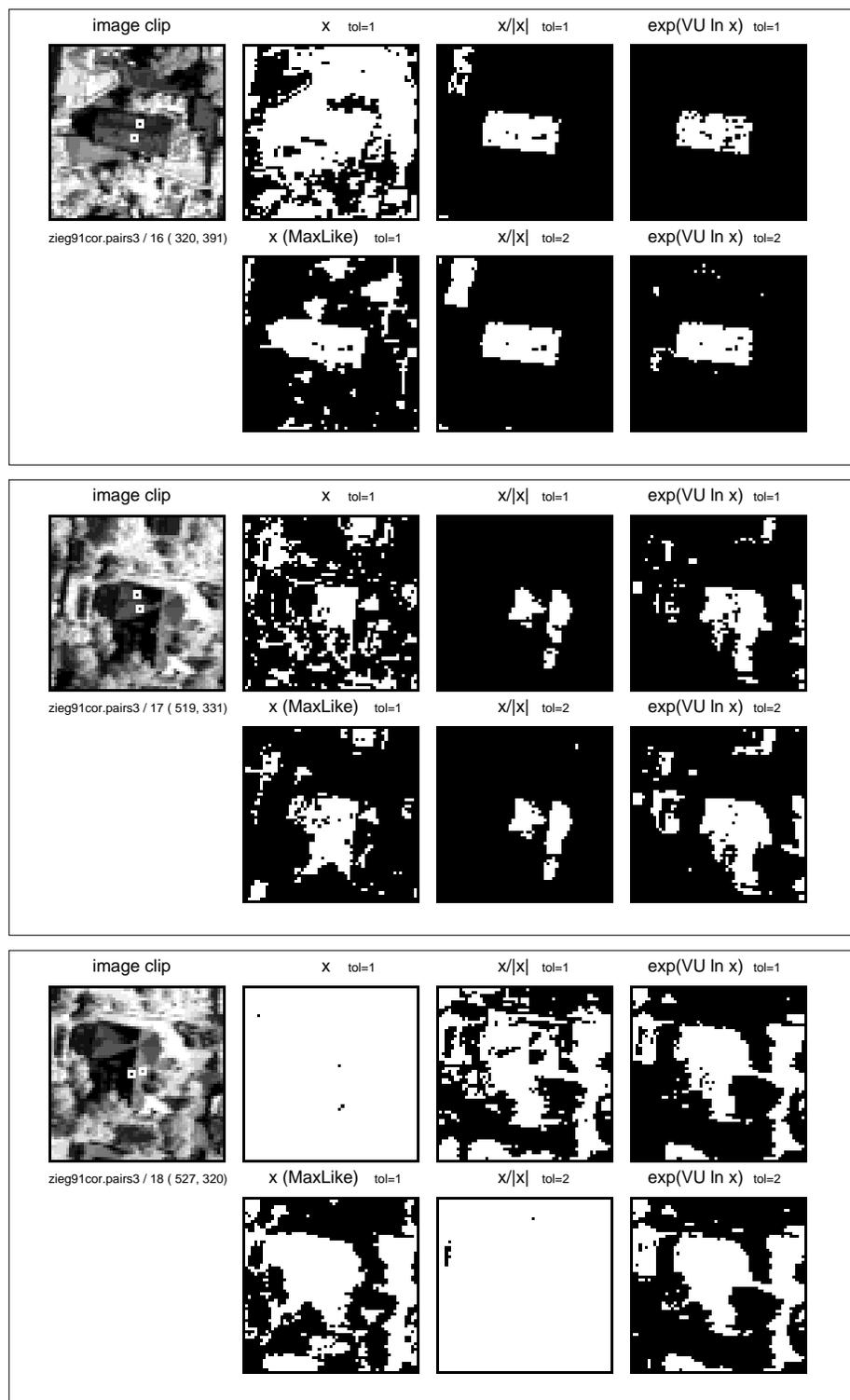


Figure C.4 (cont'd): Multispectral classification into binary class images in various feature spaces. The threshold is always fixed such that the designated pixels (top left) are just falling into the class. (Scene 'Ziegelstein 1991')

White: belongs to class; black: belongs not to class.

C.5 Bitemporal Sample Pairs from Scene ‘Gewerbegebiet’

