

# **Biologically Inspired Visual Navigation in Indoor and Outdoor Environments**

Der Technischen Fakultät  
der Universität Bielefeld

vorgelegt von

Frank Röben

zur Erlangung des akademischen Grades  
eines Doktor-Ingenieurs

Tag der Disputation: 30. September 2009



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Outline and Contributions . . . . .	2
1.3	Insect Visual System . . . . .	3
1.4	Visual Homing . . . . .	7
1.5	Outdoor Navigation with Color Contrasts . . . . .	9
1.6	Cleaning Robot Control . . . . .	10
<b>2</b>	<b>Objectives</b>	<b>12</b>
2.1	Polarized-Light Compass . . . . .	12
2.2	Local Visual Homing Methods . . . . .	12
2.3	Outdoor Navigation with Color Contrasts . . . . .	13
2.4	Cleaning Robot Control . . . . .	14
<b>3</b>	<b>Materials and Methods</b>	<b>16</b>
3.1	Image Databases . . . . .	16
3.2	Quality Measures for Navigation . . . . .	19
3.2.1	Average Angular Error . . . . .	19
3.2.2	Median Angular Error . . . . .	20
3.2.3	Return Ratio . . . . .	20
3.3	Statistical Methods . . . . .	21
3.3.1	Jackknife . . . . .	21
3.3.2	Bootstrap . . . . .	22
3.3.3	Fisher’s Linear Discriminant Analysis . . . . .	23
3.4	Scale Invariant Feature Transform (SIFT) . . . . .	24
3.4.1	Scale-space and Detection of Extrema . . . . .	25
3.4.2	Localization of Keypoints . . . . .	27
3.4.3	Assigning an Orientation . . . . .	29
3.4.4	Computation of Keypoint Descriptor . . . . .	30
3.5	Image Parameters . . . . .	31
3.5.1	Statistics without Invariance in Rotation . . . . .	32
3.5.2	Statistics with Invariance in Rotation . . . . .	32

3.5.3	DFT Parameters without Invariance in Rotation . . . . .	33
3.5.4	DFT Parameters with Invariance in Rotation . . . . .	34
<b>4</b>	<b>Polarized-Light Compass</b>	<b>36</b>
4.1	Hardware . . . . .	37
4.2	Tests . . . . .	37
4.2.1	Polarized Filter Test . . . . .	38
4.2.2	Polarized-Light vs. Electromagnetic Compass . . . . .	38
4.2.3	Measuring the <i>e</i> -vector Field of the Sky . . . . .	41
4.2.4	Discussion . . . . .	43
4.3	Summary and Conclusion . . . . .	43
<b>5</b>	<b>Local Visual Homing Methods</b>	<b>45</b>
5.1	SIFT Homing . . . . .	45
5.1.1	Movement Vector for Correlated Keypoints . . . . .	45
5.1.2	Home Vector for Correlated Keypoint Descriptors . . . . .	46
5.1.3	Overall Home Vector . . . . .	49
5.1.4	Tests . . . . .	50
5.1.5	Discussion . . . . .	54
5.2	Parameter Models . . . . .	56
5.2.1	Tests . . . . .	58
5.2.2	Discussion . . . . .	66
5.3	Summary and Conclusion . . . . .	67
<b>6</b>	<b>Outdoor Navigation with Color Contrasts</b>	<b>68</b>
6.1	Five-Channel Hand-held Sensor . . . . .	68
6.2	Spectral Contrast Analysis for Natural Objects . . . . .	71
6.2.1	Data Collection . . . . .	71
6.2.2	Discrimination . . . . .	71
6.2.3	Statistical Tests . . . . .	75
6.2.4	Discussion . . . . .	75
6.3	Spectral Contrast Analysis for Artificial Objects . . . . .	77
6.3.1	Data Collection . . . . .	77
6.3.2	Discrimination . . . . .	77
6.3.3	Statistical Tests . . . . .	77
6.3.4	Discussion . . . . .	81
6.4	2D- Scene Contrast Analysis . . . . .	82
6.4.1	Data Collection . . . . .	83
6.4.2	Contrast Stability over Time . . . . .	86
6.4.3	Contrast Stability over Time after Skyline Separation . . . . .	89
6.4.4	Discussion . . . . .	91

6.5	Visual Homing with Skyline Information . . . . .	91
6.5.1	Exemplary Skyline Detection . . . . .	91
6.5.2	Navigation Dataset . . . . .	99
6.5.3	Discussion . . . . .	100
6.6	Summary and Conclusion . . . . .	100
<b>7</b>	<b>Cleaning Robot Control</b>	<b>102</b>
7.1	Trajectory Controller . . . . .	102
7.2	Tests . . . . .	103
7.3	Discussion . . . . .	106
7.4	Summary and Conclusion . . . . .	107
<b>8</b>	<b>Overall Conclusion</b>	<b>108</b>
<b>9</b>	<b>Future Research</b>	<b>110</b>
9.1	Polarized-Light Compass . . . . .	110
9.2	Local Visual Homing . . . . .	110
9.3	Outdoor Navigation . . . . .	111
9.4	Cleaning Robot Control . . . . .	111
	<b>Bibliography</b>	<b>112</b>
<b>A</b>	<b>Polarized-light Compass, Additional Figures</b>	<b>120</b>
A.1	Compass Filter . . . . .	120
A.2	Curcuit Diagram . . . . .	121
A.3	Additional Drawings . . . . .	123
<b>B</b>	<b>SIFT Homing</b>	<b>124</b>
B.1	Additional Tables . . . . .	124
<b>C</b>	<b>Parameter-Navigation</b>	<b>126</b>
C.1	Homing Using Non-Rotation Invariant Parameter . . . . .	126
C.2	Homing Using Rotation Invariant Parameter . . . . .	128
C.3	Parameter-Navigation, Additional Figures . . . . .	129
<b>D</b>	<b>Contrasts, Additional Figures</b>	<b>135</b>
D.1	Natural Contrasts . . . . .	135
D.2	Artificial Contrasts . . . . .	136
D.3	2D-Contrast Measures . . . . .	137

# List of Figures

1.1	Reflected irradiance of different flowers . . . . .	4
1.2	Peak sensitivities of the bees' receptors . . . . .	5
1.3	3-D representation of the pattern of polarization . . . . .	6
1.4	Classification of homing methods . . . . .	7
1.5	Irradiance of blue sky and vegetation . . . . .	10
1.6	Object and sky samples (log G, log UV) . . . . .	11
2.1	Example of topological map building . . . . .	14
3.1	ActiveMedia Pioneer 2-DX and 3-DX . . . . .	16
3.2	SIFT scale space . . . . .	26
3.3	Detection of extrema in scale space . . . . .	27
3.4	SIFT keypoint descriptor . . . . .	30
3.5	Concentrical rings on panoramic images . . . . .	32
4.1	Polarized-light compass, relative filter sensitivity, normalized . .	38
4.2	Voltage output of the POL-OP-Units . . . . .	39
4.3	Test setup polarized-light compass vs. electromagnetic compass	39
4.4	Averaged compass output, manual rotation . . . . .	40
4.5	Manual rotation of the test setup without averaging . . . . .	41
4.6	Polar plots of the measured $e$ -vector field . . . . .	42
4.7	Polar transformation diagram . . . . .	43
5.1	Top view of the homing situation . . . . .	47
5.2	Side view of the homing situation . . . . .	48
5.3	Example angle histograms, <i>original</i> database . . . . .	49
5.4	$AAE_*$ , $MAE_*$ , and $RR_*$ for Butterworth cutoff from 0.03 to 1.0	51
5.5	$AAE_*$ and $RR_*$ for different image sizes, <i>original</i> database . . .	52
5.6	Results cross database tests . . . . .	55
5.7	Example home vector array and potential function . . . . .	57
5.8	Design of tests for parameter homing methods . . . . .	59
5.9	The eight best parameter methods for <i>original</i> database . . . . .	61
5.10	The eight best parameter methods for <i>hall1</i> database . . . . .	63
5.11	The eight best parameter methods for <i>uni</i> database . . . . .	65

6.1	Five-channel hand-held sensor . . . . .	69
6.2	Glass filter spectral sensitivity . . . . .	70
6.3	UV-G and UV-IR sample data . . . . .	73
6.4	Plotted data for natural objects . . . . .	74
6.5	Plotted data for artificial objects . . . . .	79
6.6	2-D pan-tilt sensor . . . . .	82
6.7	Stitched DSLR camera image dataset one . . . . .	83
6.8	Raw image data, panorama dataset one . . . . .	84
6.9	Raw image data, panorama dataset two . . . . .	85
6.10	logNRMSD . . . . .	87
6.11	logMSEstd . . . . .	88
6.12	logMSEvar . . . . .	88
6.13	logNRMSD for skyline separated contrasts . . . . .	89
6.14	logMSEstd for skyline separated contrasts . . . . .	90
6.15	logMSEvar for skyline separated contrasts . . . . .	90
6.16	Separated dataset one . . . . .	93
6.17	Separated dataset two . . . . .	94
6.18	Separation images UV, dataset one . . . . .	95
6.19	Separation images B, dataset one . . . . .	96
6.20	Separation images G, dataset one . . . . .	96
6.21	Separation images R, dataset one . . . . .	97
6.22	Separation images UV, dataset two . . . . .	97
6.23	Separation images B, dataset two . . . . .	98
6.24	Separation images G, dataset two . . . . .	98
6.25	Separation images R, dataset two . . . . .	99
6.26	SIFT home vectors of all contrasts, square test . . . . .	99
6.27	Outdoor UV channel images for homing . . . . .	100
7.1	Triangulation trajectory controller . . . . .	103
7.2	Ideal cleaning trajectories . . . . .	104
7.3	Cleaning results long lanes . . . . .	105
7.4	Cleaning results short lanes . . . . .	106
A.1	Polarized-light compass, unnormalized relative filter sensitivity . . . . .	120
A.2	Polarized-light compass circuit diagram . . . . .	121
A.3	Exploded assembly drawing of the polarized-light compass . . . . .	123
C.1	$RR_*$ and $AAE_*$ values ranking 9 - 12 from <i>original</i> database . . . . .	129
C.2	$RR_*$ and $AAE_*$ values ranking 13 - 19 from <i>original</i> database . . . . .	130
C.3	$RR_*$ and $AAE_*$ values ranking 9 - 16 from <i>hall1</i> database . . . . .	131
C.4	$RR_*$ and $AAE_*$ values ranking 17 - 19 from <i>hall1</i> database . . . . .	132
C.5	$RR_*$ and $AAE_*$ values ranking 5 - 12 from <i>uni</i> database . . . . .	133

C.6	$RR_*$ and $AAE_*$ values ranking 13 - 19 from <i>uni</i> database . . . .	134
D.1	Histograms of the projected natural data points. . . . .	135
D.2	Histograms of the projected artificial data points . . . . .	136
D.3	MSE and PSNR . . . . .	137
D.4	CC . . . . .	138
D.5	MSE . . . . .	138
D.6	PSNR and CC . . . . .	139
D.7	One-day dataset: 11:27:04, 12:06:19, 12:46:02, 13:25:05 . . . .	140
D.8	One-day dataset: 14:07:52, 14:47:21, 15:26:36, 16:07:15 . . . .	141
D.9	One-day dataset: 16:46:10, 17:26:10, 18:44:09, 19:24:41 . . . .	142
D.10	One-day dataset: 20:03:57, 20:43:36 . . . . .	143
D.11	Dual-channel contrasts one-day dataset: 11:27:04, 12:06:19 . . . .	144
D.12	Dual-channel contrasts one-day dataset: 12:46:02, 13:25:05 . . . .	145
D.13	Dual-channel contrasts one-day dataset: 14:07:52, 14:47:21 . . . .	146
D.14	Dual-channel contrasts one-day dataset: 15:26:36, 16:07:15 . . . .	147
D.15	Dual-channel contrasts one-day dataset: 16:46:10, 17:26:10 . . . .	148
D.16	Dual-channel contrasts one-day dataset: 18:44:09, 19:24:41 . . . .	149
D.17	Dual-channel contrasts one-day dataset: 20:03:57, 20:43:36 . . . .	150
D.18	Difference of dual-channel contrasts: 12:06:19, 12:46:02 . . . .	151
D.19	Difference of dual-channel contrasts: 13:25:05, 14:07:52 . . . .	152
D.20	Difference of dual-channel contrasts: 14:47:21, 15:26:36 . . . .	153
D.21	Difference of dual-channel contrasts: 16:07:15, 16:46:10 . . . .	154
D.22	Difference of dual-channel contrasts: 17:26:10, 18:44:09 . . . .	155
D.23	Contrasts dataset one . . . . .	156
D.24	Histogram dataset one . . . . .	157
D.25	Contrasts dataset two . . . . .	158
D.26	Histogram dataset two . . . . .	159

# List of Tables

3.1	Characteristics of the used image databases . . . . .	17
5.1	Min-Warping number of steps, search range, and resolution . . .	53
5.2	Min-warping scale factors and thresholds . . . . .	53
5.3	Results SIFT homing . . . . .	53
5.4	Results 2D-warping homing . . . . .	54
5.5	Ranking parameter-homing, <i>original</i> database . . . . .	60
5.6	Ranking parameter-homing, <i>hall1</i> database . . . . .	62
5.7	Ranking parameter-homing, <i>uni</i> database . . . . .	64
6.1	Glass filter combination, sensitivity peak and bandwidth . . . .	70
6.2	Single-channel and dual-channel contrasts of natural objects . .	72
6.3	Significances for $J$ , $J_\sigma$ , and $n_f$ for natural objects . . . . .	76
6.4	Single-channel and dual-channel contrasts of artificial objects . .	78
6.5	Significances for $J$ , $J_\sigma$ , and $n_f$ for artificial objects . . . . .	80
A.1	List of components . . . . .	122
B.1	Average vector approach <i>original</i> and <i>uni</i> . . . . .	124
B.2	Maximum approach <i>original</i> and <i>uni</i> . . . . .	124
B.3	Average vector approach <i>hall1</i> and <i>hall2</i> . . . . .	125
B.4	Maximum approach <i>hall1</i> and <i>hall2</i> . . . . .	125



# Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr.-Ing. Ralf Möller for supporting me the years I have spent in his group. The many discussions we had, especially at the end of my PhD, were informative and helpful. I appreciate that Prof. Dr.-Ing. Ralf Möller and Dr. Wolfgang Stürzl agreed to review this thesis. In addition, I am in dept to Prof. Dr. Ellen Baake, Prof. Dr. Martin Egelhaaf, and Dr. Carsten Gnörlich for joining my thesis committee.

Furthermore, I would like to thank my wife Paulina Röben for loving and supporting me all the time, even if I was in a bad temper. She made me free from housework, so I could concentrate on my thesis during the last months.

My parents Gerhard and Ulrike Röben play an decisive role in my life. I have to thank them for their support. They permitted me a free choice of career and believed in me.

I would like to thank Dr.-Ing. Wolfram Schenck for giving me support for the group internal development environment, so called “PROG-Pfad”. He managed compilation errors that would have cost me many hours of debugging. My colleagues Dr.-Ing. Tim Köhler and Lorenz Gerstmayr always had an open ear for me if I had to think out loud. Special thanks go to my office colleague Martin Krzykowski who managed the administrative support on his own, which normally is one of my jobs. He always motivated me when I was down.

Angelika Deister was a big help with organizational problems that appeared during the preparation of my thesis. Moreover, I would like to thank the other PhD “comrades-in-arms” for the time we spent at several coffee breaks. I received technical assistance from Hans Brinkmann in several mechanical works, and from Klaus Kultz in electrical engineering. This work was supported by the Faculty of Technology at Bielefeld University.



# Chapter 1

## Introduction

### 1.1 Motivation

The field of robotics research is extensive. The key aspects of research activity are focused on design, manufacture, and application of robots. Thus, it covers fields like electrical engineering, mechanical engineering, artificial intelligence, control theory, computer science, and many more.

Concerning robotics research, actuators and sensors are very important. Sensors are necessary as information acquisition devices for robotic systems. Actuators allow the robot to react to stimuli that have been measured by sensors. A mobile robot gives a good example for the combination of actuators, sensors, mechanical engineering, and control theory. If a mobile robot is equipped with a distance sensor like an ultrasonic sensor or a laser rangefinder, the actuators of the robot, typically some sort of motors that power the robot, can be controlled by an algorithm to perform obstacle avoidance if the distance measured by the sensor is too short.

Regarding a mobile robot there are various tasks that need to be solved. For example, a mobile robot needs to know his position, which is a typical localization problem. A lot of research effort has been invested into robot localization. Methods like simultaneous localization and mapping (SLAM) give a mobile robot the ability to localize its position absolutely in the environment it is working in. Typically, if SLAM is performed, the robot builds a metrical representation.

Navigation from one position to another in an environment can be dealt of as localization problem, but it is also possible to resolve the navigation problem without knowing the absolute position of the robot. This can be done by visual navigation methods, which mostly use only visual information from the environment, often captured by a panoramic vision system. Based on this information, a topological map is created. The nodes of the map contain the visual information. The reachability of two positions from each other on a direct path is represented by an edge between the nodes. Without knowing the metrical

positions of the nodes, a local visual homing approach can be used to navigate a mobile robot from one node to the other.

Over the last two decades, robotics research has been increasingly taking inspiration from biological models (in a field called “biorobotics”). In their evolutionary development biological systems have been exquisitely adapted to the habitats they are living in and the task they have to perform. For the most part, the evolutionary process bears intelligent, elaborated, convenient, and simple solutions for complex problems. Insects are able to travel long distances from a food location back to their nest. For example, ants use pheromones to mark their way back to the anthill. Flying insects make use of non-contact information like visual cues. During the last century, biologists observed in many empirical studies that bees make use of visual information to guide their way back to the hive. Especially, insects solve many problems which are of importance also for robot navigation, for example illumination invariance, with a very small number of neurons.

The following main questions are addressed in this thesis:

- Is it possible to extend the existing insect-inspired visual homing methods to achieve a better performance?
- How can different spectral frequencies of the sunlight be used to achieve illumination invariance for visual navigation in outdoor applications (especially in combination with ultraviolet light which can be perceived by insects)?
- Is it possible to use the polarization of the natural light for a visual compass in a similar way as insects do, even if the sky is cloudy and only small patches of sky are visible?
- How can local visual homing methods be used to improve the cleaning quality of a cleaning robot?

## 1.2 Outline and Contributions

Chapter 1 is the introductory part of this thesis. It comprises an overall motivation, the outline, and a detailed motivation for the different subtopics: polarized-light compass, local visual homing, outdoor navigation with color contrasts, and the cleaning robot application. The objectives for these are described in chapter 2. Chapter 3 contains the materials and methods that were used in this thesis. Detailed information on the databases used and the image acquisition systems is given. Furthermore, quality measures for local visual homing methods and

statistical methods are presented. Also some of the feature selection methods published by other researchers are recapitulated here.

The range of topics related to the polarized-light compass is presented in chapter 4. The built hardware and the test setup are a major part of this chapter. Chapter 5 deals with local visual homing. In particular, the main topics, SIFT homing and homing with parameter models, are presented, including experiments on image databases. The third main topic, outdoor navigation with color contrasts, is presented in chapter 6. This chapter describes the prototype construction of a five-channel hand-held sensor for measuring spectral frequencies. It also covers the discrimination and statistical analysis of the collected data samples for natural and artificial objects. Furthermore, the chapter provides information on the collection and analysis of two-dimensional image-like samples. An application that makes use of some of the previously described concepts is addressed in chapter 7. The cleaning robot task tries to explain how the theoretical research results could lead to a consumer application. In particular, a strategic and efficient way of cleaning is presented in this chapter. The work that is presented in this chapter was carried out in close collaboration with Lorenz Gerstmayr, Martin Krzykawski and Prof. Ralf Möller.

The overall conclusion is given in chapter 8. A review of the material is presented, which summarizes the presented topics. At this point, an interpretation of the main results is provided. Chapter 9 presents the future work that is necessary in order to further exploit the suggestions made in this thesis. Some unresolved questions that occurred during the analysis of the test results are stated.

## 1.3 Insect Visual System

The design of animal eyes vary from species to species. Land and Nilsson (2004) introduce a summary of the polymorphic vision systems, which covers aquatic eyes, lens eyes on land, apposition compound eyes, and superposition eyes.

The insect visual system has been investigated by many researchers. von Frisch (1914, 1919) concentrated his activities on investigating the honeybee. His pioneering work took more than half a century and was later on extended and verified by many other scientists. His work was focused on the visual system, orientation, and communication of the honeybee (*Apis mellifera*).

The vision system of the honeybee consists of two compound eyes with a large field of view. The angle of aperture of one compound eye is horizontally  $171^\circ$  and vertically  $216.5^\circ$ ; both eyes cover horizontally  $313^\circ$  (Seidl, 1982),

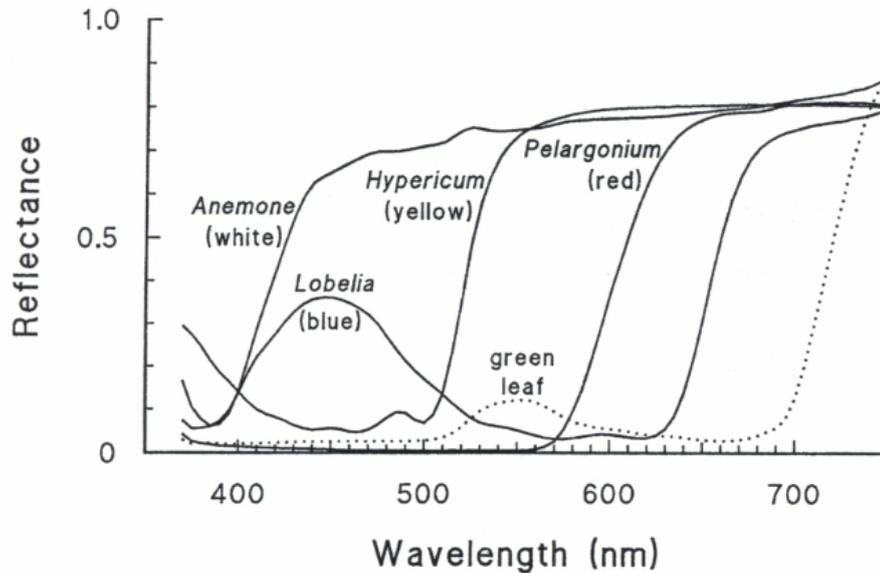


Figure 1.1: Reflected irradiance of *Anemone*, *Lobelia*, *Hypericum*, and *Elargonium* flowers. According to the colors of the flowers, the reflected spectral irradiance is a superposition of different frequencies. The green leaf is plotted for comparison. The figure is taken from Land and Nilsson (2004).

which is close to  $360^\circ$ . For this reason, robotic studies concerning mobile visual navigation use panoramic vision systems to obtain insect-vision-like images.

Briscoe and Chittka (2001) reviewed more than 140 publications on the physiological, molecular, and neural mechanisms of insect color vision. Most insects have receptors for at least three spectral sensitivities. The receptors of insects are sensitive to ultra-violet (UV), blue (B), and green (G). Some insects are also sensitive to the red (R) light. Between different insect species, the peak sensitivities of the different receptor classes vary in wave length. One reason for the differing peak sensitivities might be found in the food preferences of the species, another reason is evolution: “Individual variation is the substrate for evolution. It provides populations with the genetic resources to respond to changing environments and colonize new habitats” (Briscoe and Chittka, 2001). Hence, the evolution tried to adapt the visual systems to the environment the species is living in. The visual system of sawfly (*Symphyta*) and ichneumon wasps (*Ichneumonidae*) is adapted in such a way that blue receptors are absent.

By mixing different color stimuli, Daumer (1956) showed that the Graßmanian laws<sup>1</sup> apply to the honeybee color vision and verified that UV (350nm) and bluish-green (490nm) are complementary color stimuli, and that 490nm are indistinguishable from a mixture of 65% 590nm and 35% 440nm. Thus, Backhaus

<sup>1</sup>“The basic laws for additional colours and colour-matching for human vision trichromatic colour theory” (Malacara-Hernandez, 2002).

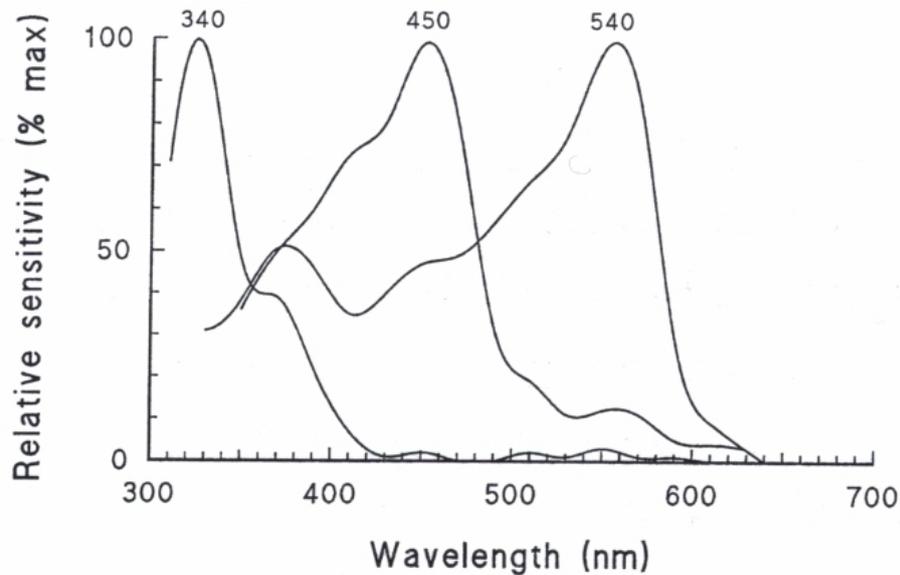


Figure 1.2: Peak sensitivities of the bees' receptors. The peaks are at 340nm (UV), 450nm (B), and 540nm (G). The figure is taken from Land and Nilsson (2004).

and Menzel (1987) state that “bees may establish similarity relations between perceived colors in a manner that is analogous to that of the human color vision system, irrespective of whether bees have subjective impressions of color or not.” Figure 1.1 shows the reflected irradiance of flowers. It can be seen that different flowers reflect different light wave lengths. All flowers reflect light wave lengths greater than 650nm. Figure 1.1 also shows the reflectance of a green leaf for comparison, which also reflects wave lengths greater than 700nm. Thus, green leaves and flowers are only distinguishable below light wave lengths of 700nm which could be a reason that only a few insects are equipped with red receptors. Furthermore, this confirms the suspicion that insects, especially bees, make use of the visual contrast between green leaves and a flower if they try to approach a flower surrounded by green leaves. The peak sensitivities of the bees' receptors are depicted in figure 1.2.

Another important fact on flying insects like bees is that they can detect the polarization of light (von Helversen and Edrich, 1974). The polarization pattern of the sky is formed because the sunlight is scattered at the atmosphere. The pattern of skylight polarization is usually characterized by the angle of polarization (e-vectors). Figure 1.3 visualizes the pattern of polarization experienced by an observer. It presents a two-dimensional representation of the e-vector direction and the degree of polarization for two different sun elevations. Lambrinos et al. (1997, 2000) showed in computer simulations and robotic implementations that it is possible to obtain compass information from three large-field polarization

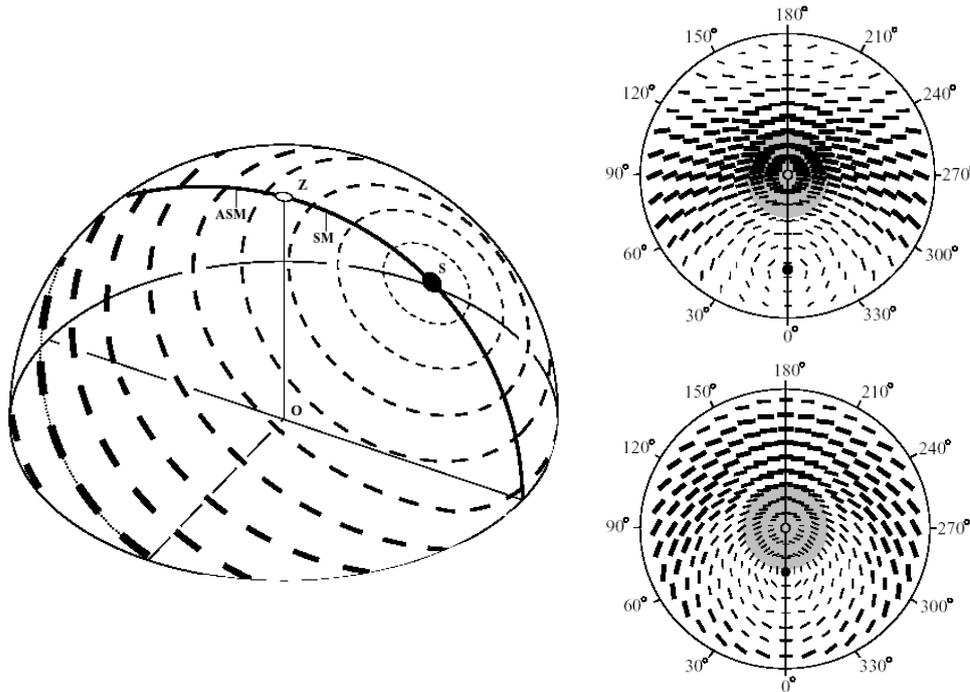


Figure 1.3: The left part of the figure shows a 3-D representation of the pattern of polarization in the sky experienced by an observer in point  $O$ ,  $S$  is the sun,  $Z$  the zenith,  $SM$  the solar meridian, and  $ASM$  the anti-solar meridian. The right part of the figure visualizes two 2-D representations of the e-vector directions and the degree of polarization, which are indicated by the orientation and width of the black bars. **Top:** Sun elevation of  $25^\circ$ . **Bottom:** Sun elevation of  $60^\circ$ . The figure is adapted from Lambrinos et al. (2000).

analyzers. The polarized-light (POL) sensors were all directed to the zenith. Three pairs of POL sensors were used; for each pair, the polarizing axis of one POL sensor was adjusted perpendicular to the other. According to the insect layout “three pairs of POL sensors were adjusted, such that the polarizing axes of the positive channel were 0, 60, and 120 degrees with respect to the robots transverse axis” (Lambrinos et al., 1997).

Pomozi et al. (2001) concentrated on the proportion  $P$  of the celestial polarization pattern. “The parameter  $P$  gives the proportion of the skylight pattern that can be used by the insect to reliable e-vector orientation” (Pomozi et al., 2001). It was pointed out that the e-vector can even be computed under cloudy sky: “clouds decrease the extent of skylight polarization useful for animal orientation much less than hitherto assumed” (Pomozi et al., 2001). Towne and Moscrip (2008) motivate that “bees know the sun’s pattern of movement in relation to an inclusive representation of the landscape, not only familiar flight routes.” It is also stated that “bees learn the sun’s daily pattern in the relation to the entire landscape panorama around their nests.” Hence, bees are able to

remember the stored daily pattern as an effective backup system if the pattern is covered on cloudy days. It was moreover shown that the solar ephemeris is connected to the skyline panorama.

## 1.4 Visual Homing

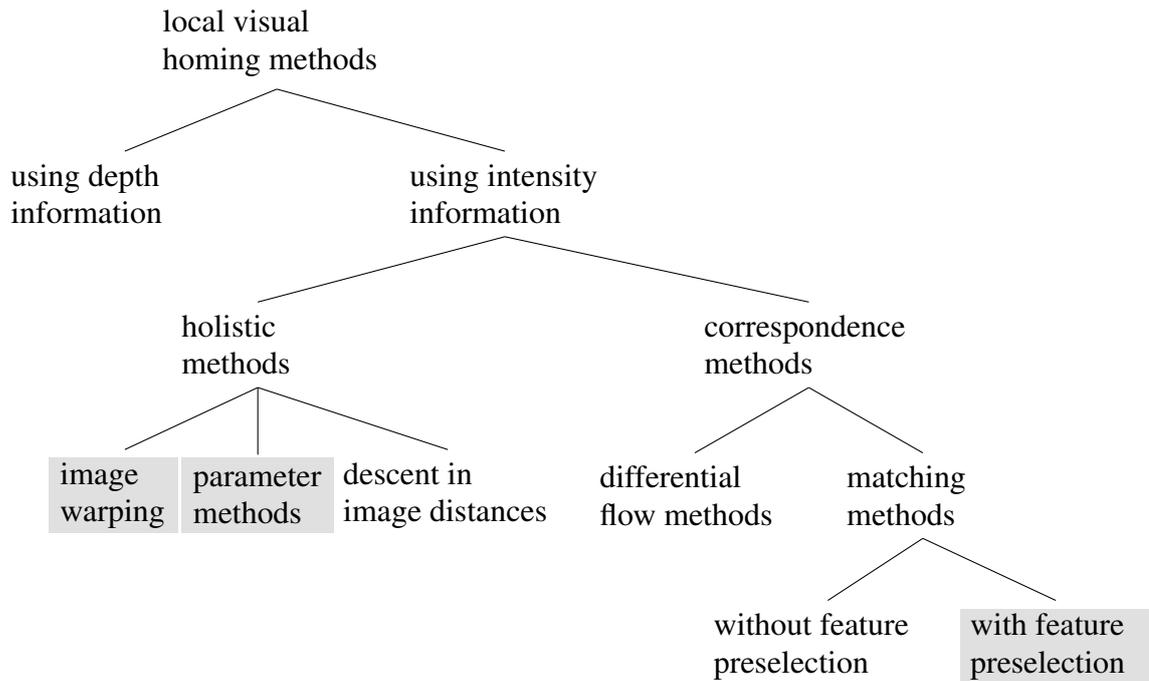


Figure 1.4: Classification of homing methods, visualized as tree (adapted from Möller and Vardy (2006)). The gray boxes mark the branches of homing methods which are dealt with in this thesis.

Visual homing is the ability to use visual information to move from the current position, where the current view is captured, to a second position, in the following referred to as snapshot position, where another image, the snapshot, was taken. Local visual homing methods have been well categorized by Möller and Vardy (2006). Figure 1.4 shows a treelike representation of the homing methods. The first level of the tree distinguishes between methods that either use depth information or intensity information. Methods that belong to the class that use depth information are disregarded here; this thesis concentrates on the class of methods that use intensity information.

Intensity information based methods can be divided into holistic and correspondence methods on the second level of the homing tree. Image warping, parameter methods, and descent in image distances belong to the group of holistic methods which treat the image as a whole and do not solve the correspondence problem. Franz et al. (1998) suggested the warping method which was extended

by Möller (2009). Franz et al. (1998) use one-dimensional panoramic images, which are extracted from the horizon of two-dimensional images. The current view image is warped for several virtual movements of the agent. The virtual movement parameters for the warped image which has the smallest difference to the snapshot are used to compute the corresponding home vector. This method has an intrinsic compass; hence, the images need not be rotationally aligned. Möller (2009) extended the method to two-dimensional input images. In order to take vertical pixel displacement into account, which occurs when landmarks are not on the horizon of the image, scaled copies of the input images are used.

Parameter methods are addressed by Lambrinos et al. (2000), Möller (2000, 2001), and Menegatti et al. (2004). Parameter methods make use of a condensed description of the image data which is referred to as “parameter vector”. The parameter vectors are used for navigation purpose. The average landmark vector model (Lambrinos et al., 2000; Möller, 2000) computes unit vectors from the current position to each of the visible landmarks. The mean of all landmark vectors gives the average landmark vector (ALV). The difference of the snapshot ALV and the current view ALV is the home vector if the input images are compass-aligned. The contour model (Möller, 2001) uses the length and the eccentricity of the contour of the landmark panorama. A distance function is used to build a potential field where the negative gradient is used as the home vector. Menegatti et al. (2004) compute line-by-line Fourier coefficients from the input images. Only the amplitudes of the first  $k$  Fourier coefficients are stored for navigation. Again, a distance function is used to build a potential field, where the negative gradient gives the home vector. Since the Fourier coefficients are rotationally invariant, no compass is needed. Stürzl and Mallot (2006) also use Fourier coefficients as a parameter signature and close the gap between parameter models and warping. The authors state that the presented method is a warping method which utilizes Fourier coefficients instead of image intensities.

The third class of holistic methods deals with methods that use descent in image distances. Zeil et al. (2003) showed that a simple measure like the root mean square difference of two images can be used to compute a potential field with a minimum at the snapshot position. The difference becomes the smaller, the closer current view and snapshot get. The main disadvantage of methods based on pixel intensities is that they are prone to illumination changes. Moreover, the method needs aligned images. Möller and Vardy (2006) and Möller et al. (2007) extended the decent in image distances (DID) method to get rid of test steps, which used to be necessary to compute the gradient of the potential field.

The second part of the second level of figure 1.4 comprises of the correspondence methods. Correspondence methods establish correspondences between local regions of the input images. A pair of correspondences describes a vector

representing the movement of a features that arises from the displacement of the capture locations of the input images. From the movement vectors an overall movement vector is computed, which is used for navigation purpose. As shown in figure 1.4, two subclasses of correspondence methods exist: matching methods and differential flow methods. Flow methods are based on a first-order (intensity) or second-order (intensity gradient) Taylor approximation of the correspondence functions. Provided that only small displacements in the image appear, the Taylor approximation can be applied. Even if this assumption is violated, in many cases a sufficient number of correct flow vectors is computed, and an adequate home vector is obtained (Vardy and Möller, 2005).

The snapshot model, which is the oldest model, belongs to the matching methods with feature preselection and was published by Cartwright and Collett (1983, 1987). It is based on the theory that a biological agent tries to minimize the difference between landmark bearing and size in current view and snapshot. In empirical studies, bees were trained to locate a food location in-between different cylindric landmarks. If the arrangement of cylinders was changed, the bees searched at the wrong position, which implicates that the bees learned the alignment of the cylinders. The horizontal input images are separated into dark and light sectors. The images are compass-aligned and every sector from the snapshot is associated with the closest sector from the current view. The correspondence pairs are represented by two unit vectors. One describes the angular difference, the other the difference in apparent size. Accumulating all unit vectors yields the home vector. Establishing correspondences between snapshot and current view is widely used for visual homing. Some approaches try to extract distinctive features (Hong et al., 1991; Se et al., 2001, 2002; Brown and Lowe, 2002; Fiala and Basu, 2004; Hayet et al., 2007; Pons et al., 2007), others use less unique features (Cartwright and Collett, 1983, 1987; Weber et al., 1999; Lambrinos et al., 2000). Vardy and Oppacher (2003, 2004) concentrate on features that have been extracted by the Harris corner detector. Furthermore, colored regions have also been employed for navigation by Gourichon et al. (2003) and Goedemé et al. (2004).

In addition to differential methods, several homing methods are presented by Vardy and Möller (2005): blockmatching, homing by intensity and gradient matching. Homing by intensity and gradient matching are variants of the blockmatching method. All presented methods belong to the class of correspondence methods without feature preselection. The blockmatching method does not extract single landmarks; instead, correspondences are computed among similar regions in the images. Searching the whole images for corresponding blocks of similar gray values leads to correspondence vectors. If the agent moves solely on a planar area the correspondence vectors can be accumulated to directly compute an overall home vector. A couple of the correspondences might be wrong,

but the majority of them will be correct in most cases. The “democracy effect” ensures that the correct correspondences exert a dominating influence.

## 1.5 Outdoor Navigation with Color Contrasts

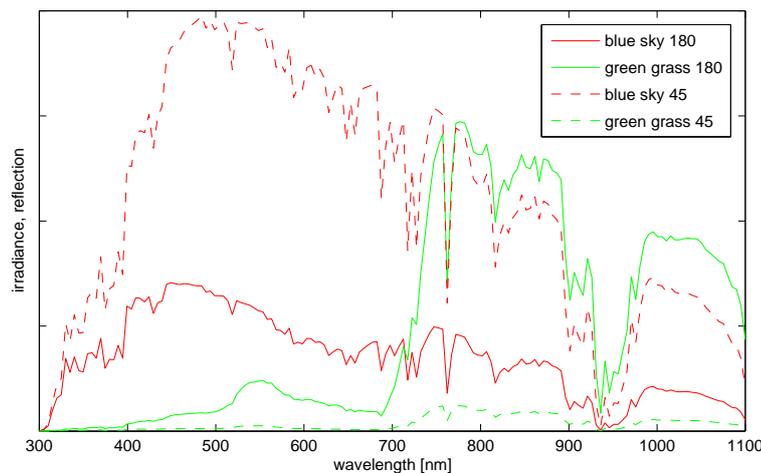


Figure 1.5: Irradiance of blue sky and of light reflected from vegetation (green grass) in energy units, obtained from a numerical method. Solid curves, azimuth of  $180^\circ$  between viewing direction and sun; dotted curves  $45^\circ$ , adapted from Kollmeier et al. (2007).

As introduced in section 1.4, a lot of biological research has been invested into local visual homing of animals. Fukushi (2001) published experimental results showing that wood ants use the skyline of a panoramic image as a navigational guidemark. The habitat surrounding the nest was nearly featureless only the crowns of some distant trees and the troughs between them was visible from the hive. The ants were able to find their nest.

Furthermore, Möller (2002b) conjectured that insects use UV-green contrast for landmark navigation. Figure 1.5 shows the irradiance of blue sky and of light reflected from vegetation. It can be seen that skylight has a higher proportion of UV radiation than the reflected light from vegetation. Figure 1.6 shows that it is possible to distinguish foreground and sky objects with a linear separation process. For this purpose, samples were taken from objects and sky with a sensor that measures UV and green irradiance (Möller, 2002b).

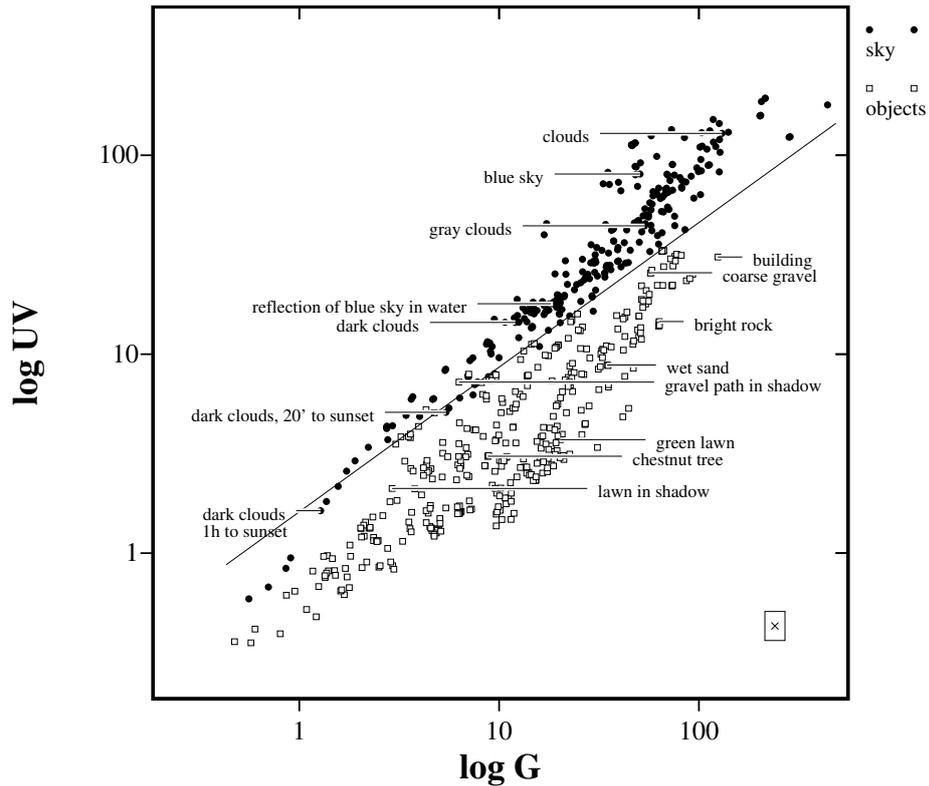


Figure 1.6: Samples ( $\log G$ ,  $\log UV$ ) taken from different objects and sky. Unless otherwise noted, the labeled object samples were taken under direct sunlight on a clear day. The threshold line (slope = 0.73) was computed such that the error of misclassified points is as small as possible (22 of 628). The maximal sensor noise is depicted by the rectangle in the lower right corner; the maxima over all samples of the standard deviations in the UV and green channel in each sample sequence is visualized. The figure is adapted from Möller (2002b).

## 1.6 Cleaning Robot Control

The main task of autonomous cleaning robots is to completely clean the whole area the robot can access. The quality of cleaning can be defined by the proportion of repeated coverage to the accessible area and time. How to achieve a small proportion depends on the size of the cleaning area.

For large buildings, large and heavy-weight machines exist. Typically, these machines are equipped with strong batteries, large computational power, sophisticated sensor systems and make use of SLAM (Prassler et al., 2000; Prassler and Kosuge, 2008).

In order to clean an apartment the domestic robot needs to be small and agile. Unlike industrial cleaning robots, domestic robots must clean areas below furniture as well as open areas. To advertise a domestic robot to the consumer the robot needs to be convenient. A low priced robot can only be equipped

with a relatively small battery, low computational power, and a small number of cheap sensors. Thus, the length of a cleaning run of a domestic robot is limited due to the battery power which implies that the domestic robot needs a cleaning strategy that decreases the repeatedly covered areas. Today's most purchasable cleaning robots use a randomized cleaning strategy (Prassler and Kosuge, 2008). The advantage of such an strategy is that it does not need an internal representation of the area, thus the computational power and the effort for memorizing information of the environment is low.

Sensors like panoramic image sensors have not been used for cleaning robots, mostly, infrared sensors are used. Using sensors like panoramic image sensors enables the robot to build a detailed representation of the environment which can be used to implement a specific cleaning strategy that yields a good cleaning result.

# Chapter 2

## Objectives

### 2.1 Polarized-Light Compass

Lambrinos et al. (1997, 2000) describe a physical model of the polarized-light compass and test the model on a mobile robot. The compasses that were built by Lambrinos et al. (1997, 2000) consists of six photodiodes in combination of a linear polarizer and a blue transmitting filter.

The first aim of this work is to build a polarized-light compass similar to the one by Lambrinos et al. (1997, 2000). This sensor has to be tested with respect to how well it measures the direction and the strength of polarized light and how accurate the derived compass information is. For this purpose, an experimental setup with the polarized-light compass attached to an electromagnetic compass is used. After verifying that the polarized-light compass works as accurate as the original compass (Lambrinos et al., 1997, 2000), this thesis complements the original work by testing whether it is possible to determine the azimuth orientation of the sun from the polarization pattern of the sky even if only small patches of blue sky are visible. For this purpose, a pan-tilt unit is used to point the compass towards the sky in different directions. The pan-tilt unit can be used to measure the angular displacement of the compass, since it is very accurate ( $0.01^\circ$  position resolution). In this context it is also analyzed how clouds affect the measurements.

### 2.2 Local Visual Homing Methods

Existing local visual homing methods were outlined in section 1.4. Even though a lot of work has been invested, there are still unresolved problems and room for improvements of the visual local homing methods; for example computational speed and homing accuracy. The visual local homing method needs to be computational fast and precise. Computational speed is not crucial if the homing method is used in simulation, but in real-time robot experiments it is

essential. The computational time is limited due to the motion speed of the robot. The catchment area of a visual local homing method needs to be as large as possible, alternatively, long distance homing can only be realized by map building. Furthermore, the local visual homing method should be able to cope with illumination changes.

One question that is considered in this thesis is if it is possible to use parameter models for navigation. The big advantage of parameter models is that the amount of memory is reduced and that some of the used parameters are not effected by illumination changes. The average landmark vector model shows that it is possible, but are there any other parameter models that will work as well or better? ALV only takes as input a panoramic image that has one pixel height and which is extracted at the horizon of the original camera image. The input image is parametrized and gives the ALV's. Thus, scene changes that appear at the horizon of the image have a large influence on the parameter set. It would be good to compute parameters that are more robust and easy to compute. It might be possible that statistical image parameters are also well suited for representing locations by image information, by using rotation-invariant parameters. Moreover, it would be interesting to see whether it is possible to eliminate the compass which is required by many of the local visual homing methods (for example flow-based methods, but not warping).

Furthermore, the goal is to design a local homing method that relies on distinct features which are unaffected by illumination changes and also invariant when seen from different angles and distances. For this purpose, the "Scale Invariant Feature Transformation" (SIFT) by David Lowe is used. The quality of the designed method can be compared to other published methods like 2D-warping.

## 2.3 Outdoor Navigation with Color Contrasts

Visual navigation in outdoor environments is more difficult to achieve than in indoor environments. In outdoor applications, the intensity of natural light can vary over several decades. As pointed out by Fukushi (2001) and Möller (2001), it might be possible to use the skyline of a natural scene for navigation. The skyline separates foreground objects from sky (background). The main question is: How can such a skyline be computed? Color contrasts were analyzed by Möller (2002b). The author focuses on green and UV. Since this publication only deals with natural samples and two color frequencies, which are biologically motivated (bees have UV, blue, and green receptors), it would be interesting to test whether the UV-green contrast actually leads to the best separation between sky and terrestrial objects, or whether other frequency combinations fare better. So,

a sensor that is able to measure UV, blue, green, red, and infrared frequencies is constructed. By measuring more than two frequencies it is possible to compute a multitude of contrasts and check which performs best. Furthermore, it needs to be considered, whether artificial objects are also separable by these contrasts. Due to structural alteration, a typical outdoor scene might partly consist of natural (e.g. trees) and artificial objects (e.g. houses).

If case natural and artificial terrestrial objects behave in the same manner and are separable from sky, the contrast sensor has a high potential to be used in a wide variety of applications. By extending the analysis to two-dimensional images, a skyline can be extracted. This skyline is used as input for local visual navigation methods.

## 2.4 Cleaning Robot Control

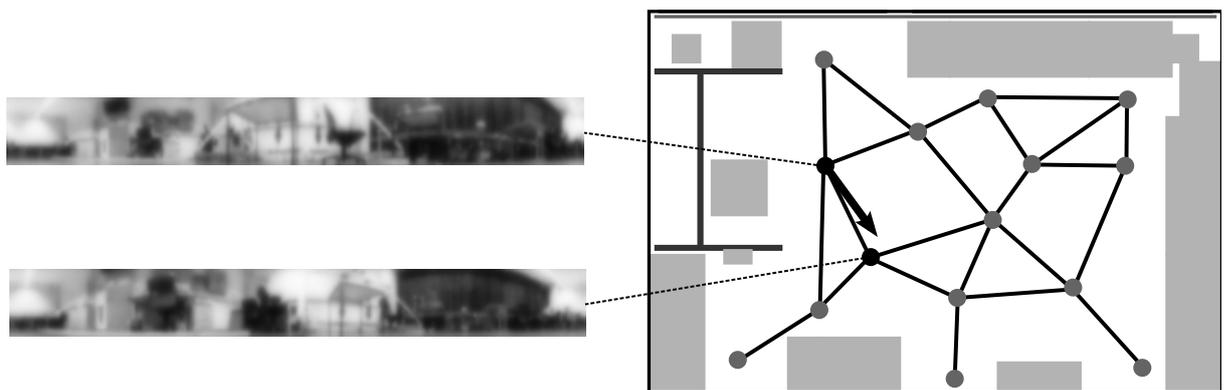


Figure 2.1: Example of a topological map that is build during a run of a cleaning robot. The nodes depict positions where panoramic images have been stored. An edge which connects two nodes represents the ability to move from one position to the other. Local visual navigation methods are used to move from one node to another.

A randomized cleaning strategy does not need memorized information of the environment. To enhance the cleaning quality, the complete accessible area needs to be cleaned and the proportion of the repeated overlap of the accessible area needs to be minimized. Random strategies tend to have a large overlap. Meandering cleaning trajectories, which produce less overlap, are more efficient than random walks. Thus, it is recommended to use a structured cleaning strategy concerning the environment.

An exemplary cleaning run with no observable cleaning strategy is shown in figure 2.1. To memorize the visited locations, the system is extended with a topological map. During a cleaning run, the robot captures images and stores

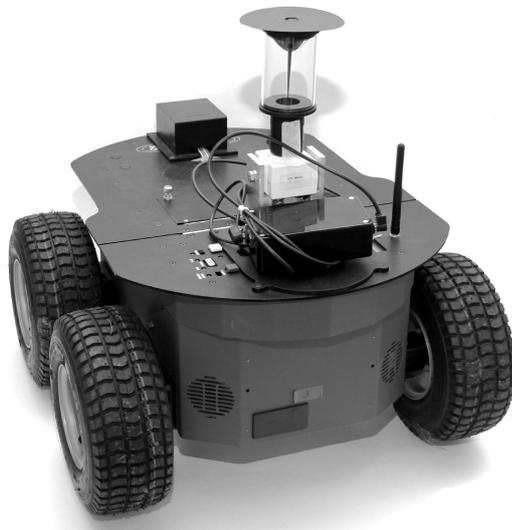
the data in the topological map. The nodes of the topological map represent the visited positions. An edge between two nodes specifies the reachability from one node to the other. Navigation between successive nodes is done by local visual homing methods.

Given a cleaning robot with a topological representation of the environment, a meandering cleaning strategy can be implemented. For this purpose, the meanders are adjusted with respect to the home vectors computed by the visual homing methods. The straighter the trajectories are, the better is the cleaning result. For this purpose, a trajectory controller was developed which keeps the inter-lane distance close to a predefined value. The inter-lane distance normally should be the width of the vacuum cleaning unit of the robot.

# Chapter 3

## Materials and Methods

### 3.1 Image Databases



ActivMedia Pioneer 2-DX



ActivMedia Pioneer 3-DX

Figure 3.1: The Pioneer robots used for database collection. On top of each robot, a camera system with a hyperbolic mirror is mounted. The geometry of the mirrors is different (2-DX: normal-view mirror; 3-DX: wide-view mirror).

