

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Diplom-Physiker Volker Hilsenstein
born in Mannheim, Germany
Oral examination: 5. May 2004

Design and Implementation of a Passive Stereo-Infrared Imaging System for the Surface Reconstruction of Water Waves

Referees: Prof. Dr. Bernd Jähne
Prof. Dr. Kurt Roth

Zusammenfassung

Eine quantitative Beschreibung der Austauschprozesse zwischen Ozean und Atmosphäre erfordert ein Verständnis des Einflusses von Wellen auf diese Prozesse. Diese Arbeit stellt ein bildgebendes, passives Infrarot-Stereokamerasystem zur Rekonstruktion wellenbewegter Wasseroberflächen vor. Da keine Lichtquelle unterhalb der Wasseroberfläche benötigt wird, ist das System feldgänglich. Zunächst werden bestehende Verfahren zur Wellenvisualisierung besprochen und die Hauptprobleme der stereo-photogrammetrischen Rekonstruktion von Wasseroberflächen herausgearbeitet: Transparenz, unzureichende Bildstruktur und spiegelnde Reflektion. Es wird aufgezeigt, daß sich durch Bildaufnahme im infraroten Wellenlängenbereich viele dieser Probleme vermeiden lassen. Nach einer Wiederholung der Infrarotradiometrie werden die wichtigsten Komponenten der auf dem Stereoprinzip beruhenden Oberflächenrekonstruktion erläutert: Kamerakalibration, Epipolargeometrie sowie Disparitätssuche. Im Anschluß wird das verwendete Stereokamerasystem beschrieben. Eine experimentelle Validierung des Systems erfolgte am Heidelberger Wind-Wellenkanal. Anhand mehrerer dort aufgenommener Infrarot-Stereobildsequenzen von Wasserwellen wird gezeigt, daß das Instrument eine dichte Rekonstruktion der Wasseroberfläche ermöglicht. Eine experimentelle Beurteilung der Genauigkeit des Verfahrens erfolgt anhand der Rekonstruktion einer ruhenden Wasseroberfläche, welche als Referenzebene verwendet wird.

Abstract

To quantify air-sea exchange processes, an understanding of how they are influenced by water waves is necessary. This work presents a passive, infrared stereo-imaging system for the reconstruction of a wavy water surface. The system does not require a submerged light source, so it is suitable for field operation. The structure of the thesis is as follows. Previous work on water wave imaging is reviewed and the major problems with stereo-based reconstruction of water surfaces are identified: transparency, lack of texture and specular reflections. It is shown that many of these problems can be avoided by imaging at infrared wavelengths. Following a review of infrared radiometry, the key ingredients of surface reconstruction using the stereo principle are explained, including camera calibration, epipolar geometry and disparity estimation. A description of the stereo infrared camera system used in this work is given. An experimental validation of the system was performed at the Heidelberg wind-wave channel. Several stereo infrared image sequences of water waves recorded at this facility are used to demonstrate that a dense surface reconstruction of water waves is possible using this system. The accuracy of the reconstruction is experimentally assessed using a flat water surface as a reference plane.

Contents

1	Introduction	7
1.1	Motivation	7
1.1.1	Air-Sea Interactions and their Effects on Climate . . .	7
1.1.2	Factors Influencing Air-Sea Exchange Processes . . .	9
1.1.3	Heat Flux Measurements using Thermography	11
1.2	Related work	12
1.2.1	Stereo Measurements of Water Waves	13
1.2.2	Slope Measurements of Waves	16
1.3	Why is wave imaging difficult ?	18
1.4	Aim and own contribution	21
1.5	Thesis Outline	23
2	Basics of Infrared Imaging	25
2.1	Radiometry	26
2.1.1	Definitions	26
2.1.2	Electromagnetic Radiation of a Blackbody	30
2.1.3	Emissive Properties of Real Surfaces	34
2.1.4	Optical Properties of Water in the Infrared Region . .	35
2.2	Infrared Detectors	40
2.2.1	Types of detectors	41
2.2.2	Sensitivity	42
2.2.3	Detector Output and Temperature Measurements . .	43
3	Geometry	47
3.1	Projective Geometry	48
3.1.1	The Projective Plane \mathbb{P}^2	48
3.1.2	The Projective Space \mathbb{P}^3	49
3.1.3	Homographies	49

3.2	Single View Geometry and Camera Models	50
3.2.1	Pinhole camera model	50
3.2.2	World Coordinate System	52
3.2.3	CCD-type cameras	53
3.3	Non-Linear Distortion	53
3.3.1	Modeling Non-Linear Distortion	54
3.3.2	Correcting for Non-Linear Distortion	55
3.4	Estimation of Camera Parameters using Zhang's Method	55
3.4.1	The Calibration Process	57
3.4.2	Initial Guess through Closed-Form Solution	58
3.4.3	Full Solution through Minimization of Geometric Error	61
3.5	Two View Geometry	62
3.5.1	Calibration of a Stereo Camera System	62
3.5.2	Epipolar Constraint and the Fundamental Matrix	65
3.5.3	Image Rectification	67
3.5.4	Projective Distortion-Minimizing Rectification	69
3.5.5	Triangulation	73
4	Solving the Correspondence Problem	75
4.1	Classification of Matching Algorithms	75
4.1.1	Feature-based Stereo Matching	76
4.1.2	Area-Based Stereo Matching	76
4.2	Prerequisites for Area-Based Matching	79
4.2.1	Fronto-Parallel	79
4.2.2	Lambertian Surface	80
4.2.3	Opacity	81
4.2.4	Texture	82
4.2.5	Conclusions Regarding Wave Imaging	82
4.3	Matching Algorithm	83
4.3.1	Matching Score	84

<i>Contents</i>	3
4.3.2 Efficient Implementation	84
4.3.3 Computation of Disparity	87
4.3.4 Sub-Pixel Refinement	87
4.3.5 Validation of Matches	88
4.3.6 Multi-Scale Approach	89
5 Image Pre- and Postprocessing	93
5.1 Non-Uniformity Correction and Radiometric Calibration . . .	93
5.2 Outlier Removal	95
5.3 Regularization: Filling in the Gaps	96
5.3.1 Theory	96
5.3.2 Filling in Defective Pixels	98
5.3.3 Regularization of Disparity Estimates	99
6 Experimental Setup and Procedures	101
6.1 Infrared Cameras	101
6.1.1 Specifications	101
6.1.2 Stereo Setup	103
6.2 Blackbody	104
6.3 Acquisition System	104
6.3.1 Frame Grabber	104
6.3.2 Camera Synchronization	106
6.3.3 PC and RAID	106
6.4 Geometric Calibration Target	108
6.5 Aeolotron	109
6.6 Experimental Procedure	110
6.6.1 Radiometric Calibration Procedure	110
6.6.2 Geometric Calibration Procedure	110
6.6.3 Deployment	111
6.6.4 Acquisition of Image Sequences	112

7	Results	113
7.1	Radiometric Calibration and Non-Uniformity Correction . . .	113
7.1.1	Thermosensorik CMT Camera	114
7.1.2	Amber Radiance Camera	115
7.2	Geometric Camera Calibration	118
7.2.1	Interior Parameters	120
7.2.2	Exterior Parameters	122
7.3	Rectification	126
7.4	Disparity Estimation	128
7.4.1	Test Image Results	128
7.4.2	Multi-Scale Disparity Estimation	129
7.5	Regularization	130
7.6	Depth Reconstruction	133
7.7	Measurements of Water Waves	134
7.8	Discussion of Accuracy	138
7.8.1	Assessing the Total System Accuracy	138
8	Conclusions	145
8.1	Summary	145
8.2	Discussion and Outlook	145
A	Reconstruction Accuracy	149
A.1	Factors Influencing the Reconstruction Accuracy	149
A.2	Range Resolution of Triangulation	150
A.3	Estimation of a Reference Plane	152
A.3.1	Least-Squares Estimate	152
A.3.2	Robust Estimate	153
B	Estimating Planar Homographies	155
B.1	Problem statement	155
B.2	Initial Guess	156
B.3	Minimizing Geometric Error	157

<i>Contents</i>	5
Index	160
Bibliography	163

1 Introduction

1.1 Motivation

1.1.1 Air-Sea Interactions and their Effects on Climate

In 2000, an international and interdisciplinary research program, the *Surface Ocean Lower Atmosphere Study (Solas)*, was founded. The program's goal is

“to achieve quantitative understanding of the key biogeochemical-physical interactions and feedbacks between the ocean and the atmosphere, and how this coupled system affects and is affected by climate and environmental change”. [69]

These ocean-atmosphere interactions are being studied partly in order to understand whether the changes in the world's climate, with a rise in the global mean temperature being the most prominent, can be attributed to human activities and to be able to predict future effects of these activities on climate. As the first point is already largely answered – according to the third assessment report of the *Intergovernmental Panel on Climate Change* [53] “there is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities” – the emphasis of current research is to monitor and better quantify the ongoing changes and to predict future changes in the world's climate system.

The anthropogenic release of carbon dioxide (CO_2), as well as other greenhouse gases, into the atmosphere is known to be one of the major factors driving climate change. When deriving global budgets for atmospheric CO_2 concentrations, the coupling between the atmosphere and the oceans has to be taken into account because the oceans are a major sink for CO_2 . The net flux of CO_2 from the atmosphere into the oceans during the period from 1990 to 1999 was about 2 PgC^1 per year, which is roughly one third of the yearly anthropogenic emissions into the atmosphere (see IPCC [53]). The temporal and regional variability of this CO_2 air-sea flux is not well understood. The available data on this variation is from measurements obtained over the last three decades assimilated by Takahashi [97, 98].

One of the parameters influencing the magnitude of these fluxes is the transport velocity across the air-sea interface. This transport velocity quantifies the kinetics of gas exchange and its value is in turn affected by many physical and biogeochemical processes such as diffusion, near-surface turbulence, surfactants and wave breaking.

¹PgC stands for petagrams of carbon. One petagram equals 10^{15} grams or one gigaton.

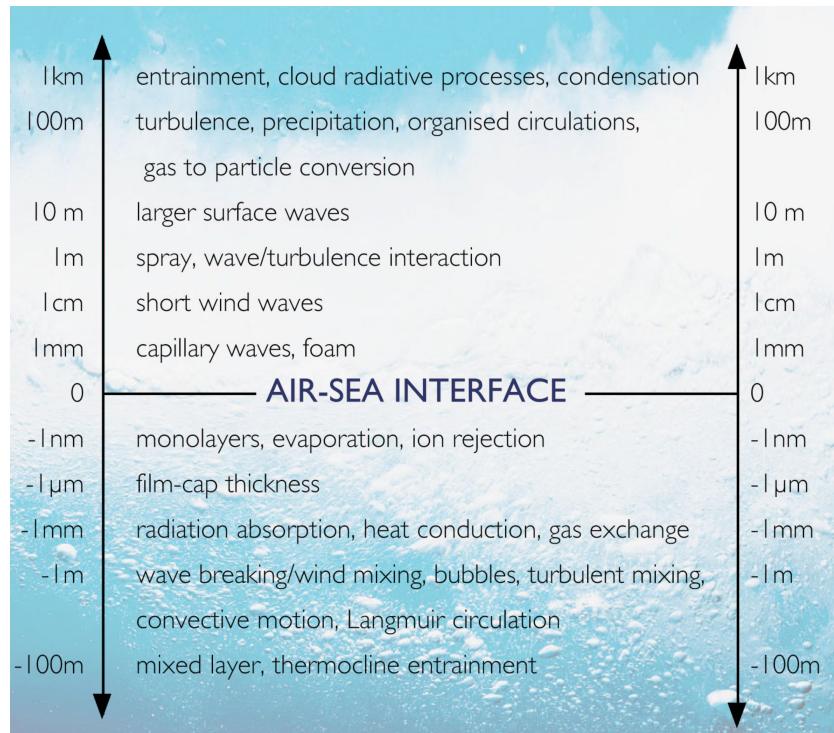


Figure 1.1: The different length scales (height/depth) of processes occurring at the air-sea interface. Taken from [69].

The study of the processes influencing transport across the air-sea interface is one of the main foci of the *Solas* program. Figure 1.1 summarizes some of these processes together with their characteristic length scales. Close to the surface, waves of different scales and wave breaking play an important role in the exchange processes. A primary goal of the current studies is to derive a physically sound parameterization of the transfer velocity, based on observable physical and biogeochemical properties of the interface. Such a parameterization can then be used to improve the predictive capabilities of climate models, and, provided that the parameters can be observed by satellite remote sensing, to derive global maps of air-sea fluxes.

Such a parameterization of air-sea transport velocity is not only of interest for the exchange of CO_2 , but also for other climate-relevant gases such as nitrous oxide (N_2O) and dimethyl sulfide (DMS), as well as for the transport of heat and momentum. In the oceans, heat energy can be transported by oceanic circulations and thereby alters the global distribution of energy.

1.1.2 Factors Influencing Air-Sea Exchange Processes

To quantify the oceans' role as a sink for CO₂, other gases, and heat, it is necessary to understand the exchange of gases across the air-sea interface. Note that the processes controlling air-sea transport are similar for gases, heat, and momentum. Therefore the following discussion is independent of the transported substance. In this section, some of the terms used in the subsequent discussion are defined.

Boundary Layer Air-sea exchange is driven by the concentration difference between the atmosphere and the ocean, and its kinetics are usually expressed in terms of a single number, the transfer velocity k . For most gases, as well as for heat, the transfer velocity is controlled by processes within a thin aqueous boundary layer. In this layer, molecular diffusion is the dominant mode of transport, because turbulent eddies cannot reach through the surface.

Surface Renewal Model It has been observed that a range of processes, such as wind stress, waves, wave breaking, buoyancy fluxes and interactions with subsurface flows can significantly enhance gas transfer. The common feature of these processes is that they create subsurface turbulence. Surface films, which reduce subsurface turbulence by damping waves and reducing the momentum input of wind stress, have the opposite effect and reduce air-sea transfer.

A simple surface renewal model describes the effects of subsurface turbulence on the boundary layer very well. The model assumes that turbulent eddies randomly replace parts of the diffusion-controlled boundary layer with water from the well-mixed bulk. The higher concentration difference between the bulk water and the atmosphere enhances the transport velocity. Figure 1.2 illustrates this model for the case of heat transfer. A net heat flux cools down the water surface and a temperature difference develops between the well-mixed bulk water and the diffusion-controlled boundary layer. The colder surface layer has a typical thickness on the order of half a millimeter and is referred to as the *cool skin of the ocean*[79]. Turbulent eddies randomly replace parts of this cool skin, bringing up warmer bulk water, which subsequently equilibrates its temperature by diffusion.

At higher wind speeds, bubble entrainment due to wave breaking and sea spray further enhance gas transfer.

Water Waves As already mentioned, water waves play an important role in the spatial and temporal variability of gas transfer. Short capillary and capillary-gravity waves dissipate their energy by microscale breaking, creating subsurface turbulence. Larger gravity waves affect the spatial and temporal distribution of

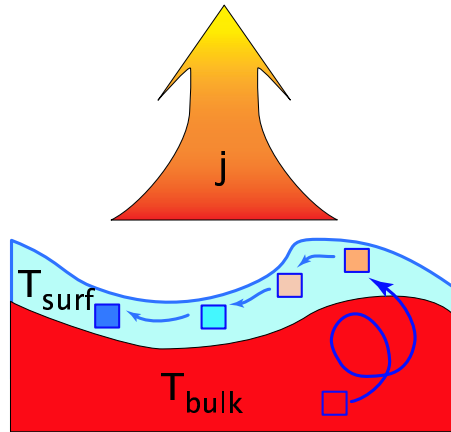


Figure 1.2: Surface renewal model. A heat flux j cools down the water surface. Directly below the surface, a cold water layer with a typical thickness of about 0.5 mm develops, the “cool skin of the ocean” [79]. Turbulent eddies randomly replace parts of this cool water layer with warmer fluid parcels from the well mixed bulk. These warmer fluid parcels subsequently equilibrate their temperature by diffusion. (Image courtesy of C. Garbe, IWR Heidelberg)

this microscale breaking, that is breaking without entrainment of bubbles, by modulating the steepness and propagation velocity of the smaller waves. At larger scales and higher wind speeds the distribution of whitecap and wave breaking with entrainment of bubbles is similarly modulated by long gravity waves and swell [23].

Parameterizations Even with a perfect understanding of air-sea gas exchange, it would not be feasible to account for all the processes influencing the transfer velocity when modeling fluxes on a regional or global scale. Therefore, a parameterization of transfer velocity, using only a few observable variables, and based on sound physical principles, is desirable.

Most parameterizations of transport velocity used in climate models are based on the single parameter wind speed (see Liss and Merlivat [70], Wanninkhof [104]). Part of the popularity of wind speed as a parameter is probably due to the fact that it can be measured easily. However, it is known that for a given wind speed, gas transfer velocities can fluctuate significantly in the presence of surfactants [30, 31]. Such surface films, either from anthropogenic sources or produced by marine plankton, are ubiquitous, so using wind speed as a single parameter for the transfer velocity appears questionable.

It has been suggested that wave slope might be a more suitable parameter to characterize sub-surface turbulence and its effect on gas exchange. Measurements by Jähne et al. [60] and Bock and Hara [9] have shown that a parameterization based on mean squared wave slope can be used to estimate the transfer velocity, independent of surface films. Mean squared wave slope also determines the intensity of microwave scattering, making it a parameter that can be remotely sensed by microwave scatterometry from airplanes or satellites [95].

1.1.3 Heat Flux Measurements using Thermography

Modern infrared cameras can image the temperature distribution on the water surface with high spatial and temporal resolution. The smallest temperature differences that they can resolve typically lie around 20 mK. This allows direct observation of surface renewal events at the air-sea interface, as the temperature difference between the cool skin layer and the warm bulk water is typically on the order of 100 mK (see figure 1.2). Therefore, such infrared cameras are a powerful tool to study air-sea gas exchange.

Typical examples of infrared images of a water surface, acquired at the Heidelberg *Aeolotron* (see section 6.5), are given in figure 1.3. In the first image of the sequence the water is in thermal equilibrium with the airspace, which results in an almost uniform temperature. At the time this first image was taken, a net heat flux was created by flushing the airspace with dry air, but the effect of this flux is not yet observed. Wind stress from the air conditioning system and buoyancy fluxes due to density differences generate subsurface turbulence. This subsequently leads to surface renewal events bringing up warmer bulk water, which shows up as bright streaks. A complex temperature pattern develops as seen in figure 1.3b-f. The total temperature difference between bright and dark regions in this image sequence is approximately 0.5 K.

Over the last decade, a number of image processing techniques have been developed, which allow the estimation of net heat flux and the transfer velocity directly from infrared image sequences. These methods are based on measuring the temporal decrease in temperature of fluid parcels at the water surface. To do this, the fluid parcels have to be tracked between image frames, using motion estimation algorithms. Recent reviews are given by Haußecker et al. [46] and Garbe et al. [36].

Because the transport processes for heat and gases are essentially the same, the transport velocities for gases can be calculated from the transport velocity obtained for heat, using appropriate scaling that corrects for the different diffusivities of heat and gases. This is known as *Schmidt number scaling* [see for example 46].

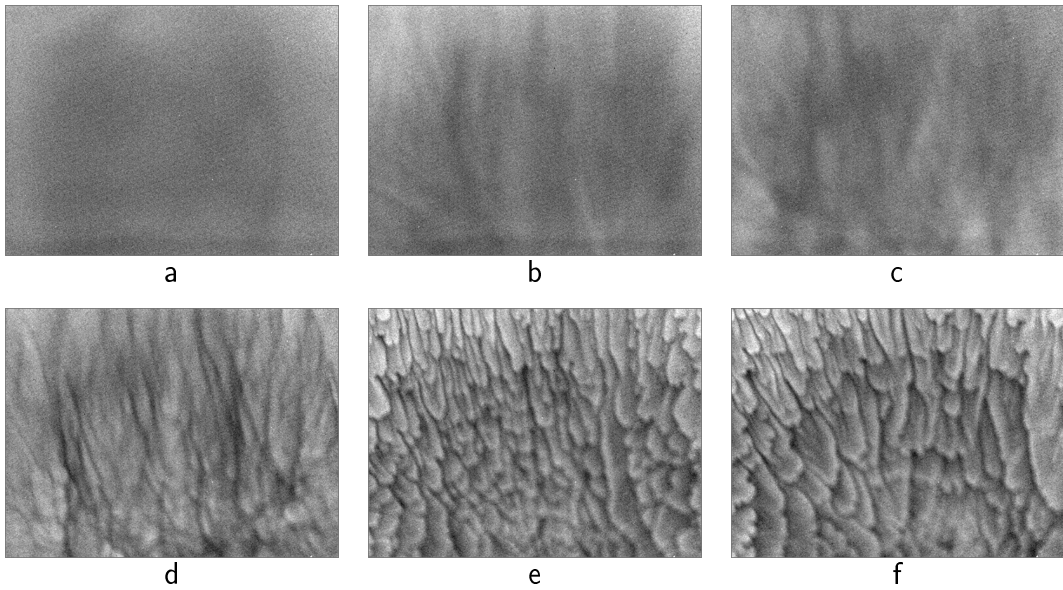


Figure 1.3: Infrared image sequence of a flat water surface acquired in the Heidelberg Aeolotron wind-wave facility. When the first image was captured the water surface was still in thermal equilibrium with the airspace. As a heat flux was created by flushing the airspace with cool dry air complex temperature patterns develop at the water surface, driven by subsurface turbulence (b-f). The time between consecutive images is about 1.5s. Surface temperature is expressed in terms of grey-value with brighter values representing higher temperatures. The temperature difference between dark and light pixels is about 0.5 K. The imaged area measures approximately $18\text{cm} \times 13.5\text{cm}$.

1.2 Related work

This section gives a two-part summary of previous approaches to wave imaging. The first part reviews techniques that use stereo imaging to recover the wave surface topography, while the second part reviews techniques that measure the wave slope or statistical parameters thereof.

Other reviews of wave imaging can be found in Redweik [76, chapter 2] with an emphasis of stereo imaging of sea swell, Jähne et al. [59] which focuses on imaging short wind waves, and Luhmann and Tecklenburg [72], whose focus is on reconstructing the shape of bow waves created by ships in a model tank.

1.2.1 Stereo Measurements of Water Waves

Stereo photogrammetry is a method that can be used to reconstruct the shape of objects given two images taken from different positions. It is based on the principle of triangulation depicted in figure 1.4; the position of a point P can be calculated by intersecting the lines of sight.

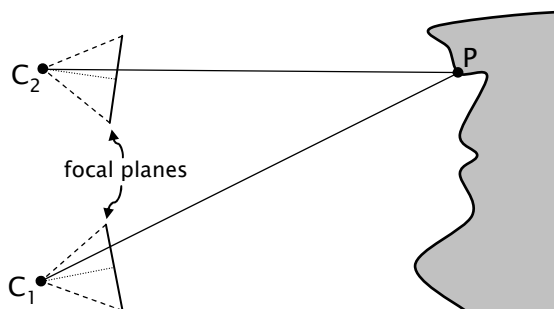


Figure 1.4: Triangulation: If a point P on a surface is imaged by two cameras with known optical centers C_1 and C_2 , its position can be calculated by intersecting the two rays of view.

The idea of measuring the shape of water waves using stereo imaging is not new. In fact, in 1903, shortly after Pulfrich built the *Zeiss stereo-comparator* in Jena (see figure 1.5) that made routine photogrammetric analysis of stereo image pairs feasible, Rottok and Kohlschütter [77] suggested using the newly available instrument to measure wave topography on the ocean.

In the summer of 1904 Kohlschütter embarked on the *S.M.S. Hyäne*, equipped with a stereo camera setup designed by Pulfrich, for a cruise on the Kieler Förde. The primary task of the cruise was to evaluate the suitability of stereo photogrammetry for mapping shorelines, but image pairs of swell were collected as well. Although the measurements were apparently hampered by the fact that the alignment of the two photographic plates was distorted due to vibrations on the ship, Kohlschütter [68] notes that his experiments on stereo wave imaging produced “günstige Ergebnisse” (favorable results).

Also in 1904, Laas, during a cruise on the Southern Atlantic Ocean, collected several stereo image pairs of water waves and was able to produce topographic maps of wave height. Such a map is depicted in figure 1.6. The image analysis had to be done manually using a stereo-comparator, which is very time consuming and therefore hardly feasible for large numbers of image pairs.

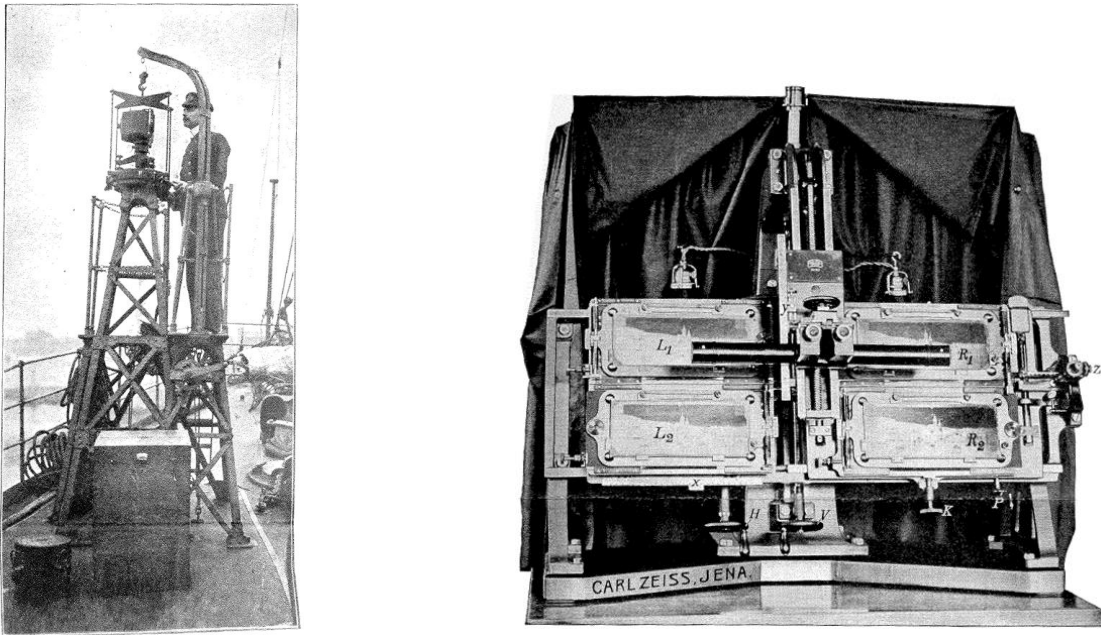


Figure 1.5: Left: One of the cameras, mounted close to the bow, aboard the *S.M.S Hyäne*, which were used for the stereo photogrammetric measurements in 1904. The second camera (not shown) was located close to the stern. Right: The Zeiss stereo-comparator, which was used to analyze stereo image pairs. (Both images from [68]).

Schumacher used similar stereo camera setups on cruises aboard the *S.M.S. Meteor* in 1925–1927 [87] and aboard the *Europa* in 1939. Due to the outbreak of war the results of these cruises weren't published until 1950 [88].

In the early 1950s, attempts were made to obtain aerial stereo images of ocean waves, in France by the *Institut Géographique National* [18], and in the United States during the *Stereo Wave Observation Project, SWOP* [15]. The images were obtained from two cameras mounted on separate airplanes flying in parallel at the same height. To guarantee simultaneous image exposure, the cameras were synchronized with a radio transmitter. Both attempts were largely a failure; in the French project the contrast of the images was not good enough to allow for a proper analysis and the SWOP team could evaluate only two of about one hundred image pairs taken.

In 1993, Redweik [76] evaluated whether correlation-based matching can be used for the automated analysis of stereo image pairs of waves. For a selected set of images acquired on a North Sea platform under favorable natural lighting conditions, without sun glint, he produced digital elevation models of the wave fields.

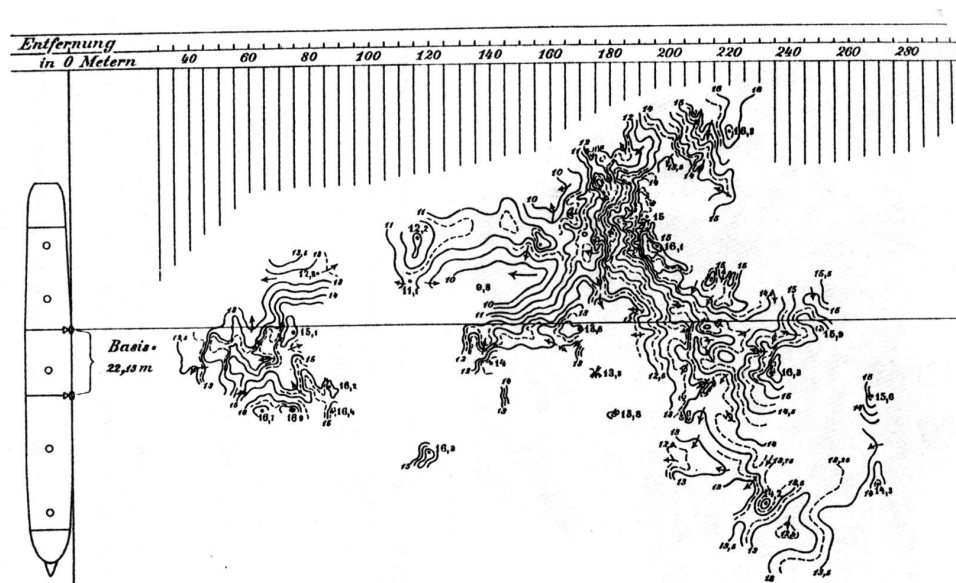


Fig. 11.
Topographic der Meeresoberfläche.
Aufnahme von Professor Laas 1904. Messung von Topograph Seliger 1905.

Figure 1.6: Surface topography of large gravity waves reconstructed from a stereo image pair taken in the Southern Atlantic in 1904 by Laas.

He demonstrated that automated image matching was suited to the task and, within the uncertainties, produced identical results as a manual analysis on a stereo-planigraph.

In 2002 Santel et al. [78] used a stereo camera system to reconstruct the shape of water waves in the surf zone. They manually labeled a number of corresponding points that were used as seed points for a correlation-based matching algorithm. Correlation-based matching is feasible in this case because large parts of the water surface in the surf zone are covered with foam, which provides surface texture and has almost Lambertian reflection properties (see Figure 1.8c,e and section 2.1.1).

In the works described so far, the observed swell and wave fields were typically on the order of hundred meters wide with wave heights on the order of meters. Research on stereo imaging applied to short gravity and capillary wind waves has been performed by Shemdin et al. [91],[90] as well as Banner et al. [5], who produced directional wave spectra from stereo image pairs. However, a critical analysis reveals that their results may be unreliable, as they claim a much higher resolution than theoretically possible [6, 55, 59].

In the last decade, Waas [101, 102, 103] developed a combined wave height and

slope gauge that uses the stereo principle to compute the average wave height. For this instrument, called the *Reflective Stereo Slope Gauge*, a LED light source, creating specular glints, is mounted next to each camera and polarizing filters ensured that each camera would only image light originating from the light source mounted near the other camera. The symmetric design of the system ensures that specular glints appearing in the two images correspond to the same surface patch, making these glints good features to match. The glint patterns are also used to obtain statistical data about the wave slope distribution. This system was later refined by Dieter [20, 21] and used for field measurements from *Scripps' Pier* in La Jolla, California. The disadvantage of this approach is that only the small fraction of the water surface covered by specular glints can be reconstructed. Furthermore, the system is sensitive to ambient light and can therefore only be operated at night.

Other researchers have used a different strategy altogether, actively altering the optical properties of the water surface by applying aluminum powder or other substances to it, thereby reducing its specularity [see for example 1, 73]. For example, this was used for the study of hydrodynamic models of harbors. However, exchange processes at the air-sea interface are strongly influenced by surface films, and on the open ocean the marker substance would have to be constantly resupplied, so it is obvious that this method can not be used for the present application. Common to all previous approaches to stereo wave imaging is that they operate at visible wavelengths.

1.2.2 Slope Measurements of Waves

Another class of methods for obtaining information about the shape and motion of water waves is based on measuring the surface slope, by using either the directional dependence of the light intensity *reflected* at the air-water interface or the directional dependence of the *refracted* light that passes through the interface.

Reflection-Based Slope Measurements The seminal work for methods to characterize wave fields based on reflection of light at the water surface dates back to 1954 and is due to Cox and Munk [16, 17]. They derived wave slope statistics from sun glint patterns, using the fact that the sun only covers a small solid angle of the sky and therefore the slope of facets that specularly reflect sunlight towards an observer is determined within a narrow range.

Stilwell [96] later developed a generalization of this technique, which used an optical Fourier transform to generate wave energy spectra directly from photographs of the water surface, obtained in situations with a uniform sky illumination.

In computer vision, a technique to reconstruct the shape of Lambertian surfaces based on reflected intensity is known as *shape-from-shading* [see 52]. Using this shape-from-shading approach, Schultz [85, 86] proposed and simulated a system consisting of four synchronized cameras that should allow a complete reconstruction of the surface slope under natural lighting conditions. One camera was pointed to the sky to map the hemispherical radiance distribution. Based on this radiance map and the irradiance values obtained by the other three cameras, imaging the same patch of water from three different directions, the two components of surface slope could be estimated using a specular surface model. Presumably owing to the complexity of this proposed system it was never implemented [55].

Refraction-Based Slope Measurements The other class of wave slope imaging techniques is based on refraction at the air-water interface. The current state of the art for wave slope measurements in wind-wave tanks is the so-called *Color Imaging Slope Gauge* (CISG) [3, 4, 33, 61]. The key to this technique, depicted in figure 1.7, is an extended light source. The intensity of this light source is color coded, such that every point on the light source has a different color value. With this optical setup, each value of the surface slope corresponds to a unique color value that is recorded by the camera. This technique allows very accurate slope measurements and can resolve small capillary waves, as depicted in figure 1.7 on the right.

Based on the good performance of these instruments in laboratory settings, several refraction-based instruments have been built for use in the field.

Klinke and Jähne [66, 67] used an imaging slope gauge on a freely drifting buoy, with cameras mounted above the water surface and a submerged array of light emitting diodes (LEDs) to create an intensity-coded area light source. Due to interference from ambient light, the use of the instrument is restricted to dusk and night.

Bock and Hara [8] have employed the refractive technique using a submerged scanning laser and a position-sensitive photodiode for field measurements on a catamaran. The submerged laser assembly is smaller than the extended light source used by Klinke but can still interfere with the wave field.

Both instruments have provided valuable data, but they also demonstrate a general problem with using the refractive technique for routine measurements. As either the camera or (preferably) the light source must be submerged, they can usually not be operated directly from a research vessel, but have to be mounted on a smaller external float, like a buoy or a catamaran. Therefore, the system has to be deployed separately, which is very labor intensive and can become difficult at

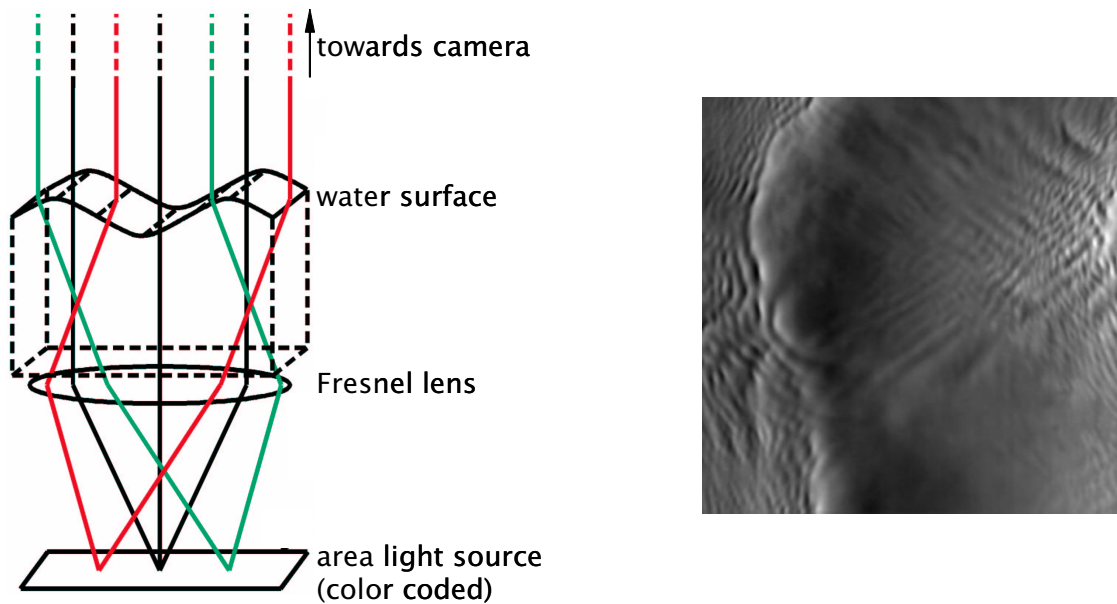


Figure 1.7: Color Imaging Slope Gauge. *Left: Optical Setup. The light received by the camera depends on the slope of the water surface. Due to the color coded area light source, the slope can be calculated from the color value of an image point. Right: Wave slope image of capillary waves obtained with a CISG. The imaged area is approximately $15\text{ cm} \times 15\text{ cm}$. (Image courtesy of Jochen Klinke of Scripps Institution for Oceanography)*

high wind speeds or with an agitated sea. Furthermore, the whole setup has to be very rugged and watertight to withstand sea water. The submerged parts of such a system can also interfere with the larger gravity waves.

1.3 Why is wave imaging difficult ?

When analyzing the previous attempts to reconstruct the surface topography, surface slope and motion of water waves reviewed in the preceding section, it becomes apparent that the optical properties of water in the visible wavelength range pose severe problems under natural lighting conditions.

The difficulties are best illustrated with a few images of water surfaces under different conditions, as shown in figure 1.8. Depending on view angle and illumination conditions the following effects are observed:

- **Specularity:** Light reflection off the water surface is largely specular, or mirror-like. That means that the reflected intensity is very large when the

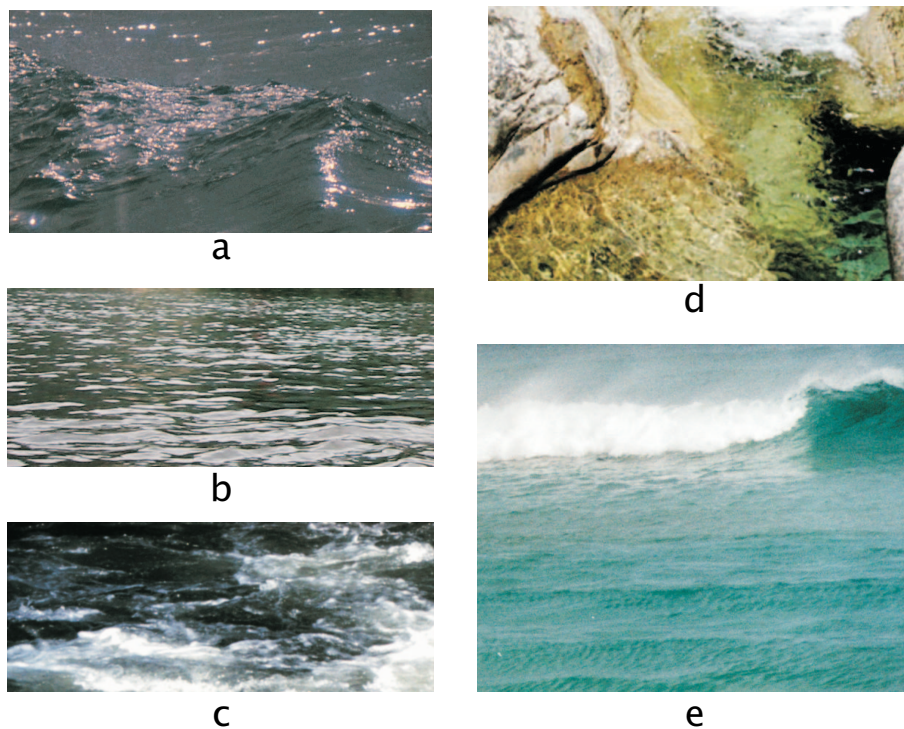


Figure 1.8: Images of water waves taken under different illumination conditions and observation angles illustrate some of the major optical properties, which make wave imaging difficult (see text).

incidence angle equals the observation angle. In bright sunlight, this leads to sun glints originating from the narrow range of surface slopes that reflect direct sunlight as depicted in figure 1.8a. For overcast sky conditions the specularity leads to a situation as in figure 1.8b.

- Transparency: For observation angles close to the surface normal, water is largely transparent in the visible region of the electromagnetic spectrum (figure 1.8d). This becomes important when imaging small scale capillary waves in the field, as the observation angle has to be close to the surface normal in this case. Otherwise, the capillary waves are more likely to be occluded by larger gravity waves.
- Lack of texture: Smooth parts of the water surface are often mostly featureless (figure 1.8a,b,c).
- Large intensity variations: The specularity causes large variations in image irradiance, with some surface patches reflecting direct sunlight and other surface patches reflecting light from darker areas of the sky or showing up-

welling light (figure 1.8a). Due to their limited dynamic range, most cameras cannot cover the whole range of irradiance observed with good intensity resolution.

- Non-uniform reflectance properties: Strongly aerated water, as found in breaking waves (figure 1.8c,e), has very different reflectance properties from non-aerated water. Due to multiple reflections at the interface of bubbles, breaking waves are to a good approximation Lambertian (see section 2.1.1).

Specularity, transparency and lack of texture pose a serious problem for stereo photogrammetry. To reconstruct the surface using triangulation, points in images taken from different positions corresponding to the same point on the surface have to be identified. Stereo matching algorithms usually fail to find such correspondences where the water surface appears without texture or transparent.

Specular reflection is view angle dependent, so it can introduce a systematic error when finding correspondences. For example, a flat part of the water surface reflects radiation from different parts of the sky into the two cameras. The stereo algorithm will then try to match features in the sky, which do not correspond to the same point on the water surface. This situation is illustrated in figure 1.9. For a wavy surface, this effect can lead to a systematic bias in the estimated wave height as well as missing correspondences, as is shown in Jähne et al. [59].

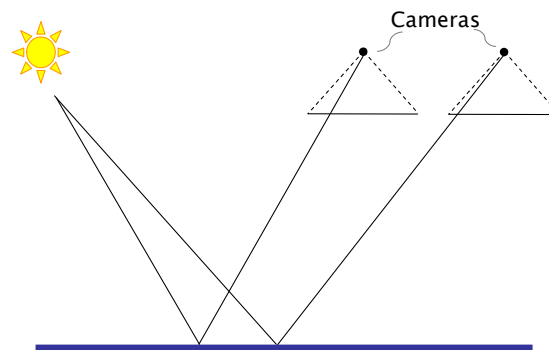


Figure 1.9: Due to the view-angle dependence of specular reflection, corresponding features seen in the two camera images originate not from the same patch on the water surface, but from the same region of the sky.