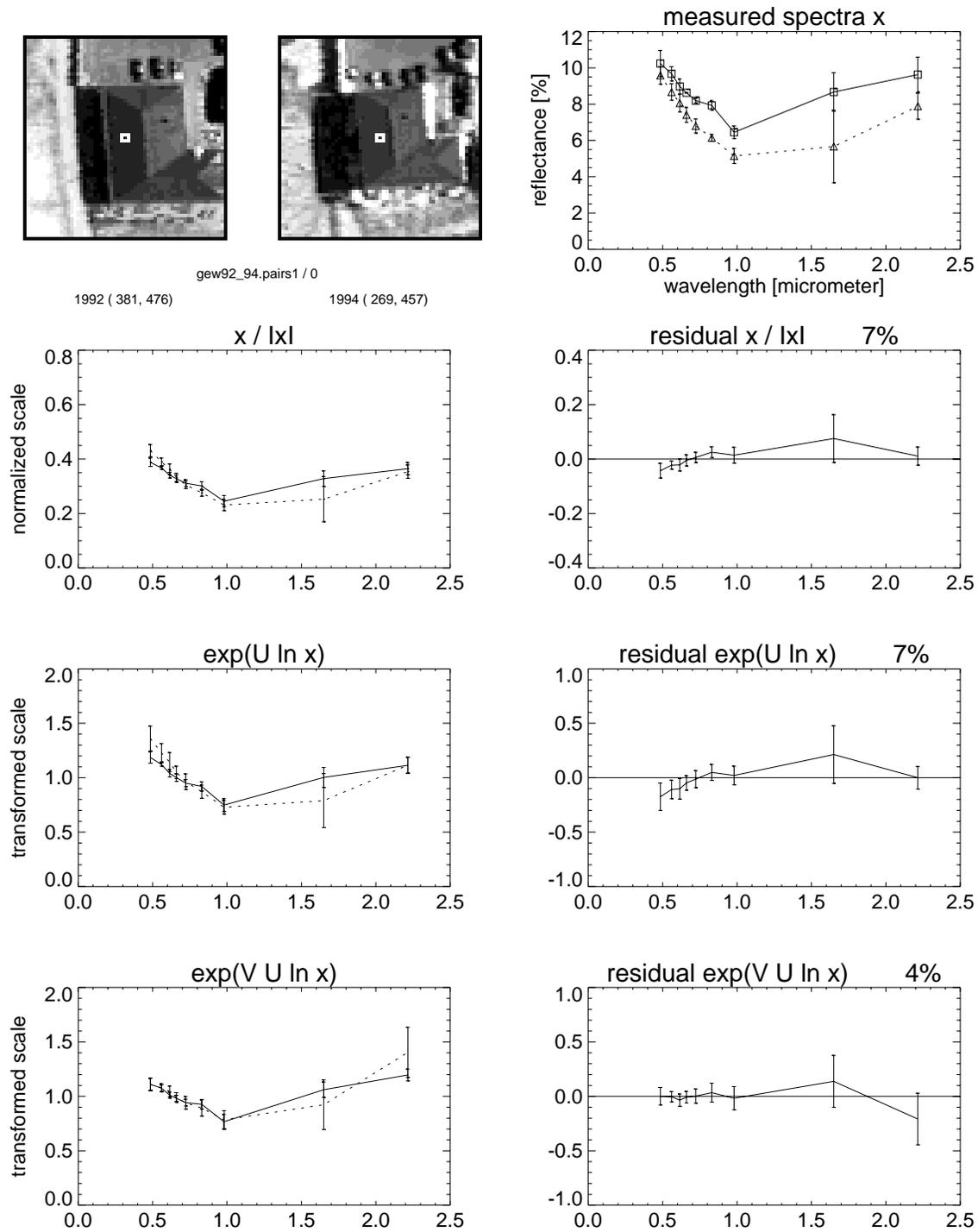


Figure C.5: Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

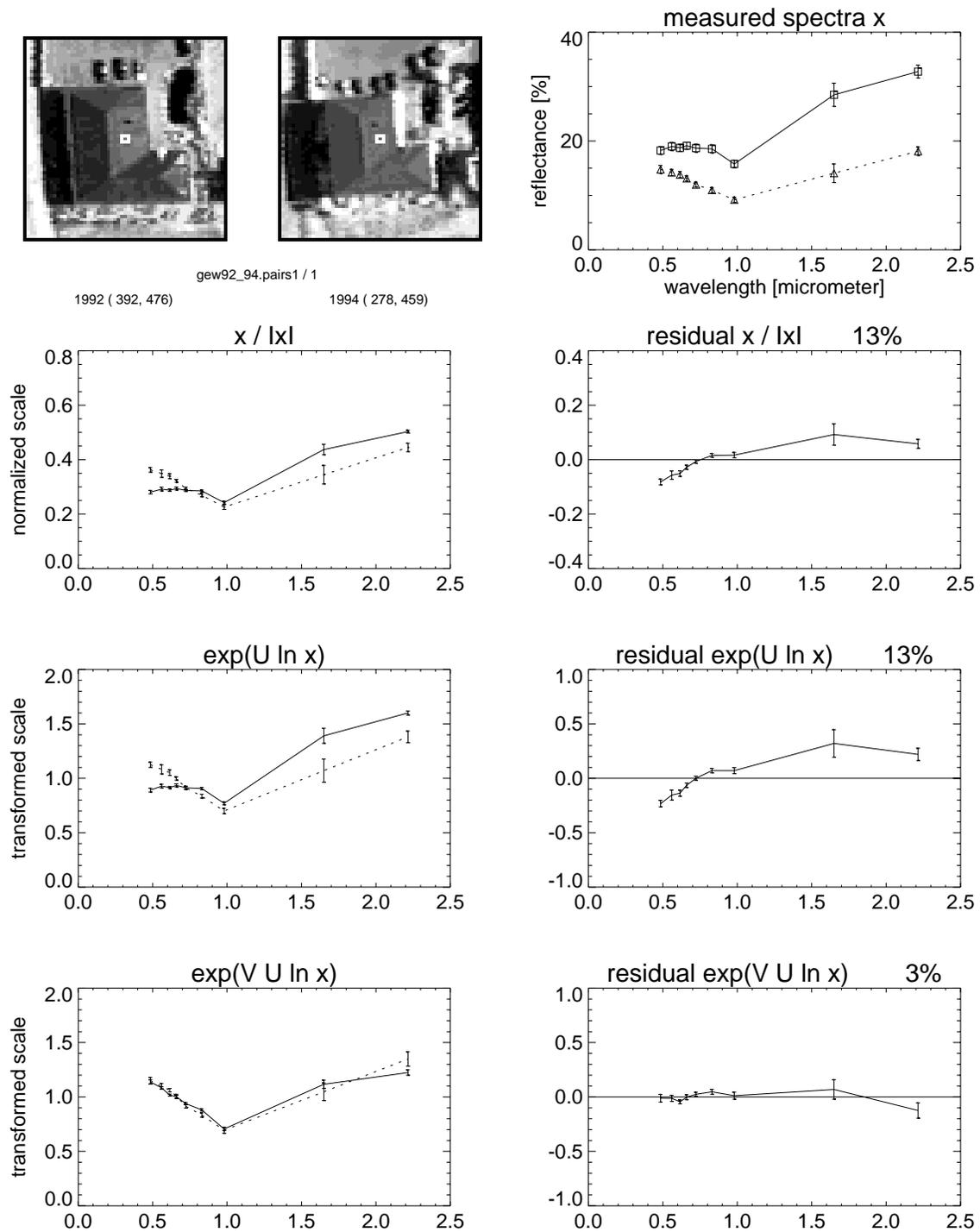


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene 'Gewerbegebiet' 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

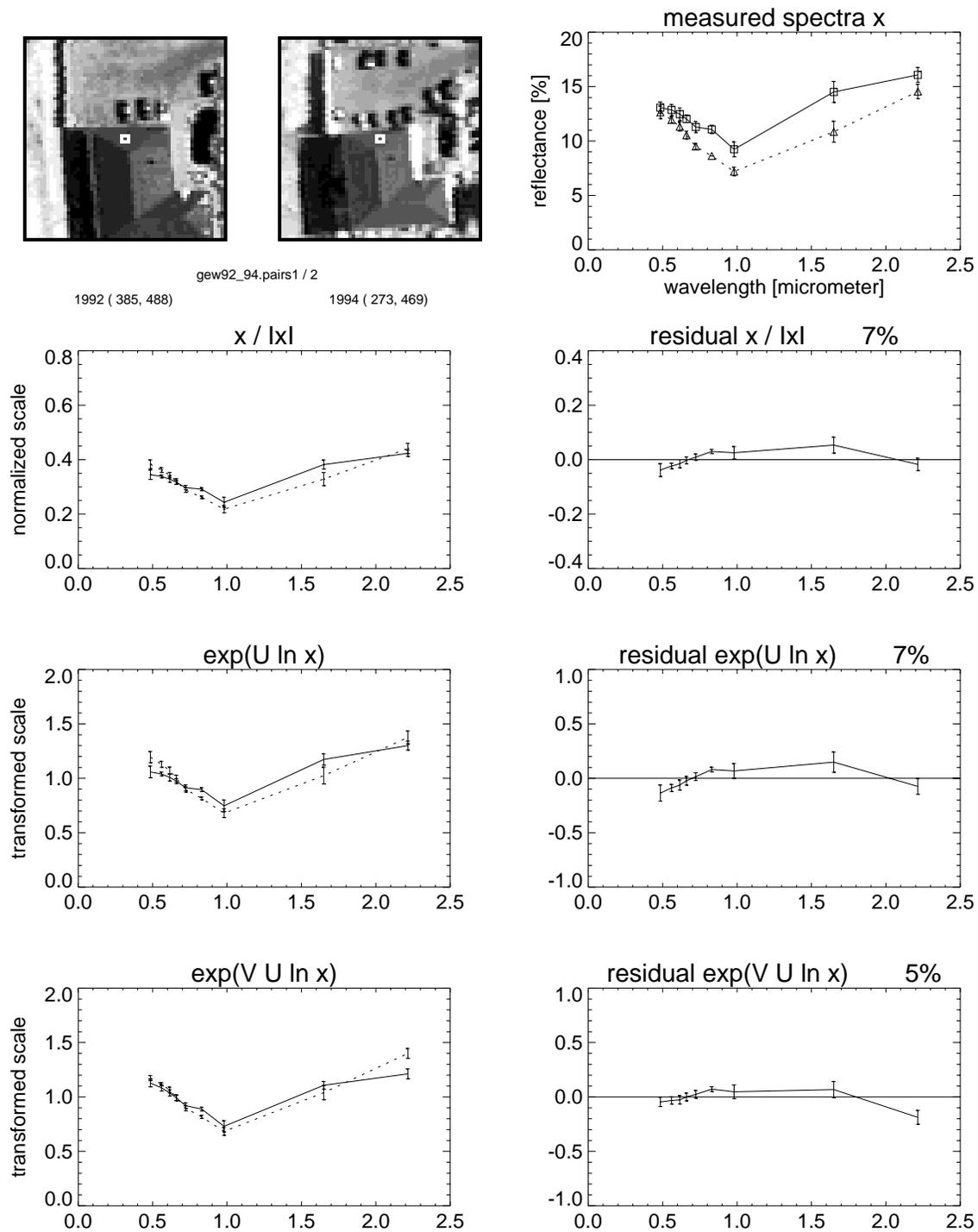


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene 'Gewerbegebiet' 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