The image databases are collected for offline simulation purposes. By using the same databases for several navigation methods, a reliable comparison of the homing quality can be achieved. The database **uni** was captured by using the four-wheeled outdoor Pioneer 2-DX; all the other image databases were captured with a two-wheeled indoor Pioneer 3-DX (figure 3.1). Mounted on top of the pioneer 2-DX, a camera (The Imaging Source DFK 4304) pointed towards a hyperbolic mirror (normal-view mirror from Accowle Ltd.) to capture panoramic images. The Pioneer 3-DX is equipped with the same camera but with an ultra-wide-view mirror from Accowle Ltd. For the ultra-wide-view

database	grid size	area	spacing	#images	used mirror
<b>original</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>arboreal</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>chairs</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>day</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>doorlit</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>original screen</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>twilight</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>winlit</b>	10×17	2.7m×4.8m	0.3m	170	ultra-wide
<b>hall1</b>	10×20	4.5m×9.5m	0.5m	170	ultra-wide
<b>moeller1</b>	11×22	1.0m×2.1m	0.1m	242	ultra-wide
<b>roeben1</b>	11×36	1.0m×3.5m	0.1m	396	ultra-wide
<b>uni</b>	10×17	9m×16m	1.0m	170	normal-view

Table 3.1: Characteristics of the used image databases.

mirror, the upper-limit of the visual field above the horizontal plane is about 26~46 degrees (taken from Accowle product specification), for the normal-view mirror about 10~26 degrees. Table 3.1 summarizes the characteristics of the databases relevant for this thesis.

## original

The **original** database was collected by Andrew Vardy (Vardy, 2005; Vardy and Möller, 2005) in the laboratory of the computer engineering group at Bielefeld University. The database covers an area of 2.7m×4.8m. The images were collected on a rectangular grid with a spacing of 30cm between each node of the grid. The size of the grid is 10×17; thus, the database holds 170 images. No additional items were placed in the lab during the image acquisition, the door and the curtains of the room were closed. Only the lighting mounted at the ceiling of the room illuminated the laboratory. Metrical information about the image acquisition positions were obtained from robot odometry.

The setup of the laboratory was rearranged after the collection of this database. Therefore, there are variations of the **original** database which can be used for cross-database tests. The cross-database test is performed by taking the snapshot from one database and the current view from a different database. The cross-database test gives an indication how well a homing method copes with environment or illumination changes. The name of the modified databases give a hint what has been added to the laboratory.

- **arboreal**  
An arboreal plant was placed in the center of the laboratory.
- **chairs**  
Chairs have been added to the lab.
- **day**  
The curtains of the room were open; thus, the natural illumination changes over day affect the database images.
- **doorlit**  
The illumination of the lab consists of two light-bands at the ceiling mounted parallel to the windows. The light-band next to the door was on and the light-band next to the window was off.
- **original screen**  
A white projection screen was added to the lab.
- **twilight**  
The curtains were open and the data was collected in the twilight.
- **winlit**  
The light-band next to the window was on and the light-band next to the door was off.

## **hall1**

The **hall1** database was also collected by Andrew Vardy. The location is the main hall of the Bielefeld University. The database was collected in the deserted main hall over night and covers an area of  $4.5\text{m} \times 9.5\text{m}$ , with a spacing of 50cm. The database consists of 170 images. The environment is relatively achromatic and the illumination is sparse because there are only few fluorescent lamps.

## **moeller1**

The **moeller1** database was collected in a living room and covers an area of  $1.0\text{m} \times 2.1\text{m}$ . The grid size is  $11 \times 22$ . The spacing is 10cm. During the image collection process the robot was supervised and the position was corrected semi-automatically (Kreft, 2007).

## roeben1

The **roeben1** database was captured in a living room combined with a kitchen. It covers a grid size,  $11 \times 36$ , again with a 10cm spacing. The covered area is  $1.0\text{m} \times 3.5\text{m}$ . The images were taken close to the furniture such that the images at the beginning of a lane differ considerably from those at the end (Kreft, 2007).

## uni

The **uni** database covers an area of  $9\text{m} \times 16\text{m}$  and is located outside the north entrance of the university. During the collection time, the area was in the shadow. The spacing of the grid is 1m, and the grid consists of 170 images (Schatz, 2006).

## 3.2 Quality Measures for Navigation

To compare different visual homing algorithms, the same quality measure is determined for competing algorithms. As measures, average angular error, median angular error, and return ratio are used in this theses.

### 3.2.1 Average Angular Error

The average angular error is one measure for the performance of a homing algorithm. It was used for different robot homing studies (Weber et al., 1999; Hafner, 2001; Vardy and Möller, 2005). By testing the homing algorithm with known positions of the snapshots and the current views, it is possible to compute the true home vectors. The average angular error is the absolute angular difference between the observed home vectors and the true home vectors.

Given the position  $\mathbf{c}$  of the current view and the snapshot position  $\mathbf{s}$ , the ideal home vector  $\mathbf{v}_{\text{Ideal}}(\mathbf{s}, \mathbf{c})$  for  $\mathbf{s} \neq \mathbf{c}$  is defined by

$$\mathbf{v}_{\text{Ideal}}(\mathbf{s}, \mathbf{c}) = \frac{\mathbf{s} - \mathbf{c}}{\|\mathbf{s} - \mathbf{c}\|} \quad (3.1)$$

The angular error (AE) between an ideal home vector  $\mathbf{v}_{\text{Ideal}}$  and a home vector  $\mathbf{v}_{\text{Method}}$  computed by a homing method between the positions  $\mathbf{c}$  and  $\mathbf{s}$  is defined as

$$AE(\mathbf{s}, \mathbf{c}) = \arccos \left( \mathbf{v}_{\text{Ideal}}^T(\mathbf{s}, \mathbf{c}) \cdot \mathbf{v}_{\text{Method}}(\mathbf{s}, \mathbf{c}) \right). \quad (3.2)$$

Concerning a single snapshot position  $\mathbf{s}$ , the average angular error (AAE) can be computed by finding the average value from equation 3.2 over all current view positions. For statistical analysis of the homing performance, a large set

of data is used. For this purpose, an image collection with a grid size of width  $m$  and height  $n$  is used. The position coordinates are the same as the indices of the  $m \times n$  capture grid. For snapshot position  $\mathbf{s}$  the AAE is defined as

$$AAE_s = \frac{1}{mn - 1} \sum_{c_x=0}^{m-1} \sum_{c_y=0}^{n-1} AE(\mathbf{s}, \mathbf{c}), \quad (3.3)$$

where  $\mathbf{s} = (s_x, s_y)^T$ ,  $\mathbf{c} = (c_x, c_y)^T$ , and  $c_x, c_y$  are the grid coordinates of the image. The AE at the snapshot position is excluded.

Averaging the  $AAE_s$  over all snapshot positions gives a overall measure for a homing method. This measure is used to compare two competing homing methods

$$AAE_* = \frac{1}{mn} \sum_{s_x=0}^{m-1} \sum_{s_y=0}^{n-1} AAE_s, \quad (3.4)$$

where  $\mathbf{s} = (s_x, s_y)^T$ .

### 3.2.2 Median Angular Error

The median angular error (MAE) was used by Möller (2009). The advantage of the MAE in comparison to the AAE is that home vectors with a large deviation in the AE have a smaller influence.

The median angular error of an ordered observation  $AE_1, \dots, AE_{mn}$  with  $AE_i \leq AE_{i+1} \forall i \in \{1, \dots, mn - 1\}$  is given by

$$MAE_s = \begin{cases} AE_{\frac{mn+1}{2}} & \text{for } mn \text{ odd} \\ \frac{1}{2}(AE_{\frac{mn}{2}} + AE_{\frac{mn}{2}+1}) & \text{otherwise} \end{cases}, \quad (3.5)$$

$m$  and  $n$  denote the height and width of the image collection grid as explained in section 3.2.1. Again, an overall measure can be computed by building the median over all  $MAE_s$

$$MAE_* = \begin{cases} MAE_{\frac{mn+1}{2}} & \text{for } mn \text{ odd} \\ \frac{1}{2}(MAE_{\frac{mn}{2}} + MAE_{\frac{mn}{2}+1}) & \text{otherwise} \end{cases}. \quad (3.6)$$

### 3.2.3 Return Ratio

The return ratio (RR) is a measure for successful homing of a virtual agent. In the biological literature, the catchment area describes the homing success. The catchment area is “the area surrounding the goal within which the agent will be attracted to the spot where the snapshot was taken” (Cartwright and Collett,

1987). Vardy and Möller (2005) remark that the “average angular error is an indirect measure of the catchment area”. But as the AAE is not independent of the position where it is computed, “the size of the catchment area will depend very much on where errors occur”.

One way to analyze the return ratio is to perform a homing trial with a virtual agent that moves according to the home vectors in steps of length  $l$ . The step length  $l$  is typically set to 0.5 grid units of the database (Vardy and Möller, 2005). To check whether the trial of the agent is finished, the condition  $\|s - c\| < 1$  is tested. The trial is also finished if the predefined maximum traveling length is exceeded. The maximum traveling length is set to half of the circumference of the image collection capture grid. If the agent reaches the goal, the homing trial is considered successful and marked with 1, otherwise it is marked with 0. The traveling length of the homing trial is defined by the number of steps the robot needs to reach the goal. If the maximum is reached, the trial is aborted and marked as unsuccessful. Accumulating the number of successful homing trials divided by the total number of homing attempts, one obtains the return ratio. As used by Vardy and Möller (2005), the return ratio for snapshot position  $(x, y)$  is referred to as  $RR_{(x,y)}$  and the quantity is in the range  $[0, 1]$ , where 1 indicates perfect homing.

### 3.3 Statistical Methods

This section gives a brief introduction to the statistical methods used in this thesis.

#### 3.3.1 Jackknife

The jackknife is a historically earlier alternative to the bootstrap for calculating standard errors that is less computation intensive. In Efron (1979, p. 1) it was shown “that the jackknife can be thought of as a linear expansion method for approximating the bootstrap”.

In the initial phase, the statistics are computed from the full sample for all  $n$  observations ( $\theta^*$ ). Later on, the sample size is stepwise reduced by removing one data point ( $\theta_{-i}^*$ ). For each step the statistic is calculated. Pseudo-values can be computed for each observation in the original sample as:

$$\tilde{\theta}_i = n\theta^* - (n - 1)\theta_{-i}^* \quad (3.7)$$

where  $\theta_{-i}^*$  is the statistic calculated from the the sample with observation  $i$  omitted (Quinn and Keough, 2002).

Two estimates accumulate one pseudo-value, one based on the whole sample  $\theta^*$ , the other based on the removal of a particular observation  $\theta_{-i}^*$ . The mean of the pseudo-values is the jackknife estimate of the parameter. The standard deviation of the jackknife estimate is:

$$\sqrt{\frac{n-1}{n} \sum_{i=1}^n (\theta^* - \tilde{\theta}_i)^2} \quad (3.8)$$

(Quinn and Keough, 2002). These calculations are based on the assumption that the pseudo-values are independent of each other, which is not true in most real cases. If the original sample size is small, the jackknife is usually not used for confidence intervals.

### 3.3.2 Bootstrap

The base idea of bootstrap (Efron, 1979, 1987) is that, without knowing the theoretical distribution, the sampling distribution of the statistic is determined empirically by randomly resampling, with replacement, from the original sample, usually with the same original sample size. As the sampling is with resampling, the same observation can obviously be resampled, such that the bootstrap examples will be different from each other. The desired statistics can be determined from each bootstrapped sample and the sampling distribution can be derived from each statistic determined. The mean of the statistics from the bootstrapped samples is the bootstrap estimate of the parameter. The standard deviation of the bootstrap estimate is the standard deviation of the statistics from the bootstrapped samples.

To get a good estimate of the distribution, the bootstrap sample size should not be less than one thousand samples. Increasing the sample size yields a more precise estimate of the distribution.

#### Bootstrap for Navigation Tests

Homing simulation on image databases is a time-consuming task, therefore the quantity of computed  $AAE_*$  and  $RR_*$  values is small. Vardy and Möller (2005) made use of 36 sample positions. The empirical distribution of such a small quantity can be covered by statistical tools like bootstrap.

Let's assume that two known navigation methods should be compared against each other. The estimates  $\hat{\theta}_{AAE}$  and  $\hat{\theta}_{RR}$  are computed by

$$\hat{\theta}_{AAE} = AAE_*^{\text{Method1}} - AAE_*^{\text{Method2}} \quad (3.9)$$

$$\hat{\theta}_{RR} = RR_*^{\text{Method1}} - RR_*^{\text{Method2}} \quad (3.10)$$

where  $AAE_*^{\text{Method}i}$  and  $RR_*^{\text{Method}i}$  are the averaged values over the samples.  $\hat{\theta}_{AAE}$  and  $\hat{\theta}_{RR}$  are estimates of the real differences  $\theta_{AAE}$  and  $\theta_{RR}$

The sampling distribution is built from the  $AAE$  and  $RR$  measures. For every sample, new values for  $\hat{\theta}_{AAE}(i)$  and  $\hat{\theta}_{RR}(i)$  are recomputed, where  $i = 1 \dots B$ . The size of the sampling distribution is  $B = 2000$  samples.

### Confidence Intervals

For checking if two homing methods differ significantly, it is essential to look at the confidence intervals of the bootstrap distribution. Efron and Tibshirani (1994) present several different ways to calculate the confidence intervals. The  $BC_\alpha$  intervals proved to be best, since they correct biases of the statistical estimates.

### 3.3.3 Fisher's Linear Discriminant Analysis

Fisher's linear discriminant analysis (LDA) describes a metric for the quality of separability of two observation classes. Fisher's LDA tries to determine a vector  $\mathbf{w}$  such that, when the data points of both classes are projected onto this vector, their projected means are as far apart as possible, and the projected data in both classes are scattered over the smallest possible region (Alpaydin, 2004). Thus, the ratio of the variance between the classes to the variance within the classes characterizes the separation defined by Fisher.

Let  $\mathbf{x}_i$  denote a data point with  $r_i \in \{0, 1\}$  being its classification ( $r_i = 1$  for class 1;  $r_i = 0$  for class 2) and  $\mathbf{m}_1$ ,  $\mathbf{m}_2$  be the means of the two classes. Then the projected means are  $m_1 = \mathbf{w}^T \mathbf{m}_1$  and  $m_2 = \mathbf{w}^T \mathbf{m}_2$ . The variances of the two projected classes are defined as

$$s_1^2 = \frac{1}{N_1} \sum_i (\mathbf{w}^T \mathbf{x}_i - m_1)^2 r_i \quad (3.11)$$

$$s_2^2 = \frac{1}{N_2} \sum_i (\mathbf{w}^T \mathbf{x}_i - m_2)^2 (1 - r_i) \quad (3.12)$$

where  $N_1$  and  $N_2$  are the numbers of data points in the classes 1 and 2. The linear discriminant is the vector  $\mathbf{w}$  that maximizes the criterion

$$J(\mathbf{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \quad (3.13)$$

By introducing the covariance matrices

$$\mathbf{S}_1 = \frac{1}{N_1} \sum_i (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T r_i \quad (3.14)$$

$$\mathbf{S}_2 = \frac{1}{N_2} \sum_i (\mathbf{x}_i - \mathbf{m}_2)(\mathbf{x}_i - \mathbf{m}_2)^T (1 - r_i) \quad (3.15)$$

and their sum  $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$ , which is called the within-scatter of the classes,  $J(\mathbf{w})$  can be rewritten as

$$J(\mathbf{w}) = \frac{[\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)]^2}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}. \quad (3.16)$$

This yields a closed-form solution for the maximum (Alpaydin, 2004):

$$\mathbf{w} = c \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2), \quad (3.17)$$

where  $c$  is an arbitrary constant.

### Adding Noise to LDA

The LDA criterion  $J$  (3.16) introduced in section 3.3.3 is modified (Kollmeier, 2006). Gaussian receptor noise with a standard deviation  $\sigma$  is added. The modified criterion is labeled  $J_\sigma$ . The modified covariance matrix for class  $i$  is defined as follows

$$\mathbf{S}'_i = \mathbf{S} + \sigma^2 \mathbf{I}, \quad (3.18)$$

where  $\mathbf{I}$  denotes the unit matrix.

The modified within-class covariance matrix is given by

$$\mathbf{S}'_W = \mathbf{S}'_1 + \mathbf{S}'_2 \quad (3.19)$$

$$= \mathbf{S}_W + 2\sigma^2 \mathbf{I} \quad (3.20)$$

Then, the criterion is given by

$$J_\sigma(\mathbf{w}) = \frac{[\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)]^2}{\mathbf{w}^T \mathbf{S}'_W \mathbf{w}} \quad (3.21)$$

$$= \frac{[\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)]^2}{\mathbf{w}^T (\mathbf{S}_W + 2\sigma^2 \mathbf{I}) \mathbf{w}} \quad (3.22)$$

and the optimal discriminant vector is obtained by

$$\mathbf{w} = c \mathbf{S}'_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \quad (3.23)$$

## 3.4 Scale Invariant Feature Transform (SIFT)

This section explains the method for feature selection. The “Scale Invariant Feature Transformation” (SIFT) by David Lowe is introduced. In this thesis, SIFT features are used for visual homing with feature preselection. The original idea of SIFT with the extraction of robust features for object recognition in mind was presented by Lowe (1999). Robustness of the features is given by the invariance concerning translation, scale and rotation. The scale is given by the relative alteration of the size of an object according to two images of the same scene with different resolutions (Mikolajczyk, 2002); translation and rotation describe object displacements.

The extraction of the keypoints and their associated features is done in four steps (Lowe, 2004):

- **Detection of scale-space extrema:** The detection of scale-space extrema is done by searching all scales and all image pixels. During the search, all scale-space invariant pixels are detected which serve as candidate keypoints.
- **Localization of keypoints:** In order to eliminate the unstable keypoints, the inter-pixel position and the scale of every candidate keypoint are considered.
- **Assignment of an orientation:** By computing the gradients around a keypoint pixel within a fixed window size, one or more orientations can be assigned to a keypoint. This leads to features which are invariant with respect to rotation.
- **Computation of keypoint descriptor:** A feature vector for every keypoint is computed. Since an orientation is added to the feature vector, a local invariance is achieved concerning rotation, deformation, illumination, and noise.

### 3.4.1 Scale-space and Detection of Extrema

Witkin (1983) proposed that the scale should be treated as a continuous parameter. Later on, the representation of image structures via scale-space was elaborated by Witkin’s scale-space theory. Koenderink (1984) and Lindeberg (1994) analyzed the scale-space characteristics in detail. It was shown that the scale-space of an image can be computed in an optimal way by using the Gaussian function. Thus, the scale-space  $L(x, y, \sigma)$  of an image is defined as follows:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3.24)$$

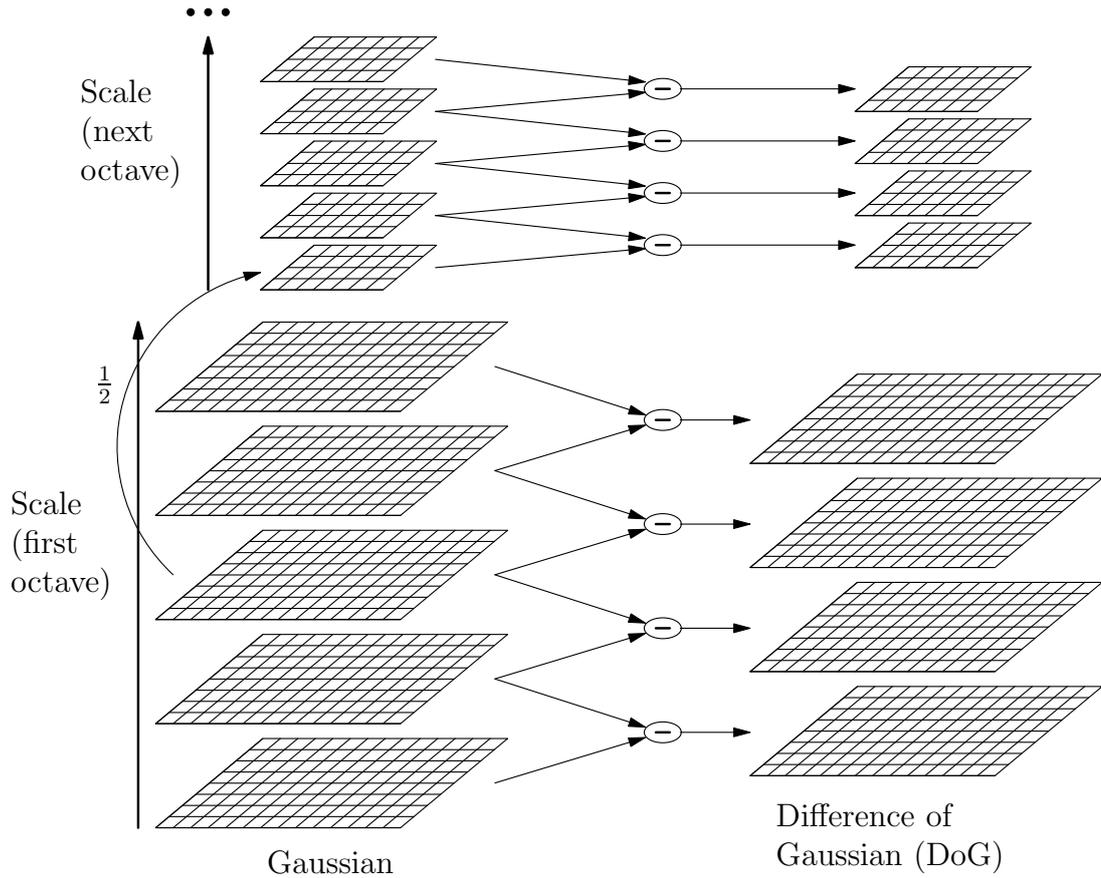


Figure 3.2: The left stack of images shows the first octave and the next, which is downsampled by a factor of 2. For each octave, the initial image is repeatedly convolved with a Gaussian filter. The difference of Gaussians is computed from adjacent images. The figure was adapted from Lowe (2004).

where  $*$  is the convolution,  $I(x, y)$  is the image pixel intensity at position  $x, y$ , and  $G(x, y, \sigma)$  is a Gaussian function of the form:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (3.25)$$

Lowe (1999) suggested that an efficiently detection of stable keypoints could be accomplished by looking for extrema in the difference of Gaussians of two neighbored scale-planes. The neighboring scale-planes are separated by a constant factor  $k$ .

$$D(x, y, \sigma) = (G(x, y, k \cdot \sigma) - G(x, y, \sigma)) * I(x, y) \quad (3.26)$$

$$= L(x, y, k \cdot \sigma) - L(x, y, \sigma) \quad (3.27)$$

On one hand,  $D$  can be calculated very efficient since the Gaussian images  $L$  have to be computed anyway and only need to be subtracted from each other. On

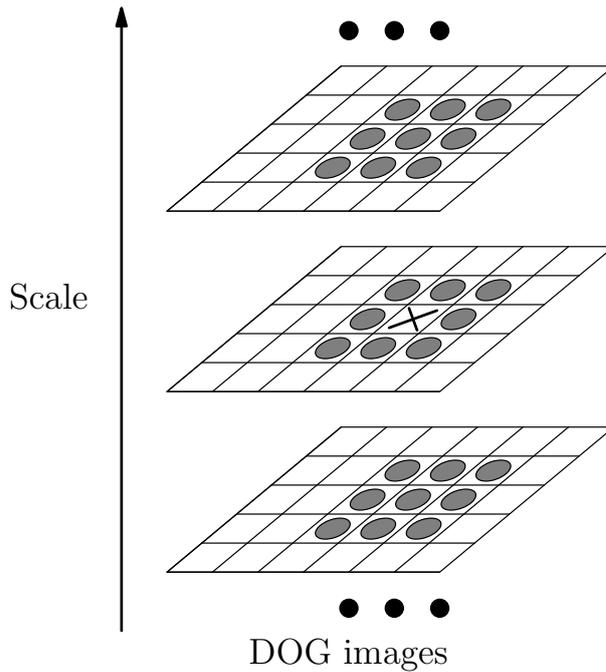


Figure 3.3: The extrema are computed through a  $3 \times 3$  neighborhood in the scale space octave. The figure is adapted from Lowe (2004).

the other hand, the difference of Gaussian is an approximation for the Laplace operator  $\sigma^2 \nabla^2 G$  normalized to scale. Lindeberg (1994) showed that invariance in scale is not achieved without this normalization. According to Mikolajczyk (2002), the keypoints, computed from the difference of Gaussian extrema, are the most stable compared to other methods.

Searching pixels which are a scale-space extrema is realized by comparing each pixel with its neighbors (e.g. figure 3.3). At first the neighboring pixel of the same DOG are compared. If the tested pixel is a minimum or a maximum in the DOG, the neighbors of the neighboring DOG's are checked. If the pixel is a minimum or maximum of the 26 neighbors, it is marked as a possible candidate keypoint.

### 3.4.2 Localization of Keypoints

Brown and Lowe (2002) describe how the localization of the keypoints can be approximated by a Taylor series. The approximation improves the stability of the keypoints. An extremum at  $\mathbf{p}_0 = (x_0, y_0, \sigma_0)^T$  can be represented by a quadratic Taylor polynomial at point  $\mathbf{p}$

$$D(\mathbf{x}) \approx D(\mathbf{p}) + \nabla D(\mathbf{p})^T (\mathbf{x} - \mathbf{p}) + \frac{1}{2} (\mathbf{x} - \mathbf{p})^T \mathbf{H}(\mathbf{p}) (\mathbf{x} - \mathbf{p}) \quad (3.28)$$

with  $\mathbf{x} = (x, y, \sigma)^T$  and the Hessian matrix  $\mathbf{H}(\mathbf{p})$ :

$$\mathbf{H}(\mathbf{p}) = \begin{bmatrix} D_{xx}(\mathbf{p}) & D_{xy}(\mathbf{p}) & D_{x\sigma}(\mathbf{p}) \\ D_{yx}(\mathbf{p}) & D_{yy}(\mathbf{p}) & D_{y\sigma}(\mathbf{p}) \\ D_{\sigma x}(\mathbf{p}) & D_{\sigma y}(\mathbf{p}) & D_{\sigma\sigma}(\mathbf{p}) \end{bmatrix} \quad (3.29)$$

$D_{**}$  are partial derivatives. The Hessian matrix  $\mathbf{H}$  as well as the gradient of  $D$  are approximated by the difference of neighboring image pixels. In this case, the derivation of equation 3.28 gives

$$D'(\mathbf{x}) \approx \nabla D(\mathbf{p}) + \mathbf{H} \cdot (\mathbf{x} - \mathbf{p}). \quad (3.30)$$

Rearranging equation 3.30 yields:

$$\hat{\mathbf{x}} = \mathbf{x} - \mathbf{p} = -\mathbf{H}^{-1} \nabla D(\mathbf{p}) \quad (3.31)$$

which computes the offset  $\hat{\mathbf{x}}$ . If all components of  $\hat{\mathbf{x}}$  are larger than 0.5, the keypoint is closer to another image point and the extremum needs to be corrected, which leads to a recomputation of the keypoint offset. Extremum and offset are accumulated and constitute the location of the keypoint. Furthermore, a scale is added to the keypoint, which is computed by:

$$l = \sigma \cdot \left( \frac{n + \hat{\mathbf{x}}_3}{s} \right)^2, \quad (3.32)$$

with  $\sigma = 1.6$  (Lowe, 2004),  $n$  being the index of the corresponding DOG image,  $s$  the number of images per octave, and  $\hat{\mathbf{x}}_3$  the third component of the offset. Furthermore, keypoints with low contrast are eliminated, since they are sensitive to noise. Combining equation 3.31 and equation 3.28 gives

$$D(\mathbf{p} + \hat{\mathbf{x}}) = D(\mathbf{p}) - \nabla D(\mathbf{p})^T \mathbf{H}^{-1} \nabla D(\mathbf{p}) \quad (3.33)$$

$$\begin{aligned} & + \frac{1}{2} (\mathbf{H}^{-1} \nabla D(\mathbf{p}))^T \mathbf{H} (\mathbf{H}^{-1} \nabla D(\mathbf{p})) \\ & = D(\mathbf{p}) - \nabla D(\mathbf{p})^T \mathbf{H}^{-1} \nabla D(\mathbf{p}) + \frac{1}{2} \nabla D(\mathbf{p})^T \mathbf{H}^{-1} \nabla D(\mathbf{p}) \\ & = D(\mathbf{p}) - \frac{1}{2} \nabla D(\mathbf{p})^T \mathbf{H}^{-1} \nabla D(\mathbf{p}) \\ & = D(\mathbf{p}) + \frac{1}{2} \nabla D(\mathbf{p})^T \hat{\mathbf{x}}. \end{aligned} \quad (3.34)$$

Keypoints with  $|D(\mathbf{p} + \hat{\mathbf{x}})| < 0.03$  are rejected; pixel values are normalized within the parameters of zero and one.

The difference of Gaussians computes large values for pixels at both edges and corners. Since SIFT only concentrates on corners, the pixels on edges should be rejected. Note, normally the maximum is a corner but there are special situations where this is not the case. For this purpose the change of gradient needs to be considered. The bending of the DOG function can be measured by the Hessian matrix

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \quad (3.35)$$

where  $D_{**}$  are the second order partial derivatives, which are approximated by the difference of the neighbours. The larger the ratio of eigenvalues is, the worse the surrounding of the pixel is. Since it is only necessary to take a look at the ratio, the eigenvalues do not need to be computed. The Hessian matrix is  $2 \times 2$ , thus it has 2 eigenvalues. If  $\alpha$  is the larger one and  $\beta$  the smaller one then the trace and the determinant is defined by

$$\text{trace}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta \quad (3.36)$$

$$\det(\mathbf{H}) = D_{xx}D_{yy} - D_{xy}^2 = \alpha \cdot \beta. \quad (3.37)$$

The pixel is rejected if the determinant is negative. Suppose there is a correlation of the eigenvalues  $\alpha = r\beta$ , thus the following equation can be obtained

$$\frac{\text{trace}(\mathbf{H})^2}{\det(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha \cdot \beta} = \frac{(r \cdot \beta + \beta)^2}{\alpha \cdot \beta^2} = \frac{(r + 1)^2}{r}. \quad (3.38)$$

If the eigenvalues are equal,  $\frac{(r+1)^2}{r}$  is as small as possible and increases with  $r$ . Thus, for a given  $r$  the eigenvalues are not needed. It is necessary to check whether

$$\frac{\text{trace}(\mathbf{H})^2}{\det(\mathbf{H})} < \frac{(r + 1)^2}{r} \quad (3.39)$$

is true or not. Lowe (2004) suggests that pixels with a ratio of the eigenvalues larger than 10 should be rejected.

### 3.4.3 Assigning an Orientation

The orientation of a keypoint, computed from the local qualities of the image at the keypoint position, can be used to relatively rotate the keypoint descriptor. Thereby, the computed keypoint descriptor is invariant to image rotation. Since the keypoints are computed in  $D(x, y, \sigma)$ , the smoothed image  $L(x, y, \sigma)$  with

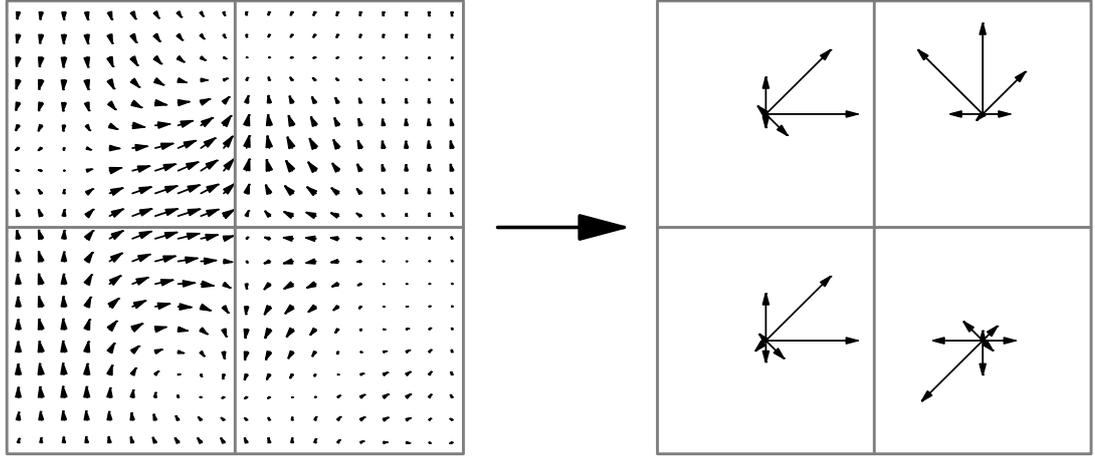


Figure 3.4: A keypoint descriptor is computed by first computing the gradient magnitude and orientation in a region around the keypoint location, shown on the left. Then, the samples are accumulated into orientation histograms, shown on the right side. The length of each arrow corresponds to the sum of the gradient magnitudes. The figure shows a  $2 \times 2$  descriptor array and is adapted from Lowe (2004)

the same  $\sigma$  as the keypoint is selected and used for estimating the bearing  $\theta(x, y)$  and the magnitude  $m(x, y)$  of the gradient. Let's assume

$$\Delta x = L(x + 1, y, \sigma) - L(x - 1, y, \sigma) \quad (3.40)$$

$$\Delta y = L(x, y + 1, \sigma) - L(x, y - 1, \sigma) \quad (3.41)$$

then  $m(x, y)$  and  $\theta(x, y)$  are defined by

$$m(x, y) = \sqrt{\Delta x^2 + \Delta y^2} \quad (3.42)$$

$$\theta(x, y) = \arctan\left(\frac{\Delta y}{\Delta x}\right). \quad (3.43)$$

Using the bearings of the keypoint, a histogram of orientations can be generated. The histogram has 36 bins, covering  $360^\circ$ , every bin represents  $10^\circ$ . For every pixel in the neighbourhood of the keypoint at the bearing direction, the magnitude of the gradient is accumulated, weighted by the distance to the keypoint and a Gaussian function, and added to the histogram. The influence of a magnitude gets smaller, the larger the distance to the keypoint is.

### 3.4.4 Computation of Keypoint Descriptor

The keypoint descriptor represents the scale invariant features. It is a feature vector which summarizes all the necessary features (keypoint location, scale,

and orientation). The keypoint descriptor is composed from the histograms of orientations (e.g. figure 3.4). The histograms are computed on a rectangular region around the keypoint. The magnitudes of the gradients are weighted by the Gaussian-weighted distance to the keypoint. The  $\sigma$  of the Gaussian function depends on the grid size of the histograms (e.g. figure 3.4,  $\sigma = 0.5 \cdot 2 = 1$ ). Not all gradients are relevant for the descriptor, thus only gradients are considered which are in the radius

$$r = \sqrt{2} \cdot l \cdot \frac{i + 1}{2}, \quad (3.44)$$

where  $l$  is the scale of the keypoint and  $i$  the grid size of the descriptor. Afterwards, an orientation can be assigned to the descriptor as pointed out in section 3.4.3. In the last step, the descriptor is normalized.

### 3.5 Image Parameters

Concerning homing with parameter methods, statistical image parameters and image parameters derived from the Fourier transformation are introduced. The input images are represented by their corresponding parameter set, and only the parameter sets are used by the parameter homing method. Thus, the parameter set of an input image is a compressed data representation of the image.

Due to the fact that some image parameters (e.g. center of gravity) highly reduce the image information that is carried by the parameter set, it is necessary to extend the parameter set for an image by segmenting the input image by rings as depicted in figure 3.5. Concentric rings on panoramic images are equivalent to image rows in unfolded images. Except for the center of gravity, which is computed on the rings of the original panoramic image, the statistical parameters are computed on the rows of the unfolded image.

A functional transformation like the discrete Fourier transform (DFT) can be used to reduce the amount of data and to achieve an invariance in rotation if only the amplitude information is considered (Menegatti et al., 2004). Since the unfolded panoramic image is closed in x-direction, the one-dimensional discrete Fourier transform is applied line-by-line.

$$\mathcal{F}_k(I) = \sum_{i=0}^{N-1} I_i \exp(-\iota k i \frac{2\pi}{N}), \quad (3.45)$$

where  $I$  is an image row with  $N$  pixels,  $k$  the frequency,  $\iota$  the imaginary entity.

The following statistical measures are used for parameter homing tests. The image parameters are categorized by four classes; statistics with and without

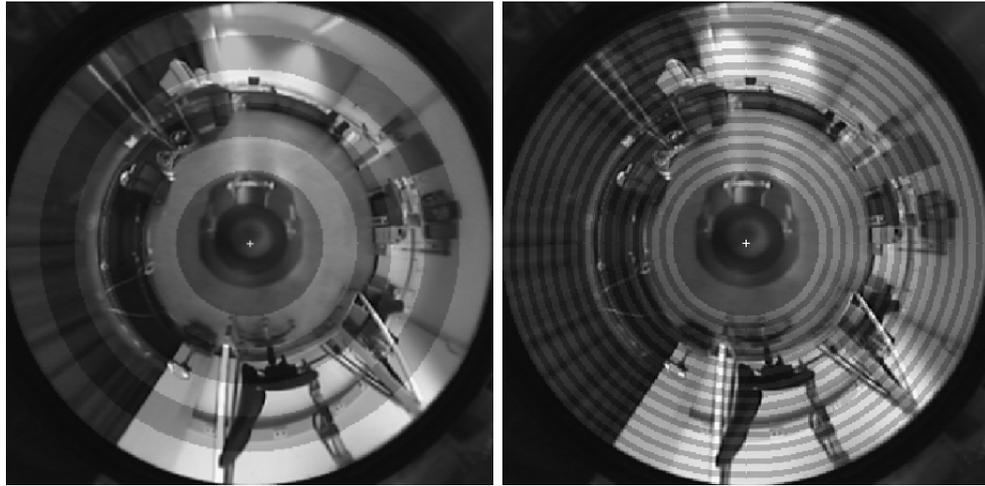


Figure 3.5: Panoramic images from **original** database, position (0,0) with no preprocessing. The light and the dark textured areas visualize the ring segmentation, which is used to extend the number of parameters for the signature. Concentric rings on panoramic images are equivalent to image rows in unfolded images. The widths of the rings are equal. **Left:** 5 rings; **Right:** 25 rings. (Adapted from Fischer (2006).)

invariance in rotation, and DFT parameters with and without invariance in rotation. If the computed parameter set is invariant in rotation, two parameter sets could be compared even if their orientation is not aligned which supersedes a compass.

The panoramic image  $P$ , which is captured by the optical system, has  $R$  rows and  $C$  columns. The unfolded image  $I$  has  $M$  columns and  $N$  rows.  $i$  depicts the index of the column,  $j$  the index of the row. The bold ciphers surrounded by squared brackets are used as identifier for the used parameter so that the tests results are easier to interpret.

### 3.5.1 Statistics without Invariance in Rotation

#### Center of Gravity [cog]

The center of gravity  $\mathbf{s} = (i_s, j_s)^T$  is computed on the original panoramic image. Its components are defined as follows:

$$i_s = \frac{1}{E} \sum_{ij} P_{ij} \cdot i \quad \text{where} \quad E = \sum_{ij} P_{ij}. \quad (3.46)$$

$$j_s = \frac{1}{E} \sum_{ij} P_{ij} \cdot j \quad \text{where} \quad E = \sum_{ij} P_{ij}. \quad (3.47)$$

**Image [image]**

For the sake of completeness, here the parameter is the complete image. Later on, this parameter is used to compare the different homing methods.

**3.5.2 Statistics with Invariance in Rotation****The Length of the Center of Gravity [cog\_l]**

$$| \mathbf{s} | = | (i_s, j_s)^T | = \sqrt{i_s^2 + j_s^2} \quad (3.48)$$

**Average [mean]**

$$\bar{I} = \mu = \frac{1}{MN} \sum_{ij} I_{ij} \quad (3.49)$$

**Standard Deviation [std]**

$$\sigma = \sqrt{\frac{1}{MN} \sum_{ij} (I_{ij} - \bar{I})^2} \quad (3.50)$$

**Skewness [skew]**

The skewness is a measure for the asymmetry of a probability distribution.

$$v = \frac{\mu_3}{\sigma^3} = \frac{\sum_{ij} (I_{ij} - \bar{I})^3}{MN \cdot \sigma^3}, \text{ where } \mu_3 = E[(X - \mu)^3] \quad (3.51)$$

**Kurtosis [kurt]**

The kurtosis is a measure for the sharpness of a distribution.

$$g = \frac{\mu_4}{\sigma^4} - 3 = \frac{\sum_{ij} (I_{ij} - \bar{I})^4}{MN \cdot \sigma^4} - 3, \text{ where } \mu_4 = E[(X - \mu)^4] \quad (3.52)$$

**Histogram [hist]**

A histogram with  $B$  bins is computed from the image.

$$h_b = \text{number of pixels in the bin } b, \text{ where } (0 \leq b \leq B - 1) \quad (3.53)$$

**Entropy [ent]**

The entropy is a statistical measure of randomness that can be used to characterize the texture of the image.

$$S = - \sum_{b=0}^{B-1} h_b \cdot \ln(h_b) \quad (3.54)$$

**3.5.3 DFT Parameters without Invariance in Rotation**

The discrete Fourier transformation presented in equation 3.45 can also be used to compute image parameters. Again, the parameters can be designed with or without rotation invariance.

**Fourier Coefficient [fc]**

$$\mathcal{F}_k(I) = \sum_{i=0}^{N-1} I_i \exp(-\iota k i \frac{2\pi}{N}) \quad (3.55)$$

**DFT of the Column Average with Phase [mean\_ft]**

$$\mathcal{F}_k(\mu) = \sum_{i=0}^{N-1} \mu_i \exp(-\iota k i \frac{2\pi}{N}), \quad \text{where } \mu_i = \frac{1}{M} \sum_j I_{ij} \quad (3.56)$$

**DFT of the Standard Deviation of the Columns with Phase [std\_fft]**

$$\mathcal{F}_k(\sigma) = \sum_{i=0}^{N-1} \sigma_i \exp(-\iota k i \frac{2\pi}{N}), \quad \text{where } \sigma_i = \sqrt{\frac{1}{M} \sum_j (I_{ij} - \mu_i)^2} \quad (3.57)$$

**DFT of the Center of Gravity of the Rows with Phase [cog\_fft]**

$$\mathcal{F}_k(j_s) = \sum_{i=0}^{N-1} j_{s_i} \exp(-\iota k i \frac{2\pi}{N}), \quad \text{where } j_{s_i} = \frac{\sum_j I_{ij} \cdot j}{\sum_j (I_{ij})} \quad (3.58)$$

**3.5.4 DFT Parameters with Invariance in Rotation****Fourier Coefficient without Phase Information [fc\_abs]**

$$|\mathcal{F}_k(\mu)| = \left| \sum_{i=0}^{N-1} I_i \exp(-\iota k i \frac{2\pi}{N}) \right| \quad (3.59)$$

**DFT of the Column Average without Phase Information [mean\_fft\_abs]**

$$|\mathcal{F}_k(\mu)| = \left| \sum_{i=0}^{N-1} \mu_i \exp(-\iota k i \frac{2\pi}{N}) \right|, \text{ where } \mu_i = \frac{1}{M} \sum_j I_{ij} \quad (3.60)$$

**DFT of the Standard Deviation of the Columns without Phase [std\_fft\_abs]**

$$|\mathcal{F}_k(\sigma)| = \left| \sum_{i=0}^{N-1} \sigma_i \exp(-\iota k i \frac{2\pi}{N}) \right|, \text{ where } \sigma_i = \sqrt{\frac{1}{M} \sum_j (I_{ij} - \mu_i)^2} \quad (3.61)$$

**DFT of the Center of Gravity of the Rows without Phase [cog\_fft\_abs]**

$$|\mathcal{F}_k(j_s)| = \left| \sum_{i=0}^{N-1} j_{s_i} \exp(-\iota k i \frac{2\pi}{N}) \right|, \text{ where } j_{s_i} = \frac{\sum_j I_{ij} \cdot j}{\sum_j (I_{ij})} \quad (3.62)$$

## Chapter 4

### Polarized-Light Compass

The polarization pattern of the sky is formed because the sunlight is scattered at the atmosphere. The directions of polarization form a regular pattern which generates concentric rings around the sun. The polarization-plane at the zenith is perpendicular to the solar-meridian plane spanned by the observer, the sun and the zenith (Lambrinos et al., 1997, 2000). To detect the polarization pattern, a POL sensor is needed. A POL sensor can be build by a photosensitive device, a linear polarizer and spectral filters. The output of a POL-Sensor is described by

$$s(\phi) = KI(1 + d \cos(2\phi - 2\phi_i)), \quad (4.1)$$

where  $K$  is an unknown constant,  $I$  the measured light intensity,  $d$  the grade of polarization,  $\phi$  is the current orientation with respect to the solar meridian  $\phi_i$  is the value of  $\phi$  that maximizes  $s$  (Lambrinos et al., 1997, 2000).

Two POL sensors build a polarization-opponent unit (POL-OP-Unit) “that are functionally similar to the POL-neurons found in insects. The POL-OP-Unit enhances the e-vector contrasts sensitivity and makes the e-vector response insensitive to fluctuations of light intensity” (Lambrinos et al., 2000). The output of a POL-OP-Unit is described by

$$p_i(\phi) = \log \left( \frac{1 + d \cos(2\phi - 2\phi_i)}{1 - d \cos(2\phi - 2\phi_i)} \right). \quad (4.2)$$

To delogarithmize equation 4.2, the sigmoidal function  $\tilde{p}_i(\phi) = \frac{1}{1+10^{p_i(\phi)}}$  is applied and gives

$$1 - 2\tilde{p}_i(\phi) = d \cos(2\phi - 2\phi_i). \quad (4.3)$$

The grade of polarization  $d$  is unknown in equation 4.3 but can be eliminated by combining two POL-OP-Units; the relative orientation  $\phi$  towards the sun can be computed. The output of two combined POL-OP-Units is given by

$$\tan(2\phi) = \frac{f_{ij} \cos(2\phi_j) - \cos(2\phi_i)}{\sin(2\phi_i) - f_{ij} \sin(2\phi_j)} \quad (4.4)$$

$$\phi = \frac{1}{2} \arctan \left( \frac{f_{ij} \cos(2\phi_j) - \cos(2\phi_i)}{\sin(2\phi_i) - f_{ij} \sin(2\phi_j)} \right), \quad (4.5)$$

where  $f_{ij} = \frac{1-2\tilde{p}_i}{1-2\tilde{p}_j}$ .

The function  $\phi(\tilde{p}_i, \tilde{p}_j)$  should be in the range  $[-\pi, \pi]$  but the result of equation 4.5 is in the range  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ . To extend the range of usual principal values to  $[-\pi, \pi]$  the periodicity of the arctan function is used and gives

$$\varphi(\tilde{p}_i, \tilde{p}_j) = \begin{cases} (\phi(\tilde{p}_i, \tilde{p}_j), \phi(\tilde{p}_i, \tilde{p}_j) - \pi) & \text{if } \phi(\tilde{p}_i, \tilde{p}_j) \geq 0 \\ (\phi(\tilde{p}_i, \tilde{p}_j), \phi(\tilde{p}_i, \tilde{p}_j) + \pi) & \text{if } \phi(\tilde{p}_i, \tilde{p}_j) < 0 \end{cases} \quad (4.6)$$

The result of equation 4.6 is a pair of candidate orientations  $(\varphi, \varphi + \pi)$ , but only one represents the real situation. Thus, it is necessary to decide which of the two possible orientations is the right one. For this purpose an error function is used which compares the computed and the desired result. The orientation with the lower error is chosen to be the correct orientation.

## 4.1 Hardware

The polarized-light compass is built with six photodiodes from Hamamatsu (S5821-1, TO-18 package with lens) with a spectral sensitivity from 300nm to 1100nm and an angle of aperture of  $\pm 10^\circ$ . A standard linear polarized filter foil is added on top of every photodiode. The orientation of the polarized filter is analog to Lambrinos et al. (2000, 1997). The polarized filter set of each POL-OP-Unit is rotated with respect to the first filter set by  $+60^\circ$  and  $+120^\circ$ . The filters within one POL-OP-Unit are mounted with  $90^\circ$  rotation with respect to each other. The spectral sensitivity is limited by a spectral glass filter combination, depicted by figure 4.1 because the transmission maximum of linear polarized-light is at a spectral frequency of 550nm (Horváth and Varjú, 2004). The polarized-light compass is shown on the right side of figure 4.3. Figure A.2 shows the circuit diagram for the polarized-light compass.

## 4.2 Tests

The tests are used to verify that the compass works as expected. For this purpose, an artificial test setup proves the quality of the POL-OP-Units. After-

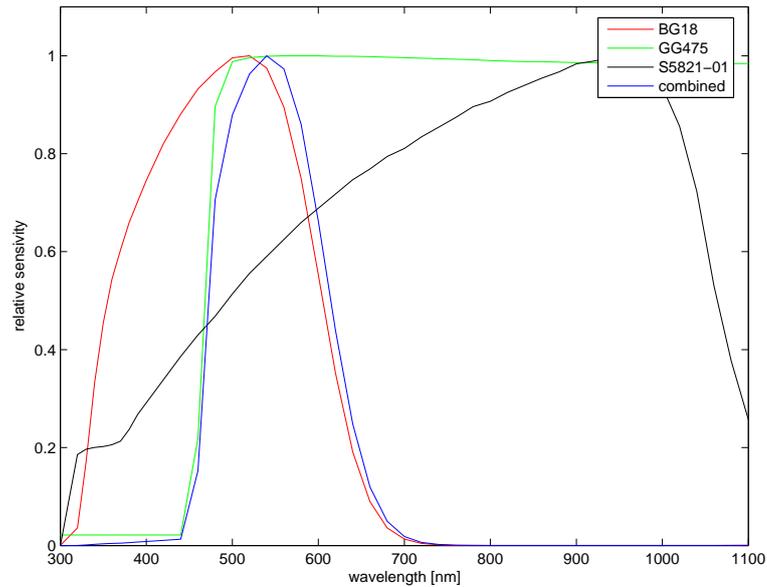


Figure 4.1: Spectral sensitivity (normalized to a maximum of 1) of the used Schott glass filters for the polarization compass, retrieved from Schott’s program Filter2004. Only the wavelengths from 300 to 1100nm are considered since the diode cutoff is at 1150nm. The spectral sensitivity of the photodiodes is depicted by the black line. Figure A.1 shows the unnormalized spectral sensitivity.

ward, the polarized-light compass is compared to an electromagnetic compass. Finally, the e-vector array of the sky is measured and shown.

### 4.2.1 Polarized Filter Test

For this test a non-polarized indirect lighting was used. An additional polarized filter was mounted onto the polarized-light compass. The foil was counterclockwise rotated. The compass measured 100 samples at every successive orientation. The step size was  $3^\circ$ . The averaged output of the POL-OP-Units is shown in figure 4.2. It can be seen, that the maximum output voltages of the POL-OP-Units are slightly different, because of the production tolerances of the electrical components (photodiodes, resistors) used for the sensors. The tolerances are compensated by the software that drives the compass. The blue color depicts the first POL-OP-Unit, green the second, and red the third.

### 4.2.2 Polarized-Light vs. Electromagnetic Compass

This test compares the output of the polarized-light compass to a electromagnetic compass (ActiveMedia Robotics TCM2). For this purpose, the polarized-

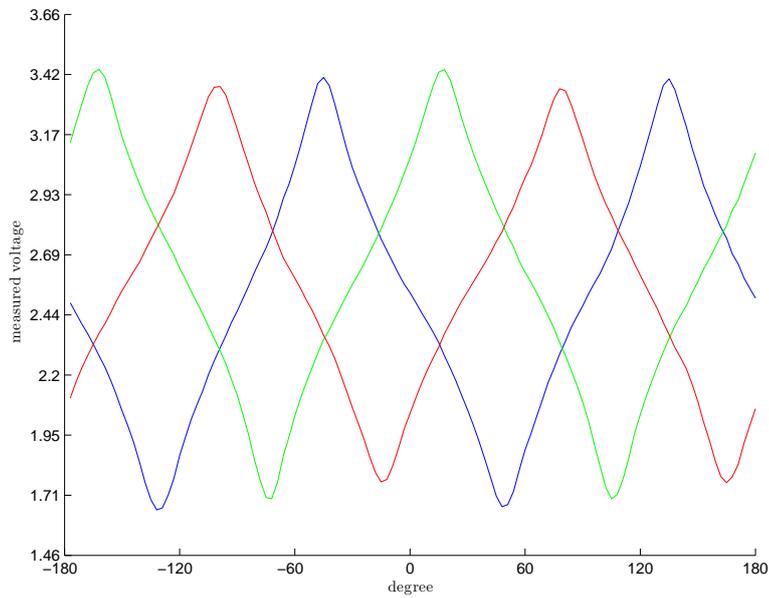


Figure 4.2: Voltage output of the POL-OP-Units, blue depicts the first POL-OP-Unit, green the second, red the third.

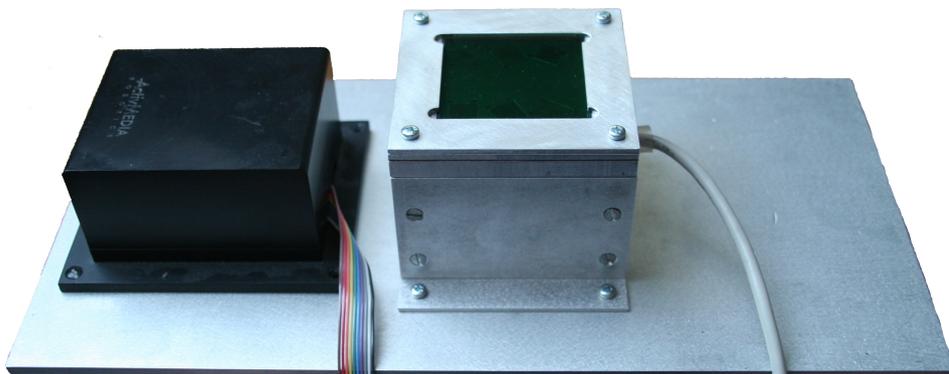


Figure 4.3: Test setup for comparison of polarized-light compass and electro-magnetic compass.