1.4 Aim and own contribution

Section 1.1 illustrated the need for simultaneous measurements of air-sea fluxes and the parameters, such as wave slope, that influence the transport across the interface. Specifically, combining heat flux measurements using thermographic imaging with wave measurements is desirable. In the laboratory environment this can be achieved by combining gas and heat flux measurements with slope measurements using a Color Imaging Slope Gauge (see section 1.2.2) (see for example [106]).

However, existing field data from combined measurements of the water surface and air-sea fluxes is scarce, largely due to the difficulties with wave imaging. To help fill this gap, the aim of the present work was to design an instrument that can be used to reconstruct the shape and motion of short to medium gravity waves on the ocean, either from a fixed platform or from a research vessel. In this respect, the present work addresses some of the aims put forth in the science plan of the Solas program [69].

Based on the experience from previous work (discussed in section 1.2) the following features are desirable for such an instrument:

1. The instrument should be field deployable for use off the boom of a ship during a research cruise, from a pier or from an offshore platform. Rigging should be easy and fast.
2. The system should be operable over a wide range of wind speeds and other meteorological conditions.
3. The instrument should be rugged enough to withstand the harsh operating conditions at sea.
4. The water waves should not be disturbed, for example by placing parts of the instrument underneath the surface.
5. The instrument should be as insensitive as possible to variations in natural illumination conditions.

The approach taken in this work is a *stereo infrared imaging wave gauge*, which consists of two infrared cameras and uses stereo photogrammetry to recover the surface shape of short gravity waves. In addition to wave reconstruction, the infrared image sequences acquired with this instrument can be used to estimate air-sea heat and gas fluxes as described in section 1.1.3.

Operating in the infrared wavelength region from 3–5 μm solves many of the problems of previous stereo wave imaging systems, arising from the optical properties of water at visible wavelengths (mentioned in section 1.3), as follows:

- Textured: Infrared images of the water surface show a rich texture when a heat flux cools down the surface. An example was given in figure 1.3.
- Opaque: The maximum penetration depth of infrared radiation in water is less than 0.1 mm in the 3–5 μm range.
- Emissive: Every warm object radiates electromagnetic energy according to Planck’s law (see section 2.1.2).
- Lambertian: In the 3–5 μm region, the brightness of the radiation emitted by a water surface changes only by about 5% for viewing angles up to 60° relative to the surface normal, see section 2.1.4.

The temperature patterns observed at the water surface whenever a heat flux is present create a rich texture in the infrared images, with features that can be used to match corresponding points in the stereo image pairs automatically (see section 4.2.4).

Opacity and almost Lambertian surface characteristics ensure that these features are bona fide surface features for a large range of viewing angles. In contrast, the transparency and specularity of water in the visible region can give rise to visible image features that are below the water surface or are view-dependent mirror images of surrounding features like clouds, respectively (see sections 4.2.3 to 4.2.4).

The infrared radiation detected by the two cameras is emitted by the water surface itself, according to Planck’s Law (see section 2.1.2). Therefore, no light source is needed, which means that the infrared stereo wave gauge is a *passive* remote sensing device.

Although the present system has only been evaluated in the laboratory, monocular infrared image sequences of similarly good quality (high contrast and a rich texture) were obtained on previous research cruises (CoOp 1995,1997 [7] and GasEx 2001) under different wind speed regimes, as reported by Schimpf [83]. However, the best results are obtained at dusk or at night, when the negative heat flux cooling the surface is usually strongest due to the lessened solar radiative flux. Stronger cooling makes for a larger temperature difference between the bulk water and the top thermal skin, giving higher contrast in the thermal images.

Measuring at night also avoids specular sun glints in the images. These can, however, still be a problem, although they are less severe than at visible wavelengths, where the maximum of the solar blackbody spectrum is located.

Previous experience with thermographic imaging at sea also shows that the infrared cameras can be made rugged enough to withstand and operate under harsh

conditions by placing them in a protective casing as described by Garbe [35, section 12.2].

The image sequences obtained from each of the two infrared cameras making up the stereo system can be used to analyze the heat flux at the water surface, using the techniques described by Schimpf et al. [84] and Garbe et al. [37]. In fact, the additional depth information obtained with the stereo camera setup should help to improve the accuracy of the heat flux and transfer velocity estimates obtained using the methods described in [37]. These methods rely on the accurate calculation of *optical flow* from the infrared image sequences. The wave motion changes the distance of the water surface from the camera, introducing a diverging or converging component to the optical flow. Using the depth information obtained from stereo imaging, it is possible to correct for this effect.

In summary, the infrared stereo wave gauge presented in this thesis can be classified as an imaging, passive, close-range, remote-sensing device, where the individual attributes have the following meaning:

- Imaging: In contrast to point measurements, spatial variations can be recorded.
- Passive: No light source is required, in contrast to active remote sensing devices.
- Close-range: The distance of the object from the camera is only one order of magnitude larger than the baseline of the system.
- Remote-sensing: The instrument does not touch or interfere with the surface.

1.5 Thesis Outline

Chapter 2 introduces the basic concepts and terminology of radiometry, with an emphasis on infrared radiation. It is shown that a warm water surface radiates electromagnetic energy similarly to a blackbody and how, based on this radiation, infrared cameras can be used to measure the temperature distribution at the water surface. The optical properties of water in the infrared region are also presented.

Chapters 3 and 4 explain the foundations of stereo computer vision. As this work might be of interest to physical oceanographers without a background in computer vision, most of the algorithms used are stated explicitly. The problem of stereo computer vision can be divided into two main parts: the geometric concepts needed to extract metric information from image pairs; and the automatic matching of corresponding points across the two images, known as disparity estimation. The geometric aspects are the focus of chapter 3, where camera calibration and two-view

geometry are presented using the notation of homogeneous coordinates, a useful concept from projective geometry. Disparity matching is the topic of chapter 4, where the matching algorithm used for the stereo infrared wave gauge is outlined, following a review of existing algorithms and underlying model assumptions.

The image sequences obtained from the infrared cameras cannot be used directly for the algorithms outlined in chapters 3 and 4, as they first have to be corrected for the different sensitivities of the sensor elements in the camera. Also, gaps in the images due to defective camera pixels must be interpolated. The output data of the stereo algorithm also contains gaps in image regions where no disparity estimate could be obtained. Both input and output images are interpolated in a similar manner, using a technique known as regularization, detailed in chapter 5. Chapter 5 also discusses some other well-established image processing methods that have been used in this work, namely non-uniformity correction and outlier removal.

The experimental setup and an overview of the experiments performed with the infrared stereo wave gauge at the Heidelberg wind-wave channel are detailed in chapter 6 and the results are presented in chapter 7.

The conclusion in chapter 8 contains a critical discussion of the results obtained and an outlook on possible future developments.

For better readability, some mathematical derivations that would otherwise disrupt the text, are presented in appendices. Appendix A discusses the factors that influence the accuracy of the 3D points reconstructed with the infrared stereo camera system. The theoretical limit on range resolution achievable using triangulation is analyzed. To obtain an experimental assessment of the achievable accuracy, a method for robustly fitting a reference plane to the reconstructed points of a flat water surface is presented.

Appendix B provides a detailed explanation of one of the steps involved in camera calibration, namely the estimation of a projective transform, or homography, from a set of point correspondences.

2 Basics of Infrared Imaging

This chapter discusses what can be seen in the images obtained by an infrared camera. When trying to analyze infrared images, the thermal radiation of warm bodies as described by Planck’s law, the thermal and optical properties of the emitting surfaces and the environment, and the interaction of the radiation with the detector become important.

The use of infrared imaging, or thermography, distinguishes this work from the majority of stereo computer vision applications, which usually operate in the visible wavelength region. The main difference is that in the infrared region of the electromagnetic spectrum, every object emits radiation (as will be seen in section 2.1.3). That is, every scene object is also a “light source”, whereas at visible wavelengths, objects usually reflect or re-emit the radiation from a light source.

In section 2.1.1, some radiometric terms needed for the subsequent sections are defined. The nomenclature and units follow the recommendations of the *International Association of the Physical Sciences of the Ocean* (IAPSO) (see, for example, [49]).

Section 2.1.2 reviews the physics governing the thermal emission of electromagnetic radiation for the idealized case of a blackbody and introduces Planck’s law. The thermal emission and radiometric properties of real world objects are then addressed in section 2.1.3. Due to their importance for the application to wave imaging, the properties of water in the infrared region are stated in section 2.1.4; some of the properties of water in the visible region have already been mentioned in section 1.3. Specifically, the assumption that the water surface is approximately *Lambertian* in this wavelength region, which will be used for the stereo matching described in chapter 4, is justified.

Section 2.2 describes the physical principles by which the cameras used in the present application detect infrared radiation and defines the parameter $NE\Delta T$ that is commonly used to specify the thermal sensitivity of a detector. For current infrared imagers, the sensitivity shows strong variations between individual pixels, so a radiometric calibration must be performed, as explained in sections 2.2.3 and 5.1. Radiometric calibration also forms the basis for temperature measurements using an infrared detector.

2.1 Radiometry

2.1.1 Definitions

In this section, the definitions of several radiometric quantities are given, which provide the basis for the subsequent sections. The concept of a Lambertian surface is also introduced. For a treatment that goes beyond the bare definitions, the reader is referred to the summary by Haußecker [45]. The notation follows the recommendations of the *International Association of the Physical Sciences of the Ocean* (see [49]) .

Preceding the definition of the actual radiometric quantities, the terms *spherical coordinates* and *solid angle* are defined, as they are subsequently used to specify directional dependence. Table 2.1 summarizes the definitions and the units of the different radiometric quantities.

Spherical Coordinates Using spherical coordinates, any point in three dimensional space can be expressed by two angles θ, ϕ and a distance r from a point of reference, as shown in figure 2.1. To represent a direction, only the angles θ and ϕ are needed.

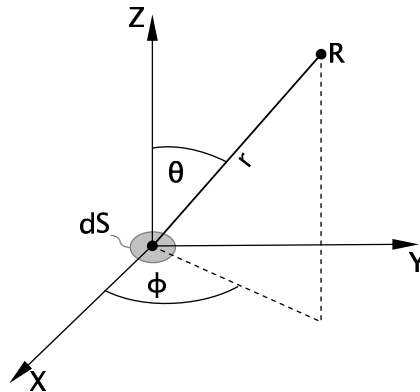


Figure 2.1: Spherical coordinates: In spherical coordinates a point \mathbf{R} is represented by the angles ϕ, θ and the distance r . The directional dependence of the flux emanating from a surface patch dS is also expressed in terms of ϕ and θ .

Solid Angle It is straightforward to extend the notion of an angle in two dimensions, defined as the length of a segment on the unit circle, to the three dimensional case where the *solid angle* is defined as the area of a surface patch on the unit

sphere. This surface patch can be interpreted as the projection of an arbitrarily shaped patch onto the unit sphere, as is depicted in figure 2.2. Since the solid angle is defined as a ratio of two areas, it is dimensionless. As this can lead to confusion, the artificial unit *sterad* [sr] is used to mark quantities with a directional dependence.

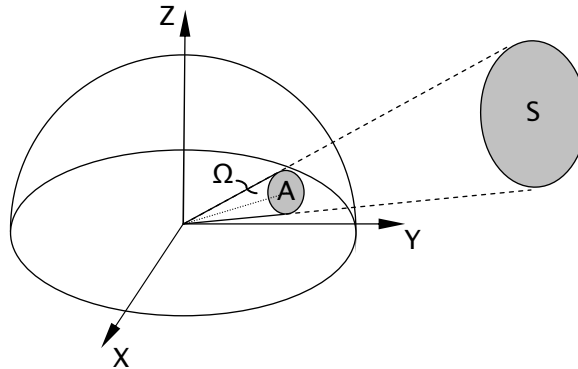


Figure 2.2: Solid angle: The surface patch S covers a solid angle Ω . The solid angle is defined as the fractional area of surface of the unit sphere covered by the projection A of S .

Radiant Energy and Flux The total *radiant energy* emitted or received by an object is denoted by Q and is measured in Joule [J]. The temporal derivative of radiant energy is the *radiant flux* or *radiant power* Φ , which has the units Watt [W] or [Js⁻¹].

Exitance and Irradiance *Radiant exitance* M is a measure of the radiative flux emitted per unit surface area. The corresponding quantity for incident (as opposed to emitted) flux is *irradiance* E . Both M and L are used to describe an extended area.

$$\text{Exitance} \quad M = \frac{d\Phi_{\text{emitted}}}{dS} \left[\frac{\text{W}}{\text{m}^2} \right] \quad (2.1)$$

$$\text{Irradiance} \quad E = \frac{d\Phi_{\text{incident}}}{dS} \left[\frac{\text{W}}{\text{m}^2} \right] \quad (2.2)$$

In general M and E are functions of the position on the surface $M(\mathbf{x}), E(\mathbf{x})$.

Radiant Intensity *Radiant intensity* I is defined as the radiant flux emitted per unit solid angle Ω from a point in space.

$$\text{Radiant intensity} \quad I = \frac{d\Phi_{\text{emitted}}}{d\Omega} \quad \left[\frac{\text{W}}{\text{sr}} \right] \quad (2.3)$$

Radiance *Radiance* L is a function of both position and direction. It describes the radiant flux per *foreshortened* (projected) unit area per unit solid angle emanating from a given position in a given direction. The foreshortened area specifies a surface element perpendicular to the propagation direction of the radiation.

$$\text{Radiance} \quad L = \frac{d\Phi}{d\Omega dS_{\perp}} = \frac{d\Phi}{d\Omega dS \cos(\theta)} \quad \left[\frac{\text{W}}{\text{m}^2 \text{sr}} \right] \quad (2.4)$$

If the radiance is known, the other radiometric quantities introduced above can easily be obtained from it by integration. For example, combining equations (2.1) and (2.3) with equation (2.4) yields the following relations:

$$L \cos \theta = \frac{dI}{dS} \quad (2.5)$$

$$L \cos \theta = \frac{dM}{d\Omega}. \quad (2.6)$$

Integrating dM over the hemisphere surrounding the point of reference, and integrating dI over the surface patch S , yields

$$M(\mathbf{x}) = \int_{\text{hemisphere}} L(\mathbf{x}, \theta, \varphi) \cos \theta d\Omega = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} L(\mathbf{x}, \theta, \varphi) \cos \theta \sin \theta d\varphi d\theta \quad (2.7)$$

and

$$I(\theta, \varphi) = \int_S L(\mathbf{x}, \theta, \varphi) \cos \theta dS, \quad (2.8)$$

respectively. Note that equation (2.8), is an extension of the original definition of intensity (2.3), which describes point sources only.

Lambertian Surfaces A surface which looks equally “bright” from all directions is said to be *Lambertian* or *ideal diffuse*. More precisely, this means that its radiance is independent of view angle, that is $L(\mathbf{x}, \theta, \varphi) = L(\mathbf{x})$. From equation (2.8) it follows that for a Lambertian surface the angular distribution of intensity obeys the relation

$$I(\theta) = \int_S L(\mathbf{x}) \cos \theta dS = I_0 \cos \theta, \quad (2.9)$$

known as *Lambert’s cosine law*. This is depicted in figure 2.3.

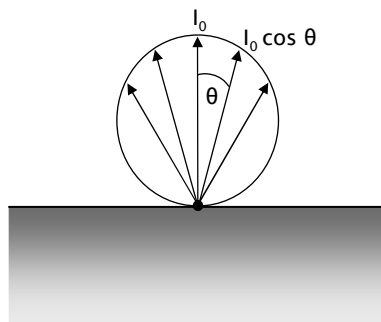


Figure 2.3: Lambert’s cosine law: For a Lambertian surface the angular variation of emitted intensity follows the functional relationship $I(\theta) = I_0 \cos \theta$, which is schematically represented by the lengths of the arrows.

The concept of a Lambertian surface is idealized, but some materials like cotton cloth or matte paper at visible wavelengths, and (as is shown in section 2.1.4) water in the infrared region are reasonably close to this ideal. The last example is important for the present application since the correlation-based stereo matching described in chapter 4, like many computer vision algorithms, is based on the assumption that the observed object has Lambertian surface characteristics. (see section 4.2.2).

Lambert’s cosine law appears, at first glance, to contradict the definition of a Lambertian surface as being equally “bright” from all view angles because of the cosine factor. However, when calculating the flux the detector receives from a given solid angle, the cosine factor is canceled by the foreshortening effect: for an oblique view angle θ , the observed surface patch is larger by a factor of $1/\cos \theta$ as depicted in figure 2.4. Since radiance, by definition, accounts for this foreshortening effect, it is constant independent of the angle θ . Note that in this discussion it was assumed that only the emitting surface is tilted, while the detector surface (CCD chip) remains perpendicular to the view direction. For a discussion of arbitrary source-detector geometry see for example Jähne et al. [57, section 5.2.1].

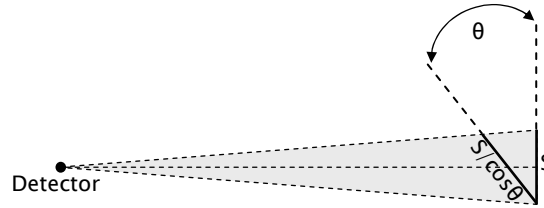


Figure 2.4: Effect of foreshortening: For a given solid angle, a detector receives flux from a larger area of a surface tilted by angle θ than from a surface perpendicular to the view direction at the same distance.

Spectral Distribution The radiant energy Q is generally comprised of radiation in a spectrum of different wavelengths. To account for this, the *spectral radiant energy* Q_λ is defined as

$$Q_\lambda(\lambda) = \frac{dQ(\lambda)}{d\lambda} \quad [\text{J m}^{-1}], \quad (2.10)$$

which is the radiant energy within the infinitesimal wavelength interval $[\lambda, \lambda + d\lambda]$. For all of the radiometric quantities introduced above, a corresponding spectral quantity can be defined analogous to equation (2.10). These spectral quantities will hereafter be denoted with a subscript λ .

Photon Flux The infrared cameras used for the present application are photon detectors. As will be shown in section 2.2.3, their detector output depends on the number of incident photons. Therefore, it is sometimes useful to formulate the radiometric quantities defined above in terms of the number of photons rather than in terms of energy. For the spectral quantities this can be done by dividing the energy-based radiometric quantities by the energy of a single photon at the given wavelength, $E_{\text{photon}} = \frac{hc}{\lambda}$, to obtain a photon count. These photon-based quantities are denoted with a subscript p. For example, the spectral exitance expressed in terms of the number of photons emitted per unit time and unit area is

$$M_{p,\lambda} = \frac{\lambda}{hc} M_\lambda. \quad (2.11)$$

2.1.2 Electromagnetic Radiation of a Blackbody

In 1800, during his experiments with sunlight, Herschel discovered that there is some sort of “heat radiation” that is not tied to visible light. With a prism

Quantity	Symbol	Unit	Definition
Radiant energy	Q	J	Total energy emitted by a source or received by a detector
Radiant flux	Φ	W	Total power emitted by a source or received by a detector
Radiant exitance	M	Wm^{-2}	Power emitted per unit surface area
Irradiance	E	Wm^{-2}	Power received at unit surface element
Radiant intensity	I	Wsr^{-1}	Power leaving a point on a surface into unit solid angle
Radiance	L	$\text{Wm}^{-2}\text{sr}^{-1}$	Power leaving unit projected (i.e. perpendicular to direction of travel) surface area into unit solid angle

Table 2.1: Summary of radiometric quantities and their units.

and a thermometer he observed this “heat radiation” close to the red part of the visible spectrum and named it “infrared radiation”. With Maxwell’s theory of electromagnetism (1864) this radiation could be explained as the electromagnetic radiation emitted due to the acceleration of charged particles by thermal motion.

Planck’s law of Radiation In 1900, Planck [74] derived an explanation for the spectral distribution of thermal radiation, based on the assumption that the emitter can only lose discrete quanta of energy. This marked the start of quantum mechanics and modern physics.

Other than temperature, the spectral distribution of electromagnetic radiation also depends on the physical properties of the emitting surface (see section 2.1.3). To address this, Kirchhoff introduced the concept of an ideal surface that neither transmits nor reflects, but absorbs all incident radiation independent of wavelength and direction. Such a perfect absorber of radiation is called a *blackbody*. By the principle of conservation of energy, a blackbody must also be a perfect emitter of radiation. This is a particular case of *Kirchhoff’s law* (see below).

Planck’s law for the spectral distribution of radiation by a blackbody is given by:

$$\begin{aligned}
 M_\lambda(T)d\lambda &= \frac{2\pi hc^2}{\lambda^5} (e^{hc/k_b T \lambda} - 1)^{-1} d\lambda \\
 M_\nu(T)d\nu &= \frac{2\pi h\nu^3}{c^3} (e^{h\nu/k_b T} - 1)^{-1} d\nu,
 \end{aligned}
 \tag{2.12}$$

where T is the absolute temperature in Kelvin [K], $c = 3 \cdot 10^8$ [ms⁻¹] is the speed of light, $k_b = 1.38 \cdot 10^{-23}$ [J/K] is *Boltzmann's constant* and $h = 6.62 \cdot 10^{-34}$ [Js] is *Planck's constant*.

To switch between the formulations of equation (2.12) for wavelength λ and frequency ν the relation $\frac{d\lambda}{d\nu} = -c/\nu^2$ is used.

In equation (2.12), Planck's law is stated for spectral radiant exitance, in contrast to many physics textbooks that formulate Planck's law in terms of the spectral energy density $\rho_\lambda(T)$, which has the units [Jsm⁻³] and is related to the spectral radiant exitance by $\rho_\lambda(T) = 4cM_\lambda(T)$.

The spectral radiant exitance of a blackbody at different temperatures is given in figure 2.5 (for simplicity, these spectral distributions will be referred to as blackbody curves). With increasing temperature, the maximum of the blackbody curve, λ_{\max} , shifts to shorter wavelengths. If the temperature is high enough, radiation is also emitted in the visible wavelength range. This effect forms the basis for the incandescent lamp. Note that the blackbody curves for different temperatures do not intersect, meaning that the spectral radiant exitance increases monotonically with temperature for all wavelengths. This monotonic increase is important, as it leads to the possibility of measuring the temperature of a blackbody surface from its radiant exitance within any spectral subregion of the blackbody curve, as explained in section 2.2.3.

Wien's Displacement Law The temperature dependence of the wavelength λ_{\max} , at which the blackbody curve has its maximum, is known as *Wien's displacement law*:

$$\lambda_{\max} \cdot T = \text{constant} = 2891 \mu\text{m} \cdot \text{K}. \quad (2.13)$$

Stefan-Boltzmann Law The total radiant exitance of a blackbody is obtained by integrating equation (2.12) over all wavelengths, leading to the *Stefan-Boltzmann law*:

$$M_{\text{total}}(T) = \int_0^{\infty} M_\lambda(\lambda, T) d\lambda = \sigma \cdot T^4 \quad (2.14)$$

where $\sigma = 5.67 \cdot 10^{-8} \text{Wm}^{-2} \text{K}^{-4}$.

Kirchhoff's Law As mentioned in the introduction to blackbodies, a perfect emitter is also a perfect absorber. In fact, for a surface in thermodynamic equilibrium, there is a more general rule: its emissivity is always equal to its absorptivity independent of direction, temperature and wavelength. This is known as *Kirchhoff's*

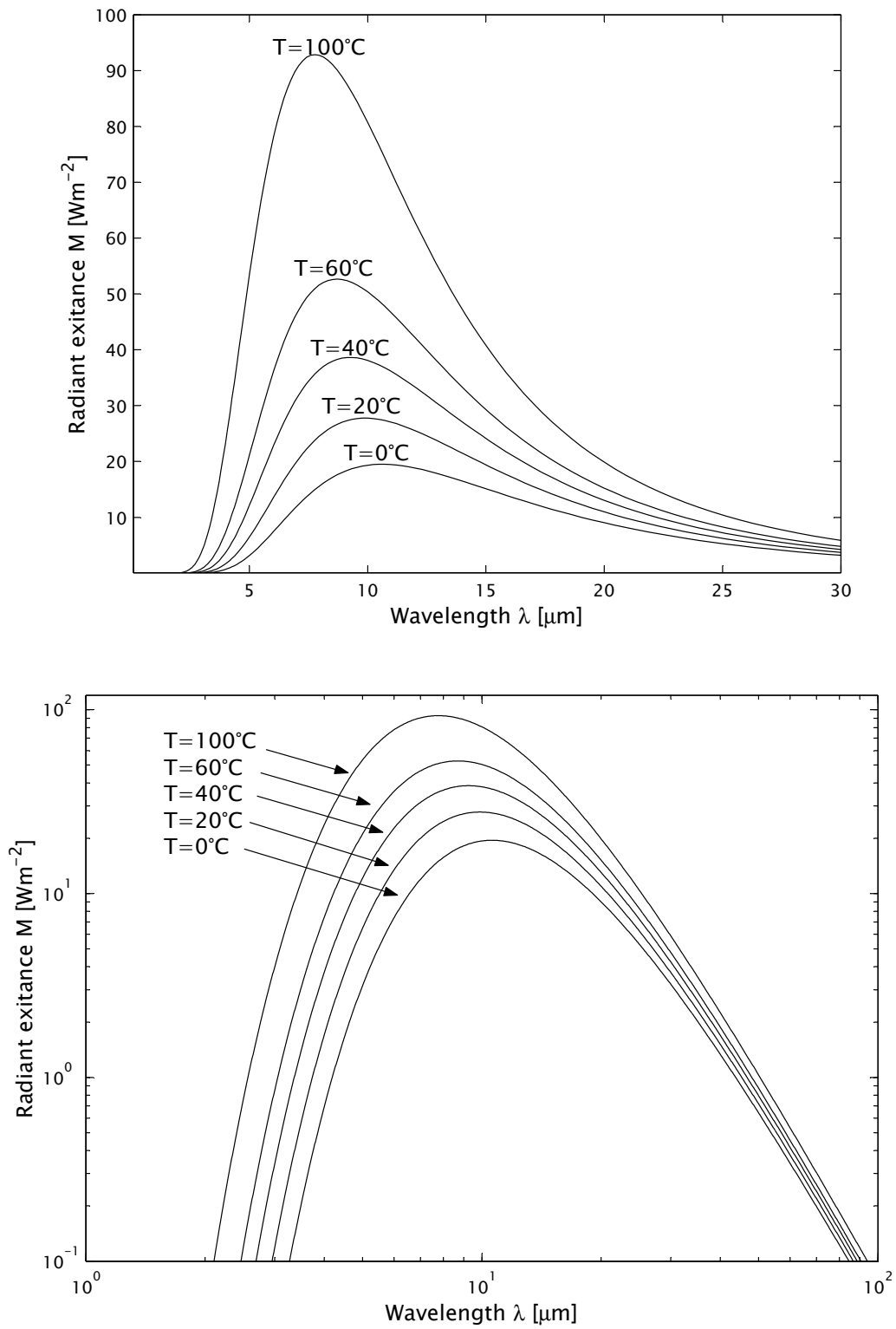


Figure 2.5: Blackbody curves at different temperatures.

Law and is due to the conservation of energy. If there were an imbalance between absorptivity and emissivity, the surface would either lose energy to, or gain energy from, its environment, violating the assumption that it is in thermal equilibrium.

2.1.3 Emissive Properties of Real Surfaces

Emission from real objects differs from blackbody emission, because the physical surface properties affect the angular distribution and the spectral characteristics of the emitted radiation. Furthermore, a real surface does not absorb all of the incident radiation, but reflects or transmits some fraction of it. Surface properties also change with temperature, complicating the situation further.

The radiometric effects of the surface properties are described by the four parameters *emissivity* ϵ , *absorptivity* α , *reflectivity* ρ and *transmissivity* τ .

Emissivity is defined as the ratio of emitted radiation for the surface in question to the emitted radiation of a blackbody at the same absolute temperature:

$$\epsilon = \frac{M_{\text{surface}(T)}}{M_{\text{blackbody}(T)}}. \quad (2.15)$$

Absorptivity is defined as the fraction of the total irradiance absorbed by the surface:

$$\alpha = \frac{E_{\text{absorbed}}}{E_{\text{total}}}. \quad (2.16)$$

From Kirchoff's Law it follows that $\alpha = \epsilon$.

Reflectivity is similarly defined as the fraction of the total irradiance reflected by the surface:

$$\rho = \frac{E_{\text{reflected}}}{E_{\text{total}}}. \quad (2.17)$$

Finally, transmissivity is a parameter describing transparent surfaces and is defined as the fraction of the total irradiance transmitted through the surface:

$$\tau = \frac{E_{\text{transmitted}}}{E_{\text{total}}}. \quad (2.18)$$

As indicated above, these four parameters generally have an angular as well as a spectral dependence, and for some objects they vary with temperature (e.g. the emissivity of glowing hot objects approaches 1). So, in general, these parameters are functions of wavelength and direction, for example in the case of emissivity $\epsilon = \epsilon(\lambda, \theta, \phi, T)$.

By conservation of energy, it follows that

$$\rho + \alpha + \tau = 1. \quad (2.19)$$

For all non-transparent surfaces and for sufficiently thick bulk material, in which all radiation will eventually be absorbed, this simplifies to

$$\alpha = \epsilon = 1 - \rho. \quad (2.20)$$

The nomenclature used here reflects the common use in the infrared community, although it is slightly incorrect:

“The emissivity, reflectivity and absorptivity are properties of ideal materials. Real world materials have defects, surface irregularities, and may contain a variety of trace materials. Real materials are characterized by emittance, reflectance, and absorptance. Unfortunately, the infrared community indiscriminately mixes emissivity with emittance. Since emissivity is used more often this book will use emissivity for real materials.” Holst [51]

Non-blackbody surfaces can be classified according to their emissivity as follows:

- *Greybodies* have an emissivity $\epsilon < 1$ that is independent of angle and wavelength. That is, their spectral emissivity differs from the blackbody curve only by a constant factor, namely ϵ . For a greybody $\rho = 1 - \epsilon$.
- *Selective emitters* are the most general case; their emissivity varies with wavelength.

Figure 2.6 compares the spectral radiant exitance of a blackbody, a greybody and a selective emitter at the same temperature.

2.1.4 Optical Properties of Water in the Infrared Region

This section states some optical properties of water at infrared wavelengths that are important for the present application. The values for the reflectivity and the penetration depth are derived from data on the complex refractive index of water in the infrared region that was compiled by Downing and Williams [22]. As this derivation is based on the Fresnel equations, these are briefly reviewed. It is shown that the reflectivity of water at infrared wavelengths is very small and depends only slightly on the angle of incidence. Therefore, for many practical purposes the water surface can be considered to be Lambertian.

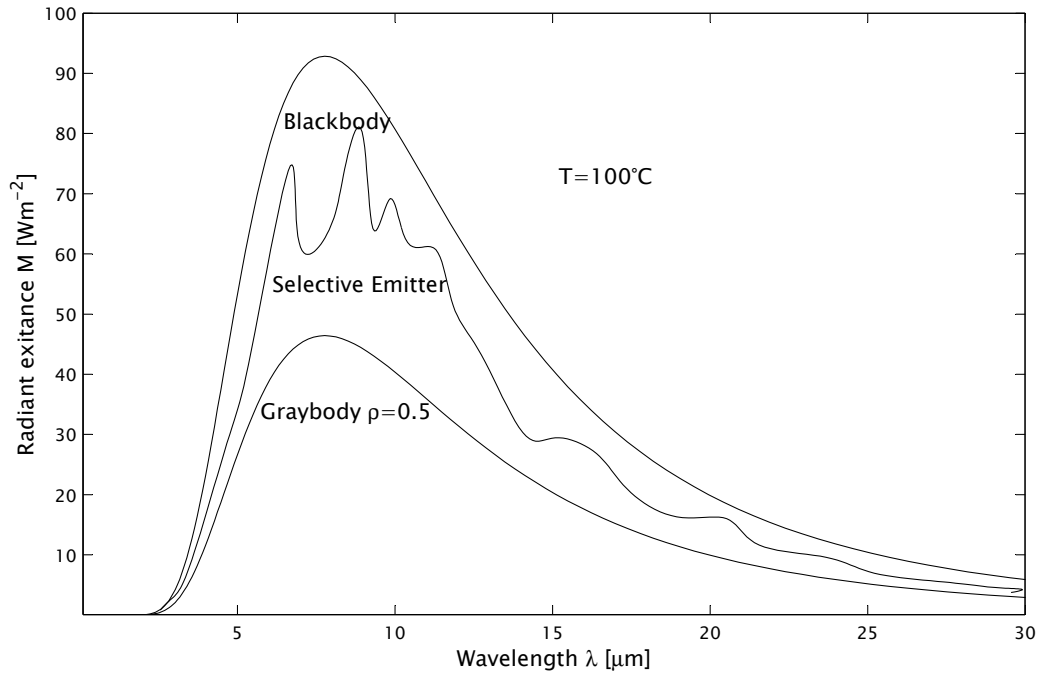


Figure 2.6: Spectral radiant exitance for real objects. The spectra of a selective emitter (schematic) and a greybody with an emissivity of 0.5 are compared to a blackbody at the same temperature (100°C).

Complex Index of Refraction

When an electromagnetic field is incident on the surface of a material, it exerts a force on particles with non-zero electric moments (such as electrons and polar molecules). These particles are thereby accelerated, re-emitting some of the incident radiation and absorbing some. The details of this interaction are fairly complex and depend, among other things, on the crystal structure, and the electronic and molecular excitation levels of the material.

For bulk matter, the average effects of these processes are described by the *dielectric function* $\epsilon(\lambda)$, not to be confused with the emissivity, and the *complex refractive index*

$$N(\lambda) = n(\lambda) - ik(\lambda) \quad (2.21)$$

derived from it [see e.g. 28, 64, 100].

The real and imaginary parts $n(\lambda)$ and $k(\lambda)$ of the complex refractive index for water at infrared wavelengths are displayed in figure 2.7.

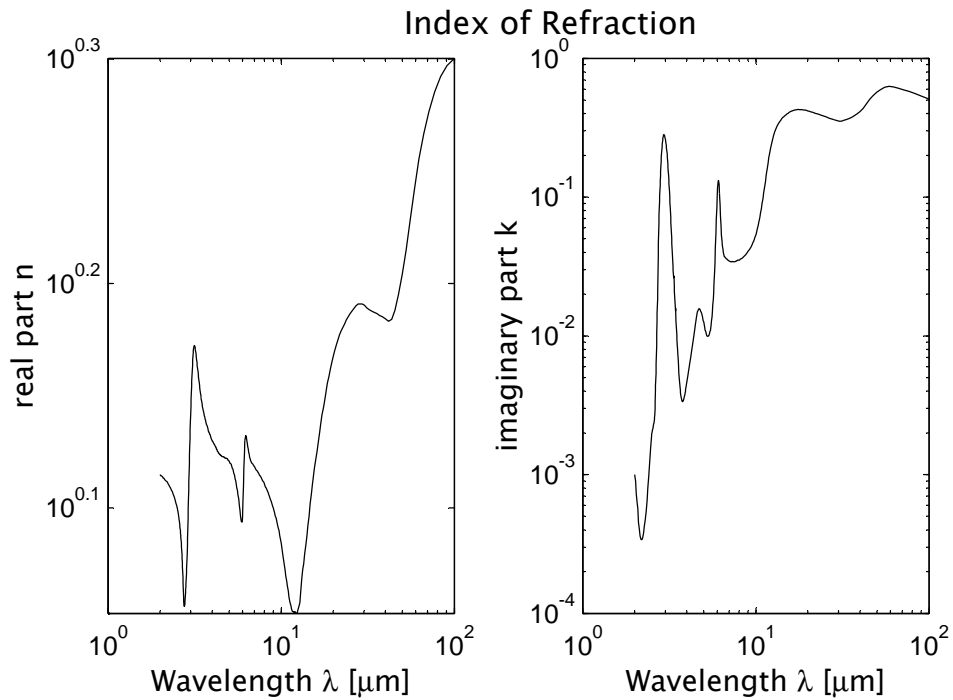


Figure 2.7: Real part, n , and imaginary part, k , of the complex index of refraction for water at infrared wavelengths. From data collected and tabulated by Downing and Williams [22]

Reflectivity The reflectivity of a material can be calculated from the index of refraction using the *Fresnel equations*:

$$\rho_{\perp}(\theta, \psi) = \left| \frac{N \cos \theta - \cos \psi}{N \cos \theta + \cos \psi} \right|^2 \quad \text{and} \quad (2.22)$$

$$\rho_{\parallel}(\theta, \psi) = \left| \frac{\cos \theta - N \cos \psi}{\cos \theta + N \cos \psi} \right|^2, \quad (2.23)$$

where the angles θ and ψ are defined as in figure 2.8, and ρ_{\perp} and ρ_{\parallel} are the reflectivity for radiation polarized perpendicular to and parallel to the plane of incidence, respectively. The plane of incidence is the plane spanned by the propagation

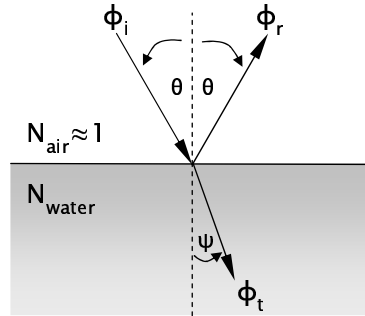


Figure 2.8: Incident flux, ϕ_i , reflected flux, ϕ_r and transmitted (refracted) flux, ϕ_t , at the air-sea interface. The angle of incidence θ is equal to the angle of reflection. The angle of refraction is denoted ψ .

direction of the electromagnetic wave and the surface normal. For unpolarized radiation, the reflectivity is obtained by taking the mean:

$$\rho = \frac{1}{2} (\rho_{\perp} + \rho_{\parallel}). \quad (2.24)$$

The value of $\cos \psi$ can be derived using *Snell's law* of refraction (see, for example [100]):

$$\cos \psi = \sqrt{1 - \frac{\sin^2 \theta}{N^2}}. \quad (2.25)$$

Although many textbooks only state equations (2.22) and (2.23) for a purely real refractive index $N = n$, they also hold if the imaginary part k is non-vanishing (see [40]).

The angular dependence of the reflectivity of water in the infrared region, computed using equation (2.22), is displayed in figure 2.9. The real and imaginary parts of the refractive index used to calculate this dependence were obtained by averaging $n(\lambda)$ and $k(\lambda)$, as tabulated in Downing and Williams [22], over the $3 - 5 \mu\text{m}$ wavelength region. Note that the reflectivity is very small and almost constant for angles of up to 55° . Therefore, water at these wavelengths has surface properties that are similar to a Lambertian surface.

Penetration Depth If the imaginary part of the complex refractive index is non-zero, the radiative flux is attenuated in the medium. After a distance z in the medium, an initial flux Φ_0 is reduced to:

$$\Phi_z(\lambda) = \Phi_0 e^{-\beta(\lambda)z}, \quad (2.26)$$

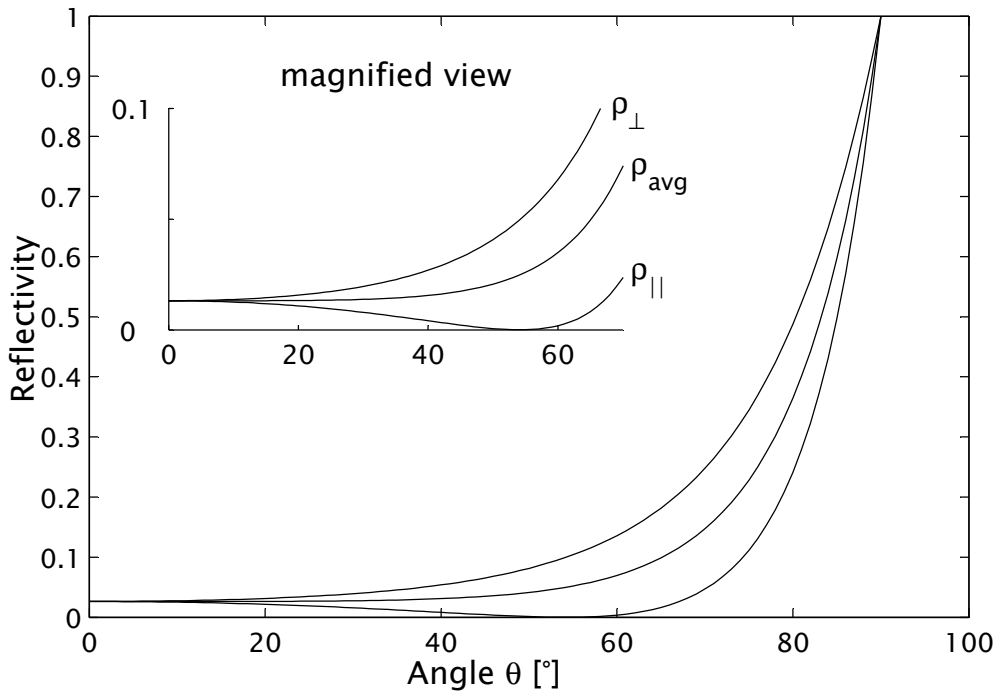


Figure 2.9: Angular dependence of reflectivity of water obtained from equations (2.22), (2.24) and the complex index of refraction (2.21). The values were calculated with the real part and the imaginary part of the refractive index chosen as $n=1.384$ and $k=0.0412$ respectively. These values are the mean of n and k over the $3-5 \mu\text{m}$ interval.

where β is the *absorption coefficient*. It is related to the imaginary part k of the refractive index by

$$\beta(\lambda) = 4\pi \frac{k(\lambda)}{\lambda}. \quad (2.27)$$

The *penetration depth*, $\zeta(\lambda)$, defined as the distance in the medium over which the flux is reduced by a factor of e , is given by the inverse of β :

$$\zeta(\lambda) = \frac{1}{\beta(\lambda)}. \quad (2.28)$$

For water in the $3-5 \mu\text{m}$ wavelength region, which is the region in which the cameras used for the present research are sensitive (see section 6.1), the penetration depth varies considerably, as can be seen in figure 2.10. However, for this application the important result is that the penetration depth in this region is always less than $100 \mu\text{m}$. This means that if water is imaged in this wavelength range,

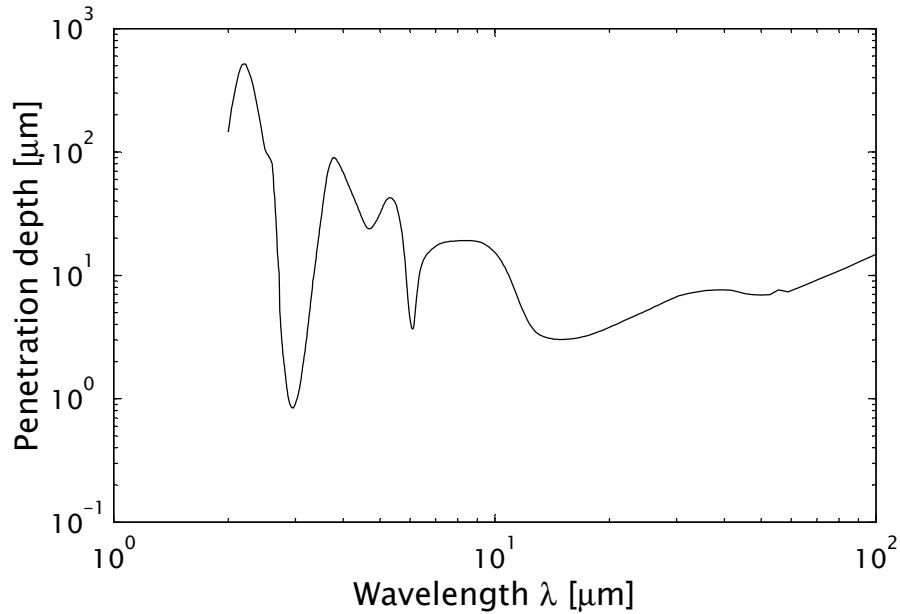


Figure 2.10: Spectral dependence of the penetration depth of infrared radiation in water. The penetration depth is lower than $100\ \mu\text{m}$ throughout the $3 - 5\ \mu\text{m}$ wavelength region. From data published by Downing and Williams [22].

only the temperature distribution at the surface is observed, fulfilling an important prerequisite for stereo matching (see section 4.2.3). This is in stark contrast to water in the visible part of the spectrum, where the transparency at normal incidence is quite high. Exact values in the visible spectrum vary considerably (see Hoejerslev [49]), depending on dissolved substances and suspended particles, but everyday experience shows that one can look through several meters of clear water and, for example, recognize the tiles at the bottom of a swimming pool.