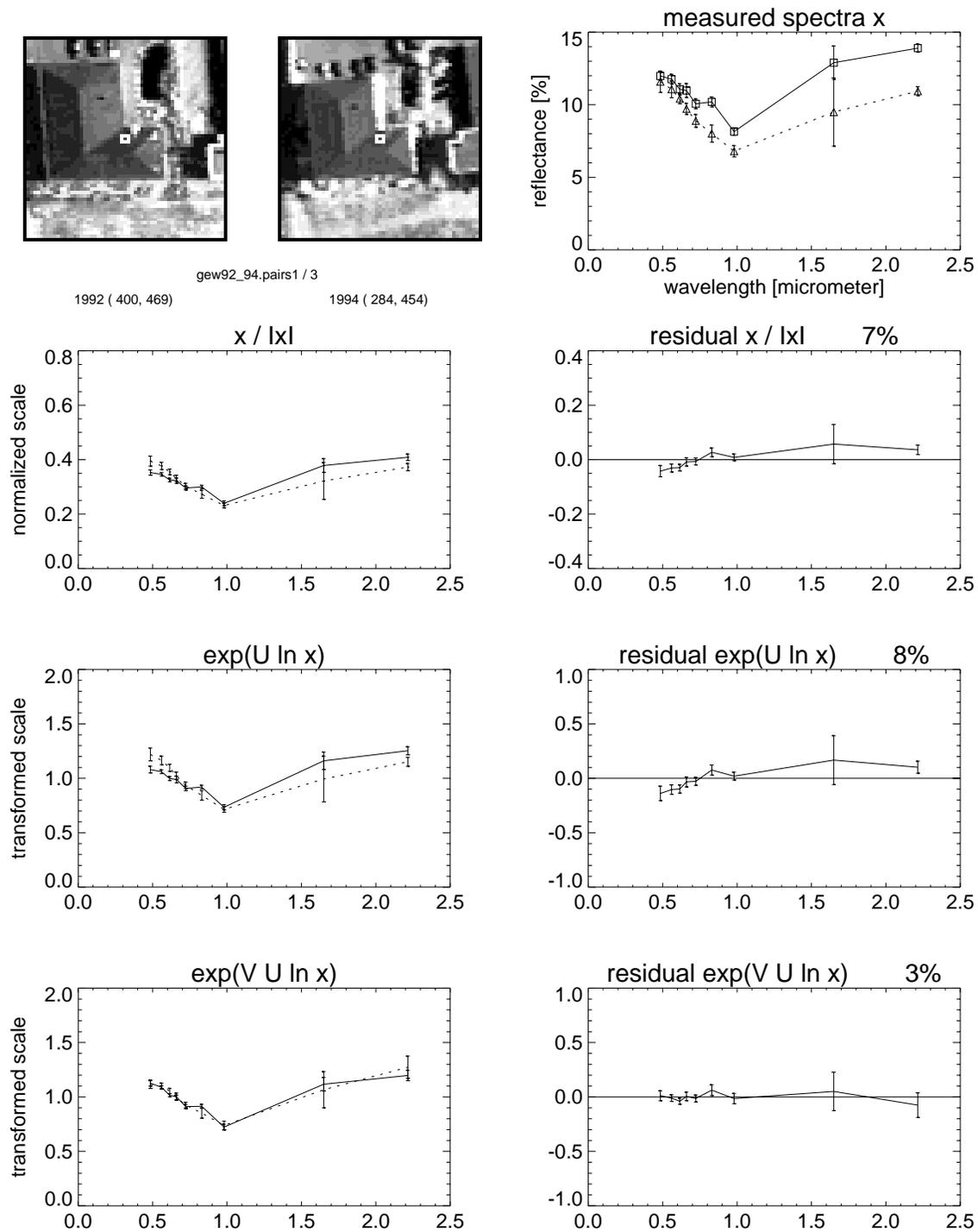


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene 'Gewerbegebiet' 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

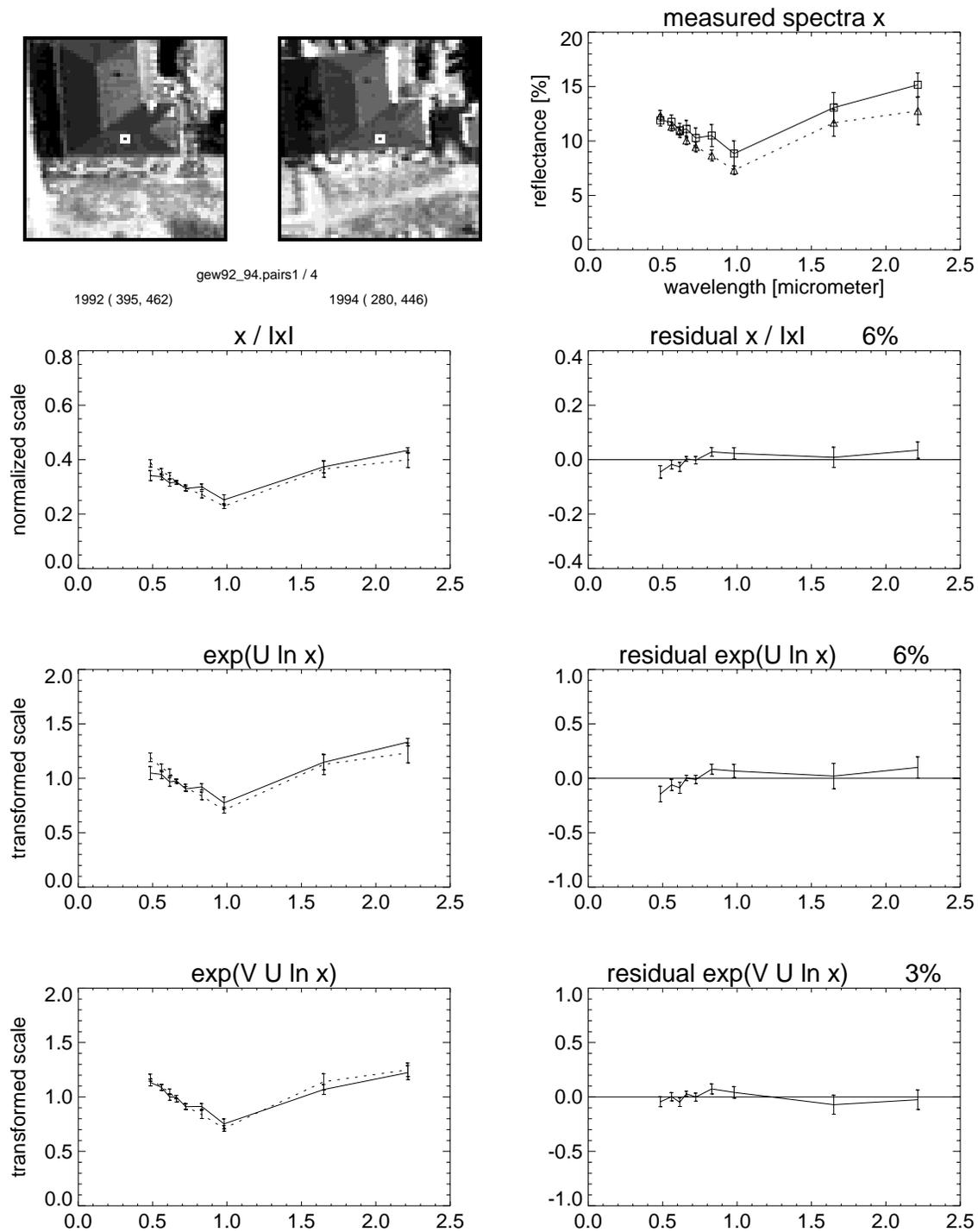


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene 'Gewerbegebiet' 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

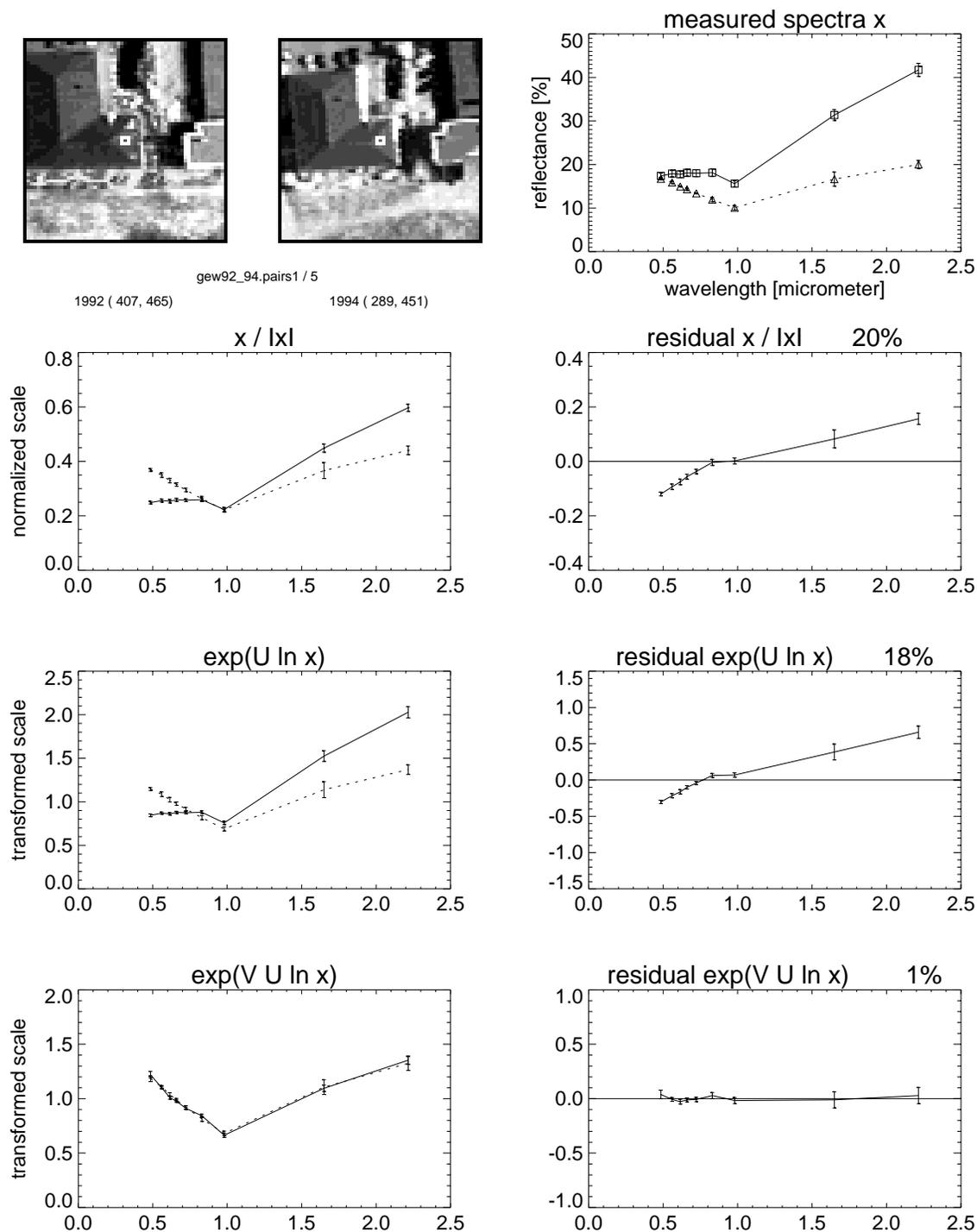


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene 'Gewerbegebiet' 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