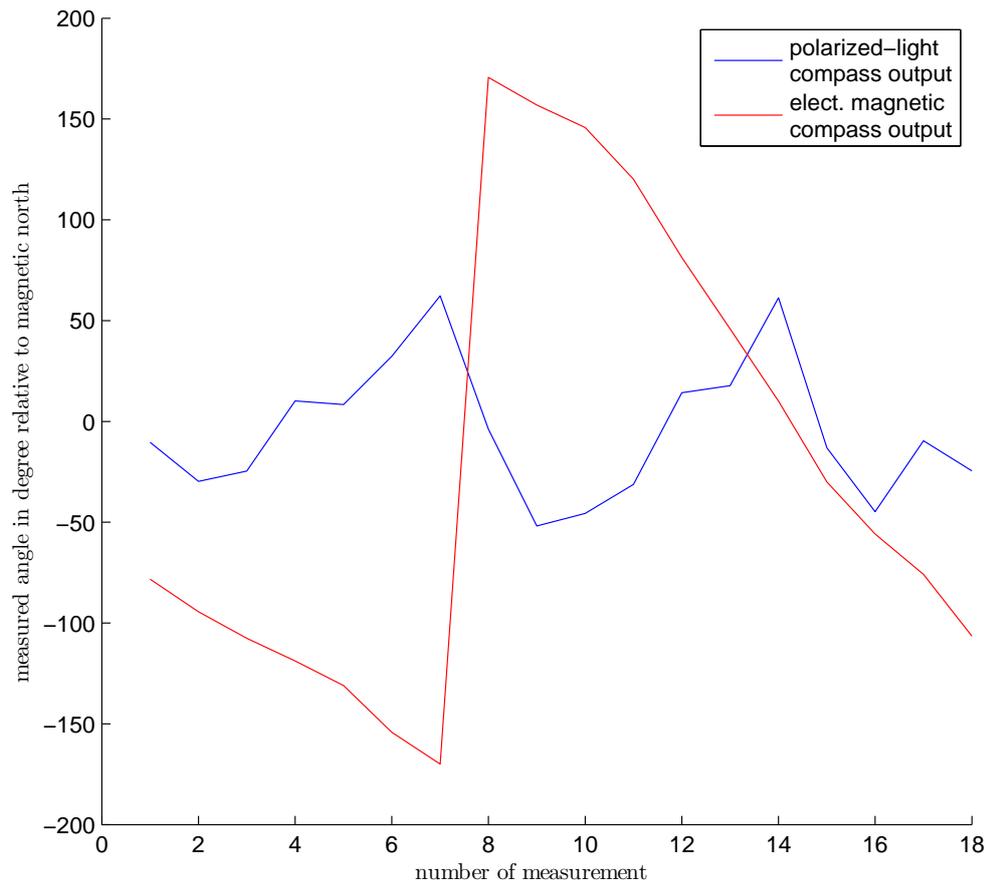


Figure 4.4: 18 manually rotated polarized-light compass positions are used for measurement. At every position, 100 successive measurements are taken. The measured compass output was averaged and plotted. The blue curve depicts the averaged measures, the red curve depicts the output of the electromagnetic compass.

light compass and the electromagnetic compass have been mounted on a plate. Figure 4.3 shows the setup. The test setup was pointed towards the blue sky. For the first test, 18 manually rotated compass positions are used. At every position, 100 successive measurements are taken and the output is averaged and visualized in figure 4.4. The second test is shown in figure 4.5. Again, the sensor is rotated manually, but this time, the output is not averaged.

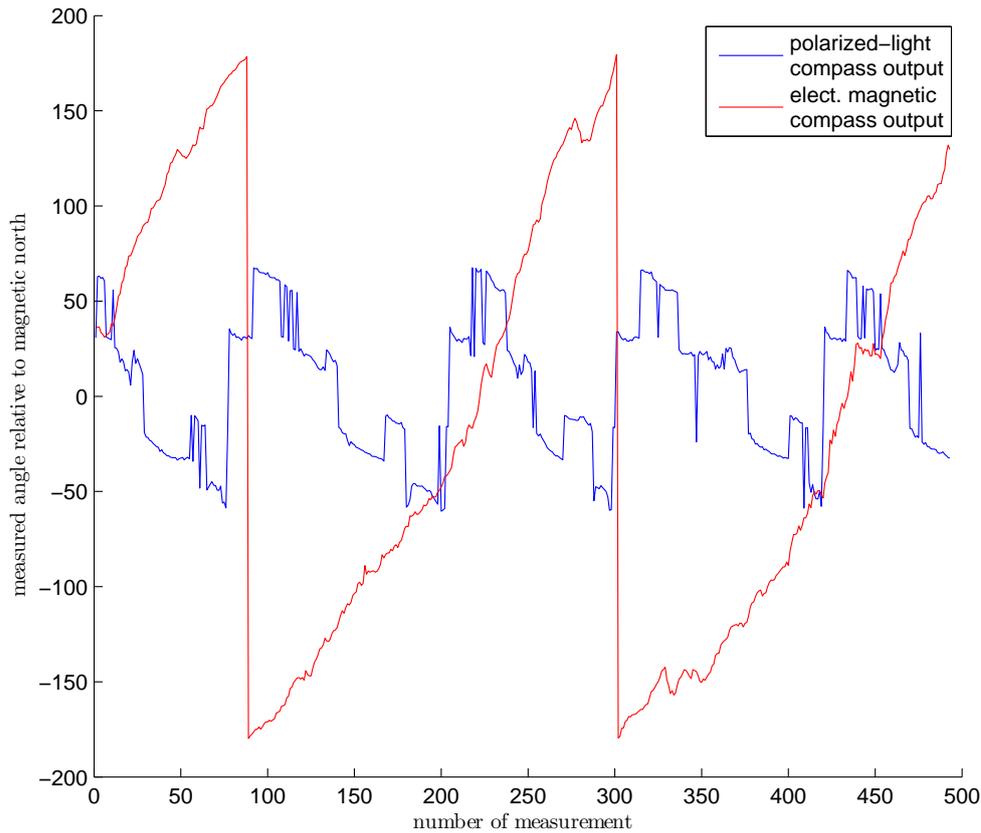


Figure 4.5: The test setup was rotated manually. The color coding is analog to figure 4.4.

### 4.2.3 Measuring the $e$ -vector Field of the Sky

The polarized-light compass was mounted on a Directed Perception pan-tilt unit (PTU-46-17.5). The position resolution of the encoders is  $0.01^\circ$ . The entire unit was mounted on a tripod. A systematic gridlike sampling of the sky can be achieved by rotating around the two axes of the pan-tilt unit. The measured data is shown in figure 4.6. The data is visualized by applying the polar transformation to the pan and tilt angle. The bars depict the output of the polarized-light compass for each pan-tilt position. The length of each bar represents the degree of polarization (see figure 4.7 for a schematized visualization).

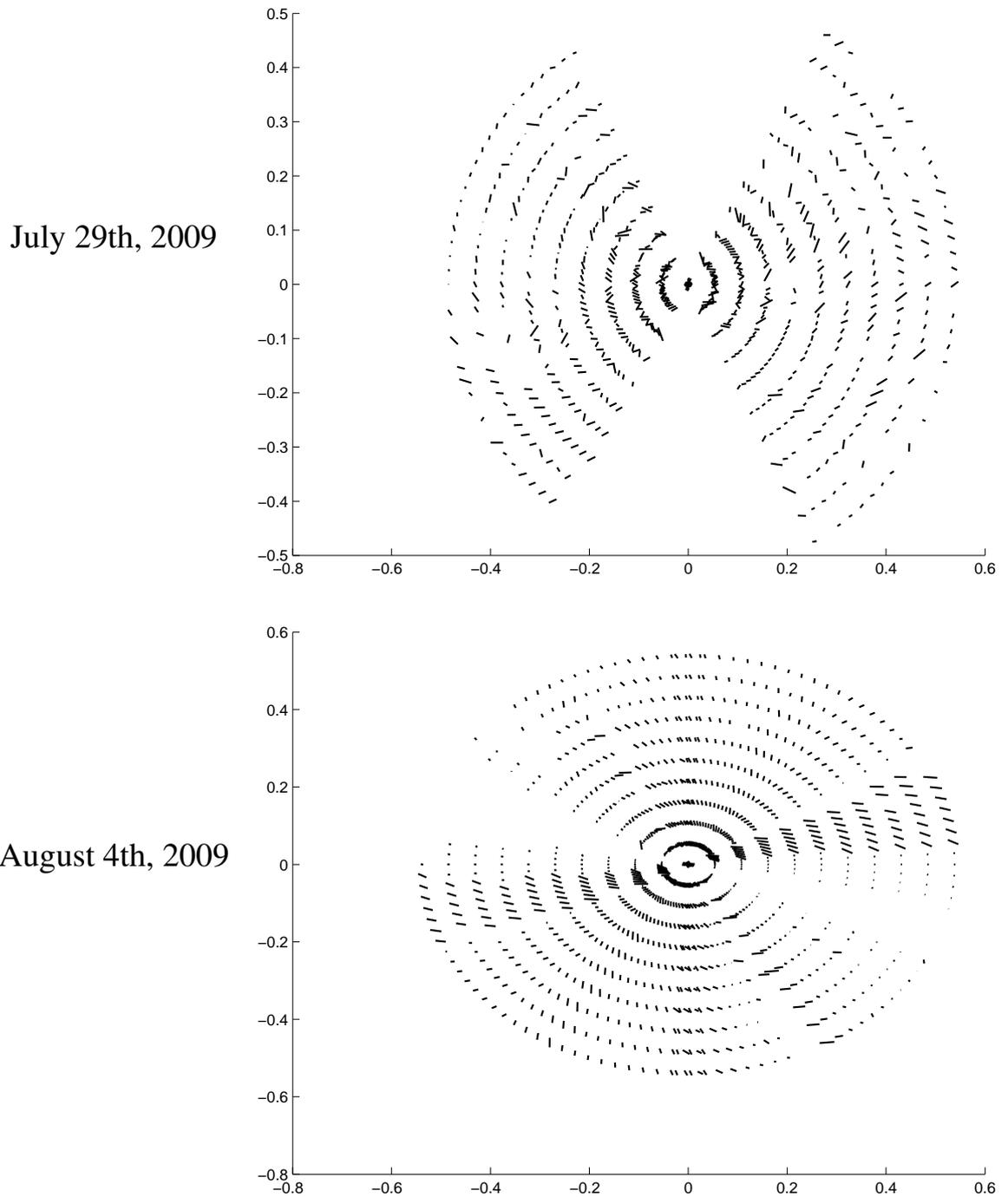


Figure 4.6: Polar plots of the measured  $e$ -vector field. The length of the bars depicts the degree of polarization. The position of each bar is calculated by applying the polar transformation to pan and tilt. The orientation of the bar is depicted by the measured compass angle.

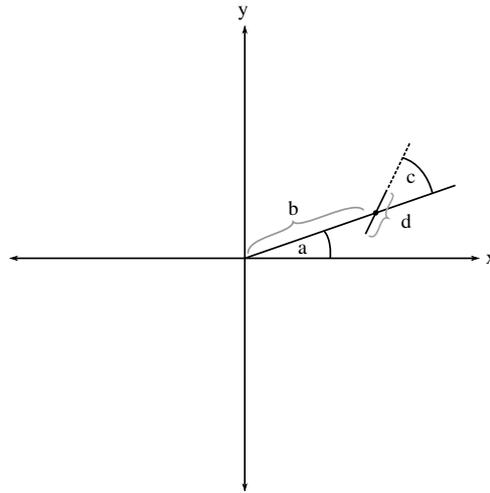


Figure 4.7: The diagram schematizes the polar transformation that is used to calculate the  $e$ -vector fields in figure 4.6.  $a$  depicts the pan angle, the radius  $b$  is the tilt angle,  $c$  is the measured angle of the polarized-light compass at the position (pan,tilt), and  $d$  depicts the measured grade of polarization.

#### 4.2.4 Discussion

Figure 4.2 shows the output of the polarized-light compass. The polarized filter set of each POL-OP-Unit is rotated with respect to the first filter set by  $+60^\circ$  and  $+120^\circ$ . As expected, the outputs of each POL-OP-Unit are also shifted by  $60^\circ$ . According to the tests with the electromagnetic compass, the polarized-light compass works as expected. The polarized-light compass can not distinguish between  $-180^\circ$  and  $0^\circ$  as shown in figure 4.5. The reason for this problem is given by the axial symmetry of equation 4.6. The angular output is given by  $\varphi, \varphi + \pi$ . Even so, the output of the polarized-light compass is as expected.

The  $e$ -vectored fields visualized in figure 4.6 show the grade of polarization of the sky. The  $e$ -vector field measured on August 4th, 2009 is from cloudless blue sky, on July 29, 2009 some clouds of vertical development appeared during data collection, which could be a reason for the scattered grade of polarization that is depicted by the orientation of the bars of the figure. In the  $e$ -vector field collected on August 4th, 2009 the orientations of the bars are aligned regularly. A systematic model of the orientation can be seen.

### 4.3 Summary and Conclusion

In this chapter, the polarized-light compass is presented. The output of the polarized-light compass was verified by an electromagnetic compass. Because of the pair of candidate orientations, it is necessary to decide which of the two

possible orientations is the right one. Without an adequate error function the correct orientation can not be determined. Afterward, the polarized-light compass was used to measure the  $e$ -vector field of the sky. The data of two  $e$ -vector fields were presented. Because of the limited sky data, a systematic analysis was not possible. Especially, the effect of clouds could not be analyzed.

# Chapter 5

## Local Visual Homing Methods

### 5.1 SIFT Homing

Realizing visual homing with scale invariant feature transformation can be achieved by computing SIFT features in current view and snapshot. The features in the current view need to be matched in the snapshot so that a local movement vector can be computed. Mapping the movement vector of the corresponding keypoints to the “real world” gives an individual home vector with respect to the local position of the panoramic image. The resulting home vector is calculated from all individual local home vectors.

#### 5.1.1 Movement Vector for Correlated Keypoints

Let’s assume that the current view image  $c$  and the snapshot image  $s$  are aligned; thus, the images have been captured in the same orientation or have been rotated using a compass system. After extracting the SIFT keypoint descriptor (section 3.4) for current view and snapshot, the correlation has to be computed, and the local movement vector is calculated. The quality of the correlation for two keypoint descriptors  $p' \in s$  and  $p \in c$  can be defined by computing the Euclidean distance of the keypoint descriptors:

$$d(p, p') = \|des_p - des_{p'}\| = \sqrt{\sum_j (des_p(j) - des_{p'}(j))^2}, \quad (5.1)$$

whereas  $des_p$  and  $des_{p'}$  are descriptors and  $j$  iterates over all dimensions of the descriptor. To find the corresponding keypoint  $\hat{p} \in s$  for every keypoint descriptor  $p \in c$ , an exhaustive search is used:

$$\hat{p} = \arg \min_{p' \in E_q(p)} d(p, p'), \quad (5.2)$$

where  $E_q(p)$  is the neighborhood with radius  $q$

$$E_q(p) = \{\tilde{p} \in s \mid \|p, \tilde{p}\| < q\}. \quad (5.3)$$

The condition decreases the complexity of the search since the correlating keypoints should be as close as possible to each other. The resulting local movement vector is defined by

$$\delta = \begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = \begin{pmatrix} \Delta_x \cdot (p_x - \hat{p}_x) \\ \Delta_y \cdot (p_y - \hat{p}_y) \end{pmatrix}, \quad (5.4)$$

where  $\Delta_x$  and  $\Delta_y$  are factors for angle per pixel in x- and y-direction. Hence, the local movement vector of two correlated keypoint descriptors characterizes the movement of the descriptors in the panoramic images.

### 5.1.2 Home Vector for Correlated Keypoint Descriptors

The local movement vectors which are computed in the panoramic images need to be transformed to world movements. The following method was suggested by Vardy and Möller (2005). A top view of the homing situation is schematized in figure 5.1(a), where  $s$  and  $c$  represent the positions of snapshot and current view. The circles indicate the panoramic images as seen from the positions  $s$  and  $c$ . It is assumed that the images are aligned, i.e. captured in the same direction, which is depicted by the arrows in the schematic. The landmark  $l$  stands for an object that is visible from both positions and which produces a pair of correlated keypoint descriptors in the panoramic images. In the snapshot image, the angle towards the landmark is  $\theta_x$ . The images are rotationally aligned as if the robot's orientation did not change during its movement. Hence, the landmark appears in the current view at  $\theta_x + \delta_x$ ;  $d_c$  and  $d_s$  are the unknown distances of the landmark to the positions  $s$  and  $c$ .  $\beta = \pi + \alpha$  is the home angle of the needed home vector, which has to be calculated. Even though the distances are unknown,  $\beta$  can be computed by finding the ratio of the distances. Figure 5.2 shows that  $l$  appears at  $\theta_y$  at position  $s$  and  $\theta_y + \delta_y$  at position  $c$  with a distance  $h_l$  to the horizon. Computing the tangent gives

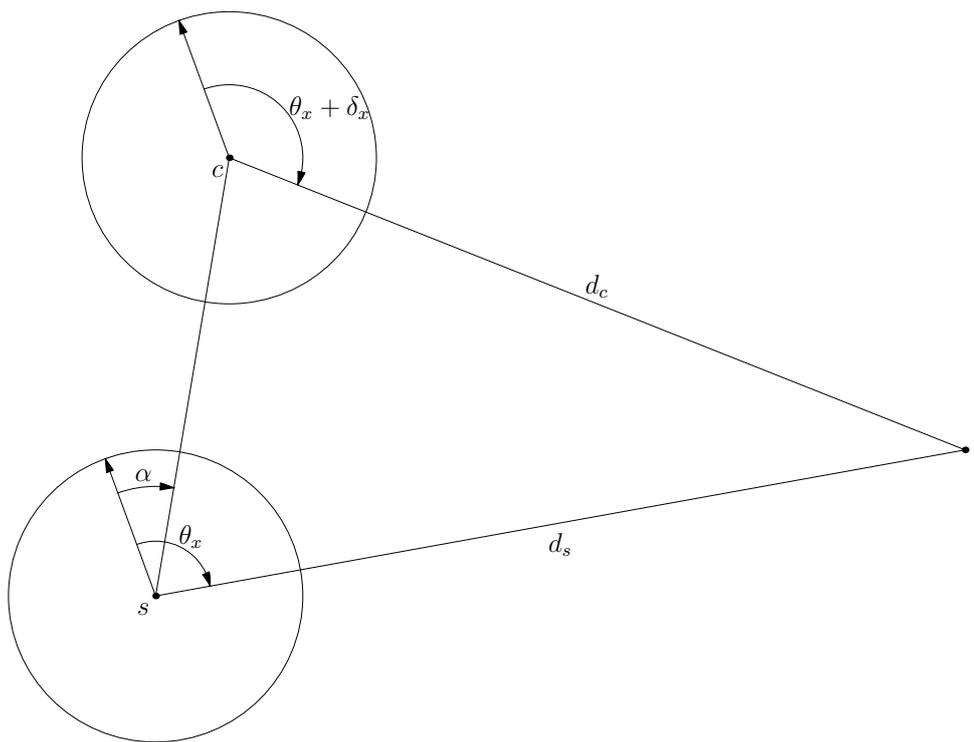
$$\tan \theta_y = \frac{h_l}{d_s} \quad (5.5)$$

$$\tan(\theta_y + \delta_y) = \frac{h_l}{d_c}. \quad (5.6)$$

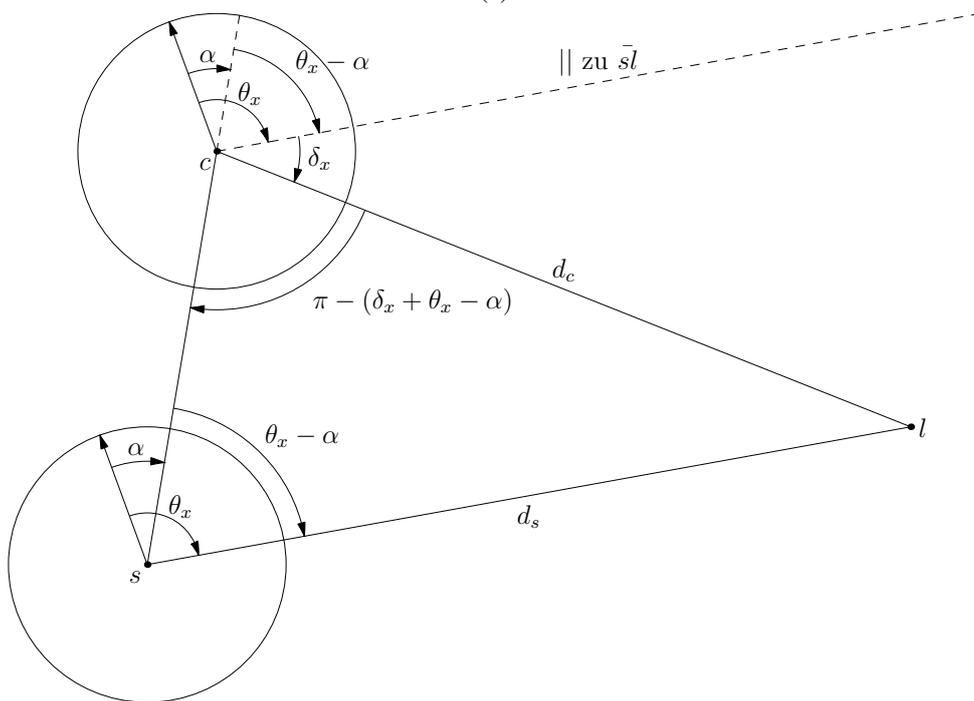
Now the ratio is defined by

$$\frac{d_s}{d_c} = \frac{\tan(\theta_y + \delta_y)}{\tan \theta_y}. \quad (5.7)$$

Keypoints with  $|\theta_y| \leq 0.01$  are ignored for numerical stability reasons. After computing the ratio of  $\frac{d_s}{d_c}$ , the homing problem schematized in figure 5.1(b) can



(a)



(b)

Figure 5.1: Top view of the homing situation,  $s$  and  $c$  depict the snapshot and the current view position; the circles constitute the panoramic view. The landmark  $l$  is a placeholder for an object that produces correlated keypoint descriptors in the images captured at the positions  $s$  and  $c$ . This figure was taken from Vardy and Möller (2005).

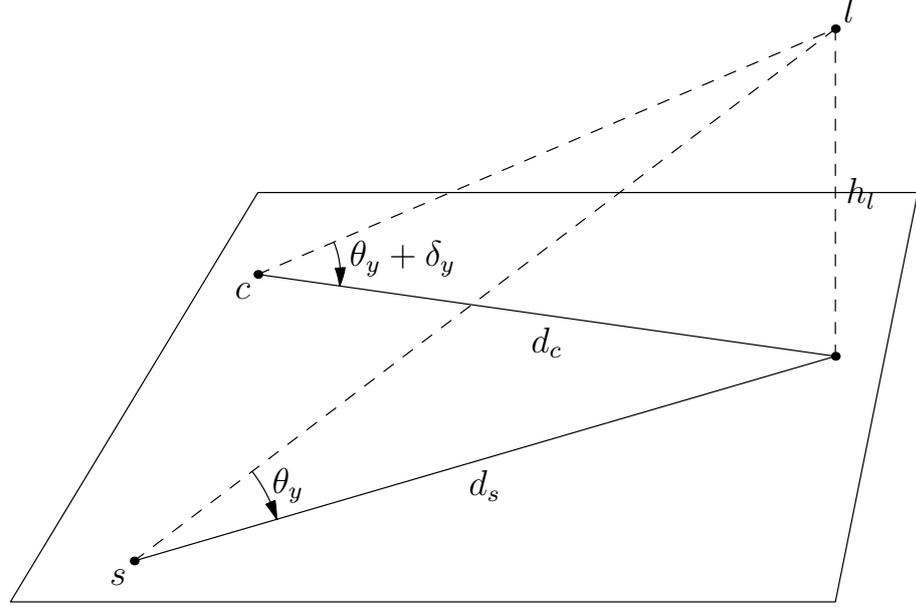


Figure 5.2: Side view of the homing situation,  $s$  and  $c$  depict the snapshot and current view position. The landmark  $l$  appears at corresponding angles above or below the horizon. This figure was taken from Vardy and Möller (2005)

be solved. Concerning the triangle  $\triangle scl$ , the angles  $\angle c$  and  $\angle s$  are relevant.  $\angle s$  is

$$\angle s = \theta_x - \alpha. \quad (5.8)$$

$\angle c$  is defined by

$$\angle c = \pi - (\delta_x + \theta_x - \alpha). \quad (5.9)$$

Thus, the law of sines gives

$$\frac{\sin(\pi - (\delta_x + \theta_x - \alpha))}{d_s} = \frac{\sin(\theta_x - \alpha)}{d_c}. \quad (5.10)$$

After trigonometric conversions, which are presented in Vardy and Möller (2005), follows

$$\beta = \theta_x - \arctan\left(\frac{\sin \delta_x}{\frac{d_s}{d_c} - \cos \delta_x}\right) + \pi. \quad (5.11)$$

Substituting equation 5.7 gives

$$\beta = \theta_x - \arctan\left(\frac{\sin \delta_x}{\frac{\tan(\theta_y + \delta_y)}{\tan \theta_y} - \cos \delta_x}\right) + \pi, \quad (5.12)$$

which can be used to compute a local home vector of two correlated keypoints by

$$\mathbf{u}(\theta_x, \theta_y) = \begin{pmatrix} \cos \beta \\ \sin \beta \end{pmatrix}. \quad (5.13)$$

### 5.1.3 Overall Home Vector

The overall home vector needs to be computed from the local home vectors for correlated keypoints (section 5.1.2). The extraction of SIFT keypoint descriptors provides a set of local home vectors which are more or less correct. They have to be combined to one global home vector without knowing which of the local home vectors are correct and which might be wrong. Two possible methods are presented; both rely on angle histograms of the home angles.

#### Home Vector Computed by Maximum of the Angle Histogram

Schatz (2006) recommends to search for the maximum in the angle histogram. To get rid of possible local maxima in the angle histogram, the maximum is searched indirectly. A window of width 10 is moved through the histogram. The values in the bins in the window are summed. The center of the window with the largest sum depicts the global maximum. Figure 5.3(b) displays an example search. According to the angle histogram in figure 5.3(a), the result of this method is  $-46^\circ$  which is close to the correct angle of  $-45^\circ$  computed from the snapshot and current view location.

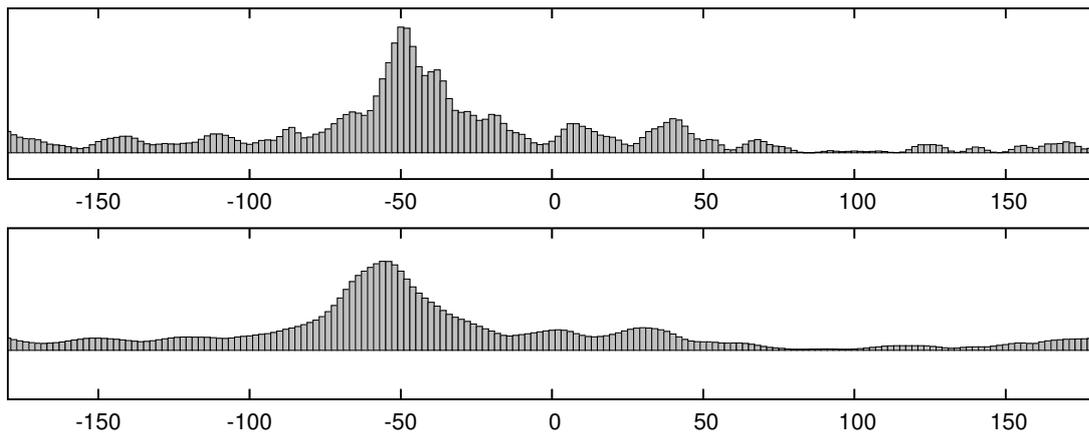


Figure 5.3: Figure (a) shows an angle histogram from the **original** database with snapshot position at (0,0) and current view at (6,6). Figure (b) shows the values for the search of the maximum of a search window width 10 (adapted from Schatz (2006)).

### Home Vector Calculated by the Average Vector

Vardy and Möller (2005) propose a method which sums all individual home vectors and normalizes the result to compute the final home vector. The method can be improved by weighting all associated home vectors of the histogram entries by the histogram entries themselves. The weighted home vectors are added and normalized

$$\mathbf{u} = \sum_{i=0}^{N-1} h(i) \cdot \mathbf{u}_i \quad (5.14)$$

$$\hat{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|} \quad (5.15)$$

where  $h(i)$  is  $i$ -th the histogram entry,  $\mathbf{u}_i$  is the home vector associated to this histogram entry, and  $N$  is the number of bins in the histogram.

This method expects the angular errors to be uncorrelated and that the democracy effect (Vardy and Möller, 2005) takes care of the outliers. For the example in figure 5.3, the result of this method is  $-44^\circ$  which is also close to the correct angle of  $-45^\circ$ .

#### 5.1.4 Tests

##### The Effect of Image Size and Butterworth Cutoff Frequency

The aim of the first test is to check whether the size of the input images affects the quality of the homing. To test whether the applied Butterworth filter also influences the quality, the cutoff frequency was varied from 0.03 to 1.0 where 1.0 depicts unfiltered images. A cutoff frequency below 0.03 was rejected because SIFT homing was not able to finish its computations. The step size from 0.03 to 0.1 cutoff frequency was 0.01. For cutoff frequencies larger than 0.1 the step size was 0.1.

The  $AAE_*$  and  $RR_*$  for the **original** database were computed; 169 snapshot and current view pairings ( $10 \times 17 - 1$ ) were considered. The final home vector was calculated by the average vector as presented in section 5.1.3. The results are visualized in figure 5.4. Furthermore, it was checked if a histogram equalization improves the quality of the homing. For this test, the original database was used. Schatz (2006) performed a parameter optimization, which shows that the best performance appears with a search window with the size 20% – 25% of the image width. It is also mentioned, that the best performance does not make use of orientation coded SIFT features.

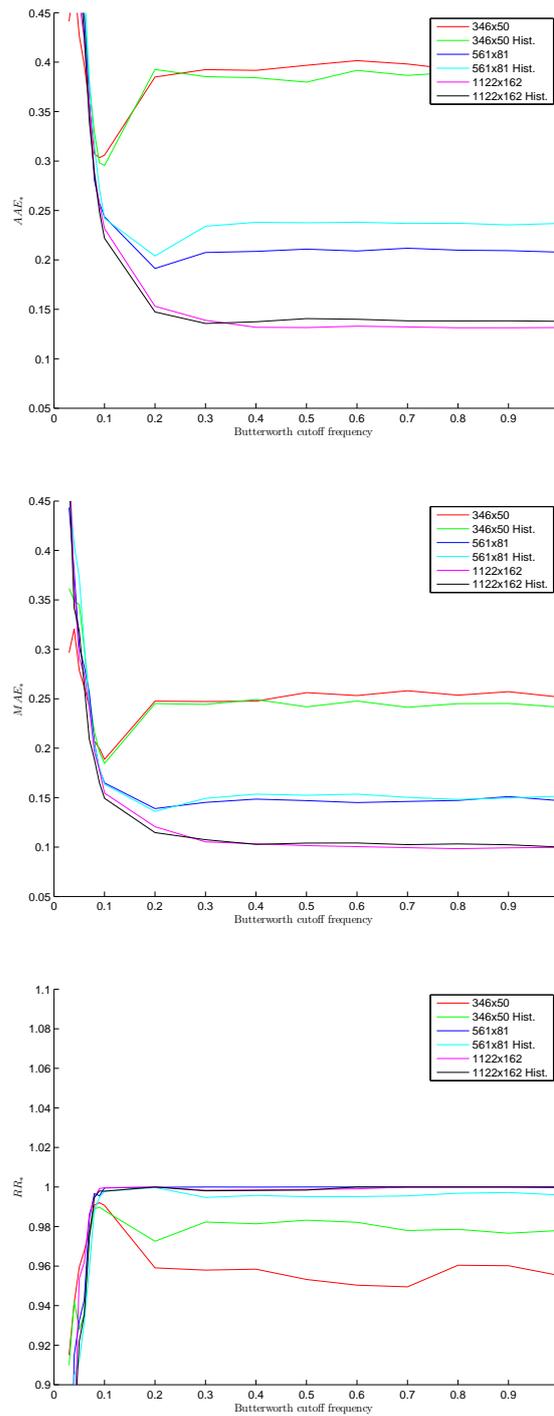


Figure 5.4:  $AAE_*$ ,  $MAE_*$ , and  $RR_*$  for different Butterworth cutoff frequencies diversified from 0.03 to 1.0, where the step size is 0.01 for cutoff frequencies smaller than 0.1; otherwise 0.1. The size of the input images was  $346 \times 50$ ,  $561 \times 81$ , and  $1122 \times 162$ . Histogram equalization was used to improve the intensity values of the input images. The input images belong to the **original** database.

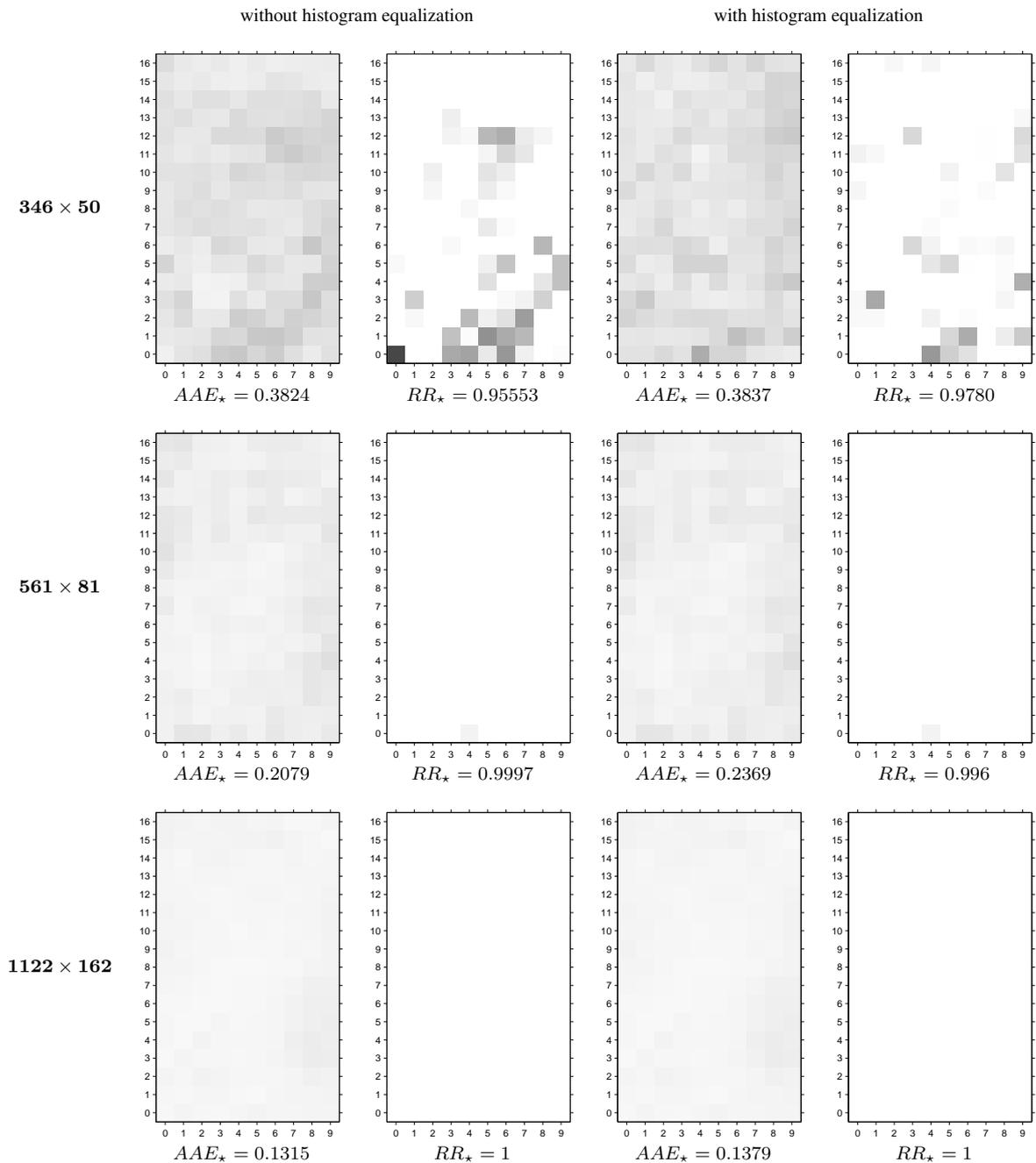


Figure 5.5:  $AAE_*$  and  $RR_*$  for  $346 \times 50$ ,  $561 \times 81$ , and  $1122 \times 162$  image size. The measures on the left side of the figure are computed without histogram equalization, the right with histogram equalization. For all computed measures the Butterworth cutoff frequency was 1.0, the SIFT parameters are oriented, and the search window size was fixed to 250 Pixel. The measures were calculated for the **original** database.

### Comparison to 2D-Warping

The second test verifies the sift homing quality for different image databases. Figure 5.4 shows that the  $AAE_*$  does not markedly increase if the image is filtered with decreasing Butterworth cutoff frequencies. For this test, the databases **original**, **hall1**, **moeller1**, **roeben1** and **uni** were used. The Butterworth cutoff frequency for SIFT homing is 1.0, thus the images were not modified. It was shown by Schatz (2006), that the home vector which is calculated by the average vector (see section 5.1.3) performs as good as the maximum approach. Hence, only the average vector approach is used for the tests. For comparison a 2D-warping method is used. Min-warping is an extended version of the original 2D-warping that uses a Butterworth filter of third grade (Möller, 2009; Möller et al., 2010).

Table 5.1 and table 5.2 describe the parameter set for the min-warping method. The original input images are Butterworth filtered with a 0.1 cutoff frequency; afterwards, the images are unfolded to the size  $346 \times 50$ .

search steps	search range	resolution
$n_\alpha$	$0^\circ \dots 355^\circ$	$\Delta\alpha = 5^\circ$
$n_\psi$	$0^\circ \dots 355^\circ$	$\Delta\psi = 5^\circ$
$n_\rho$	$0 \dots 0.95$	$\Delta\rho = 0.05$

Table 5.1: Number of steps, search range, and resolution used for the three warp parameters  $\alpha, \psi, \rho$  used in the min-warping method.

$\sigma_k$	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
$\tau_k$	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	2.0

Table 5.2: The scale factors  $\sigma_k$  and the thresholds  $\tau_k$  used for the min-warping.

	<b>original</b>	<b>hall1</b>	<b>moeller1</b>	<b>roeben1</b>	<b>uni</b>
$AAE_*$	0.095	0.2112	0.5509	0.3422	0.2899
$MAE_*$	0.0741	0.1601	0.3617	0.1261	0.1891
$RR_*$	1.0	0.8961	0.6267	0.6494	0.6240

Table 5.3: Results SIFT homing.

	<b>original</b>	<b>hall1</b>	<b>moeller1</b>	<b>roeben1</b>	<b>uni</b>
$AAE_*$	0.2663	0.1092	0.6800	0.4848	0.4844
$MAE_*$	0.0316	0.0623	0.3271	0.1030	0.1934
$RR_*$	0.9053	0.9912	0.6636	0.6112	0.6749

Table 5.4: Results 2D-warping homing.

The tables 5.3 and 5.4 show the computed quality measures for the SIFT homing and the 2D-warping method. For the **original** database, the return ratio of SIFT homing is 1.0. The 2D-warping method only has an return ratio of 0.9052. For all other databases, the warping performance is slightly better. The homing results for 2D-warping presented in table 5.4 differ from the results presented by (Möller et al., 2010).

### Cross Database Test

To show, how well a local homing method copes with different illumination situations, a cross database test is used, where the snapshot image is from a different database than the current view image. Figure 5.6 shows the results of the cross database tests, The measures  $AAE_*$ ,  $MAE_*$  and the  $RR_*$  are presented.

### 5.1.5 Discussion

Figure 5.4 shows that applying a lowpass filter to the input images only increases the quality of the SIFT homing method if the image size is small ( $346 \times 50$ ). If the Butterworth cutoff frequency is below 0.1, the error increases again, the return ratio for all tested image sizes drops below 98%. With image size  $346 \times 50$ , the return ratio only reaches 98%. If the input images are larger than  $561 \times 81$ , the return ratio for the **original** database is 100%. Since the SIFT keypoints are computed within the scale-space, it is not surprising that large images yield a better homing performance.

The tables B.1, B.2, B.3, and B.4, taken from Schatz (2006), show the results for the average vector approach and the maximum approach for the databases **original**, **uni**, **hall1**, and **hall2**. The results for  $AAE_*$  and  $RR_*$  slightly differ due to a different preprocessing of the input images. It can be seen that the number of SIFT keys increases the larger the images are. Furthermore, the  $AAE_*$  increases the smaller the input images are.

Figure 5.5 shows the quality measures for the different snapshot positions in the **original** database. Again, it can be seen that the return ratio increases if the input image size increases; the average angular error decreases. The best  $AAE_*$  is computed if the input image size is  $1122 \times 162$ . Furthermore, it can be

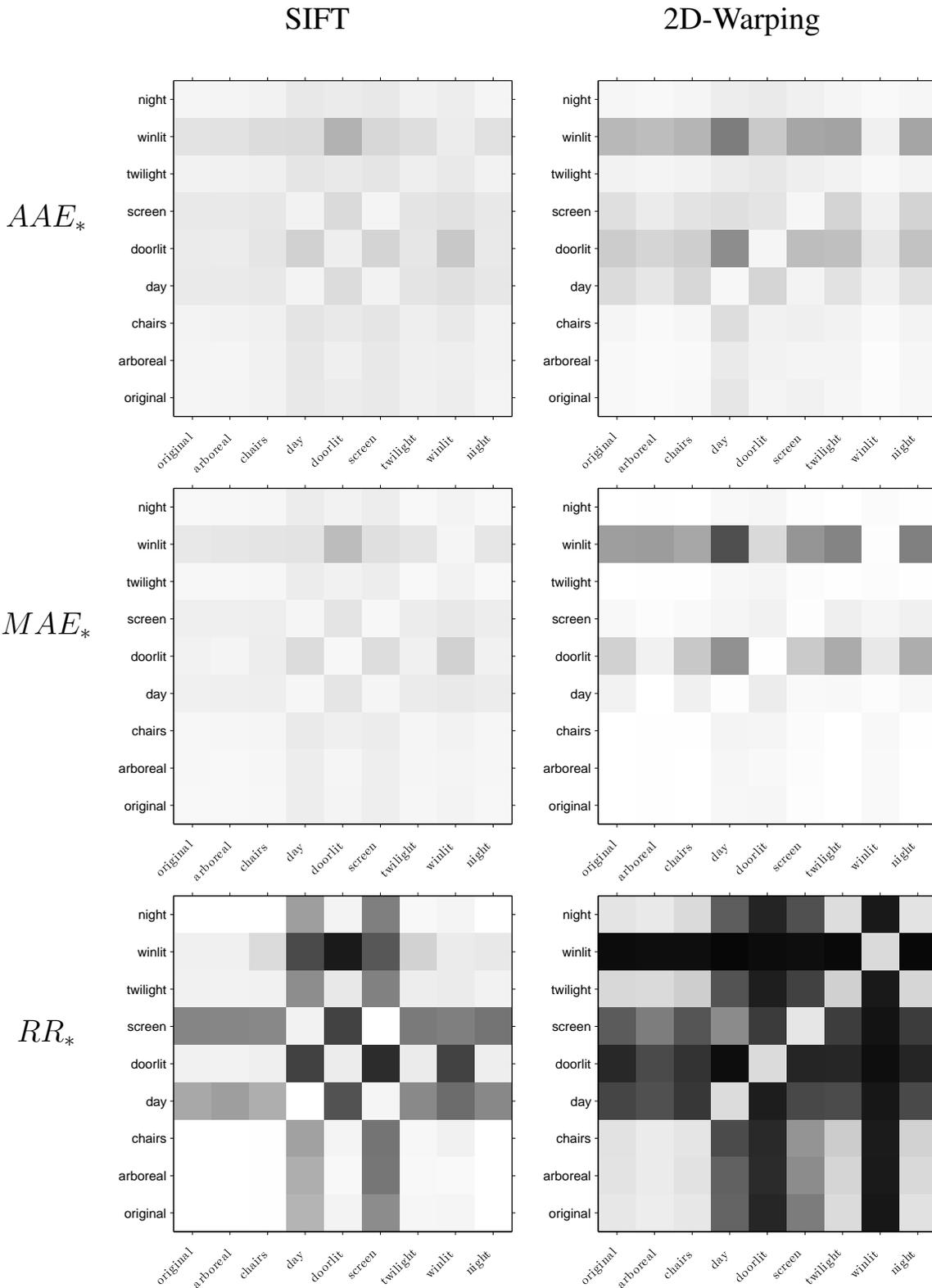


Figure 5.6: Results of the cross database tests. The ordinate of the gray value plots depicts the snapshot database, the abscissa represents the current view database. Light colors depict good results, dark colors depict bad results. The results are normalized for visualization.

seen that if the pixel intensities of the input images are corrected by the use of a histogram equalization, the average angular error increases. Thus, a histogram equalization does not make sense for SIFT homing. Again, the reason for that can be explained by the theory that SIFT features are based on.

The quality of a local visual homing method can only be verified if the method performs good for more than one environment. For that purpose, the SIFT homing method was compared to 2D-warping. The tables 5.3 and 5.4 show the results. It can be seen that the performance of two local visual homing methods is close to each other. For the **original** database, the return ratio is less worse than the SIFT homing return ratio. The return ratio of the **hall1** database is at 89% if SIFT homing is used, respectively 99% if 2D-warping is used. The large dark environment seems to have a positive effect on the 2D-warping results. The return ratio performance of SIFT homing is at 90%. The absence of light flooded areas leads to fewer detectable features. The warping performance of the **moeller1**, **roeben1**, and **uni** database is slightly better than the performance of SIFT homing.

The cross database test is a good indication how well a visual local homing method copes with illumination changes. Figure 5.6 show the results of the cross database tests. It can be seen that 2D-warping has problems if the illumination changes dramatically. The database variations **winit** and **doorlit** produce a bad return ratio. Since 2D-warping is based upon comparisons of pixel intensities, it is not remarkable that 2D-warping fails. The SIFT homing has no problems with the changing illumination. The detected features seem to be stable enough to cope with the **winit** and **doorlit** illumination situation. Surprisingly, both methods perform good on the databases **arboreal** and **chairs** but worse on the **screen** database. The only explanation or this behavior could be given by the color of the objects. The plant and the chairs are dark colored whereas the screen is white. The dark objects inconspicuously fit in the environment; the screen is conspicuous.

## 5.2 Parameter Models

Visual homing with parameter models uses image parameters which are extracted from the images. Instead of the original captured images, the parameters are stored and represent the images. This section is mainly taken from the diploma project of Fischer (2006) which was supervised by the author.

A parameter set of the image  $I$  with the size  $L$  is defined by

$$\mathbf{f}(I) = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_L \end{pmatrix} \quad (5.16)$$

To perform visual homing with image parameters, a feature parameter set for current view and snapshot is computed. The easiest way to get from current view to snapshot position is to calculate a potential function  $p(\mathbf{x})$  with respect to the Euclid distance of the parameter sets

$$p(\mathbf{x}) = \sum_{i=1}^L (f_i(S) - f_i(C(\mathbf{x})))^2. \quad (5.17)$$

Figure 5.2 shows that it is possible to reach the snapshot position by using a gradient descent. The home vector is given by the negative gradient of the potential function

$$h(\mathbf{x}) = -\nabla_{\mathbf{x}}p(\mathbf{x}). \quad (5.18)$$

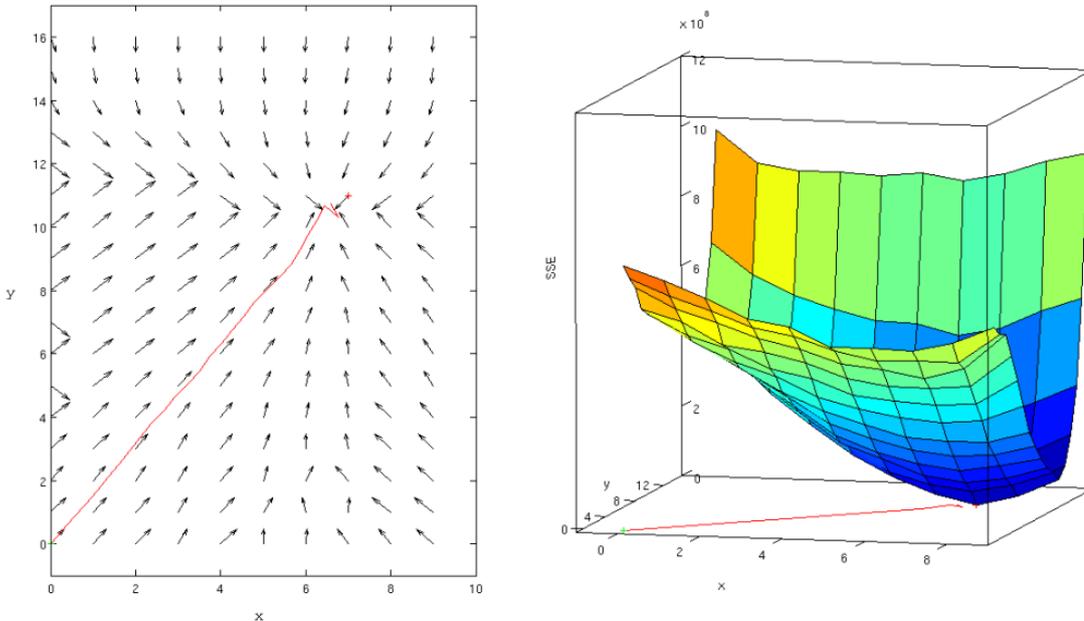


Figure 5.7: Example home vector array and the corresponding potential function. The home vector array is derived from the potential function, hence the home vectors approximate the negative gradient of the potential function (Fischer, 2006).

The gradient is not derivable at the local position, thus an approximation of two neighbouring positions and their corresponding potential functions has to be computed

$$\tilde{h}(\mathbf{x}) = - \left( \begin{array}{c} p([x + \Delta x, y]^T) - p([x, y]^T) \\ p([x, y + \Delta y]^T) - p([x, y]^T) \end{array} \right). \quad (5.19)$$

One disadvantage appears when the approximation of the gradient needs to be calculated. Explorative test steps need to be done in at least two directions to compute the gradient. To solve this problem, Möller and Vardy (2006) pointed out that it is possible to avoid test steps by predicting image changes that correspond to movements of the agent. Thus, the gradient can be computed from the predicted images. For small movements, the gradient is computable by using “flow-templates” (Möller and Vardy, 2006) and the intensity of the image pixels. It was shown that the “matched-filter method” can be used for any navigation method that uses parameter signatures. Since the rotation invariant matched-filter method for parameter homing (appendix C.2) was already described in Möller and Vardy (2006), it is necessary to derive a matched-filter method version for parameter homing that is not rotation invariant, which is carried out in detail in the appendix C.1.

### 5.2.1 Tests

The homing tests have been carried out on the databases **original**, **hall1** and **uni**. A large variety of parameter combinations was tested, hence more than 15000 tests had to be simulated (figure 5.8). The parameter variation was based on image size, preprocessing and the number of rings respectively the number of Fourier coefficients. The preprocessing was applied to the original input image; afterwards the image was unfolded. The following parameters have been tested in detail:

- **image size:**  $150 \times 25$  (**25**),  $300 \times 50$  (**50**),  $600 \times 100$  (**100**)
- **preprocessing:** none (**none**), Butterworth 1 with a 0.03 cutoff frequency (**bw1**), Butterworth 2 with a 0.015 cutoff frequency (**bw2**), Butterworth 1 with edge detector (**bw1e**), Butterworth 2 with edge detector (**bw2e**), histogram equalization (**he**)
- **number of**
  - **rings R** (statistical parameters): **1, 5, 10, 15, 20, 25**
  - **Fourier coefficients K** (Fourier-based): **1-20, 25, 30, 35, 40**

The bold term enclosed with round brackets is the identifier which is used in the result tables. A test calculates  $RR_*$  and  $AAE_*$  as mentioned in section 3.2. The results are summarized in table 5.5, 5.6 and 5.7.