2.2 Infrared Detectors

This section reviews the principles by which infrared radiation is detected and how the detector output can be used to sense temperature distributions remotely.

Two common detector types, thermal detectors and photon detectors, are presented in section 2.2.1. The thermal sensitivity of infrared detectors is often specified in terms of the *noise equivalent temperature difference*, a performance parameter that is explained in section 2.2.2. Section 2.2.3 discusses how the detector output is related to the temperature and to the surface properties of the

imaged object. This discussion is for photon detectors, as the cameras used in this work (see section 6.1) are based on this detector type, but similar principles apply to thermal detectors.

2.2.1 Types of detectors

There are several different types of infrared detectors. Imaging infrared detectors are often referred to as *focal plane array*, or FPA. The most common types of FPAs are *thermal detector arrays* and *photon detector arrays*.

Thermal detectors detect infrared radiation by secondary effects due to the heat deposited by the radiation incident on the detector. Typical examples of thermal detectors are bolometers. In contrast to thermal detectors, photon detectors detect infrared radiation directly, through charges created by incident photons.

Bolometers Bolometers typically consist of two material types, one with high absorptivity and one with a heat-dependent resistance. Incoming infrared radiation deposits energy in the highly absorptive material, such that heat is generated. This absorbed heat causes a change in the electrical resistance of the second material, which is in thermal contact with the absorber. The change in electrical resistance can be transformed into a voltage signal. For use in infrared imagers, a number of small bolometers are arranged to form a micro-bolometer array, where each bolometer accounts for one image pixel. Most commercially available micro-bolometer arrays are sensitive in the $8 - 14 \mu m$ wavelength region and can be operated without cooling.

Photon Detectors Photon detectors directly transform incoming infrared radiation into electrical charges. The detector consists of a semiconductor, which has a band gap between the valence band and the conduction band. For infrared detectors, the semiconductor material is chosen such that the potential energy difference between these bands is on the same order of magnitude as the energy of infrared photons in the desired wavelength range. Incident infrared photons are absorbed and can excite electrons from the valence band into the conduction band, creating free electrical charges and thus a measurable electronic signal.

The fraction η of photons that excite an electron into the conduction band is called the *quantum efficiency* of a detector and is generally a function of wavelength.

As electrons can also be thermally excited into the conduction band, infrared photon detectors are cooled to improve the signal-to-noise ratio. Typical operating

temperatures are in the range between 70 K and 77 K and are achieved by Stirling coolers or by using liquid nitrogen. Common detector materials are Indium Antimonide (InSb), Platinum Silicide (PtSi) and Mercury Cadmium Telluride (HgCdTe). Another group of photon detectors are *quantum well inter-sub-band photodetectors*, or QWIPs. These are based on sandwiched layers of AlGaAs and GaAs. The layers form quantum wells, which give rise to quantized electron bands. The band gap between different states, and therefore the infrared wavelength region in which the detector is sensitive, can be tailored by varying the thickness of the different layers.

Detectors based on InSb and PtSi typically operate in the 3 – 5 μm wavelength region, also called mid-wave infrared (MWIR), while HgCdTe-based and QWIP-based detectors are available for the 8 – 12 μm , or long-wave infrared (LWIR), wavelength region as well as for MWIR.

2.2.2 Sensitivity

The thermal sensitivity of an infrared camera is often specified in terms of the performance parameter $\text{NE}\Delta\text{T}$ or NEDT, the *noise equivalent temperature differential*. This can be understood as follows: In a setup in which an infrared camera receives radiation from an ideal blackbody, assume that the detector output is expressed as a voltage V . With such a setup, the $\text{NE}\Delta\text{T}$ can be defined as the temperature difference ΔT of the blackbody that causes a difference in detector output ΔV of equal magnitude as the noise level V_{noise} of the detector. In other words, the $\text{NE}\Delta\text{T}$ is the smallest temperature difference that can be resolved by the detector.

It is important to note that the $\text{NE}\Delta\text{T}$ is not a fixed performance parameter, but itself depends on the background temperature T_{bg} . This is because the difference in power radiated by the blackbody is not linearly related to the temperature difference but follows the Stefan-Boltzmann law $\Delta M = \sigma((T_{\text{bg}} + \Delta T)^4 - T_{\text{bg}}^4)$. Moreover, the noise level also depends on the transmittance of the optics and the integration time of the sensor. Therefore, $\text{NE}\Delta\text{T}$ is only meaningful as a performance parameter if the integration time and the background temperature are also specified with it¹. The $\text{NE}\Delta\text{T}$ s for the cameras used in this work are listed in table 6.1 on page 102.

¹Unfortunately, it is common that manufacturers don't state at which integration time or frame rate their claimed $\text{NE}\Delta\text{T}$ was measured. It is often safe to assume that the specified value does not hold for the shortest possible integration time.

2.2.3 Detector Output and Temperature Measurements

In this section, the dependence of the output of an infrared detector on the incident flux is analyzed. It is shown that, if a blackbody is imaged, the detector output increases monotonically with the temperature of the blackbody. Because the cameras used for the present work are photon-detectors, the discussion of the detector output will focus on such detectors. However, for bolometer cameras the detector output can be derived in an analogous way by using the energy-based rather than the photon-based radiometric quantities.

Detector Output The output of an infrared detector depends on the incident irradiance, the attenuation of the optical system and the response of the sensor. All these factors influencing the detector output may exhibit a spectral dependence, so that the output V_{out} is given as an integral over the spectral range $\lambda_{\text{min}} \dots \lambda_{\text{max}}$ in which the detector is sensitive

$$V_{\text{out}} = k \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} \eta(\lambda) E_{\text{p},\lambda}(\lambda, T) d\lambda. \quad (2.29)$$

In this integral, $E_{\text{p},\lambda}(\lambda, T)$ is the spectral incident irradiance, expressed in terms of the number of photons, and $\eta(\lambda)$ is the quantum efficiency of the detector system, that is the quantum efficiency of the sensor multiplied by the attenuation factor of the optical system. The effects of the electronic read-out circuit are represented by a constant factor k .

For a camera imaging a blackbody, the irradiance is given as the exitance of the blackbody, according to Planck's law (2.12). As noted in section 2.1.2, the exitance of a blackbody increases monotonically for every spectral subregion. Assuming that the quantum efficiency of the detector is approximately constant over the spectral range $\lambda_{\text{min}} \dots \lambda_{\text{max}}$, the integral (2.29) also monotonically increases with temperature, resulting in a one-to-one correspondence between the blackbody temperature and the detector output. Therefore, it is possible to sense the temperature of a blackbody remotely with a radiometrically calibrated detector.

Radiometric Calibration Radiometric calibration of a detector is performed by recording its output for several accurately known temperatures of an imaged blackbody source. As the sensitivity of each pixel on an infrared detector chip is different, such a calibration has to be performed on a per-pixel basis. For a given



Figure 2.11: *Effect of emissivity: The left image shows a color photograph of a bottle with a white sticker, aluminum coated tape, and black duct tape. The bottle is filled with warm water. In the infrared image on the right the white sticker and the aluminum tape appear darker because of their lower emissivity compared to the black tape and the bottle glass.*

detector output, the temperature of a blackbody source can then be determined by interpolating between the different calibration points. In the present work, a polynomial is fitted through the calibration points to obtain a functional relationship between temperature and detector output, see sections 5.1 and 7.1.

Non-Blackbody Objects For objects that are not blackbodies, the situation is more difficult. Because their emissivity is smaller than one, such objects do not emit infrared radiation as effectively as blackbodies. Moreover, they also reflect infrared radiation from their environment towards the detector. In such cases the detector output is given by

$$V_{\text{out}} = k \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} \epsilon(\lambda)\eta(\lambda)M_{\text{p},\lambda}^{\text{obj}}(T_{\text{obj}}) + k \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} \rho(\lambda)\eta(\lambda)M_{\text{p},\lambda}^{\text{env}}(T_{\text{env}})d\lambda, \quad (2.30)$$

where ϵ and ρ are the emissivity and the reflectivity of the object respectively. Therefore, the exact emissivity of the object and the temperature of the environment, T_{env} , have to be known in order to perform absolute temperature measurements. The temperature of a blackbody that would produce the same grey-value as the imaged object with emissivity $\epsilon < 1$ is often called the *apparent temperature* of that object.

The effect of the emissivity on the output of an infrared detector is demonstrated in figure 2.11, which displays images of a bottle of warm water acquired with a “normal” camera and an infrared camera. Although the bottle has a homogeneous surface temperature, the irradiance received by the infrared camera, and therefore the apparent temperature, differs for those parts of the bottle that are covered with tape. This can be attributed to the variation in emissivity between the different types of tape.

3 Stereo Computer Vision I: Geometry

This chapter introduces the geometric concepts necessary to perform a 3D reconstruction with a stereo camera setup. Section 3.1 briefly outlines the concept of projective spaces, and introduces *homogeneous* coordinates, since these allow the formulation of the imaging process in a linear framework. The focus of section 3.2 is a detailed analysis of the imaging process with a single camera. This analysis is presented first for a simple *pinhole model* in section 3.2.1 and is then extended to a more realistic model involving a CCD-type camera and lens distortion in sections 3.2.3 and 3.3.

To describe a real camera with such a model, the values of the free model parameters have to be estimated for the camera. This process is called *camera calibration* and is addressed in section 3.4, where a calibration algorithm proposed by Zhang [108] is presented. Camera calibration is essential if metric information about a 3D scene is to be derived from 2D images of that scene.

Section 3.5 reviews the imaging geometry for the stereo case, in which two cameras are used, with a presentation of the epipolar constraint in section 3.5.2. If a point in one image is given, this constraint can be used to restrict the search space for a corresponding image point in the second image to a line.

The topic of sections 3.5.3 and 3.5.4 is rectification; that is, a projective transform of two stereo images that brings corresponding image points onto the same image scan-lines. Rectification is an elegant method to make use of the epipolar constraint during disparity estimation, discussed in chapter 4,

Once both cameras of a stereo setup have been calibrated and correspondences between points in both images have been established, 3D reconstruction can be performed using triangulation, described in section 3.5.5.

In this work, homogeneous vectors are distinguished from non-homogeneous vectors by a tilde. Vectors representing points in 3D space are denoted by upper-case letters whereas vectors representing image points are denoted by lower-case letters. Table 3.1 gives examples and summarizes the notation for geometric entities used in this work.

Type	Example
2D Image	\mathcal{J}, \mathcal{j}
Homogeneous 3-vector (point or line in \mathbb{P}^2)	$\tilde{\mathbf{a}}, \tilde{\mathbf{b}}$
Rectified homogeneous 3-vector	$\bar{\mathbf{a}}, \bar{\mathbf{b}}$
Homogeneous 4-vector (point or plane in \mathbb{P}^3)	$\tilde{\mathbf{A}}, \tilde{\mathbf{B}}$
2-vector (point in \mathbb{R}^2 , e.g. image point)	\mathbf{a}, \mathbf{b}
2-vector (undistorted image point with respect to principal point)	$\hat{\mathbf{a}}, \hat{\mathbf{b}}$
2-vector (distorted image point)	$\check{\mathbf{a}}, \check{\mathbf{b}}$
3-vector (point in \mathbb{R}^3 , e.g. scene point)	\mathbf{A}, \mathbf{B}
Matrices	\mathbf{A}, \mathbf{B}

Table 3.1: Summary of the notation used in chapter 3.

3.1 Projective Geometry

In this section, homogeneous coordinates, a concept used in projective geometry, are introduced. Homogeneous coordinates are important for the present application, as they can be used to express the imaging process with a camera, which is a projection of the three-dimensional scene space onto a two-dimensional image plane, in a linear form.

For an in-depth introduction to projective geometry in the context of computer vision, the reader is referred to the recent monographs by Faugeras and Luong [27] and Hartley and Zisserman [44].

3.1.1 The Projective Plane \mathbb{P}^2

The projective plane \mathbb{P}^2 corresponds to the Euclidean plane \mathbb{R}^2 . However, in contrast to the Euclidean case, where a point in the plane is represented by a 2-vector $(x, y)^T$ in \mathbb{R}^2 , a point in the projective plane is represented by a 3-vector $(\tilde{x}, \tilde{y}, \tilde{w})^T$ in \mathbb{R}^3 without $(0, 0, 0)^T$. This 3-vector is scale invariant, that is, for any factor $\alpha \neq 0$, $(\tilde{x}, \tilde{y}, \tilde{w})^T$ and $(\alpha\tilde{x}, \alpha\tilde{y}, \alpha\tilde{w})^T$ are equivalent and correspond to the same point in \mathbb{P}^2 . This scale-invariant representation is known as a *homogeneous* vector. As mentioned above, in this work, homogeneous vectors and elements of homogeneous vectors are denoted with a tilde.

A Euclidean point $\mathbf{x} = (x, y)^T$ can be represented in homogeneous notation by the projective point $\tilde{\mathbf{x}} = (\alpha x, \alpha y, \alpha)^T$. It is common to choose $\alpha = 1$ and to augment the Euclidean 2-vector by a 1 in the third coordinate when going from a Euclidean to a projective representation of a point. Going back from a projective

representation to a Euclidean representation of a point is achieved by dividing out the scale factor, $\mathbf{x} = (\tilde{x}/\tilde{w}, \tilde{y}/\tilde{w})^T$, for $\tilde{w} \neq 0$. Projective points with $\tilde{w} = 0$ are called *points at infinity* and have no counterpart in the Euclidean plane. The use of homogeneous vectors in \mathbb{P}^2 can be illustrated by considering a line on a plane. A line on the Euclidean plane can be represented by a 3-vector $(a, b, c)^T$. A point $\mathbf{x} = (x, y)^T$ lies on the line if $ax + by + c = 0$. Using homogeneous coordinates, this constraint defining the line can be expressed by an inner product

$$\tilde{\mathbf{l}}^T \tilde{\mathbf{x}} = (a, b, c) (x, y, 1)^T = 0. \quad (3.1)$$

Note that the set of points $\tilde{\mathbf{x}}$ fulfilling (3.1) does not change if the 3-vector (a, b, c) is scaled by an arbitrary, non-zero factor α . Therefore lines in \mathbb{P}^2 can also be represented as homogeneous 3-vectors.

It is easy to show (see for example [44]) that the point of intersection $\tilde{\mathbf{x}}$ of two lines $\tilde{\mathbf{l}}_1, \tilde{\mathbf{l}}_2$ in \mathbb{P}^2 is given by their cross-product $\tilde{\mathbf{x}} = \tilde{\mathbf{l}}_1 \times \tilde{\mathbf{l}}_2$. Similarly, the line $\tilde{\mathbf{l}}$ connecting two points $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2$ is given by their cross-product $\tilde{\mathbf{l}} = \tilde{\mathbf{x}}_1 \times \tilde{\mathbf{x}}_2$. This interchangeability between points and lines is known as *duality*.

3.1.2 The Projective Space \mathbb{P}^3

Analogous to the extension of the Euclidean plane to the projective plane discussed in the previous section, it is possible to extend the concept of the Euclidean 3-space \mathbb{R}^3 to the projective space \mathbb{P}^3 . In \mathbb{P}^3 , points are represented by homogeneous 4-vectors in \mathbb{R}^4 without $(0, 0, 0, 0)^T$. As in the 2-dimensional case, the Euclidean representation of a projective point $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{W})^T$ can be obtained by dividing by $\mathbf{X} = (\tilde{X}/\tilde{W}, \tilde{Y}/\tilde{W}, \tilde{Z}/\tilde{W})^T$, for $\tilde{W} \neq 0$. Again, points with $\tilde{W} = 0$ are called points at infinity and have no counterpart in the Euclidean space. A point $\mathbf{X} = (X, Y, Z)^T$ in Euclidean space can be represented in the projective space by augmenting its vector with a one, that is $\tilde{\mathbf{X}} = (X, Y, Z, 1)^T$.

Planes in \mathbb{P}^3 are also represented by homogeneous 4-vectors, analogous to the role of lines in \mathbb{P}^2 .

In this work, vectors and elements of vectors in the Euclidean space \mathbb{R}^3 and in the projective space \mathbb{P}^3 are denoted with capital letters to distinguish them from vectors in the Euclidean and projective planes \mathbb{R}^2 and \mathbb{P}^2 .

3.1.3 Homographies

In projective geometry, non-singular, linear transformations in projective spaces are called *homographies*. A homography can be described by a non-singular matrix.

For example, the transformation of a homogeneous vector in \mathbb{P}^2 can be expressed as $\tilde{\mathbf{x}}' = \mathbf{H}\tilde{\mathbf{x}}$, where \mathbf{H} is a 3×3 matrix. Because homogeneous vectors are scale invariant, the transformation matrix is also only defined up to an arbitrary scale factor.

3.2 Single View Geometry and Camera Models

3.2.1 Pinhole camera model

Figure 3.1 illustrates a simple pinhole camera model that can be used to describe the central projection of points in 3-dimensional space onto a 2-dimensional *focal plane* or *image plane*. The image \mathbf{m} of a point \mathbf{M} in the scene is determined by the intersection of the image plane and the line connecting \mathbf{M} with the *center of projection* or *camera center* \mathbf{C} .

One can define an orthonormal coordinate system, with its origin located at the camera center and base vectors $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, such that the focal plane is perpendicular to the \mathbf{Z} -direction. This coordinate system is called the *camera coordinate system*. In the focal plane, a 2-dimensional *image coordinate system* is defined by the base vectors \mathbf{u} and \mathbf{v} . The \mathbf{Z} -direction is aligned with the view direction of the camera and is referred to as the *principal axis*. The intersection of the *principal axis* with the focal plane is called the principal point \mathbf{p}_0 with the image coordinates $(u_0, v_0)^T$. The distance of the focal plane from the origin is called the *focal length* f . Denoting image coordinates relative to the principal point with a circumflex ($\hat{\cdot}$), such that $\hat{\mathbf{x}} = \mathbf{x} - \mathbf{p}_0$, the projection of a scene point $\mathbf{M} = (X, Y, Z)^T$ can be calculated as

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} fX/Z \\ fY/Z \end{pmatrix}. \quad (3.2)$$

The image coordinates are then given by

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} fX/Z + u_0 \\ fY/Z + v_0 \end{pmatrix}. \quad (3.3)$$

The resulting image point \mathbf{x} corresponds to the homogeneous point

$$\tilde{\mathbf{x}} = (fX/Z + u_0, fY/Z + v_0, 1)^T \quad (3.4)$$

in the projective plane, which is equivalent to $\tilde{\mathbf{x}} = (fX + Zu_0, fY + Zv_0, Z)^T$ due to the scale invariance of homogeneous vectors. Therefore, the central projection

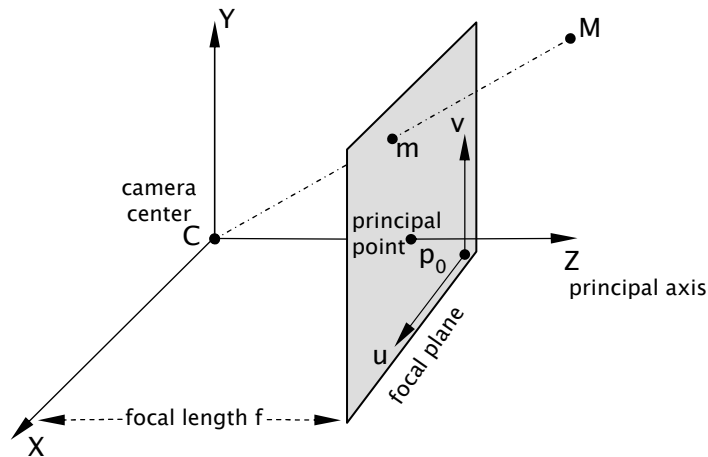


Figure 3.1: Pinhole camera model. The focal plane is located at a distance f from the center of projection C , which is the origin of the camera coordinate system $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$. The axis perpendicular to the focal plane is called the principal axis and intersects it at the principal point \mathbf{p}_0 . A coordinate system in the image plane is spanned by the vectors \mathbf{u}, \mathbf{v} . The image \mathbf{m} of a point \mathbf{M} in the scene is given by the intersection of the line connecting \mathbf{M} and C .

of a point $(X, Y, Z)^T$ onto the image plane can be expressed by a linear mapping from \mathbb{P}^3 to \mathbb{P}^2 as follows

$$\tilde{\mathbf{x}} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{pmatrix} = \begin{bmatrix} f & s & u_0 & 0 \\ & f & v_0 & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (3.5)$$

Equation (3.5) can be rewritten as

$$\tilde{\mathbf{x}} = \mathbf{K} [\mathbf{1} \mid \mathbf{0}] \tilde{\mathbf{X}}_{\text{camera}} \quad (3.6)$$

using the 3×3 camera calibration matrix

$$\mathbf{K} = \begin{bmatrix} f & & u_0 \\ & f & v_0 \\ & & 1 \end{bmatrix}. \quad (3.7)$$

The symbols $\mathbf{1}$ and $\mathbf{0}$ denote the 3×3 identity matrix and the 3×1 null vector, respectively.

3.2.2 World Coordinate System

In many cases, 3D scene points are expressed with respect to a *world coordinate system* that differs from the camera coordinate system described in the previous section.

The world coordinates of a point can be transformed into the camera coordinate system, with the transformation

$$\mathbf{X}_{\text{camera}} = \mathbf{R}(\mathbf{X}_{\text{world}} - \mathbf{C}), \quad (3.8)$$

where \mathbf{C} is the position of the camera center in the world coordinate system and \mathbf{R} is the 3×3 rotation matrix that aligns the axes of the world and the camera coordinate systems. The parameters that describe the six degrees of freedom for the translation and the rotation are called the *exterior camera parameters*. In homogeneous notation this coordinate transformation can be expressed as

$$\tilde{\mathbf{X}}_{\text{camera}} = \left[\begin{array}{c|c} \mathbf{R} & -\mathbf{R}\mathbf{C} \\ \hline \mathbf{0}^T & 1 \end{array} \right] \tilde{\mathbf{X}}_{\text{world}}. \quad (3.9)$$

Introducing the translation vector

$$\mathbf{T} = -\mathbf{R}\mathbf{C}, \quad (3.10)$$

the projection of points expressed as homogeneous vectors in the world coordinate system can be written as

$$\tilde{\mathbf{x}} = \mathbf{K} [\mathbf{R} \mid \mathbf{T}] \tilde{\mathbf{X}}_{\text{world}}. \quad (3.11)$$

With the 3×4 *camera projection matrix* \mathbf{P} , defined as

$$\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{T}], \quad (3.12)$$

equation (3.11) can be written in the concise form

$$\tilde{\mathbf{x}} = \mathbf{P} \tilde{\mathbf{X}}_{\text{world}}. \quad (3.13)$$

Thus, the imaging process with a projective camera can be expressed as a linear transformation from \mathbb{P}^3 to \mathbb{P}^2 .

3.2.3 CCD-type cameras

When using CCD-type cameras, the basic pinhole camera model has to be slightly modified. The term CCD-type is used here to refer to cameras with a detector that has a number of individual sensor elements arranged in a rectangular array. Factors such as the type of integrated circuit technology used and the electronic read-out mechanism of the detector, are not important for this discussion¹, and so the results of this section also hold for the focal plane array detectors commonly used in infrared cameras and CMOS cameras.

For a CCD-type detector chip, the image coordinates are usually expressed in terms of pixel coordinates. This can be accounted for by multiplying the parameter focal length f in the camera matrix K by an appropriate scale factor. This scale factor is determined by the center-to-center spacing of the pixels on the detector chip, the so-called *pixel pitch*. As the horizontal pixel pitch p_x may differ from the vertical pixel pitch p_y for some detectors, the two directions have to be considered separately. With $\alpha_x = f/p_x$ and $\alpha_y = f/p_y$ the camera calibration matrix for a CCD-type camera can be rewritten as

$$K = \begin{bmatrix} \alpha_x & s & u_0 \\ & \alpha_y & v_0 \\ & & 1 \end{bmatrix}. \quad (3.14)$$

The parameter s is called *skew* and accounts for cases where the rows and columns of pixels on the detector chip are not perpendicular to each other. With modern cameras this is usually not a concern, so that one can assume that $s = 0$.

The free parameters of K are commonly referred to as *interior camera parameters*.

3.3 Non-Linear Distortion

The pinhole model for CCD-type cameras is useful, as it allows a linear, and therefore simple, formulation of the imaging process. However, it is only an approximation, since real cameras and lenses exhibit distortion, which cannot be accounted for with a linear model. The effect of such non-linear distortion is demonstrated in figure 3.2 where the straight lines on a checkerboard appear curved in the image. Section 3.3.1 presents a non-linear model to describe lens distortion and

¹Strictly speaking, the term CCD, short for *charge-coupled device*, only refers to the way in which the image information is stored and read out from the detector chip. Some imaging detectors use a different read-out mechanism and are therefore not CCD detectors in the strict sense of the original definition.

section 3.3.2 provides an algorithm that makes it possible to “un-distort” images obtained with a real camera. The relation between world points and image coordinates in these undistorted images can then be used with the linear model, equation (3.11).

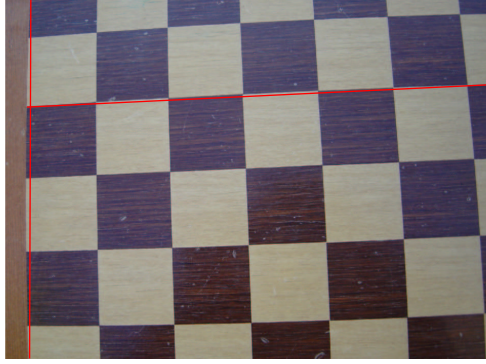


Figure 3.2: Due to non-linear distortion introduced by the lens, straight lines in the scene appear curved in this image taken with a consumer digital camera. For reference, straight lines are shown in red.

3.3.1 Modeling Non-Linear Distortion

To account for image distortion caused by real lenses, a non-linear distortion term can be added to the undistorted coordinates calculated using the pin-hole model. In this section, appropriate non-linear correction terms for modeling *radial* and *tangential distortion* are presented.

Most real lenses exhibit radial distortion, which causes a radial displacement between the image coordinate \mathbf{x} , onto which a scene point \mathbf{X} would be projected obtained by an ideal linear camera, and the distorted coordinate $\check{\mathbf{x}}$ of the projection actually observed. The center of this radial displacement is usually assumed to coincide with the principal point. Using undistorted image coordinates, expressed relative to the principal point, $\hat{u} = u - u_0, \hat{v} = v - v_0$, the radial displacement $(\delta u_r, \delta v_r)^T$ can be approximated by a series expansion

$$\begin{pmatrix} \delta u_r \\ \delta v_r \end{pmatrix} = \begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} (\kappa_1 r^2 + \kappa_2 r^4 + \dots) \quad (3.15)$$

where

$$r = \sqrt{\hat{u}^2 + \hat{v}^2} \quad (3.16)$$

is the distance from the center of distortion.

In the manufacturing process, it may happen that the elements of a lens are not perfectly centered on the optical axis. This causes a distortion that has both a radial and a tangential component. The radial component can be accounted for by (3.15) and the displacement due to tangential distortion can be modeled as:

$$\begin{pmatrix} \delta u_t \\ \delta v_t \end{pmatrix} = \begin{pmatrix} 2\tau_1 \hat{u}\hat{v} + \tau_2 (r^2 + 2\hat{u}^2) \\ \tau_1 (r^2 + 2\hat{v}^2) + 2\tau_2 \hat{u}\hat{v} \end{pmatrix}. \quad (3.17)$$

Finally, the distorted image coordinate $(\check{u}, \check{v})^T$ is obtained by adding the radial and the tangential displacement vectors to the idealized, undistorted coordinate.

$$\begin{pmatrix} \check{u} \\ \check{v} \end{pmatrix} = \begin{pmatrix} u + \delta u_r + \delta u_t \\ v + \delta v_r + \delta v_t \end{pmatrix} \quad (3.18)$$

This particular choice of parameters for modeling non-linear distortion is also used in [10, 47] and goes back to Brown [11, 12].

The distortion parameters are considered to be additional interior camera parameters, complementing the set of interior parameters of the camera calibration matrix \mathbf{K} , described in section 3.2.3.

3.3.2 Correcting for Non-Linear Distortion

In section 3.3.1, it was shown how to calculate the mapping between a 3D point and its real (distorted) image coordinates. In practice, the reverse problem usually arises: one acquires images with a lens system that introduces distortion, and would like to derive metric information from the images using the linear model of equation (3.13). Therefore, a mapping from the distorted coordinates $(\check{u}, \check{v})^T$ to the idealized undistorted coordinates $(u, v)^T$ is required before applying the pin-hole model. If both tangential and radial distortion are included in the model, there is generally no analytical solution to this inverse mapping problem. However, by employing a look-up table approach, a corrected image can be calculated using algorithm 3.1.

3.4 Estimation of Camera Parameters using Zhang's Method

In the previous section, it was shown how a projective CCD camera model can be parameterized by a set of intrinsic and extrinsic parameters. In this section,

Given: A distorted image J_{dist} .

1. Allocate image space J_{corr} to hold the corrected image (usually the same size as the original image).
2. Loop over all the image coordinates (u,v) in J_{corr}
 - (a) calculate distorted coordinate (\check{u}, \check{v})
 - (b) Fill in image pixel (u,v) in J_{corr} with intensity (color or grey-value) of pixel (\check{u}, \check{v}) from J_{dist} (in general this will involve interpolation)

Algorithm 3.1: *Algorithm for removing non-linear image distortion.*

it is shown how these parameters can be estimated for a given real camera. The process of estimating the camera parameters is usually called *camera calibration*. A variety of methods for camera calibration can be found in the literature; see [14] for an overview.

The common idea underlying these methods is to derive constraints on the parameters using known correspondences between features in the 3D scene and features in the image. This can be done either by the use of existing features in the 3D scene, such as straight lines and right angles, which are often found as architectural features, or through the use of known feature points on a dedicated calibration target which is put in the scene.

For the present application, a calibration method is used in which a planar calibration target was imaged from several unknown orientations. This method was proposed by Zhang [108]. The calibration target used is a plane with a checkerboard pattern. The corners of the checkers are high-contrast feature points that can be easily and precisely located in the images, for example with a Harris corner detector [41]. This calibration method was chosen for its ease of use, keeping in mind that the calibration might eventually have to be performed on a ship, where other methods (such as moving a target on a linear positioner) might be less feasible. For a description of the calibration target used in the present application and the results of the calibration see sections 6.4 and 7.2 respectively. The implementation of the algorithm used for this work is due to Bouguet [10].

The presentation in this section outlines the main steps of the calibration algorithm described in Zhang's paper.

3.4.1 The Calibration Process

For the calibration, a planar calibration target with easily identifiable point features (corners, edges) \mathbf{Q}_i at known locations is needed. During the calibration process, images of this target are collected from several different orientations, either by moving the camera and leaving the target fixed or vice versa. This is illustrated in figure 3.3. The points \mathbf{q}_i corresponding to \mathbf{Q}_i are then extracted from the images and are used for the calibration.

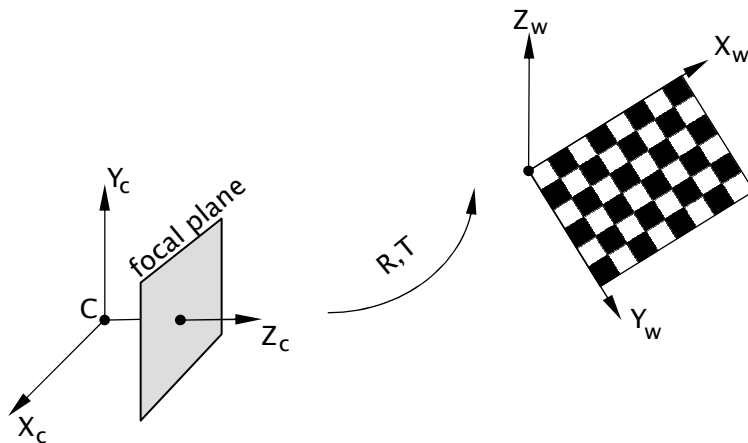


Figure 3.3: Calibration process. A planar calibration target, for example a checkerboard pattern, defines the world coordinate system. The position of the checker corners \mathbf{Q}_i with respect to this world coordinate system are known. For the calibration, images of the calibration target are collected from several orientations, either by keeping the camera fixed and moving the calibration target or vice versa. For each orientation j of the calibration target, a set of exterior parameters, describing the rotation \mathbf{R}_j and translation \mathbf{T}_j between world and camera coordinate system, has to be estimated in addition to the interior parameters of the camera.

The main calibration step, described in section 3.4.3, is an optimization that finds the set of interior camera parameters and exterior parameters with respect to the calibration plane that minimizes the re-projection error. The re-projection error is the distance between the observed image points \mathbf{q}_i and the image points $\check{\mathbf{q}}_i$ calculated by projecting the known \mathbf{Q}_i with the estimated camera parameters.

To perform this optimization, it is necessary to have an initial guess for the camera parameters. Without this initial guess, it is likely that the optimization will converge to some local minimum of the re-projection error. An algebraic solution for finding this initial guess is presented in the following section 3.4.2.

3.4.2 Initial Guess through Closed-Form Solution

For the initial guess, only a linear pinhole model without radial distortion is used. The following description of how to calculate the initial solution closely follows the steps outlined in Zhang's paper [108].

Notation and World Coordinate System Throughout this section, the subscript j is used to denote different orientations of the calibration target and the subscript i is used to denote feature points on the calibration target. The world coordinate system is chosen such, that the planar calibration target lies in the $Z = 0$ plane. The i^{th} feature point on the j^{th} orientation of the calibration target is denoted by the vector $\mathbf{Q}_{ji} = (X_{ji}, Y_{ji}, Z_{ji})^T$ and its image is denoted by $\mathbf{q}_{ji} = (u_{ji}, v_{ji})^T$. The Z_{ji} component of the feature point vector is, by definition of the world coordinate system, always zero, reflecting the fact that a planar calibration target is intrinsically two-dimensional.

Homography between the Model Plane and its Image The feature points $\tilde{\mathbf{Q}}_{ji}$ lie on the model plane (that is, the calibration target) and their respective images $\tilde{\mathbf{q}}_{ji}$ also lie on a plane, namely the focal plane of the camera. Therefore, they are related by a planar homography, which is a projective transform in the projective plane \mathbb{P}^2 . This homography can be represented by a 3×3 matrix \mathbf{H}_j as

$$\tilde{\mathbf{q}}_{ji} = \mu \begin{pmatrix} u_{ji} \\ v_{ji} \\ 1 \end{pmatrix} = \mathbf{H}_j \begin{pmatrix} X_{ji} \\ Y_{ji} \\ 1 \end{pmatrix}, \quad (3.19)$$

for some scale factor μ . If at least four correspondences between the feature points $\tilde{\mathbf{Q}}_{ji}$ and their images $\tilde{\mathbf{q}}_{ji}$ are established, the homography \mathbf{H}_j can be calculated. The details of how this is done are given in appendix B.

Constraints on the Camera Parameters For each relative orientation j between the calibration target and the camera, the homography \mathbf{H}_j gives two constraints on the interior camera parameters. If at least three different homographies are obtained, the entire camera matrix can be determined. Once the interior camera parameters are determined, the exterior parameters for each orientation of the model plane can be extracted from \mathbf{H}_j .

To derive a relation between the homographies \mathbf{H}_j and the camera parameters, consider the projection of the feature points $\tilde{\mathbf{Q}}_{ji}$ into the image (see equation (3.11))

$$\tilde{\mathbf{q}}_{ji} = \mathbf{K} [\mathbf{R}_j \mid \mathbf{T}_j] \tilde{\mathbf{Q}}_{ji}, \quad (3.20)$$

with the camera matrix \mathbf{K} and the exterior parameters $\mathbf{R}_j, \mathbf{T}_j$. Representing the rotation matrix \mathbf{R}_j by its column vectors this becomes

$$\lambda \begin{pmatrix} u_{ji} \\ v_{ji} \\ 1 \end{pmatrix} = \mathbf{K} [\mathbf{R}_{j1} \quad \mathbf{R}_{j2} \quad \mathbf{R}_{j3} \quad \mathbf{T}_j] \begin{pmatrix} X_{ji} \\ Y_{ji} \\ Z_{ji} = 0 \\ 1 \end{pmatrix}. \quad (3.21)$$

Using the fact that by definition of the world coordinate system $Z_{ji} = 0$ for all the feature points (the planar calibration target lies in the $Z=0$ plane), the notation can be shortened by leaving out the Z coordinate to

$$\lambda \begin{pmatrix} u_{ji} \\ v_{ji} \\ 1 \end{pmatrix} = \mathbf{K} [\mathbf{R}_{j1} \quad \mathbf{R}_{j2} \quad \mathbf{T}_j] \begin{pmatrix} X_{ji} \\ Y_{ji} \\ 1 \end{pmatrix}, \quad (3.22)$$

representing the $\tilde{\mathbf{Q}}_{ji}$ by 3-vectors $(X_{ji}, Y_{ji}, 1)^T$. A comparison of the relation between the feature points and their images expressed by the planar homography, equation (3.19), and the same relation derived using the camera projection matrix, equation (3.22), shows that

$$\mathbf{H}_j = [\mathbf{H}_{j1} \quad \mathbf{H}_{j2} \quad \mathbf{H}_{j3}] = \lambda \mathbf{K} [\mathbf{R}_{j1} \quad \mathbf{R}_{j2} \quad \mathbf{T}_j], \quad (3.23)$$

where the \mathbf{H}_{jk} , $k = 1 \dots 3$, represent the column vectors of \mathbf{H}_j . Again, the constant λ denotes some scale factor, though not necessarily the same as the λ as in (3.22).

Using $\mathbf{K}^{-1} \mathbf{H}_{1,2} = \lambda \mathbf{R}_{1,2}$ and the orthonormality of the rotation matrix \mathbf{R} , that is $\mathbf{R}_k^T \mathbf{R}_l = \delta_{kl}$, equation (3.23) leads to the following two constraints on the camera matrix \mathbf{K}

$$\mathbf{H}_{j1}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{H}_{j2} = 0 \quad (3.24)$$

$$\mathbf{H}_{j1}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{H}_{j1} = \mathbf{H}_{j2}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{H}_{j2}. \quad (3.25)$$

The camera matrix \mathbf{K} has five degrees of freedom. As each homography only gives two constraints, at least three homographies are needed to calculate the intrinsic parameters.

Closed-form solution In this subsection, it is shown how the above constraints can be used to obtain a solution for \mathbf{K} given at least three homographies.

Equations (3.24) and (3.25) involve the matrix product $\mathbf{K}^{-T}\mathbf{K}^{-1}$. For the calculations below, it is useful to express this product as

$$\mathbf{B} = \mathbf{K}^{-T}\mathbf{K}^{-1} \equiv \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix}. \quad (3.26)$$

\mathbf{B} is symmetric and can therefore be represented by a 6-vector

$$\mathbf{B} = (B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33})^T. \quad (3.27)$$

If the k^{th} column vector for a given homography \mathbf{H}_j is defined to be $\mathbf{H}_k = (H_{k1}, H_{k2}, H_{k3})^T$, one can define a vector \mathbf{V}_{kl} as follows:

$$\mathbf{H}_k^T \mathbf{B} \mathbf{H}_l = \mathbf{V}_{kl}^T \mathbf{B} \quad (3.28)$$

with

$$\mathbf{V}_{kl} = \begin{pmatrix} H_{k1}H_{l1} \\ H_{k1}H_{l2} - H_{k2}H_{l1} \\ H_{k2}H_{l2} \\ H_{k3}H_{l1} + H_{k1}H_{l3} \\ H_{k3}H_{l2} + H_{k2}H_{l3} \\ H_{k3}H_{l3} \end{pmatrix}. \quad (3.29)$$

Using this definition, the constraints (3.24) and (3.25) can be recast in the form

$$\begin{bmatrix} \mathbf{V}_{12}^T \\ (\mathbf{V}_{11} - \mathbf{V}_{12})^T \end{bmatrix} \mathbf{B} = \mathbf{0} \quad (3.30)$$

Given a single homography, this system is under-determined, but if n homographies corresponding to different orientations of the model plane are available, there are n equations of type (3.30). These n equations can be stacked, yielding the equation system

$$\mathbf{V}\mathbf{B} = \mathbf{0} \quad (3.31)$$

with the $2n \times 6$ matrix \mathbf{V} . Leaving degenerate configurations aside, this system can be solved for \mathbf{B} if at least three homographies are given. A least-squares solution of (3.31) can be obtained by the singular value decomposition of \mathbf{V} , with \mathbf{B} given as the vector corresponding to the smallest singular value.