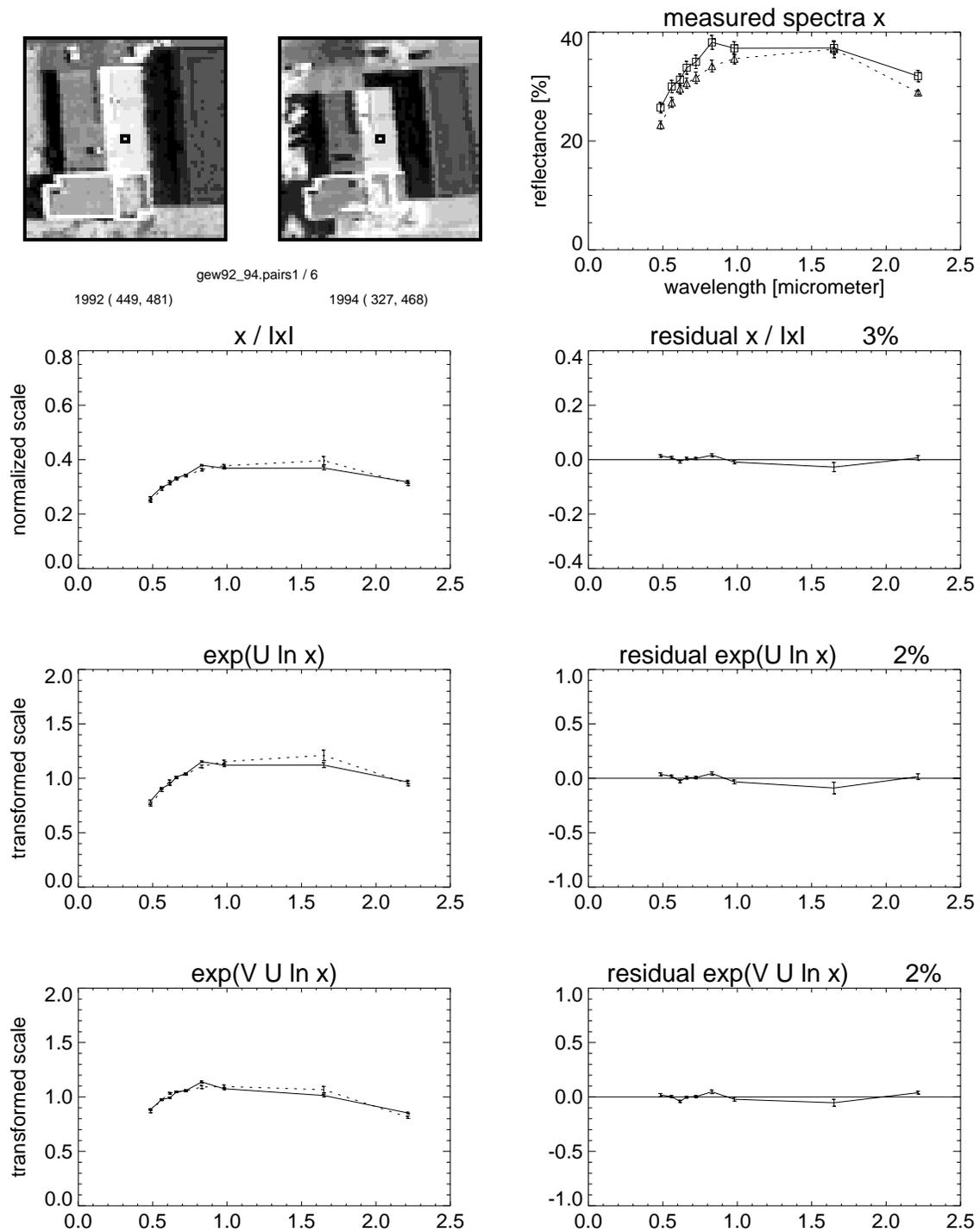


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene 'Gewerbegebiet' 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