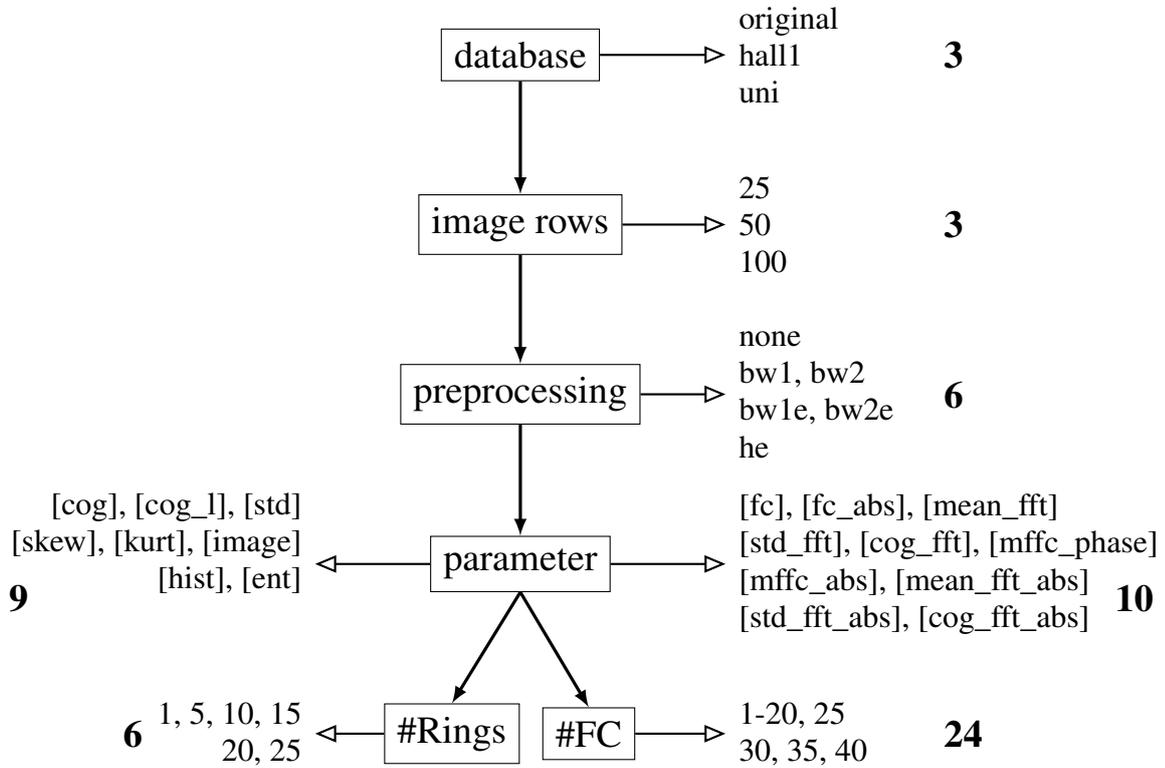


Figure 5.8: Schematic drawing of the design of the tests. The various combinations of the test parameters are depicted. The parameter sets based on Fourier coefficients are presented on the right side of the figure and those based on rings are presented on the left side. The bold numbers reflect the quantity of parameter sets obtained by varying the given parameter. The overall quantity of tests is beyond 15000. The figure is adapted from Fischer (2006).

**Results original Database**

Rank	Parameter	RR*	AAE*	#Rings	#FC	Image size	Pre-process.	rot.-invar.
1	[image]	0.9483	0.4626	-	-	25	bw1	n
2	[fc]	0.9460	0.4437	-	4	25	he	n
3	[mffc_phase]	0.9120	0.4403	-	7	25	bw2e	n
4	[mean_fft]	0.8405	0.5966	-	8	100	he	n
5	[cog]	0.8195	0.6935	20	-	25	he	n
6	[std_fft]	0.8108	0.6497	-	7	25	bw2	n
7	[mffc_abs]	0.7748	0.7763	-	5	25	bw2e	j
8	[cog_fft]	0.7473	0.7650	-	5	25	bw2	n
9	[fc_abs]	0.6611	0.8033	-	7	100	none	j
10	[mean]	0.5411	0.8343	25	-	50	none	j
11	[std_fft_abs]	0.5161	0.9967	-	4	100	bw2	j
12	[hist]	0.5102	0.9740	20	-	25	bw2	j
13	[std]	0.4574	0.9970	20	-	25	bw2	j
14	[cog_fft_abs]	0.4201	1.2405	-	8	25	bw2	j
15	[cog_l]	0.4131	1.1323	20	-	25	he	j
16	[mean_fft_abs]	0.4062	1.1046	-	10	25	bw2e	j
17	[skew]	0.3565	1.2477	20	-	100	bw2e	j
18	[kurt]	0.3197	1.0319	25	-	100	none	j
19	[ent]	0.2994	1.3158	25	-	50	bw2e	j

Table 5.5: Ranking of the different parameter homing methods that have been implemented. The results are computed for the **original** database. The quality of the return ratio assesses the ranking order. The table is adapted from Fischer (2006).

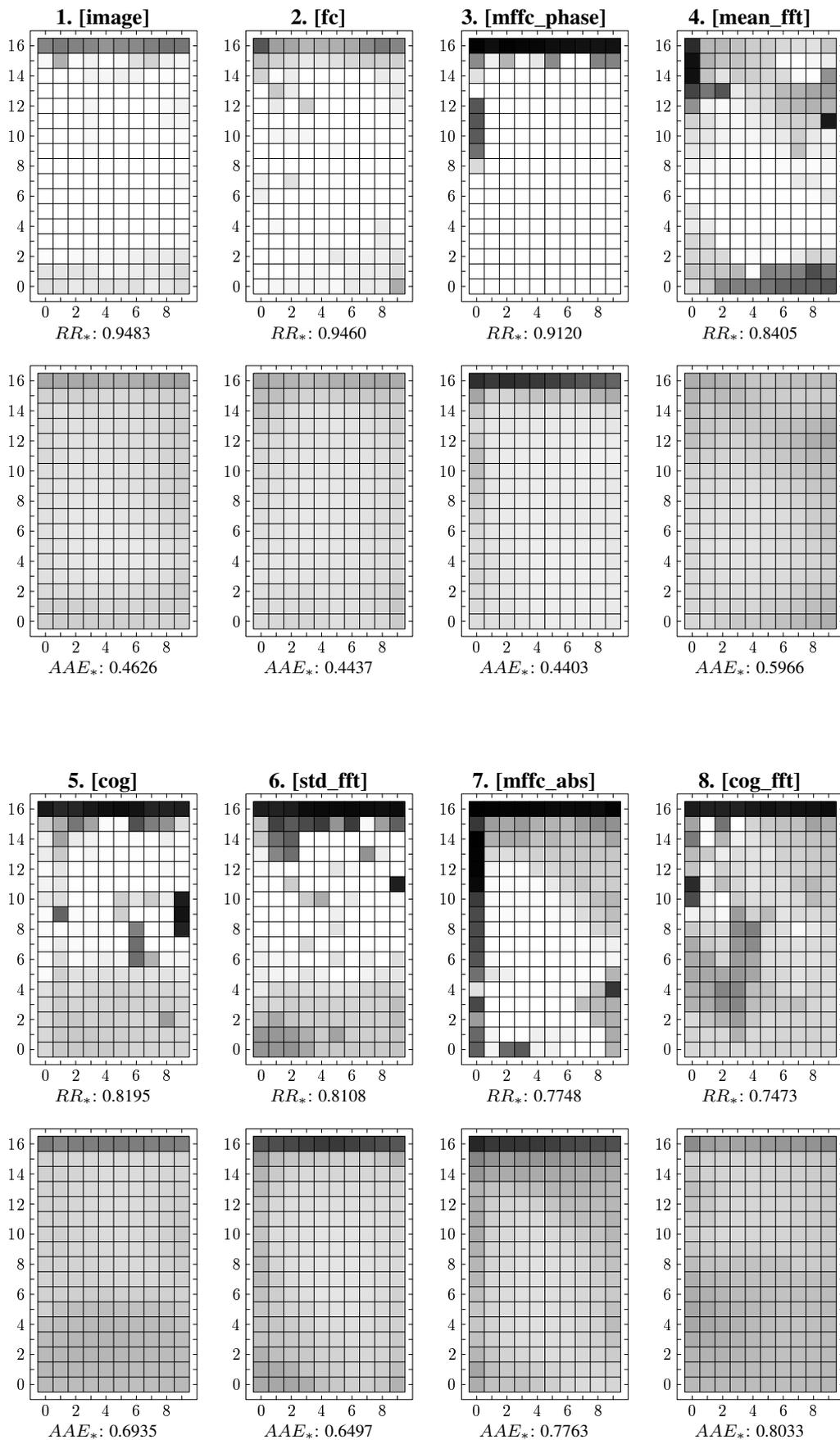


Figure 5.9:  $RR$  and  $AAE$  values for the eighth best parameter homing methods according to the **original** database, adapted from Fischer (2006).

**Results hall1 Database**

Rank	Parameter	RR <sub>*</sub>	AAE <sub>*</sub>	#Rings	#FC	Image size	Pre-process.	rot.-invar.
1	[mffc_phase]	0.9988	0.2420	-	9	50	he	n
2	[mffc_abs]	0.8793	0.4956	-	10	25	bw1e	j
3	[fc]	0.8367	0.5468	-	9	100	bw1	n
4	[image]	0.7880	0.6446	-	-	25	bw2	n
5	[cog]	0.6792	0.7060	25	-	100	bw1e	n
6	[std_fft]	0.6216	0.8101	-	7	100	none	n
7	[mean_fft]	0.6075	0.7862	-	6	50	none	n
8	[cog_fft]	0.5857	0.8938	-	9	100	bw2	n
9	[fc_abs]	0.5511	0.7890	-	7	100	bw2	j
10	[hist]	0.4347	1.0096	25	-	100	none	j
11	[mean_fft_abs]	0.4271	0.8894	-	6	100	he	j
12	[std_fft_abs]	0.4157	0.9723	-	9	100	he	j
13	[mean]	0.3850	1.0238	25	-	100	none	j
14	[std]	0.3695	1.0255	20	-	100	he	j
15	[cog_l]	0.3592	1.3129	25	-	50	bw2	j
16	[skew]	0.2936	1.2260	25	-	50	bw1	j
17	[cog_fft_abs]	0.2902	1.3268	-	19	50	bw2e	j
18	[ent]	0.2689	1.1081	25	-	100	bw1	j
19	[kurt]	0.2598	1.2821	20	-	50	bw1	j

Table 5.6: Ranking result of the parameter homing methods for **hall1** database, adapted from Fischer (2006).

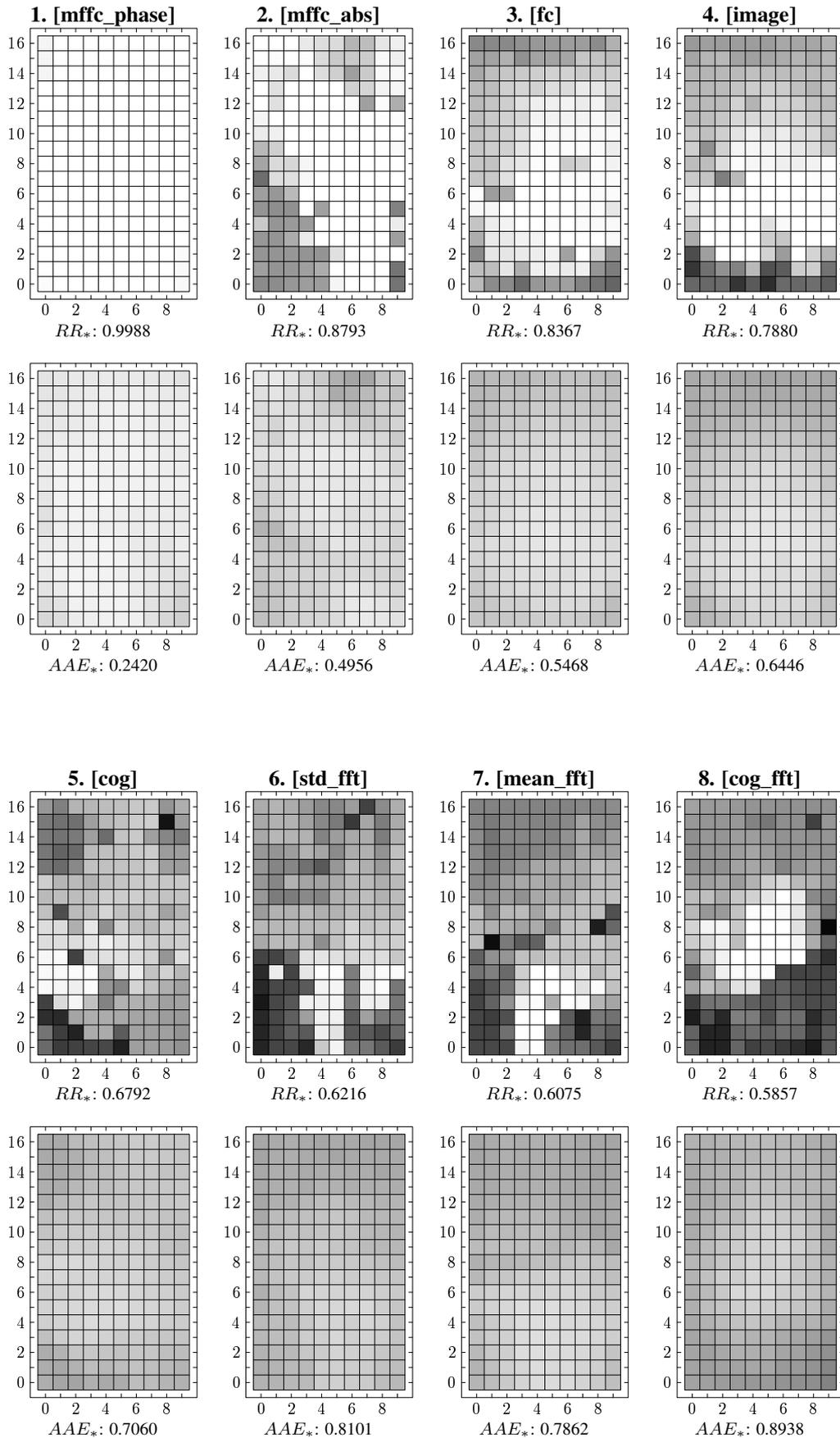


Figure 5.10:  $RR$  and  $AAE$  values of the eight best parameter homing methods for **hall1** database, adapted from Fischer (2006).

**Results uni Database**

Rank	Parameter	RR <sub>*</sub>	AAE <sub>*</sub>	#Rings	#FC	Image size	Pre-process.	rot.-invar.
1	[mffc_phase]	0.7989	0.5143	-	15	25	bw1e	n
2	[mffc_abs]	0.4852	0.8952	-	20	25	he	j
3	[fc]	0.4381	0.8564	-	16	100	bw1e	n
4	[image]	0.4251	0.8538	-	-	50	bw1e	n
5	[mean_fft]	0.3474	1.0714	-	18	50	he	n
6	[fc_abs]	0.3370	1.1633	-	40	50	he	j
7	[std_fft]	0.2858	1.1519	-	15	100	bw1e	n
8	[cog_fft]	0.2508	1.3073	-	35	100	bw2e	n
9	[cog]	0.2448	1.0606	25	-	25	none	n
10	[mean_fft_abs]	0.2222	1.2833	-	40	100	he	j
11	[kurt]	0.2141	1.2318	25	-	100	he	j
12	[std]	0.2040	1.2167	20	-	100	he	j
13	[mean]	0.1953	1.1981	25	-	25	bw1e	j
14	[std_fft_abs]	0.1843	1.3328	-	40	25	bw1e	j
15	[skew]	0.1763	1.2603	15	-	50	he	j
16	[hist]	0.1750	1.2461	25	-	100	he	j
17	[ent]	0.1704	1.2485	20	-	100	bw2	j
18	[cog_l]	0.1684	1.2235	25	-	25	none	j
19	[cog_fft_abs]	0.1664	1.3743	-	30	100	he	j

Table 5.7: Ranking result of the parameter homing methods for **uni** database, adapted from Fischer (2006).

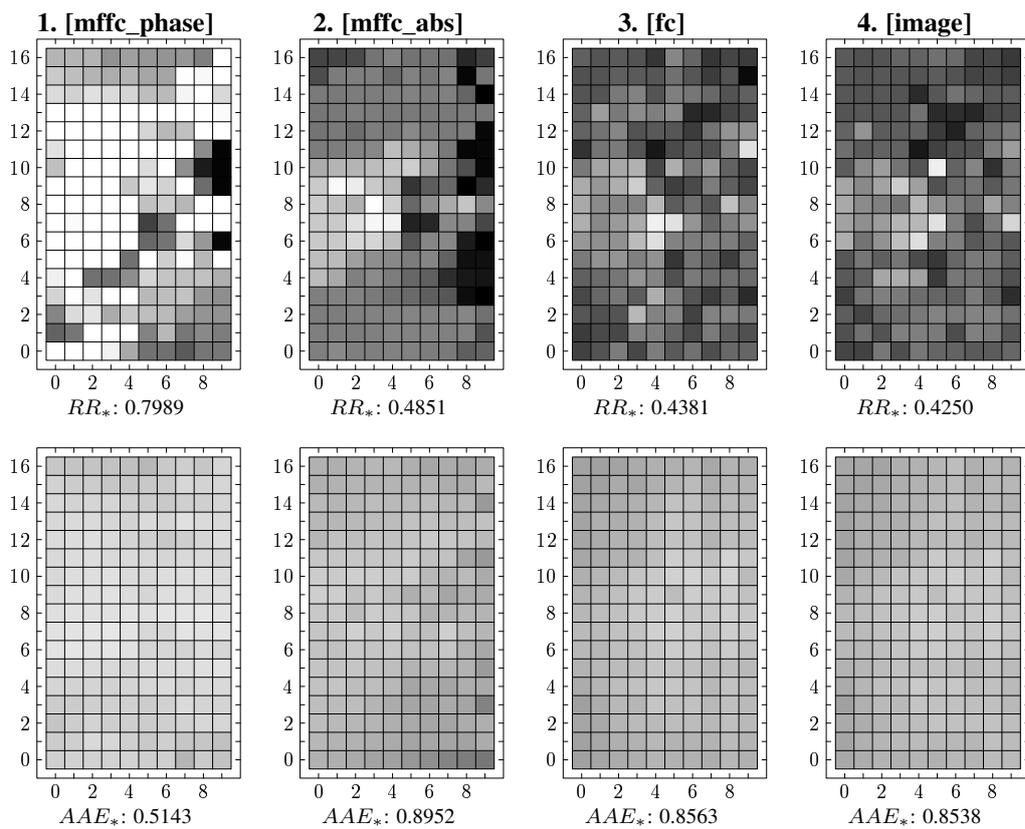


Figure 5.11:  $RR$  and  $AAE$  values of the fourth best parameter homing methods for **uni** database, adapted from Fischer (2006).

### 5.2.2 Discussion

Table 5.5 shows the homing results for the **original** database. The best performance ( $RR_* = 0.9483$ ,  $AAE_* = 0.4626$ ) is achieved by the **[image]** parameter set with a 0.03 Butterworth cutoff frequency. The parameter set is not invariant in rotation. The second best ( $RR_* = 0.9460$ ,  $AAE_* = 0.4437$ ) is the **[fc]** parameter set with four Fourier coefficients and a histogram equalization. The best parameter set **[mffc\_abs]** which is invariant in rotation is at rank seven ( $RR_* = 0.7748$ ,  $AAE_* = 0.7763$ ). All parameter features with a rank greater than eight are invariant in rotation. The purely statistical parameters skew, kurtosis, and entropy only attain return ratios less than 40%. The center of gravity **[cog]** is the only statistical parameter that achieves a return ratio of 0.8195, but it is not invariant in rotation. The rotation invariant pendant **[cog\_l]** attains only a return ratio of 41%. Figure 5.9 shows the  $RR$  and  $AAE$  values for the first eight best parameter homing results. Depicted by the black squares in the figure, it can be seen that all parameter sets have problems with images located at position  $(x,16)$  of the capture grip introduced in section 3.1. The reason for this might be the database. The snapshots at position  $(x,16)$  are located under a table which entirely changes the illumination of the scene.

The parameter homing results of the **hall1** database are presented in table 5.6. The best is **[mffc\_phase]** ( $RR_* = 0.9988$ ,  $AAE_* = 0.2420$ ), which is not invariant in rotation. Surprisingly, the second best is **[mffc\_abs]** ( $RR_* = 0.8793$ ,  $AAE_* = 0.4956$ ), which is invariant in rotation. Again, the rotation invariant parameters are located on the ranks greater than eight. To attain a good return ratio the number of rings should be as large as possible. The number of used Fourier coefficients should be between six and ten. Again, the kurtosis and the entropy achieve a return ratio less than 30%. Figure 5.10 shows that the used parameters have different problems with the database. It is noticeable that there is an accumulation of bad return ratios in the corners of the database grid. A possible explanation for such a behavior could be explained by anomalies in the database like disappearing lamps or hidden regions behind pillars.

Table 5.7 shows the results of the parameter-homing for the **uni** database. Again, the best performance gives **[mffc\_phase]** ( $RR_* = 0.7989$ ,  $AAE_* = 0.5143$ ) followed by **[mffc\_abs]** ( $RR_* = 0.4852$ ,  $AAE_* = 0.8952$ ). In contrast to the results for the **hall1** database, the number of used Fourier coefficients increased. The rotation invariant parameters are located on ranks greater than nine. Conspicuously, the parameters perform worse for this database than for **original** and **hall1**. The statistical parameters never attain a return ratio greater than 25%. Again, the number of rings and Fourier coefficients increased. As visualized in figure 5.11, the return ratios are worse. At some positions of the capture grid e.g.  $(10,10)$  or  $(5,7)$ , even the best parameter set fails to perform

successful homing. The reason for such considerably worse return ratios might be found by the normal view hyperbolic mirror (see section 3.1) or by the outdoor environment, which is not limited like an indoor environment. .

### 5.3 Summary and Conclusion

In this chapter, two local visual homing methods are presented; SIFT homing and homing with parameter models. Several tests are presented. It was tested if a lowpass filter leads to better homing performance by SIFT homing. It was shown, that images which are not filtered yield the best performance for images larger than  $1122 \times 162$ . Because of this result, the further navigation tests are carried out with the image size  $1122 \times 162$ . SIFT homing was compared to 2D-warping with respect to the **original**, **hall1**, **moeller1**, **roeben1**, and **uni** database. It was shown, that the performance of SIFT is comparable to 2D-warping. A cross database test was used to prove how well the homing methods cope with illumination changes. It can be seen that SIFT homing is more stable than 2D-warping. A major shortcoming was noticed. The computational speed of SIFT homing is not competitive with 2D-warping. Although, warping extensively uses image difference computations, it is quite evident that it is faster than SIFT homing. The database tests showed that SIFT homing is about three times slower. The small input image size of 2D-warping might be one reason for the speed-up.

Extensive tests have been carried out for the parameter models. The tests concentrate on the **original**, **hall1**, and **uni** databases. Rotation invariant parameters have been analysed as well as rotation variant ones. It can be seen that at most database tests the return ration of rotation invariant parameters get not better than 87%.

## Chapter 6

# Outdoor Navigation with Color Contrasts

A method for illumination invariant Outdoor navigation with UV-green contrast was described by Möller (2002b). A two-channel sensor, one channel sensitive to UV the other sensitive to green, is used to measure the UV and green light that is reflected by different objects or emitted by the sky. The measured amount of light is logarithmized and afterwards digitized. The UV-green contrast is computed by the difference of the digitized values. Because of the logarithmic presentation of the data, the UV-green contrast is not affected by the intensity of skylight or light reflected from objects.

This thesis focuses on contrasts that are computed from UV, blue, green, red, and IR. For this purpose, a five-channel hand-held sensor (figure 6.1) was built. The sensor was used to collect data from sky, natural and artificial objects. The collected data is used to learn a linear separation that separates foreground (objects) and background data (sky).

Additionally, the hand-held sensor, mounted on a pan-tilt unit, is used to collect panoramic data samples. The data samples are stored in a image structure; one image for every channel. The pixel that represents the corresponding data sample is coded by the the number of the pan-step ( $x$ ) and the tilt-step ( $y$ ). Afterwards, the images are subtracted to calculate contrast images. The learned separation is used to separate foreground and background data from the images. The result is used for local visual homing. Furthermore, it is checked how stable color contrasts are over time.

## 6.1 Five-Channel Hand-held Sensor

The five-channel hand-held sensor was built in the diploma project by Thomas Kollmeier (Kollmeier, 2006) which was supervised by the author. The results were published by Kollmeier et al. (2007).



Figure 6.1: Five-channel hand-held sensor

The sensor is equipped with a microcontroller unit with an 8051 architecture and 1MBit static RAM. The microcontroller (ADuC812) drives six 12-bit analog-to-digital converters each measuring the logarithmized signal of one channel. Thus, the output of each photodiode is logarithmized by a Burr Brown (LOG112) log-ratio amplifier by feeding the log-ratio amplifiers with a reference signal and the photodiode outputs.

The five channels UV, blue, green, red, and infrared are each measured by a Hamamatsu photodiode (S5821, without lense). The spectral sensitivity of the Hamamatsu photodiode ranges from 300nm to 1100nm. Glass filter combinations are used to produce the desired spectral sensitivity curve in each channel; table 6.1 lists the filter combinations and figure 6.2 shows the spectral sensitivity of each channel. The glass filter combinations for UV and green are biologically motivated by the receptors of *Cataglyphis bicolor* (Briscoe and Chittka, 2001). The glass filters are mounted in front of each associated photodiode. To reduce the opening angle of each channel, a UV-grade plano-convex quartz glass lens (diameter 6.35mm, focal length 9.5 at 589nm wavelength) is mounted in front of the glass filter.

Channel	Filter Type (Schott) and Thickness	Peak	Bandwidth
UV	UG11 (3mm), BG40 (1mm)	350nm	50nm
B	BG12 (2mm),GG400 (1mm) BG28 (1mm)	440nm	80nm
G	BG7 (2mm),GG475 (2mm)	500nm	70nm
R	RG645 (1mm),KG5 (1mm) RG665 (1mm)	680nm	90nm
IR	RG780 (1mm),KG4 (3mm)	800nm	200nm

Table 6.1: Filter combination, sensitivity peak, and bandwidth (50% to 50% sensitivity) for all five sensor channels UV, B, G, R, IR; the sensitivity curves are obtained from the combined filter transmittance (measured data) and the spectral sensitivity of the photodiodes (Hamamatsu S5821, catalog data) (Kollmeier et al., 2007).

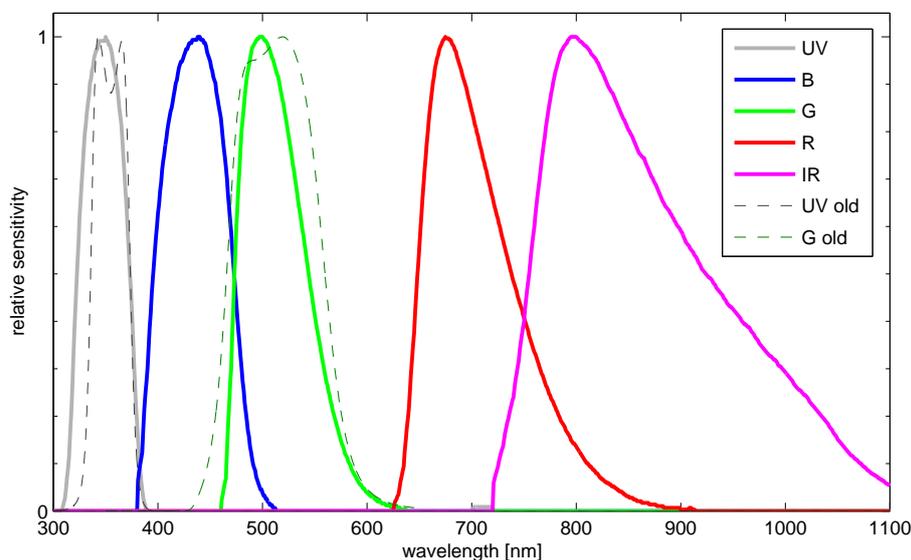


Figure 6.2: Spectral sensitivity (normalized to a maximum of 1) of all five spectral channels, obtained from combined filter transmittance (measured data) and photodiode sensitivity (catalog data). The dashed curves refer to the filters and sensors used by Möller (2002b).

## 6.2 Spectral Contrast Analysis for Natural Objects

### 6.2.1 Data Collection

The database contains 239 readings from natural objects and 233 readings from the sky. All data points were collected between May, 9th and June, 20th, 2006 in the vicinity of Bielefeld, Germany. The data were collected by pointing the sensor towards the sample region. 10 subsequent samples were taken in short succession (each  $507\mu s$ ). The mean of the subsequent samples constitutes one data point. The standard deviation was used to analyze the noise of the sensor. The distance to objects was typically between 5m and 30m.

Object samples include grass, shrubs, trees, tar, dead leaves, and dead wood as well as minerals like rocks, gravel, sand, and soil. The objects were sampled from different views and different angles; especially the angle to the sun was varied, and the sensor position varied from shadow to sun-flooded places. Sky samples cover blue sky and numerous different types of clouds (e.g. high-level to low-level clouds, and clouds of vertical development). The weather conditions were ranging from warm, cloudless days to cold days with closed cloud cover and drizzle.

### 6.2.2 Discrimination

The collected data samples are discriminated by the Linear Discriminant Analysis mentioned in section 3.3.3. Figure 6.3 depicts the biologically motivated UV-green contrast and the UV-IR contrast which is the dual-channel contrast producing the largest values in the discriminant criteria  $J$  and  $J_\sigma$  (table 6.2). The green points represent data samples of foreground objects; the red ones represent background objects. The solid blue line delineates the linear separation obtained by maximizing  $J$ ; the dashed line specifies the linear separation obtained by maximizing  $J_\sigma$  and is an indication for the effect of Gaussian receptor noise added to each channel. The larger the value of the discriminant criterion is, the better is the separation of the classes, and thus, the separation error (misclassified data points) should be small.

	UV	B	G	R	IR	All
$J$	4.36	3.35	2.37	0.67	0.11	8.99
$J_\sigma$	4.26	3.27	2.32	0.66	0.11	6.95
$n_f$ [%o]	34	47	76	248	379	15

J	B	G	R	IR	$J_\sigma$	B	G	R	IR	$n_f$	B	G	R	IR
UV	5.15	5.54	6.39	7.30	UV	4.51	4.97	5.92	6.82	UV	44	42	28	15
B		5.37	6.03	6.67	B		3.90	5.34	6.09	B		36	17	8
G			5.38	5.97	G			4.34	5.25	G			21	15
R				5.26	R				2.68	R				55

Table 6.2: Discriminant criterion  $J$  from equation 3.16 and  $J_\sigma$  from equation 3.22 (with  $\sigma = 30$ ), and the frequency of misclassified data points  $n_f$  (in %o) for different contrasts. **Top:** Single-channel contrasts and five-channel contrasts (“All”). **Bottom:** Dual-Channel contrasts.

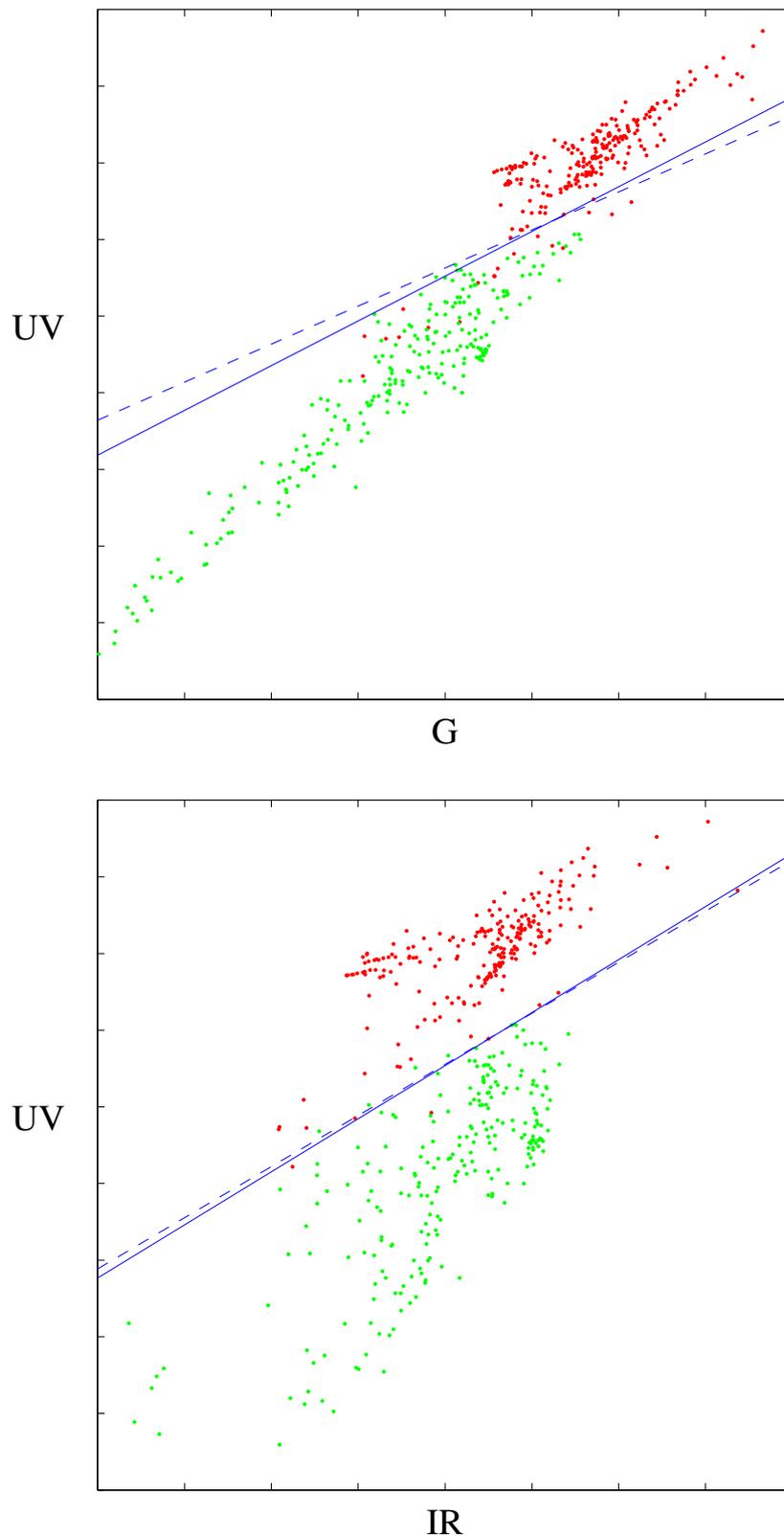


Figure 6.3: UV-G and UV-IR data, green depicts foreground samples, red depicts background samples. The solid blue line is the optimal separation  $J(\mathbf{w})$ , the dashed blue line is  $J_\sigma(\mathbf{w})$ .

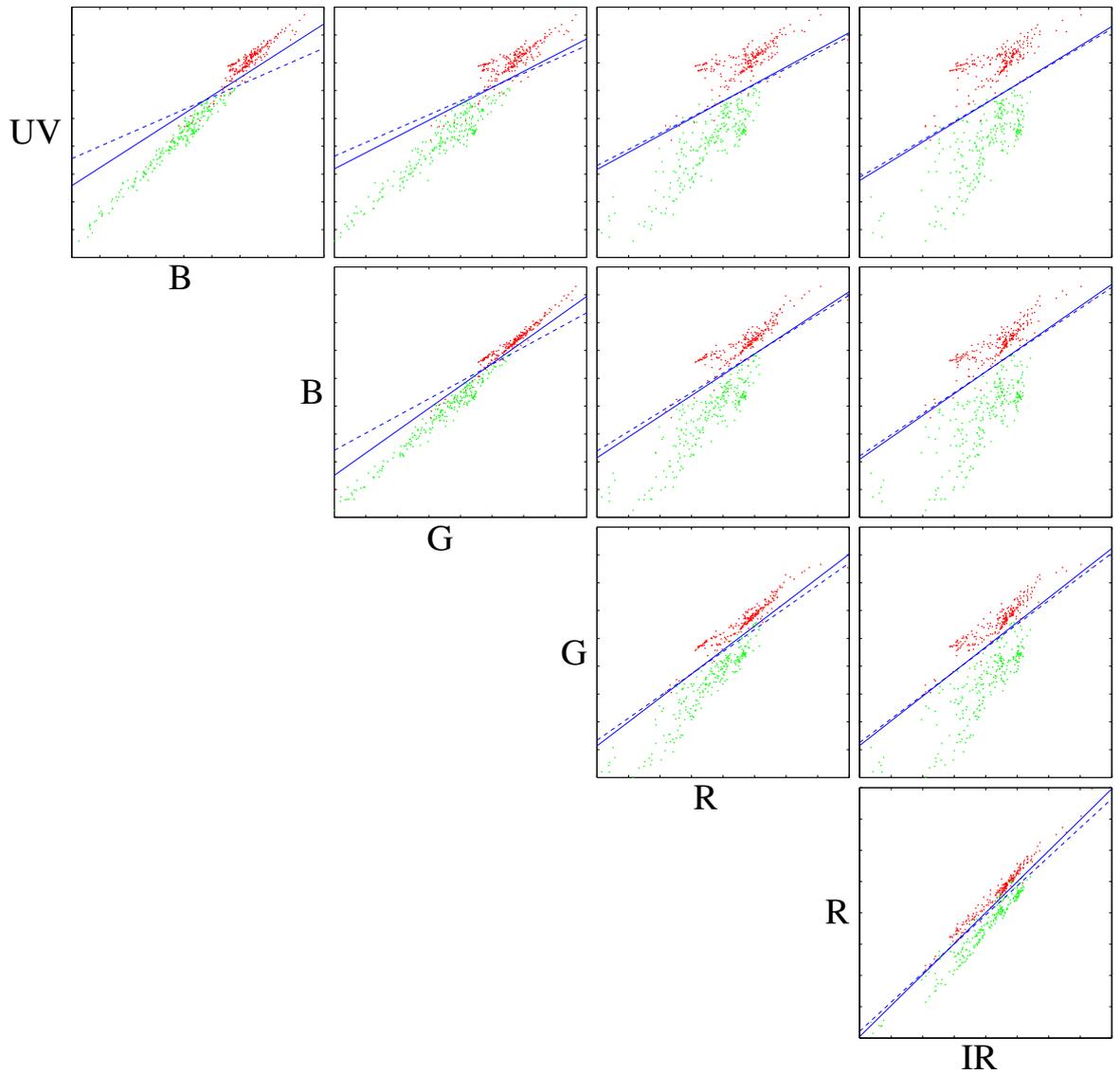


Figure 6.4: Plotted data for natural objects. The color coding is analog to figure 6.3.

### 6.2.3 Statistical Tests

Different combinations of spectral channels were analyzed with respect to the discriminant measures  $J(\mathbf{w})$ ,  $J_\sigma(\mathbf{w})$ , and to the relative frequency of misclassifications  $n_f$ , which were presented in table 6.2. The results are statistical estimates. To test the statistical hypotheses that the estimates differ significantly between different contrasts, bootstrap (see section 3.3.2) is used. The confidence interval for hypothesis testing can be determined from the percentiles of the resulting bootstrap distribution. A bootstrap distribution based on 10000 bootstrap data sets was generated. The endpoints of the confidence intervals were varied from  $\alpha = 5\%$ ,  $1\%$  to  $0.1\%$ .

### 6.2.4 Discussion

Figure 6.3 and figure 6.4 depict the results of the discrimination. It can be seen that the sky samples and the foreground samples are well separated. Table 6.2 shows that the best separation is achieved by spectral contrasts with a large distance between the wavelength of the two channels. UV-IR separates best and produces the largest value for  $J$  and  $J_\sigma$ . The misclassified data points are the smallest for the B-IR contrast. The results of table 6.2 are also visualized in figure 6.4. The larger the distances of the wavelengths, the more the data is scattered. Interestingly, the data points are separable even if the distance is very small. For example, the data samples from the R-IR contrast have a small distance in wavelength but are separable. The biologically motivated UV-G contrast is only at rank six in  $J$  and  $J_\sigma$ .

The results of the statistical tests for natural objects are presented in table 6.2.3. UV-IR and B-IR, the two contrasts on the diagonal in figure 6.4 below UV-IR, are the next best two contrasts. According to the bootstrap results, they do not significantly differ from each other. They also do not differ from the R-IR contrast, when additional noise is considered in the  $J_\sigma$  criterion. In the  $J_\sigma$  criterion, the performance of R-IR drops drastically and is not competitive with UV-R. The contrasts from the second diagonal, UV-G, B-R, and G-IR, also do not differ significantly from each other in  $J$  and  $J_\sigma$ . Furthermore, the lowest diagonal, the UV-B, B-G, G-R, and R-IR contrasts, are not significantly different in  $J$ . The single channel classification performs worse compared to the dual-channel contrasts. The best single channel classification is achieved by the UV. The single channels do not differ significantly from each other in  $J$  and  $J_\sigma$ .

The single channels have the lowest quality in separation  $J$  without noise. In the  $J_\sigma$  criterion, the discrimination quality of the contrasts B-G and R-IR drop below that of the UV single channel. Until the second diagonal, UV-G, B-R, and



G-IR, of the dual-channel contrasts of table 6.2.3, the rank of the contrasts does not change if noise is added. UV-B and R-IR are sensitive to added noise; their  $J_\sigma$  drops below that of UV. The misclassified data is smallest for B-IR, followed by UV-IR and G-IR. Again, the UV single channel is the best single channel and beats the dual channel contrasts B-G, UV-G, UV-B, and R-IR. The five channel contrast “all” beats the single channels and the dual-channel contrasts in  $J$  and  $J_\sigma$ . The frequency of misclassified data is equal to UV-IR and G-IR but larger than B-IR.

For technical applications, the UV-IR, B-IR or UV-R contrasts will yield the best results. The biologically motivated UV-G contrasts does not reach the performance of the superior contrasts, but even with noise ( $J_\sigma$ ), the separation quality is still better than the best single channel, UV. The two non-UV contrasts B-R and G-IR are not significantly better than UV-G in  $J$  and  $J_\sigma$ . The separation quality of the single channels is best for UV. G, R and IR are not acceptable in their separation quality.

## 6.3 Spectral Contrast Analysis for Artificial Objects

### 6.3.1 Data Collection

The database contains 233 readings from sky, which were already used in the natural data collection setup, and 124 readings from artificial objects. The artificial data points were collected between January 14th and May 10th, 2007 and also between May 10th and May 15th, 2008 in the vicinity of Bielefeld, Germany. The data was collected in the same manner as in section 6.2.1, but this time with smaller distances to the objects, typically between 50cm and 2m. Object samples include plastic, lacquered metal, and painted surfaces in different views and different angles.

### 6.3.2 Discrimination

Again, the Linear Discriminant Analysis is used for discrimination. The method which is used in section 6.2.2 also was used for the artificial objects.

### 6.3.3 Statistical Tests

The statistical tests are realized analog to section 6.2.3.

	UV	B	G	R	IR	All
$J$	4.14	1.56	1.0	0.40	0.30	6.12
$J_\sigma$	3.99	1.52	0.98	0.39	0.30	5.31
$n_f$ [%o]	60	120	183	291	312	47

J	B	G	R	IR	$J_\sigma$	B	G	R	IR	$n_f$	B	G	R	IR
UV	5.08	5.12	5.22	5.65	UV	4.58	4.71	4.87	5.22	UV	53	57	55	50
B		2.26	2.04	2.23	B		1.76	1.87	2.04	B		107	115	105
G			1.39	1.51	G			1.22	1.34	G			150	133
R				0.52	R				0.43	R				252

Table 6.4: Discriminant criteria  $J$  from equation 3.16 and  $J_\sigma$  from equation 3.22 (with  $\sigma = 30$ ), and the frequency of misclassified data points  $n_f$  (in %) for different contrasts. Top: Single-channel contrasts and five-channel contrasts (“All”). Bottom: Dual-Channel contrasts.

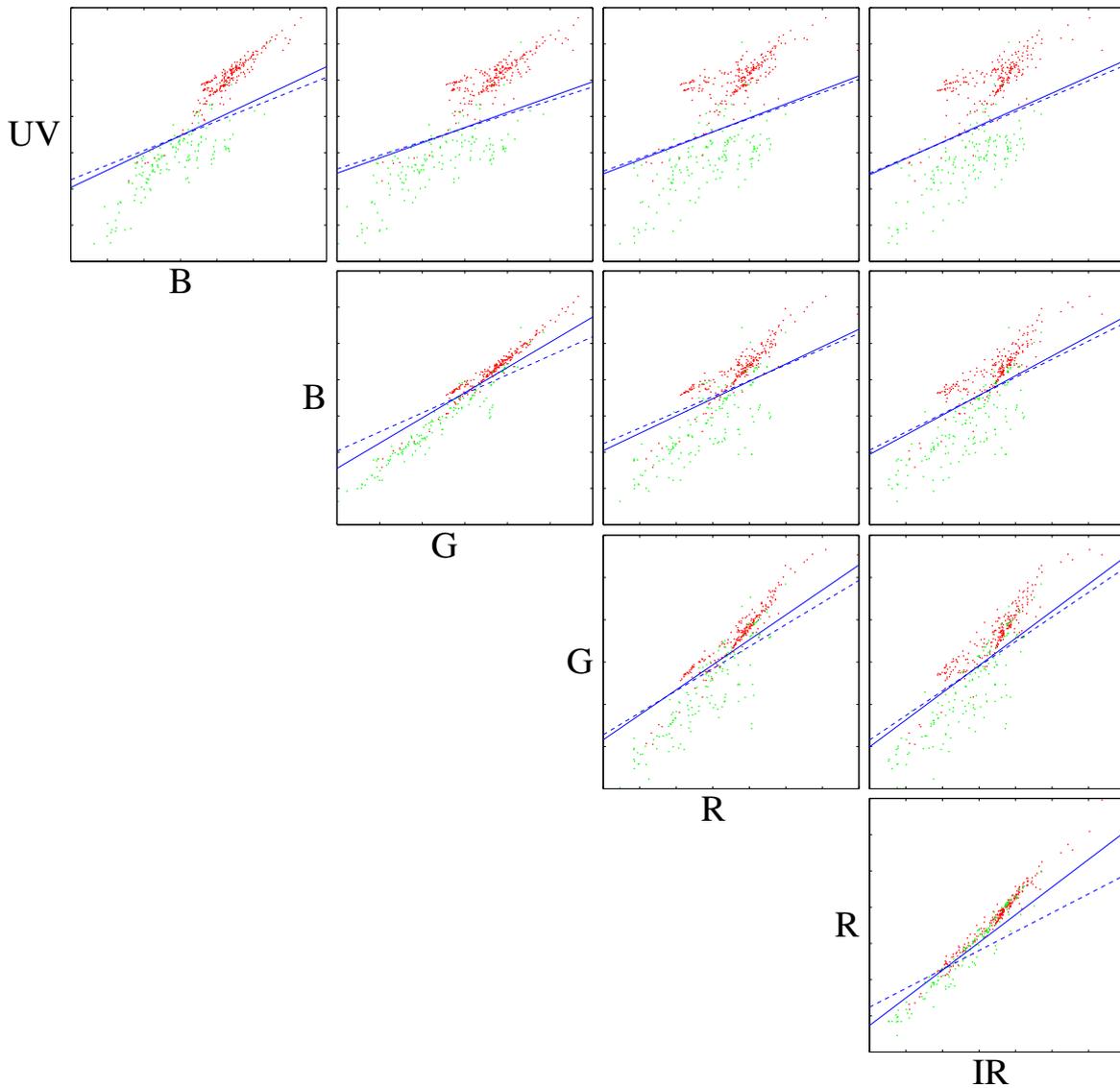


Figure 6.5: Plotted data for artificial objects. The color coding is analog to figure 6.3. The separation criteria  $J$  and  $J_\sigma$  that are depicted by the blue lines have been recomputed.

All	○	All	○	All	○
UV-IR	***	UV-IR	***	UV-IR	***
UV-R	***	UV-R	***	UV-R	***
UV-G	***	UV-G	***	UV-G	***
UV-B	***	UV-B	***	UV-B	***
UV	***	UV	***	UV	***
B-G	***	B-G	***	B-G	***
B-IR	***	B-IR	***	B-IR	***
B-R	***	B-R	***	B-R	***
B	***	B	***	B	***
G-IR	***	G-IR	***	G-IR	***
G-R	***	G-R	***	G-R	***
G	***	G	***	G	***
R-IR	***	R-IR	***	R-IR	***
R	***	R	***	R	***
IR	***	IR	○	IR	○

Table 6.5: Significances for  $J$ ,  $J_\sigma$ , and  $n_f$  (\*\*\* 0.1%, \*\* 1%, \* 5%) for artificial objects.

### 6.3.4 Discussion

Table 6.4 shows the results from the discrimination. It can be seen that the UV-contrasts perform the best in all three categories ( $J$ ,  $J_\sigma$ ,  $n_f$ ). Looking at the single channels, the UV-channel is the best by far. The ranking of the natural contrasts qualities was ordered in diagonal direction from top left to bottom right in table 6.2. The artificial objects are ordered from top to bottom (see table 6.4).

The UV-IR and UV-B do not differ significantly from each other in the discriminant criterion  $J$ . Except for the UV-B contrast, UV-IR differs significantly from every other contrast. The UV single channel is at rank five, after the UV-contrasts. The UV-contrasts also perform best for the  $J_\sigma$  criterion. UV-G and UV-B do not differ significantly if noise is added.

When looking at the frequency of misclassified data, UV-IR is best. It can be seen that the UV-contrasts are not significantly different from each other. Again, the UV single channel is at rank five. Furthermore, the B-contrasts do not differ significantly. As in the natural object discrimination in section 6.2.4, the UV-contrasts yields the best results. For the single channels only the UV channel is competitive with the two-channel contrasts.

## 6.4 2D- Scene Contrast Analysis



Figure 6.6: 2-D pan-tilt sensor, a combination of a Directed Perception pan-tilt unit (PTU-46-17.5) and the five-channel handheld sensor.

The data acquisition sensor from section 6.1 was extended with a Directed Perception pan-tilt unit (PTU-46-17.5). The position resolution of the encoders is  $0.01^\circ$ . Because the pan-tilt unit has a payload of approximately 0.45kg, the sensor had to be mounted on the rotation axis of the tilt motor. As depicted in figure 6.6, the sensor was mounted on top of the pan-tilt unit with a small offset. The entire unit was mounted on a tripod. Rotating around the two axes of the pan-tilt unit yields a systematic gridlike sampling of the surrounding environment.



Figure 6.7: Stitched wide view images of the two test scenes, captured with a Canon 20D DSLR camera. (a) shows the image of dataset one, (b) the image of dataset two. Due to a small aperture angle of the DSLR, in (b) the image height is smaller than in the results obtained from the spectral sensor. The sky of the scene is not present in the DSLR image.

## 6.4.1 Data Collection

### Two Example Datasets

The data collection of the two-dimensional datasets was done on different days. The data acquisition system was placed in different locations. Two example locations are shown in figure 6.7.

The first location is between a summer house and a garage. The summer house is located on the right side of the scene, the garage on the left side. The sensor stood on the pavement between the garage and the summer house. The center of the image points towards the north.

The second location was in a backyard. The plants of the garden were close to the sensor (approximately 2m). In the north of the sensor location was a summer house as seen in figure 6.7 (b). The center of the reference image points towards the northeast. The aperture of the DSLR is smaller than the tilt angle of the data acquisition system. Thus, the reference image shown in figure 6.7 (b) is cut off at the top. This is no problem since the image is only presented for better understanding of what can be seen in the raw data images.

Both datasets were captured on October 24th, 2008. The weather was normal, the ceiling was covered with white clouds, and there was no rain. The first dataset was captured at 11:20 am, the second at 11:50 am. The example datasets contain sample points from 113 pan and 26 tilt angles. The step size in pan direction was  $2^\circ$ , in tilt direction  $3^\circ$ ; the scan order is tilt-pan. Thus, the data is collected column by column. Capturing one dataset takes 20 minutes. The sensor was externally triggered. The measured data was not averaged; if the hand-held sensor would capture with averaging, the capture time would increase by factor 10.

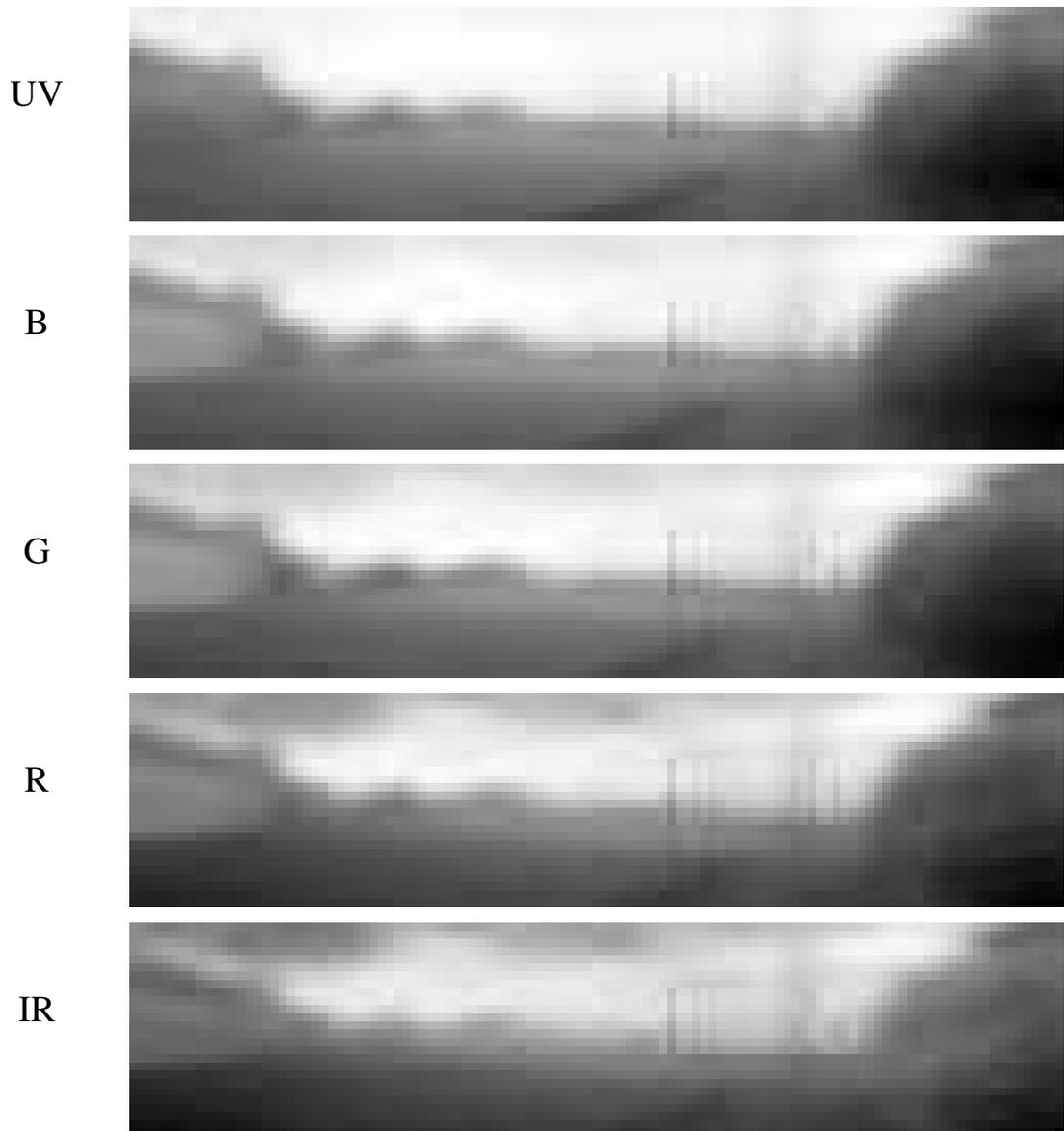


Figure 6.8: Raw image data of panorama dataset one. The measured data is normalized for visualization and shown as gray value image. The images are ordered by their spectral frequencies, UV, B, G, R, and IR. The light gray bars at position (66,y) of the images are no errors. The sensor was pointed towards a street, where cars passed by during the capture.