Once \mathbf{B} is estimated, the intrinsic camera parameters can be calculated using equations (3.14) and (3.26), yielding

$$\begin{aligned}
v_0 &= (B_{12}B_{13} - B_{11}B_{23}) / (B_{11}B_{22} - B_{12}^2) \\
\lambda &= B_{33} - (B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})) / B_{11} \\
\alpha_x &= \sqrt{\lambda / B_{11}} \\
\alpha_y &= \sqrt{\lambda B_{11} / (B_{11}B_{22} - B_{12}^2)} \\
s &= -B_{12}f_x^2 f_y / \lambda \\
u_0 &= sv_0 / f_x - B_{13}f_x^2 / \lambda.
\end{aligned} \tag{3.32}$$

If \mathbf{K} has been calculated, the exterior parameters follow from equations (3.14) and (3.23) as

$$\mathbf{R}_{j1} = \lambda \mathbf{K}^{-1} \mathbf{H}_{j1} \quad \mathbf{R}_{j2} = \lambda \mathbf{K}^{-1} \mathbf{H}_{j2} \quad \mathbf{R}_{j3} = \mathbf{R}_{j1} \times \mathbf{R}_{j2} \quad \mathbf{T} = \lambda \mathbf{K}^{-1} \mathbf{H}_{j3}. \tag{3.33}$$

3.4.3 Full Solution through Minimization of Geometric Error

The algebraic error minimized by solving equation (3.31) in a least-squares sense has no meaningful physical interpretation. This is because the matrix \mathbf{B} contains parameters with different physical units, for example skew and focal length. Also, the solution obtained is only a linear solution that does not include the parameters of the distortion model described in section 3.3.

The final calibration step is therefore a non-linear optimization, which varies the set of all parameters (exterior, intrinsic, distortion) to find the parameters that minimize a geometrically meaningful cost function. This cost function is

$$\epsilon(\mathbf{K}, \kappa, \tau, \mathbf{R}_j, \mathbf{T}_j, \mathbf{Q}_i) = \sum_{\substack{j=1 \\ \text{plane} \\ \text{orientations}}}^n \sum_{\substack{i=1 \\ \text{points}}}^m \|\mathbf{q}_{ij} - \check{\mathbf{q}}(\mathbf{K}, \kappa, \tau, \mathbf{R}_j, \mathbf{T}_j, \mathbf{Q}_i)\|^2, \tag{3.34}$$

that is, the sum of squared differences between the image coordinates of the extracted grid corners \mathbf{q}_{ji} and the coordinates calculated by re-projecting the known grid-corner positions \mathbf{Q}_{ji} using the set of exterior, interior and distortion parameters. The sums run over all orientations j of the planar calibration target and over all the grid points i .

The term $\|\mathbf{q}_{ji} - \check{\mathbf{q}}(\mathbf{K}, \kappa, \tau, \mathbf{R}_j, \mathbf{T}_j, \mathbf{Q}_i)\|$ is called the *re-projection error* for each pixel. It has a meaningful interpretation as the distance between an extracted point

coordinate and the respective coordinate obtained by re-projection, and therefore ϵ is a meaningful geometric error function that must be minimized.

In the present application, an implementation by Bouguet [10] was used, in which the optimization is performed with a gradient-descent algorithm (see for example [75]). As the optimization algorithm may become trapped in some irrelevant local minimum it is important to have an initial estimate of the parameters that is close to the correct solution. To this end, the results obtained using the algebraic method described in section 3.4.2 were used as an initial guess for the internal and exterior camera parameters, and the initial values of the distortion parameters were set to zero.

3.5 Two View Geometry

In this section the geometry of a stereo system consisting of two cameras is reviewed.

First, section 3.5.1 shows how the camera calibration method described in section 3.4 can be used to calibrate a stereo camera system and to determine the relative orientations of two cameras in addition to their intrinsic parameters.

Section 3.5.2 discusses the *epipolar constraint*, which is one the most important geometrical constraint in two-view geometry. Given a point in one image, the epipolar constraint limits the possible locations for the corresponding point in the second image to a line. Using homogeneous coordinates, the epipolar constraint can be expressed elegantly in terms of the *fundamental matrix*.

The epipolar constraint can be exploited to reduce the search space of candidate image points in stereo matching (see chapter 4). For this, it is often convenient to apply projective transformations to the original stereo image pair, which aligns the epipolar lines with the image scan-lines. This so-called *rectification* process is described in sections 3.5.3 and 3.5.4.

3.5.1 Calibration of a Stereo Camera System

An important parameter describing a stereo camera system is the relative orientation of the two cameras with respect to each other. It is required for the calculation of the 3D position of a point by triangulation (see section 3.5.5), and, if known, allows a simple determination of the fundamental matrix (see section 3.5.2).

The relative orientation of two cameras with respect to each other can easily be determined if the two cameras are geometrically calibrated with respect to the

same world coordinate system. Therefore, the calibration process described in section 3.4.1 should be performed for both cameras together, using the same feature points and the same orientations of the reference plane.

The calibration procedure as described in section 3.4.1 uses a world coordinate system with the origin at one of the corners of the calibration target. For a stereo camera system, where usually only the relative position of the cameras is of interest, it is more appropriate to let the world coordinate system coincide with the camera coordinate system of one of the cameras. The position and orientation of the second camera is then expressed with respect to this camera system, as discussed below.

Assume that the rotation matrices R_1, R_2 and the translation vectors T_1, T_2 describe the exterior orientation of the two cameras with respect to the same world coordinate system, defined by the calibration pattern. This situation is depicted in figure 3.4.

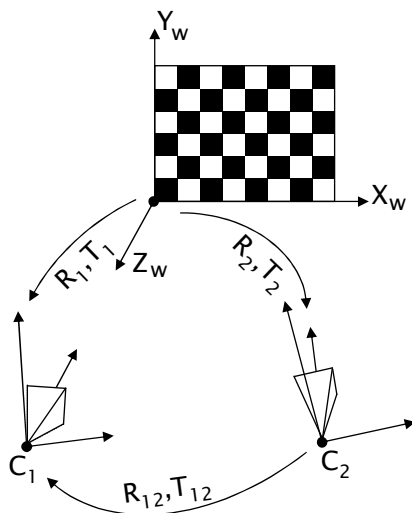


Figure 3.4: Calibration of a stereo camera system. The projective centers of two cameras comprising a stereo system are denoted by C_1 and C_2 . The world coordinate frame, denoted by the subscript w , is defined by the calibration target. The exterior orientation of the two cameras with respect to this world frame is represented by the rotation matrices R_1, R_2 and the translation vectors T_1, T_2 . The position of the second camera with respect to the camera coordinate system of the first camera can be calculated using equations (3.35) and (3.36) and is represented by R_{12} and T_{12} .

The translation vectors T_1, T_2 are related to the positions of the two camera centers C_1, C_2 by $T_1 = -R_1 C_1$ and $T_2 = -R_2 C_2$, respectively (see equation

(3.10)). The position of the second camera center with respect to the first camera center can be expressed by the vector sum $\mathbf{C}'_2 = -\mathbf{C}_1 + \mathbf{C}_2$. The corresponding translation vector is therefore

$$\begin{aligned}\mathbf{T}_{12} &= -\mathbf{R}_1(\mathbf{R}_1^{-1}\mathbf{T}_1 - \mathbf{R}_2^{-1}\mathbf{T}_2) \\ &= -\mathbf{T}_1 + \mathbf{R}_1\mathbf{R}_2^{-1}\mathbf{T}_2\end{aligned}\quad (3.35)$$

The rotation of the camera coordinate system of the second camera with respect to the camera coordinate system of the first camera is obtained by two successive rotations:

- Rotating it into alignment with the world coordinate system (defined by the calibration target) using \mathbf{R}_2^{-1} .
- Rotating it from the world coordinate system into alignment with the camera coordinate system of the first camera using \mathbf{R}_1 .

The combined rotation matrix is therefore

$$\mathbf{R}_{12} = \mathbf{R}_1\mathbf{R}_2^{-1}. \quad (3.36)$$

As only the relative orientations of the two cameras and the calibration target are involved, it makes no difference whether the stereo camera system is moved and the calibration target remains fixed or whether the stereo camera system remains fixed and the calibration target is moved.

For each orientation j of the reference plane, a complete set of exterior parameters $\mathbf{R}_{j1}, \mathbf{T}_{j1}, \mathbf{R}_{j2}, \mathbf{T}_{j2}$ is obtained for both cameras. Each of these sets can be used to determine the relative orientation $\mathbf{T}_{12}, \mathbf{R}_{12}$ between the two cameras. Because of uncertainties, there will be some variation in the estimated relative orientations. The implementation of the stereo calibration, due to Bouguet [10], used in the present work, therefore uses the component-wise median of all the translation vectors. As the entries of a rotation matrix are not all independent, one cannot simply take the median of each entry. Therefore, each rotation is expressed in terms of a rotation vector, whose direction specifies the axis of rotation and whose magnitude specifies the angle of rotation. Then the median of each component of the rotation vector is taken, to obtain a “median rotation”.

As a final step, the orientation parameters are further refined using a non-linear optimization, as in section 3.4.3, that minimizes the re-projection error of the points in both images by variation of the relative orientation and the interior parameters of both cameras.

3.5.2 Epipolar Constraint and the Fundamental Matrix

Epipolar Constraint The diagram in figure 3.5 shows the geometry of two cameras imaging the same scene point P .

Assuming that the image of P in one of the cameras is given, for example p_1 , then the line connecting the camera center C_1 and P is known. The image point p_2 , which is the projection of P into the other camera, must lie on the projection of the line connecting C_1 and P . This projection is a line in the image plane (named l_2 in figure 3.5) and is called the *epipolar line* corresponding to p_1 .

The configuration is symmetric, there is also an epipolar line l_1 in the first camera corresponding to the image p_2 of P in the second camera. Together with the two camera centers C_1 and C_2 , every scene point P defines a plane, the so called *epipolar plane*. The intersections of this plane with the two focal planes are just the epipolar lines corresponding to P . All epipolar planes contain the line that connects the two camera centers C_1 and C_2 . This line intersects the focal planes at the so-called epipoles e_1 and e_2 . In each image all epipolar lines intersect each other at the epipole.

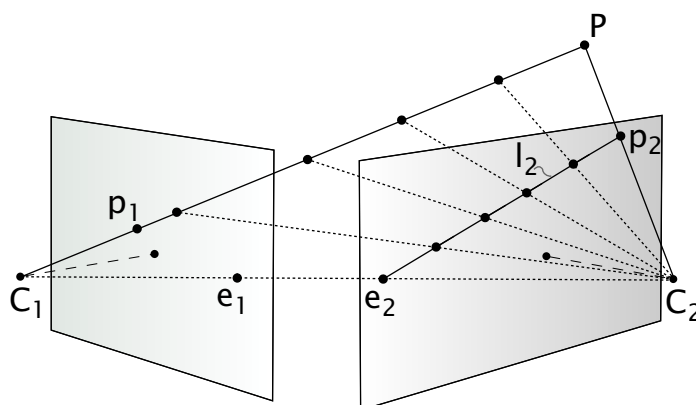


Figure 3.5: Two-view geometry. The scene point P is projected to the image points p_1, p_2 through the projective centers C_1 and C_2 respectively. The camera center C_1 and the image point p_1 define a line of sight. The projection of this line onto the image plane of the second camera forms the epipolar line l_2 corresponding to the point p_1 . The image point p_2 corresponding to the same scene point P as the image point p_1 in the other view must lie on the epipolar line l_2 . For illustration, the projections of some points lying on the line connecting C_1 to P onto the epipolar line are shown. The plane spanned by the point P and the two camera centers is called the epipolar plane. The projection of the two camera centers into the respective other image defines the epipoles e_1, e_2 .

The Fundamental Matrix and its Properties The epipolar constraint finds its algebraic expression in the so-called *fundamental matrix* \mathbf{F} , a 3×3 matrix of rank 2. Without derivation this section lists some properties of the fundamental matrix (for a derivation see [27, 44]).

Using the homogeneous notation for points and lines in the projective plane \mathbb{P}^2 , introduced in section 3.1.1, the fundamental matrix establishes a relationship between a point $\tilde{\mathbf{x}}_1$ in image \mathcal{J}_1 and its corresponding epipolar line $\tilde{\mathbf{l}}_2$ in image \mathcal{J}_2

$$\tilde{\mathbf{l}}_2 = \mathbf{F}\tilde{\mathbf{x}}_1. \quad (3.37)$$

For a point $\tilde{\mathbf{x}}_2$ in image \mathcal{J}_2 , the corresponding epipolar line $\tilde{\mathbf{l}}_1$ in image \mathcal{J}_1 can be calculated using the transpose of the fundamental matrix

$$\tilde{\mathbf{l}}_1 = \mathbf{F}^T \tilde{\mathbf{x}}_2. \quad (3.38)$$

If the two points $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2$ are the images of the same scene point \mathbf{X} $\tilde{\mathbf{x}}_2$ must lie on the epipolar line $\tilde{\mathbf{l}}_2$ corresponding to $\tilde{\mathbf{x}}_1$. It follows that

$$\tilde{\mathbf{x}}_2^T \tilde{\mathbf{l}}_2 = \tilde{\mathbf{x}}_2^T \mathbf{F} \tilde{\mathbf{x}}_1 = 0, \quad (3.39)$$

because when using homogeneous coordinates, the inner product $\tilde{\mathbf{p}}^T \tilde{\mathbf{l}}$ is zero for every point $\tilde{\mathbf{p}}$ on the line $\tilde{\mathbf{l}}$.

Because all epipolar lines in each image intersect at the epipole of their respective image, the epipoles are the null-space of \mathbf{F} and its transpose,

$$\begin{aligned} \mathbf{F}\tilde{\mathbf{e}}_1 &= \mathbf{0} \\ \mathbf{F}^T \tilde{\mathbf{e}}_2 &= \mathbf{0} \end{aligned} \quad (3.40)$$

where $\mathbf{0}$ is the null-vector $(0, 0, 0)^T$.

Calculating the Fundamental Matrix The fundamental matrix can be computed from point correspondences across the two images, for example using the algorithm outlined by Hartley [42].

If intrinsic and extrinsic camera parameters of the stereo system are already known, for example as the result of a camera calibration process as outlined in 3.4, the fundamental matrix can be derived from these parameters. Assuming that the camera coordinate system of the first camera coincides with the world coordinate system, and the relative position and orientation of the second camera is expressed by a translation vector \mathbf{T} and a rotation matrix \mathbf{R} , that is $\mathbf{P}_1 = K_1 [1 \mid \mathbf{0}]$, $\mathbf{P}_2 = K_2 [\mathbf{R} \mid \mathbf{T}]$, then the fundamental matrix is given as

$$\mathbf{F} = K_2^{-T} [\mathbf{T}]_{\times} \mathbf{R} K_1^{-1}. \quad (3.41)$$

The notation $[\mathbf{T}]_{\times}$ stands for the matrix for which $[\mathbf{T}]_{\times} \mathbf{X} = \mathbf{T} \times \mathbf{X}$ for any \mathbf{X} , that is

$$[\mathbf{T}]_{\times} = \begin{bmatrix} 0 & -T_3 & T_2 \\ T_3 & 0 & T_1 \\ -T_2 & -T_1 & 0 \end{bmatrix}. \quad (3.42)$$

3.5.3 Image Rectification

For a verging stereo camera setup like the one depicted in figure 3.5, the epipoles lie either in the image or at a finite distance from the image. As the epipoles are the points where all epipolar lines intersect, this results in non-parallel epipolar lines as depicted in figure 3.6.

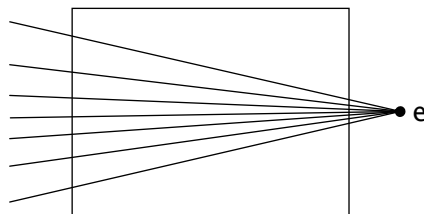


Figure 3.6: Epipolar lines for a configuration where the epipole e lies within a finite distance of the image.

For the implementation of the stereo matching process, described in chapter 4, it is more convenient to have parallel epipolar lines that coincide with the image scan-lines. This can be achieved if the images of both cameras lie in the same focal plane. For such a setup, the line connecting the two camera centers must be parallel to the common focal plane. Therefore the epipoles, which are the projections of the two camera centers into the respective other image, lie at infinity, resulting in parallel epipolar lines. However, it is often not practical to use such a setup, because the overlap of the fields-of-view is smaller than for a verging stereo camera setup (see also section 6.1.2).

Instead of using a stereo setup with coplanar focal planes, one can use a verging stereo system, and apply projective transformations to map the image points obtained with this setup onto two *rectified* images. These rectified images share a virtual image plane that is parallel to the baseline connecting the camera centers, as shown in figure 3.7. This process of re-projecting the image is known as *rectification*.

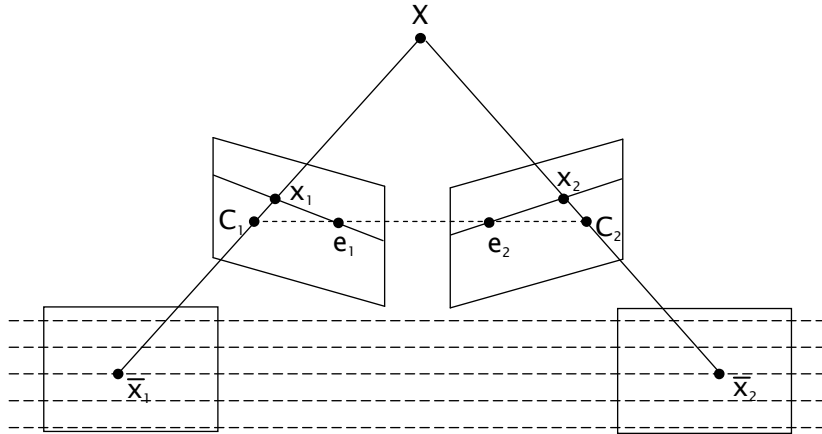


Figure 3.7: Rectification of stereo image pairs. The image points \mathbf{x}_1 and \mathbf{x}_2 of \mathbf{X} are re-projected onto a common image plane, which is parallel to the baseline connecting the two camera centers \mathbf{C}_1 and \mathbf{C}_2 . The rectified image points are denoted $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$. In contrast to the original epipoles $\mathbf{e}_1, \mathbf{e}_2$, the epipoles corresponding to the new image planes lie at infinity, resulting in parallel epipolar lines (dashed).

Using homogeneous coordinates, the rectification can be performed by applying homographies \mathbf{H}_1 and \mathbf{H}_2 to all points in the original images \mathcal{J}_1 and \mathcal{J}_2 respectively:

$$\bar{\mathbf{x}}_1 = \mathbf{H}_1 \tilde{\mathbf{x}}_1 \text{ and } \bar{\mathbf{x}}_2 = \mathbf{H}_2 \tilde{\mathbf{x}}_2, \quad (3.43)$$

where $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ are the image coordinates in the rectified images.

This results in the following transformation of the epipolar constraint (3.39) for two points $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2$ in images $\mathcal{J}_1, \mathcal{J}_2$ corresponding to the same scene point $\tilde{\mathbf{X}}$:

$$\begin{aligned} \tilde{\mathbf{x}}_2^T \mathbf{F} \tilde{\mathbf{x}}_1 &= 0 \\ \bar{\mathbf{x}}_2^T \underbrace{\mathbf{H}_2^{-T} \mathbf{F} \mathbf{H}_1^{-1}}_{\bar{\mathbf{F}}} \bar{\mathbf{x}}_1 &= 0, \end{aligned} \quad (3.44)$$

with $\bar{\mathbf{F}}$ being the fundamental matrix for the rectified image pair.

For image rectification, the homographies $\mathbf{H}_1, \mathbf{H}_2$ have to be chosen in a way that the epipolar lines corresponding to $\bar{\mathbf{F}}$ are parallel and aligned with the image scan-lines. It is easy to verify that the fundamental matrix

$$\bar{\mathbf{F}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3.45)$$

fulfills these requirements. Its epipoles lie at $\bar{\mathbf{e}} = (1, 0, 0)^T$, which is a point at infinity, therefore the epipolar lines are parallel. Corresponding points, for which $\bar{\mathbf{x}}_2^T \bar{\mathbf{F}} \bar{\mathbf{x}}_1 = 0$, also have the same y-coordinate that is they lie on the same scan-line.

3.5.4 Projective Distortion-Minimizing Rectification

Several algorithms for calculating the rectifying homographies given the fundamental matrix or the projection matrices of the two cameras have been proposed, see for example [25, 32, 38, 71].

The choice of homographies $\mathbf{H}_1, \mathbf{H}_2$ that fulfill

$$\bar{\mathbf{F}} = \mathbf{H}_2^{-T} \mathbf{F} \mathbf{H}_1^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3.46)$$

is not unique. A geometric explanation for this is, that there are infinitely many planes that are parallel to the camera baseline.

The particular choice of homographies can affect the quality of the rectified image pair, because the digital images have to be resampled when the projective transformations are applied. This is illustrated in figure 3.8, where a stereo pair is rectified with two different sets of homographies. Clearly, the choice of homographies for the stereo pair in the middle is unfortunate, as some parts of the images are compressed, resulting in loss of detail, whereas other parts of the image are stretched.

For the present application, the algorithm by Loop and Zhang [71] has been implemented to calculate a set of homographies that minimizes a measure of distortion for the rectified images. The derivation of the algorithm is quite lengthy and can be found in the original paper by Loop and Zhang [71]; only the main steps of the algorithm are outlined here.

In the following discussion, the homographies $\mathbf{H}_1, \mathbf{H}_2$ are represented by

$$\mathbf{H} = \begin{bmatrix} \tilde{\mathbf{u}}^T \\ \tilde{\mathbf{v}}^T \\ \tilde{\mathbf{w}}^T \end{bmatrix} = \begin{bmatrix} u_a & u_b & u_c \\ v_a & v_b & v_c \\ w_a & w_b & 1 \end{bmatrix} \quad (3.47)$$

where w_c is set to 1 as homographies are only defined up to a scale factor. Indices of 1 and 2, as in $u_{a,1}$, represent elements of \mathbf{H}_1 and \mathbf{H}_2 respectively; without indices the discussion holds for both homographies.

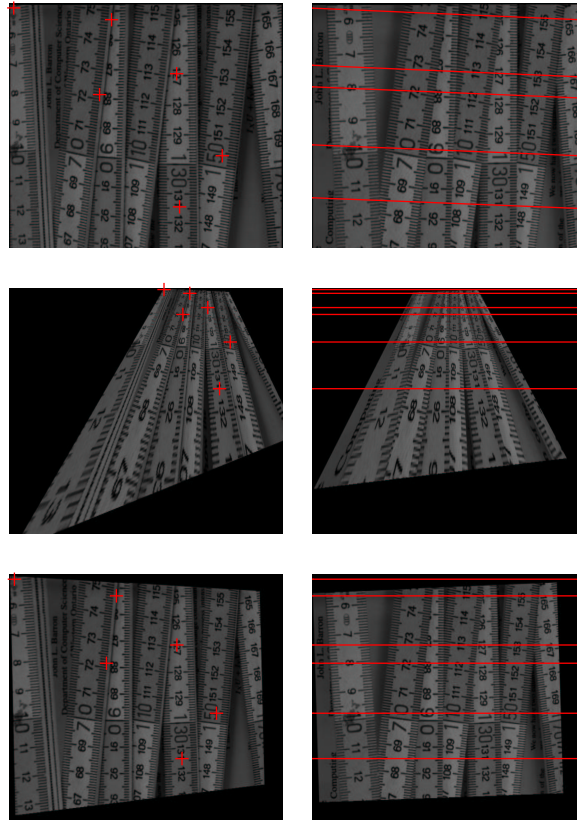


Figure 3.8: Effect of the choice of rectifying homographies on the rectified image pairs. The top row shows the original image pair, and the middle and bottom row show rectified image pairs obtained with a bad and a good choice of rectifying homographies. A few sample points and their corresponding epipolar lines are shown in red.

Decomposition of Homographies The main idea of the algorithm is to decompose the rectifying homographies into a projective part H_p and an affine part H_a , and to choose the free parameters of the projective part such that a cost function representing projective distortion is minimized. The affine transform is further subdivided into a similarity H_r , that is a rotation and a translation, and a shearing transform H_s . This results in the following two homographies

$$H_1 = H_{s,1}H_{r,1}H_{p,1} \quad \text{and} \quad H_2 = H_{s,2}H_{r,2}H_{p,2} \quad (3.48)$$

The projective part H_p of the transforms is used to bring the epipoles to infinity. The similarity transform H_r subsequently rotates them into alignment with the horizontal image scan-lines, that is the $(1, 0, 0)^T$ direction. One of the similarities is also used to translate one of the images vertically, such that corresponding points

lie on the same scan-line in each image. After this step, the images are rectified. The final shearing transform H_s , reduces the perspective distortion, if possible, by preserving the perpendicularity of lines and aspect ratio. It is also used to adjust the scale of the rectified images.

Projective Part H_p and Distortion Minimization The projective parts of the transform are used to map the epipoles to points at infinity and take the form

$$H_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ w_a & w_b & 1 \end{bmatrix}. \quad (3.49)$$

H_p will transform a point $\tilde{\mathbf{x}}_i = (x_i, y_i, 1)^T$ to $(\frac{x_i}{\gamma_i}, \frac{y_i}{\gamma_i}, 1)$, where γ_i is given by the dot product of the third row of H_p and $\tilde{\mathbf{x}}_i$,

$$\gamma_i = (w_a, w_b, 1) (x_i, y_i, 1)^T. \quad (3.50)$$

If the γ_i were the same for all points $\tilde{\mathbf{x}}_i$, there would be no projective distortion, and the transform H_p would be affine. However, it is not possible to use an affine transform to map the epipole to infinity unless it is already there. Therefore, the γ_i , which can be thought of as projective weight factors, will show some variation for different points $\tilde{\mathbf{x}}_i$.

The idea behind the distortion minimization used in [71] is to find the projective transform that maps the epipole to a point at infinity and is as close to an affine transform as possible. To decide whether a projective transform is close to affine, the variation of the γ_i over all image points is used as a criterion. This results in a cost function

$$f(w_{a,1}, w_{b,1}, w_{a,2}, w_{b,2}) = \sum_{i=1}^{n_1} \left(\frac{\gamma_{i,1} - \gamma_{c,1}}{\gamma_{i,1}} \right)^2 + \sum_{i=1}^{n_2} \left(\frac{\gamma_{i,2} - \gamma_{c,2}}{\gamma_{i,2}} \right)^2, \quad (3.51)$$

where the first sum runs over all pixels n_1 in the first image and the second sum runs over all pixels n_2 in the second image. The weights $\gamma_{c,1}$ and $\gamma_{c,2}$ are the weights of the center pixels of image J_1 and J_2 respectively. When minimizing this function, the fact that the parameters $w_{a,1}, w_{b,1}$ and $w_{a,2}, w_{b,2}$ are not fully independent, because they are related by the fundamental matrix, must be taken into account. The set of parameters that minimizes f , and therefore minimizes the variation of the weights γ_i over both images, yields the projective transforms that are closest to affine and therefore reduce distortion.

Similarity Transform The similarity transform \mathbf{H}_r is used to rotate the epipoles into alignment with the image scan-lines. Based on the assumption that $w_{a,1}, w_{b,1}$ and $w_{a,2}, w_{b,2}$ have been determined by minimizing equation (3.51), Loop and Zhang arrive at the following results

$$\mathbf{H}_{r,1} = \begin{bmatrix} F_{32} - w_{b,1}F_{33} & w_{a,1}F_{33} - F_{31} & 0 \\ F_{31} - w_{a,1}F_{33} & F_{32} - w_{b,1}F_{33} & F_{33} + v_{c,2} \\ 0 & 0 & 1 \end{bmatrix} \quad (3.52)$$

and

$$\mathbf{H}_{r,2} = \begin{bmatrix} w_{b,2}F_{33} - F_{23} & F_{13} - w_{a,2}F_{33} & 0 \\ w_{a,2}F_{33} - F_{13} & w_{b,2}F_{33} - F_{23} & v_{c,2} \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.53)$$

where the F_{ij} are the elements of the fundamental matrix. The parameter $v_{c,2}$ should be chosen such that corresponding points lie on the same scan-line.

Shearing Transform The final shearing transform is of the form

$$\mathbf{H}_s = \begin{bmatrix} \delta_1 & \delta_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.54)$$

It is used to reduce the projective distortion introduced by \mathbf{H}_p . This is done by adjusting the free parameters δ_1 and δ_2 such that the perpendicularity of lines and the aspect ratio of each image is preserved. Given points $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ in the original image as defined in figure 3.9 and their counterparts $\mathbf{a}', \mathbf{b}', \mathbf{c}', \mathbf{d}'$ in the image transformed by $\mathbf{H}_r\mathbf{H}_p$, one can define the Euclidean 2-vectors \mathbf{h}, \mathbf{v} connecting the mid-points of the transformed image borders.

With $\tilde{\mathbf{h}} = (\mathbf{h}^T, 0)$ and $\tilde{\mathbf{v}} = (\mathbf{v}^T, 0)$ perpendicularity after applying the shearing transform is ensured if

$$(\mathbf{H}_s\tilde{\mathbf{h}})^T(\mathbf{H}_s\tilde{\mathbf{v}}) = 0. \quad (3.55)$$

The original aspect ratio is preserved after the shearing transform if

$$\frac{(\mathbf{H}_s\tilde{\mathbf{h}})^T(\mathbf{H}_s\tilde{\mathbf{h}})}{(\mathbf{H}_s\tilde{\mathbf{v}})^T(\mathbf{H}_s\tilde{\mathbf{v}})} = \frac{n_w^2}{n_h^2}, \quad (3.56)$$

where n_w and n_h are the width and the height of the original image. The values of δ_1, δ_2 for which \mathbf{H}_s satisfies equations (3.55) and (3.56) are given in Loop and Zhang [71] as

$$\delta_1 = \frac{n_h^2 h_y^2 + n_w^2 v_y^2}{n_w n_h (h_y v_x - h_x v_y)} \quad \text{and} \quad \delta_2 = \frac{n_h^2 h_y h_x + n_w^2 v_x v_y}{n_w n_h (h_x v_y - h_y v_x)}. \quad (3.57)$$

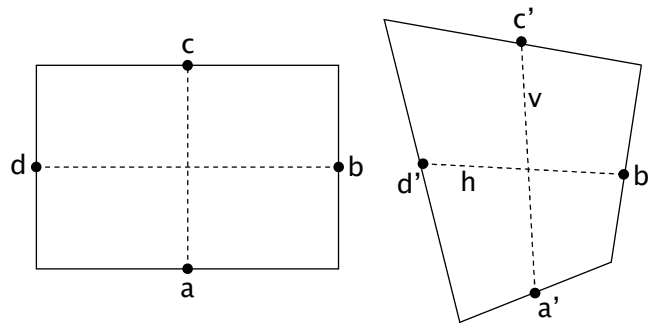


Figure 3.9: Definition of the midpoints a, b, c, d of the image borders in the original image (left) and their counterparts a', b', c', d' in the image transformed by $H_r H_p$ (right).

The shearing transforms for both images can be calculated this way, yielding the rectifying homographies $H_1 = H_{s,1}H_{r,1}H_{p,1}$ and $H_2 = H_{s,2}H_{r,2}H_{p,2}$. In general, the rectified images resulting from these homographies still have to be scaled. An appropriate scale factor can be chosen by requiring the combined area of the rectified images to be the same as for the original image pair.

3.5.5 Triangulation

The goal of stereo imaging is usually to obtain a 3D-reconstruction of an object, for example the water surface in the present application. To obtain the 3D-coordinates of a point \mathbf{X} , which projects to the image points \mathbf{x}_1 and \mathbf{x}_2 in two images \mathcal{J}_1 and \mathcal{J}_2 , one has to find the intersections of the two lines of sight that go through the image points and their respective camera centers.

In general, the positions of \mathbf{x}_1 and \mathbf{x}_2 are not known exactly, but are subject to uncertainties arising from noise or errors in the search for corresponding points. Therefore, the lines of sight might not intersect and an approximate solution must be sought.

Several methods for choosing an approximate solution are summarized and assessed by Hartley and Sturm [43]. For the present application a linear triangulation method, described below, was chosen. As demonstrated in [43], this method does not perform very well in cases where the camera projection matrices are only known up to a projective transform. As the camera projection matrices are known in this work, this is not a critical issue, and the use of the linear triangulation method is justified.

Linear Triangulation Method By substituting the projection matrices P_1, P_2 of the two cameras and a pair of corresponding image points $\tilde{\mathbf{x}}_1 = w_1(x_1, y_1, 1)^T$ and $\tilde{\mathbf{x}}_2 = w_2(x_2, y_2, 1)^T$ into equation (3.13) one obtains

$$\tilde{\mathbf{x}}_1 = P_1 \tilde{\mathbf{X}} = \begin{bmatrix} P_{1,a}^T \\ P_{1,b}^T \\ P_{1,c}^T \end{bmatrix} \begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ \tilde{W} \end{pmatrix} \quad (3.58)$$

and an analogous equation for $\tilde{\mathbf{x}}_2$ and P_2 . If the rows of the projection matrix are expressed by the 4-vectors $P_{a\dots c}^T$ as above, this linear system yields the following three equations:

$$\begin{aligned} w_1 x_1 &= P_{1,a}^T \tilde{\mathbf{X}} \\ w_1 y_1 &= P_{1,b}^T \tilde{\mathbf{X}} \\ w_1 &= P_{1,c}^T \tilde{\mathbf{X}}. \end{aligned} \quad (3.59)$$

The third equation for w_1 can be substituted into the other equations, yielding two equations for three unknowns. An analogous substitution can be applied to the equations corresponding to the second camera, and the resulting system of four linear equations can be arranged into a matrix form as

$$\mathbf{A} \tilde{\mathbf{X}} = \begin{bmatrix} x_1 P_{1,c}^T - P_{1,a}^T \\ y_1 P_{1,c}^T - P_{1,b}^T \\ x_2 P_{2,c}^T - P_{2,a}^T \\ y_2 P_{2,c}^T - P_{2,b}^T \end{bmatrix} \tilde{\mathbf{X}} = \mathbf{0}. \quad (3.60)$$

The least-squares solution for $\tilde{\mathbf{X}}$ yields the position of the scene point corresponding to $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2$ in homogeneous coordinates.

In the present implementation, this equation is solved by computing the eigenvector corresponding to the smallest eigenvalue of $\mathbf{A}^T \mathbf{A}$ using the singular value decomposition (see for instance Golub and van Loan [39] or any other textbook on matrix computation).

Finally, the Euclidean coordinates of the scene point corresponding to the image points $\mathbf{x}_1, \mathbf{x}_2$ are obtained by $\mathbf{X} = 1/\tilde{W}(\tilde{X}, \tilde{Y}, \tilde{Z})^T$.

If, as in the present work, the stereo image pair is rectified by a pair of homographies H_1, H_2 , the triangulation can be performed on pairs of rectified points. In this case, the original projection matrices have to be replaced by the effective projection matrices after rectification, namely $P_{1r} = H_1 P_1$ and $P_{2r} = H_2 P_2$.

4 Stereo Computer Vision II: Solving the Correspondence Problem

The concepts outlined in the previous chapter can be used to infer the 3D coordinates of a scene point given its image points in two different views.

In this chapter, the problem of how to identify a pair of image points $\mathbf{p}_1, \mathbf{p}_2$, corresponding to a common point \mathbf{P} in the scene, given two images of a scene $\mathcal{J}_1, \mathcal{J}_2$ is addressed. Such corresponding points are sometimes referred to as *homologous* points and the problem of identifying them is known as the *stereo correspondence problem*. In the computer vision literature, a vast number of stereo matching algorithms for solving the correspondence problem can be found. Section 4.1 classifies these algorithms, which fall into two main categories: feature-based, briefly discussed in section 4.1.1, and area-based, analyzed in more detail in section 4.1.2. Based on the recent taxonomy of area-based stereo matching algorithms by Scharstein and Szeliski [82], it is shown how these algorithms can be broken up into a few characteristic building blocks.

Area-based matching algorithms are based on the premise that the respective local neighborhoods around two homologous image points look similar. This premise can sometimes be violated. Therefore, section 4.2 reviews the prerequisites for area-based matching and shows that, for a water surface, they are better fulfilled for infrared images than for images acquired at visible wavelengths.

The matching algorithm used in the present application and its implementation are presented in section 4.3.

Note that the difference between the image coordinates of two homologous points is commonly referred to as *disparity*. Therefore the process of matching corresponding points in stereo images is also known as *disparity estimation*.

4.1 Classification of Matching Algorithms

In order to motivate the choice of matching algorithm used in the present work, this section gives a brief overview of existing stereo algorithms. A more extensive review can be found in the recent paper by Brown et al. [13]. The paper by Dhond and Aggarwal [19], though outdated in some points, can still serve as a good introduction. The work by Scharstein and Szeliski [82] focuses on area-based methods.

As mentioned above, stereo matching algorithms can be divided into two major categories. Feature-based algorithms seek correspondences only for salient image features, such as edges or corners, whereas area-based algorithms try to find correspondences for every image point. The focus of this overview is on area-based algorithms, as they allow a dense surface reconstruction, which is a requirement for the present application.

4.1.1 Feature-based Stereo Matching

In feature-based stereo matching, the complexity of the matching process is reduced by a pre-processing step in which prominent features like corners (points with high grey value gradient in all directions) and edges are extracted from the image. The matching is then performed on these features only. The reduction in the number of candidate points significantly speeds up the algorithm. Furthermore, the chance of ambiguous matches is reduced.

One application of feature-based stereo matching in a field related to the study of water waves is *stereo particle tracking velocimetry*. This technique has been employed at the Institute for Environmental Physics, University of Heidelberg, to analyze the subsurface turbulence generated by water waves [see 24, 34, 105] and to study the effects of ship-induced turbulence on sediment transportation in waterways [see 62, 65, 92, 94]. In both applications, moving particles (tracer or sediment particles) create streaks in the images of both cameras. These streaks are then segmented from the background and matched, eventually leading to a 3D-reconstruction of the particles' trajectories and velocities.

Feature-based algorithms generate only a sparse depth map and are therefore not suitable for the present application.

4.1.2 Area-Based Stereo Matching

In area-based disparity matching, small area patches around a point in one image are compared to area patches around every point on the corresponding epipolar line (see section 3.5.2) in the other image, usually by computing the correlation of the image intensities (grey-values). Based on the assumption that the areas surrounding two image points corresponding to the same scene point are similar, points with a higher correlation score are assigned a higher probability of being the correct match. For this assumption to be justified, a number of conditions, discussed in detail in section 4.2, have to be met. With area-based matching, a match for every image pixel is sought, leading to a dense 3D reconstruction.

In the following paragraphs, the main building blocks of area-based algorithms are outlined. First the representation of matches using a disparity map is defined. This is followed by a dissection of the area-based algorithms into three main steps based on work by Scharstein and Szeliski [82].

Representation of Matches It is common to represent matching points by their shift in image coordinates (from one image to the other), expressed in terms of a uni-valued *disparity map* $\mathcal{D}(x, y)$. This disparity map is defined as

$$\mathcal{J}_1(x, y) \overset{\text{matches}}{\longleftrightarrow} \mathcal{J}_2(x + \mathcal{D}(x, y), y). \quad (4.1)$$

This means that for every pixel coordinate (x, y) in image \mathcal{J}_1 , the corresponding image point in image \mathcal{J}_2 can be found by adding the disparity $\mathcal{D}(x, y)$ to the x-coordinate. Disparity can therefore be interpreted as a shift in pixels.

Note that one image of the stereo pair plays the role of the reference image with respect to which the disparity is defined. Here, \mathcal{J}_1 was arbitrarily chosen as the reference image; it is of course also possible to use \mathcal{J}_2 as the reference.

The disparity is only added to the x-coordinate, taking into account that the match must lie on the epipolar line (see section 3.5.2), and assuming that the epipolar lines have been aligned with the x-axis through a rectification step as described in section 3.5.3.

Algorithm Building Blocks To provide a framework for their taxonomy of area-based matching algorithms, Scharstein and Szeliski [82] break down the algorithms into three major building blocks:

1. computation of a correlation score for each possible match,
2. aggregation of the matching score over local neighborhoods, and
3. computation, and possibly optimization, of the disparity.

Scharstein and Szeliski then classify the matching algorithms based on the approaches used to perform the above steps. Most stereo algorithms fit into this classification scheme, although many algorithms combine one or more of the steps into a single step for implementation reasons. For example, the algorithm chosen for the present application, described in section 4.3, combines the computation of the matching score with the aggregation step.

Matching Score The first step of a stereo algorithm consists of calculating a similarity measure for every pair of candidate points. This similarity measure is

often referred to as matching score or correlation score. Common matching scores used as a similarity measure include absolute differences (AD), squared differences (SD) and cross-correlation of image grey-values.

The matching scores for all image pixels can be represented as a data volume $\mathcal{C}_0(x, y, d)$. This data volume is referred to as *disparity space image* [82]. As an example, for a matching score of absolute differences, the disparity space image is calculated as

$$\mathcal{C}_0(x, y, d) = |\mathcal{J}_1(x, y) - \mathcal{J}_2(x + d, y)|. \quad (4.2)$$

Depending on the image size and the range of disparities that is searched, the data volume $\mathcal{C}_0(x, y, d)$ can take up large amounts of memory. For efficiency reasons, many implementations, including the present application, therefore only store the matching scores for a single image line (see 4.3.2).

The major differences between different matching scores is their sensitivity to grey-value transformations (scaling and offset) and to outliers. However, the findings of Scharstein and Szeliski [82] and Faugeras et al. [26] suggest that the result of the stereo matching processes is not strongly dependent on the particular choice of matching score.

Aggregation of Matching Score In general, matches cannot be found based on comparing only single pixels. This is because the grey-value is affected by camera noise and view angle, and because other, non-corresponding, pixels might have a similar grey-value.

Therefore, it is necessary to consider pixel neighborhoods. For two pixels that are a correct match, it is likely that their respective neighboring pixels will also have a high degree of similarity. However, for two pixels that don't match, the matching score of their respective neighboring pixels is likely to be lower. By taking the neighboring pixels into account, the comparison is effectively of area patches and not single pixels. A simple way to take the neighborhood into account is to sum up the matching score over a fixed window of support around each pixel. This can be done by convolving the disparity space image $\mathcal{C}_0(x, y, d)$ with a filter mask $w(x, y, d)$ that computes a local average, for example a rectangular correlation window (see 4.3.2) or a Gaussian. Mostly, the aggregation is only performed along the x- and y-coordinates, but some algorithms also aggregate over neighboring disparity values. The general case looks as follows.

$$\mathcal{C}(x, y, d) = w(x, y, d) * \mathcal{C}_0(x, y, d). \quad (4.3)$$

More elaborate aggregation schemes use, for example, adaptive filter sizes, applying a larger filter in areas with little texture. Other algorithms, for example [109], use iterative aggregation schemes that apply, in effect, anisotropic smoothing (see, for example [80]) to the disparity space image.