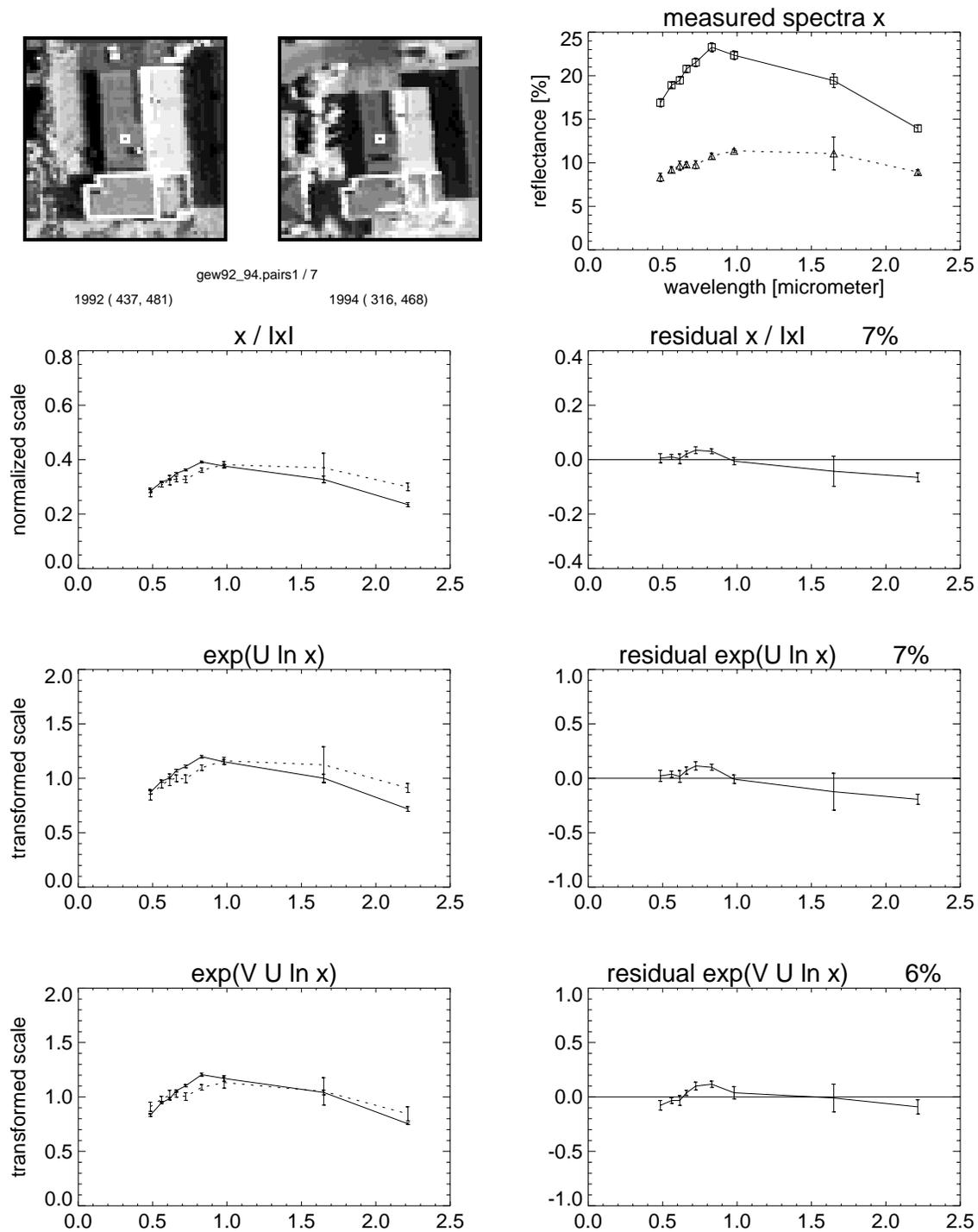


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

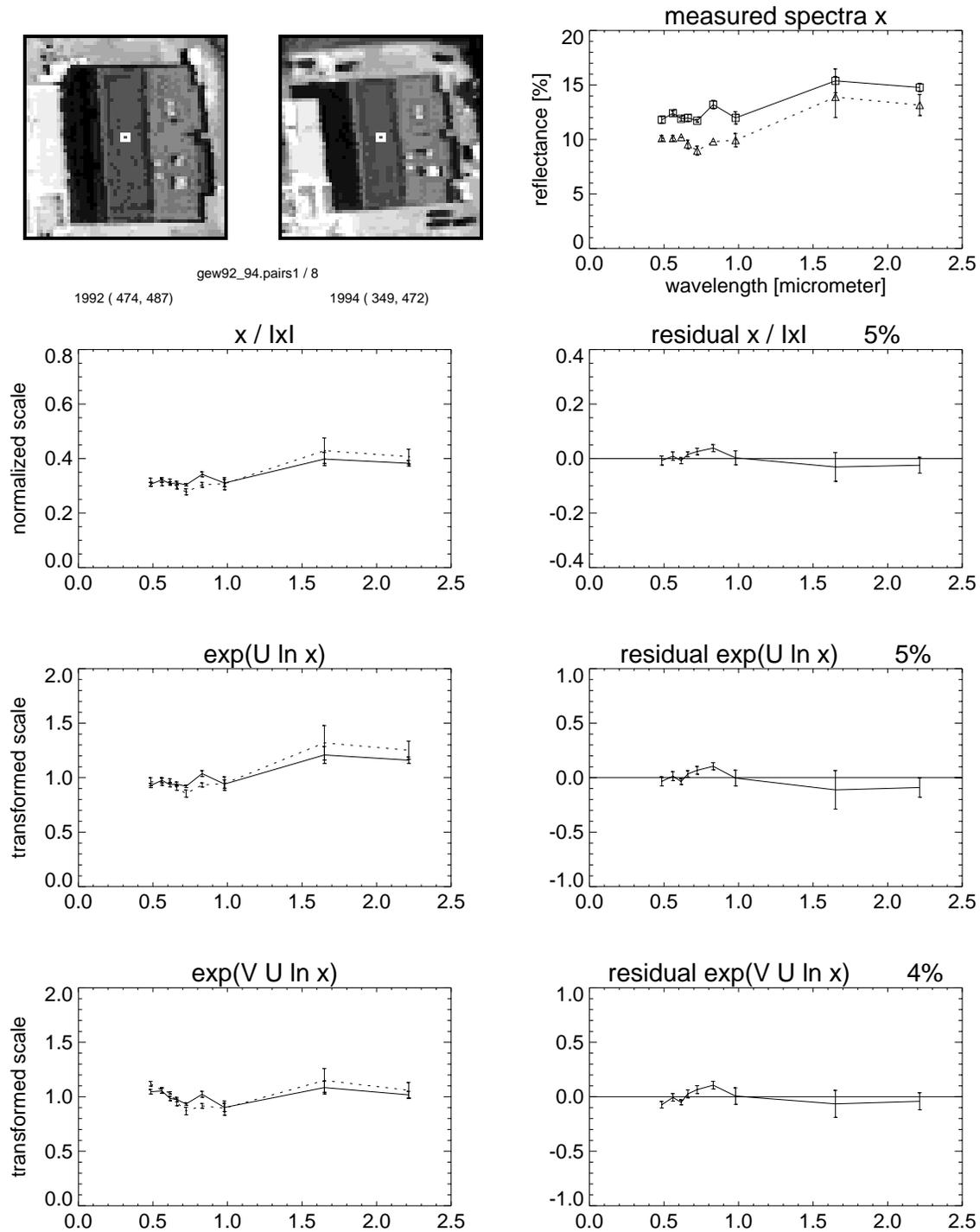


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).

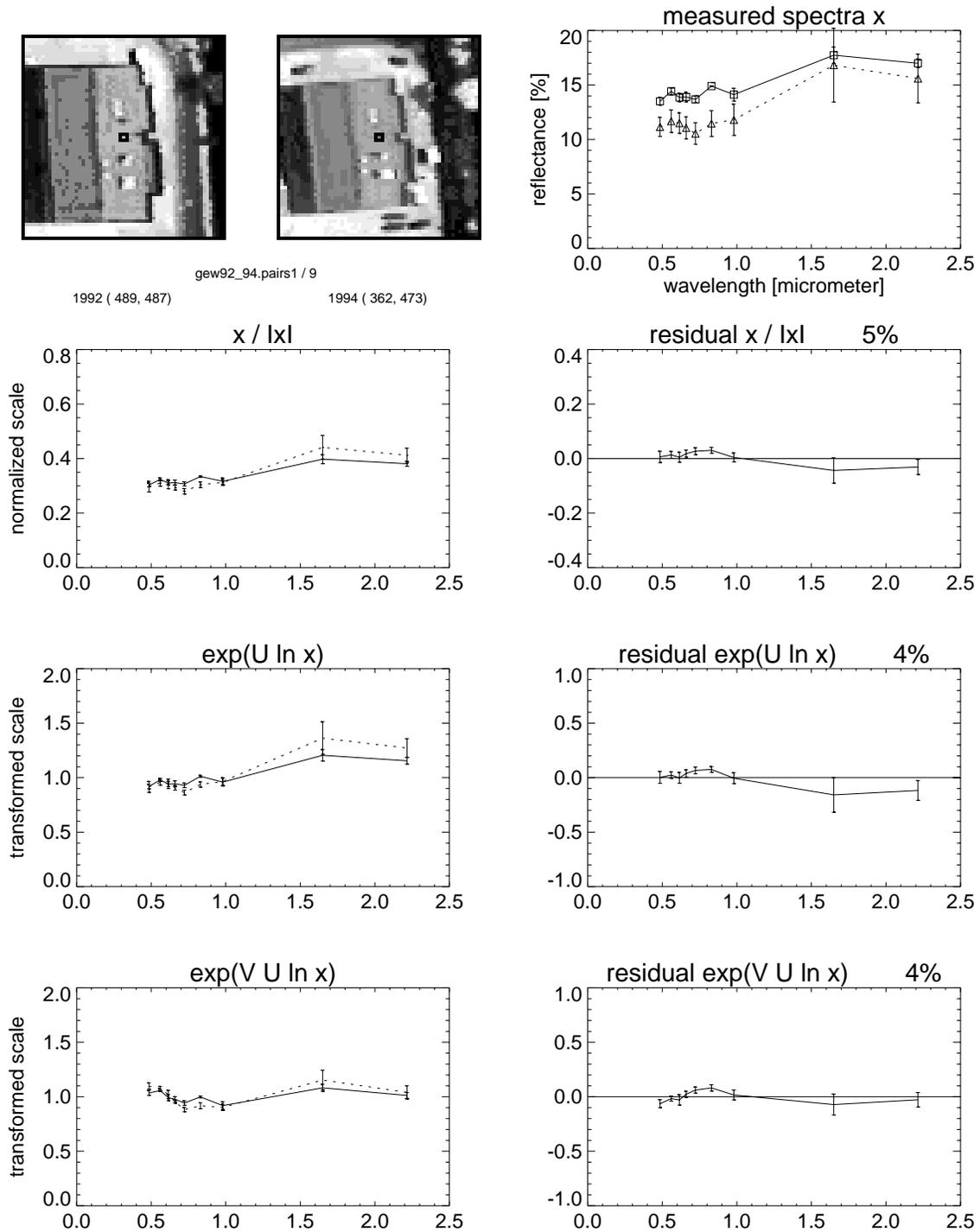


Figure C.5 (cont'd): Bitemporal spectral sample pair from scene ‘Gewerbegebiet’ 1992 and 1994 (Fig. 6.2, and Fig. 6.3 on page 121).



Brief Curriculum Vitae

Rafael Wiemker, married, two daughters

wiemker@informatik.uni-hamburg.de

<http://kogs-www.informatik.uni-hamburg.de/~wiemker>

<http://kogs-www.informatik.uni-hamburg.de/projects/Censis.html>

- 1966 born in Reinbek
two sisters
- 1986 Abitur in Lingen / Ems
- 1986–1987 assistant at the ‘Ravensburger’ publishing house
with Dr. Friedbert Stohner
- 1987–1989 2-year-volunteer program in Israel
in various Arab and Jewish projects
(Aktion Sühnezeichen / Friedensdienste)
- 1989–1993 student in physics and astronomy at the Universität Hamburg
scholarship of the CUSANUS-Foundation
- 1991 Vordiplom Physics with Astronomy
- 1991–1992 graduate student at the Georgia State University, Atlanta, USA,
with Prof. Douglas Gies (Center for High Angular Resolution Astronomy)
- 1992 Master of Science in Physics with Astronomy Concentration
*‘Tomographic Reconstruction of Faint Secondary Spectra
in O-type Binary Systems’*
- 1993–1997 PhD student at the Universität Hamburg
with Prof.s Hartwig Spitzer, Johann Bienlein, Leonie Dreschler-Fischer
(CENSIS)
research assistantship of the Volkswagen-Foundation
- 1994 marriage with Susanne Paulsen
- 1995 scholarship of the Universität Hamburg
- 1995 birth of our daughter Veronika
- 1997 birth of our daughter Thordis
- 1997 Doctorate in Physics
‘The Color Constancy Problem in Multispectral Remote Sensing’

Bibliography

- Accetta, J.S. and D.L. Shumaker, editors (1993). *The Infrared And Electro-Optical Systems Handbook*. Environmental Research Institute of Michigan (ERIM) and SPIE Optical Engineering Press, Ann Arbor and Bellingham, 1993.
- Albert, A. (1972). *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York, 1972.
- Albertz, J. (1991). *Grundlagen der Interpretation von Luft- und Satellitenbildern*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1991.
- Anderson, G.P., F.X. Kneizys, J.H. Chetwynd, J. Wang, M.L. Hoke, L.S. Rothman, L.M. Kimball, R.A. McClatchey, E.P. Shettle, S.A. Clough, W.O. Gallery, L.W. Abreu, and J.E.A. Selby (1995). FASCODE / MODTRAN / LOWTRAN : Present / Past / Future. In *Proceedings of the 18th Annual Review Conference on Atmospheric Transmission Models, June 6–8, 1995*.
- Ball, G. and D. Hall (1967). A Clustering Technique for Summarizing Multivariate Data. *Behavioral Sciences* **12**, 153–155, 1967.
- Ballard, D. and C. Brown (1982). *Computer Vision*. Prentice-Hall, Englewood Cliffs, 1982.
- Baraldi, A. and F. Parmiggiani (1995). A Neural Network for Unsupervised Categorization of Multivalued Input Patterns: An Application to Satellite Image Clustering. *IEEE Transactions on Geoscience and Remote Sensing* **33** (2), 305–316, 1995.
- Barnard, K., G. Finlayson, and B. Funt (1996). Colour Constancy for Scenes with Varying Illumination. In *European Conference on Computer Vision – ECCV '96*, Heidelberg, New York, 1996, pages 3–15. Springer.
- Bartsch, B., S. Bakan, and J. Fischer (1994). Remote Sensing of Water Vapour within the Solar Spectrum. In *Atmospheric Sensing and Modelling*, volume 2311, pages 197–206. SPIE, 1994.