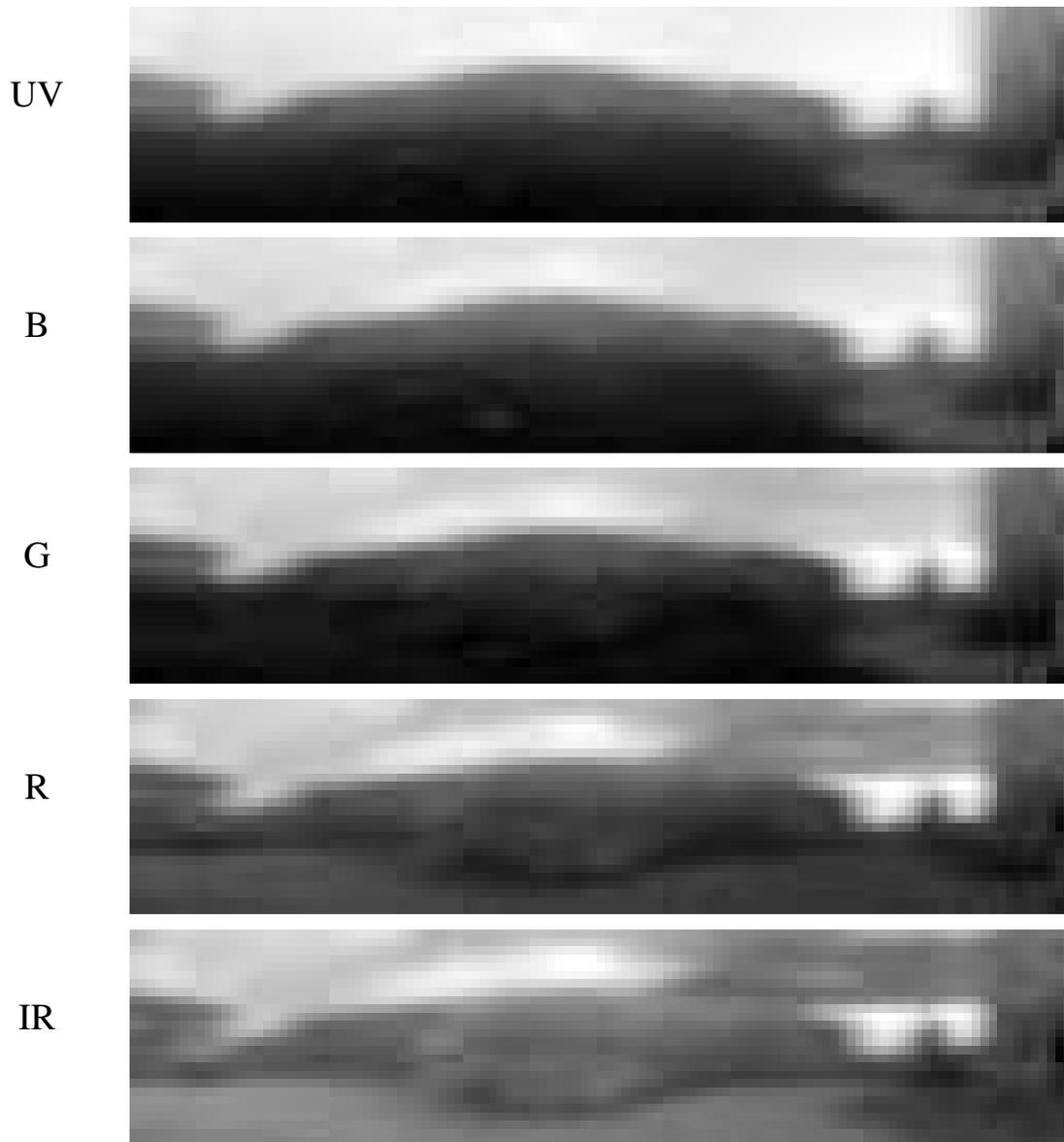


Figure 6.9: Raw image data of panorama dataset two. The measured data is normalized for visualization and shown as gray value image. The images are ordered by their spectral frequencies, UV, B, G, R, and IR.

Figure 6.8 and figure 6.9 show the raw values for the five channels for both datasets. The datasets were normalized for visualization and shown as gray value images. The analysis and skyline computation is done on the original data. The intensity of a pixel depicts the measured spectral intensity. A measured high spectral intensity relates to bright pixels, low intensity correspond to dark pixels. At position (66,y), light gray vertical bars appear. The reason for this vertical bars are cars that drove past the data acquisition system during the measurement.

### Data Collection over one Day

Additionally to the two example datasets, data were collected over more than 9 hours from the same position. The data acquisition sensor was again placed at the backyard. The location is identically to the dataset two from section 6.4.1. The pan resolution was set to one degree and covered  $200^\circ$ , the tilt resolution was set to  $2^\circ$  and covered  $102^\circ$ . Due to the sample resolution, the sample time of one dataset is 40 minutes. The scene was sampled fourteen times and covers the time from 11:27am to 20:43pm. The sunset was close to 22:00pm. The weather was variable and the look of the scene changed markedly. Heavy thick clouds moved over the sky when the data was collected. Due to the large size of the data set, it is presented in the appendix D.3 and is referred to as “one-day dataset”. This data collection is used for a stability analysis of the contrasts.

## 6.4.2 Contrast Stability over Time

To show that the contrasts are stable over time, different measures are computed. The first dataset (11:27am) which was captured is taken as reference input. Seven measures are computed for every time-dataset captured after the reference dataset. For single channel, the single channel of the reference is used; for two-channel contrasts, the contrasts of the reference are used. The contrasts are computed by subtracting the measured logarithmic values. Every single channel, and every two-channel combination is computed.

The measures for two inputs ( $A$  and  $B$ ), where the input is either a single channel or a contrast, are computed as follows:

$$MSE(A, B) = \text{mean}((A - B)^2) \quad (6.1)$$

$$RMSD(A, B) = \sqrt{MSE(A, B)} \quad (6.2)$$

$$\log NRMSD(A, B) = \log_{10} \frac{RMSD(A, B)}{\max(A) - \min(A)} \quad (6.3)$$

$$PSNR(A, B) = 10 \cdot \log_{10} \left( \frac{\max(A)^2}{MSE(A, B)} \right) \tag{6.4}$$

$$CorrCoeff(A, B) = \frac{\frac{1}{N} \sum (A_{ij} - \bar{A})(B_{ij} - \bar{B})}{\left( \frac{1}{N} \sum (A_{ij} - \bar{A})^2 \right) \left( \frac{1}{N} \sum (B_{ij} - \bar{B})^2 \right)} \tag{6.5}$$

$$\logMSEstd(A, B) = \log_{10} \frac{MSE(A, B)}{\sqrt{VAR(A)VAR(B)}} \tag{6.6}$$

$$\logMSEvar(A, B) = \log_{10} \frac{MSE(A, B)}{VAR(A)} \tag{6.7}$$

Only the logarithmized normalized root mean squared difference (logNRMSD), the logMSEstd, and the logMSEvar are visualized in figure 6.10, 6.12, and 6.11. For the sake of completeness, the results for the other measures are shown in appendix D.3.

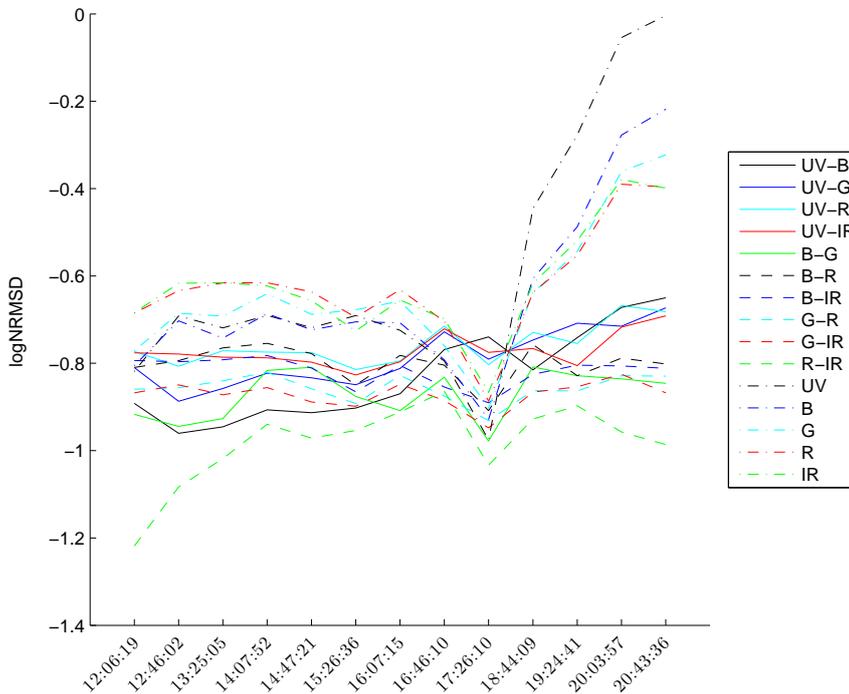


Figure 6.10: The measure logNRMSD as defined by equation 6.3.

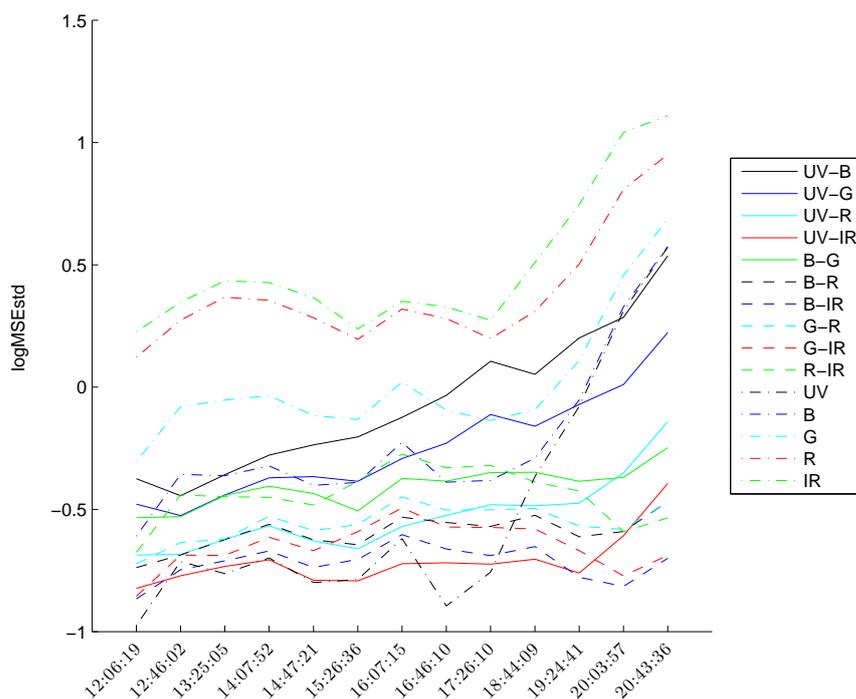


Figure 6.11: The measure  $\log\text{MSEstd}$  as defined by equation 6.6.

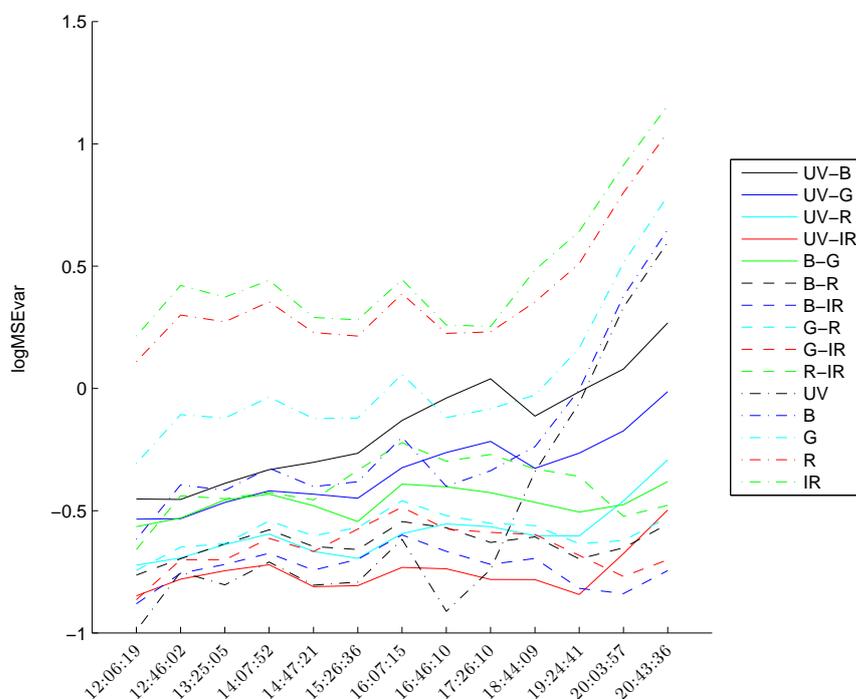


Figure 6.12: The measure and  $\log\text{MSEvar}$  as defined by equation 6.7.

### 6.4.3 Contrast Stability over Time after Skyline Separation

To compute the skyline, the foreground and background data samples need to be separated. In section 6.2, a discriminant analysis of natural and artificial objects is presented. The sample points were used to learn a linear separation line between foreground and background samples leading to a skyline of terrestrial objects (foreground) in front of the sky (background). The discriminant criterion  $J_\sigma$  of the natural objects (see section 6.2) was used for separation. Again, the measures which are presented in section 6.4.2 are computed.

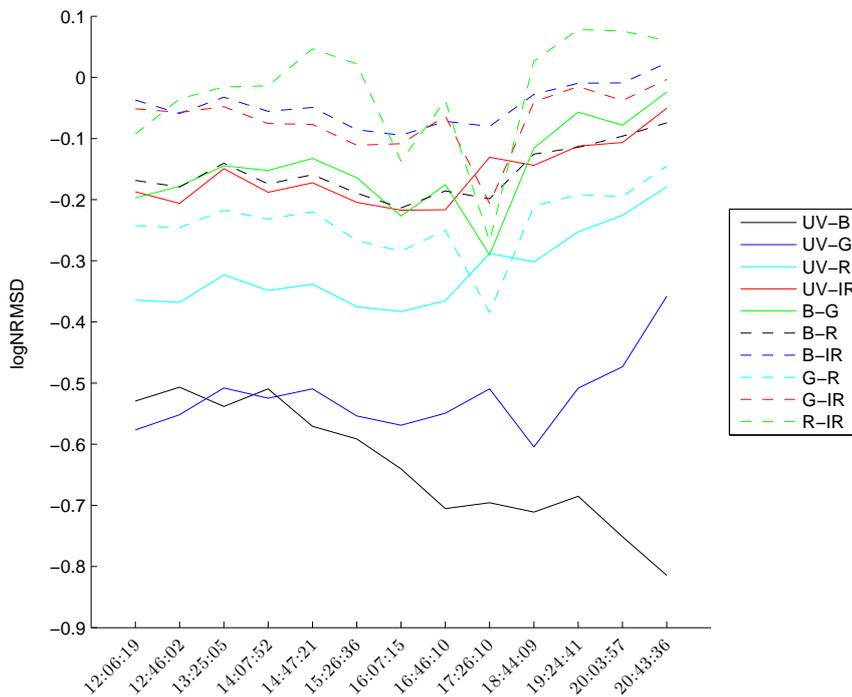


Figure 6.13: The measures  $\log\text{NRMSD}$  as defined by equation 6.3 for skyline separated contrasts.

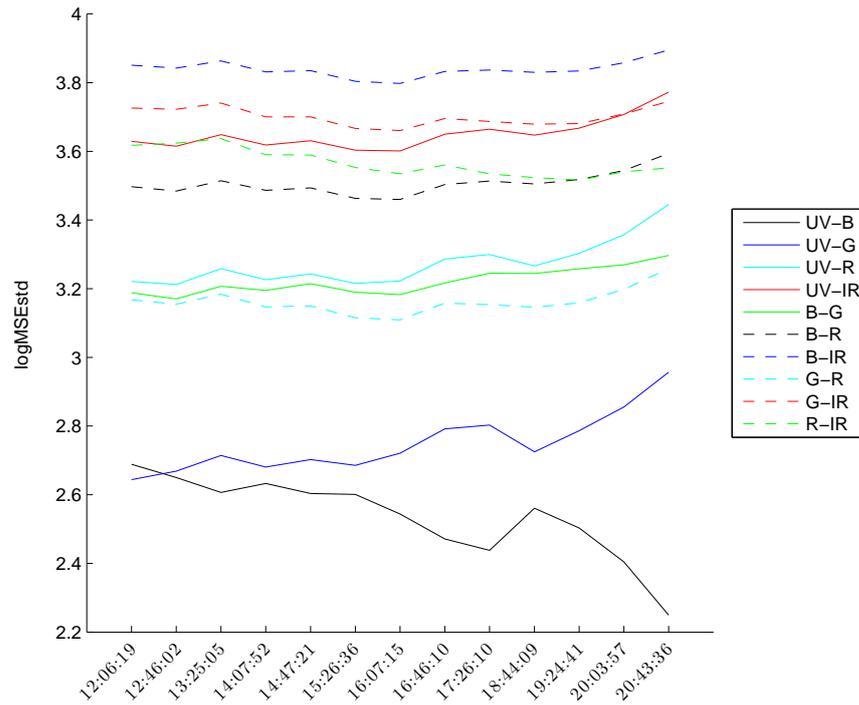


Figure 6.14: The measures  $\log\text{MSEstd}$  as defined by equation 6.6 for skyline separated contrasts.

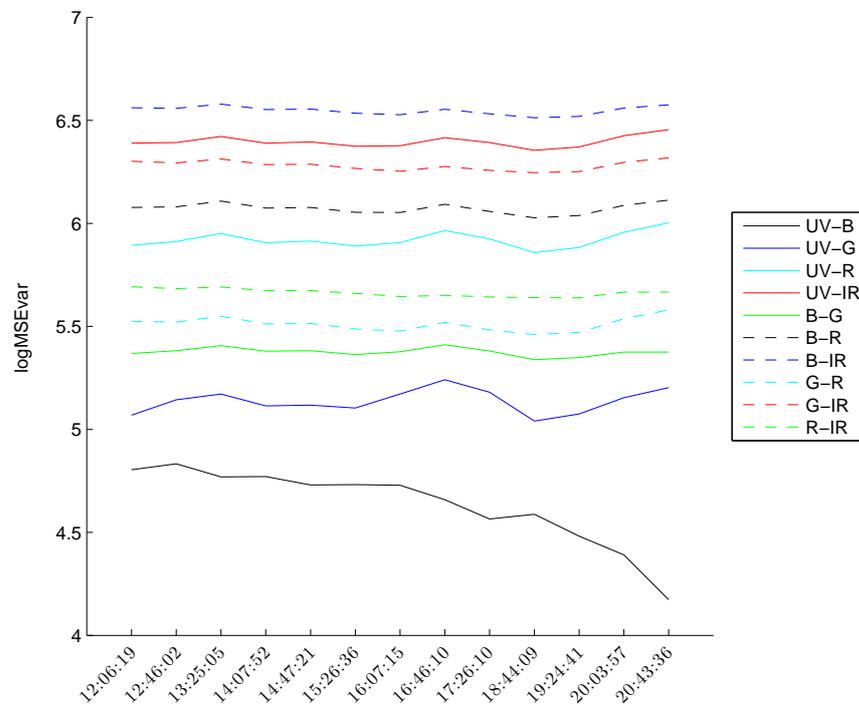


Figure 6.15: The measures  $\log\text{MSEvar}$  as defined by equation 6.7 for skyline separated contrasts.

### 6.4.4 Discussion

The figures 6.10, 6.12, and 6.11 show the computed image measures for the dual-channel contrasts and single spectral channels over a day. The logNMRSD measure shows that the single channels UV, B, G, R and IR produce a larger value than the two-channel contrasts. After 5 hours, the single channels increase markedly. The scene changed strongly, because the sky cleared off and the sun appeared in the evening. Obviously, the single channels are not invariant to illumination changes. The dual-channel contrasts perform nearly equal. At the beginning, the R-IR contrast has the smallest value but its increasing quickly to nearly the same values of all contrasts. The dual-channel contrasts are nearly constant over time, even when the sky cleared up in the evening. It is surprising that there is no difference between the UV-dual channel contrasts and that R-IR performs best. Assuming that this is due to the maximum and minimum used in this measure which may impair the comparability of different channels and contrasts. Therefore, the two alternative measures logMSEstd and logMSEvar which normalize the MSE to the variances of input channel A or to the standard deviation of both channel A and B, respectively, give a clearer picture. The logMSEstd and logMSEvar show that the contrast differences with the same first contrast are ordered by the difference of the wavelength. It can be seen that the variance of the contrast B does not have an influence on the measure since the plots of logMSEstd and logMSEvar are nearly equal. The standard deviation in the equation 6.6 only smooths the plotted values.

The figures 6.13, 6.14, and 6.15 show the computed image measures for the contrasts which have been separated by the  $J_\sigma$  criterion. The logMSEstd and logMSEvar measures show that all contrast combinations perform nearly constant over time. The UV, B, G and R contrasts are all ordered by their spectral difference. The order of the dual-channel UV contrasts changes. In figure 6.11 and 6.12 the order is UV-B, UV-G, UV-R, and UV-IR, which changes to UV-IR, UV-R, UV-G, and UV-B (figure 6.14 and 6.15).

## 6.5 Visual Homing with Skyline Information

### 6.5.1 Exemplary Skyline Detection

To detect the skyline, the foreground and background data samples need to be separated. In section 6.2, a discriminant analysis of natural and artificial objects is presented. The sample points were used to learn a linear separation line between foreground and background samples leading to a skyline of terrestrial objects (foreground) in front of the sky (background). The discriminant criterion  $J_\sigma$  of the natural objects (see section 6.2) was used for separation and the

skyline can be easily computed afterward. An exemplary skyline separation is presented for the dataset one and dataset two. Figure 6.16 and figure 6.17 show the separated data samples, green samples depict foreground samples and red depict sky samples. The separated images are presented from figure 6.18 to figure 6.25. The black pixels depict pixels classified as background (sky). The white pixels depict pixels classified as foreground (terrestrial objects).

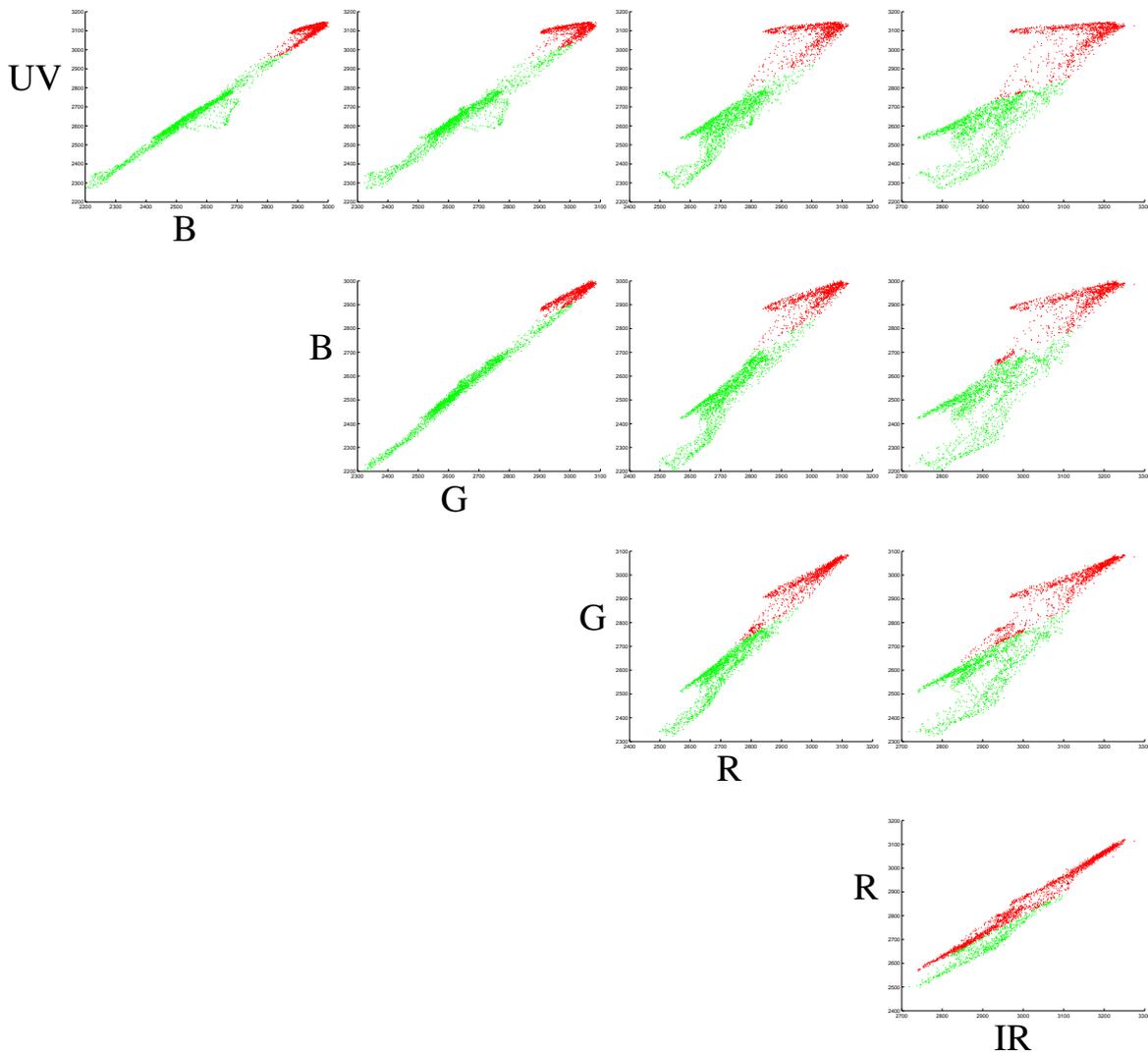


Figure 6.16: Separated contrasts for all pixels of dataset one. The red dots depict samples classified as sky (background). The green dots depict samples classified as terrestrial objects (foreground).

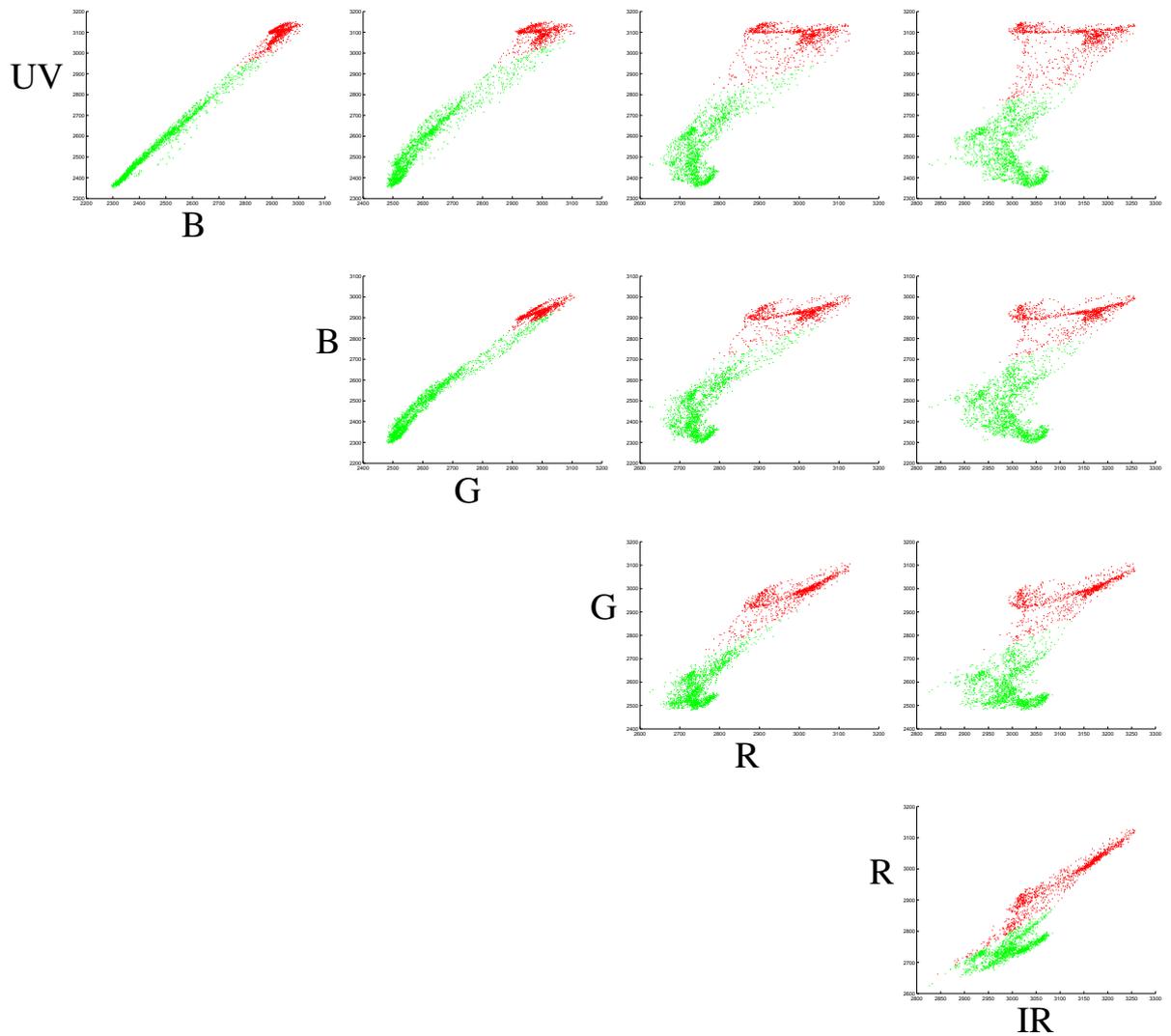


Figure 6.17: Separated contrasts for all pixels of dataset two. The red dots depict samples classified as sky (background). The green dots depict samples classified as terrestrial objects (foreground).

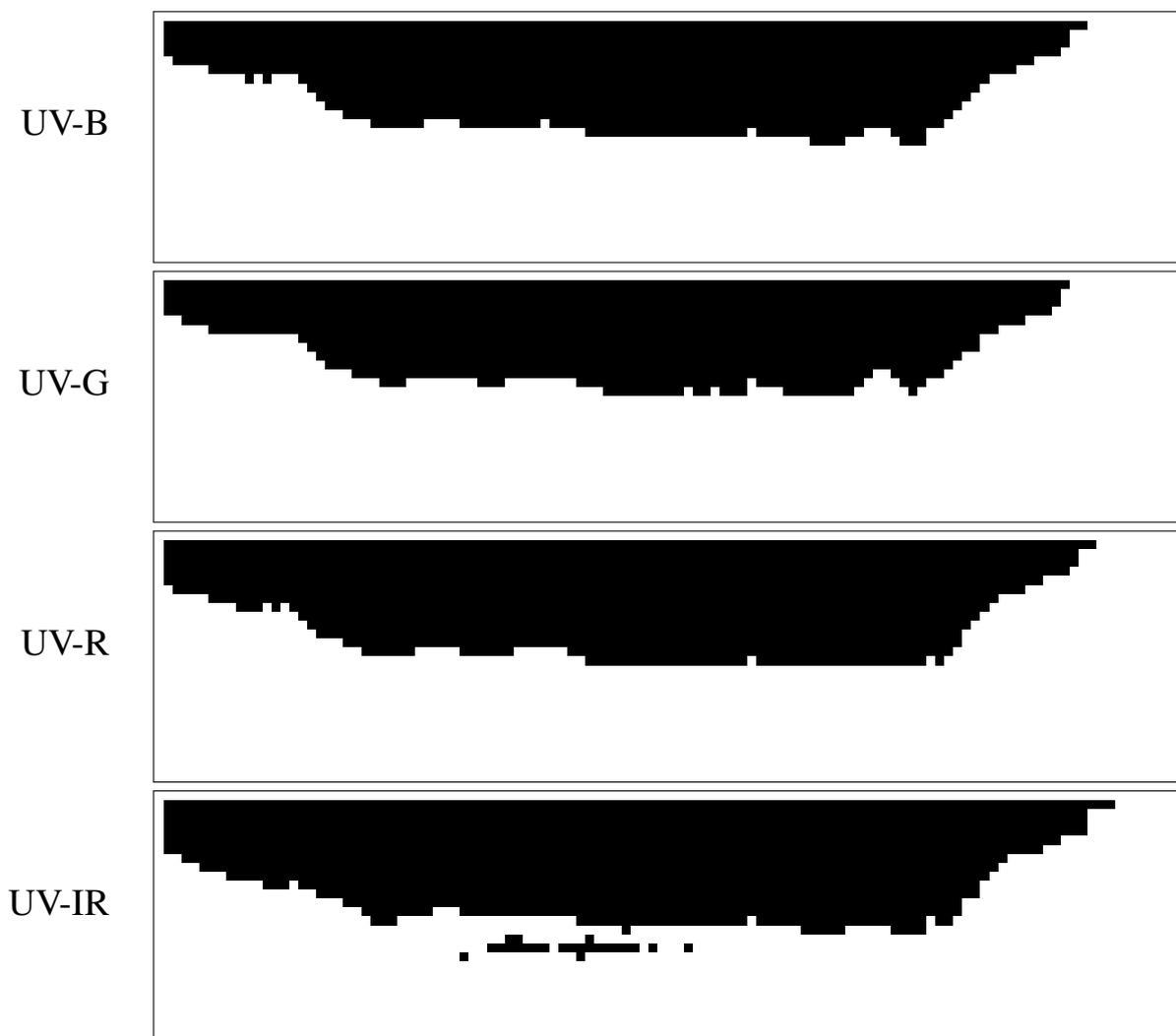


Figure 6.18: Separation images (UV-B, UV-G, UV-R, UV-IR) for dataset one. Black pixels depict pixels classified as background (sky), white ones depict pixels classified as foreground (terrestrial objects).



Figure 6.19: Separation images (B-G, B-R, B-IR) for dataset one. For explanation see caption of figure 6.18.

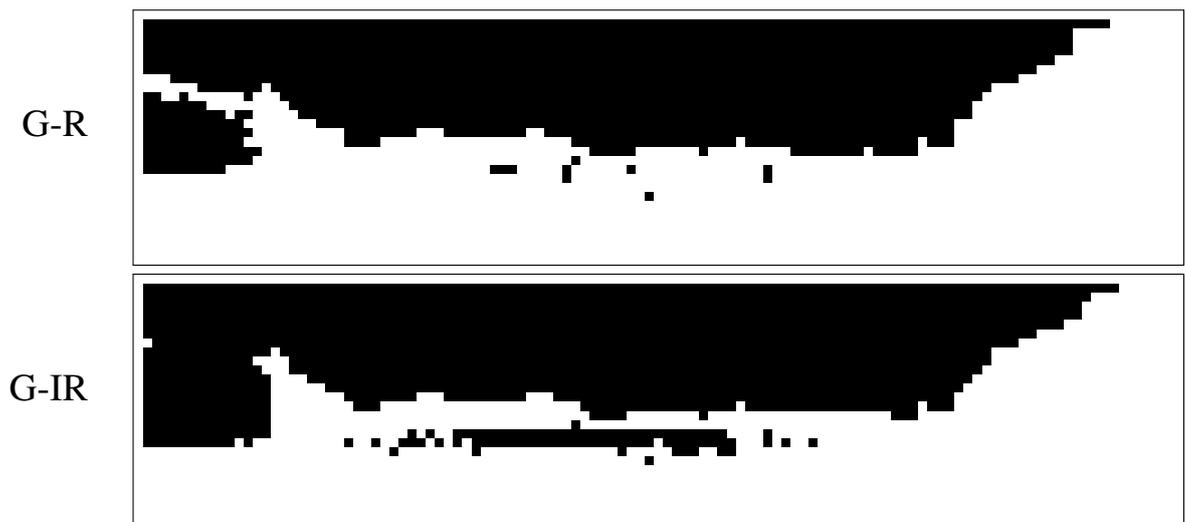


Figure 6.20: Separation images (G-R, G-IR) for dataset one. For explanation see caption of figure 6.18.



Figure 6.21: Separation images (R-IR) for dataset one. For explanation see caption of figure 6.18.



Figure 6.22: Separation images (UV-B, UV-G, UV-R, UV-IR) for dataset two. Black pixels depict the foreground, white ones depict samples of the background.



Figure 6.23: Separation images (B-G, B-R, B-IR) for dataset two. For explanation see caption of figure 6.22.



Figure 6.24: Separation images (G-R, G-IR) for dataset two. For explanation see caption of figure 6.22.

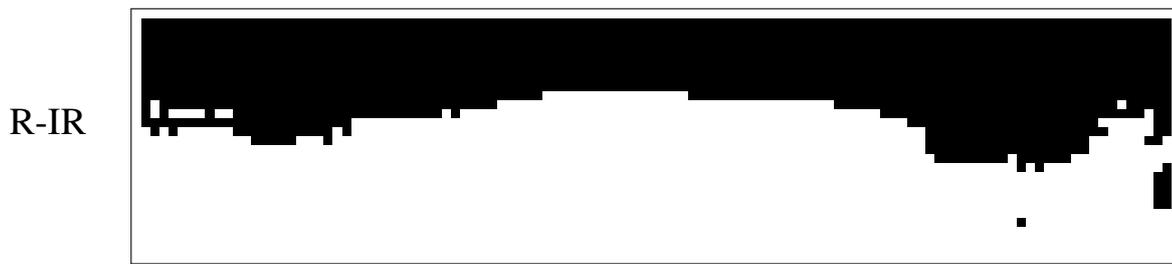


Figure 6.25: Separation images (R-IR) for dataset two. For explanation see caption of figure 6.18.

## 6.5.2 Navigation Dataset

The example datasets one and two can not be used to perform visual navigation; both datasets only consist of one position. For navigation purpose a very small database was collected. The Directed Perception pan-tilt unit can not pan  $360^\circ$ , so the sensor captured  $180^\circ$ , was manually rotated and captured  $180^\circ$ , again. The images were collected in the backyard arranged in a square. The images are captured at the corners of the square with a side length of 2m. The SIFT visual local homing method was used to compute home vectors. Computing the average angular error over all contrasts gives  $AAE = 1.3427$ . Figure 6.26 shows the computed home vectors. Figure 6.27 shows the captured UV images of the positions.

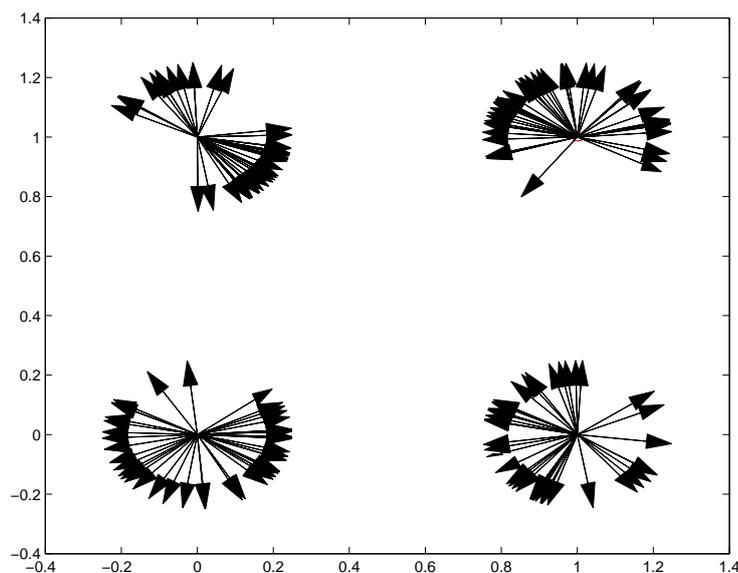


Figure 6.26: SIFT home vectors of all contrasts, square test.

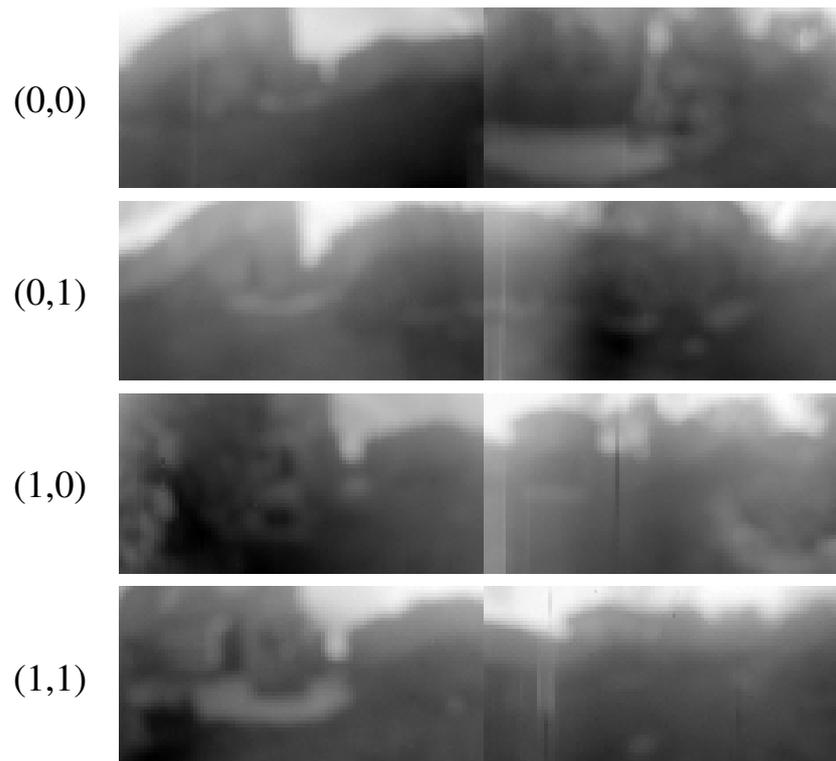


Figure 6.27: UV channel images of the corner positions (0,0), (0,1), (1,0), and (1,1).

### 6.5.3 Discussion

Even so, the average angular error is  $AAE = 1.3427$  the home vectors of all contrasts shown in figure 6.26 are really bad. One reason is the image data of the four positions itself. Each image exists of 2 single data measures of  $180^\circ$ . Furthermore, the images are far apart from each others, which also might be a reason why the SIFT homing fails. To improve the homing, a better and faster data acquisition system is needed. Unfortunately, the results are not usable at all. Further work has to be done.

## 6.6 Summary and Conclusion

A five-channel hand-held sensor which is sensitive to UV, blue, green, red, and infrared was presented. The sensor was used to measure the spectral reflectance of different natural and artificial objects. Furthermore, samples of sky with different cloud configurations were captured. The objects (foreground) and the sky (background) are discriminated by the Linear Discriminant Analysis. The UV-IR dual-channel contrast yields the best separation for natural and artificial objects.

A two-dimensional scene was captured with the five-channel hand-held sensor. For this purpose the sensor was mounted on a pan-tilt unit. The data acquisition setup generates two-dimensional images of each spectral frequency. By analysing the different dual-channel contrasts combinations over time, the stability of the contrasts is demonstrated. It is also pointed out, that the contrasts stability over time increases if the learned separation is used to separate foreground and background pixels. Seven measures are presented to verify the stability. Furthermore it was tested if the contrast images can be used to perform visual local homing in outdoor environments. Due to the noisy input images that have been captured, the visual homing tests in “outdoor environments” failed. Further test for “visual homing in outdoor environments” that make use of dual-channel contrasts need to be performed.

# Chapter 7

## Cleaning Robot Control

Today's commercial available autonomous cleaning robots use randomized cleaning strategies. To achieve a good cleaning result, a cleaning strategy with a small overlap of cleaned areas and a full coverage of the work space is needed. By moving the robot along parallel cleaning lanes, the cleaning performance is increased. For this purpose, a trajectory controller based on local visual homing methods was designed.

### 7.1 Trajectory Controller

The trajectory controller based on local visual homing methods tries to control the distance of meandering parallel lanes (Kreft, 2007; Möller et al., 2008; Gerstmayr et al., 2009). A topological map is built to obtain long-range navigation abilities. To build the map, the robot captures panoramic snapshots every 10cm, which are successively added to the map. When the robot reaches the end of one lane, it turns by  $90^\circ$ . It moves forward (by a predefined distance), turns again by  $90^\circ$ , and starts a new lane parallel to the previous one.

Figure 7.1 depicts the trajectory control approach. The inter-lane distance is 30cm and should be set to the width of the vacuum cleaning unit of the robot.  $SS_1$  and  $SS_2$  are two snapshots on the previous lane which are distance  $d$  apart. CV depicts the current view position. From this position, the home vectors to both snapshots are computed and yield the home vector angles  $\alpha_1$  and  $\alpha_2$ ; both are relative to the orientation of the robot. Knowing the angles and the distance  $d$ , the distance  $h$  which is perpendicular to  $d$  can be computed. The estimated distance  $h$  and the inter-lane distance (e.g. 30cm) give the displacement error  $e$  of the robot. The displacement error  $e$  is used to compute a movement vector for steering the robot closer to the desired lane.

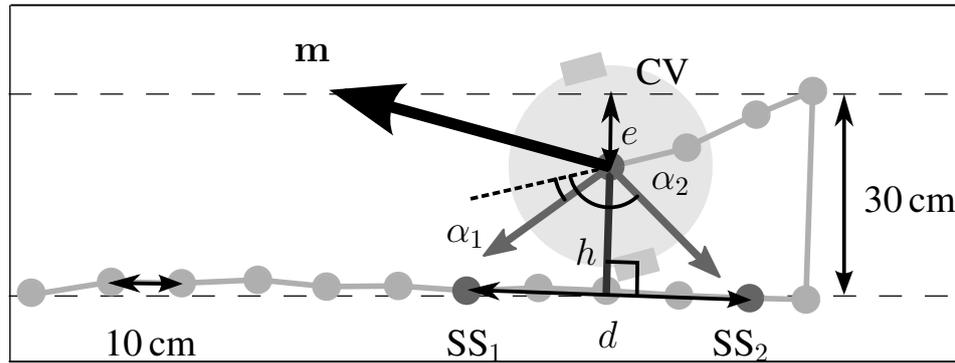


Figure 7.1: Triangulation trajectory controller for keeping the robot on a lane parallel to the previous lane (adapted from Gerstmayr et al. (2009)). The snapshots are depicted by  $SS_1$  and  $SS_2$ , CV is the current view,  $\alpha_1$  and  $\alpha_2$  are the home vector angles.  $d$  is the distance between both snapshots measured by the odometry,  $h$  is the computed distance of the robot to the lane and  $e$  is the error to the desired inter-lane distance.  $m$  depicts the computed robot movement vector by the trajectory controller.

## 7.2 Tests

Figure 7.2 shows a schematic setup of the tests. The first setup (figure 7.2(a)) uses seven long lanes (4m each) which start at the bottom right corner of the cleaning area. The second setup (figure 7.2(b)) has fourteen short lanes (2m each) and starts at the top right corner. The cleaning area is not extended to the whole accessible floor because the external tracking system, which is used for analyzing the cleaning runs, does not cover the whole floor.

To show that the trajectory controller works and that the results are reproducible, ten cleaning runs for both setups have been collected. For this purpose, the robot was manually placed on the starting point at the beginning of every run. The orientation of the robot at the starting point was manually adapted such that the initial orientation for all runs of one test setup was as accurate as possible. The first lane was driven with a wall following approach. A static error of 10% was added to the wheel speeds (+5% on the left, -5% on the right); after the half of the tests the error was switched to -5% on the left wheel and +5% on the right wheel. The same error was considered in the simulated odometry of the robot. The size of the unfolded panoramic images was  $335 \times 40$ . A variant of two-dimensional warping (min-warping, Möller (2009)) with parameters which perform well for this application (Gerstmayr et al., 2009) was used.

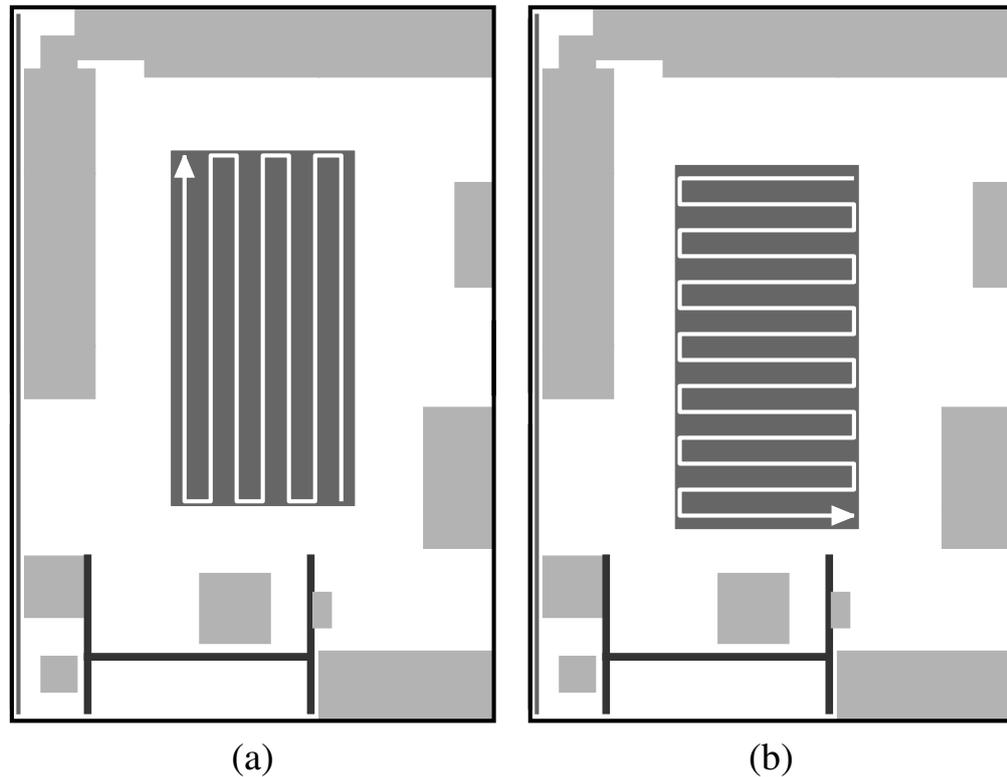


Figure 7.2: The ideal cleaning trajectories for the laboratory room are shown. A cleaning run with seven long lanes is shown in (a), (b) depicts a cleaning run with fourteen short lanes. The gray areas depict furniture, this area is not accessible. The black “H” shape is the support of a robot arm and also not accessible (adapted from Gerstmayr et al. (2009)).

To analyze the quality of a cleaning run, a model of the cleaning device was fitted to the trajectory. The width of the cleaned area of a trajectory was modeled by the width of the cleaning unit. The light gray area symbolizes the area that was cleaned once by the robot. The white parts depict the uncleaned areas. A perfect trial would yield 100.0% coverage, 0.0% overlap, and 0.0% uncovered area.

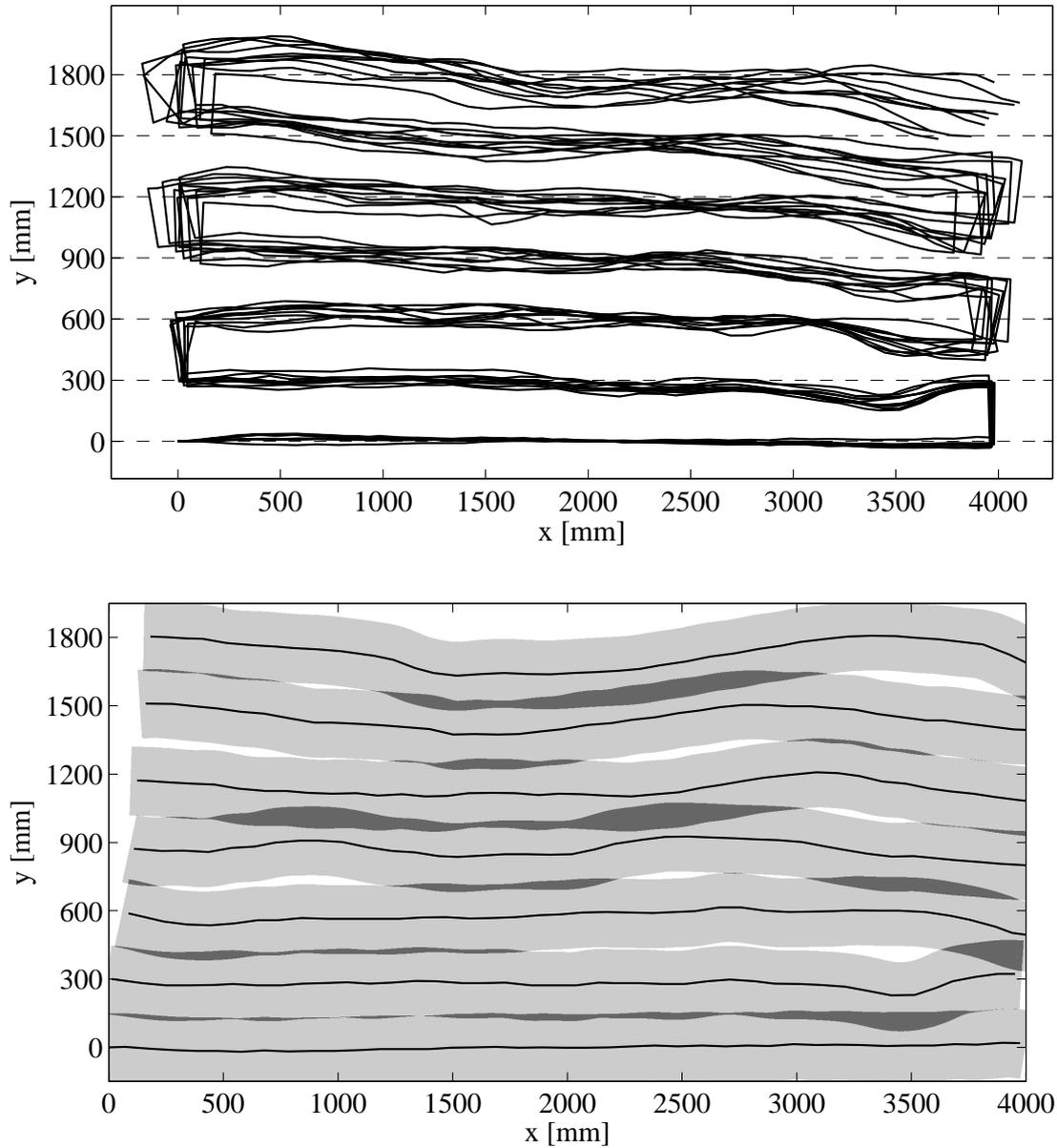


Figure 7.3: Cleaning results for the long lane test setup (adapted from Gerstmayr et al. (2009)). **Top:** Trajectories of ten cleaning runs, tracked by an external tracking system. For comparison reasons, the first lane of each trajectory is aligned to the abscissa to compensate for orientation errors at the starting point. **Bottom:** Overlap analysis of one cleaning run. White depicts uncleaned areas, light gray depicts cleaned areas, and dark gray depicts overlap areas.