Computation and Optimization of Disparity In the final step of the algorithm, the disparity for each pixel in the reference image is computed. A common approach, implemented in the present application, is to choose the value of d with the maximum correlation score:

$$\mathcal{D}(x, y) = \underset{d}{\operatorname{argmax}} \mathcal{C}(x, y, d). \quad (4.4)$$

In other approaches, the correlation score is used as part of a global cost function. Such a cost function can for example be used to enforce a smooth spatial variation of the disparity estimates, similar to the regularization technique described in section 5.3.3. The disparity map is found as the solution that minimizes the overall cost function.

4.2 Prerequisites for Area-Based Matching

As mentioned in section 1.3, stereo matching algorithms can fail if the imaged surface has specular reflection characteristics. Area-based stereo matching is based on the assumption that the neighborhoods centered around homologous image points look similar. In addition to specular reflections, there are other situations for which this assumption is violated. Such situations arise when the image of a surface is strongly dependent on view angle. This view angle dependence can cause patches centered around corresponding points to look dissimilar in the images of two cameras.

In this section, the assumptions about the scene properties underlying area-based stereo matching algorithms will be reviewed in detail.

4.2.1 Fronto-Parallel

If a surface S is observed at a grazing angle, small changes of the viewing angle have a large effect on the size of the image of S .

To quantify this effect, consider a small solid angle Ω (see section 2.1.1) as seen from a camera imaging the surface. This solid angle Ω corresponds to an image

area a_Ω . The size of the surface patch in the 3D scene corresponding to a_Ω will be denoted by A_\perp when S is perpendicular to the view direction ($\theta = 0$). When the surface is tilted by an angle θ , the covered scene area $A(\theta)$ corresponding to the same fixed image area a_Ω is enlarged by a factor of $1/\cos(\theta)$:

$$A(\theta) = \frac{A_\perp}{\cos \theta}. \quad (4.5)$$

The view angle dependent change in size of this area is given by the derivative:

$$\frac{dA(\theta)}{d\theta} = A(\theta) \tan \theta. \quad (4.6)$$

Therefore, at grazing angles (large θ), a small change in viewing angle will have a large foreshortening effect on the image, whereas for close to normal angles (small θ) such a change will not have a major effect.

The effect of this view angle dependence on a stereo camera system is illustrated in figure 4.1, which depicts two cameras imaging a surface that is tilted with respect to the camera baseline. Due to foreshortening, the solid angle covered by the surface patch A is larger for the left camera than for the right camera. The dissimilarity in image size reduces the correlation between the image patches resulting in a decreased performance of correlation-based disparity estimation.

Therefore, the so-called *fronto-parallel assumption* underlies most correlation-based algorithms: it is assumed that the objects in the scene are imaged frontally and that their surfaces are oriented more or less parallel to the baseline connecting the two camera centers.

4.2.2 Lambertian Surface

In section 2.1.1, a Lambertian surface was defined as a surface for which the radiance is constant, independent of view angle. This is a property that is very desirable in stereo vision, because a non-Lambertian surface introduces a dependence of the grey-value on the view angle in addition to the foreshortening effect described above.

In particular, specular surfaces cause problems, because they exhibit highlights when the angle of incidence of the light source is equal to the view angle. These highlights are very distinct features, and a correlation-based algorithm will therefore often match them across the images. This leads to errors in the reconstruction

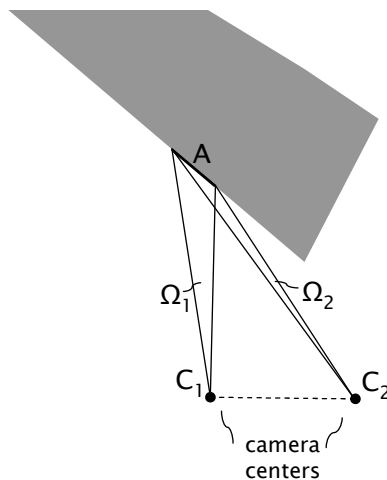


Figure 4.1: Foreshortening effect for a surface that is tilted with respect to the baseline. The solid angles Ω_1 and Ω_2 covered by the same surface patch A differ significantly in size.

as the highlights observed from different view angles do not usually correspond to the same point on the surface. An example for the errors introduced by specular reflection is also given in the introduction, see figure 1.9.

For mirror-like surfaces, the disparity algorithm will match the reflection of the surroundings, which depends on the angle of incidence. This situation is depicted in figure 1.9 on page 20.

4.2.3 Opacity

A fundamental assumption in correlation-based matching is that the surfaces of the objects being imaged are opaque. When transparent surfaces are present in the scene, the radiance received by the camera along a line of sight may be the cumulative radiance originating from several points along that line, as depicted in figure 4.2. For different view angles, the image of a point on a transparent surface will thus have different grey-values depending on what lies behind it on the particular line of sight. Therefore, the correlation based on the grey-value will fail in such cases. In fact, independent of the matching approach, stereo vision cannot recover a 3D structure in the general case involving transparent media. This problem can only be solved using a volume reconstruction technique such as tomography.

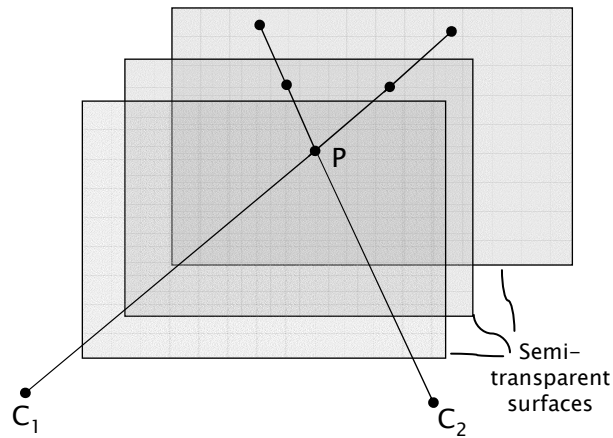


Figure 4.2: When transparent objects are present, the radiance received by the cameras C_1 and C_2 from the direction of point P is the sum of the contributions from several points along the line of sight.

4.2.4 Texture

Correlation-based matching is ineffective on textureless surfaces with a constant grey-value, because this leads to a constant matching score everywhere on the surface. For such a case, the assumption that a match is characterized by a maximum of the correlation score is wrong.

Therefore, it is necessary that objects in the scene be textured; that is, that they show some spatial variation in grey-value. Whether the correlation score exhibits a single pronounced maximum depends on the statistical properties of the grey-value pattern. In particular, spatially periodic textures produce several local extrema in the matching score, which makes the determination of the true maximum difficult or impossible.

4.2.5 Summary and Conclusions Regarding Wave Imaging

Many of the assumptions presented regarding scene properties that are used in area-based stereo matching algorithms are clearly violated for water at visible wavelengths. Water in the visible region is transparent, has specular reflection properties, and does not have a lot of surface texture. Examples of this were given in section 1.3.

In contrast, water in the $3 - 5 \mu m$ infrared wavelength region exhibits the surface properties required for stereo matching; as shown in section 2.1.4 it is opaque and has a rich surface texture, demonstrated in figure 1.3 on page 12. Therefore, water in this wavelength region is ideally suited for stereo matching.

4.3 Multi-scale, Area-Based Matching Algorithm using Shiftable Windows

Many disparity estimation algorithms are tailored for situations often found in computer vision applications, such as robot navigation or the reconstruction of architectural features. In these applications, the scene is usually comprised of multiple objects, giving rise to occlusions and surface discontinuities. Moreover, these objects often also have different surface properties. This causes problems, as aggregation of the matching score across object boundaries tends to smooth out the discontinuities. Some algorithms address this issue and try to preserve sharp object boundaries by using adaptive filter windows or iterative aggregation of support, as mentioned in section 4.1.2.

The situation encountered when reconstructing water waves using infrared imagery is more favorable, insofar as the water surface is a single “smooth” surface without sharp discontinuities. Given this situation, it is possible to use a comparatively simple algorithm using fixed window sizes for spatial aggregation of support and without a computationally expensive optimization stage.

This section discusses the algorithm employed for the present application, which is an adaption of an algorithm described by Faugeras et al. [26]. This algorithm was chosen because it is computationally efficient, easy to implement and it produced good results. The computational efficiency stems from the fact that the computation of the matching score and the spatial aggregation are combined into a single step that is implemented using shiftable correlation windows as described in section 4.3.2.

The algorithm computes one disparity map using the first and one disparity map using the second image of the stereo pair as the reference image. Matches that are not consistent across the two maps are discarded to produce a final disparity map. This validation procedure is explained in section 4.3.5.

Sub-pixel accuracy is achieved by interpolating the correlation score between discrete values of disparity with a parabola as described in section 4.3.4.

A simple multi-scale extension using *image pyramids* is presented in section 4.3.6. For this, the basic algorithm described in sections 4.3.1 to 4.3.5 is applied to several

down-sampled versions of the stereo pair. This has a similar effect to running the algorithm with differently-sized correlation windows, but it is computationally more efficient. The matches found on coarser scales can be used to fill in regions where a match cannot be found at the original resolution, for example due to lack of texture.

4.3.1 Matching Score

Faugeras et al. [26] describe several different choices for the matching score, and note that three out of the four correlation criteria they tested perform about equally well for cases in which the grey-value distributions of the two images are similar. The difference between the correlation criteria they tested lies mainly in their sensitivity to differences in the grey-value histograms of both images. In the present application, the radiometric calibration process and the non-uniformity correction, described in section 5.1 ensure that the grey-value distributions of the two images are matched, so that the particular choice of the matching score is not critical. For this application, cross-correlation (matching score C_2 in Faugeras et al. [26]) was used as the matching score:

$$\mathcal{C}(x, y, d) = \frac{\sum_{i,j} J_1(x+i, y+j) \cdot J_2(x+d+i, y+j)}{\sqrt{\sum_{i,j} J_1(x+i, y+j)^2} \cdot \sqrt{\sum_{i,j} J_2(x+d+i, y+j)^2}}. \quad (4.7)$$

The summations run over a $(2n+1) \cdot (2m+1)$ rectangular window. This function has higher values for patches that are better correlated.

Note that the computation of equation (4.7) combines two of the steps outlined in section 4.1.2, that is, the computation of the correlation score and the spatial aggregation. Here, spatial aggregation is controlled empirically by selection of the window size, given by the two parameters m and n . When choosing this size, a balance has to be found between blurring details by choosing too big a window versus not having a good enough signal-to-noise ratio in regions with low texture if the window is too small.

4.3.2 Efficient Implementation using Shiftable Windows

Computation of the Correlation Score To evaluate the correlation score (4.7), one has to iterate over the variables x, y and d . This involves many redundant

multiplications, if the sums in equation (4.7) are evaluated by explicitly iterating over all $(2n + 1) \cdot (2m + 1)$ elements of the rectangular window for each combination of x, y and d . This is very inefficient and leads to long processing times if correspondences are sought in large images and over a wide range of disparities.

To obtain a computationally efficient implementation, redundant multiplications must be avoided. This section presents a simple strategy for doing so, using shiftable correlation windows.

In equation (4.7), the first sum under the square root in the denominator is independent of the disparity d . Therefore, this sum is constant for a fixed x and y . As a constant factor is of no interest when searching the maximum over d one can consider the simplified matching score

$$\mathcal{C}(x, y, d) = \frac{\sum_{i,j} \mathcal{J}_1(x + i, y + j) \cdot \mathcal{J}_2(x + d + i, y + j)}{\sqrt{\sum_{i,j} \mathcal{J}_2(x + d + i, y + j)^2}}. \quad (4.8)$$

Next, the numerator and the denominator of expression (4.8) are considered separately:

$$\mathcal{C}(x, y, d) = \frac{N(x, y, d)}{\sqrt{M(x + d, y)}} \text{ with} \quad (4.9)$$

$$N(x, y, d) = \sum_{i,j} \mathcal{J}_1(x + i, y + j) \cdot \mathcal{J}_2(x + d + i, y + j) \text{ and} \quad (4.9)$$

$$M(x, y) = \sum_{i,j} \mathcal{J}_2(x + d + i, y + j)^2 \quad (4.10)$$

The strategy to avoid redundant computations is best explained using figure 4.3. For both the numerator and the denominator, first the elements of the product term along columns of length $2m + 1$ starting in the first row are summed up. Next, starting in the first column, $2n + 1$ of these column sums are summed to obtain the sum over a complete $(2n + 1) \times (2m + 1)$ window. Then, proceeding along the row, this window sum can be updated for each step by subtracting the column sum on the left and adding the next column on the right.

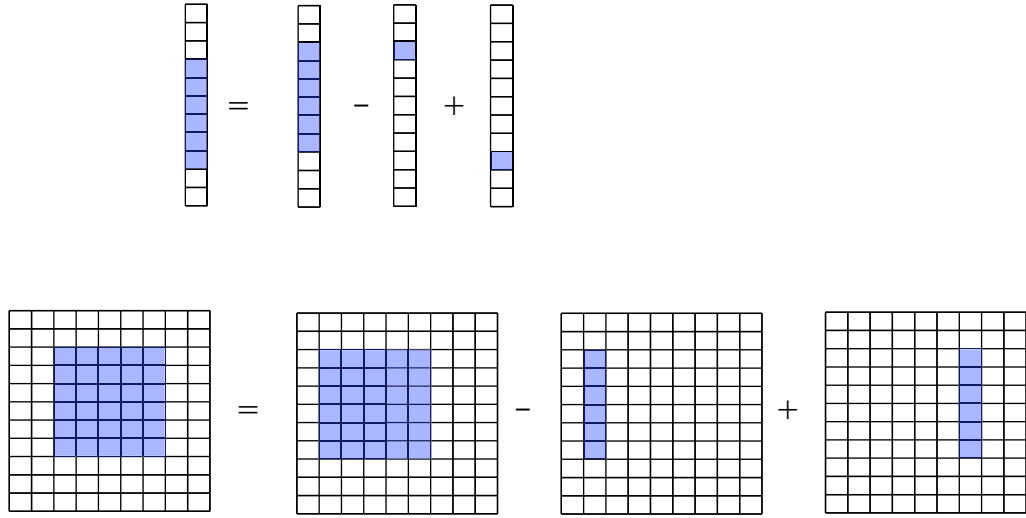


Figure 4.3: Efficient computation of matching score using shiftable windows. Product terms (blue squares) from the next column/row are added and pixels from the first column/row are subtracted from the window when proceeding to the next column/row.

For the numerator $N(x, y, d)$ this procedure leads to the following equations: Let $P_1(x, y, d) = J_1(x, y) \cdot J_2(x + d, y)$ be a shorthand for the product term in the sum of the numerator, and let

$$Q_1(x, 0, d) = \sum_{j=-m}^m P_1(x, j, d) \quad (4.11)$$

be the first column sum at position x . Once the column sums of the first row have been calculated for all x , the correlation window sum for the first pixel can be calculated as

$$N(0, y, d) = \sum_{i=-n}^n Q_1(i, y, d). \quad (4.12)$$

All the other pixels can then be calculated by iterating along the rows and columns using the following update rules

$$\begin{aligned} N(x + 1, y, d) &= N(x, y, d) - Q_1(x - n, y, d) + Q_1(x + n + 1, y, d) \\ Q_1(x, y + 1, d) &= Q_1(x, y, d) - P_1(x, y - m, d) + P_1(x, y + m + 1, d). \end{aligned} \quad (4.13)$$

The same scheme can be applied to the denominator, yielding the following algorithm: With $P_2(x, y) = \mathcal{J}_2(x, y)^2$, the column sum in the denominator is initialized as

$$Q_2(x, 0) = \sum_{j=-m}^m P(x, j) \quad (4.14)$$

for each column. The row sum is similarly initialized as

$$M(0, y) = \sum_{i=-n}^n Q_2(i, y). \quad (4.15)$$

The update rules for the denominator are

$$\begin{aligned} M(x+1, y) &= M(x, y) - Q_2(x-n, y) + Q_2(x+n+1, y) \\ Q(x, y+1) &= Q_2(x, y) - P_2(x, y-m) + P_2(x, y+m+1). \end{aligned} \quad (4.16)$$

For the image borders, some of the sums above are not well defined, as they include pixels coordinates that are outside the actual images. A common approach to deal with this, is to pad the image borders with zeros, which is the method used in the present application. Other approaches, such as taking the value of the pixel closest to the border for padding, can also be used.

4.3.3 Computation of Disparity

Once the correlation score for a given point has been calculated for a given image point $(x_0, y_0)^T$ over the whole disparity search range $d_{\text{low}} \dots d_{\text{high}}$, the value of d that maximizes function (4.8) is chosen as the disparity value for this image point:

$$\mathcal{D}(x_0, y_0) = \operatorname{argmax}_{d \in [d_{\text{low}} \dots d_{\text{high}}]} \frac{N(x_0, y_0, d)}{\sqrt{M(x_0 + d, y_0)}}. \quad (4.17)$$

4.3.4 Sub-Pixel Refinement

Using the above method, the correlation score is only computed for discrete, whole-numbered, values of disparity.

To estimate the disparity with sub-pixel precision, a parabola is used to interpolate between the discrete disparity values in the neighborhood of the detected maximum.

Assuming that d_{\max} is the discrete disparity value that maximizes (4.8) for a given pixel $(x, y)^T$, the interpolating parabola $P(d)$ is determined by d_{\max} and its neighboring points:

$$\begin{aligned} P(d_{\max} - 1) &= \mathcal{C}(x, y, d_{\max} - 1) \\ P(d_{\max}) &= \mathcal{C}(x, y, d_{\max}) \\ P(d_{\max} + 1) &= \mathcal{C}(x, y, d_{\max} + 1). \end{aligned} \quad (4.18)$$

The final sub-pixel disparity estimate is then given as the value d for which $P(d)$ has its maximum; this value is easily computed using elementary calculus.

4.3.5 Validation of Matches

Left/Right Consistency Check An effective test for checking the validity of the found matches can be applied by computing the correlation score (4.8) twice, using each of the images as the reference image in turn.

Correct matches are expected to be consistent for both runs of the algorithm. In contrast, if a match is found to be inconsistent when reversing the role of the reference image it is labeled as erroneous and is discarded. The positions of matches that fail the consistency check are marked in a binary image mask.

The rationale for this consistency check is as follows: Assume that, for a given pixel \mathbf{p}_1 in image \mathcal{J}_1 , the corresponding point in the other image cannot be found, for example due to occlusion. The correlation algorithm will then, more or less at random, assign a pixel \mathbf{p}_2 in the second image \mathcal{J}_2 as a match. However, \mathbf{p}_2 might have a corresponding pixel in \mathcal{J}_1 and therefore it is not likely that it will be matched to \mathbf{p}_1 if the roles of the images are reversed.

However, note that this consistency check cannot detect wrong matches that are due to similar looking pixels that do not correspond to the same surface point; a situation that can arise from specular highlights.

Mathematically this consistency criterion can be formulated as follows. Let \mathcal{D}_1 and \mathcal{D}_2 be the disparity maps obtained when using images \mathcal{J}_1 and \mathcal{J}_2 as the reference image, respectively. The consistency criterion then becomes

$$\mathcal{D}_1(x, y) \stackrel{!}{=} -\mathcal{D}_2(x + \mathcal{D}_1(x, y), y). \quad (4.19)$$

In this work, it was implemented such that all pixels for which

$$|\mathcal{D}_1(x, y) + \mathcal{D}_2(x + \mathcal{D}_1(x, y), y)| > T, \quad (4.20)$$

where T is a manually selected threshold, were discarded. The value of T was set to 0.5. As the consistency check is performed after the sub-pixel refinement, $\mathcal{D}_1(x, y)$ is generally not an integer value, therefore $\mathcal{D}_2(x + \mathcal{D}_1(x, y), y)$ is linearly interpolated when evaluating equation (4.20).

Confidence Measure In addition to the above mentioned left/right consistency check, Faugeras et al. [26] propose a confidence measure associated with each disparity estimate. This confidence measure is based on how pronounced the maximum d_{\max} of the correlation score is. For this, they suggest using the difference between the two largest local maxima as a measure. The rationale for choosing this measure is, that, in a situation with several possible candidate matches, for example due to a periodic texture, they are likely to have a similar correlation score. The confidence in such a match will therefore be low.

However, it is unclear how this confidence measure could be properly integrated into a multi-scale approach, as described in the following section. Therefore, it was not used in the present work.

4.3.6 Multi-Scale Approach

Sometimes no correspondences can be found for some areas of an image because of noise or lack of texture. In these cases, it is often possible to find a match if the correlation is performed on a down-sampled version of the original image.

Gaussian pyramids Therefore, the algorithm described in the previous sections was extended to process the original image pair at multiple scales. For this, so called *Gaussian image pyramids* (see, for example Jähne [56]) were used. Starting at the original image resolution, a Gaussian pyramid is created by low-pass filtering the image and then sub-sampling it by taking pixels of every other column on every other row. This is repeated several times to obtain images of different resolutions, which can be visualized as different levels of a pyramid, see figure 4.4.

The low-pass filtering is required to avoid aliasing effects such as Moiré patterns when sub-sampling. In the present work, the low-pass filtering was implemented as a convolution of the image with a 5-entry Binomial filter tap, $\frac{1}{16} (1, 4, 6, 4, 1)$, before sub-sampling and another convolution with a 3-entry Binomial filter tap, $\frac{1}{4} (1, 2, 1)$, after sub-sampling. Both convolutions were performed in x- and y-direction.

Disparity at Different Scales After Gaussian pyramids for both images of a stereo pair have been calculated, the correlation-based disparity estimation as

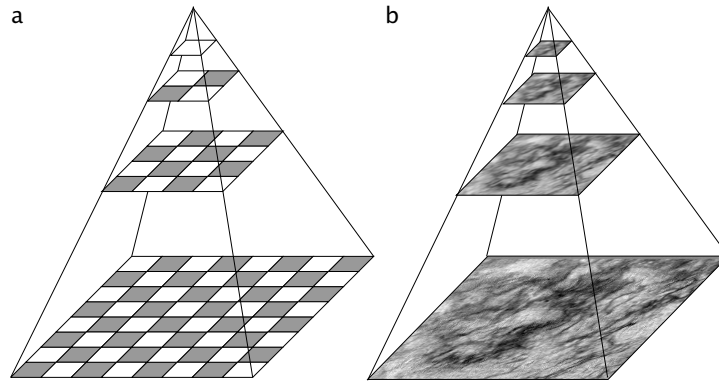


Figure 4.4: Gaussian pyramid. With a Gaussian pyramid, an image is represented at several scales. The original image is low-pass filtered and then down-sampled to obtain the next pyramid level. A schematic representation of this, in which pixels are represented by checkers, is given on the left. With each step up the pyramid, the number of pixels is reduced by a factor of four. On the right, the Gaussian pyramid obtained by down-sampling an infrared image of a water surface is shown. Image courtesy of Uwe Schimpf, University of Heidelberg.

described in sections 4.3.1 to 4.3.5 is applied to each pyramid level. This yields several disparity maps $\mathcal{D}_0, \dots, \mathcal{D}_{l_{\max}}$, where the pyramid level is denoted by the subscript and level 0 represents the original resolution. Note that this meaning of the subscript is different than in the previous section, where it was used to denote which image was used as reference. The combined disparity map is obtained using algorithm 4.1.

```

For each pixel  $(x, y)$ 
   $l = 0$  // pyramid level
  While ( $\mathcal{D}_l(x, y)$  is NOT valid) AND  $l \leq l_{\max}$ 
     $l = l + 1$  // next level
  EndWhile
  If  $l \leq l_{\max}$ 
     $\mathcal{D}_{\text{final}}(x, y) = 2^l \mathcal{D}_l \cdot (2^{-l}x, 2^{-l}y)$ 
  Else // no valid match at any scale
     $\mathcal{D}_{\text{final}}(x, y) = \text{not valid}$ 
  EndIf
Next pixel

```

Algorithm 4.1: Algorithm for combining disparity estimates at different pyramid levels.

The algorithm first tries to find a valid match at the highest resolution. If no

valid match is found, it proceeds to the higher pyramid levels until a valid match is found or the highest pyramid level is reached. The pixel positions for which no valid match was found on any scale are stored using a binary image. Strictly, the disparity estimates from higher pyramid levels would have to be interpolated when they are used to fill in gaps at the lowest level. Without interpolation, one obtains some block artifacts, as a disparity estimate at a coarse scale fills in up to four pixels at the next finer scale with the same value.

The results obtained with the stereo matching algorithm described in sections 4.3.1 to 4.3.6 on infrared image sequences of water waves are described in sections 7.4 and 7.7.

5 Image Pre- and Postprocessing

This chapter presents image processing techniques for non-uniformity correction of infrared images, for removing outliers from disparity maps and for filling in missing data. The technique for filling in missing data, *regularization*, is used both for the preprocessing of the infrared images and for the post-processing of the disparity image sequences. For this reason, the treatment of both pre- and post-processing techniques is combined in this chapter.

Radiometric calibration, briefly introduced in section 2.2.3, is used to correct for the non-uniform response of infrared detector chips. This pre-processing step, described in section 5.1, ensures that corresponding structures in two images of a stereo pair have similar grey-value distributions, a prerequisite for correlation-based disparity estimation.

A simple technique for identifying outliers in measured data is presented in section 5.2. This method is used in a post-processing step to remove erroneous matches from estimated disparity maps.

Missing data is interpolated using a so-called *membrane model*, discussed in section 5.3. This regularization method finds a data estimate that is controlled by a smoothness constraint and the available measurements. The theoretical background of the membrane model is presented in section 5.3.1.

Section 5.3.2 explains how the membrane model can be applied to infrared images, where it is used to fill in missing pixels, resulting from defective sensor elements.

Section 5.3.3 discusses how the disparity estimates obtained with the stereo matching algorithm described in chapter 4 can be regularized. For this particular application, an extension of the membrane model to image sequences rather than individual frames is presented. With this extension, temporal smoothness can be enforced during the regularization process and the convergence of the iterative interpolation method is improved.

5.1 Non-Uniformity Correction and Radiometric Calibration

As mentioned in section 2.2.3, the detector response of an infrared focal plane array generally varies from sensor element to sensor element. As a result of this, the grey-value images obtained with an infrared camera may have a non-uniform grey-value distribution when a blackbody source with a spatially homogeneous temperature distribution is imaged.

Sample images illustrating such a non-uniform sensor response are given in figure 7.1, page 114, and figure 7.2, page 115. The latter image shows that the spatial grey-value variations due to non-uniformity can be much larger than those attributable to the variation of the actual signal. Therefore, the non-uniformity has to be corrected for in order to obtain images that are suited for visual interpretation or correlation-based stereo matching.

Radiometric calibration, introduced in section 2.2.3, establishes a relationship between the detector output and the incident irradiance. This relationship can be used for a non-uniformity correction by changing the grey-values in such a way that the new grey-value of each pixel is proportional to the equivalent blackbody temperature.

Polynomial Fit For the radiometric calibration, it is assumed that n images of a blackbody at different temperatures T_1, \dots, T_n are available, as described in section 2.2.3. To obtain a functional relationship between temperature and detector output, the dependence of the grey-value output on the temperature is modeled as a polynomial for each pixel. The following discussion is for a single pixel.

The relationship between the detector output, expressed as grey-value G , and blackbody temperature T is modeled as

$$T(G) = \sum_{i=0}^m a_i G^i, \quad (5.1)$$

where the polynomial is chosen as third order ($m=3$), as suggested in [35]. The coefficients a_i are determined by a linear regression that minimizes the sum of the squared distances between equation (5.1) evaluated at the grey-values G_1, \dots, G_n of the calibration points, and the blackbody temperatures T_1, \dots, T_n at which these calibration points were obtained. Once the coefficients have been determined for every sensor pixel, the non-uniformity of the camera images can be corrected for by replacing the original grey-value with a grey-value that is proportional to the apparent temperature obtained using equation (5.1). In this work, the image intensities were represented as floating point values, such that the value of the apparent temperature is used directly as the corrected grey-value.

Identifying the Defective Pixels On a typical infrared detector chip, there are usually some “dead” or “bad” pixels, that either do not function at all, or that have a very noisy output signal.

In the present work, the parameters and the quality of the radiometric fit are used as a criterion to identify and label these defective pixels.

For sensor elements that produce a noisy or random output, it is likely that the deviation of the fit curve from the calibration points is larger than for a regular pixel. Pixels that exceed a given threshold of the mean squared distance between the calibration points

$$\frac{1}{n} \sum_{i=1}^n (T(G_i) - T_i)^2 > D_{\text{threshold}}, \quad (5.2)$$

were labeled as bad pixels, see section 7.1.

Some defective sensor elements produce a constant grey-value output. These pixels can be identified by requiring a minimum slope of the fit curve.

The positions of identified defective pixels are stored using a binary mask of the same size as the original image; good pixels are represented by a one and defective pixels are represented by a zero.

Adjustment of Image Intensities The correlation score employed by the stereo matching algorithm used in this work, equation (4.8), is not normalized with respect to the grey-value of the two stereo images. Therefore, it is important that corresponding patches in the two images have a similar grey-value distribution. Ideally, if the water surface was a perfect blackbody, corresponding points in the two images would have the same grey-value after radiometric calibration. In practice, this is not always the case. As the water surface is not a perfect blackbody and the spectral responses of the two cameras used for this work differ slightly, the apparent absolute temperatures of the water surface obtained with the two cameras may differ slightly as well. However, the variation of the apparent temperature at the surface (typically several 0.1 K) should be similar for both cameras.

To obtain similar image intensities for corresponding points in both images, the mean of the apparent temperature over all pixels has been calculated for both images and subtracted from the grey-values after radiometric calibration.

5.2 Outlier Removal

Despite the validation of disparity estimates, described in section 4.3.5, wrong matches occasionally occur in the disparity maps. Often, these wrong matches have a disparity that differs significantly from the mean of the correct matches.

Using the simplistic assumption that the correct matches are normally distributed, outliers can be identified as follows:

The mean disparity $\langle \mathcal{D} \rangle$ over the whole image or a subregion Ω is calculated as

$$\langle \mathcal{D} \rangle_{\Omega} = \frac{1}{n} \sum_{\Omega} \mathcal{D}(x, y), \quad (5.3)$$

where the sum runs over all n pixels in Ω . The standard deviation σ_{Ω} is defined as

$$\sigma_{\Omega}^2 = \frac{1}{n} \sum_{\Omega} (\mathcal{D}(x, y) - \langle \mathcal{D} \rangle)^2. \quad (5.4)$$

Outliers can be removed by setting a threshold on the maximum distance, in numbers of σ_{Ω} , a disparity value may have. In this work, the threshold distance was empirically set to $2.3 \times \sigma_{\Omega}$ and the subregions Ω were chosen as the lines of the disparity images.

5.3 Regularization: Filling in the Gaps

When working with image data, a common problem is that the data contains gaps. Examples for such gaps are defective pixels in infrared images or invalid matches in disparity estimates. If the measured quantity is not completely random but has some spatial or temporal coherence, it is possible to fill in the gaps by interpolating between neighboring measurements.

This section presents a regularization technique that can perform such an interpolation. The technique is physically motivated by the elastic properties of a membrane and leads to a partial differential equation. The theory of regularization is described in section 5.3.1, while sections 5.3.2 and 5.3.3 deal with the application of this technique to the above mentioned problems of filling in gaps in the infrared images and in the disparity estimates, respectively.

5.3.1 Theory

This section briefly introduces regularization using the *membrane model*, in a similar manner as in Spies [93]. A more extensive treatment of regularization is given by Tschumperle [99].

Problem Statement Assume that the image $\mathcal{D}(x, y)$ contains measured data, and that the confidence in the measurements is given as $\kappa(x, y)$, where higher values of κ represent a higher level of confidence. Image regions where measured data is missing completely have a value of $\kappa = 0$.

The goal is to find a differentiable regularized estimate $\mathcal{E}(x, y)$ that varies smoothly everywhere and stays close to the available measurements. The required smoothness of \mathcal{E} ensures that regions with missing data are interpolated and that noise is dampened.

Membrane Model If the image \mathcal{D} is thought of as a height field, with grey-values representing the height of each pixel, the regularized estimate $\mathcal{E}(x, y)$ can be pictured as the height field formed by a membrane which is “pulled” towards the data points. A higher confidence value results in a stronger pull towards the corresponding data point. Because of the elastic forces within the membrane, it cannot be stretched arbitrarily; that is, it cannot follow each pull. Therefore, its shape represents a balance between the competing demands of staying close to the data and smooth variation.

Mathematically, this membrane model is expressed by an energy functional that has contributions from the elastic energy and from the “pulling” forces from the measured data. The regularized estimate \mathcal{E} is given by the membrane shape that minimizes the energy functional

$$\int_{\text{Image}} \underbrace{\kappa(x, y) (\mathcal{E}(x, y) - \mathcal{D}(x, y))}_{\text{data term}} + \underbrace{\alpha \left((\partial_x \mathcal{E})^2 + (\partial_y \mathcal{E})^2 \right)}_{\text{smoothness term}} dx dy. \quad (5.5)$$

$\mathcal{L}(\mathcal{E}, \partial_x \mathcal{E}, \partial_y \mathcal{E})$

In this functional, the first term accounts for the “pull” by the available data and the second term enforces smoothness by punishing large derivatives. The balance between these terms is regulated by the ratio of the confidence measure κ and the the elasticity α , which controls the smoothness.

From the calculus of variations, it is well known that the solution that minimizes the integral (5.5) satisfies the *Euler-Lagrange* equation

$$\frac{\partial \mathcal{L}}{\partial \mathcal{E}} - \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial (\partial_x \mathcal{E})} - \frac{d}{dy} \frac{\partial \mathcal{L}}{\partial (\partial_y \mathcal{E})} = 0, \quad (5.6)$$

which leads to the following partial differential equation

$$\kappa \mathcal{E} - \kappa \mathcal{D} + \alpha \Delta \mathcal{E} = 0. \quad (5.7)$$

Here, $\Delta = \partial_x^2 + \partial_y^2$ is the well known *Laplace* operator.

Discretization For a discretization of equation (5.7) it is useful to note that a discrete Laplace operator can be approximated by a local mean, for example by a binomial filter mask, minus the central value: $\Delta\mathcal{E} = \langle\mathcal{E}\rangle - \mathcal{E}$ (for an explanation, see for example Jähne [56]). Substituting this approximation into (5.7) yields

$$(\kappa + \alpha)\mathcal{E} = \alpha\langle\mathcal{E}\rangle + \kappa\mathcal{D}, \quad (5.8)$$

which leads to an iterative solution of equation (5.7) using the following update rule:

$$\mathcal{E}^{n+1} = \frac{\alpha}{\kappa + \alpha}\langle\mathcal{E}\rangle^n + \frac{\kappa\mathcal{D}}{\kappa + \alpha}. \quad (5.9)$$

5.3.2 Filling in Defective Pixels

As a result of defective sensor elements, image data collected by infrared cameras contains gaps, which must be filled in. The membrane model can be used to fill in these holes in a pre-processing step.

In the present application, the grey-values of the dead pixels are completely unknown and therefore have a confidence value of zero. In contrast, the grey-values of the good pixels are known with very high confidence, as ideally they should be only affected by noise. Therefore, the good pixels should have a value for κ that is much larger than α . In other words, the grey-values of the good pixels remain fixed. This leads to a simplification of the iterative update rule (5.9)

$$\begin{aligned} \mathcal{E}^{n+1} &= \langle\mathcal{E}\rangle^n \text{ for "dead pixels" } (\kappa = 0) \\ \mathcal{E}^{n+1} &= \mathcal{E}^n \text{ for "good pixels" (large } \kappa), \end{aligned} \quad (5.10)$$

suggesting a two-step iterative procedure:

1. Smooth the whole image, using a filter that calculates a local mean, such as a Gaussian or binomial filter.
2. Replace the grey-values of the "good pixels" with their original grey-value.

Because the holes due to defective pixels are not very large, convergence is typically reached after a few iterations.

5.3.3 Regularization of Disparity Estimates

The membrane model, equation (5.5), also lends itself to the regularization of disparity estimates. Holes in disparity estimates can occur for a variety of reasons, such as occlusions, lack of texture or noisy images.

There are two for the confidence measure κ that regulates the influence of the data term; either a confidence measure obtained from the disparity matching algorithm, as mentioned in section 4.3.5, or a binary confidence measure, setting κ to one for disparity estimates satisfying the consistency check between left and right image (see section 4.3.5) and setting κ to zero for estimates that fail the consistency check.

The regularization of the disparity data can be performed on a frame-by-frame basis using the membrane model. However, this method has several drawbacks:

First, if the disparity data contains large holes, as is the case for some frames in the current work, the iterative solution for the membrane model as in equation (5.9) does not have good convergence properties. The iterative approach is essentially a “diffusion” of disparity data with high confidence into regions with low confidence. The problem is, that diffusion is a relatively slow process and therefore large holes take many iterations to be closed.

Second, the temporal change of the water surface is not abrupt but rather gradual, and this should be taken into account by the regularization procedure. Therefore, a temporal smoothness constraint should be included in addition to the spatial smoothness term in equation (5.5).

Temporal Smoothness Constraint To see how the wave motion constrains the temporal change of disparity, the height of a simple sinusoidal wave

$$h(t, \mathbf{x}) = h_0 e^{i(\omega t - \mathbf{k}\mathbf{x})} \quad (5.11)$$

can be examined, where h_0 is the amplitude of the wave, ω is the frequency and \mathbf{k} is the wave vector. Taking the temporal derivative leads to

$$\begin{aligned} \frac{\partial h}{\partial t} &= i h_0 \omega e^{i(\omega t - \mathbf{k}\mathbf{x})} \\ &= i h_0 k c_k e^{i(\omega t - \mathbf{k}\mathbf{x})}, \end{aligned} \quad (5.12)$$

with the wavenumber k and the propagation velocity c_k . The dependence of c_k on k is given by the so-called dispersion relation. For water waves, the dispersion relation can be found, for example, in Kinsman [63].

If the time between consecutive image frames is T_{frame} , equation (5.12) gives an upper bound for the difference between frames: the maximum change in amplitude is $h_0 k c_k T_{\text{frame}}$. The exact value of this upper bound depends on the wavenumbers present in the wavefield.

Disparity Regularization of Wave Image Sequences The upper bound on the temporal derivative of the wave height is, in effect, a physically-based temporal smoothness constraint, which also limits the temporal change of stereo disparity.

This temporal smoothness constraint can be taken into account by extending the membrane model of equation (5.5) in the time domain and applying it to a whole disparity image sequence rather than to individual frames. The temporal smoothness constraint is integrated into the model in the same way as the spatial smoothness constraint, leading to the following functional to be minimized:

$$\int_{\text{Image sequence}} \underbrace{\kappa (\mathcal{E}(x, y, t) - \mathcal{D}(x, y, t)) + \alpha \underbrace{((\partial_x \mathcal{E})^2 + (\partial_y \mathcal{E})^2)}_{\text{spatial smoothness}} + \underbrace{\alpha_t (\partial_t \mathcal{E})^2}_{\text{temporal smoothness}}}_{\mathcal{L}(\mathcal{E}, \partial_x \mathcal{E}, \partial_y \mathcal{E}, \partial_t \mathcal{E})} dx dy dt. \quad (5.13)$$

Analogous to equations (5.6) and (5.7), this minimization problem leads to the following partial differential equation:

$$\kappa \mathcal{E} - \kappa \mathcal{D} + \alpha \Delta_{xy} \mathcal{E} + \alpha_t \partial_t^2 \mathcal{E} = 0 \quad (5.14)$$

This equation can be further simplified if one assumes that the factors controlling the spatial and temporal smoothness have the same value, that is $\alpha = \alpha_t$. By the same reasoning as in section 5.3.1, the regularized solution $\mathcal{E}(x, y, t)$ can be found using the same iterative update rule, equation (5.9). The only difference to the 2D case is, that the local average is taken over a 3D neighborhood, with the third dimension being time.

This iterative approach can also be used for $\alpha \neq \alpha_t$ if the filter tap that is used to calculate the local average is adjusted appropriately.

In the present application, using a 3D membrane model also helps to overcome the above mentioned problem of slow convergence when large holes are present in the disparity data. The reason for the improved convergence is that holes which are large in the x- and y- direction are not necessarily large in the temporal direction as well. Therefore, the diffusion process can quickly fill in such holes with disparity information from previous and subsequent frames.

6 Experimental Setup and Procedures

This chapter gives a detailed description of the instruments used and the procedures followed for the experiments with the infrared stereo camera system at the Heidelberg Aeolotron facility.

The infrared cameras that were used in this work are presented in section 6.1, and their integration into a stereo system is the topic of section 6.1.2. Before each deployment of the system, image sequences to be used for radiometric calibration were obtained with a laboratory blackbody, which is described in section 6.2.

Section 6.3 introduces the different components of the acquisition and data storage system. This includes subsection 6.3.2 which deals with the synchronization of the two cameras.

Geometric calibration of the cameras was performed using a calibration target, presented in section 6.4, that was designed to have high-contrast feature points in the infrared region.

An overview of Heidelberg wind-wave facility *Aeolotron*, at which the experiments were performed, can be found in section 6.5.

Finally, section 6.6 gives an account of the procedures that were followed to perform the experiments.

6.1 Infrared Cameras

For a stereo system, the use of two identical cameras is desirable. However, due to the high cost of infrared imagers, a stereo system with two different cameras (which were already on hand) was built. The cameras used were one *Thermosensorik CMT 384* and one *Raytheon Amber Radiance*.

6.1.1 Specifications

Table 6.1 summarizes the manufacturers' specifications for these two cameras. Although the two cameras are based on different detector technology, Indium Antimonide for the Amber Radiance camera and Cadmium Mercury Telluride for the Thermosensorik camera, they are both sensitive in approximately the same wavelength range ($3 - 5\mu\text{m}$). They differ slightly in terms of their $\text{NE}\Delta\text{T}$, number of pixels and maximum frame rate. Both cameras were equipped with lenses of 50 mm focal length. Despite identical focal lengths, owing to the different aperture



Factory specification	Amber Radiance	TS CMT384
Detector type	Indium Antimonide	Cadmium Mercury Telluride
Spectral bandpass	3-5 μm	3.5-5 μm
Noise equivalent ΔT	≤ 25 mK at 300K	≤ 20 mK at 293K
Maximum frame rate	60 Hz	150 Hz at 1ms int. time
Number of Pixels	256x256	384x288
Pixel size	unknown	20x20 μm^2
Pixel pitch	38x38 μm^2	24x24 μm^2
Dynamic range bits/pixel	12	14
Dimensions WxHxD	112x183x262 mm ³	130x150x260 mm ³
Focal length	50 mm	50 mm
Field-of-view	11.1° vert./horiz.	8.8° horiz., 6.0° vert.
Non-uniformity correction	on-board, two-point	not included
Dead pixel correction	nearest neighbor	not included

Table 6.1: Manufacturers' specifications of the Raytheon Amber Radiance and Thermosensorik CMT 384 infrared cameras.

and detector sizes of the two cameras their *field-of-view*, that is the maximum horizontal and vertical angular extent viewed, was different. Another consequence of the difference in optics is a marked variation in *depth-of-field*, that is the distance range in which objects are well focused ([see 57, sec. 4.6.2] for details).