- Bartsch, B. (1996). *Fernerkundung des Wasserdampfgehalts der Atmosphäre über Land aus rückgestreuter Sonnenstrahlung*. PhD thesis, Meteorologisches Institut Universität Hamburg, 1996.
- Ben-Dor, E., F. A. Kruse, A. B. Lefkoff, and A. Banin (1994). Comparison of Three Calibration Techniques for Utilization of GER 63-Channel Aircraft Scanner Data of Makhtesh Ramon, Negev, Israel. *Photogrammetric Engineering and Remote Sensing* **60** (11), 1339–1354, 1994.
- Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B* **48** (3), 259–302, 1986.
- Bezdek, J.C. (1973). *Fuzzy Mathematics in Pattern Classification*. PhD thesis, Applied Math Center, Cornell University, Ithaca, 1973.
- Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, London, 1981.
- Borel, C.C., S.A.W. Gerstl, and B.J. Powers (1991). The radiosity method in optical remote sensing of structured 3-D surfaces. *Remote Sensing of Environment* **36**, 1991.
- Brandt, S. (1992). *Datenanalyse*. BI-Wissenschaftsverlag, Mannheim, Leipzig, Wien, Zürich, 1992.
- Bronstein, I.N. and K.A. Semendjajew (1991). *Taschenbuch der Mathematik*. Teubner, Stuttgart, Leipzig, 25. edition, 1991.
- Buhmann, J. and H. Kühnel (1993). Vector Quantization with Complexity Costs. *IEEE Transactions on Information Theory* **39** (4), 1133–1145, 1993.
- Chavez, P. S. (1989). Radiometric Calibration of Landsat Thematic Mapper Multispectral Images. *Photogrammetric Engineering and Remote Sensing* **55**, 1285–1294, 1989.
- Dixon, E.R. (1978). Spectral Distribution of Australian Daylight. *Journal of the Optical Society of America* **68** (4), 437–450, 1978.
- Dreschler-Fischer, L., C. Drewniok, H. Lange, and C. Schröder (1993). A Knowledge-Based Approach to the Detection and Interpretation of Changes in Aerial Images. In Fujimura, S., editor, *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'93, Tokyo, August 1993, IEEE*, volume I, pages 159–161, 1993.
- Drew, M. S. and L. L. Kontsevich (1994). Closed-Form Attitude Determination Under Spectrally Varying Illumination. In *Proc. Conf. on Computer Vision and Pattern Recognition*, Seattle, Washington, June 1994, pages 985–990. IEEE Computer Society Press, Los Alamitos, CA, 1994.
- Duda, R. O. and P. E. Hart (1973). *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.

- Ehlers, M. (1994). Geometric Registration of Airborne Scanner Data Using Multiquadric Interpolation Techniques. In *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition, Strasbourg, Ann Arbor, 1994*, volume II, pages 492–502. Environmental Research Institut of Michigan.
- Eklundt, L. and A. Singh (1993). A Comparative Analysis of Standardized and Unstandardized Principal Components Analysis in Remote Sensing. *International Journal of Remote Sensing* **14** (7), 1359–1370, 1993.
- Fechner, Gustav Theodor (1860). *Elemente der Psychophysik*. Breitkopf und Haertel, Leipzig 1860, reprint Amsterdam 1964, 1860.
- Finlayson, G., S.S. Chatterjee, and B.V. Funt (1996). Color Angular Indexing. In *European Conference on Computer Vision – ECCV '96*, Heidelberg, New York, 1996, pages 16–27. Springer.
- Fischer, G. (1984). *Lineare Algebra*. Vieweg, Braunschweig, 1984.
- Forsyth, D.A. (1990). A Novel Algorithm for Color Constancy. *International Journal of Computer Vision* **5** (1), 5–36, 1990.
- Freeman, W. T. and D. H. Brainard (1995). Bayesian Decision Theory, the Maximum Local Mass Method, and Color Constancy. In *Proceedings of the 5th International Conference on Computer Vision – ICCV '95*, pages 210–217. IEEE, 1995.
- Fung, T. and E. LeDrew (1987). Application of Principal Components Analysis to Change Detection. *Photogrammetric Engineering and Remote Sensing* **53** (12), 1649–1658, 1987.
- Funt, B.V., M.S. Drew, and J. Ho (1991). Color Constancy from Mutual Reflection. *International Journal of Computer Vision* **6** (1), 5–24, 1991.
- Gershon, R., A.D. Jepson, and J.K. Tsotsos (1986). Ambient Illumination and the Determination of Material Changes. *Journal of the Optical Society of America* **3** (10), 1700–1707, 1986.
- Gerthsen, C., H.O. Kneser, and H. Vogel (1977). *Physik*. 13. edition, Springer, Heidelberg, 1977.
- Gonzalez, R.C. and P.A. Wintz (1987). *Digital Image Processing*. Addison-Wesley, Reading, 1987.
- Gustafson, D.E. and W. Kessel (1979). Fuzzy Clustering with a Fuzzy Covariance Matrix. In Fu, K.S., editor, *Proceedings IEEE-CDC*, Piscataway, New Jersey, 1979, volume 2, pages 761–766. IEEE Press.
- Häckel, H. (1985). *Meteorologie*. Ulmer Verlag, Stuttgart, 1985.
- Hall, F. G., D. E. Strebel, J. E. Nickeson, and S. J. Goetz (1991). Radiometric Rectification: Toward a Common Radiometric Response Among Multidate, Multisensor Images. *Remote Sensing of Environment* **35**, 11–27, 1991.

- Hämmerlin, G. and K.-H. Hoffmann (1994). *Numerische Mathematik, 4. edition*. Springer, Heidelberg, 1994.
- Hapke, B. (1993). *Theory of Reflectance and Emittance Spectroscopy*. Cambridge University Press, Cambridge, 1993.
- Harris, R. (1975). *A Primer of Multivariate Statistics*. Academic Press, New York, 1975.
- Hartigan, J.A. (1975). *Clustering Algorithms*. Wiley, New York, 1975.
- Harwit, M. (1988). *Astrophysical Concepts*. Springer, New York, Berlin, Heidelberg, 1988.
- Healey, G. (1986). Using Color for Geometry-Insensitive Segmentation. *Journal of the Optical Society of America* **6** (6), 920–937, 1986.
- Heipke, C. and C. Piechullek (1994). Towards Surface Reconstruction Using Multi Image Shape Form Shading. In Ebner, H., C. Heipke, and K. Eder, editors, *Proceedings of the ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision, Munich 1994*, volume 30 part 3/1 of *International Archives of Photogrammetry and Remote Sensing*, pages 361–369. SPIE volume 2357, 1994.
- Heipke, C. (1992). Integration of Digital Image Matching and Multi Image Shape Form Shading. In Pöppel, S.J. and H. Handels, editors, *Mustererkennung 1993. Proceedings of the 15. DAGM Symposium*, Informatik aktuell, pages 367–374. Springer, 1992.
- Helmholtz, H. (1896). *Handbuch der physiologischen Optik*. Voss, Hamburg, 1896.
- Hepp, T. (1994). Erzeugung multispektraler Reflektanzbilder zur automatisierten Bildauswertung, Diplomarbeit, Universität Hamburg, II. Institut für Experimentalphysik, CENSIS-REPORT-10-94, 1994.
- Ho, J., B. V. Funt, and M. S. Drew (1990). Separating a Color Signal into Illumination and Surface Reflectance Components: Theory and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12** (10), 966–977, October 1990.
- Horn, B. K. P. (1986). *Robot Vision*. MIT Press / McGraw-Hill, 1986.
- Hurcom, S.J., A.R. Harrison, and M. Taberner (1994). Factor Analysis of Semi-Arid Vegetation Response Using AVIRIS and Airborne Video Data. In *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition, Strasbourg*, Ann Arbor, 1994, pages II-424–436. Environmental Research Institut of Michigan.
- Jähne, B. (1993a). *Digitale Bildverarbeitung*. Springer, 1993.
- Jähne, B. (1993b). *Spatio-Temporal Image Processing*. Springer, 1993.