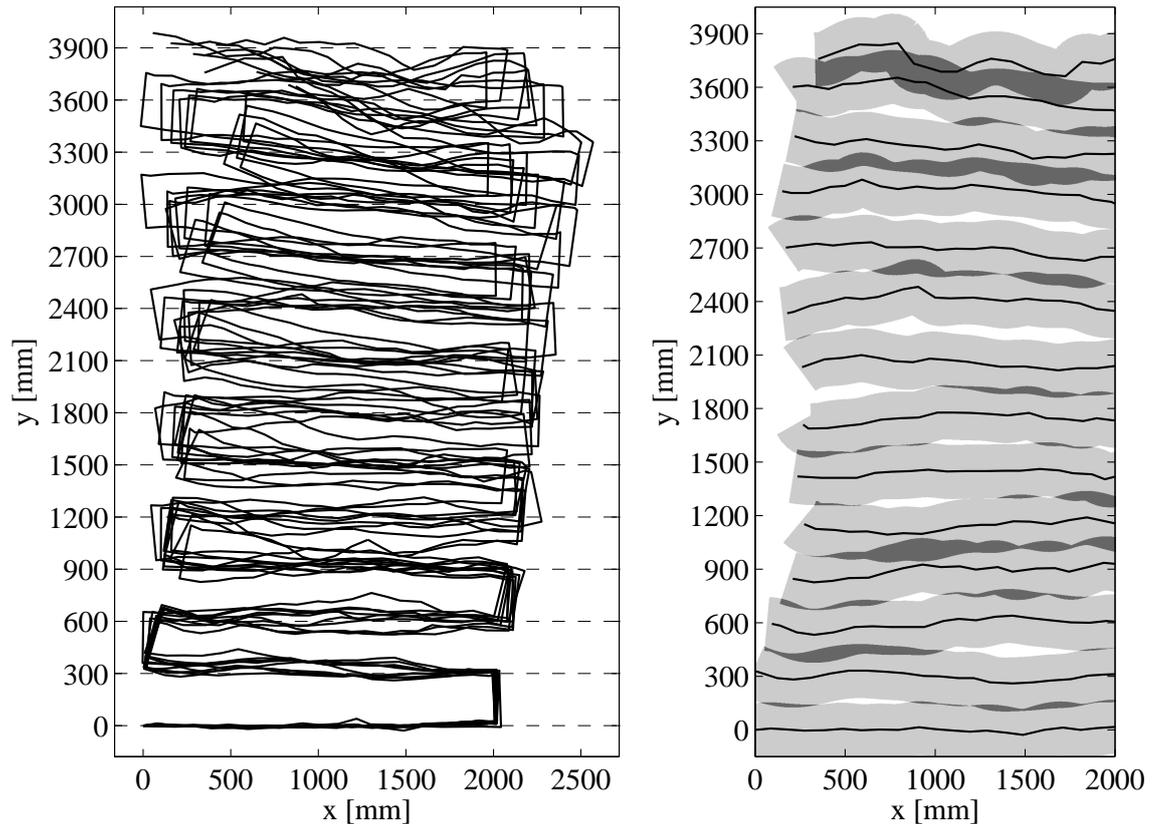


Figure 7.4: Cleaning results for the short lane test setup (adapted from Gerstmayr et al. (2009)). **Top:** Trajectories of ten cleaning runs, tracked by an external reference tracking system. For comparison reasons, the first lane of each trajectory is aligned to the abscissa to compensate for orientation errors at the starting point. **Bottom:** Overlap analysis of one cleaning run. The color coding is the same as in figure 7.3

## 7.3 Discussion

Figure 7.3 and figure 7.4 show the results of the cleaning tests. In both tests, the trajectory controller worked well. It can be seen that the lanes of one trajectory do not cross each other; thus, the trajectory controller managed to stabilize the inter-lane distance. However, if a deviation occurred in one lane, the following lanes also show the deviation. A deviation in a lane is an indication for a false computation of the distance which might occur if the measured distance  $d$  (figure 7.1) between the used snapshots is measured wrongly or if the home vector angles  $\alpha_1$  and  $\alpha_2$  computed between the current view and the snapshots are not accurate enough or incorrect.

As seen in figure 7.3, the trajectories of ten runs differ only slightly. Mostly, the lane length differs, which might be caused by some wheel slip when the robot starts moving. If the lanes are not totally straight, the curve of the lane

also has an effect on the length of the lane. The overlap analysis from figure 7.3 shows that one trajectory produces only small areas that overlap (depicted by dark gray areas in the figure). Averaged over all 10 trials of the first experiment, the coverage was 86.4%, the overlap is 6.5% and uncovered area is 7.4% (Gerstmayr et al., 2009).

Figure 7.4 shows the trajectories for the short lanes. The mean coverage is 75.2%, the overlap is 8.9%, and the uncovered area is 15.9% (Gerstmayr et al., 2009). It can be seen that the trajectory controller is able to handle more than ten parallel lanes, but the lanes get more wound from lane to lane. The errors accumulate over the lanes, i.e. an error on one of the first lanes has an influence on the quality of later lanes, too. If the computed home vector angles are not precise, the trajectory controller will reduce the error. Since the error of the inter-lane distance gets greater with increasing number of lanes, the overlapping and the uncleaned areas increase as well as the uncleaned areas do.

## 7.4 Summary and Conclusion

A trajectory controller and tests with respect to the quality of the controller were presented in this chapter. Two test scenarios have been explained and the results were discussed.

A major conclusion is that it is advisable to use a cleaning strategy that avoids a large number of parallel lanes. The strategy should consider the shape of the area that needs to be cleaned. Large areas should be cleaned by long lanes with a small number of parallel lanes. Whenever the lanes get too disturbed, which could be detected by the number of direction changes of the robot, the strategy might decide to record a new reference lane. Thus, the trajectory controller works well, but a strategy controller is needed to ensure a higher quality of cleaning.

# Chapter 8

## Overall Conclusion

This thesis is focused on four major topics:

- The Polarized-Light Compass,
- Local Visual Homing Methods,
- Outdoor Navigation with Color Contrasts,
- Cleaning Robot Control.

The introduction gives a brief overview of all four topics. It motivates the adaptation of biological solutions for complex problems and the fundamentals that have already been published. The objectives are presented in chapter two. For each major topic, the objectives are presented. It is explained that the polarized-light compass is used for measuring the  $e$ -vector field of the sky. Furthermore, it is pointed out, that two new local visual homing methods are tested and compared to 2D-warping. Then, the local visual navigation approaches are adapted to the outdoor environments. For this purpose, the input images are measured with a five-channel hand-held sensor that is mounted on a pan-tilt unit. The illumination situation in outdoor applications vary over several decades. To achieve an illumination invariant representation of the scene, color contrasts of the different spectral frequencies, measured by the hand-held sensor, are used. The cleaning robot control approach tries to combine the local visual homing methods with a cleaning application.

The materials and methods chapter describes the tools that are used during this thesis. Image databases, quality measures for navigation, statistical methods, and feature selection methods are presented. The polarized-light compass chapter shows the work that was done for this major topic. The construction of the compass and the measurement of the  $e$ -vector field is shown.

Two local visual homing methods are presented in chapter five. The performance of SIFT homing compared to 2D-warping is pointed out. Furthermore, a parameter model navigation method is shown. It is pointed out, that SIFT

homing performs as good as 2D-warping if the images are large enough, even though, the computational speed is not as fast as 2D-warping. The SIFT homing is about 3 times slower than the 2D-warping method. The parameter models are noncompetitive. The homing results of rotation invariant parameters is not close enough to 100% return ratio. In chapter six, the thesis focuses on outdoor navigation with color contrasts. A positive conclusion can be made for the color contrasts. It is shown that dual-channel color contrasts are stable over time, especially, if a learned separation of foreground and background is used to be independent from illumination changes. Regrettably, the navigation attempt in outdoor environments failed when an indoor local homing method is used for the outdoor scene. But, the reasons for this failure might not be searched at the used homing method. The quality of the input images was bad. Thus, the provisional data acquisition for panoramic images failed. The manually rotated images were too imprecise. Consolidating the four major topics into one application would lead to new knowledge. A possible application would be a self-sufficiently working outdoor lawn-mower. The lawn-mower could use visual navigation to drive parallel lanes while mowing the lawn.

# Chapter 9

## Future Research

This thesis focuses on a small part of the four research topics, since the robotic field of research is immense. The following four sections present some ideas for future research that could expand the current work.

### 9.1 Polarized-Light Compass

In this thesis, the polarized-light compass is used to visualize the e-vector field of the sky. It is shown, that it is possible to calculate a compass information from the cloudless sky. In future research, the focus has to be pointed towards the assumption, that bees can also navigate if they only see a small patch of the sky. A model could be developed that is able to handle a cloudy sky with only small unclouded patches. For this purpose, it is necessary to collect more e-vector fields of the sky. Especially, it is necessary to collect data that covers a multiplicity of different clouded skies.

### 9.2 Local Visual Homing

The different local visual homing methods bear a lot of future research options. First of all, the speed of a local homing method is a crucial factor for the usability in applications. In most cases, the speed of a homing method is proportional to the size of the input images; the smaller the size, the faster the computation. Unfortunately, for every method a parameter optimization has to be carried out. The result of the parameter optimization could yield larger input images if the homing quality needs to be increased. In this thesis, SIFT homing performed best for image sizes larger than  $1122 \times 162$ ; but warping performs best for  $346 \times 50$ .

Another aim is that a local visual homing method should cope with illumination changes. Extracting invariant features is one way to achieve illumination invariance. Local visual homing methods only can gain attention to industrial

applications like cleaning, if they are robust to illumination changes and temporal changes of a scene over time. The combination of different existing local visual homing methods might yield a better performance.

### 9.3 Outdoor Navigation

The outdoor navigation with color contrasts is not well explored at the moment. For future research, a faster data acquisition system is needed. The hand-held pan-tilt unit acquisition system, used in this thesis, is too slow to capture a large image database of one environment. It also is not able to scan 360° at “one shot”. Furthermore, a new sensor needs to be built because at the moment only a few cameras with logarithmic output offer the required spectral range. With a new fast sensor it would be possible to capture an outdoor image database, which could be used for offline local homing experiments. It has to be analyzed if the existing local homing methods are also able to work with these input images or if new visual local homing methods have to be developed.

### 9.4 Cleaning Robot Control

The future research in the field of cleaning robot control could be focused on methods that increase the autonomous behavior of the robot. The first objective is the quality of cleaning: How can different control strategies increase the cleaning quality? Second, it is necessary to extend the control algorithm with different behaviour strategies like a “clever” obstacle avoidance or a strategy for “trapped in a corner” situations. The second objective is an efficient cleaning. Efficient cleaning includes an intelligent power consumption (e.g. cleaning unit is off when moving towards already cleaned areas), the ability to discover uncleaned areas, and autonomous reacting to unknown situations. At the moment, the internal topological representation is only used to control the inter-lane distance to the previous lane. The system might be extended by an approach that drives the shortest way to a desired position like a battery-charging station.

# Bibliography

- Alpaydin, E. *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, October 2004.
- Anderson, E. Animals as Navigators. In *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pages 1003–1007, 1983.
- Backhaus, W. and Menzel, R. Color distance derived from a receptor model of color vision in the honeybee. *Biological Cybernetics.*, 55(5):321–331, 1987.
- Bianco, G., Zelinsky, A., and Lehrer, M. Visual landmark learning. In *Proceedings of the International Conference on Intelligent Robots and Systems*, volume 1, pages 227–232, Takamatsu , Japan, 2000.
- Binding, D. and Labrosse, F. Visual local navigation using warped panoramic images. In *Proceedings of Towards Autonomous Robotic Systems*, pages 19–26, University of Surrey, Guildford, 2006.
- Brandt, R. and Vorobyev, M. Metric analysis of threshold spectral sensitivity in the honeybee. *Vision Research*, 37(4):425–439, 1997.
- Bregy, P., Sommer, S., and Wehner, R. Nest-mark orientation versus vector navigation in desert ants. *Journal of Experimental Biology*, 211(12):1868–1873, 2008.
- Briscoe, A. and Chittka, L. The evolution of color vision in insects. *Annual Review of Entomology*, 46:471–510, 2001.
- Brown, M. and Lowe, D. Invariant features from interest point groups. In *British Machine Vision Conference*, pages 656–665, Cardiff, Wales, 2002.
- Cartwright, B. and Collett, T. S. Landmark learning in bees. *Journal of Comparative Physiology A*, 151(4):521–543, 1983.
- Cartwright, B. A. and Collett, T. S. Landmark maps for honeybees. *Biological Cybernetics*, (57):85–93, 1987.

- Cheung, A., Stürzl, W., Cheng, K., and Zeil, J. The information content of panoramic images: II. view-based navigation in nonrectangular experimental areas. *Journal of Experimental Psychology*, 34(1):1–14, 2008.
- Colios, C. I. and Trahanias, P. E. Landmark identification based on projective and permutation invariant vectors. In *International Conference on Pattern Recognition*, volume 4, pages 128–131, Los Alamitos, CA, USA, 2000. IEEE Computer Society.
- Daumer, K. Reizmetrische Untersuchungen des Farbsehens der Bienen. In *Journal of Comparative Physiology*, volume 38, pages 413–478. Springer Berlin/ Heidelberg, 1956.
- Duchon, A. P. Maze navigation using optical flow. In *International Conference on Simulation of Adaptive Behavior*, pages 225–232. MIT Press, 1996.
- Efron, B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.
- Efron, B. *The Jackknife, the Bootstrap, and Other Resampling Plans (CBMS-NSF Regional Conference Series in Applied Mathematics)*. Society for Industrial & Applied Mathematics, January 1987.
- Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, May 1994.
- Fiala, M. and Basu, A. Robot navigation using panoramic tracking. *Pattern Recognition*, 37(11):2195–2215, 2004.
- Fischer, S. Visuelle Navigation mit Parameter-Modellen. Diplomarbeit, AG Technische Informatik, Technische Fakultät der Universität Bielefeld, Bielefeld, 2006.
- Franz, M. O., Schölkopf, B., Mallot, H. A., and Bühlhoff, H. H. Where did I take that snapshot? Scene-based homing by image matching. *Biological Cybernetics*, (79):191–202, 1998.
- Fukushi, T. Homing in wood ants, *formica japonica*: Use of the skyline panorama. *The Journal of Experimental Biology*, 204(12):2063–2072, 2001.
- Gerstmayr, L., Röben, F., Krzykowski, M., Kreft, S., Venjakob, D., and Möller, R. A vision-based trajectory controller for autonomous cleaning robots. To appear in *Proceedings of Autonomous Mobile Systems*, 2009.

- Gerthsen, C. and Meschede, D. *Gerthsen Physik*. Springer-Verlag, Berlin, Heidelberg, 2006.
- Goedemé, T., Nuttin, M., Tuytelaars, T., and Van Gool, L. Vision based intelligent wheel chair control: The role of vision and inertial sensing in topological navigation. *J. Robot. Syst.*, 21(2):85–94, 2004.
- Gourichon, S., Meyer, J. A., Lip, A., Ieng, S. H., Smadja, L., and Benosman, R. Estimating ego-motion using a panoramic sensor: Comparison between a bio-inspired and a camera-calibrated method. In *AISB03 Symposium on Biologically Inspired Vision, Theory and Application*, pages 91–101, 2003.
- Graham, P. and Cheng, K. Which portion of the natural panorama is used for view-based navigation in the australian desert ant? *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 195(7):681–689, 2009.
- Hafner, V. Adaptive homing - robotic exploration tours. *Adaptive Behavior*, 9 (3-4):131–141, 2001.
- Hayet, J. B., Lerasle, F., and Devy, M. A visual landmark framework for mobile robot navigation. *Image Vision Computing*, 25(8):1341–1351, 2007.
- Hong, J., Tan, X., Pinette, B., Weiss, R., and Riseman, E. M. Image-based homing. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, volume 1, pages 620–625, 1991.
- Horridge, G. *The Compound Eye and Vision of Insects*, pages 121–153. Oxford: Clarendon Press, 1975.
- Horridge, G. The separation of visual axes in apposition compound eyes. *Philosophical Transactions of the Royal Society of London*, 285(1003):1–59, 1978.
- Horváth, G. and Varjú, D. *Polarized Light in Animal Vision*. Springer Verlag, 2004.
- Koenderink, J. J. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984.
- Kollmeier, T. Kontrastanalyse zur beleuchtungsinvarianten Detektion von Landmarken. Diplomarbeit, AG Technische Informatik, Technische Fakultät der Universität Bielefeld, Bielefeld, 2006.

- Kollmeier, T., Röben, F., Schenck, W., and Möller, R. Spectral contrasts for landmark navigation. *Journal of the Optical Society of America A*, 24(1): 1–10, 2007.
- Kreft, S. Reinigungstrajektorien mobiler Roboter unter visueller Steuerung. Diplomarbeit, AG Technische Informatik, Technische Fakultät der Universität Bielefeld, Bielefeld, 2007.
- Lambrinos, D., Kobayashi, H., Pfeifer, R., Maris, M., Labhart, T., and Wehner, R. An autonomous agent navigating with a polarized light compass. *Adaptive Behavior*, 6(1):131–161, 1997.
- Lambrinos, D., Möller, R., Labhart, T., Pfeifer, R., and Wehner, R. A mobile robot employing insect strategies for navigation. *Robotics and Autonomous Systems, special issue: Biomimetic Robots*, 30(1-2):39–64, 2000.
- Land, M. F. Visual Acuity In Insects. *Annual Reviews of Entomology*, 42: 147–177, 1997.
- Land, M. F. and Nilsson, D.-E. *Animal eyes*. Oxford animal biology series. Oxford Univ. Press, 2004.
- Lindeberg, T. Scale-space theory: A basic tool for analysing structures at different scales. *J. of Applied Statistics*, 21(2):224–270, 1994. (Supplement on Advances in Applied Statistics: Statistics and Images: 2).
- Lowe, D. G. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157, Corfu, Greece, 1999.
- Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- Malacara-Hernandez, D. *Color Vision and Colorimetry: Theory and Applications (SPIE Press Book)*, volume PM105, chapter Trichromatic Theory, pages 31–32. SPIE, 2002.
- McCarthy, C. and Barnes, N. Comparison of temporal filters for optical flow estimation in continuous mobile robot navigation, 2004a.
- McCarthy, C. and Barnes, N. Performance of optical flow techniques for indoor navigation with a mobile robot. In *Proceedings of the IEEE International Conference on Robotics & Automation*, pages 5093–5098, 2004b.

- Menegatti, E., Maeda, T., and Ishiguro, H. Image-based memory for robot navigation using properties of the omnidirectional images. *Robotics and Autonomous Systems*, 47(4):251–267, 2004.
- Menzel, R. Untersuchungen zum Erlernen von Spektralfarben durch die Honigbiene (*Apis mellifica*). *Zeitschrift für vergleichende Physiologie*, 56:22–62, 1967.
- Menzel, R. Evidence for color receptors in the hymenoptera eye obtained from selective adaptation experiments. *Journal of Life Sciences*, 3(3):95–100, 1973.
- Menzel, R. and Backhaus, W. Colour Vision in Insects. In Gouras, P., editor, *Vision and Visual Dysfunction*, pages 262–292. Nature Publishing Group, 1991.
- Menzel, R. and Blakers, M. Colour receptors in the bee eye morphology and spectral sensitivity. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 108(1):11–33, 1976.
- Menzel, R., Chittka, L., and Geiger, K. Can honeybees count landmarks? *Anim. Behav.*, (49):159–164, 1995.
- Menzel, R., Geiger, K., Chittka, L., Joerges, J., and Kunze, J. The knowledge base of bee navigation. *Journal of Experimental Biology*, 199(1):141–146, 1996.
- Menzel, R. and Wehner, R. Augenstrukturen bei verschiedenen großen Arbeiterinnen von *Cataglyphis bicolor* Fabr. (Formicidae, Hymenoptera). *Zeitschrift für vergleichende Physiologie*, 68(4):446–449, 1970.
- Mikolajczyk, K. *Detection of local features invariant to affine transformations*. PhD thesis, Grenoble Institute of Technology, Grenoble, 2002.
- Mitchell, T. and Labrosse, F. Visual homing: a purely appearance-based approach. In *University of Essex*, pages 101–108, 2004.
- Möller, R. Insect visual homing strategies in a robot with analog processing. *Biological Cybernetics*, 83(3):231–243, 2000.
- Möller, R. Do insects use templates or parameters for landmark navigation? *Journal of Theoretical Biology*, 210(1):33–45, 2001.
- Möller, R. A biorobotics approach to the study of insect visual homing strategies. Habilitationsschrift, Wirtschaftswissenschaftliche Fakultät der Universität Zürich, 2002a.

- Möller, R. Insects could exploit UV-green contrast for landmark navigation. *Journal of Theoretical Biology*, 214(4):619–631, 2002b.
- Möller, R. Local visual homing by warping of two-dimensional images. *Robotics and Autonomous Systems*, 57(1):87–101, 2009.
- Möller, R., Krzykawski, M., and Gerstmayr, L. Three 2d-warping schemes for visual robot navigation. *Autonomous Robots*, 2010. submitted.
- Möller, R., Lambrinos, D., Roggendorf, T., Pfeifer, R., and Wehner, R. Insect strategies of visual homing in mobile robots. In Webb, B. and Consi, T. R., editors, *Biorobotics. Methods and Applications*, pages 37–66. AAAI Press / MIT Press, 2001.
- Möller, R. and Vardy, A. Local visual homing by matched-filter descent in image distances. *Biological Cybernetics*, 95(5):413–430, 2006.
- Möller, R., Vardy, A., Gerstmayr, L., Röben, F., and Kreft, S. Neuroethological concepts at work: Insect-inspired methods for visual robot navigation. In *Biological Approaches for Engineering*, pages 91–94. Institute of Sound and Vibration Research, University of Southampton, 2008.
- Möller, R., Vardy, A., Kreft, S., and Ruwisch, S. Visual homing in environments with anisotropic landmark distribution. *Autonomous Robots*, 23(3):231–245, 2007.
- Neumann, K. Bau eines Polarisationslicht-Kompasses. Bachelor's thesis, AG Technische Informatik, Technische Fakultät der Universität Bielefeld, Bielefeld, 2007.
- Peitsch, D., Fietz, A., Hertel, H., Souza, J., Ventura, D. F., and Menzel, R. The spectral input systems of hymenopteran insects and their receptor-based colour vision. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 170(1):23–40, 1992.
- Pomozi, I., Horváth, G., and Wehner, R. How the clear-sky angle of polarization pattern continues underneath clouds: full-sky measurements and implications for animal orientation. *Journal of Experimental Biology*, 204(17):2933–2942, 2001.
- Pons, J. S., Hübner, W., Dahmen, H., and Mallot, H. A. Vision-based robot homing in dynamic environments. *Robotics and Applications*, pages 293–298, Würzburg, Germany, August 2007. ACTA Press.

- Prassler, E. and Kosuge, K. *Springer Handbook of Robotics*, chapter Domestic Robotics, pages 1253–1281. Springer, 2008.
- Prassler, E., Ritter, A., Schaeffer, C., and Fiorini, P. A short history of cleaning robots. *Autonomous Robots*, 9(3):211–226, 2000.
- Quinn, G. P. and Keough, M. J. *Experimental design and data analysis for biologists*. Cambridge University press., United Kingdom., 2002.
- Rizzi, A., Bianco, G., and Cassinis, R. A bee-inspired visual homing using color images. *Robotics and Autonomous Systems*, 25(3-4):159–164, Nov 30 1998.
- Santos-Victor, J., Sandini, G., Curotto, F., and Garibaldi, S. Divergent stereo for robot navigation: Learning from bees. In *Computer Vision and Pattern Recognition*. IEEE, 1993.
- Schatz, A. Visuelle Navigation mit Scale Invariant Feature Transform. Diplomarbeit, AG Technische Informatik, Technische Fakultät der Universität Bielefeld, Bielefeld, 2006.
- Se, S., Lowe, D., and Little, J. Vision-based mobile robot localization and mapping using scale-invariant features. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2051–2058, 2001.
- Se, S., Lowe, D., and Little, J. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21(8):735–758, 2002.
- Seidl, R. *Die Sehfelder und Ommatidien-Divergenzwinkel von Arbeiterin, Königin und Drohne der Honigbiene Apis mellifera*. PhD thesis, Technische Hochschule Darmstadt, 1982.
- Stürzl, W., Cheung, A., Cheng, K., and Zeil, J. The information content of panoramic images: I. rational errors and the similarity of views in rectangular experimental arenas. *Journal of Experimental Psychology*, 34(1):1–14, 2008.
- Stürzl, W. and Mallot, H.-P. A. Efficient visual homing based on Fourier transformed panoramic images. *Robotics and Autonomous Systems*, 54(4):300–313, 2006.
- Tistarelli, M. and Grosso, E. Dynamic stereo in visual navigation. In *1991 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'91)*, pages 186–193, 1991.

- Towne, W. and Moscrip, H. The connection between landscapes and the solar ephemeris in honeybees. *Journal of Experimental Biology*, 211(23):3729–3736, 2008.
- van Kleef, J., Berry, R., and Stange, G. Directional selectivity in the simple eye of an insect. *Journal of Neuroscience*, 28(11):2845–2855, 2008.
- Vardy, A. *Biologically Plausible Methods for Robot Visual Homing*. PhD thesis, School of Computer Science, Carleton University, Canada, 2005.
- Vardy, A. and Möller, R. Biologically plausible visual homing methods based on optical flow techniques. *Connection Science*, 17(1-2):47–89, 2005.
- Vardy, A. and Oppacher, F. Low-level visual homing. In *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)*, pages 875–884. Springer Verlag, 2003.
- Vardy, A. and Oppacher, F. Anatomy and physiology of an artificial vision matrix. In *Proceedings of the First International Workshop on Biologically Inspired Approaches to Advanced Information Technology*, pages 290–305. Springer, 2004.
- von Frisch, K. Der Farben- und Formensinn der Bienen. *Zoologische Jahrbücher der Physiologie*, 35(1-2):1–188, 1914.
- von Frisch, K. Über den Geruchssinn der Biene und seine blütenbiologische Bedeutung. *Zoologische Jahrbücher der Physiologie*, 37:1–238, 1919.
- von Helversen, O. and Edrich, W. Der Polarisationsempfänger im Bienenaug: ein Ultraviolett-rezeptor. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 94(1):33–47, 1974.
- Weber, K., Venkatesh, S., and Srinivasan, M. V. Insect-inspired robotic homing. *Adaptive Behavior*, 7(1):65–97, 1999.
- Witkin, A. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1022, 1983.
- Zeil, J., Hofmann, M. I., and Chahl, J. S. Catchment areas of panoramic snapshots in outdoor scenes. *Optical Society of America Journal A*, 20(3):450–469, March 2003.

# Appendix A

## Polarized-light Compass, Additional Figures

### A.1 Compass Filter

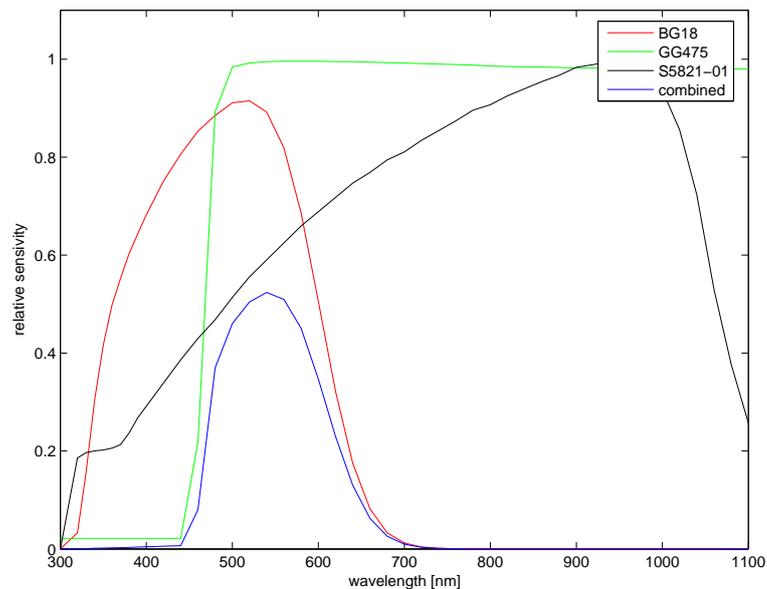


Figure A.1: Unnormalized spectral sensitivity of the used Schott glass filter for the polarization compass, retrieved from Schott's program Filter2004. Only the wavelengths from 300 to 1100nm are considered since the diode cutoff is at 1150nm.

## A.2 Circuit Diagram

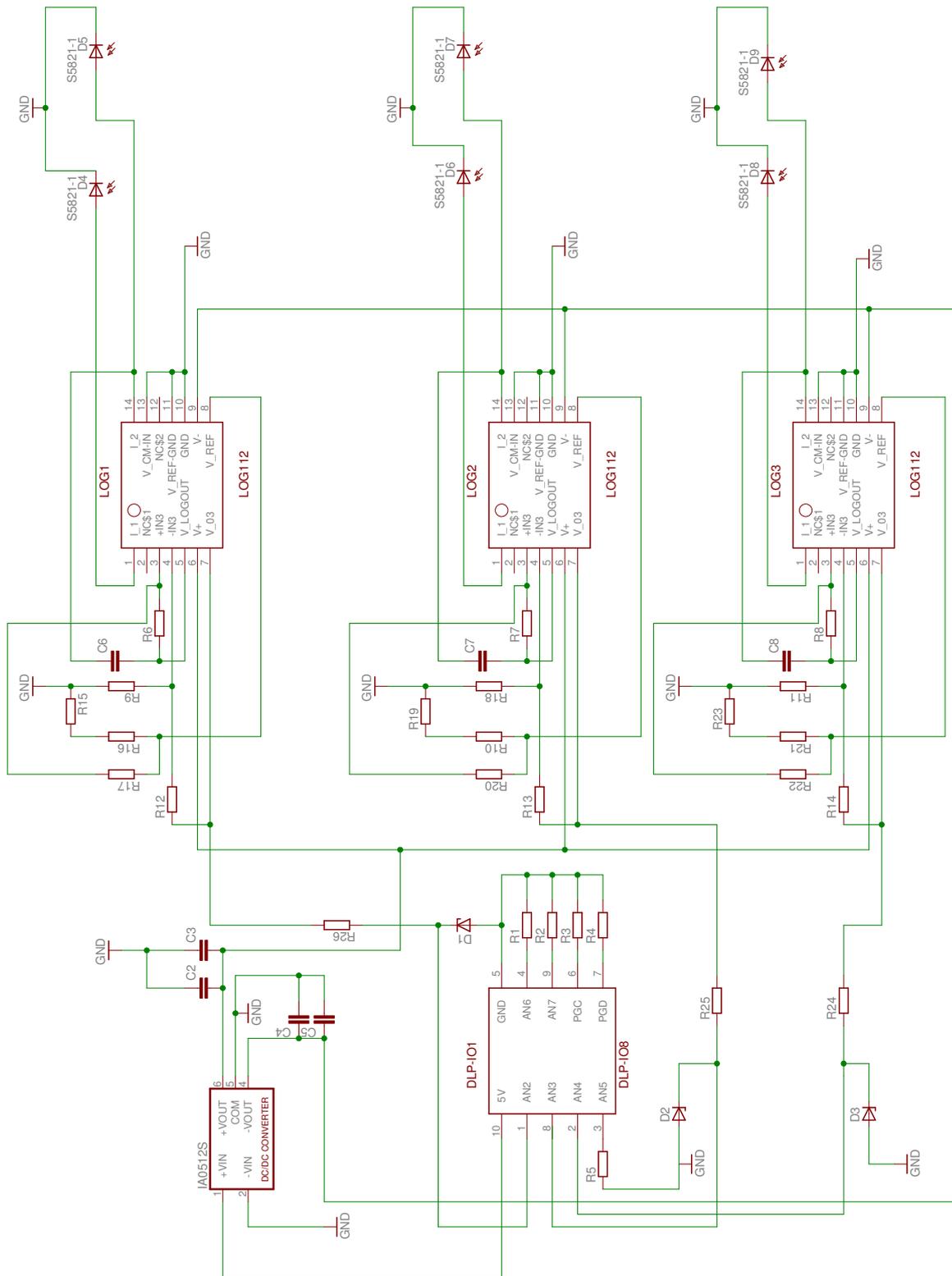


Figure A.2: Polarized-light compass circuit diagram. The values of the components are presented in table A.1 for enhanced readability of the diagram.

Component	Type	
C2, C4	$10\mu F$	capacitor
C3, C5	$1000pF$	capacitor
C6, C7, C8	$68pF$	capacitor
D1, D2, D3	6, 8V	Zener diode
D4, D5, D6, D7, D8, D9	S5821-1	Hamamatsu photodiode
DC1	IA0512S	XP Power DC/DC converter
DLP-IO1	DLP-IO8-G	8-channel data acquisition board
LOG1, LOG2, LOG3	LOG112	log ratio amplifier
R1, R2, R3, R4, R5	$4.7k\Omega$	resistor
R6, R7, R8, R17, R20, R22	$100k\Omega$	resistor
R9, R11, R12, R13, R14, R18	$15k\Omega$	resistor
R10, R15, R16, R19, R21, R23	$3.9k\Omega$	resistor
R24, R25, R26	$100\Omega$	resistor

Table A.1: List of components for the polarized-light compass circuit diagram shown in figure A.2.

### A.3 Additional Drawings

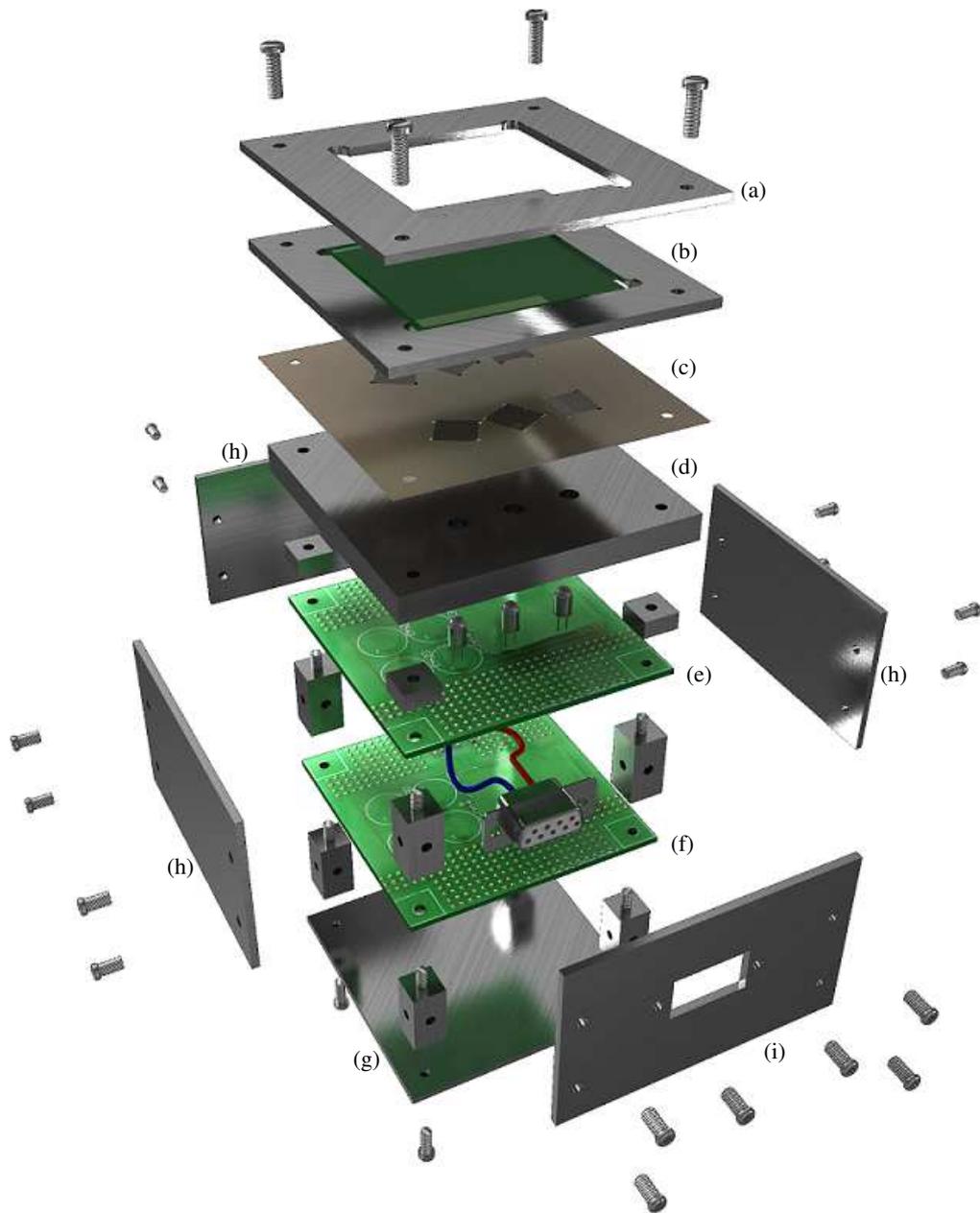


Figure A.3: Exploded assembly drawing of the polarized-light compass; adapted from Neumann (2007). The recess of the cover-plate on top (a) is slightly smaller than the clearance of the glass filter holder (b) so that the glass filter is clamped between polarized-light filter holder (c) and cover-plate (a). The polarized-light filter holder is mounted on top of the photodiode mount (d). The photodiodes are soldered to the first circuit board (e), which contains the analog logarithmic amplifiers. The second circuit board (f) contains the electric power supply. All layers of the polarized-light compass are mounted on the baseplate (g). The electric circuit boards are mounted with spacers. The parts (h) and (i) enclose the electric circuits of the compass, thus the compass is dustproof.

# Appendix B

## SIFT Homing

### B.1 Additional Tables

image size	database					
	<b>original</b>			<b>uni</b>		
	Keys	$AAE_*$	$RR_*$	Keys	$AAE_*$	$RR_*$
$1672 \times 206$	413	0.0849	1	368	0.2681	0.8424
$836 \times 103$	149	0.1213	1	165	0.3542	0.8932
$418 \times 52$	58	0.1841	0.9997	42	0.8104	0.6345
$209 \times 52$	39	0.2119	1	24	0.9974	0.5365

Table B.1: The average vector approach (see 5.1.3) for the **original** and the **uni** database with respect to the image size, taken from Schatz (2006).

image size	database					
	<b>original</b>			<b>uni</b>		
	Keys	$AAE_*$	$RR_*$	Keys	$AAE_*$	$RR_*$
$1672 \times 206$	413	0.0478	1	368	0.2942	0.8254
$836 \times 103$	149	0.0728	1	165	0.3642	0.8157
$418 \times 52$	58	0.1322	0.9943	42	1.0292	0.4583
$209 \times 52$	39	0.1634	0.9947	24	1.1969	0.3978

Table B.2: The maximum approach (see 5.1.3) for the **original** and the **uni** database with respect to the image size, taken from Schatz (2006).

image size	database					
	<b>hall1</b>			<b>hall2</b>		
	Keys	$AAE_*$	$RR_*$	Keys	$AAE_*$	$RR_*$
$1672 \times 206$	288	0.1669	1	362	0.3083	0.9881
$836 \times 103$	124	0.2212	0.9939	148	0.3248	0.9908
$418 \times 52$	54	0.7503	0.6861	48	0.4425	0.9461
$209 \times 52$	35	0.5181	0.7165	30	0.3696	0.9523

Table B.3: The average vector approach (see 5.1.3) for the **hall1** and the **hall2** database with respect to the image size, taken from Schatz (2006).

image size	database					
	<b>hall1</b>			<b>hall2</b>		
	Keys	$AAE_*$	$RR_*$	Keys	$AAE_*$	$RR_*$
$1672 \times 206$	288	0.2105	0.9250	362	0.3415	0.9728
$836 \times 103$	124	0.2634	0.9260	148	0.3527	0.9732
$418 \times 52$	54	0.9898	0.5164	48	0.5866	0.8807
$209 \times 52$	35	0.7634	0.5325	30	0.5827	0.8402

Table B.4: The maximum approach (see 5.1.3) for the **hall1** and the **hall2** database with respect to the image size, taken from Schatz (2006).

# Appendix C

## Parameter-Navigation

### C.1 Homing Using Non-Rotation Invariant Parameter

To achieve a non rotation invariant version of the Fourier approach for the matched-filter method, the line-by-line wise, one dimensional discrete Fourier transformation (DFT) of the image  $I$  used. The new parameter is referred to as [mffc\_phase].

$$F_k(I) = \sum_{i=0}^{N-1} I_i \exp(-\imath ki \frac{2\pi}{N}) \quad (\text{C.1})$$

The corresponding potential function is given by

$$p(\mathbf{x}) = \frac{1}{2} \sum_j \sum_{k=0}^{K-1} |F_k(C_j(\mathbf{x})) - F_k(S_j)|^2. \quad (\text{C.2})$$

The home vector is computed by deriving the potential function

$$\begin{aligned} \mathbf{h}(\mathbf{x}) &= -\nabla_{\mathbf{x}} p(\mathbf{x}) \\ &= - \sum_j \sum_{k=0}^{K-1} |F_k(C_j(\mathbf{x})) - F_k(S_j)| \cdot \underbrace{\nabla_{\mathbf{x}} |F_k(C_j(\mathbf{x})) - F_k(S_j)|}_{\text{inner derivative}}. \end{aligned} \quad (\text{C.3})$$

The term of the inner derivative can be further solved. The abbreviation  $C_j(\mathbf{x}) \hat{=} C_j$  is used.

$$\begin{aligned} \nabla_{\mathbf{x}} |F_k(C_j) - F_k(S_j)| &= \nabla_{\mathbf{x}} [(F_k(C_j) - F_k(S_j)) \cdot (F_k(C_j) - F_k(S_j))^*]^{\frac{1}{2}} \\ &= \frac{\nabla_{\mathbf{x}} F_k(C_j) \cdot (F_k(C_j) - F_k(S_j))^* + \nabla_{\mathbf{x}} F_k^*(C_j) \cdot (F_k(C_j) - F_k(S_j))}{2|F_k(C_j) - F_k(S_j)|} \end{aligned} \quad (\text{C.4})$$

The derivative of the Fourier coefficient  $F_k$  can be done by deriving the image itself.

$$\nabla_{\mathbf{x}} F_k(C_j) = \nabla_{\mathbf{x}} \sum_{i=0}^{N-1} C_{ij} \cdot \exp(-\imath k i \frac{2\pi}{N}) = \sum_{i=0}^{N-1} \nabla_{\mathbf{x}} C_{ij} \cdot \exp(-\imath k i \frac{2\pi}{N}) \quad (\text{C.5})$$

The spatial gradient of the images is approximated according to Möller and Vardy (2006).

$$\nabla_{\mathbf{x}} C_{ij} \approx -\frac{1}{D} \underbrace{\mathbf{B}_i^T \mathbf{\Gamma}_j}_{\text{flow-template}} \underbrace{\nabla_{\varphi} C(\varphi_{ij}, \mathbf{x})}_{\text{intensity-gradient}} := \begin{pmatrix} \Phi_{ij}(\mathbf{x}) \\ \Psi_{ij}(\mathbf{x}) \end{pmatrix} \quad (\text{C.6})$$

In equation C.3,  $\nabla_{\mathbf{x}} F_k(C_j)$  can be replaced by  $\begin{pmatrix} F_k(\Phi_{ij}(\mathbf{x})) \\ F_k(\Psi_{ij}(\mathbf{x})) \end{pmatrix}$

$$\begin{aligned} & \nabla_{\mathbf{x}} |F_k(C_j) - F_k(S_j)| = \\ & \frac{\begin{pmatrix} F_k(\Phi_j(\mathbf{x})) \\ F_k(\Psi_j(\mathbf{x})) \end{pmatrix} \cdot (F_k(C_j) - F_k(S_j))^* + \begin{pmatrix} F_k^*(\Phi_j(\mathbf{x})) \\ F_k^*(\Psi_j(\mathbf{x})) \end{pmatrix} \cdot (F_k(C_j) - F_k(S_j))}{2|F_k(C_j) - F_k(S_j)|}. \end{aligned} \quad (\text{C.7})$$

For the enumerator it can be seen that

$$(a + b\imath)(c - d\imath) + (a - b\imath)(c + d\imath) = 2(ac + bc).$$

Moreover, let  $\Re_k(I) \hat{=} \text{Re}\{F_k(I)\}$ ,  $\Im_k(I) \hat{=} \text{Im}\{F_k(I)\}$ ,  $\Phi_j \hat{=} \Phi_j(\mathbf{x})$  and  $\Psi_j \hat{=} \Psi_j(\mathbf{x})$ .

$$\begin{aligned} & \nabla_{\mathbf{x}} |F_k(C_j) - F_k(S_j)| = \\ & \frac{\begin{pmatrix} \Re_k(\Phi_j) \\ \Re_k(\Psi_j) \end{pmatrix} \cdot \text{Re}\{F_k(C_j) - F_k(S_j)\} + \begin{pmatrix} \Im_k(\Phi_j) \\ \Im_k(\Psi_j) \end{pmatrix} \cdot \text{Im}\{F_k(C_j) - F_k(S_j)\}}{|F_k(C_j) - F_k(S_j)|}. \end{aligned} \quad (\text{C.8})$$

Inserting this equation into equation C.3 gives

$$\begin{aligned} \mathbf{h}(\mathbf{x}) = & - \sum_j \sum_{k=0}^{K-1} \begin{pmatrix} \Re_k(\Phi_j) \\ \Re_k(\Psi_j) \end{pmatrix} \cdot \text{Re}\{F_k(C_j) - F_k(S_j)\} \\ & + \begin{pmatrix} \Im_k(\Phi_j) \\ \Im_k(\Psi_j) \end{pmatrix} \cdot \text{Im}\{F_k(C_j) - F_k(S_j)\} \end{aligned} \quad (\text{C.9})$$

## C.2 Homing Using Rotation Invariant Parameter

A rotation invariant method, which is based on the amplitude spectrum of the Fourier transformed images, was presented in Möller and Vardy (2006). It is referred to as [mffc\_abs].

The potential function is

$$p(\mathbf{x}) = \frac{1}{2} \sum_j \sum_{k=0}^{K-1} (|F_k(C_j(\mathbf{x}))| - |F_k(S_j)|)^2. \quad (\text{C.10})$$

and the home vector is given by

$$\mathbf{h}(\mathbf{x}) = - \sum_j \sum_{k=0}^{K-1} |F_k(C_j)|^{-1} \cdot \mathbf{A}_{jk} \cdot (|F_k(C_j)| - |F_k(S_j)|) \quad (\text{C.11})$$

where 
$$\mathbf{A}_{jk} = \begin{pmatrix} \Re_k(\Phi_j)\Re_k(C_j) + \Im_k(\Phi_j)\Im_k(C_j) \\ \Re_k(\Psi_j)\Re_k(C_j) + \Im_k(\Psi_j)\Im_k(C_j) \end{pmatrix}$$

### C.3 Parameter-Navigation, Additional Figures

The figures C.1 through C.6 depict the  $RR_*$  and  $AAE_*$ -Measures of the lower ranks from table 5.5, 5.6, and 5.7. The results are added for the sake of completeness, adapted from Fischer (2006).

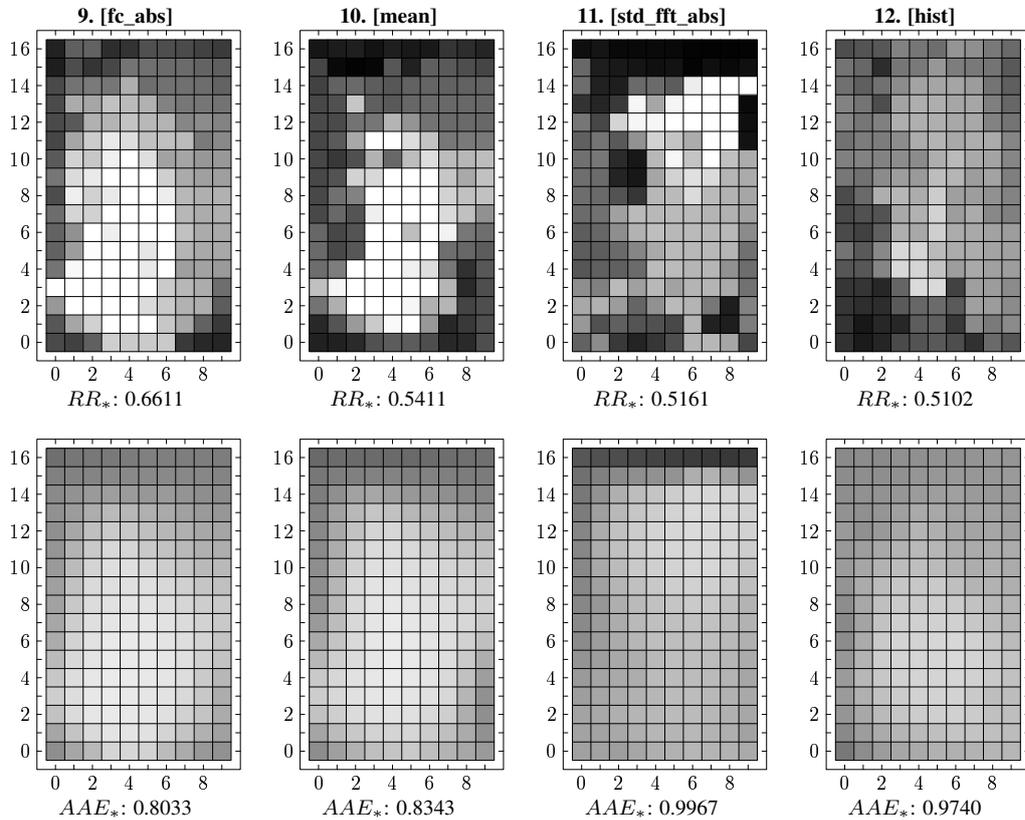


Figure C.1:  $RR_*$  and  $AAE_*$  values ranking 9 - 12 from **original** database. The figures are taken from Fischer (2006).