The Amber Radiance camera is equipped with an on-board non-uniformity correction. This non-uniformity correction is based on a radiometric calibration using a linear model, as described in 5.1. Two reference temperatures are created internally by a small plate that is heated or cooled and placed in front of the detector chip. The Amber Radiance camera automatically fills in the defective pixels with the grey-value of a neighboring pixel.

The model of the Thermosensorik camera used in this work is not equipped with an on-board non-uniformity correction.

6.1.2 Stereo Setup

To create a stereo setup, the two cameras were mounted together on a solid aluminum base plate with a thickness of 5 mm as shown in figure 6.1. The horizontal baseline of this system is about 13 cm; as the position of the optical center with respect to the camera housing is not known exactly, its value is later determined through camera calibration (see 7.2.2). Each camera was firmly attached to the plate with two screws.

The *verging angle* β of the cameras was varied by sliding one of the attachment points along a slit holding one of the screws as depicted in figure 6.2. Before deployment and geometric calibration (see section 3.4.1), the verging angle was adjusted to maximize the overlap of the two fields of view at the anticipated range, which was typically around 120 cm in these experiments. This maximization was done by visual inspection, using the PCB checkerboard pattern described in section 6.4 as a target. This adjustment of verging angle is sometimes referred to as *boresight alignment* ([see 50]). The fact that the optical axes of the two cameras are not parallel for a non-zero verging angle does not pose a problem for the stereo matching because the stereo images are rectified as described in section 3.5.3.

During experimental operation, the main disadvantage of not having identical cameras was the difference in optics, specifically the difference in depth-of-field. When the Amber Radiance camera was focused on a distance of 90 cm, everything that was further away was also in focus (focus at ∞). The Thermosensorik camera, owing to its larger aperture, has a smaller depth of field, and when focused at the same distance of 90 cm only objects in the range of 85 – 110 cm were well focused.



Figure 6.1: Infrared stereo camera setup: front (left) and top (right) views.

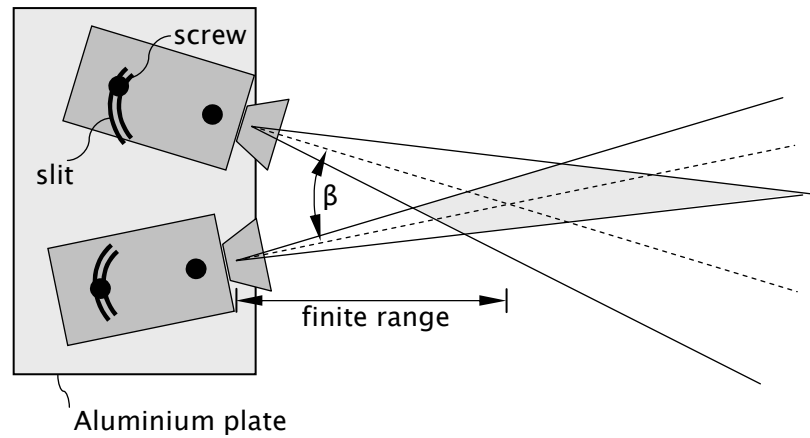


Figure 6.2: Stereo camera setup. The verging angle β can be adjusted to maximize the overlap (shaded area) of the fields of view for the anticipated distance to the water surface. This is done by sliding the attachment screws along a slit in the aluminum base plate.

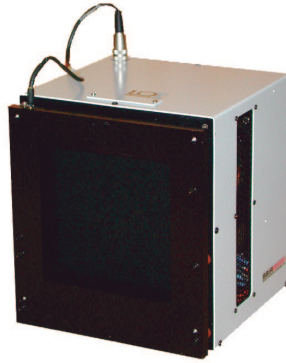
6.2 Blackbody

For the radiometric calibration (see sections 2.2.3 and 5.1), a laboratory blackbody, that is a thermal emitter that comes close to being an ideal blackbody (see section 2.1.2), manufactured by *Santa Barbara Infrared Inc.* was used. This laboratory blackbody has a highly emissive (emissivity 0.985 ± 0.014) surface whose temperature can be controlled over the temperature range $10 - 60^\circ\text{C}$. The temperature accuracy is the larger of $2.5\text{ mK} \times (T - 25^\circ\text{C})$ and 25 mK . For the experiments, radiometric calibration images were obtained over a temperature range from $19 - 25^\circ\text{C}$, in which the absolute temperature accuracy of the blackbody is 25 mK . The specifications of the blackbody are listed in table 6.2, while results of the radiometric calibration are presented in section 7.1.

6.3 Acquisition System

6.3.1 Frame Grabber

For image acquisition, two frame grabbers of the type *microEnable 2*, produced by *Silicon Software GmbH*, were used. The *microEnable 2* frame grabber, shown in figure 6.3, is based on a modular design with a programmable logic chip (FPGA)



Parameter	Factory specification
Temperature Range	10-60°C, absolute
Emissivity	0.985 ± 0.014 from 2 – 14 μm
Total system uncertainty	$\pm \text{max of } 2.5 \text{ mK} \cdot (T - 25^\circ\text{C}), 25 \text{ mK}$

Table 6.2: Santa Barbara Infrared Series 2000 Blackbody specifications.

at its core. Together with interchangeable camera port modules, the FPGA makes it possible to tailor the frame grabber to specific applications and allows some simple on-board image processing. The frame grabbers possess an additional I/O port module, which in the present application was used to route trigger signals for camera synchronization (see below). The image data is transferred to the RAM of the computer with direct memory access (DMA) via a PCI bus interface.

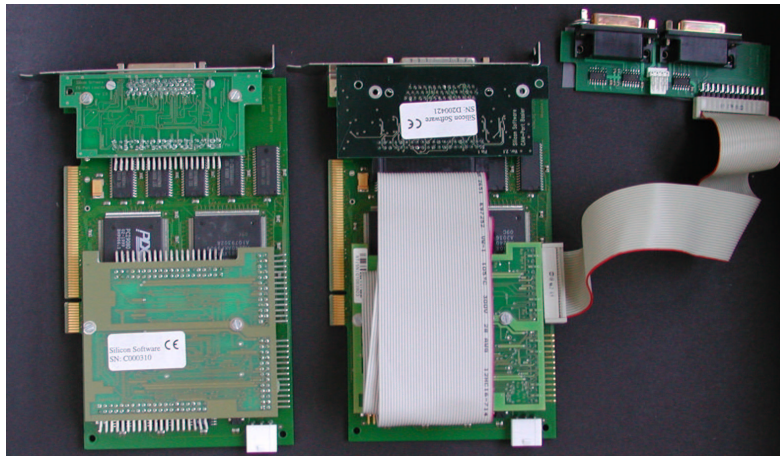


Figure 6.3: Silicon Software microEnable 2 frame grabbers: the right frame grabber is shown with the I/O port module at the upper right, which was used to route the trigger signal from one frame grabber to the other.

6.3.2 Camera Synchronization

When collecting stereo image sequences of moving objects, it is important that the two cameras be synchronized, because otherwise the disparity between the two views can be partly due to motion in the scene. In contrast to the Thermosensorik CMT camera, the Amber Radiance camera cannot be triggered externally (see section 6.1). Therefore, in these experiments, the trigger signal for the Thermosensorik CMT had to be derived from the *frame start* signal of the free-running Amber Radiance camera. After exposure of the CCD-chip, the camera sends the frame start signal, to signal the start of the image data transmission to the attached frame grabber. This frame grabber was programmed to generate a trigger signal from the frame start signal. Via an external cable, the generated trigger signal was routed to the second frame grabber, which then triggered the Thermosensorik camera. A schematic timing diagram of this setup is depicted in figure 6.4.

The timing diagram shows that there is a time lag on the order of the exposure (or integration) time T_{int} of the Amber Radiance camera (1.8ms during the experiments) before the start of exposure of the second camera. To work around this problem, attempts were made to use the frame grabber to delay the generated trigger signal by $T_{\text{frame}} - T_{\text{int}}$ where T_{frame} is the time between frames. This way the cameras would be exposed simultaneously, with their frame numbers differing by one. However, during the experiments it was found that the programmable delay did not operate as claimed by the manufacturer. Therefore the two cameras were exposed with a small time lag. For a discussion of the potential systematic error introduced by this delay, refer to the discussion in section 7.8.1.

6.3.3 PC and RAID

Both frame grabbers were mounted in a standard PC equipped with a 433 MHz Pentium II processor, 1 GB of RAM and a 270 GB striping RAID hard disk array. The operating system used was Windows NT 4.0.

For synchronous acquisition with both cameras recording at 60 frames per second, a data rate of

$$2 \left[\frac{\text{Byte}}{\text{pixel}} \right] \times \left(\underbrace{256 \cdot 256}_{\text{Amber Rad.}} + \underbrace{288 \cdot 384}_{\text{TS CMT}} \right) \left[\frac{\text{pixel}}{\text{frame}} \right] \times 60 \left[\frac{\text{frames}}{\text{s}} \right] = 20.60 \left[\frac{\text{MByte}}{\text{s}} \right] \quad (6.1)$$

must be handled. The RAID array has a nominal data rate of 50 MB/s, fast enough to write this data rate onto the hard disk in real time. Nevertheless, an occasional loss of frames cannot be completely ruled out, especially during

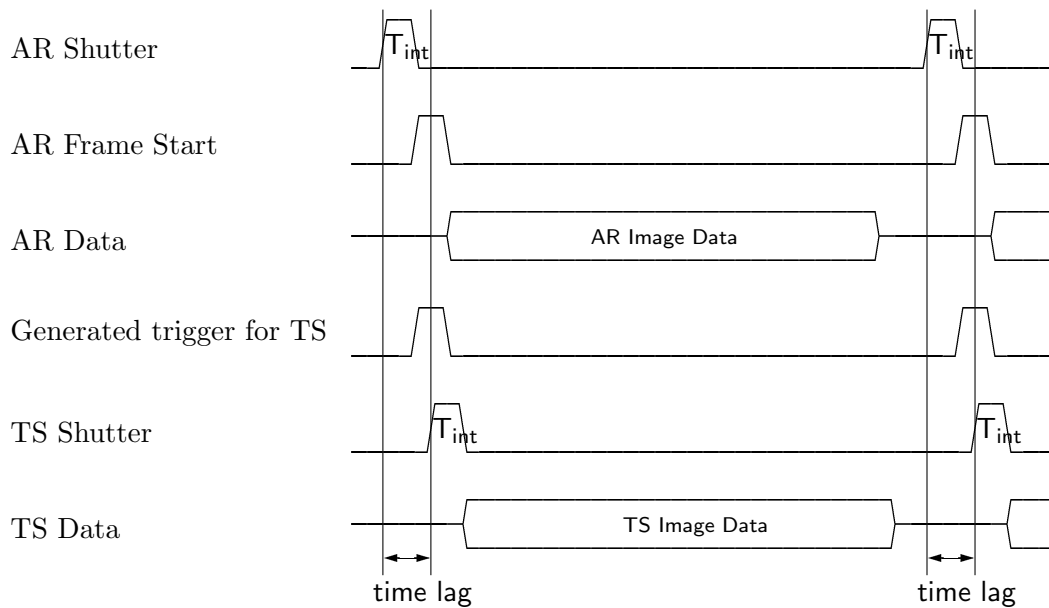


Figure 6.4: Timing diagram: The Amber Radiance camera (AR) exposes the detector chip for a time span T_{int} , after which it signals the start of a new frame to the frame grabber (AR Frame Start) and begins transmission of data. After the rising edge of AR Frame Start the frame grabber generates a trigger signal for the Thermosensorik camera which then opens its shutter for T_{int} . Note that the time lag between the exposure of the two images is on the order of T_{int} . The diagram schematically shows the sequence of events and is not to scale.

intermittent claims of system resources by other tasks running on the system. The frame grabber embeds a sequential frame number into the image data stream, so that a loss of frames can easily be detected by searching for consecutive frame numbers that differ by more than one.

However, stereo image sequences of up to approximately 30 seconds can instead be stored directly in the RAM of the PC, and then subsequently be written onto the hard disk. When doing so, care has to be taken that the RAM buffer is allocated in such a way that the operating system does not try to swap the contents of the buffer onto the hard drive.¹ This “grab-to-RAM” approach was used for the experiments, and no loss of frames occurred.

¹For Windows NT 4.0 this can be done by allocating the buffer with `VirtualAlloc()` and locking in it in physical memory with `VirtualLock()`.

6.4 Geometric Calibration Target

For the geometric camera calibration process described in section 3.4.1, a planar calibration target with precisely located and easily identifiable feature points is required. As suggested by Bouguet [10], a checkerboard pattern was used for the present work.

To facilitate the extraction of the corners from the calibration images, it is desirable to have a high contrast between the light and dark checkers. When imaging in the infrared region, this can be accomplished by using checkers made of two materials with different emissivity. Printed circuit boards (PCB) are well suited for this; the copper has a high reflectance (low emittance), whereas the base material has a high emittance (see figure 6.5). When such a target is warmed, for example with a heat gun, the base material checkers appear as bright squares in a thermal image because they are greybodies. In contrast, the copper checkers look darker, reflecting the cooler ambient temperatures like a mirror. Using PCBs has the benefit that calibration targets can quickly be produced with high accuracy (about $10\ \mu\text{m}$) and low costs by etching or milling.

For this application several targets with different checker dimensions were milled by Martin Vogel of the *Physiologisches Institut Heidelberg*. To prevent non-planar distortions due to bending or twisting, the boards were mounted onto sturdy flat plexiglass slabs. For the experiments a target with a checker size of $22 \times 22\ \text{mm}^2$ was used.

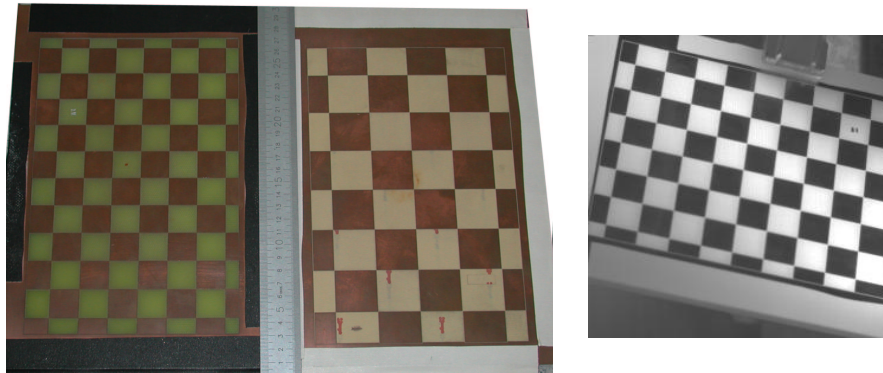


Figure 6.5: Left: Two camera calibration targets with checker size $22 \times 22\ \text{mm}^2$ and $32 \times 32\ \text{mm}^2$ respectively. The picture was taken at visible wavelengths. Right: The highly reflective copper and the highly emissive base material of the printed circuit board make for a good contrast in the infrared. This image was acquired with the Amber Radiance camera after heating the target with a heat gun.

6.5 Aeolotron

The experiments were performed at the wind-wave interaction facility *Aeolotron*¹, located at the Institute of Environmental Physics at the University of Heidelberg. The Aeolotron is designed to study air-sea gas exchange in a controlled environment. It is a gas-tight annular channel, the lower part of which is filled with water and the upper part of which is filled with air (see Figure 6.6).

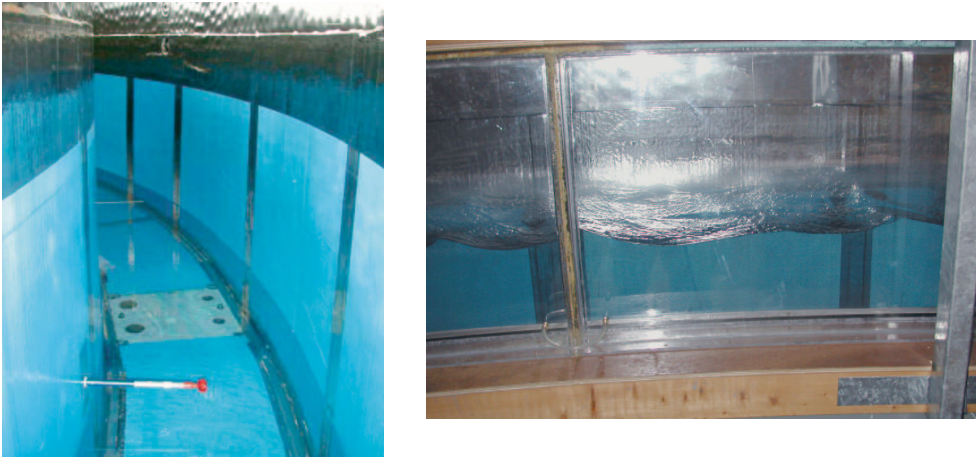


Figure 6.6: *Left: cross-sectional view of the wind wave channel. The aluminum coated foil reduces thermal emission from the walls (photo from [54]). Right: side view through the “panoramic window” into the channel with wind waves.*

Winds with speeds of up to 15 ms^{-1} can be produced by a ring of paddles circulating above the water surface. The wind stress on the water surface generates wind waves. The circular geometry gives a distinct advantage over linear channels, because it allows for a large fetch; that is, the wind can act on the water surface for a long time. Therefore, the generated wave field develops a stationary wavelength distribution after some time period of continuous wind, similar to the situation found on oceans. In contrast, in linear wind wave facilities the distribution of the observed wavelengths varies along the length of the channel. The channel has a mean circumference of 29.2 m, a width of 0.61 m and a height of 2.4 m. The water depth at the time of experiment was 0.90 m. The channel is gas-tight, permitting the study of air-sea gas exchange by mass-balance methods.

The air conditioning system provides independent control over humidity and air temperature. By warming or cooling, positive and negative heat fluxes can be

¹Named after *Aeolus*, the Greek god of the winds.

created on the water surface. Changing the humidity affects the latent heat flux due to evaporation.

The interior walls of the channel are covered with a gas-tight, aluminum coated plastic foil. This coating has a high reflectivity and helps to reduce the thermal emission from the walls, thereby reducing the effect of the environment on the temperature measurements of the water surface.

A more detailed description of the Aeolotron is given by Jähne [54, 58].

6.6 Experimental Procedure

In this section, the experimental procedures followed and the instrument settings used during the experiments are detailed. Figure 6.7 illustrates the different stages of the measurement procedure.

6.6.1 Radiometric Calibration Procedure

First, a set of image sequences to be used for radiometric calibration was acquired. This was done by placing each camera in front of the blackbody as shown in figure 6.7a. The blackbody temperature was varied over the temperature range 19 – 25°C in steps of 0.2 K. At each temperature, an image sequence of 32 frames was acquired.

The integration times were set to 1.8 ms for the Amber Radiance camera and 1.3 ms for the Thermosensorik camera; these integration times were also used for the subsequent acquisition of images of the water surface.

6.6.2 Geometric Calibration Procedure

Second, a series of stereo image pairs of the checkerboard calibration target was acquired for geometric calibration, as explained in section 3.4.1. For each image pair, the orientation of the checkerboard with respect to the camera was changed. To enhance the contrast in the infrared images, the target was warmed up using a heat gun (see figure 6.7b) before acquiring the images.

The distance between the cameras and the calibration target was chosen to lie within ± 10 cm of the anticipated distance between the cameras and the water surface, which was around around 125 cm in these experiments (see figure 6.7c).

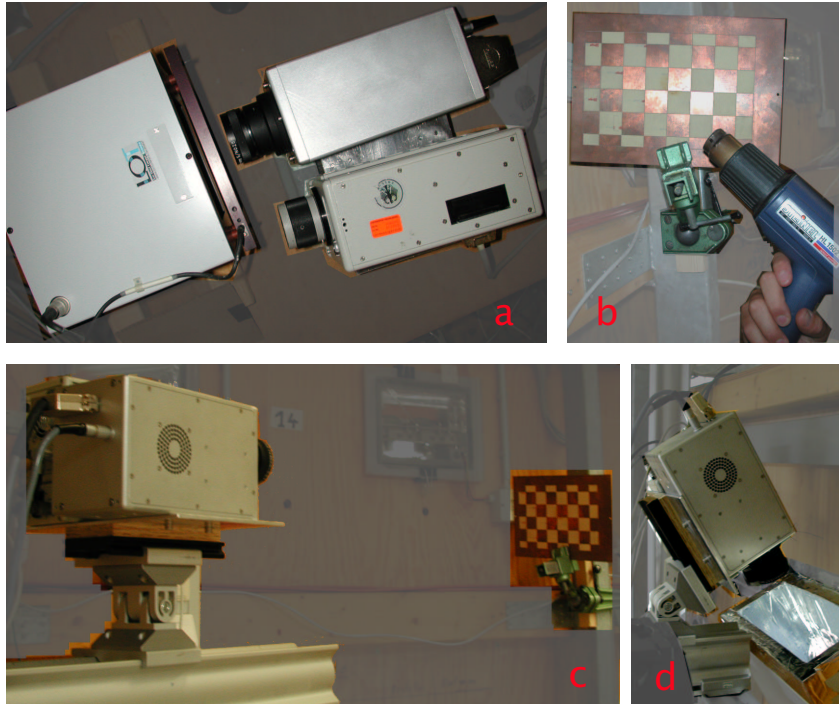


Figure 6.7: *Experimental procedures: a) Acquisition of image sequences for radiometric calibration. b) Warming the calibration target with a heat gun to enhance image contrast. c) Setup during acquisition of image sequences for geometric camera calibration. d) Mounted stereo system aimed at the water surface through an opening in the wind wave channel. Note that the calibration target shown in b) and c) differs from the one that was actually used for the experiments.*

6.6.3 Deployment

After the acquisition of image sequences for radiometric and geometric calibration, the stereo system was mounted for deployment without changing the relative orientation of the cameras or the focus setting. As shown in figure 6.7d, the system was aimed at the water surface at a slightly oblique angle, through an opening in the Aeolotron.

Using a folding rule, the distance of the cameras from the calm water surface was determined to be 125 ± 15 cm (The variation in the distance results from the oblique view angle).

6.6.4 Acquisition of Image Sequences

For different wind speeds, image sequences of the water surface were recorded. Each sequence consists of 512 frames with a frame rate of 60 frames/second. To create a heat flux, cool dry air was let into the Aeolotron using the air-conditioning system.

Before wind waves were created using the paddle ring, an image sequence of a flat water surface was recorded; the planar surface was used for assessing the accuracy of the stereo system (see section 7.8.1). Care was taken that the surface was truly calm by waiting for approximately two hours after filling the wave flume with water before the image sequences were recorded. Some frames of this flat water sequence are depicted in figure 1.3 on page 12. Because no heat flux was present at the beginning of this sequence, the first 100 frames of the sequence were not usable for disparity matching.

The sequences acquired are listed in Table 6.3. The numbering reflects the order in which the sequences were acquired and is used in section 7.7 for the presentation of the results. The emphasis during the experiments was on obtaining sequences with different wave heights, which was achieved by varying the wind speed. The listed wind speeds are rough estimates, and are only given to allow qualitative comparisons. The bulk water temperature was in the range 19.5°C to 21°C over the course of the experiments.

Sequence Number	Wind Speed
0 (Flat water)	no wind
1	medium (4-6 m/s)
2	low (2-3 m/s)
3	high (7-8 m/s)
4	low (2-3 m/s)

Table 6.3: List of the infrared image sequences of water waves acquired.

7 Results

This chapter presents the results obtained with the stereo infrared camera system described in chapter 6 and the image processing algorithms described in chapters 3 to 5.

Sections 7.1 to 7.6 document the results obtained for the individual processing steps: radiometric and geometric calibration, rectification, disparity matching, regularization and depth reconstruction.

The results obtained for the actual application, namely the reconstruction of the water surface, are presented in section 7.7. An experimental assessment of the accuracy of the surface reconstruction is given in section 7.8.

7.1 Radiometric Calibration and Non-Uniformity Correction

Radiometric calibration and non-uniformity correction were performed for both infrared cameras. At temperatures in the range of $19 - 25^{\circ}\text{C}$, image sequences of the Santa Barbara Infrared laboratory blackbody were acquired as documented in section 6.6.1. To reduce the influence of image noise, the mean over 32 frames was calculated for each pixel at each temperature step.

Using the mean images, a 3rd order polynomial fit as described in section 5.1 was performed.

Pixels for which the squared standard deviation of the calibration points from the fit curve (see equation (5.2) on page 95) exceeded an empirically chosen threshold of $5 \cdot 10^{-4} \text{ K}^2$ were marked as defective. This threshold corresponds to a standard deviation of $\sqrt{32} \cdot 5 \cdot 10^{-4} \text{ K} = 0.13 \text{ K}$ from the fit curve for a single frame. The factor of $\sqrt{32}$ accounts for the fact that the calibration points were not obtained from single image frames but from mean values calculated over sequences comprising 32 frames.

Pixels whose grey-value remained constant independent of temperature were also marked as defective.

The calibration curves thus obtained were used to apply a non-uniformity correction to the infrared image sequences by replacing the raw grey-value of each image pixel with the corresponding apparent blackbody temperature. Gaps in the images due to defective pixels were interpolated as described in section 5.3.2.

7.1.1 Thermosensorik CMT Camera

Figure 7.1 (left) shows one of the raw grey-value images collected with the Thermosensorik CMT camera to illustrate the need for a non-uniformity correction. This image was acquired with the camera imaging the surface of the Santa Barbara Infrared laboratory blackbody (see section 6.2) at a temperature of 24°C. Clearly, the uniform temperature distribution of the blackbody does not result in a uniform grey-value distribution in the raw image.

For several selected pixels, marked with red arrows in figure 7.1, the calibration points and the resulting fit curves describing the relationship between blackbody temperature and grey-value output are shown in figure 7.3. The calibration images were acquired with an integration time of 1.3 ms. Note that the grey-value level of the pixel at position (180,246) is lower than the grey-values of the other selected pixels by one order of magnitude. Also, the slope of the calibration curve is lower, resulting in a lower temperature resolution. This particular pixel was labeled as a defective pixel based on the deviation of the calibration points from the fit curve.

The pixels of the Thermosensorik camera that were labeled as defective are displayed as a binary mask in figure 7.1 on the right. The number of bad pixels was 557, that is 0.5% of the total number of pixels.

Figure 7.2 shows the results of the non-uniformity correction for an image of a water surface. The defective pixels have been filled in using the membrane model (see section 5.3.2).

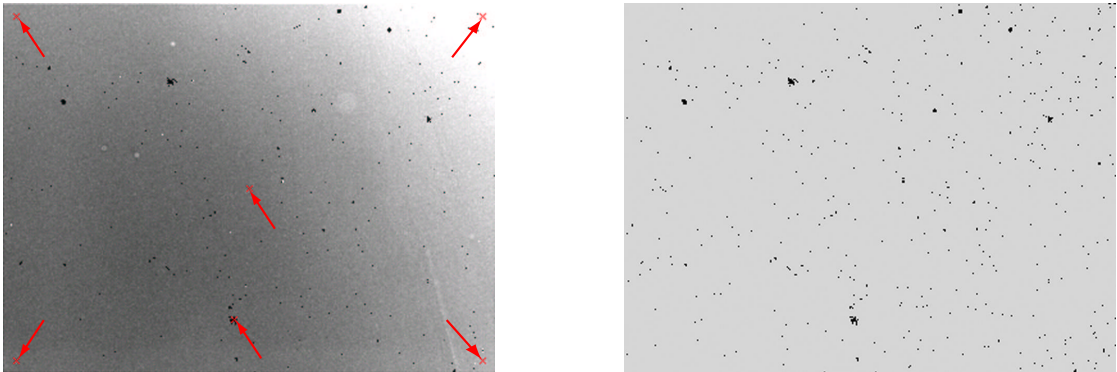


Figure 7.1: Left: Raw grey-value intensity image obtained with the Thermosensorik CMT camera imaging a laboratory blackbody at a temperature of 24°C. The calibration curves for the pixels marked with red arrows are depicted in figure 7.3. Right: Binary pixel mask for the Thermosensorik camera. Pixels labeled as defective are shown in black, good pixels are shown in grey.

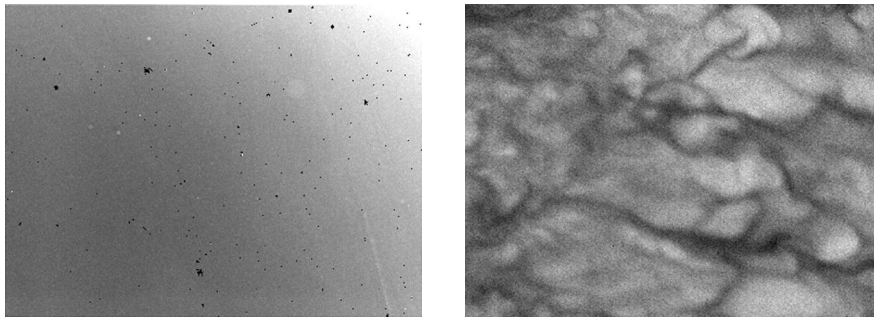


Figure 7.2: Image of the temperature distribution on a water surface before (left) and after non-uniformity correction and filling of gaps (right). The image was taken with the Thermosensorik CMT camera.

7.1.2 Amber Radiance Camera

For the Amber Radiance camera, the relationship between the grey-values of some selected pixels and the corresponding blackbody temperature is depicted in figure 7.4. The integration time was 1.8 ms. For the Amber Radiance camera, higher temperatures correspond to higher grey-values, in contrast to the Thermosensorik camera for which higher temperatures correspond to lower grey-values.

When looking at the calibration curves of the Amber Radiance camera, the fact that this camera has an on-board non-uniformity correction as described in section 6.1 must be taken into account. The camera also has a mode of operation in which it outputs raw data, but it was operated using the on-board correction as this allows immediate visual feedback using an external control monitor. This visual feedback facilitated operations such as focusing the camera on the water surface.

An additional radiometric calibration was performed on top of the on-board non-uniformity correction for two reasons. First, the internal non-uniformity correction uses a temperature-controlled plate within the camera body to perform a linear two-point calibration. Therefore, it does not take into account any non-uniformity introduced by the lens, such as vignetting.¹ Second, the radiometric calibration makes it possible to translate the grey-value output into equivalent blackbody temperatures, making it easier to compare the outputs of the two cameras.

Due to the fact that the on-board electronics of the camera replaces the output of defective pixels with the grey-value of an immediate neighbor pixel, the method of identifying bad pixels based on the quality of the fit is not applicable for this

¹The term vignetting refers to a reduced image intensity close to the borders of an image.

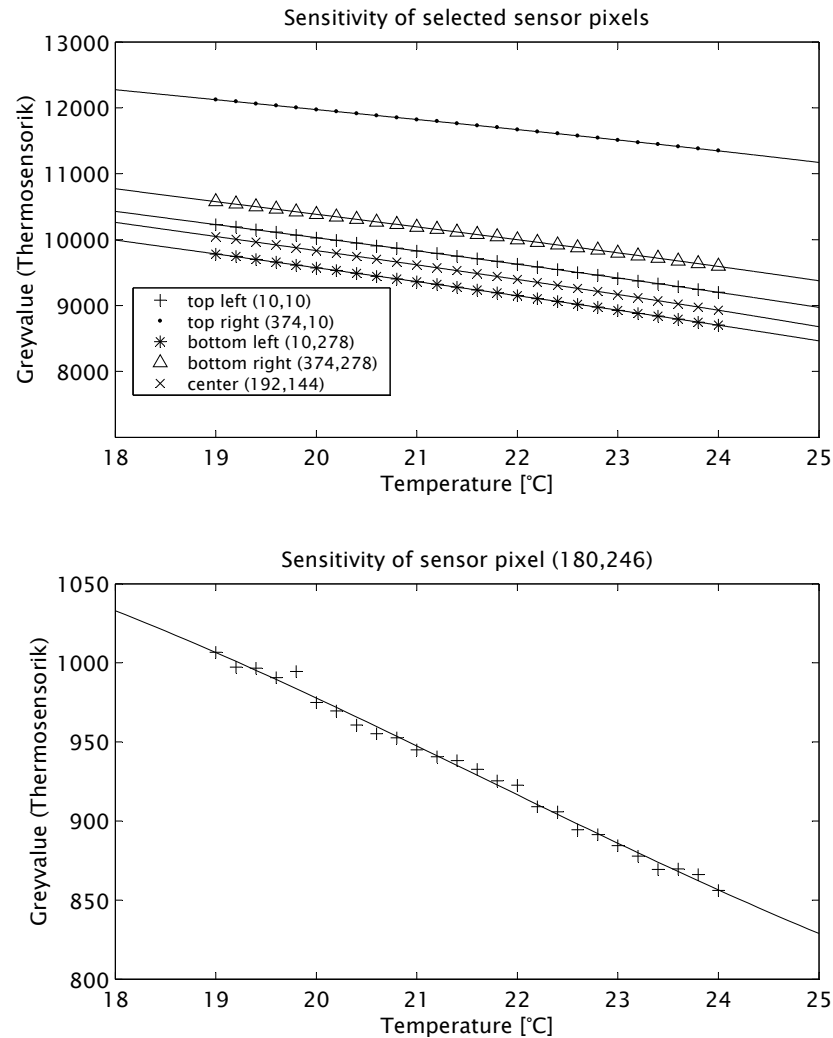


Figure 7.3: Sensitivity of selected pixels (marked in red in figure 7.1, left) for the Thermosensorik CMT camera. The mean grey-value for each pixel (averaged over 32 frames) is given as a function of the temperature of the imaged blackbody (calibration data points are marked +) together with the respective fit curve through the data points (solid lines). Note that the grey-values of the pixel at position (180,246) in the bottom graph are displayed at a different scale and differ from the grey-values of the other pixels by one order of magnitude.

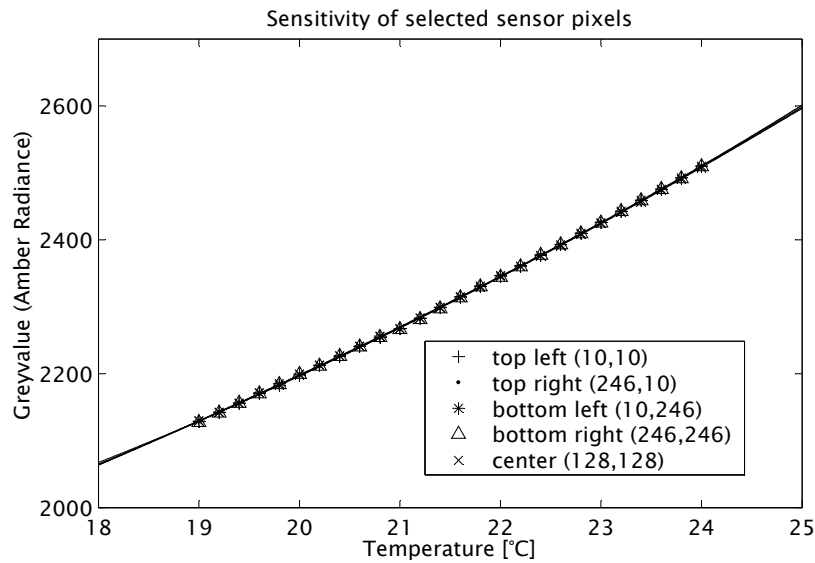


Figure 7.4: Sensitivity of selected pixels for the Amber Radiance camera. Refer to figure 7.3 for an explanation of the axes.

instrument. However, despite the on-board nearest neighbor interpolation, 26 pixels were identified as bad pixels by the distance-from-fit criterion that was also used for the Thermosensorik camera as described in section 7.1.1. These pixels are shown in figure 7.5. To fill in the intensity values of these 26 pixels, the regularization technique described in section 5.1 was used. For the other pixels, the nearest neighbor interpolation of the camera was relied upon.



Figure 7.5: Binary mask of the pixels of the Amber Radiance detector which, despite the on-board correction, were identified as bad pixels by the distance-from-fit criterion. Bad pixels are displayed in black, good pixels are displayed in grey.

7.2 Geometric Camera Calibration

This section presents the results of the geometric camera calibration obtained with the algorithm described in section 3.4.

The implementation of the algorithm used is a modified version of the *Camera Calibration Toolbox for Matlab*, which is a piece of open-source software provided by Bouguet [10]. Bouguet's implementation, however, uses a method for obtaining an initial guess for the camera parameters that differs from the method proposed by Zhang [107], and is not well documented. Therefore, Zhang's method of obtaining an initial guess, described in section 3.4.2, was implemented. However, it was found that the optimization step of the calibration (see section 3.4.3) is not very sensitive to the initial guess; for both methods of estimating the initial values, the optimization converged to the same result.

For the geometric calibration, a set of 38 stereo image pairs showing different orientations of the calibration target described in section 6.4 were acquired. This set of image pairs is depicted in figures 7.7 and 7.6.

Extraction of Grid-Corners The positions of the grid-corners in the checkerboard images were located in a semi-automatic way. The four corners delimiting the outside edges of the checkerboard were selected manually for each calibration image by clicking with the mouse. Using the information about the four corner positions, the software calculated initial guesses for the grid corner locations, which were subsequently refined using a sub-pixel corner estimation algorithm described by Harris and Stevens [41].

In some of the images, only a subset of the checkers was selected, because parts of the calibration target were not visible in the images or out of focus. Care was taken that the same subset of checkers was selected for corresponding images of the two cameras.

The calibration was initially performed using all 38 image pairs. A close inspection of the re-projection errors (see section 7.2.1, in particular figure 7.10) for this initial calibration revealed that some of the images contained gross outliers, due to inaccurate extraction of grid-corners. These images were manually deselected. The subsequent calibration was performed on the remaining 23 images comprising 942 calibration points in total.

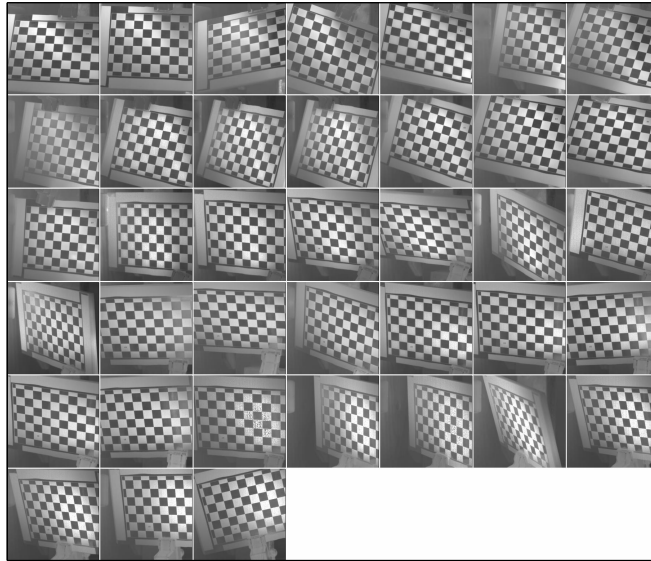


Figure 7.6: Calibration images taken with the Amber Radiance camera.

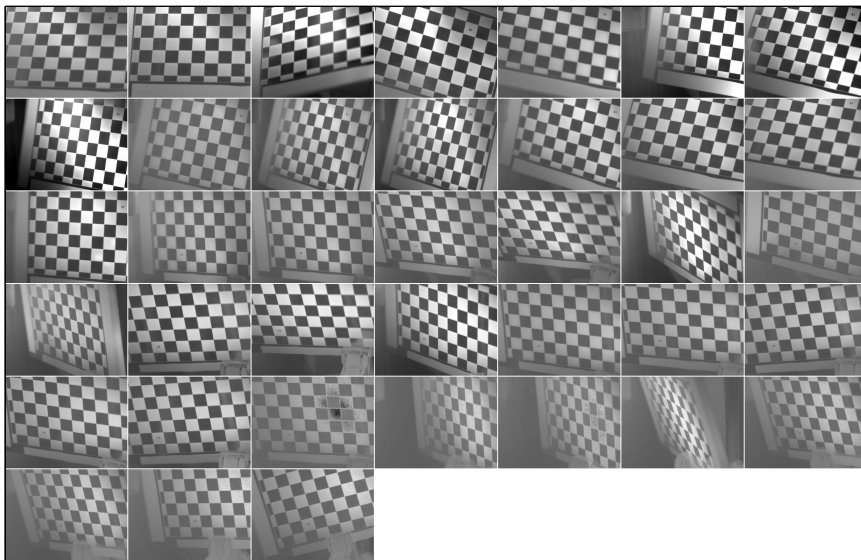


Figure 7.7: Calibration images taken with the Thermosensorik CMT camera.

7.2.1 Interior Parameters

The interior camera parameters obtained by the calibration algorithm are listed in table 7.1. The focal length is given both in terms of pixel dimensions $\alpha_{x,y}$ and in millimeters $f_{x,y}$. To calculate the focal length in millimeters, the $\alpha_{x,y}$ were multiplied by the pixel pitch, that is the center-to-center spacing of pixels on the detector chip. The pixel pitches are $38\ \mu\text{m}$ for the Amber Radiance camera and $24\ \mu\text{m}$ for the Thermosensorik camera (see table 6.1). The numerical errors associated with the parameters are derived from the inverse of the Hessian matrix used in the gradient-descent optimization of the re-projection error (see section 3.4.3) and are about three times the standard deviation.

Parameter	Amber Radiance	TS CMT384
Focal length α_x (pixel dimensions)	1405.5±6.1	2314.9±8.1
Focal length α_y (pixel dimensions)	1405.6±6.0	2316.6±7.7
Focal length f_x (mm)	53.4±0.2	55.6±0.2
Focal length f_y (mm)	53.4±0.2	55.6±0.2
Principal point u_0	143.3±7.5	275.0±15.4
Principal point v_0	123.9±8.6	118.7±11.9
Skew	not estimated	not estimated
First-order radial distortion κ_1	-0.189±0.054	-0.456±0.038
Tangential distortion τ_1	not estimated	-0.0039±0.0001
Tangential distortion τ_2	not estimated	-0.0053±0.0016
Pixel re-projection error e_x	0.052	0.065
Pixel re-projection error e_y	0.052	0.069

Table 7.1: Interior camera parameters determined by camera calibration.

Distortion Model In a first run of the calibration algorithm (see section 3.4.3), the optimization minimized the re-projection error of a camera model that included parameters for both first order radial distortion, κ_1 , and first order tangential distortion $\tau_{1,2}$. For the Thermosensorik camera, this model was adequate. However, for the Amber Radiance camera the tangential distortion parameters $\tau_{1,2}$ were found to be zero within the numerical uncertainties. Therefore, the camera parameters of the Amber Radiance were re-estimated using a distortion model that only included first order radial distortion. For both cameras, the resulting displacement between undistorted and distorted image coordinates is illustrated in figure 7.8.