- Jahnen, W. and H. Grassl (1991). Correction of Atmospheric Masking as a Prerequisite of Ecological Mapping with Satellite Data. In Putkonen, J., editor, *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS 1991*, volume 2, pages 647–650. IEEE, 1991.
- Jänich, Klaus (1993). *Lineare Algebra*. Springer, Heidelberg, 1993.
- Jhung, Y. and Philip H. Swain (1996). Bayesian Contextual Classification Based on Modified M -Estimates and Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing* **34** (1), 67–75, 1996.
- Judd, D.B. and Günter Wyszecki (1975). *Color in Business, Science and Industry*. Wiley, New York, 1975.
- Judd, D. B., D. L. MacAdam, and G. Wyszecki (1964). Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature. *Journal of the Optical Society of America* **54** (8), 1031–1040, 1964.
- Klinker, G.J., S.A. Shafer, and T. Kanade (1987). Using a Color Reflection Model to Separate Highlights from Object Color. In *First International Conference on Computer Vision — ICCV 87*, London, England, June 8–11 1987, pages 145–170. IEEE Computer Society Press, 1987.
- Klinker, G.J., S.A. Shafer, and T. Kanade (1988). The Measurement of Highlights in Color Images. *International Journal on Computer Vision* **2** (1), 7–32, 1988.
- Klinker, G.J., S.A. Shafer, and T. Kanade (1990). A Physical Approach to Color Image Understanding. *International Journal on Computer Vision* **4** (1), 7–38, 1990.
- Kohonen, T. (1977). *Associative Memory – A System-Theoretical Approach*. Springer, Berlin, Heidelberg, New York, 1977.
- Kohonen, T. (1995). *Self-Organizing Maps*. Springer, Berlin, Heidelberg, New York, 1995.
- Kollewe, M., J. Bienlein, T. Kollewe, and H. Spitzer (1996). Comparison of Multi-spectral Airborne Scanner Reflectance Images With Ground Surface Reflectance Measurements. In *Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition, San Francisco*, Ann Arbor, 1996, volume III, pages 220–228. Environmental Research Institut of Michigan.
- Kollewe, T. (1995). Vergleich multispektraler Flugzeugscanneraufnahmen mit Reflektanzmessungen am Boden, Diplomarbeit, Universität Hamburg, II. Institut für Experimentalphysik, CENSIS-Report 17-96, 1995.
- Kontsevich, L.L., A.P. Petrov, and I.S. Vergelskaya (1994). Reconstruction of Shape From Shading in Color Images. *Journal of the Optical Society of America A* **11** (10), 1047–1052, 1994.
- Kraus, K. and W. Schneider (1988). *Fernerkundung, Band 1 – Physikalische Grundlagen und Aufnahmetechniken*. Dümmler, Bonn, 1988.

- Kriebel, K.T., W. Schlüter, and J. Sievers (1975). Zur Definition und Messung der spektralen Reflexion natürlicher Oberflächen. *Bildmessung und Luftbildwesen* **43** (1), 43–50, 1975.
- Kriebel, K. T. (1978). Measured Spectral Bidirectional Reflection Properties of Four Vegetated Surfaces. *Applied Optics* **17** (2), 253–258, 1978.
- Kruse, F. A., A. B. Lefkoff, J. B. Boardman, K. B. Heidebrecht, A.T. Shapiro, P. J. Barloon, and A. F. H. Goetz (1993). The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sensing of Environment* **44**, 145–163, 1993.
- Land, E.H. and J.J. McCann (1971). Lightness and Retinex Theory. *Journal of the Optical Society of America* **61** (1), 1–11, 1971.
- Lange, H. and C. Schröder (1994). Analysis and Interpretation of Changes in Aerial Images. In Ebner, H., C. Heipke, and K. Eder, editors, *Proceedings of the ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision, Munich, September 5–9, 1994*, volume 30 of *International Archives of Photogrammetry and Remote Sensing*, pages 475–482. SPIE volume 2357, 1994.
- Lillesand, T.M. and R.W. Kiefer (1987). *Remote Sensing and Image Interpretation*. Wiley, New York, 1987.
- Lozán, J.L. (1992). *Angewandte Statistik für Naturwissenschaftler*. Parey, Berlin, Hamburg, 1992.
- MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
- Maloney, L. T. and B. A. Wandell (1986). Color Constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A* **3**, 29–33, 1986.
- Matas, J., R. Marik, and J. Kittler (1994). Illumination Invariant Color Recognition. In Hancock, E., editor, *Proceedings of the British Machine Vision Conference*. BMVA Press, 1994.
- Mather, P.M. (1987). *Computer Processing of Remotely Sensed Images*. Wiley, New York, 1987.
- Meister, G., R. Wiemker, J. Bienlein, and H. Spitzer (1996). In Situ BRDF Measurements of Selected Surface Materials to Improve Analysis of Remotely Sensed Multispectral Imagery. In *Proceedings of the XVIII. Congress of the International Society for Photogrammetry and Remote Sensing ISPRS 1996, Vienna*, volume XXXI part B7 of *International Archives of Photogrammetry and Remote Sensing*, pages 493–498, 1996.

- Meister, G. (1995). Messung der bidirektionalen Reflektanzverteilungsfunktion (BRDF) ausgewählter Oberflächen unter natürlicher Beleuchtung, Diplomarbeit, Universität Hamburg, II. Institut für Experimentalphysik, CENSIS-Report 18-96., 1995.
- Minnaert, M.G.J. (1993). *Light and Colors in the Outdoors*. Springer, New York, Berlin, Heidelberg, 1993.
- Moore, A., J. Allman, and R.M. Goodman (1991). A Real-Time Neural System for Color Constancy. *IEEE Transactions on Neural Networks* **2** (2), 237–247, 1991.
- Moran, M. S., R. D. Jackson, F. H. Galen, P. N. Slater, R. J. Bartell, S. F. Biggar, D. I. Gellman, and R. P. Santer (1990). Obtaining Surface Reflectance Factors from Atmospheric and View Angle Corrected SPOT-1 HRV Data. *Remote Sensing of Environment* **32**, 203–214, 1990.
- Nagao, K. and W.E.L. Grimson (1995). Recognizing 3D Objects Using Photometric Invariant. Technical report, Artificial Intelligence Lab, MIT, Cambridge, Massachusetts, 1995.
- Nicodemus, F.E. (1970). Reflectance Nomenclature and Directional Reflectance and Emissivity. *Applied Optics* **9** (6), 1474–1475, 1970.
- Nielsen, A.A. and R. Larsen (1994). Restoration of GERIS Data Using the Maximum Noise Fractions Transform. In *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition, Strasbourg*, Ann Arbor, 1994, volume II, pages 557–563. Environmental Research Institut of Michigan.
- Ohta, Y. and Y. Hayashi (1994). Recovery of Illuminant and Surface Colors from Images Based on the CIE Daylight. In Eklundh, J.-O., editor, *European Conference on Computer Vision – ECCV '94*, Heidelberg, New York, 1994, pages 235–246. Springer.
- Oppenheim, A. V. and R. W. Schaffer (1975). *Digital Signal Processing*. Prentice-Hall, London, 1975.
- Oren, Michael and Shree K. Nayar (1994). Seeing Beyond Lambert's Law. In Eklundh, J.-O., editor, *European Conference on Computer Vision – ECCV '94*, Heidelberg, New York, 1994, pages 269–280. Springer.
- Oren, M. and S.K. Nayar (1995). Generalization of the Lambertian Model and Implications for Machine Vision. *International Journal of Computer Vision* **14**, 227–251, 1995.
- Petrov, A.P. and L.L. Kontsevich (1994). Properties of Color Images of Surfaces under Multiple Illuminants. *Journal of the Optical Society of America A* **11** (10), 2745–2749, 1994.
- Piech, K. R. and J. E. Walker (1974). Interpretation of Soils. *Photogrammetric Engineering and Remote Sensing* **40**, 87–94, 1974.