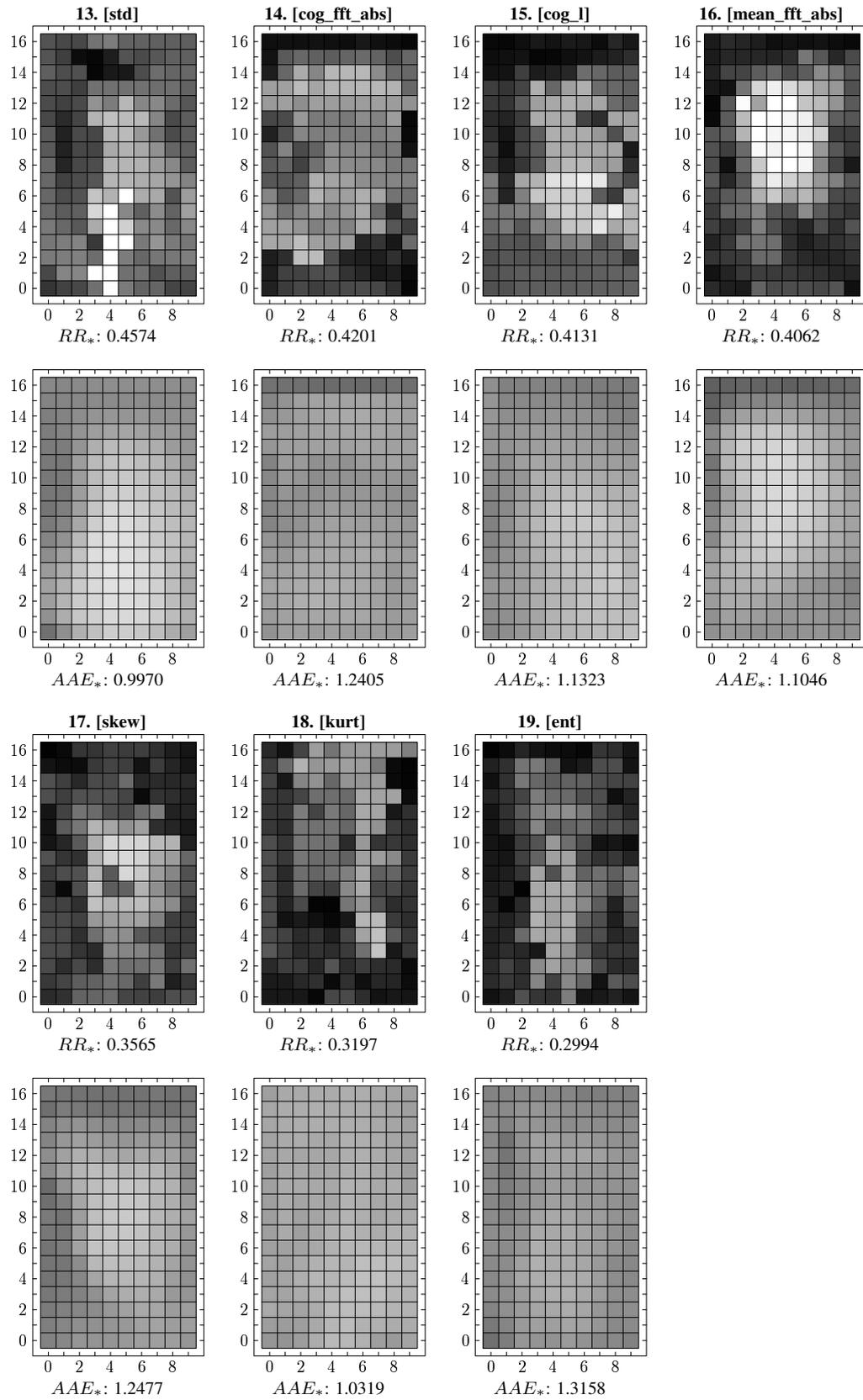


Figure C.2:  $RR_*$  and  $AAE_*$  values ranking 13 - 19 from **original** database. The figures are taken from Fischer (2006).

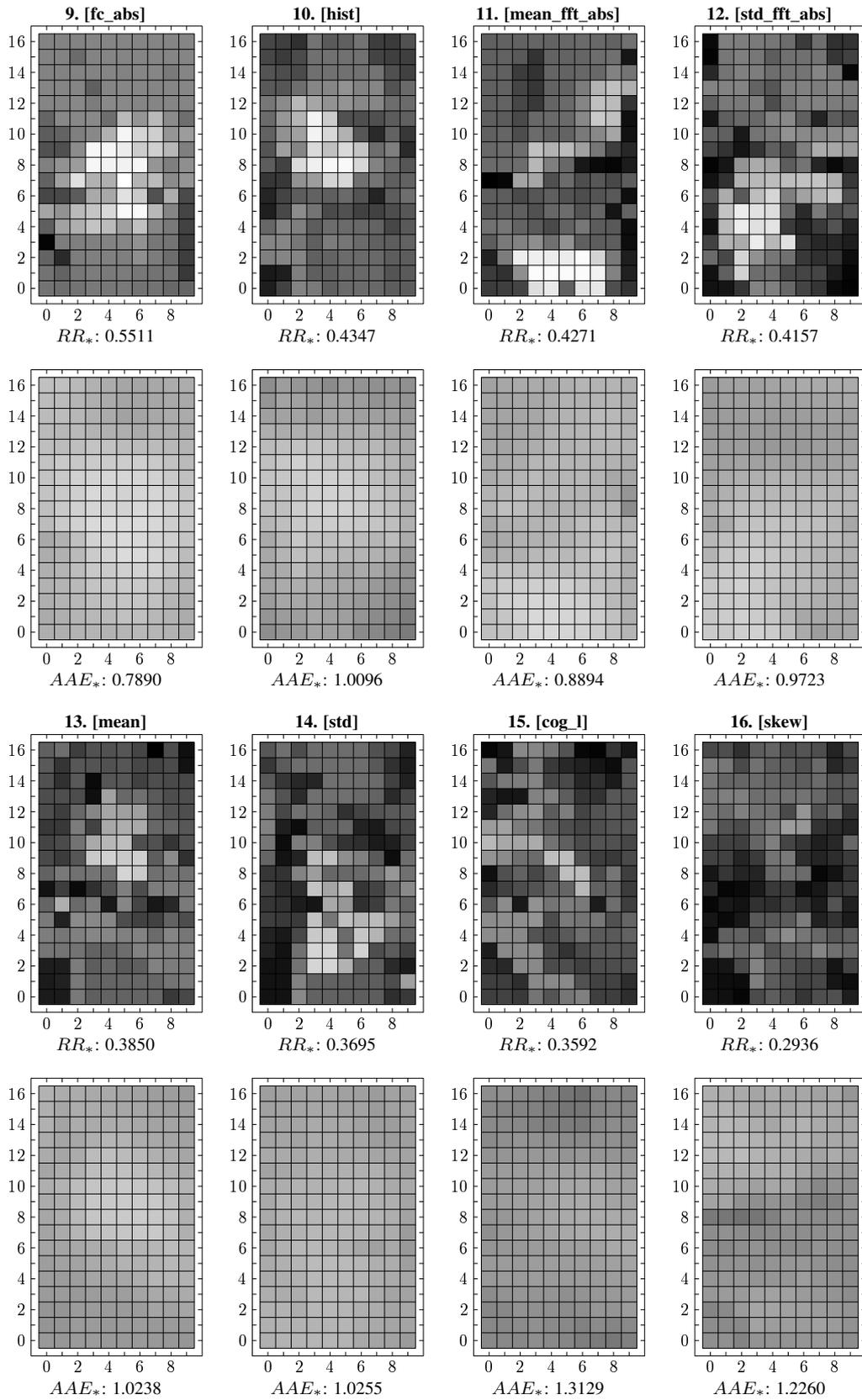


Figure C.3:  $RR_*$  and  $AAE_*$  values ranking 9 - 16 from **hall1** database. The figures are taken from Fischer (2006).

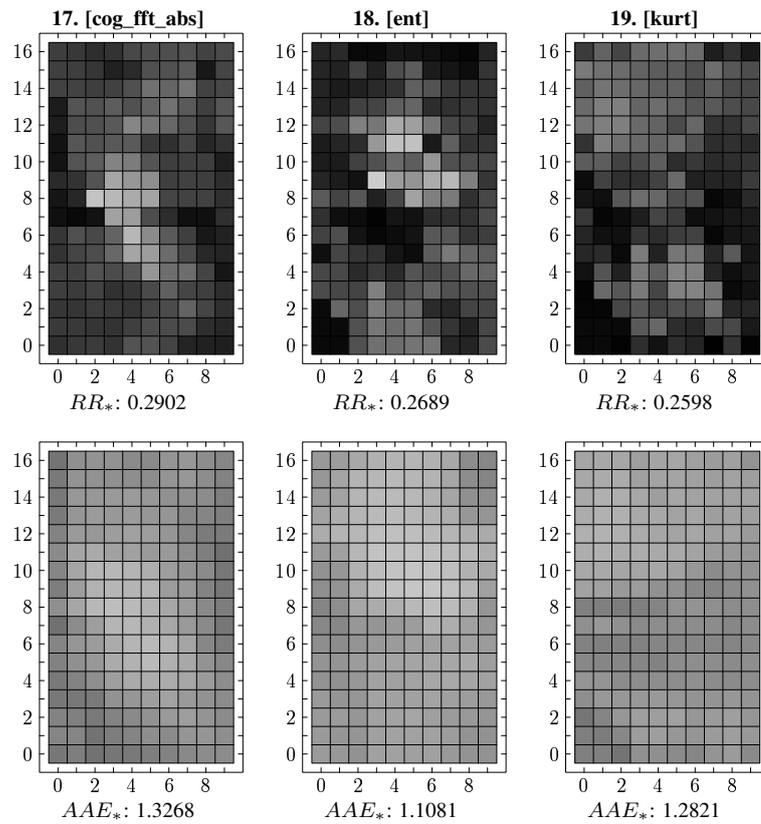


Figure C.4:  $RR_*$  and  $AAE_*$  values ranking 17 - 19 from **hall1** database. The figures are taken from Fischer (2006).

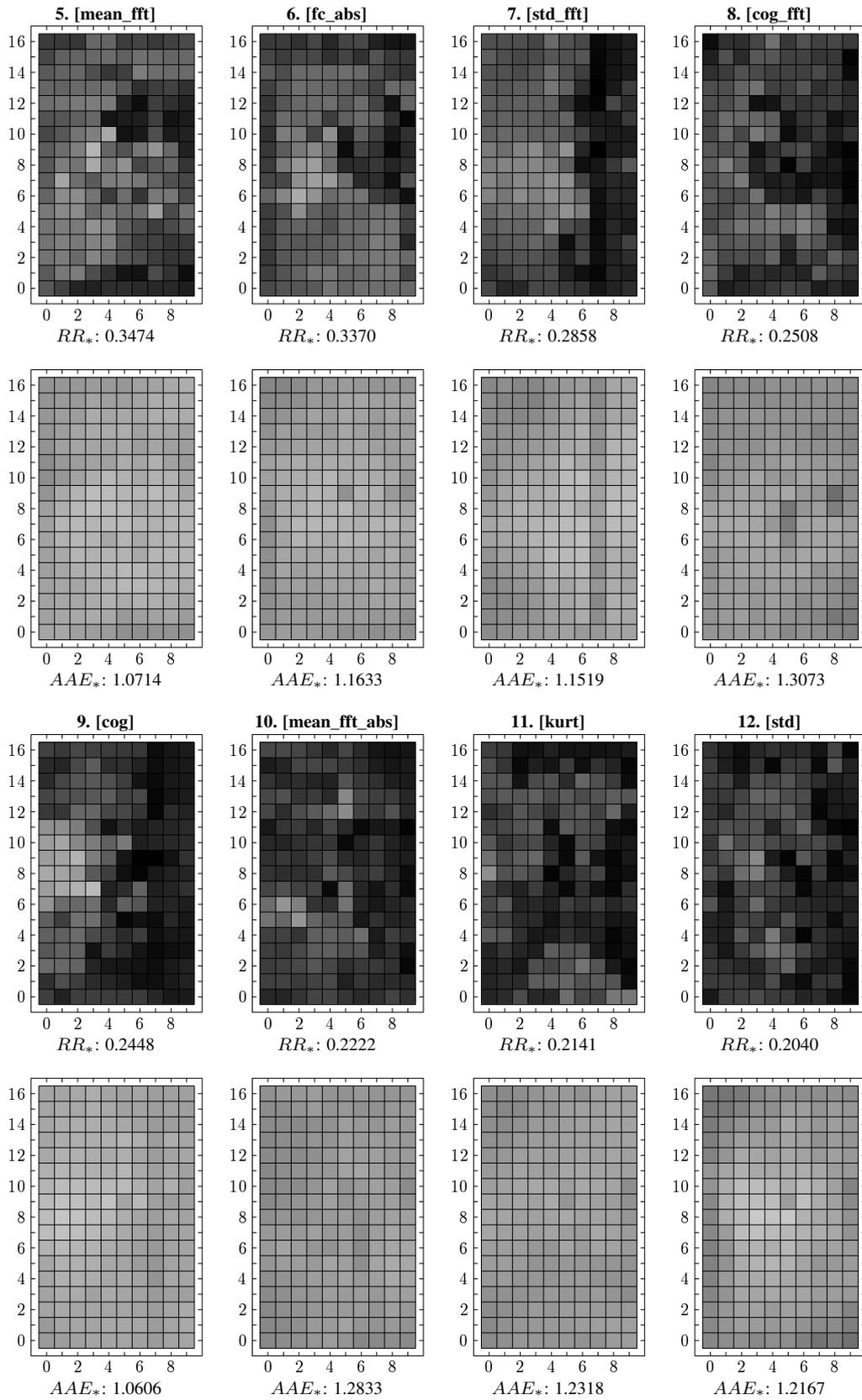


Figure C.5:  $RR_*$  and  $AAE_*$  values ranking 5 - 12 from **uni** database. The figures are taken from Fischer (2006).

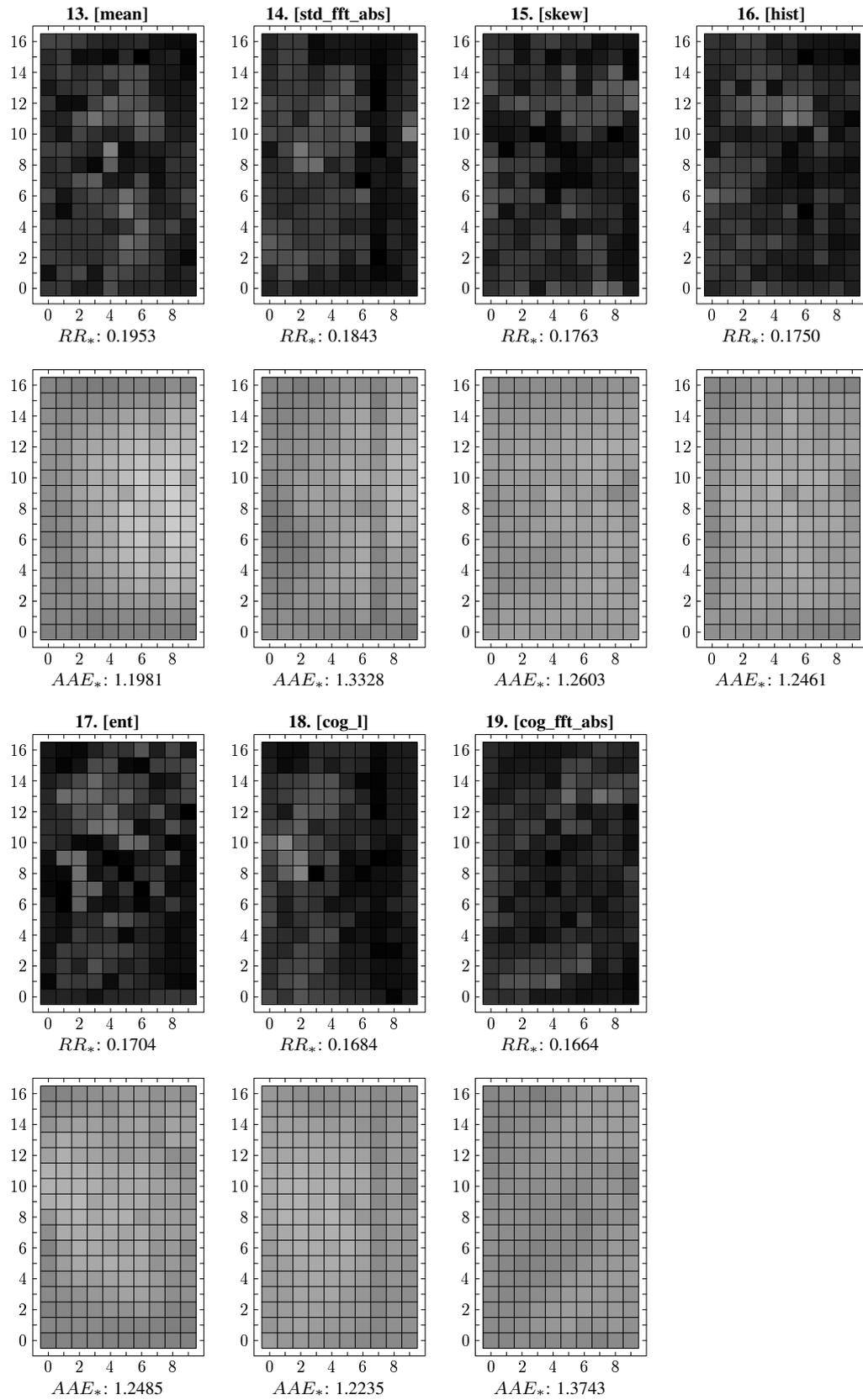


Figure C.6:  $RR_*$  and  $AAE_*$  values ranking 13 - 19 from **uni** database. The figures are taken from Fischer (2006).

## Appendix D

### Contrasts, Additional Figures

#### D.1 Natural Contrasts

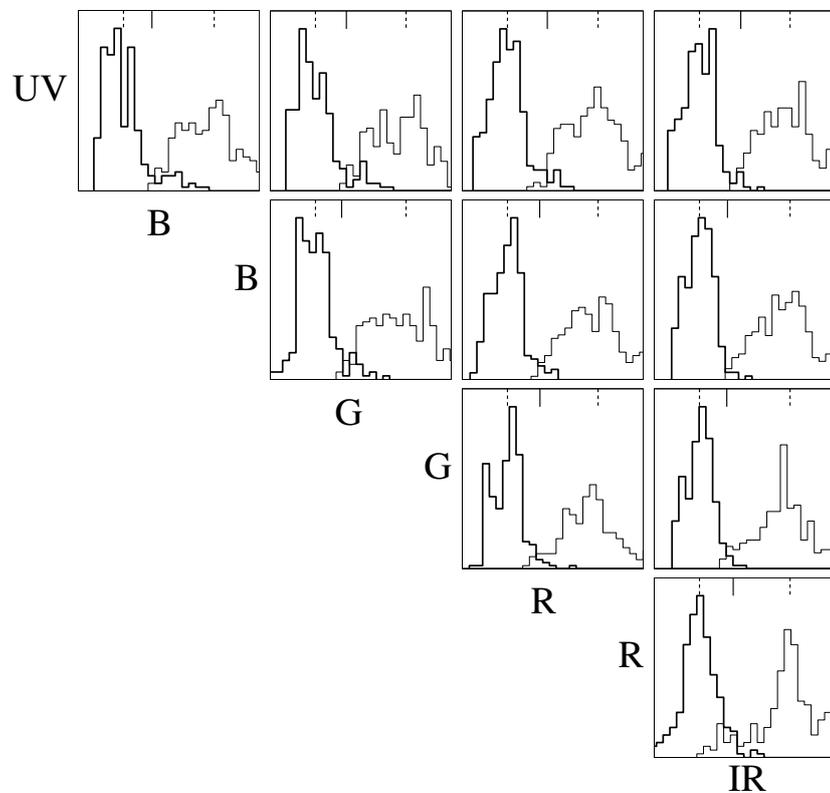


Figure D.1: Histograms of the projected natural data points; thick lines: sky, thin lines: objects. The histograms are normalized such that the projected means are at the same position in the histograms. The class border is marked by a solid line at the top margin. For each histogram, the bin size was computed by dividing the range of the projected data by  $10 \log_{10} N$ , with  $N$  being the total number of data points.

## D.2 Artificial Contrasts

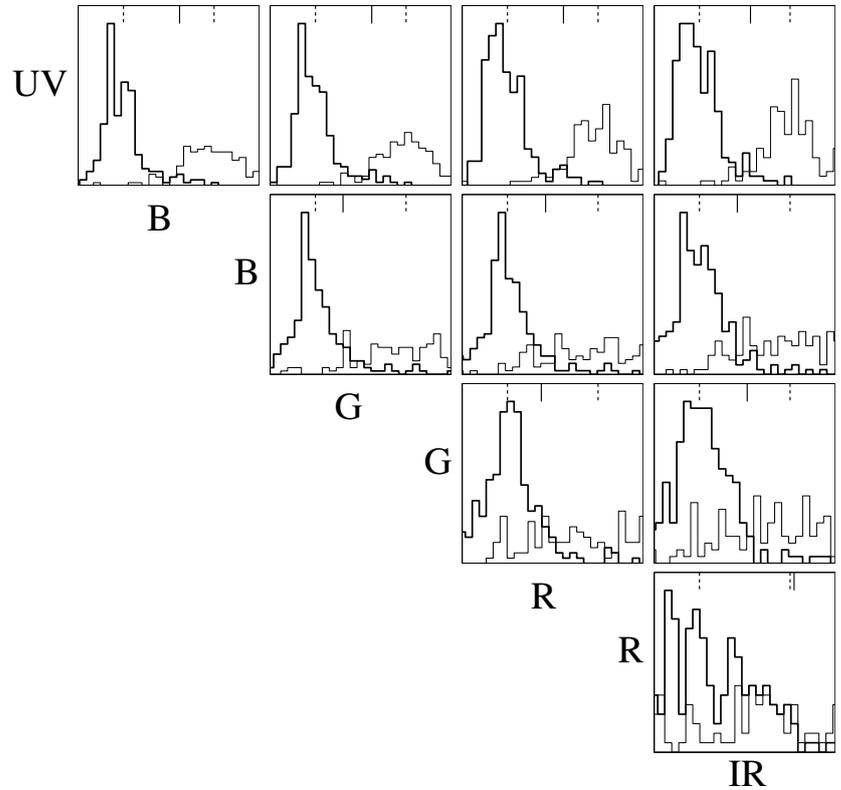


Figure D.2: Histograms of the projected artificial data points; thick lines: sky, thin lines: objects. Again, the histograms are normalized such that the projected means are at the same position in the histograms. The class border is marked by a solid line at the top margin. For each histogram, the bin size was computed by dividing the range of the projected data by  $10 \log_{10} N$ , with  $N$  being the total number of data points.

### D.3 2D-Contrast Measures

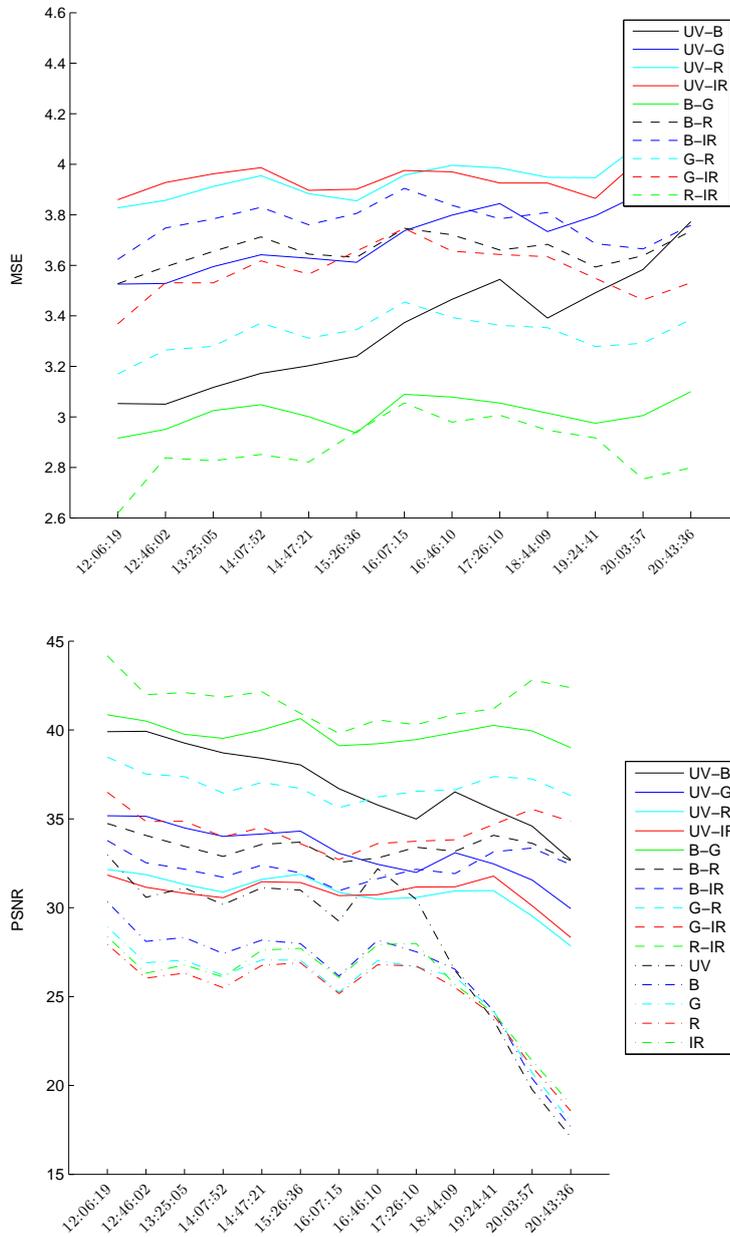


Figure D.3: The MSE and the peak signal-to-noise ratio (PSNR) measure plotted over a one day time period. The measures are computed as explained in equation 6.1 and 6.4.

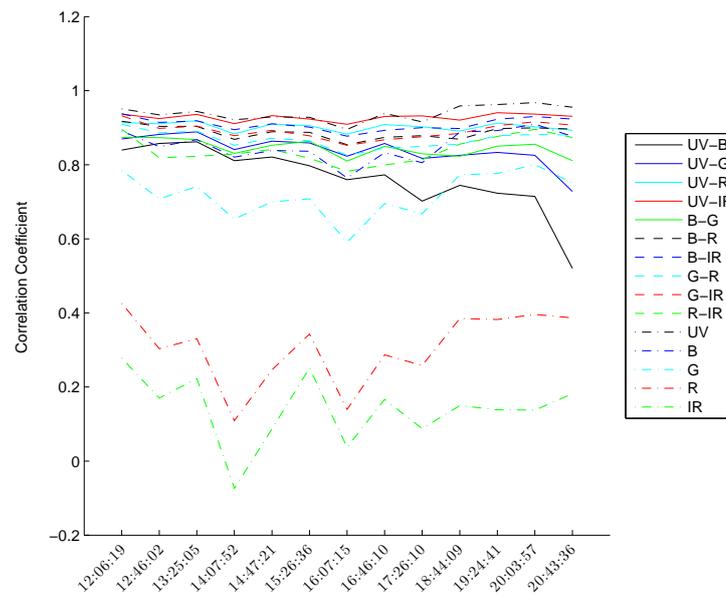


Figure D.4: The correlation coefficient (CC) is presented. The measure is computed as explained in 6.5.

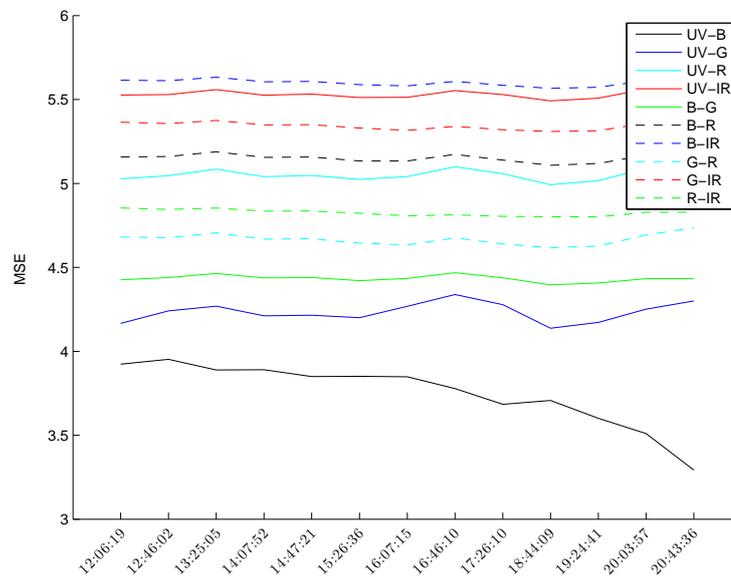


Figure D.5: The MSE plotted over a one day time period. The contrasts are linear separated by the  $J_\sigma$  criterion. The measure is computed as explained in equation 6.1.

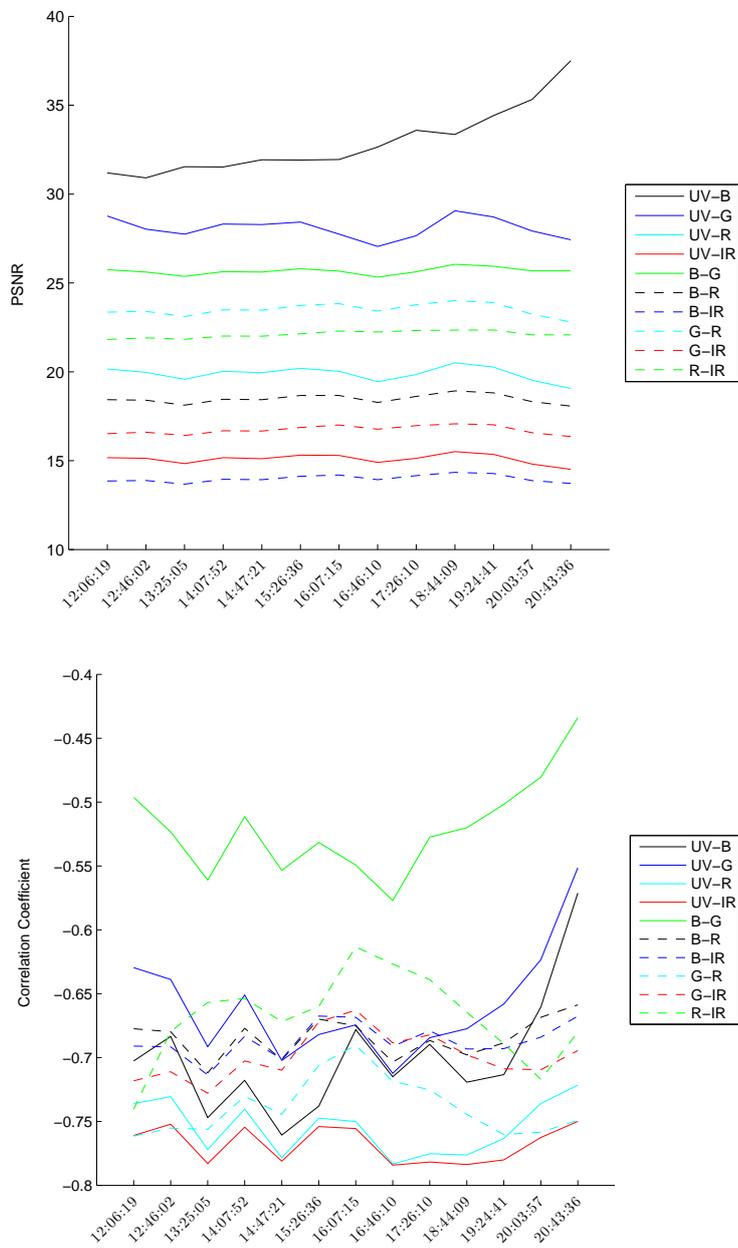


Figure D.6: The peak signal-to-noise ratio (PSNR) and the correlation coefficient (CC) are presented. The contrasts are linear separated by the  $J_\sigma$  criterion. The measures are computed as explained in equation 6.4 and 6.5.

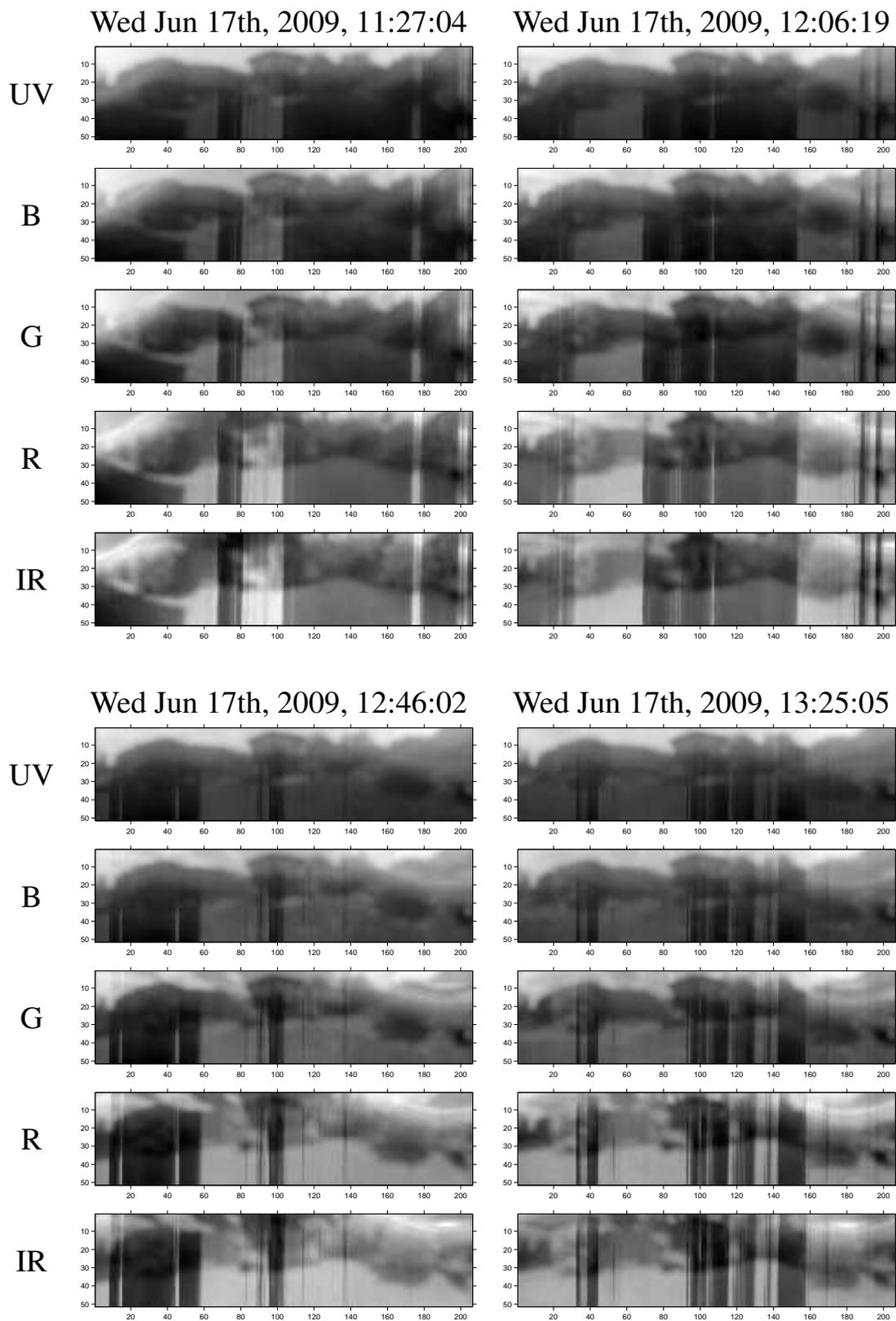


Figure D.7: Raw image data of the one-day dataset at different times. The data was collected on Wed Jun 17th, 2009, at 11:27:04, 12:06:19, 12:46:02, and 13:25:05.

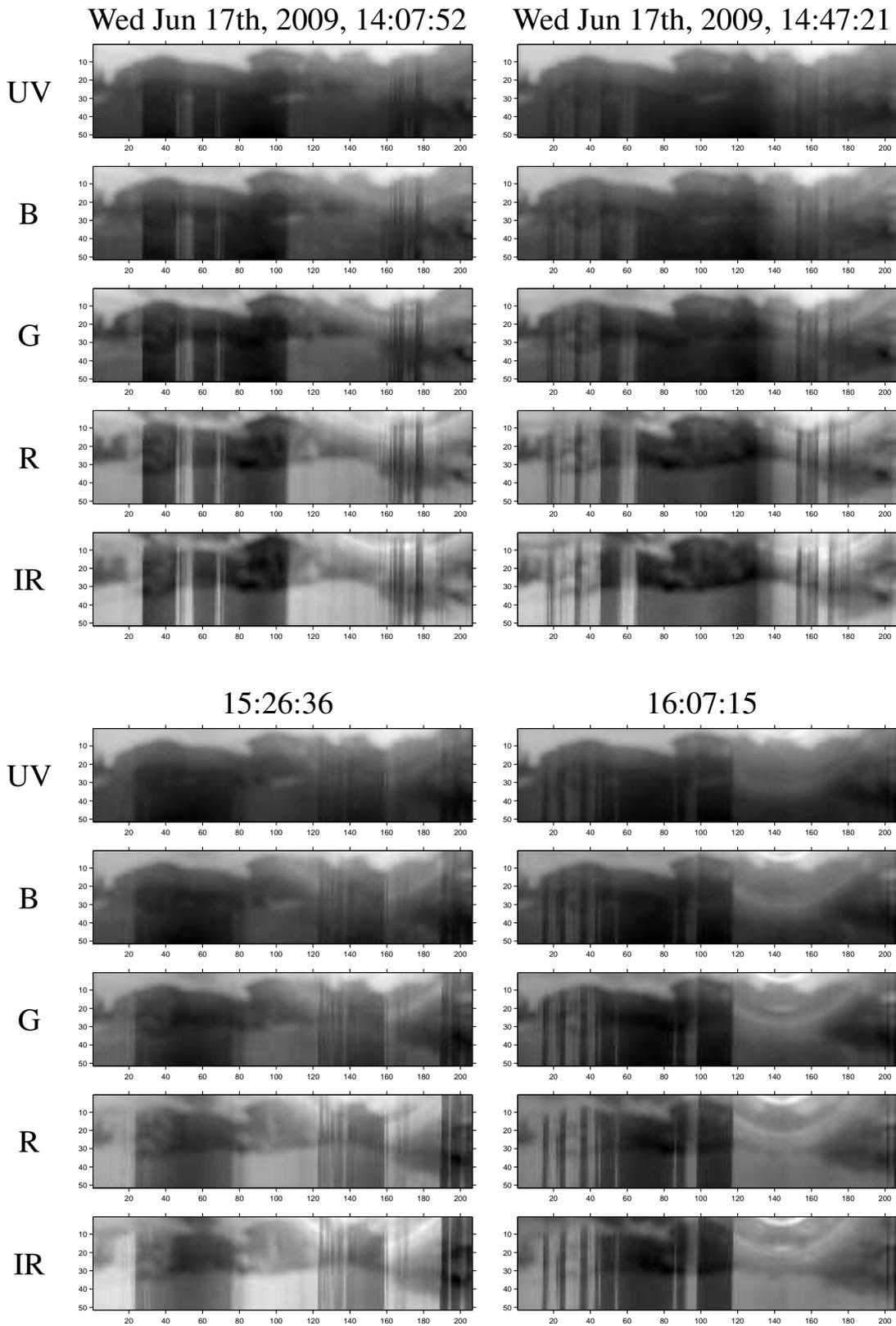


Figure D.8: Raw image data of the one-day dataset collected on Wed Jun 17th, 2009, 14:07:52, 14:47:21, 15:26:36, and 16:07:15.

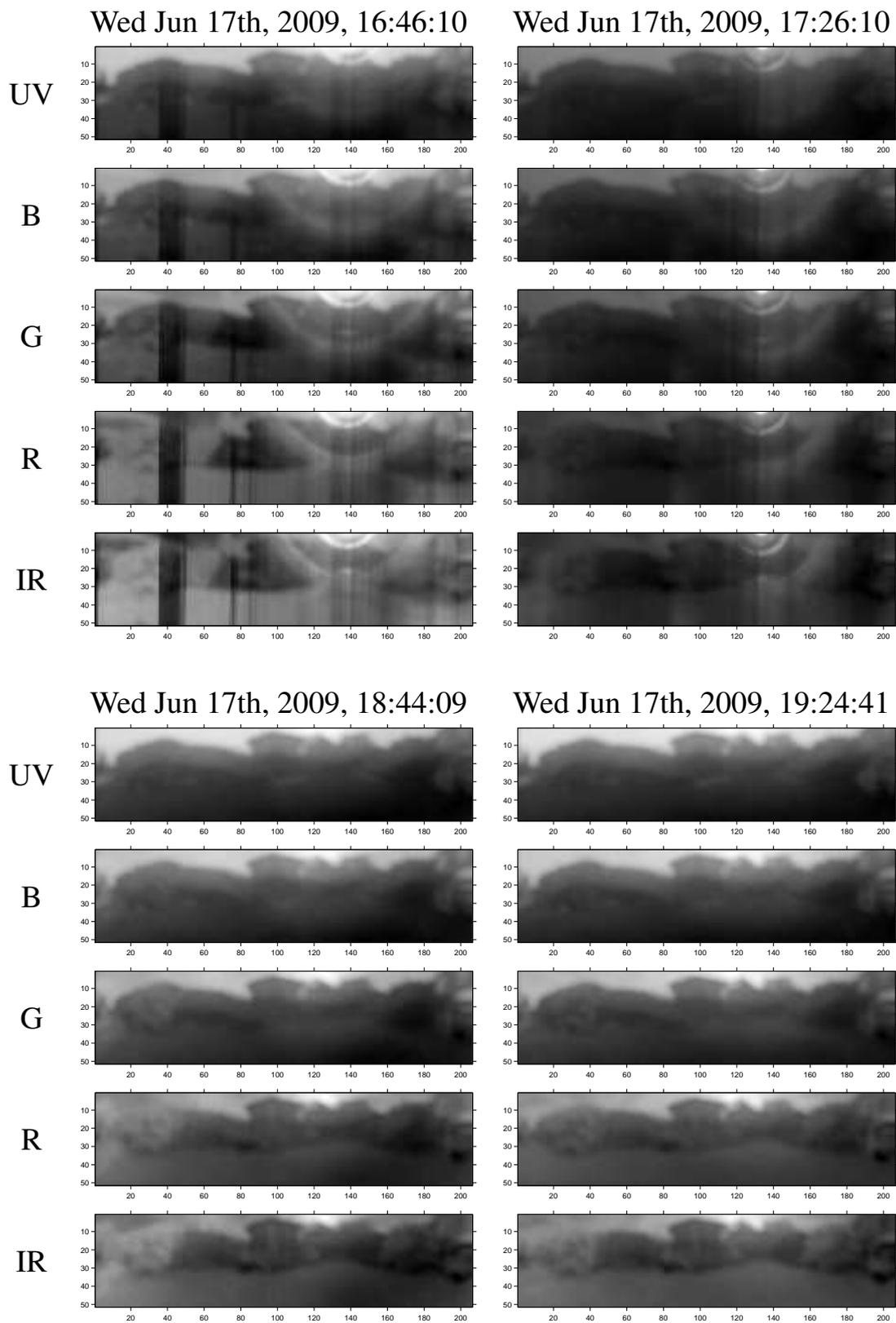


Figure D.9: Raw image data of the one-day dataset collected on Wed Jun 17th, 2009, 16:46:10, 17:26:10, 18:44:09, and 19:24:41.

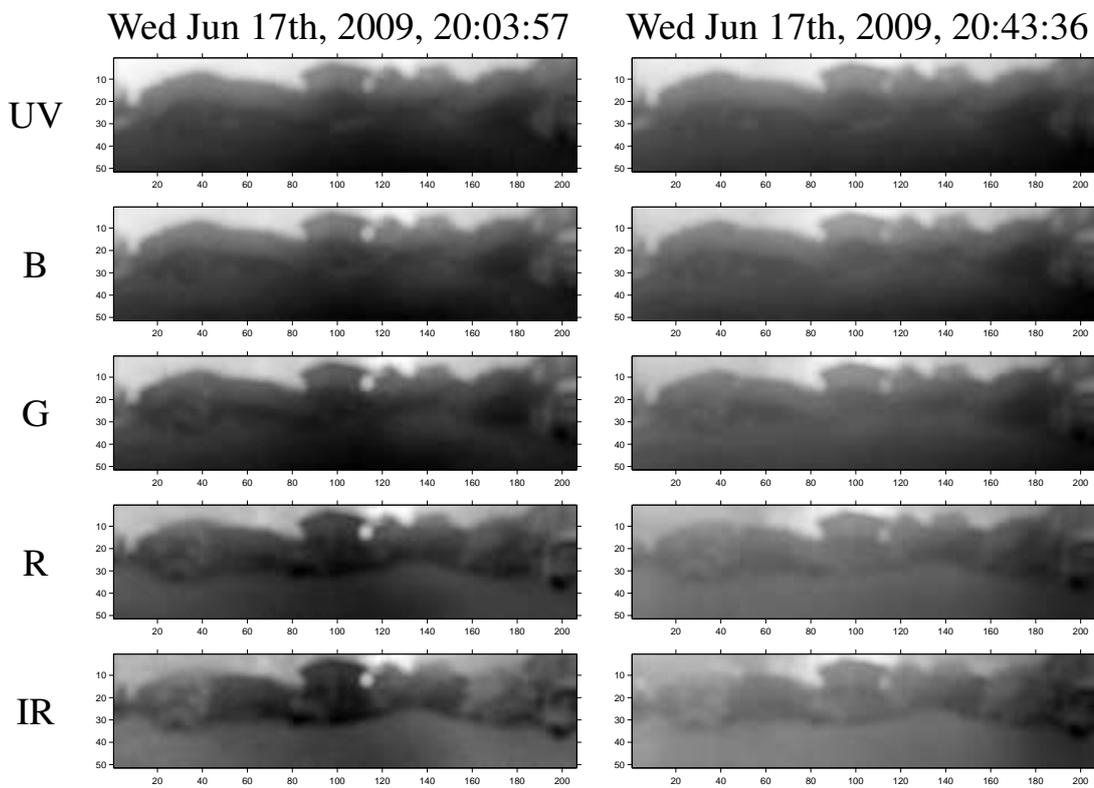


Figure D.10: Raw image data of the one-day dataset collected on Wed Jun 17th, 2009, 20:03:57, and 20:43:36.

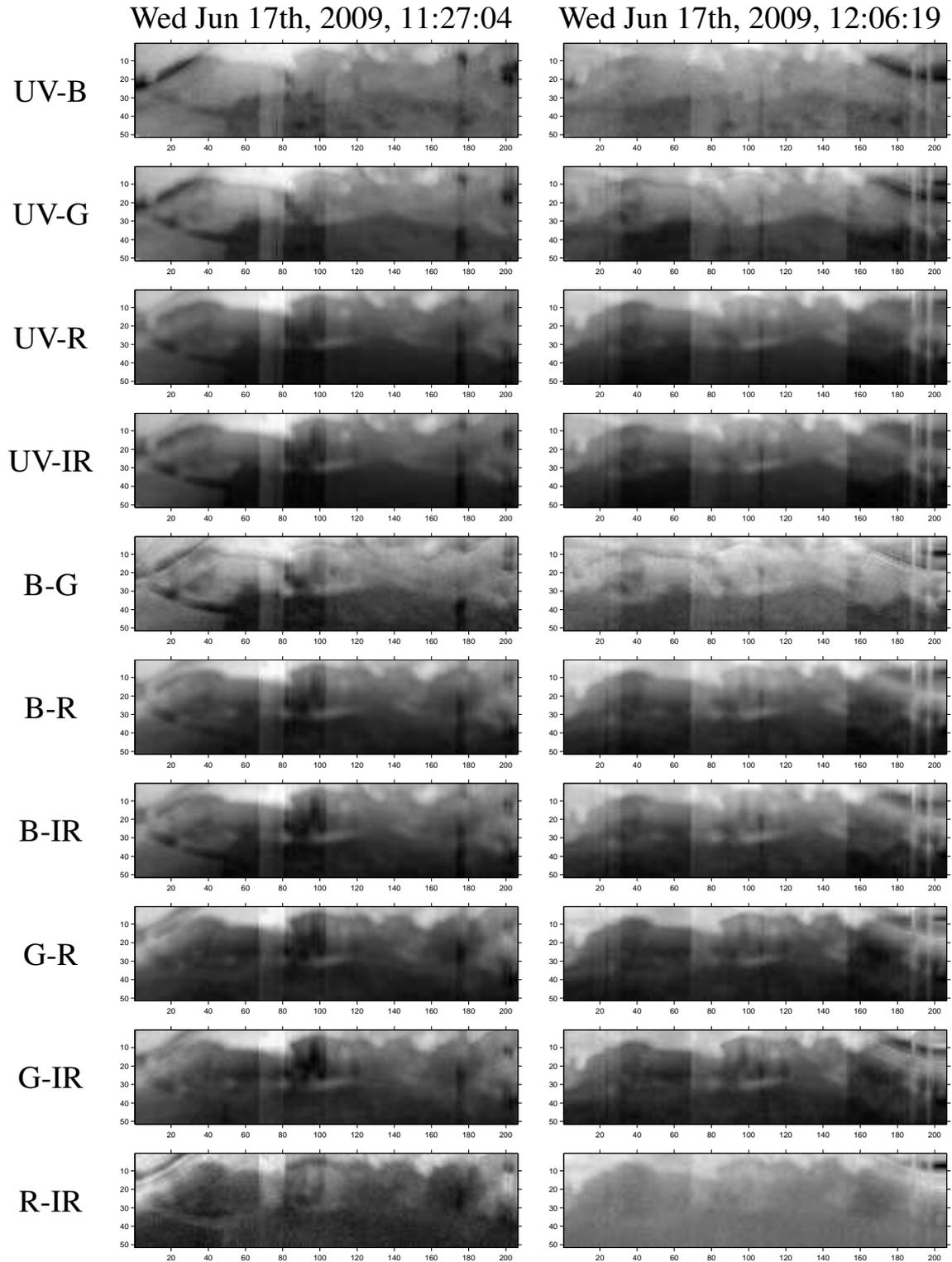


Figure D.11: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 11:27:04 and Wed Jun 17th, 2009, 12:06:19.

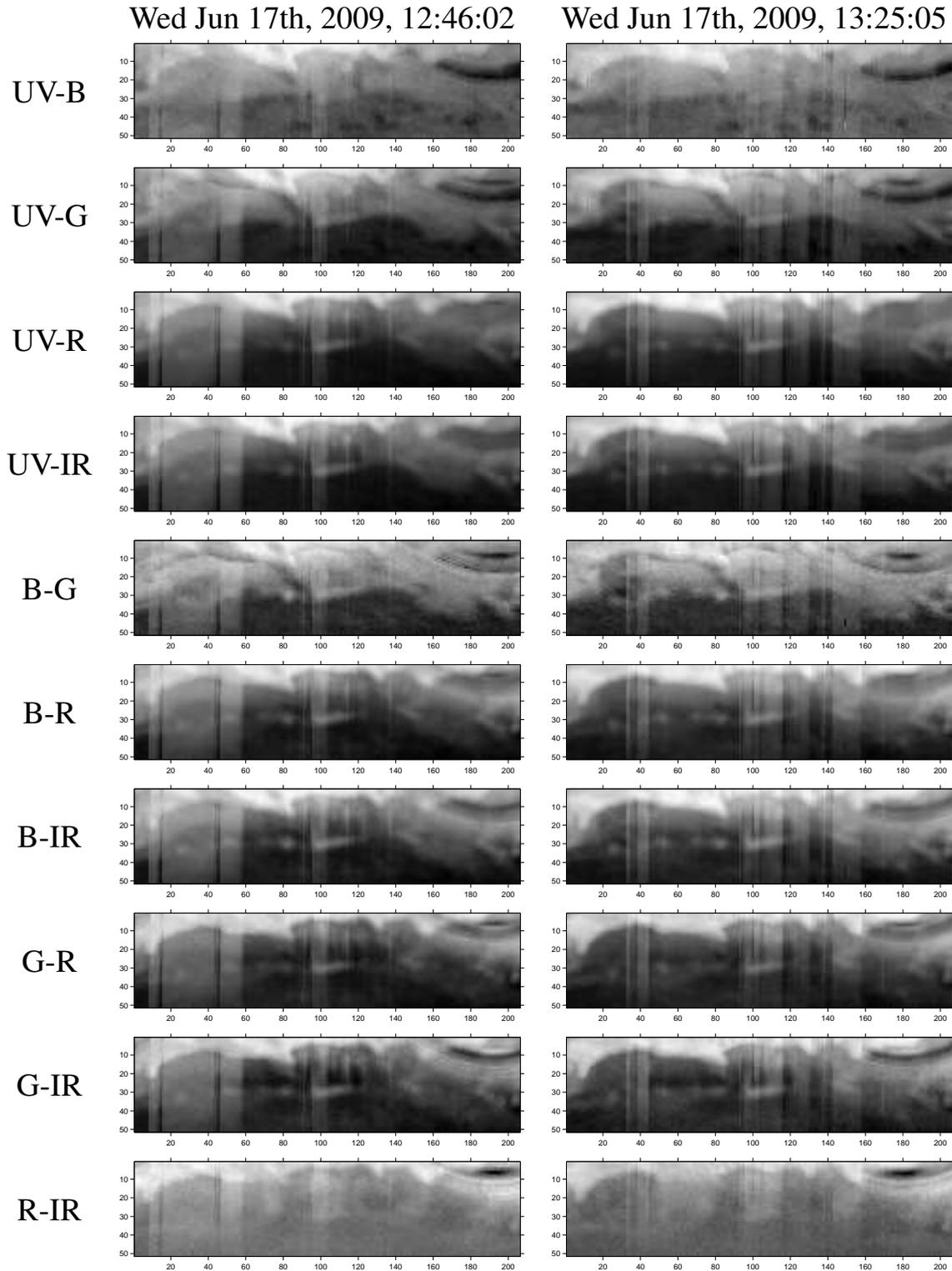


Figure D.12: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 12:46:02 and Wed Jun 17th, 2009, 13:25:05.