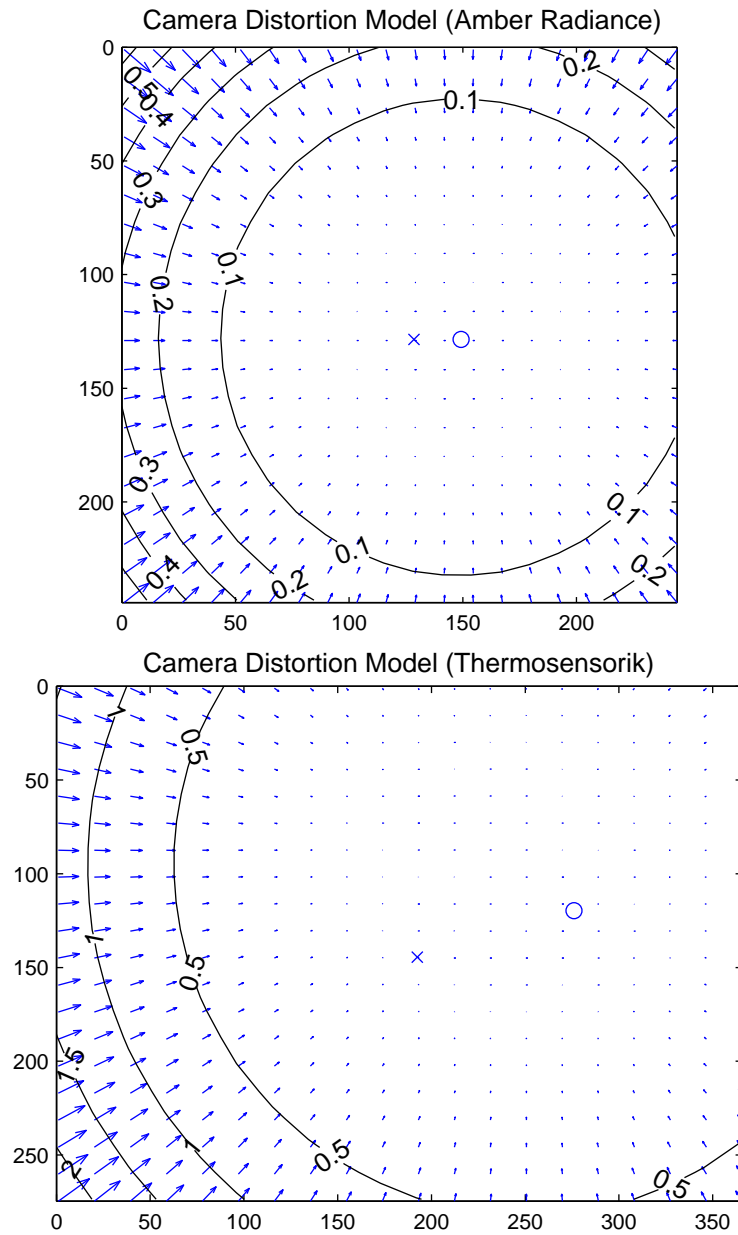


Figure 7.8: Camera distortion models for the Amber Radiance camera (top) and the Thermosensorik camera (bottom). The vectors show the displacement between distorted and corrected image coordinates (see equations (3.15)-(3.18) in section 3.3.1). The contour lines mark lines with equal magnitude of displacement. The image center and the principal point are marked \times and \circ , respectively.

Re-projection Error Table 7.1 also lists the standard deviation of the re-projection error, defined in section 3.4.3, over all calibration points in x- and y- directions. As illustrated in figure 7.9, the re-projection error is the difference between the grid corner positions extracted from the images (marked with a +) and the grid corner positions calculated by re-projecting the known locations of the feature points on the calibration target into the image with the estimated camera parameters (marked with a \circ). The re-projection errors for all feature points used for calibration are plotted in figure 7.10.

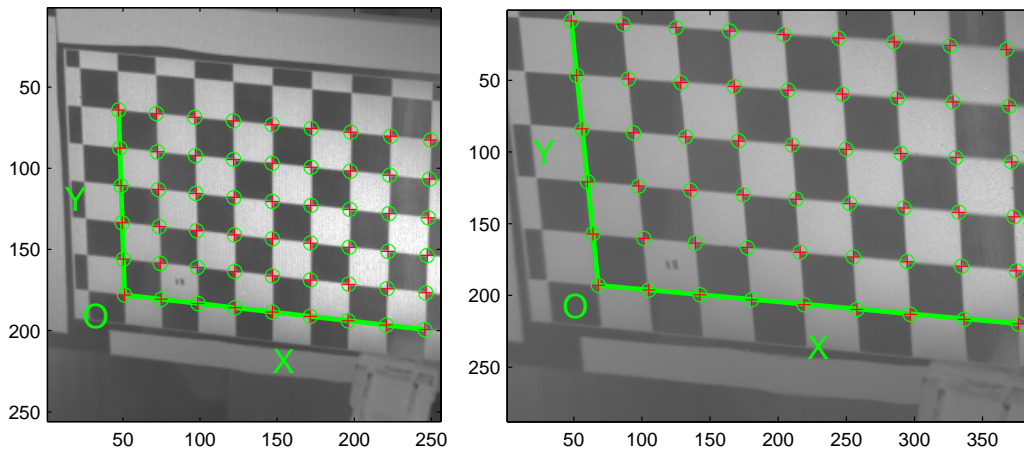


Figure 7.9: Re-projection of grid-points for one of the calibration image pairs. The extracted grid corners are marked with red crosses and the grid corner positions calculated by re-projection are marked with green circles. The axis labels denote the pixel coordinates. Left: Amber Radiance camera, right: Thermosensorik camera.

7.2.2 Exterior Parameters

The exterior parameters of the stereo setup obtained by the calibration process are listed in table 7.2 and are illustrated in figure 7.11, together with the estimated positions and orientations of the planar calibration target. The world coordinate system was chosen to coincide with the Amber Radiance camera coordinate system. The vector \mathbf{C}_{TS} represents the position of the camera center of the Thermosensorik in this coordinate system.

The rotation is represented in terms of a vector \mathbf{R} , whose direction determines the axis of rotation and whose length specifies the rotation angle. This representation was used in the gradient-descent optimization, because it consists of only three

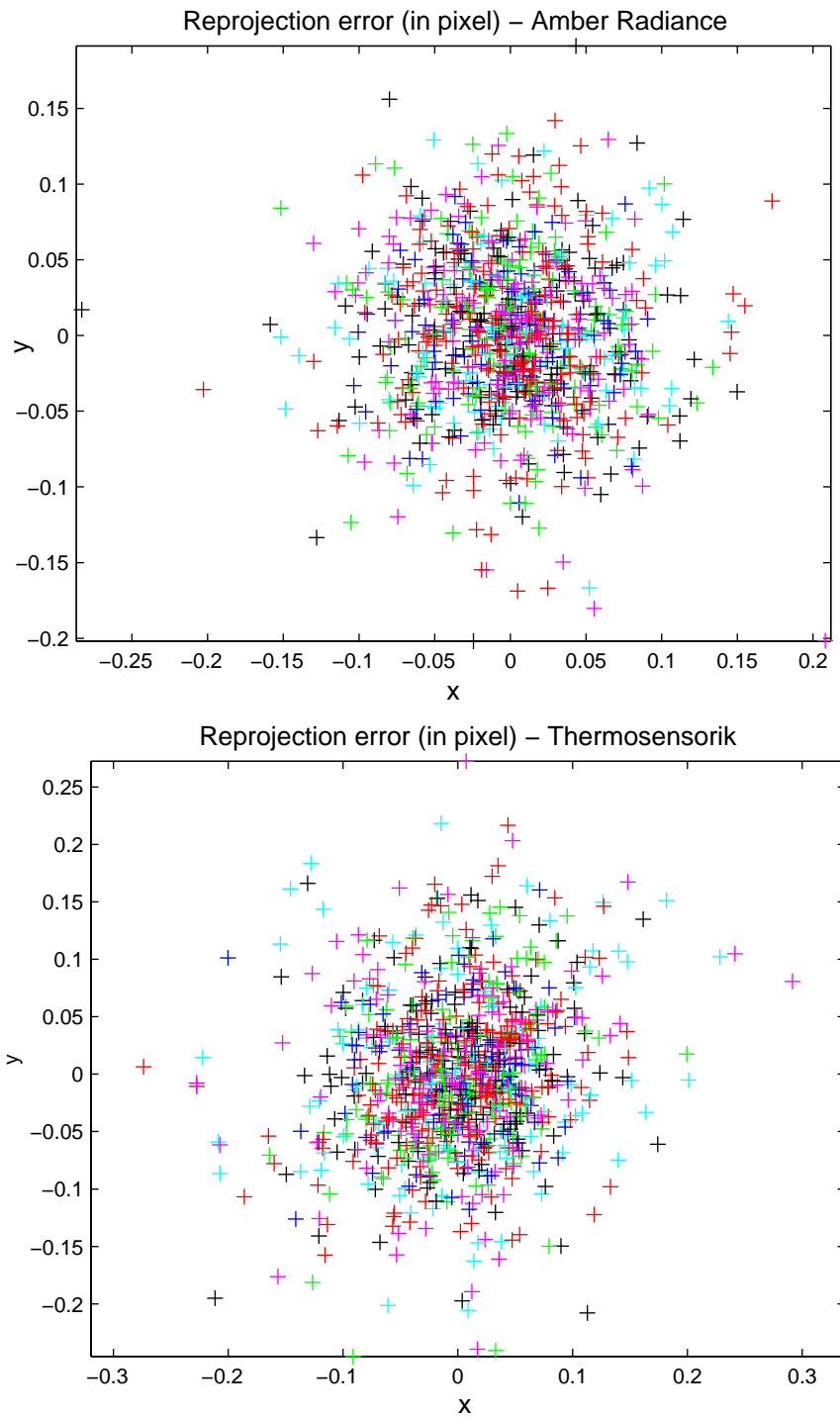


Figure 7.10: Re-projection error for the Amber Radiance camera (top) and Thermosensorik camera (bottom). For an explanation of the re-projection error, refer to section 3.4.3. The colors of the points correspond to different orientations of the calibration pattern as depicted in figure 7.11.

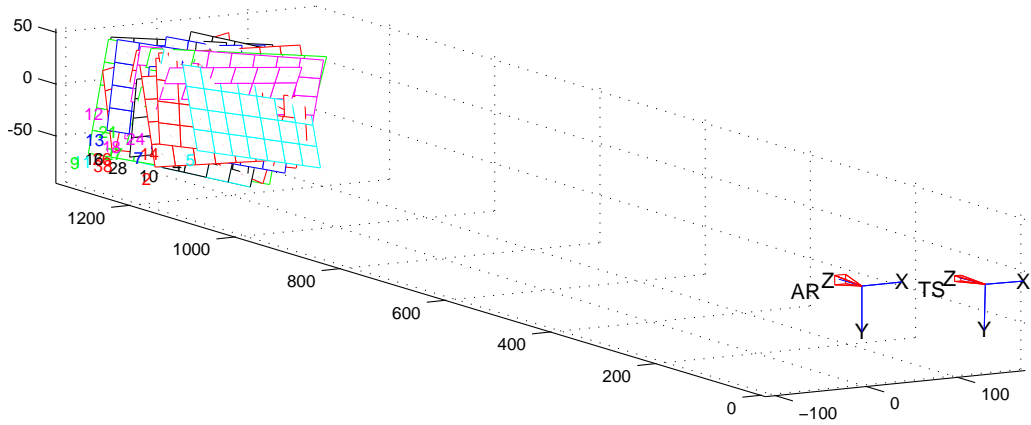


Figure 7.11: Exterior orientation obtained by the stereo calibration algorithm. The camera centers are at the origins of the small coordinate frames; AR and TS represent the Amber Radianc and the Thermosensorik infrared cameras respectively. The field-of-view of each camera is represented by a red pyramid. The positions of the calibration target and the Thermosensorik camera are relative to the camera coordinate system of the Amber Radianc, which was chosen as the world coordinate system. The axis labels are in millimeters.

Parameter	Value
Rotation vector \mathbf{R}	(0.002, 0.077, -0.0270)
Uncertainty of rotation vector	(0.007, 0.008, 0.0003)
Translation (Position of C_{TS} in world frame) [mm]	(129.7, 5.1, -18.5)
Uncertainty of C_{TS} [mm]	(0.5, 0.2, 4.8)

Table 7.2: Exterior parameters of the stereo camera setup.

elements, corresponding to the three independent degrees of freedom. The same rotation can also be expressed in terms of a rotation matrix as

$$\mathbf{R} = \begin{bmatrix} 0.9967 & 0.0271 & 0.0771 \\ -0.0269 & 0.9996 & -0.0033 \\ -0.0772 & 0.0012 & 0.9970 \end{bmatrix}. \quad (7.1)$$

Using these exterior parameters and the interior parameters presented in section 7.2.1 the projection matrices in the world coordinate system defined by the Amber Radiance camera are

$$\mathbf{P}_{\text{RAD}} = \begin{bmatrix} 1405.5 & 0 & 143.3 & 0 \\ 0 & 1405.6 & 123.9 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (7.2)$$

$$(7.3)$$

for the Amber Radiance camera and

$$\mathbf{P}_{\text{TS}} = \begin{bmatrix} 2285.9 & 63.0 & 452.6 & -295262.8 \\ -71.48 & 2315.9 & 110.8 & -9573.4 \\ -0.077 & 0.0012 & 0.997 & 18.45 \end{bmatrix} \quad (7.4)$$

for the Thermosensorik camera. These matrices were calculated by inserting the camera parameters obtained from calibration into the definition of the camera projection matrix, equation (3.12). They describe the relationship between the coordinates of 3D scene point in millimeters (augmented with a one to create a homogeneous 4-vector) and its homogeneous image coordinates in pixels.

7.3 Rectification

To facilitate the use of the epipolar constraint in the subsequent stereo matching step, the images were rectified using the algorithm described in section 3.5.4, making use of the exterior and interior parameters obtained from the camera calibration.

The outcome of the rectification is illustrated in figure 7.12 for a stereo image pair of an infrared sequence¹ of a wavy water surface. Before applying the rectifying transforms, the non-linear distortion was removed from the images using algorithm 3.1.

A visual inspection of distinctive grey-value features in several of the rectified image pairs confirmed that the rectification works correctly; apart from small variations of typically less than 1 pixel, corresponding points were aligned on the same image scanline. The deviations that were noticed are not a result of the rectification process, but are due to the limited accuracy with which the camera parameters and thus the epipolar geometry are known.

As a result of the rectification, the size ratio of the two images is changed. The different scaling corrects for the different fields-of-view of the two cameras (see section 6.1.2); corresponding structures are now approximately the same size.

The rectified left image, taken with the Amber Radiance camera, is enlarged with respect to its original size of 256×256 pixels. Its horizontal and vertical extent is now 350×350 pixels. Note that it does not fully fill a 350×350 pixel bitmap as it is tilted. In contrast, the rectified right image, corresponding to the Thermosensorik camera, shrank from its original size of 388×284 pixels to a size that fits into a rectangular bitmap image of size 310×238 pixels.

Because the rectified images do not fill a whole rectangular bitmap image, it is useful to express the scaling in terms of the total number of pixels. The 65536 original pixels of the Amber Radiance are mapped to 106323 pixels in the rectified images; this is an increase of a factor of 1.62. The 110592 original pixels of the Thermosensorik camera are mapped to 68065 pixels in the rectified images; this is a reduction by a factor of 0.615. Therefore, when disparity estimation is performed for the rectified images, one original pixel of the Amber Radiance camera is in effect compared to $1.62/0.615 = 2.63$ original pixels of the Thermosensorik camera during the evaluation of the correlation score (equation (4.7)).

¹The particular image pair is frame 1 of sequence 3 (see section 6.6.4).

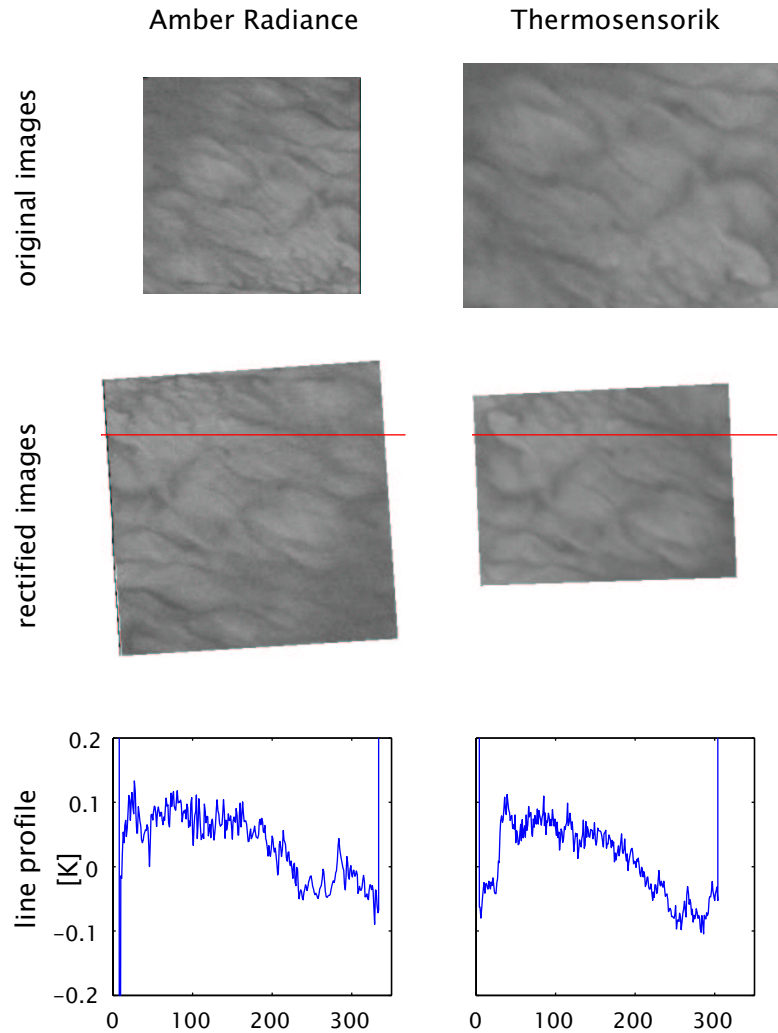


Figure 7.12: Results of the stereo rectification. An infrared image pair of a wavy water surface is shown in the top row. The middle row shows the corresponding rectified stereo pair. Note that rectified images are mirrored both vertically and horizontally, but this does not affect the subsequent processing. The bottom row presents line profiles of two corresponding lines, marked red in the rectified images. The left and right column correspond to images obtained with the Amber Radiance camera and the Thermosensorik camera, respectively.

Another negative effect of the different fields-of-view of the two cameras that is apparent in figure 7.12 is the fact that many pixels cannot be used for disparity matching due to the limited overlap of the imaged areas.

7.4 Disparity Estimation

This section documents the results obtained with the stereo matching algorithm described in section 4.3. First, in section 7.4.1, the results for a test image pair are presented. Second, in section 7.4.2, the results for an infrared image pair of a wavy water surface are presented and used to analyze the multi-scale disparity estimation.

7.4.1 Test Image Results

The disparity algorithm was tested on a standard test stereo image pair, to check the correctness of the implementation and to allow for a comparison with other algorithms.

The image pair, shown in figure 7.13, was taken from a standard test image series from the University of Tsukuba, which is available on the internet [81]. Ground truth disparity data for this image pair, with a resolution of one pixel, is shown on the bottom right of figure 7.13. The results obtained with the implementation of the algorithm described in section 4.3 are shown on the bottom left of figure 7.13. The matching was performed on three different levels of a Gaussian image pyramid using correlation window sizes of 9×9 , 7×7 and 5×5 pixels (from highest to lowest resolution). In figure 7.13, the disparity combined across scales is shown.

A comparison of the resulting disparity map with the ground truth data demonstrates that the algorithm produces reasonable disparity estimates. Problematic areas, where no matches can be found, are mostly located near object boundaries and can be attributed to occlusions. A notable exception is the wrong match located left of the depicted camera, which is probably due to the repetitive pattern of the books in the shelf. Due to the fixed size of the correlation window, small structures, such as the legs of the tripod, are fattened and disparity discontinuities are smoothed out.

The Tsukuba test image pair is not very representative for the situation encountered in the present work, because the depicted scene is comprised of many different objects, giving rise to depth discontinuities and occlusions. Also, the images have very low image noise and the whole image scene is in focus. In contrast, the images in the present work are of a single smooth surface, comparatively noisy and

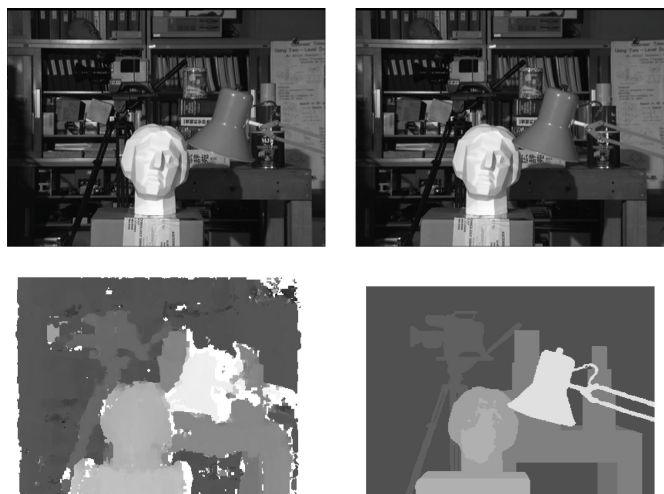


Figure 7.13: Results of the disparity matching algorithm described in section 4.3 applied to a standard test image pair. Top: Stereo image pair from the University of Tsukuba image series. The size of the images is 384×288 pixels. Bottom: The disparity map, obtained with the implemented algorithm, is shown on the left. The disparities are with respect to the right image; brighter grey-values correspond to larger disparities. Pixels for which no correspondence was found are shown in white. The ground truth disparity, with disparity levels quantized in 1-pixel steps, is shown on the right. The Tsukuba stereo image pair and the ground truth data are available on the internet [81].

sometimes blurred. However, the Tsukuba image has been used by Scharstein and Szeliski [82] and therefore it allows comparison with a large number of other algorithms. Such a comparison reveals that the implementation yields similar results as other algorithms that do not employ global disparity optimization. Algorithms using global disparity optimization generally perform better in terms of preserving disparity discontinuities across object boundaries.

7.4.2 Multi-Scale Disparity Estimation

This section presents the disparity estimate obtained for an infrared image pair¹ of a wavy water surface.

The image pair is shown at three different levels of a Gaussian image pyramid in figure 7.14, together with the disparity maps computed at each level. As mentioned in section 7.3, the image area that is usable for disparity matching fits

¹The particular image pair used for this discussion is frame 88 of sequence 4 of section 6.6.4.

into a rectangular area of 310×238 pixels. Note that the image acquired with the Thermosensorik camera is slightly out of focus.

The sizes of the correlation windows at each scale were determined empirically, such that a good balance between a dense disparity map and good resolution was achieved at each level. The correlation window sizes were 21×21 , 11×11 and 9×9 pixels in order from highest to lowest resolution. These window sizes are quite large compared to the window sizes used for the “Tsukuba” test image in the previous section. The need for large window sizes is probably attributable to image noise and the fact that the right image is slightly out of focus. These window sizes were also used to analyze the wave image sequences presented in section 7.7.

Judging by the density of the valid disparity estimates, the image regions with spatially finer texture, at the top left and bottom right of the infrared images, can be better matched, as would be expected.

The holes in the disparity map obtained at the highest resolution can to some degree be filled in by the results at lower resolution using algorithm 4.1 (page 90). This multi-scale approach yields the combined disparity shown at the bottom left of figure 7.14.

The bottom right of figure 7.14 shows the combined disparity after outliers have been removed using the method described in section 5.2; that is, for each scanline, disparity estimates that differ by more than 2.3 standard deviations from the mean disparity of that scanline were removed.

7.5 Regularization

Regularization of the image sequences was performed using the spatio-temporal membrane model described in section 5.3.3. To compute the local average, a $7 \times 7 \times 3$ pixel Binomial filter mask was used. The filter mask was made larger in the spatial dimensions than in the time dimension to give spatial smoothness a higher weight over temporal smoothness.

The elasticity parameter α , which controls the smoothness, was set to 0.6. The confidence value κ was set to zero for disparity estimates which failed the left/right consistency check (see section 4.3.5) or were marked as outliers (see section 5.2), otherwise κ was set to one.

Figure 7.15 illustrates the results obtained with this regularization technique for a disparity image taken from water wave sequence 4 (see section 6.6.4). As can be seen that convergence is reached after a few iterations; there is little change between the 6th and the 9th iteration. Close inspection of figure 7.15 reveals that

Multi-Scale Stereo Matching

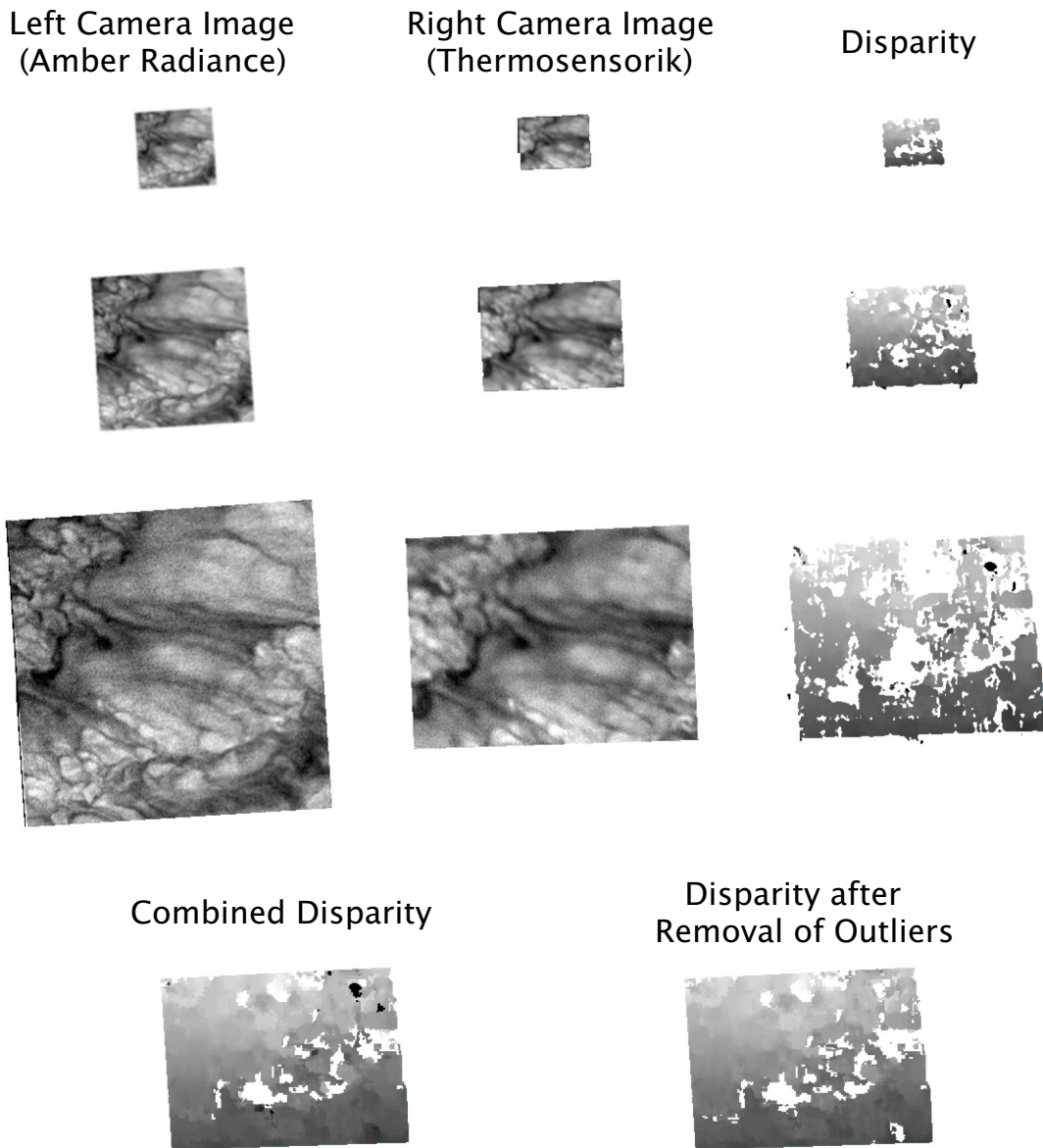


Figure 7.14: Disparity estimation at different scales, demonstrated for a stereo image pair of a water surface in the infrared region. The top three rows show the stereo image pair at different scales, together with the disparity estimates at each scale. For ease of comparison, the grey-values of disparities obtained at lower resolutions have been scaled to their equivalent disparity at the highest resolution. The combined disparity is shown on the bottom left. The final disparity map after outliers have been removed is shown on the bottom right. The range of disparities is 5 pixels to 25 pixels of displacement, with darker grey-values representing lower disparities. Points for which no valid disparity estimates were found are shown in white.

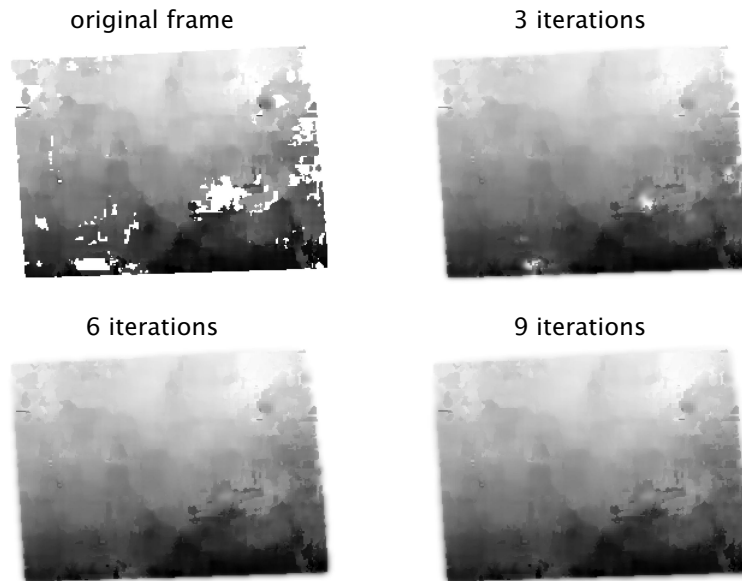


Figure 7.15: Regularization results obtained using the spatio-temporal membrane model (see section 5.3.3). The original disparity map is shown at the top left, together with regularized versions for different numbers of iterations of the update rule in equation (5.9).

the interpolation of the larger gap in the bottom right quadrant of the image is not very smooth. This is because the gap, which is relatively large in the spatial directions, was filled in with disparity estimates from neighboring frames during the first iterations as the Binomial filter also extends into the time dimension.

The fact that the interpolation in time direction produces the observed artifacts can be attributed to the frame rate of 60 Hz, with which the image sequences were acquired. This frame rate is not high enough to fulfill the sampling theorem (see Shannon [89]). Due to the resulting temporal aliasing, the discretization of the differential operators in equation (5.13), leading to the iterative update rule in equation (5.9), is not correct. The artefacts can also be observed in the space-time images presented in section 7.7, figure 7.19.

For some frames, it was observed that even a single outlier in the data that is erroneously labeled as correct ($\kappa = 1$), can severely affect the quality of the regularized solution if it is located within a larger area where no data is available. In such a case, the iterative procedure propagates the grey-value of the outlier to fill the gap.

7.6 Depth Reconstruction

Based on the disparity map calculated for the infrared stereo image pairs, a 3D reconstruction of the water surface was performed using triangulation as described in section 3.5.5.

The reconstruction was performed point by point for each valid entry in the disparity image, yielding a list containing the 3D coordinates of the reconstructed points.

Image sequences of the 3D reconstruction were produced by creating a scatter plot of the reconstructed points for each frame of the original stereo image sequence. An example of such a scatter plot, showing the reconstruction of a wavy water surface, is depicted in figure 7.16. Other visualization methods, which interpolate between the individual points to produce a continuous surface, were also evaluated; however, a small number of outliers can severely affect the visual quality of these methods, and therefore they were not found to be useful for this application.

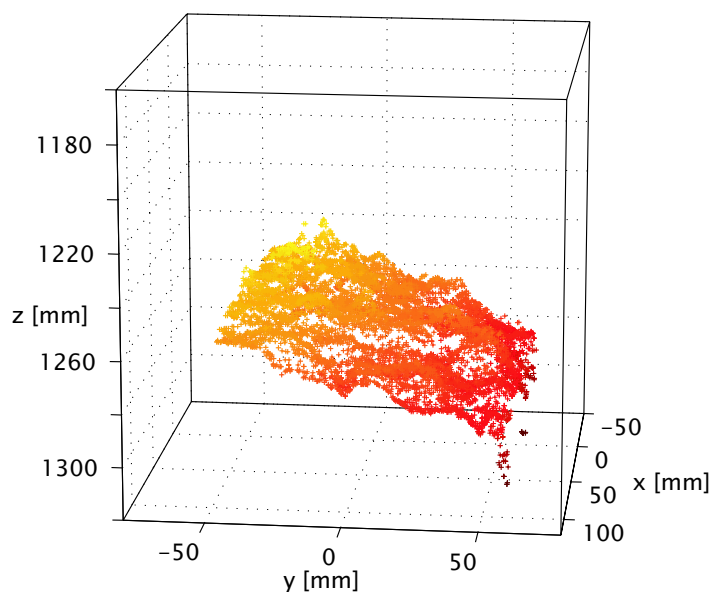


Figure 7.16: 3D Reconstruction of a wavy water surface, visualized as a scatter plot. The coordinates are with respect to the world frame defined by the Amber Radiance camera. For better visibility only a third of the total number of reconstructed points are shown. The particular image pair used to produce this example is frame 62 of sequence 3 (see section 6.6.4).

7.7 Measurements of Water Waves

In this section, the results obtained for image sequences of a wavy water surface are presented. Refer to section 6.6.4 for details about how these sequences were acquired.

Because of their inherent three-dimensionality, image sequences are difficult to present on paper¹. To address this problem, space-time images are used in this section to depict the temporal change of image grey-values. Such a space-time image shows the temporal evolution only for a single scanline rather than for a whole image to reduce the dimensionality. The velocity of a given object in the image sequence corresponds to the orientation of its grey-value trace in the space-time image. For example, an object that does not move at all creates a vertical bar in a space-time image (see [56] for a detailed explanation of space-time images).

Figure 7.17 shows the disparity estimates of four sequences recorded at different wind speeds as reported in section 6.6.4. The disparity matching was performed on three levels of a Gaussian image pyramid using the same correlation window sizes as in section 7.4.2. The image sequences are numbered according to table 6.3 on page 112. For each sequence, the disparity of a line is shown, with the time in seconds increasing from top to bottom. At the top of each space-time image a sample disparity image frame² of the sequence is used to mark the position of the scanline shown. The disparity values are in pixels, and are encoded as grey-values. Again, pixels for which no valid disparity estimate was found are shown in white. For the sequences in figure 7.17, larger disparities correspond to points that are further away from the camera. Note that the grey-value scale used for sequence 3 differs from the scale used for the other sequences.

In the space-time images of all four sequences, a distinct periodic structure, corresponding to a large gravity wave mode is visible. This wave mode is more dominant in the images sequences 1 and 3, which were taken at higher wind speeds.

In addition, some smaller periodic structures are visible. From their orientation in the space-time images, it can be observed that these structures propagate from right to left and are slower than the dominant wave mode.

It is interesting to note that the holes in the disparity estimates do not occur randomly but mostly along oriented structures that move from right to left in the space-time images. It is likely that these missing disparities arise from specular

¹The digital image sequences presented in this section can be found on the accompanying DVD or be requested from the author via email: Volker.Hilsenstein@iup.uni-heidelberg.de.

²The particular frame numbers of the sample disparity images are frame 20 for sequences 1,3,4 and frame 1 for sequence 2.

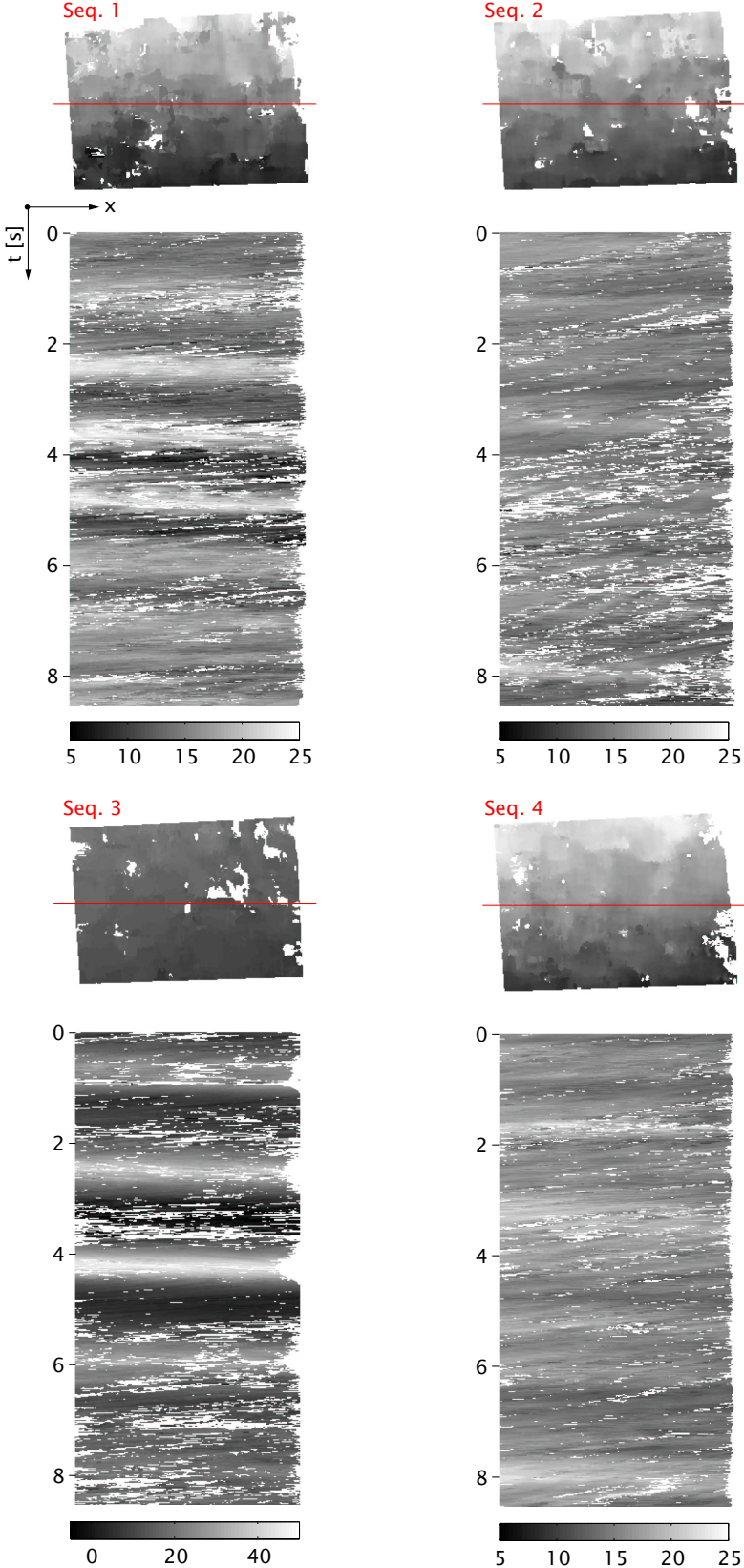


Figure 7.17: Space-time disparity images. Refer to the text of section 7.7 for details.

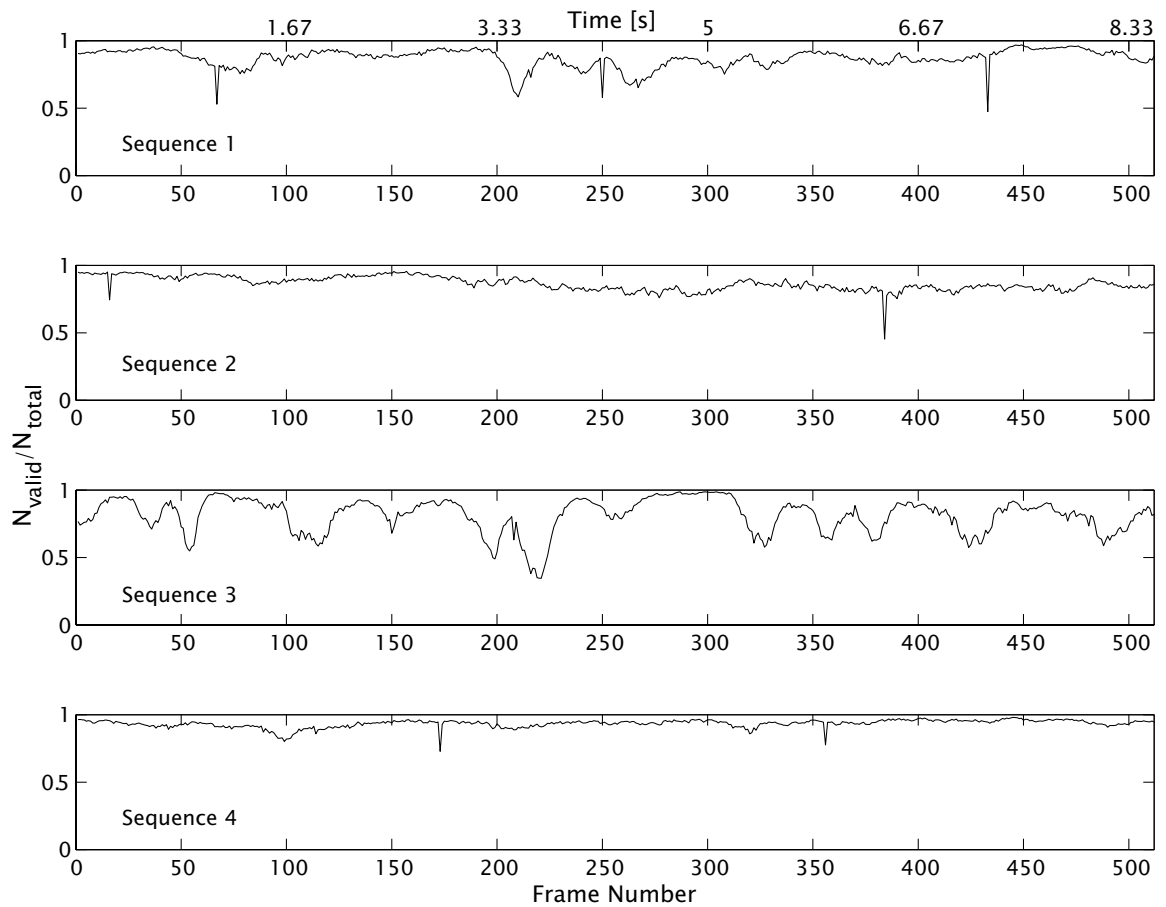


Figure 7.18: Fraction of pixels for which a valid disparity estimate could be found to the total number of usable pixels over time/frame number.

reflections. Their velocity is determined by the propagation velocity of areas with constant slope on the wave face, for a slope angle at which the radiation of some distant warm object is specularly reflected towards the camera.