- Press, S. J. (1972). *Applied Multivariate Analysis*. Holt, Rinehart and Winston, London, New York, 1972.
- Pyka, K. and K. Steinnocher (1994). Auswahl eines optimalen Datensatzes für die multispektrale Klassifizierung unter Einbeziehung von Texturmerkmalsbildern. *Zeitschrift für Photogrammetrie und Fernerkundung* **4**, 116–122, 1994.
- Rees, W.G. (1990). *Physical Principles of Remote Sensing*. Cambridge University Press, Cambridge, 1990.
- Richards, J. A. (1993). *Remote Sensing Digital Image Analysis*. Springer, Heidelberg, New York, 1993.
- Richter, R. (1990). A Fast Atmospheric Correction Algorithm Applied to Landsat TM Images. *International Journal of Remote Sensing* **11** (1), 159–166, 1990.
- Richter, R. (1992). *Radiometrische Auslegung von Sensoren und quantitative Auswertung von Fernerkundungsdaten im optischen Spektralbereich*. PhD thesis, German Aerospace Research Establishment (DLR), Oberpfaffenhofen, 1992.
- Ryan, R., P. Del Guidice, L. Smith, M. Soel, N. Fonneland, M. Pagnutti, R. Irwin, and P. Saatzter (1996). U.S. Open Skies Follow-On Sensors Evaluation Program, Multispectral Hyperspectral (MSHS) Sensor Survey. In *Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition, San Francisco, Ann Arbor, 1996*, volume I, pages 392–402. Environmental Research Institut of Michigan.
- Sabins, F. F. (1978). *Remote Sensing: Principles and Interpretation*. Freeman, New York, 1978.
- Schlüns, K. (1992). Colormetric Stereo. In Klette, R. and W.G. Kropatsch, editors, *Proceedings of the Fifth Workshop on Theoretical Foundations of Computer Vision, Buckow, Berlin, 1992*, pages 180–190. Akademie Verlag.
- Schott, J. R., C. Salvaggio, and W. J. Volchok (1988). Radiometric Scene Normalization Using Pseudoinvariant Features. *Remote Sensing of Environment* **26**, 1–16, 1988.
- Schott, J. R. (1993). Methods for Estimation of and Correction for Atmospheric Effects on Remotely Sensed Data. In Kohnle, A. and W. B. Miller, editors, *Atmospheric Propagation and Remote Sensing*, volume SPIE 1968, page 448. SPIE, April 1993.
- Shafer, S. A. (1985). Using Color to Separate Reflection Components. *COLOR Research and Application* **10** (4), 210–218, 1985.
- Shahshahani, B. M. and D. A. Landgrebe (1994). The Effect of Unlabeled Samples in Reducing the Small Size Problem and Mitigating the Hughes Phenomenon. *IEEE Transactions on Geoscience and Remote Sensing* **32** (5), 1087–1095, September 1994.
- Singh, A. (1989). Review Article: Digital Change Detection Techniques Using Remotely-Sensed Data. *International Journal of Remote Sensing* **10** (6), 989–1003, 1989.

- Singh, A. (1993). Principal Components Analysis in Remote Sensing. In Fujimura, Sadao, editor, *1993 International Geoscience and Remote Sensing Symposium (IGARSS'93)*, Tokyo, August 1993, page 1680. Institute of Electrical and Electronics Engineers, 1993.
- Skarbek, W. and A. Koschan (1994). Colour Image Segmentation – A Survey. Technical report, Technical University Berlin, 1994.
- Solberg, A.H.S., T. Taxt, and A.K. Jain (1996). A Markov Random Field Model for Classification of Multisource Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing* **34** (1), 100–113, 1996.
- Spitzer, H., R. Wiemker, A. Speck, and J. Bienlein (1997). Robuste unbeaufsichtigte Änderungsdetektion auf Multispektralbildern im Rahmen des Open-Skies-Vertrages. In Altmann, J. and G. Neuneck, editors, *Naturwissenschaftliche Beiträge zu Abrüstung und Verifikation, Verhandlungen der Fachsitzung der 61. Physikertagung der Deutschen Physikalischen Gesellschaft (DPG) in München 1997*. DPG / FONAS (Math.Seminar, Bundesstr. 55, D-20146 Hamburg), 1997. In print.
- Spitzer, H. (1997a). The Open Skies Treaty: A Cooperative Approach to Confidence Building and Verification. In Schroerer, D. and A. Pascolini, editors, *The Weapons Legacy of the Cold War – Problems and Opportunities*, pages 163–176. Ashgate, Aldershot, 1997.
- Spitzer, H. (1997b). Potential of the Open Skies Regime and Sensor Suite for Environmental Monitoring. In *Proceedings of the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen*, Ann Arbor, 1997, volume I, pages 9–16. Environmental Research Institut of Michigan.
- Torrance, K. and E. Sparrow (1967). Theory for Off-Specular Reflection from Rough Surfaces. *Journal of the Optical Society of America* **57** (9), 1105–1114, 1967.
- Trivedi, M. and J.C. Bezdek (1986). Low-level Segmentation of Aerial Images with Fuzzy Clustering. *IEEE Transactions on Systems, Man, and Cybernetics* **16** (4), 589–598, 1986.
- Valko, P. (1977). Meteorologische Strahlungsmessungen mit fahrbarer Station. In Häckel, H., editor, *Meteorologie*. Ulmer Verlag, Stuttgart (1985), 1977.
- Venkateswarlu, N. B. and R. P. Singh (1995). A Fast Maximum Likelihood Classifier. *International Journal of Remote Sensing* **16** (2), 313–320, 1995.
- Wandell, B. A. (1987). The Synthesis and Analysis of Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9** (1), 2–13, 1987.
- Warnecke, G. (1991). *Meteorologie und Umwelt*. Springer, Heidelberg, 1991.
- Wegener, M. (1990). Destriping Multiple Sensor Imagery by Improved Histogram Matching. *Remote Sensing of Environment* **11**, 859–875, 1990.

- Wiemker, R. and T. Hepp (1994). Surface Orientation Invariant Matching of Spectral Signatures. In Ebner, H., C. Heipke, and K. Eder, editors, *Proceedings of the ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision, Munich 1994*, volume 30 part 3/2 of *International Archives of Photogrammetry and Remote Sensing*, pages 916–923. SPIE volume 2357, 1994.
- Wiemker, R. and H. Spitzer (1996). Änderungsdetektion auf multispektralen Luftbildern – Perspektiven für den Open-Skies-Vertrag. In Altmann, J. and G. Neuneck, editors, *Naturwissenschaftliche Beiträge zu Abrüstung und Verifikation, Verhandlungen der Fachsitzung der 60. Physikertagung der Deutschen Physikalischen Gesellschaft (DPG) in Jena 1996*, pages 138–151. DPG / FONAS (Math.Seminar, Bundesstr. 55, D-20146 Hamburg), 1996.
- Wiemker, R., K. Rohr, L. Binder, R. Sprengel, and H.S. Stiehl (1996). Application of Elastic Registration to Imagery from Airborne Scanners. In *Proceedings of the XVIII. Congress of the International Society for Photogrammetry and Remote Sensing ISPRS 1996, Vienna*, volume XXXI part B4 of *International Archives of Photogrammetry and Remote Sensing*, pages 949–954, 1996.
- Wiemker, R., A. Speck, D. Kulbach, H. Spitzer, and J. Bienlein (1997). Unsupervised Robust Change Detection on Multispectral Imagery Using Spectral and Spatial Features. In *Proceedings of the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen, Ann Arbor, 1997*, volume I, pages 640–647. Environmental Research Institut of Michigan.
- Wiemker, R. (1995a). The Color Constancy Problem: A Fast Illumination Invariant Mapping Approach. In Hlavac, V. and R. Sara, editors, *Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns, Prague 1995, CAIP'95*, Lecture Notes in Computer Science vol. 970, pages 950–955. Springer, 1995.
- Wiemker, R. (1995b). Das Farbkonstanzproblem in der multispektralen Fernerkundung. In Rehrmann, V., editor, *1. Workshop Farbbildverarbeitung, Koblenz 1995*, number 15/95 in *Fachberichte Informatik*, pages 48–51. Universität Koblenz-Landau, 1995.
- Wiemker, R. (1995c). Improved Color Constant Classification of Remotely Sensed Multispectral Imagery. In Stein, T.I., editor, *Proceedings of the International Geoscience and Remote Sensing Symposium, Florence, IGARSS 1995*, pages 1153–1155. IEEE Catalog 95CH35770, 1995.
- Wiemker, R. (1996). Registration of Airborne Scanner Imagery Using Akima Local Quintic Polynomial Interpolation. In *Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition, San Francisco, Ann Arbor, 1996*, volume III, pages 210–219. Environmental Research Institut of Michigan.
- Wiemker, R. (1997a). An Iterative Spectral-Spatial Bayesian Labeling Approach for Unsupervised Robust Change Detection on Remotely Sensed Multispectral Imagery. In Sommer, G., K. Daniilidis, and J. Pauli, editors, *Proceedings of the 7th International Conference on Computer Analysis of Images and Patterns, Kiel 1997*,

- CAIP'97*, Lecture Notes in Computer Science vol. 1296, pages 263–270. Springer, 1997.
- Wiemker, R. (1997b). Unsupervised Fuzzy Classification of Multispectral Imagery Using Spatial-Spectral Features. In Balderjahn, I., R. Mathar, and M. Schader, editors, *Data Highways and Information Flooding, a Challenge for Classification and Data Analysis, Proceedings of the 21. Annual Meeting of the Gesellschaft für Klassifikation, GfKl'97, Potsdam, March 12–14*, Heidelberg, 1997. Springer. In print.
- Wolff, L.B. (1996). Generalizing Lambert's Law for Smooth Surfaces. In Buxton, B. and R. Cipolla, editors, *Proceedings of the European Conference on Computer Vision, ECCV'96, Springer, Berlin*, pages 40–53, 1996.
- Zhang, W., J. Albertz, and Z. Li (1994). Digital Orthoimage From Airborne Line Scanner Imagery Utilizing Flight Parameters. In Ebner, H., C. Heipke, and K. Eder, editors, *Proceedings of the ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision, Munich 1994*, volume 30 part 3/2 of *International Archives of Photogrammetry and Remote Sensing*, pages 945–950. SPIE volume 2357, 1994.