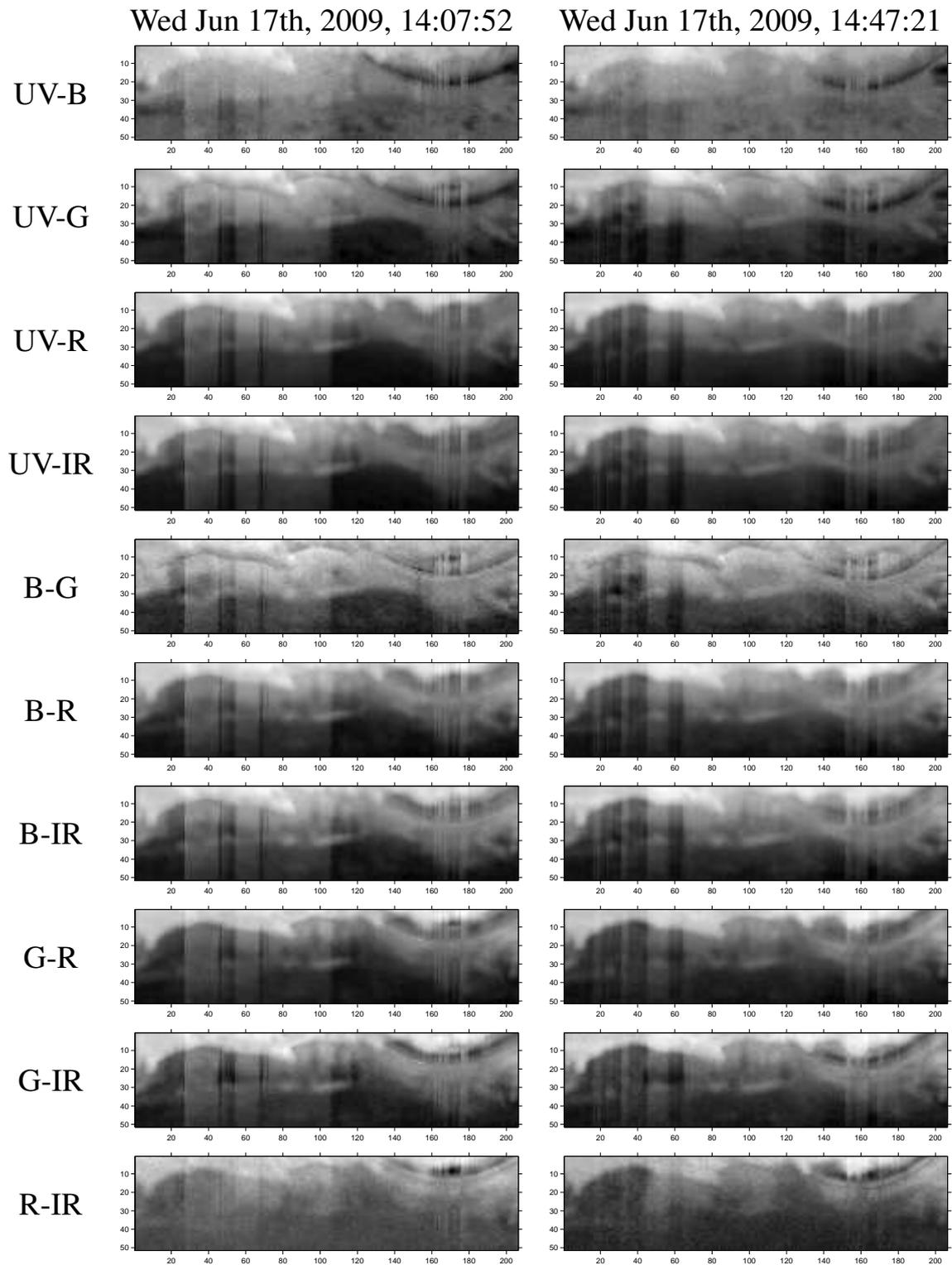


Figure D.13: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 14:07:52 and Wed Jun 17th, 2009, 14:47:21.

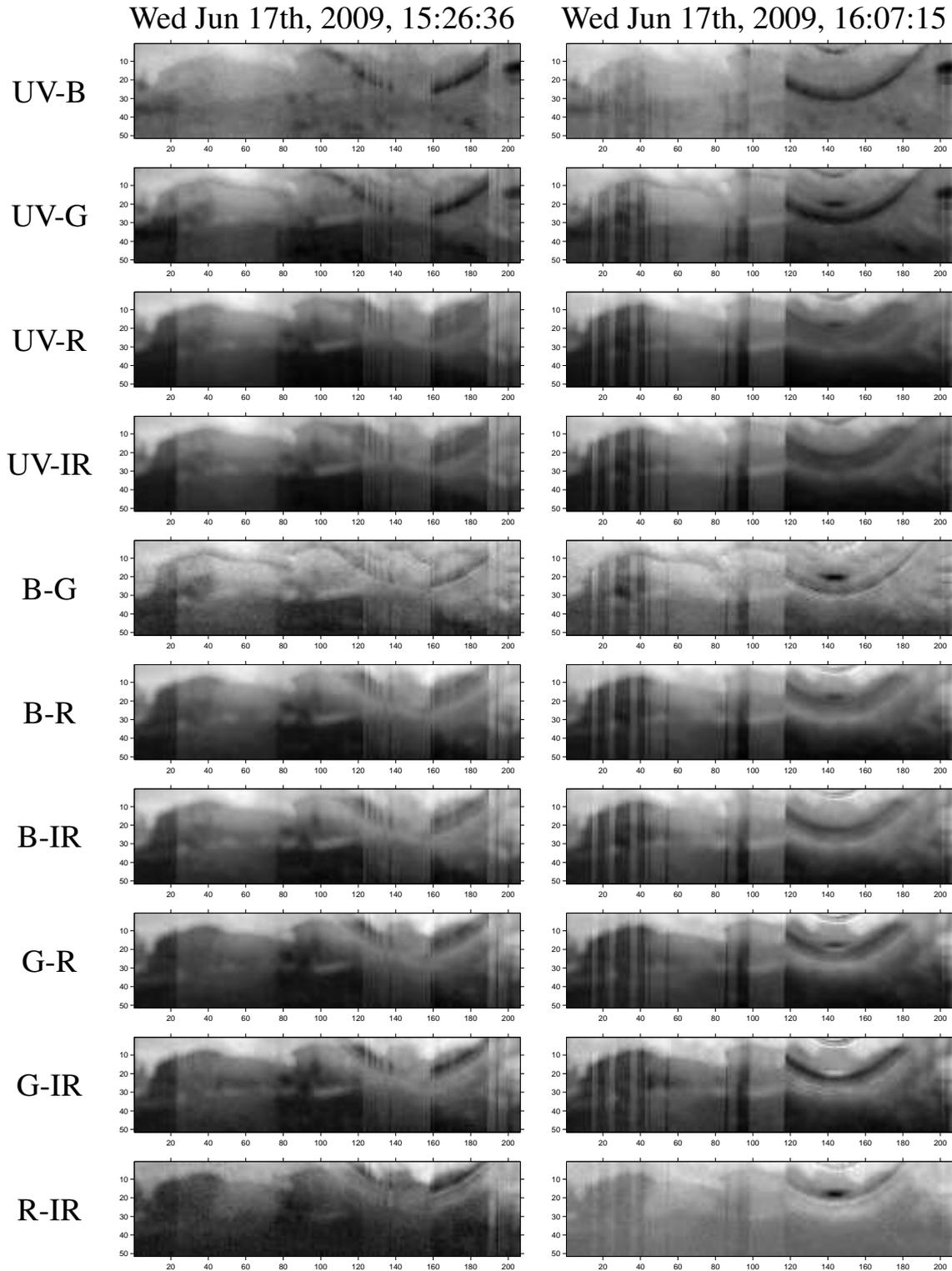


Figure D.14: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 15:26:36 and Wed Jun 17th, 2009, 16:07:15.

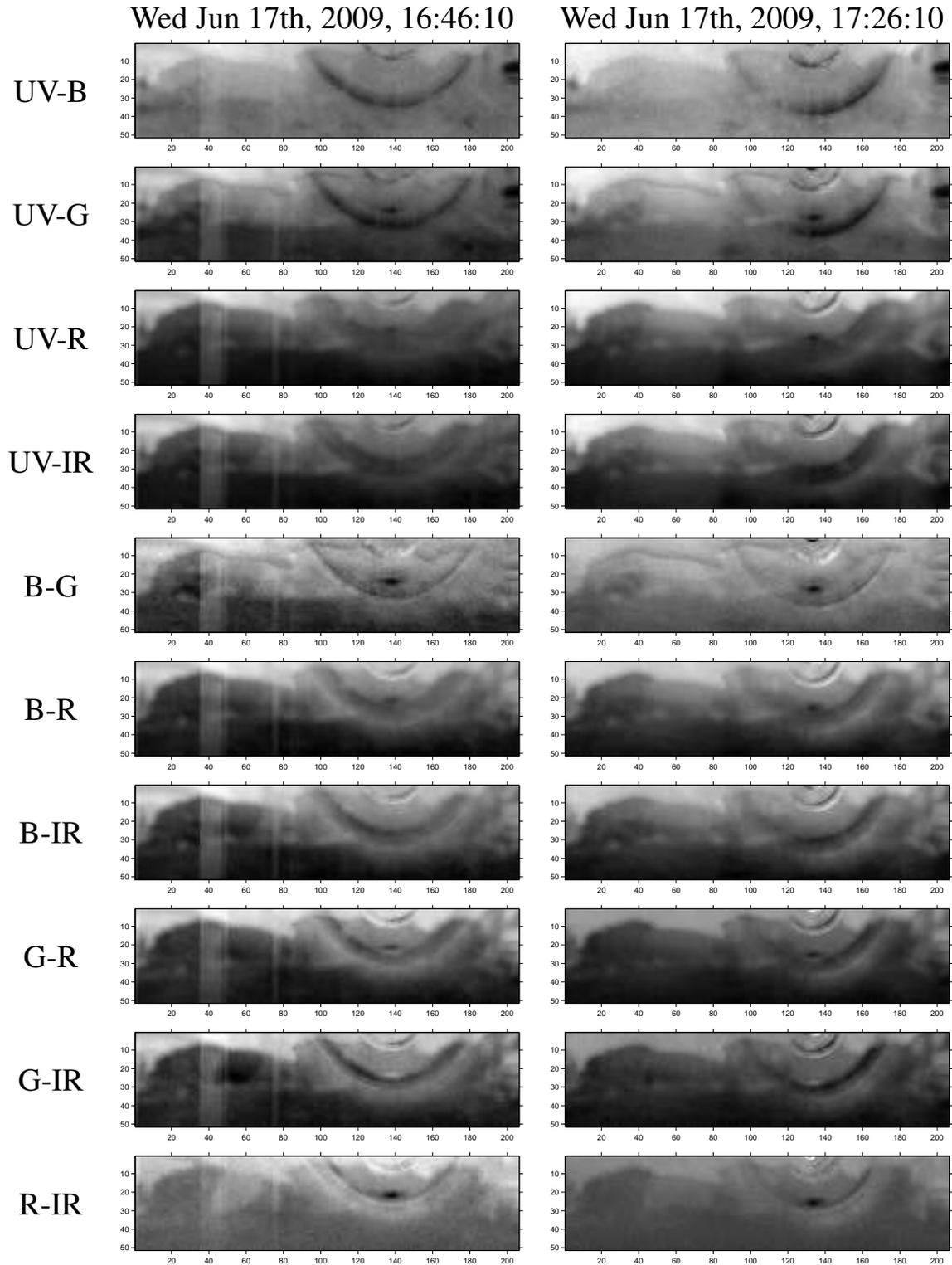


Figure D.15: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 16:46:10 and Wed Jun 17th, 2009, 17:26:10.

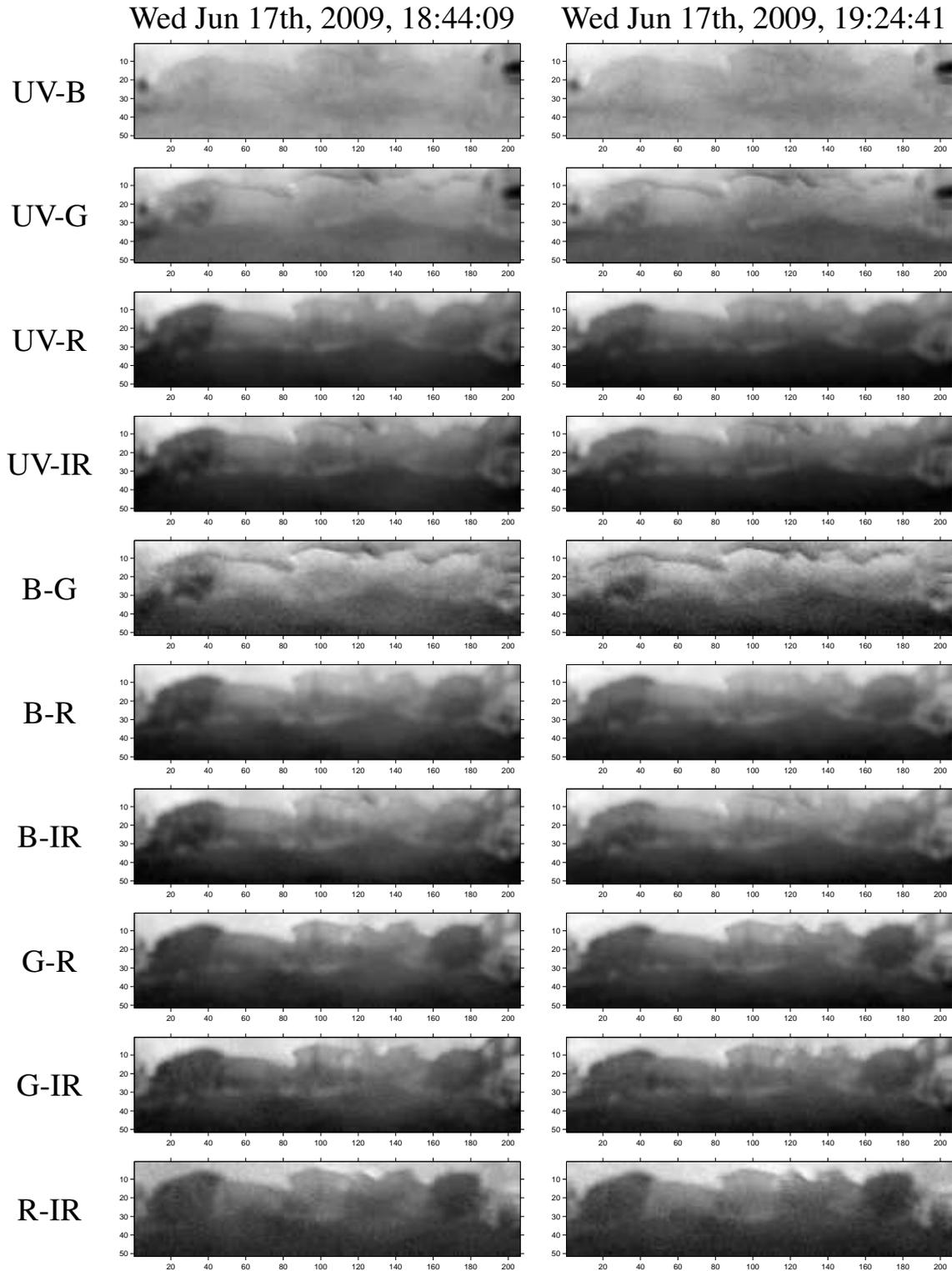


Figure D.16: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 18:44:09 and Wed Jun 17th, 2009, 19:24:41.

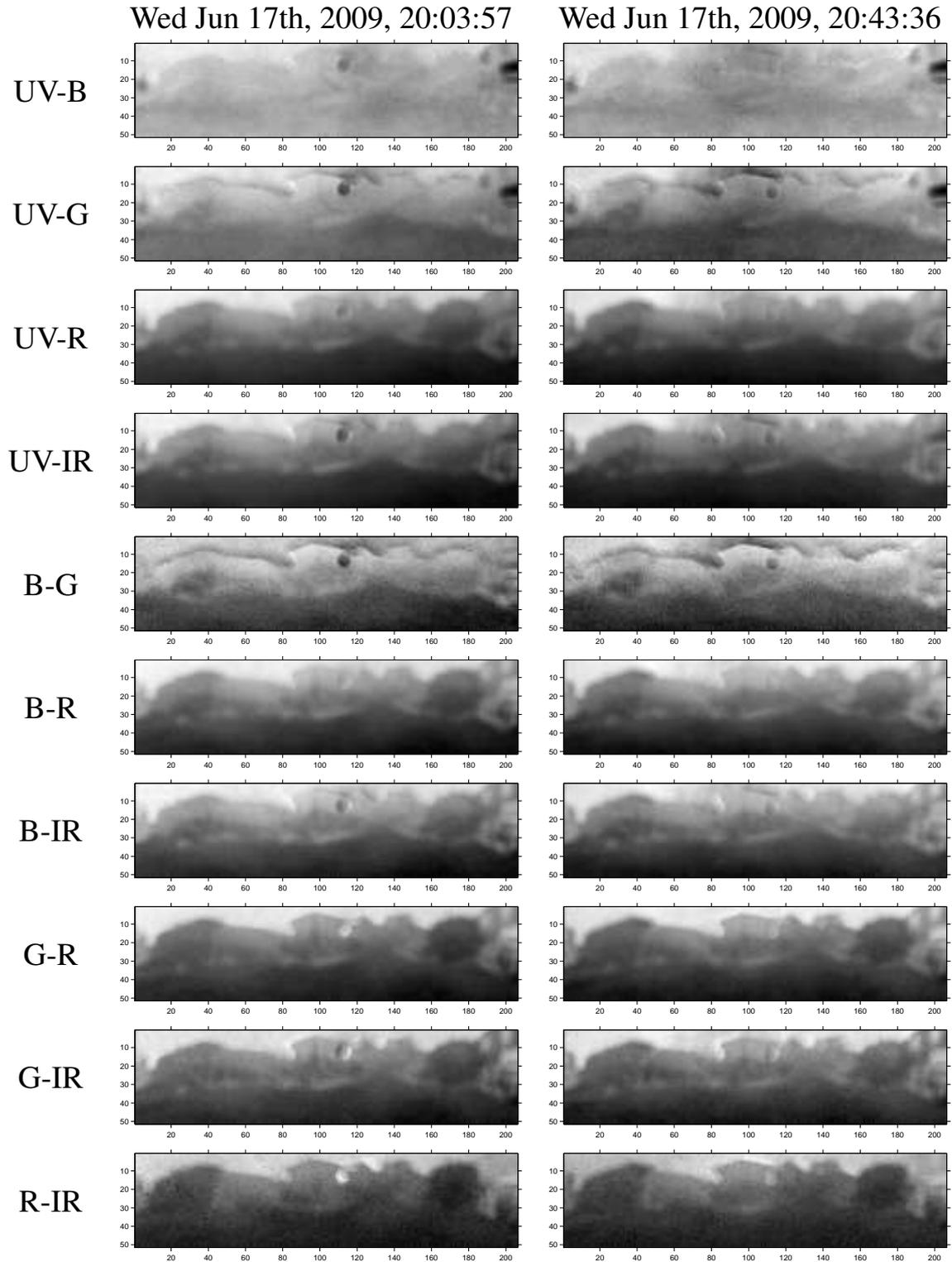


Figure D.17: Dual-channel contrasts of the one-day datasets Wed Jun 17th, 2009, 20:03:57 and Wed Jun 17th, 2009, 20:43:36.

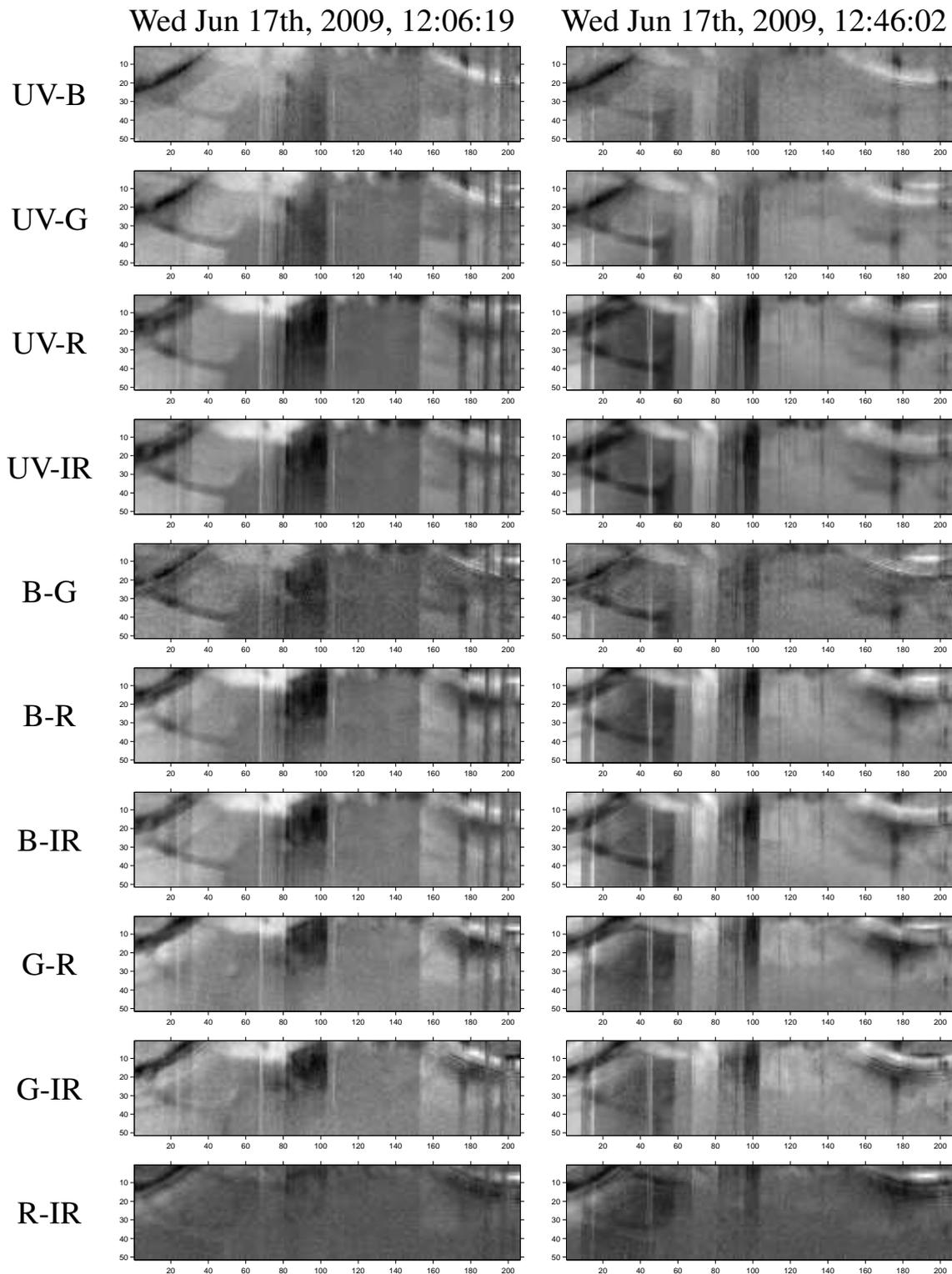


Figure D.18: Difference of dual-channel contrasts. The data collected at 11:27:04 was used as reference, such that each dual-channel contrast of the one-day dataset Wed Jun 17th, 2009, 12:06:19 and 12:46:02 were subtracted from the corresponding dual-channel reference contrast.

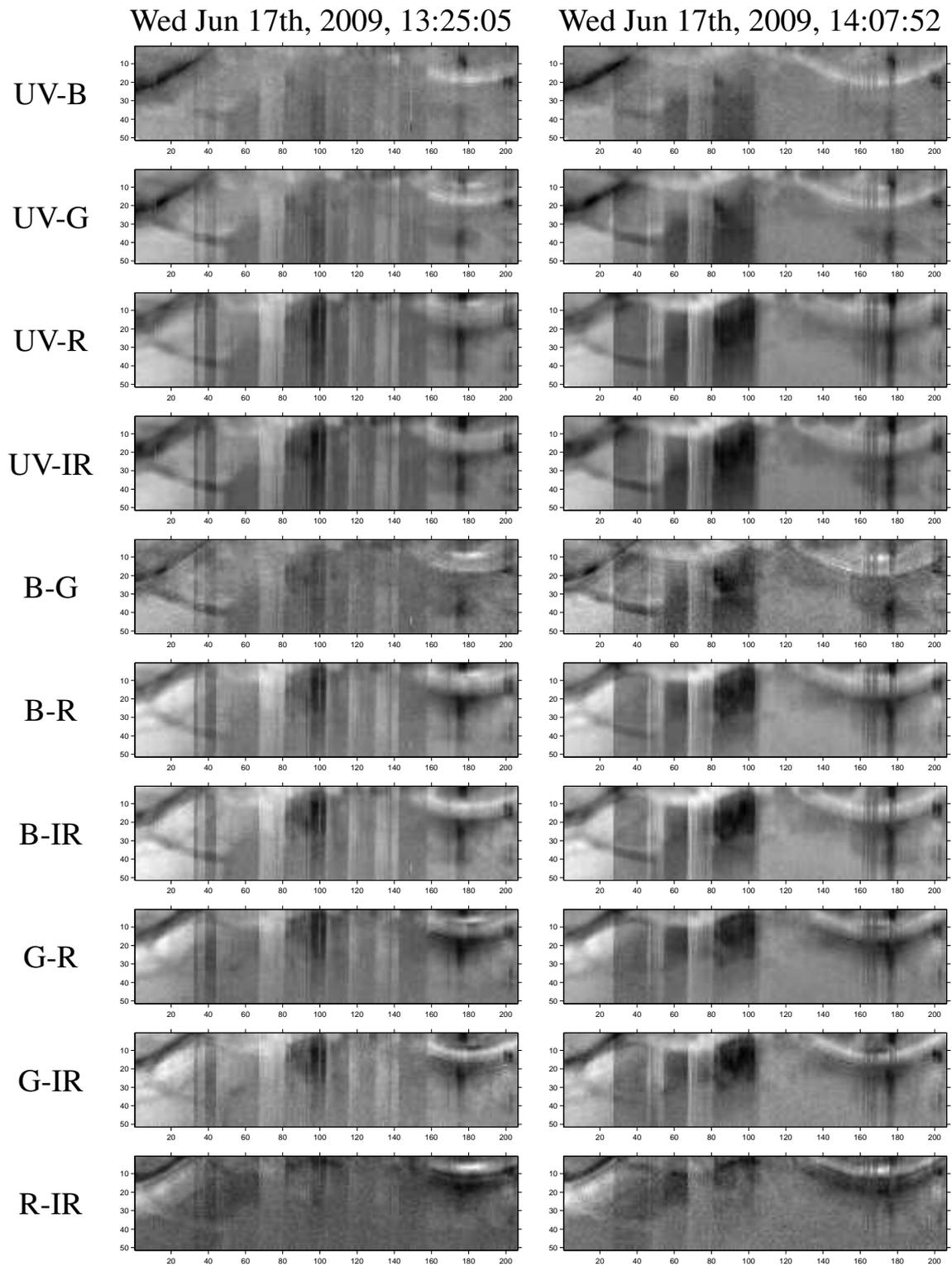


Figure D.19: Difference of dual-channel contrasts. The data collected at 11:27:04 was used as reference, such that each dual-channel contrast of the one-day dataset Wed Jun 17th, 2009, 13:25:0 and 14:07:52 were subtracted from the corresponding dual-channel reference contrast.

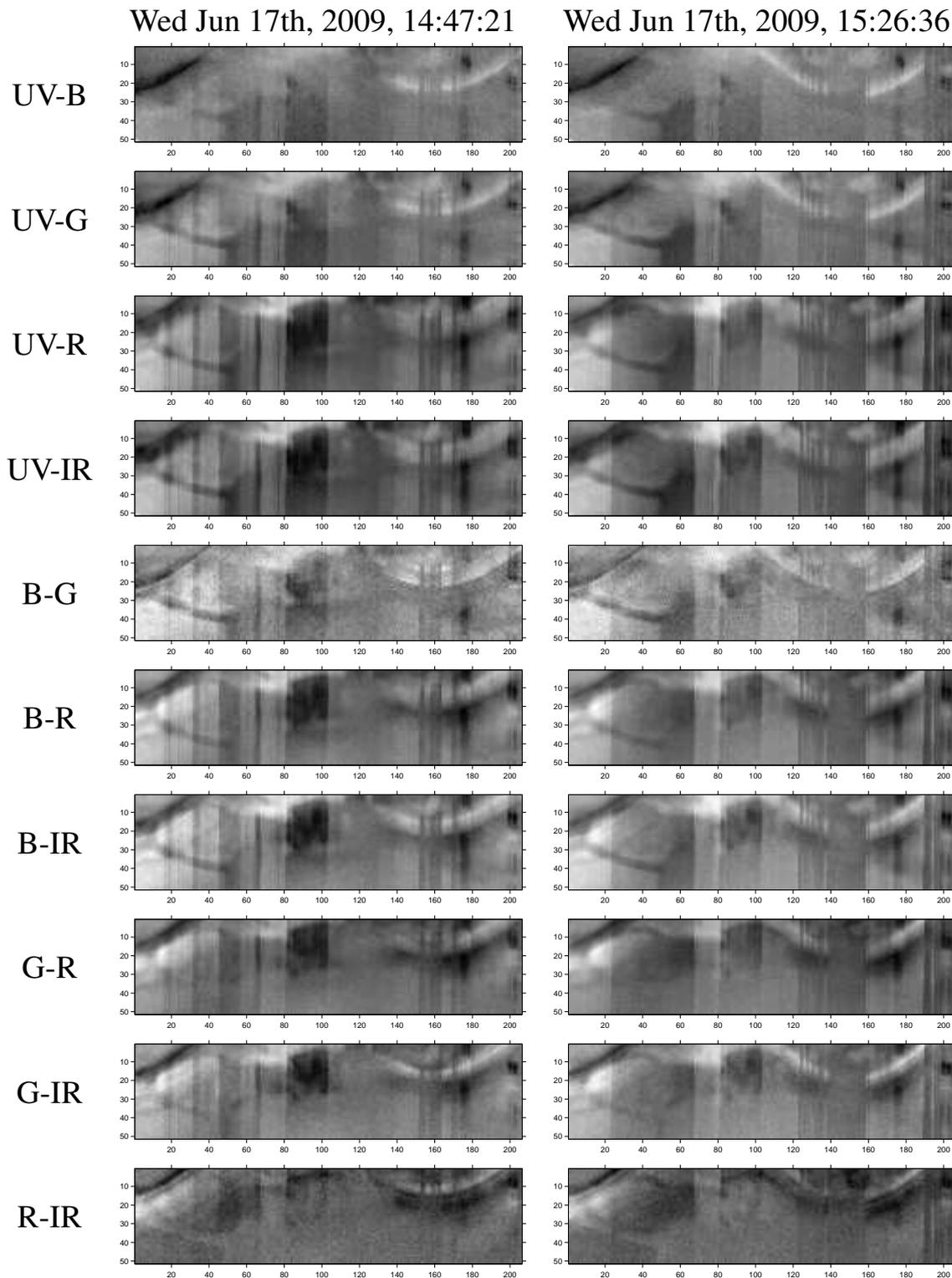


Figure D.20: Difference of dual-channel contrasts. The data collected at 11:27:04 was used as reference, such that each dual-channel contrast of the one-day dataset Wed Jun 17th, 2009, 14:47:21 and 15:26:36 were subtracted from the corresponding dual-channel reference contrast.

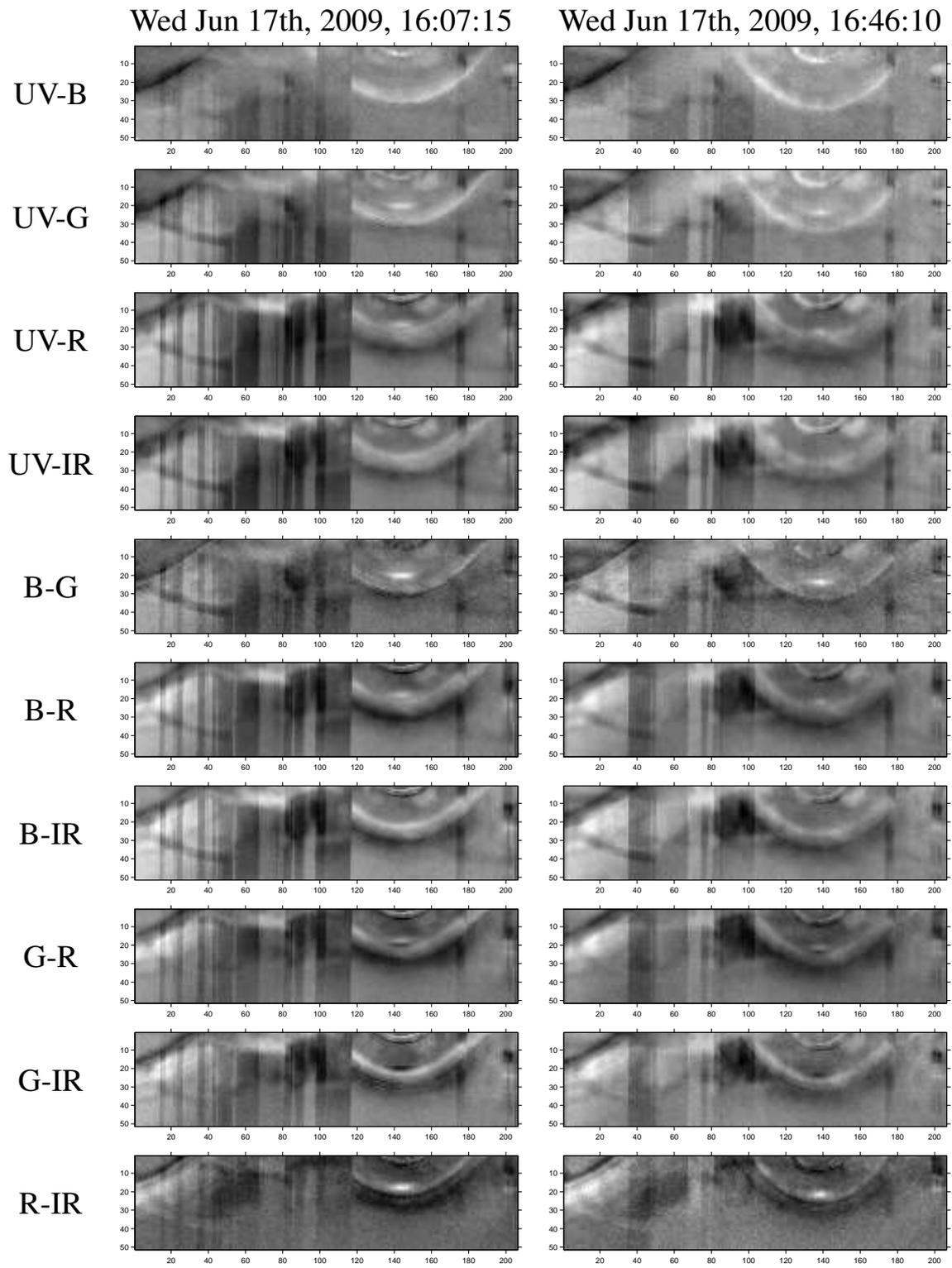


Figure D.21: Difference of dual-channel contrasts. The data collected at 11:27:04 was used as reference, such that each dual-channel contrast of the one-day dataset Wed Jun 17th, 2009, 16:07:1 and 16:46:10 were subtracted from the corresponding dual-channel reference contrast.

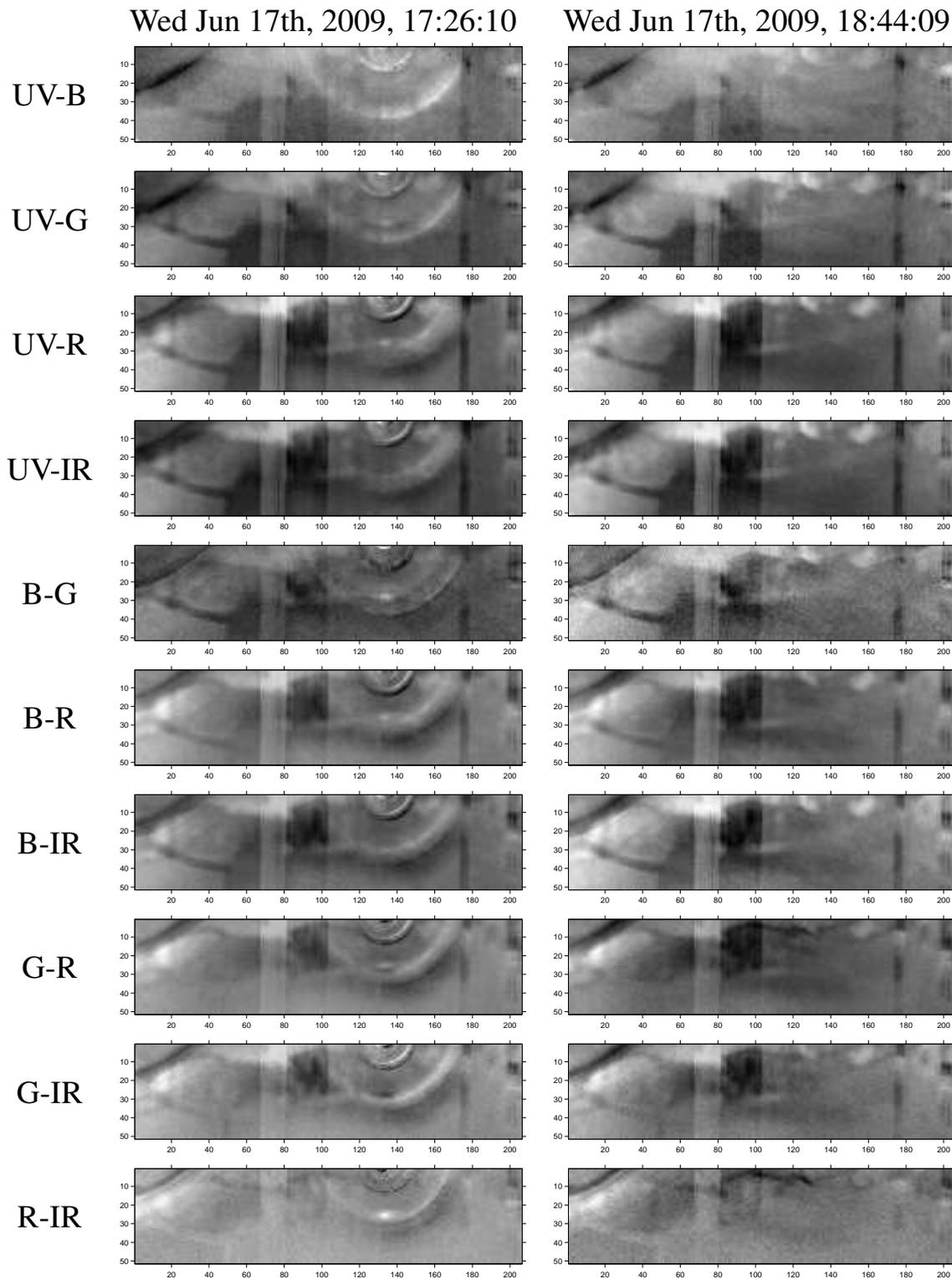


Figure D.22: Difference of dual-channel contrasts. The data collected at 11:27:04 was used as reference, such that each dual-channel contrast of the one-day dataset Wed Jun 17th, 2009, 17:26:10 and 18:44:09 were subtracted from the corresponding dual-channel reference contrast.

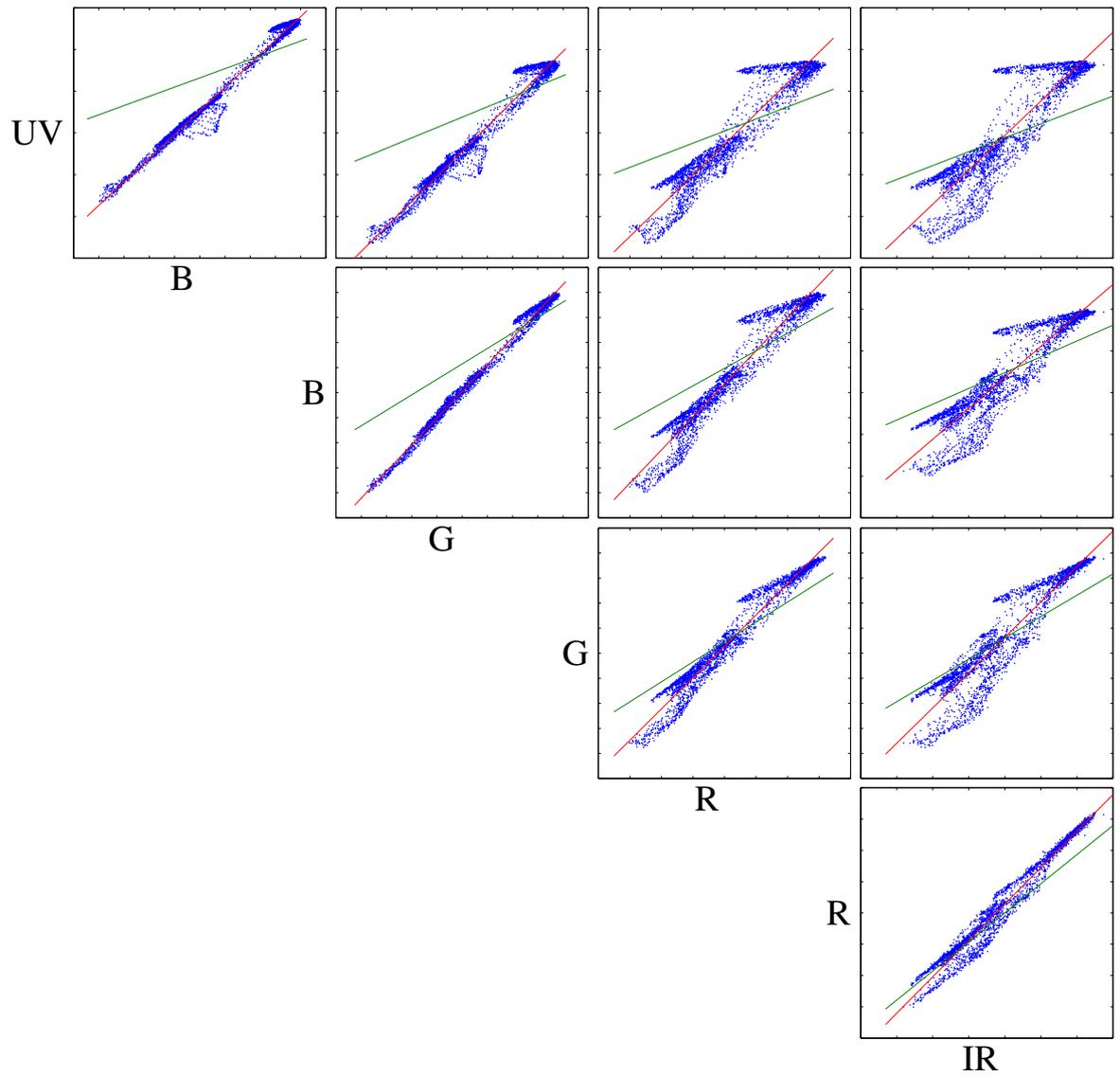


Figure D.23: Plotted contrasts for dataset one. The red line is fitted to the data. The green line depicts the learned separation line.

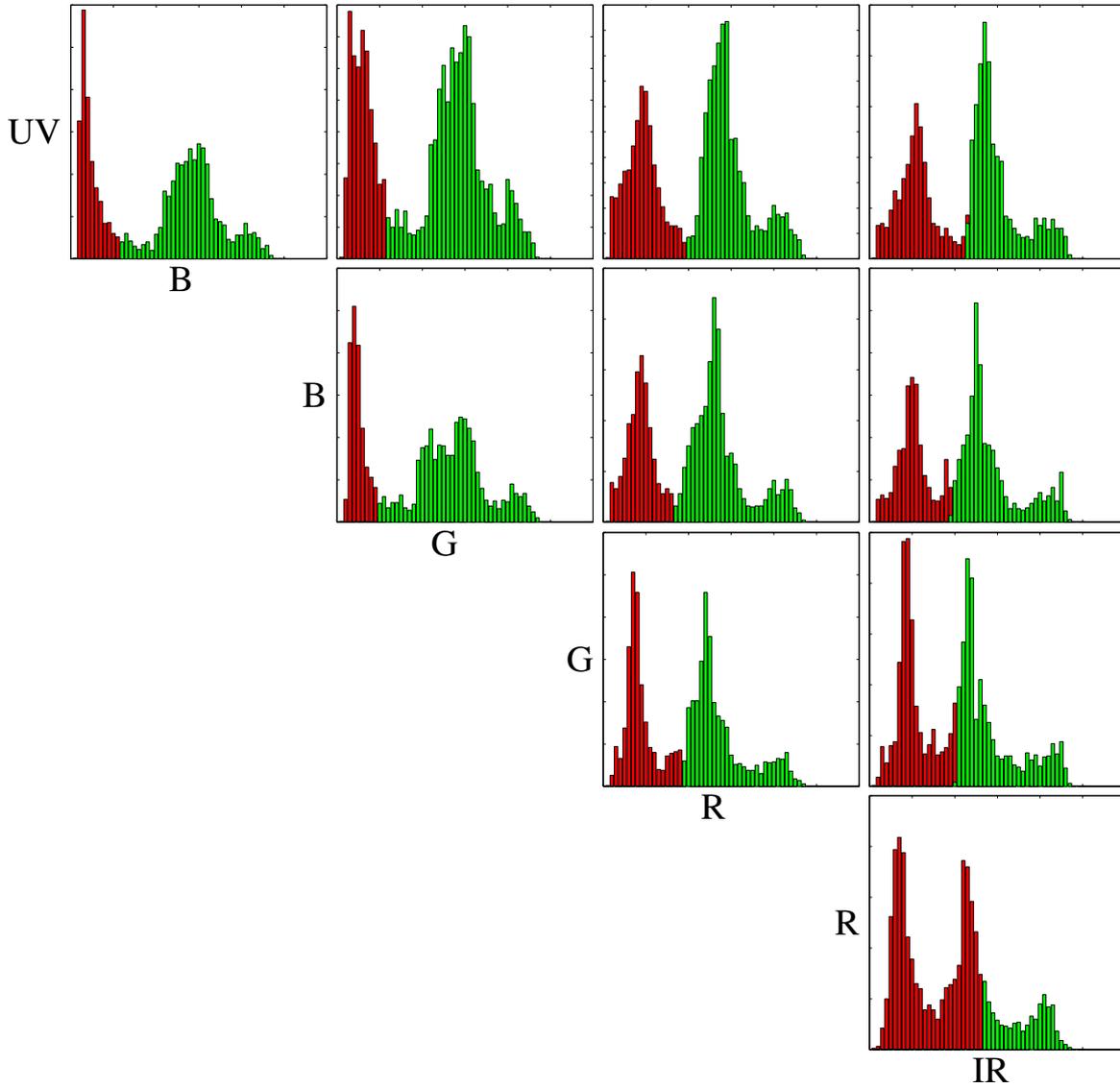


Figure D.24: Contrast histograms for dataset one. The red bars depict samples of the sky. The green depict samples of foreground.

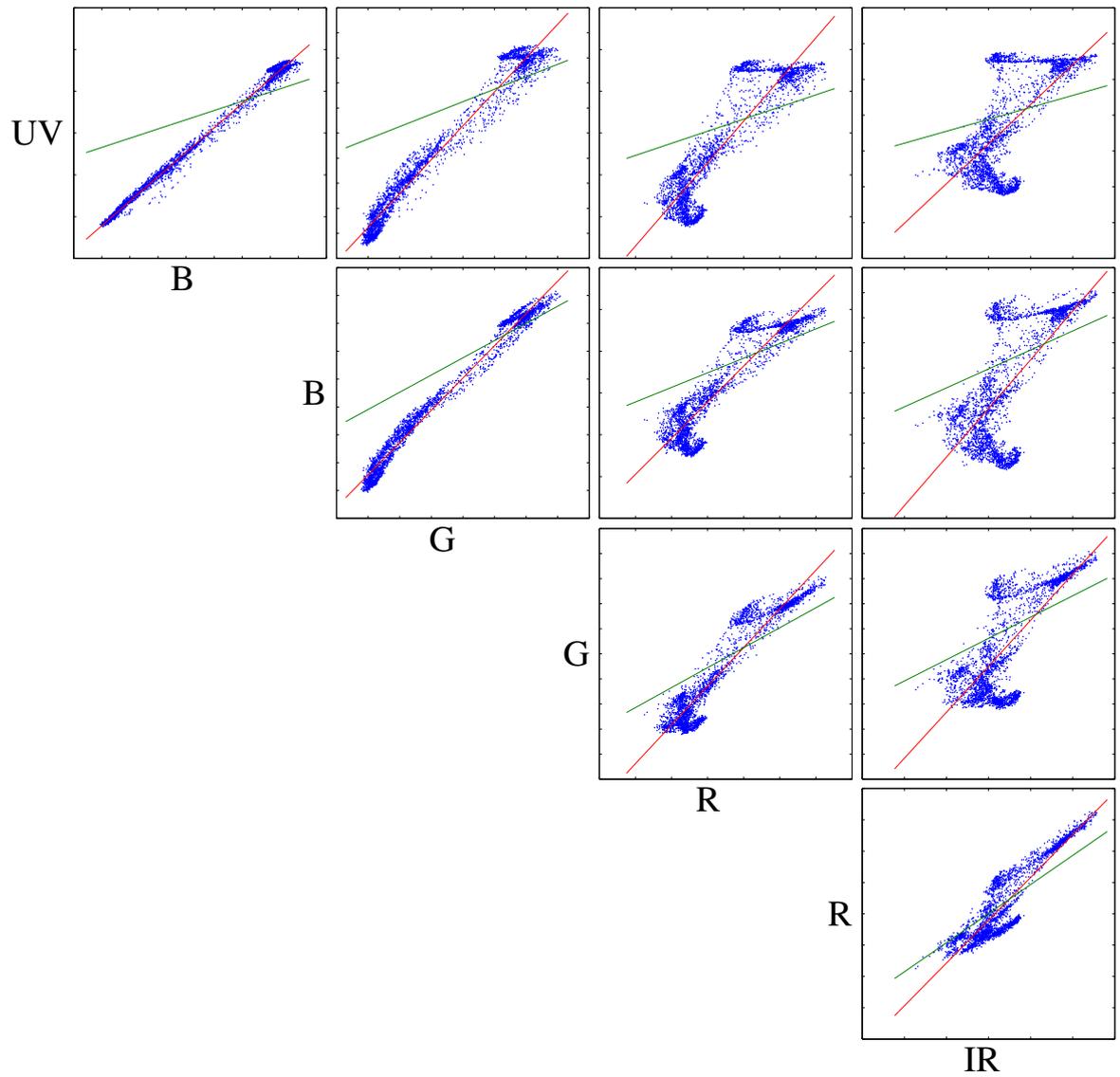


Figure D.25: Plotted contrasts for dataset two. The red line is fitted to the data. The green line depicts the learned separation line.

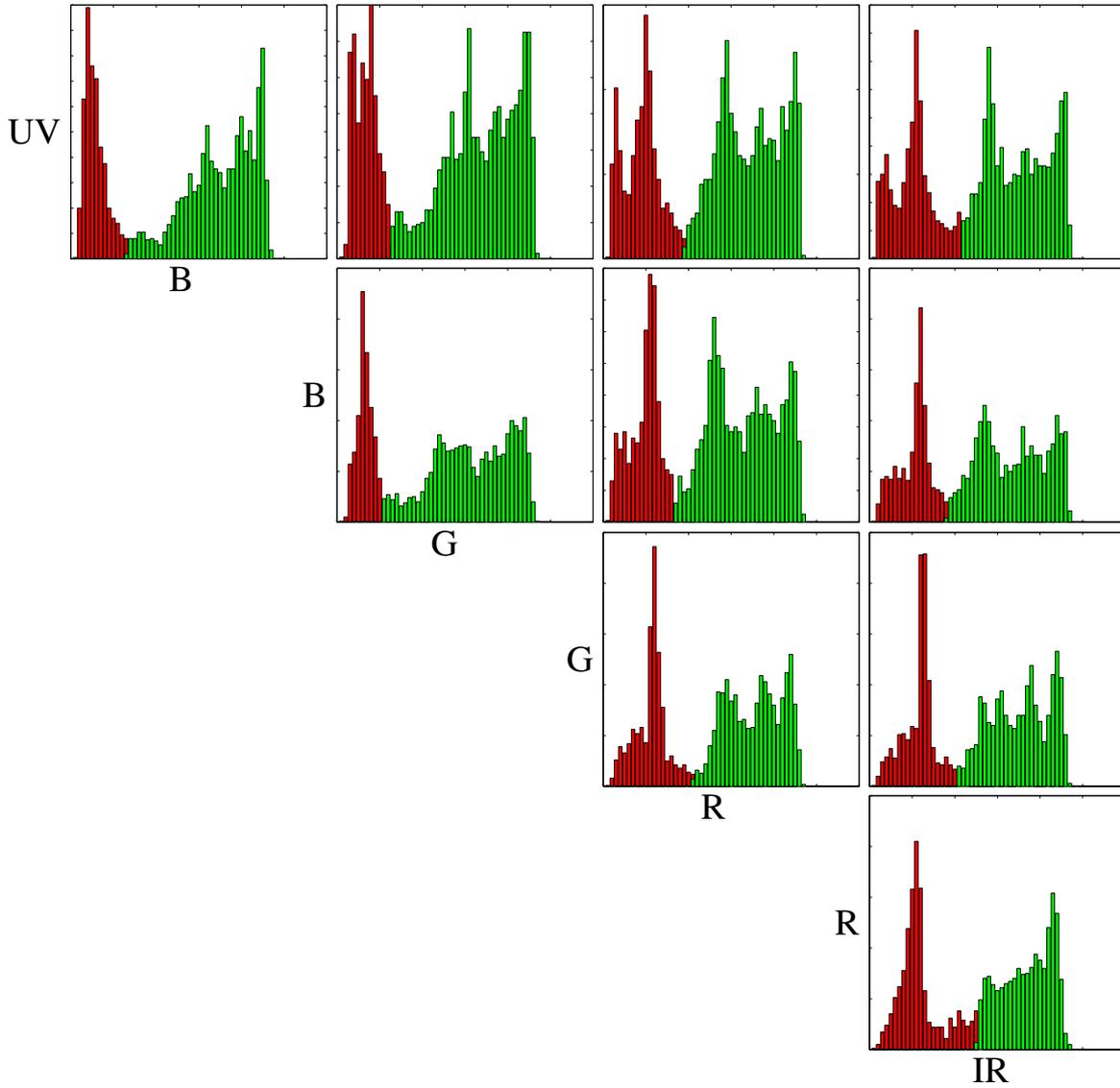


Figure D.26: Contrast histograms for dataset two. The red bars depict samples of the sky. The green depict samples of foreground.