Despite the validation of matches using the left/right consistency check (see section 4.3.5) and outlier removal (see section 5.2), careful visual inspection of the space-time images reveals some disparity estimates that are likely to be faulty. In the actual image sequences these faulty estimates appear more salient than in the printed space-time images.

The wave motion should lead to continuous grey-value trajectories in the space-time images. Clearly, this is only the case for some of the larger structures. The finer structures appear to be discontinuous. This is caused by temporal aliasing:

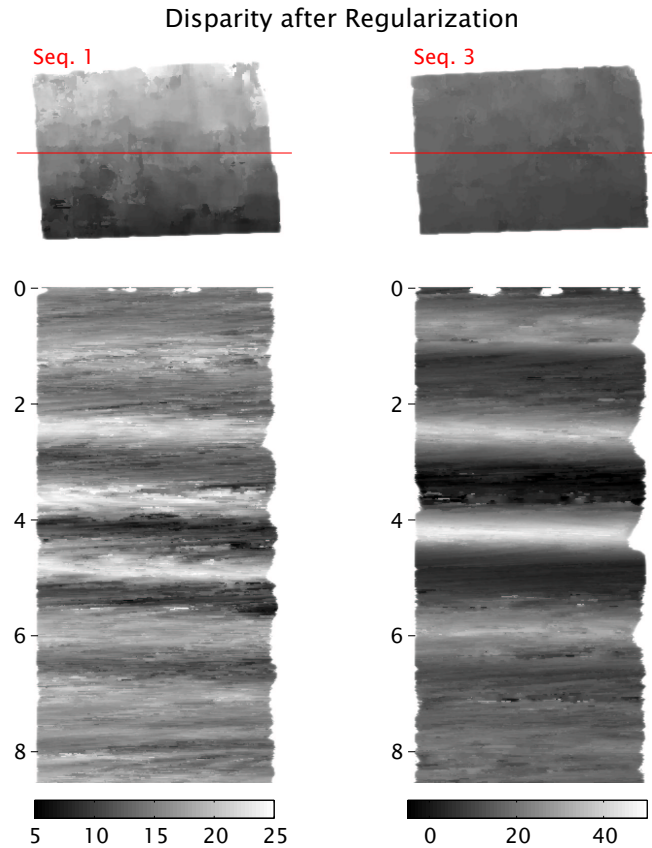


Figure 7.19: Space-time disparity images after regularization using the 3D membrane model described in section 5.3.3. The axes and sequences numbers are as in figure 7.17.

the frame rate of the camera is not high enough to resolve the motion of the smaller structures. In addition, the amplitude of shorter waves is usually smaller. Therefore, they cause smaller variations of disparity that are not well resolved given the limited spatial resolution of the cameras and the accuracy of the disparity estimation.

The fraction of pixels for which a valid disparity estimate was found to the total number of “usable” pixels was calculated for each frame. Figure 7.18 depicts this fraction for the four sequences. For each sequence, every pixel for which no valid disparity estimate was found in any of the 512 frames was marked as “unusable”, while all other pixels were marked as “usable”. The number of usable pixels was around $46 \cdot 10^3$, varying by several 100 pixels for the individual sequences.

It is apparent that the sequences which exhibit a higher variation of disparity, corresponding to larger wave heights and higher wind speeds, have a higher fraction

of invalid matches. One factor that contributes to this observed effect is the following: For larger wave heights, the range of distances between the water surface and the camera becomes larger. As a result, the water surface moves out of focus if it comes too close to the Thermosensorik camera, which has a narrow depth-of-field (see section 6.1.2). The resulting blur in the infrared images results in a less pronounced correlation of corresponding image patches and consequently leads to a lower number of found matches. This explanation is also supported by the space-time image of sequence 3 in figure 7.17. In this sequence, many gaps in the disparity estimate occur when the disparity is low, that is, when the water surface is closer to the camera. Visual inspection of the original infrared image sequences confirms this interpretation; the frames with a low number of correct matches were found to correlate with images that were out of focus.

Another possible cause that may contribute to the observed dependence of the found matches on wave height is the higher velocity that is associated with larger gravity waves. Faster motion causes blurring because the exposure of the detector takes some time. In addition, the slight delay between the exposure of the two cameras (see section 6.3.2) becomes more critical.

In figure 7.19, the space-time images of sequences 1 and 3 after applying regularization using the membrane model are shown. In these space time images, the gaps in the data have been filled in. However, as described in section 7.5, the quality of the interpolation is not very good; many image artifacts are visible.

Image sequences that show the motion of the recorded water surfaces in 3D as a series of scatter plots have been created from the disparity image sequences (these sequences can be found on the accompanying DVD). Due to the unsatisfactory quality of the regularization, the original disparity images, which contain gaps, were used to create these sequences.

7.8 Discussion of Accuracy

This section provides an analysis of the accuracy of the reconstructed 3D points that are obtained with the infrared stereo camera system described in this work. Part of this analysis is based on the theoretical considerations presented in appendix A.

7.8.1 Assessing the Total System Accuracy Using a Reference Plane

As explained in section A.2, one way of assessing the total system accuracy of a stereo setup experimentally is to reconstruct the surface of an object with known

ground truth. For the present application, no ground truth data for the wavy water surface was available for cross validation, because an alternative instrument to measure the water surface in the wind-wave channel, a Color Imaging Slope Gauge (see Fuß [33]), was still under construction at the time of measurement.

Instead, a stereo image sequence of a flat water surface was acquired to serve as ground truth. Some images of this sequence are depicted in figure 1.3 on page 12. For this image sequence, it is known that the imaged surface was level. Therefore, the reconstructed 3D points should lie in a plane. If the position and orientation of this plane is known with high accuracy, the distance of the reconstructed points from this plane can be used as a measure for the precision of the reconstruction.

Determining the Reference Plane The accurate determination of the reference plane is based on the assumption that the disparity estimates are only affected by randomly distributed errors and not by systematic errors. This assumption is justified, as the only likely cause of systematic error, the time lag between the exposure of the two cameras (see section 6.3.2), does not play a significant role for a still surface (the motion of the temperature patterns on the surface can be neglected as it is slow with respect to the time lag). With this assumption of randomly distributed errors, the accuracy of the disparity estimate can be improved by taking the mean over several frames.

The reference plane was computed as follows:

- For each pixel, the mean disparity over 400 frames of the acquired flat water sequence (see section 6.6.4) was calculated. Note that for some pixels a disparity estimate was not available for every frame of the sequence. In these cases, the mean was taken only over the available disparity estimates. The resulting disparity map is shown in figure 7.20, together with the standard deviation of the disparity for each pixel (over the frames included in calculating the mean). The mean of this standard deviation, taken over all pixels, is 0.55 pixels.
- To avoid potential boundary effects, 30 pixels at the edges of the mean disparity were cropped using a morphological erosion operator (see, for example [56]).
- For the cropped mean disparity map, a 3D reconstruction was performed using the triangulation method described in section 3.5.5. The reconstruction is shown in figure 7.23, top left. The total number of reconstructed points is 34046.

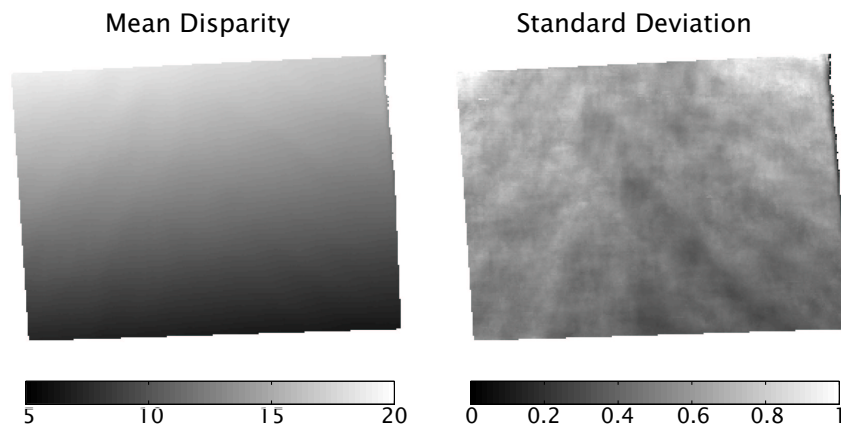


Figure 7.20: Left: Mean over 400 disparity maps calculated from a stereo image sequence of a flat water surface. The disparity is in pixels with respect to the left camera. The vertical gradient of the disparity reflects the fact that the camera was tilted with respect to the water surface. Right: Standard deviation of the disparity for each image point over 400 frames.

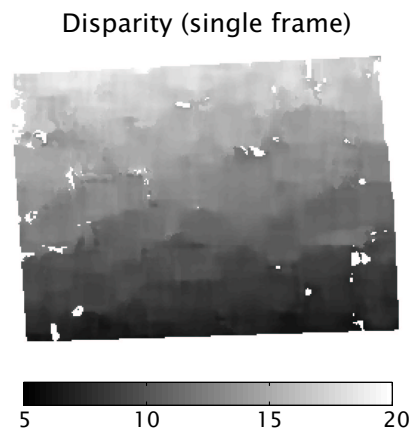


Figure 7.21: Disparity map for a single frame of the stereo image sequence of a flat water surface.

- A plane was robustly fitted to the reconstructed points using the RANSAC method, detailed in section A.3. The final consensus set of the RANSAC method comprised 97.4% of the total number of points. The homogeneous 4-vector representing the fit plane is $\tilde{\mathbf{P}}_{\text{ref}} = (1.265, 4.774, -8.185, 10^4)^T$. The standard deviation of the distance from the fit plane, calculated over all reconstructed points, was $\sigma_{d1} = 0.68$ mm. The fit plane is depicted in figure 7.22, together with the fit points, and from a different perspective in figure 7.23, top right.

Deviation from the Reference Plane The so-obtained fit plane was then used as a reference against which the points reconstructed from single frames of the disparity (in contrast to the sequence mean) were compared. For example, figure 7.23, bottom left, shows the reconstruction corresponding to the disparity map shown in figure 7.21 which was derived from a single frame of the stereo image sequence¹. The plot on the bottom right of figure 7.23 illustrates the deviation from the fit plane of the reconstructed points from this frame.

This distance from the fit plane was calculated for all points over all 400 frames of the sequence, that is, for a total number of $18.1 \cdot 10^6$ points. The standard deviation of all points from the reference plane was $\sigma_{d2} = 3.25$ mm.

¹The particular frame was number 12 of the flat water sequence.

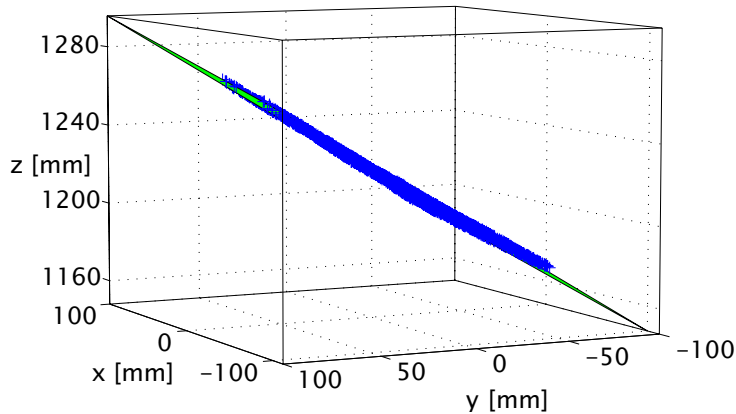


Figure 7.22: Points reconstructed from the mean disparity, shown in blue, and the reference plane fitted to these points, shown in green. The coordinates are with respect to the coordinates system defined by the Amber Radiance camera as shown in figure 7.11.

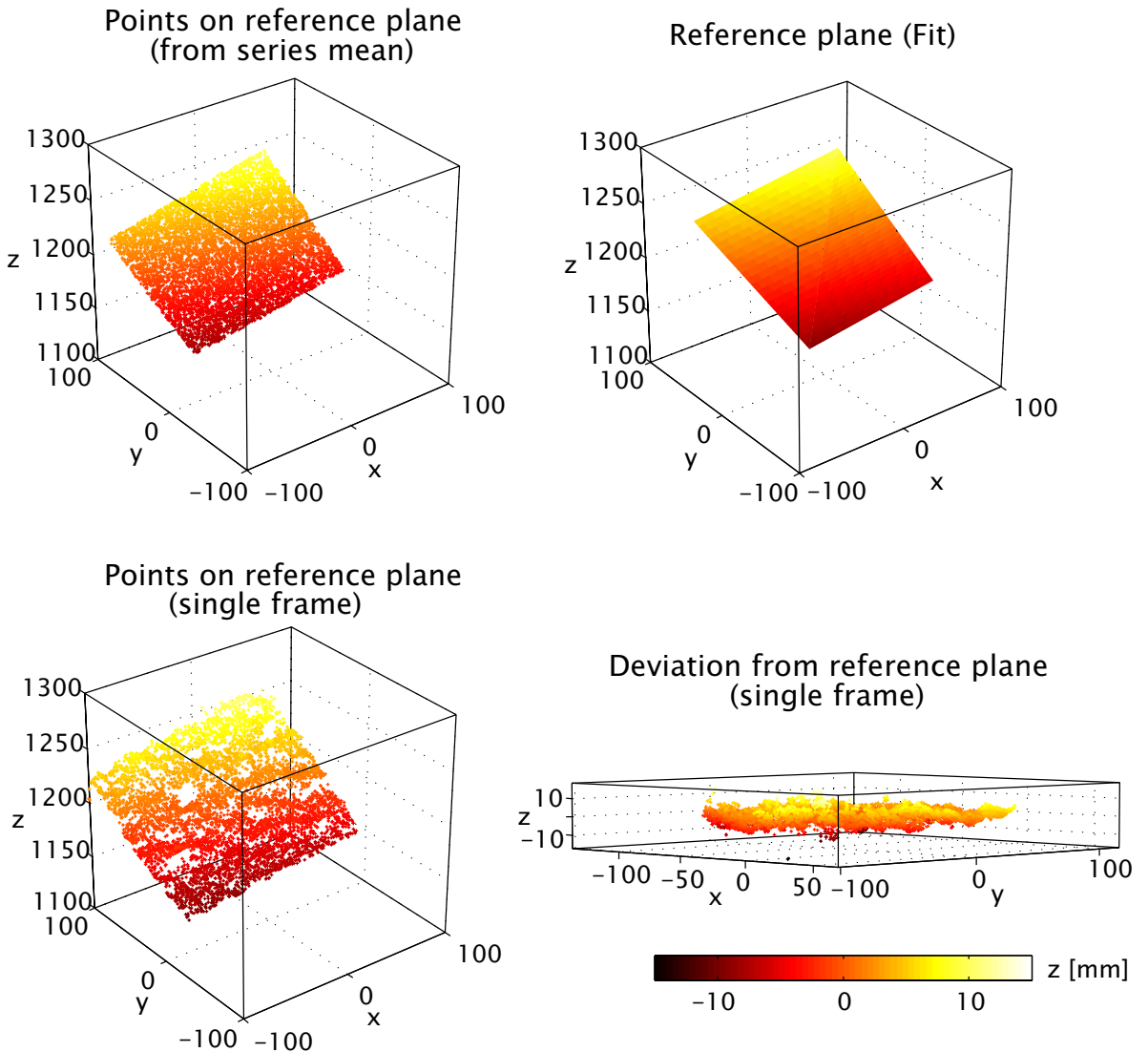


Figure 7.23: Top left: Scatter plot of points reconstructed from the mean disparity shown in figure 7.20. Top right: Reference plane fitted through the reconstructed points on the top left using the method described in section A.3. Bottom left: Scatter plot of reconstructed points from a single disparity frame of the sequence of a flat water surface. The corresponding disparity map is shown in figure 7.21. Bottom right: Deviation of the points on the bottom left from the reference plane. Except for the graphic at the bottom right, the coordinates are with respect to the coordinate system defined by the Amber Radiance camera as shown in figure 7.11. All axis labels are in millimeters. Note that the color bar only refers to the bottom right plot. For better visibility, the scatter plots show only a randomly selected 20-30% of the total number of reconstructed points.

The position of the reference plane is not known exactly (the standard deviation of the points used for the fit from the reference plane is σ_{d1}). This is accounted for by combining σ_{d1} and σ_{d2} to obtain the total uncertainty of the reconstruction

$$\sigma_d = \sqrt{\sigma_{d1}^2 + \sigma_{d2}^2} = 3.32 \text{ mm.} \quad (7.5)$$

Note that equation (7.5) is strictly not correct as σ_{d1} and σ_{d2} are not really independent.

Limitations The assessment of the achievable accuracy using the reference plane has several limitations.

First, the calculated deviation from the reference plane only includes inaccuracies in the direction normal to the plane. As the level water surface is not oriented perpendicular to the z -axis of the world coordinate frame, defined by the Amber Radiance camera, the estimated uncertainty mixes the lateral uncertainty, along the x - and y -axes, with the range uncertainty (see also figure A.2). However, the result is supported by the theoretical considerations on the range resolution achievable by triangulation presented in appendix A. For the given configuration of the stereo camera system, 10 cm on the water surface are sampled by roughly 250 pixels¹, corresponding to a lateral resolution of 0.4 mm per pixel. Assuming that the disparity estimates are accurate to half a pixel, and calculating the verging angle based on the distance to the surface of 120 cm and the baseline of 13 cm, equation (A.1) yields a value of $\Delta d = 3.7 \text{ textmm}$, which is in excellent agreement with the value $\sigma_d = 3.3 \text{ textmm}$ obtained using the reference plane.

Second, the reference plane does not take potential systematic errors into account. A possible source of systematic error, which can affect the disparity estimate, is the small time lag between the exposure of the two cameras. This is not a concern for the image sequence of a still water surface, used to determine the reference plane. However, it will affect the image sequences of water waves. As an example, consider a wave with a wave number of 20 rad/m, equivalent to a wavelength of 31 cm. Such a wave has a phase velocity of 0.7 ms^{-1} . During a time lag of 1 ms, the wave travels over a distance of 0.7 mm. For the configuration of the stereo system used in this work, where a horizontal distance of 10 cm was sampled roughly by 250 pixels, a distance of 0.7 mm can cause a systematic error of 1.75 pixels in the disparity estimate. Effects such as motion blur or defocusing are also not accounted for by the reference plane approach.

¹It is difficult to give an exact value due to the different fields-of-view of the cameras.

For the acquired wave image sequences, the uncertainty of the reconstructed points will therefore be somewhat larger than the 3.3 mm obtained with the reference plane approach.

8 Conclusions

8.1 Summary

A stereo-infrared imaging system for the 3D reconstruction of a wavy water surface was designed, implemented and tested.

The stereo-infrared wave gauge does not require a light source, that is, it is passive, and no parts of the system have to be submerged below the water surface. Therefore, it does not interfere with the wave motion and it can be used in the field, either from the boom of a ship or from a permanent structure, such as a pier. In contrast, instruments that require a submerged light source are considerably more difficult to deploy and are not suited for long-term operation from a permanent structure, as they are prone to damage from the environment.

The stereo correspondence problem was solved using a correlation-based disparity matching algorithm. Key requirements for correlation-based matching are that the imaged surfaces be opaque and have enough intensity variation, or texture. These requirements were met by operating in the $3-5\ \mu\text{m}$ region of the infrared spectrum, where water is opaque and, if a heat flux is present, exhibits rich temperature patterns at the surface, which can be recorded by infrared cameras.

For several infrared stereo image sequences of water waves at different wind speeds, acquired at the Aeolotron wind-wave flume in Heidelberg, it has been demonstrated that a dense reconstruction of the water surface can be obtained using this stereo system.

The 3D reconstruction accuracy obtained with the infrared camera system was evaluated under realistic operating conditions, that is, with similar image texture and image noise level as for the water wave sequences. To this end, a level water surface was used as a ground truth target. With this method, the accuracy of the system, given as the standard deviation of the reconstructed points from the fit plane, was found to be 3.3 mm.

8.2 Discussion and Outlook

The main contribution of this work is the use of infrared imaging to overcome some of the major problems associated with previous attempts (see sections 1.2 to 1.3) to use stereo photogrammetry for the reconstruction of wavy water surfaces. It was demonstrated that the shape of a water surface can be reconstructed based

on stereo infrared image sequences using algorithms from computer vision. As the infrared imagery can also be used to determine the heat flux at the water surface using the methods described by Haußecker et al. [46] and Garbe et al. [36], the stereo infrared wave gauge can be used to investigate the influence of short gravity waves on air-sea exchange processes in the field. The incorporation of a flexible technique for geometric camera calibration proposed by Zhang [108] permits quick adjustments of the system to adapt to different situations.

The main limitations of the system described in this work are the high cost of infrared cameras and the achievable surface resolution. With a reconstruction uncertainty of 3.3 mm for the configuration used in this work, the instrument cannot resolve capillary waves which have sub-millimeter amplitudes. Given that capillary waves contribute significantly to the mean squared slope of the water surface, this important parameter for air-sea exchange processes cannot be determined with the present system.

Several measures could be taken to improve the performance of the instrument. First, the use of two identical cameras and lenses would have several benefits. As was shown in section 7.5, the current stereo camera system does not make efficient use of the available sensor pixels due to the limited overlap of the imaged areas. With two identical cameras, it should be possible to create a greater overlap and make better use of the available sensor area. As an illustration, consider a stereo setup with two cameras of the type Thermosensorik CMT 384. If the areas imaged by the two cameras were arranged to overlap by 80%, which should be possible using identical optics, the number of usable pixels would be around 90000, roughly twice as many as in the current configuration.

Such a setup would have the additional benefit that the frame rate, which is currently limited to 60 Hz by the Amber Radiance camera, could be increased to 133 Hz. This would reduce the temporal aliasing effects observed in the image sequences, which are caused by the motion of shorter waves. As a consequence, the use of the temporal smoothness constraint for disparity regularization (see section 5.3.3) would be justified for finer spatial structures.

A potential source of systematic error, the time lag between the exposure of the two cameras, would also be eliminated for a system configuration using two Thermosensorik CMT cameras, because this camera model can be triggered externally.

Further hardware improvements are to be expected due to the steady progress in infrared detector technology. As described in Hewett [48], infrared cameras are currently making their way into consumer applications, and detectors with a larger numbers of pixels and a higher sensitivity are likely to become available.

The use of a wider stereo baseline would improve the range resolution that is achievable using triangulation (see A.2). However, due to the resulting greater difference in the view angle of the two cameras, this would also make the correlation-based disparity estimation more difficult.

Disparity estimation algorithms have received increasing attention from computer vision researchers in recent years, as is documented in a review by Scharstein and Szeliski [82]. Much of the recent work in this field is directed towards preserving disparity discontinuities, which is not of much concern for the present application. However, most of the newer algorithms make use of spatial smoothness constraints by integrating the regularization into the disparity estimation step, which would be useful for the present application. The algorithm proposed by Alvarez et al. [2] is particularly interesting, because it makes use of the known epipolar geometry to formulate the smoothness constraint based on the curvature of the reconstructed surface rather than based on the curvature of the disparity. For the application of wave imaging, such an approach would make it easier to set the parameters controlling the spatial smoothness to meaningful values derived from physical constraints rather than determining them empirically.

More accurate disparity estimates should be possible through integrating the temporal smoothness constraint based on the wave motion (see section 5.3.3) into the stereo matching algorithm. However, a prerequisite for the successful use of a temporal smoothness constraint are stereo image sequences recorded with a higher frame rate to avoid temporal aliasing.

Apart from these technical improvements, the system must be tested in the field to check whether the results obtained in the laboratory can be reproduced. In this regard, experience with field operation of infrared cameras collected during several research cruises over the past years (Schimpf [83], personal communication) is very promising, as infrared image sequences of high quality, that is, with rich texture and few specular reflections, have been recorded for a wide range of environmental conditions.

Work should also be directed towards improving the 3D visualization of the obtained results. For example, it is conceivable to use coloring to map spatially resolved heat flux estimates, computed using the image processing method proposed by Garbe et al. [37], onto the reconstructed surface. This would give previously unavailable insights into the spatial and temporal variations of air-sea exchange processes.

A Reconstruction Accuracy

In this appendix, methods to assess the accuracy of the 3D reconstruction obtained with the infrared stereo camera system are discussed.

Section A.1 presents the different factors that affect the quality of the reconstruction and their interdependencies.

In section A.2, the theoretical limit on range resolution that can be achieved using triangulation due to the finite image resolution is analyzed.

For the stereo camera system presented in this work, the achievable accuracy was assessed by experiment, using a flat water surface as a reference plane (see section 7.8.1). Section A.3 of this appendix explains the technique used to estimate the orientation and position of this reference plane robustly, given a number of reconstructed points.

A.1 Factors Influencing the Reconstruction Accuracy

The accuracy of the 3D reconstruction depends on a number of factors, such as the resolution of the cameras, the stereo baseline, the accuracy of the camera calibration, and the quality of the disparity estimates. Some of these factors in turn depend on other factors, such as the noise level and the amount of texture in the stereo image pairs. These factors often are not independent, for example the accuracy of the disparity estimates is affected by how well the epipolar geometry has been determined through geometric camera calibration.

In figure A.1 some of the factors that influence the accuracy of the 3D reconstruction and their interdependencies are identified. Adding to the difficulty of quantifying the uncertainty in the system is the fact that some of these factors, such as the image texture, are hard to provide a numerical value for. Given the complexity of the interdependencies, deriving an analytical expression that links the final accuracy of the reconstruction to all of these factors is hardly feasible.

However, it is possible to analyze individual links in this network of interdependencies. One such link, the influence of the camera resolution and the verging angle of the stereo system on the achievable range resolution, is analyzed in section A.2.

Another approach to obtain an assessment of the total system accuracy of the 3D reconstruction is by experiment. In order to do this, images of a scene containing a surface with known ground truth or marker points at known locations are recorded with the stereo camera system. The total system accuracy is then given by the

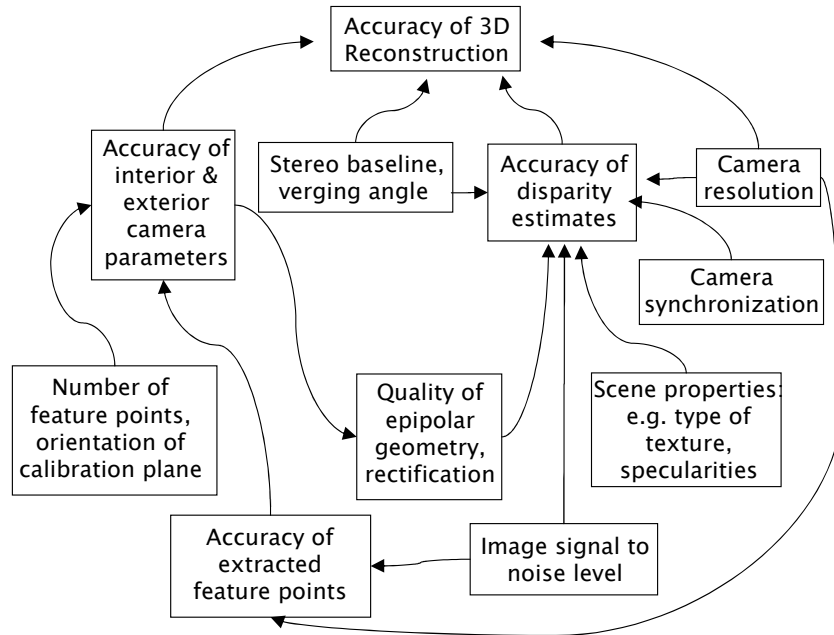


Figure A.1: Some of the factors influencing the accuracy of the 3D reconstruction and their interdependencies.

deviation of the 3D reconstruction from the known ground truth coordinates. For this experimental approach to be valid, it is important that the ground truth target is acquired under similar conditions as during the actual experiments. Specifically, the image texture and signal-to-noise ratio should be similar. In this work, a flat water surface was used as a ground truth target. Except for the lack of wind waves, it was recorded under exactly the same conditions as the sequences with water waves (see section 6.6.4). The accuracy is then given by the deviation of the reconstruction of the flat water surface from planarity. To calculate this deviation, a reference plane is robustly fitted to the reconstructed points as described in section A.3.

A.2 Range Resolution of Triangulation

As was shown in section 3.2.1, an image point and the projective center of a camera define a line of sight. Due to the limited resolution of real cameras, the coordinates of a point in the image plane are not known exactly, but are known to lie within a given region around the estimated coordinates. In most cases, the size of this

region is determined by the spacing of the sensor elements on the detector chip, or, if sub-pixel matching is used, some fraction thereof. In contrast to a single point coordinate, an uncertainty region around the estimated point no longer defines a single line of sight. Instead, it defines a solid angle.

Therefore, the range resolution of a 3D reconstruction is given by the volume defined by the intersection of the solid angles corresponding to the uncertainty regions of the image points used for triangulation. The exact size and shape of this intersection volume depends on the exterior orientation of the camera system and on the image coordinates and their uncertainties.

To obtain a rough estimate for how the range resolution Δd of a point reconstructed by triangulation depends on the image resolution, one can consider the simplified model depicted in figure A.2. Instead of central projection, this model assumes parallel projection. The lateral resolution corresponding to a single pixel is assumed to be Δx . It is apparent that the relationship between Δx and Δd depends on the vergence angle α of the camera system. Using similar triangles, the following relationship can be derived

$$\Delta d = \Delta x \frac{2 \cos \frac{\alpha}{2}}{\sin \alpha}; \quad (\text{A.1})$$

that is, Δd decreases for larger values of α . However, the uncertainty Δh in the direction perpendicular to Δd increases for larger values of α . The ratio of the uncertainties is

$$\frac{\Delta h}{\Delta d} = \tan \frac{\alpha}{2}. \quad (\text{A.2})$$

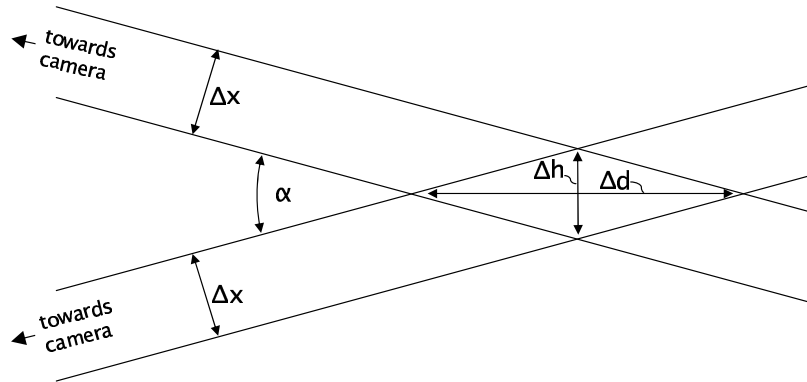


Figure A.2: Simple geometric model to calculate the uncertainty Δd of a depth estimate obtained by triangulation, assuming parallel projection with a spatial uncertainty Δx .

For a vergence angle of $\alpha = 90^\circ$, Δx and Δh are equal, minimizing the total uncertainty.

For a more detailed analysis, one would have to take into account that the camera images are formed by central projection, and that the internal and exterior camera parameters are not known exactly. Also, in the present application, the two cameras were not identical and had different resolutions.

A.3 Estimation of a Reference Plane

In section 7.8.1, the 3D reconstruction of a flat water surface was used as a reference plane to assess the accuracy of the infrared stereo camera system. This section briefly describes the algorithm that was used to robustly estimate the parameters of the reference plane, that is its position and orientation, given a number of reconstructed points $\tilde{\mathbf{X}}_1 \dots \tilde{\mathbf{X}}_n$ lying in this plane.

A.3.1 Least-Squares Estimate

Planes and points in the projective space \mathbb{P}^3 are represented by homogeneous 4-vectors, as described in section 3.1.2. Analogous to the case of points and lines in \mathbb{P}^2 , expressed in equation (3.1), a point $\tilde{\mathbf{X}}$ lies \mathbb{P}^3 on a plane $\tilde{\mathbf{P}}$ if

$$\tilde{\mathbf{X}}_i^T \tilde{\mathbf{P}} = 0, \quad (\text{A.3})$$

that is, if the inner product of the 4-vectors representing the plane and the point is zero. The first three elements of the homogeneous vector $\tilde{\mathbf{P}}$ define the surface normal of the plane, and therefore its orientation. The distance of the plane from the origin is determined by the ratio of the first three elements to the fourth element of the homogeneous vector.

Given the homogeneous coordinates $\tilde{\mathbf{X}}_1 \dots \tilde{\mathbf{X}}_n$ of $n \geq 3$ points on a plane, the homogeneous vector $\tilde{\mathbf{P}}$ representing the plane can be estimated by finding the least-squares solution

$$\mathbf{A}\tilde{\mathbf{P}} = \mathbf{0}, \quad (\text{A.4})$$

with

$$\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{X}}_1^T \\ \vdots \\ \tilde{\mathbf{X}}_n^T \end{bmatrix}. \quad (\text{A.5})$$

A.3.2 Robust Estimate

The least-squares solution of equation (A.4) is adequate if the $\tilde{\mathbf{X}}_n$ differ from the true point coordinates only by added Gaussian noise with zero mean. In practice, some of the points will be outliers to the Gaussian error distribution, that is they do not lie in the plane in the first place. For example, with a stereo system, outliers will occur if non-corresponding points are matched during disparity estimation. These outliers can severely distort the least-squares solution of equation (A.4).

In the present application, the reference plane has therefore been estimated using a so-called *robust* fit method, that is, a fit method which is insensitive to outliers. This method is based on the *Random Sample Consensus* (RANSAC) paradigm proposed by Fischler and Bolles [29] and consists of the steps outlined in algorithm A.1.

Given: A set of points $\{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n\}$.

1. Randomly select a subset of k points out of $\{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n\}$.
2. For this subset, solve equation (A.4) in a least-squares sense. This yields a solution $\tilde{\mathbf{P}}_i$ for the plane.
3. Find all the points in the whole set $\{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n\}$ that lie within a fixed distance d of the estimated plane $\tilde{\mathbf{P}}_i$. These points form the consensus set for the estimate $\tilde{\mathbf{P}}_i$.
4. Repeat the above steps N times.
5. Of the different solutions $\tilde{\mathbf{P}}_i$, select the one with the largest consensus set. Calculate the least-squares solution to equation (A.4) using all the points in this consensus set. This yields the final fit plane $\tilde{\mathbf{P}}$.

Algorithm A.1: *Robust fit of a plane to a set of points, based on the RANSAC paradigm [29].*

To perform the third step of the algorithm, which is finding the points that lie within a distance d of the estimated plane, it is useful to apply a rotation and a translation, which bring the estimated plane $\tilde{\mathbf{P}}_i$ onto the $Z = 0$ plane, to all points $\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n$. After this transformation, the consensus set is found by selecting the transformed points whose Z -coordinate lies in the interval $[-d, d]$.

For the plane fit in the present application, described in section 7.8.1, the largest consensus set out of $N = 10$ runs with subsets of $k = 1000$ out of a total of 34046 points was used. Points with $d \geq 1.5$ mm were classified as outliers. The largest consensus set, which was used for the final least-squares estimate, comprised 33156 or 97.4% of the total number of points.

B Estimating Planar Homographies

In order to perform the camera calibration procedure described in section 3.4.2, it is necessary to estimate a planar homography, that is a projective transform in \mathbb{P}^2 , from point correspondences. The estimation of a homography from point correspondences is a standard problem in computer vision and appropriate algorithms are described, for example, in [44]. For completeness, this section gives a brief summary of how to estimate a homography.

B.1 Problem statement

Given a set of points with coordinates $\mathbf{p}_i = (p_x, p_y)^T$ on a plane, and a set of corresponding points $\mathbf{m}_i = (m_x, m_y)^T$ that are related by a projective homography as

$$\tilde{\mathbf{m}}_i = \lambda \begin{pmatrix} m_x \\ m_y \\ 1 \end{pmatrix} = \mathbf{H}\tilde{\mathbf{p}}_i = \mathbf{H} \begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix}, \quad (\text{B.1})$$

the problem is to estimate the matrix \mathbf{H} describing the homography. For example, the points could correspond to two different images that are related by a projective transformation, or, as in the present application, to points on a plane in 3D space and their images obtained with a projective camera.

Leaving degenerate configurations aside, a solution for \mathbf{H} can be found if at least four point correspondences are known, because each correspondence gives two constraints on the eight degrees of freedom \mathbf{H} (3×3 minus one because projective transforms are only defined up to a scale factor).

Similar to the estimation of the camera parameters in section 3.4, the estimation of the homography consists of two main steps: First, an initial solution is found by stacking the constraints on \mathbf{H} imposed by the point correspondences and solving the resulting linear system. Second, this linear solution is used as an initial value for a non-linear optimization routine that minimizes a geometric error functional.

B.2 Initial Guess

Let the row vectors of \mathbf{H} be $\mathbf{H}_1^T, \mathbf{H}_2^T, \mathbf{H}_3^T$. Then (B.1) becomes

$$\tilde{\mathbf{m}}_i = \begin{pmatrix} u \\ v \\ w \end{pmatrix}_i = \begin{bmatrix} \mathbf{H}_1^T \\ \mathbf{H}_2^T \\ \mathbf{H}_3^T \end{bmatrix} \tilde{\mathbf{p}}_i \quad (\text{B.2})$$

Without loss of generality, it can be assumed that the homogeneous vector $\tilde{\mathbf{m}}$ is scaled such that $w = 1$. Then it is possible to introduce a factor of one into the equation $u - u = 0$ yielding

$$u - u = u - 1 \cdot u = u - wu = u - u\mathbf{H}_3^T \tilde{\mathbf{p}} = \tilde{\mathbf{p}}^T \mathbf{H}_1 - u\tilde{\mathbf{p}}^T \mathbf{H}_3 = 0. \quad (\text{B.3})$$

Similarly the equation $v - v = 0$ can be transformed into

$$\tilde{\mathbf{p}}^T \mathbf{H}_2 - v\tilde{\mathbf{p}}^T \mathbf{H}_3 = 0. \quad (\text{B.4})$$

Equations (B.3) and (B.4) are constraints on \mathbf{H} and can be written in matrix form as

$$\begin{bmatrix} \tilde{\mathbf{p}}^T & \mathbf{0}^T & -u\tilde{\mathbf{p}}^T \\ \mathbf{0}^T & \tilde{\mathbf{p}}^T & -v\tilde{\mathbf{p}}^T \end{bmatrix} \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_3 \end{pmatrix} = \mathbf{0}. \quad (\text{B.5})$$

For each pair of corresponding points one obtains a linear system of the form (B.5). These systems can be stacked to obtain the equation

$$\mathbf{L}\mathbf{H} = \mathbf{0}, \quad (\text{B.6})$$

where \mathbf{L} is a $2n \times 9$ matrix that contains two rows for each point correspondence, and where $\mathbf{H} = (\mathbf{H}_1^T, \mathbf{H}_2^T, \mathbf{H}_3^T)^T$ is a 9-vector with the elements of \mathbf{H} . Equation (B.6) can be solved for \mathbf{H} if at least four point correspondences are known. In general, there will be no exact solution due to noise in the point measurements, but the solution that minimizes $\|\mathbf{L}\mathbf{H}\|$ for a non-trivial \mathbf{H} can be found by a singular value decomposition of \mathbf{L} as the singular vector corresponding to the smallest singular value (see, for example Golub and van Loan [39]). As described by Hartley [42] for a similar type of estimation problem, the accuracy of the solution can be improved by a prior data normalization which makes the matrix \mathbf{L} better conditioned numerically.

B.3 Minimizing Geometric Error

The algebraic error functional $\|\mathbf{L}\mathbf{H}\|$ that is minimized by solving the linear system (B.6) in a least-squares sense has no intuitively meaningful geometrical interpretation. Therefore, it is common to use a non-linear optimization technique to minimize a geometrically meaningful error functional $e(\mathbf{H})$. The linear solution for equation (B.6) is used as a starting value for the non-linear optimization routine.

A suitable error functional is

$$e(\mathbf{H}) = \sum_i (\mathbf{m}_i - \check{\mathbf{m}}_i)^2, \quad (\text{B.7})$$

where the $\check{\mathbf{m}}_i$ are the \mathbf{p} transformed by \mathbf{H} as $\check{\mathbf{m}}_i = \frac{1}{\mathbf{H}_3^T \tilde{\mathbf{p}}} \begin{pmatrix} \mathbf{H}_1^T \tilde{\mathbf{p}} \\ \mathbf{H}_2^T \tilde{\mathbf{p}} \end{pmatrix}$. This choice of $e(\mathbf{H})$, called transfer error, is the sum of the squared Euclidean distances between the original points \mathbf{p}_i transformed by \mathbf{H} and their corresponding points \mathbf{m}_i as determined from the image.

The implementation of the camera calibration used in the present application, due to Bouguet [10], uses a gradient descent algorithm (see Press et al. [75]) to minimize the error functional (B.7).

Acknowledgements/Danksagung

Thanks go to ...

Bernd Jähne for giving me the opportunity to work on this interesting topic.

Kurt Roth for acting as the second referee.

Richard Hartley, for the opportunity to do part of this work in Canberra.

All members of the “Windkanaler” and image processing groups for the good company and many interesting discussions.

All the proofreaders, especially Madeleine.

My parents.

Vielen Dank !

Index

- absorptivity, 34
- accuracy
 - of reconstruction, 138, 149
- acquisition system, 104
- Aeolotron, 11, 101, 109
- algebraic error, 61, 157
- apparent
 - temperature, 94
- apparent temperature, 44
- area based stereo, 76
- blackbody, 31
 - laboratory, 104
 - Santa Barbara Infrared, 104
- bolometers, 41
- Boltzmann's constant, 32
- boresight alignment, 103
- calibration
 - camera geometry, 56
 - geometric, results, 118
 - process (camera geometry), 57
 - radiometric, 25, 43, 93, 94
 - target, 108
- camera calibration, 47, 56
- camera calibration matrix, 51
- camera center, 50
- camera projection matrix, 52, 125
- CCD-type cameras, 53
- center of projection, 50
- close-range, 23
- Color Imaging Slope Gauge, 17
- complex refractive index, 36
- cool skin of the ocean, 9
- correspondence problem, 75
- cross-correlation, 84
- dense stereo, 76
- depth-of-field, 102
- detector output, 43
- dielectric constant, 36
- dielectric function $\epsilon(\lambda)$, 36
- disparity, 75
 - map, 77
 - sub-pixel estimation, 87
 - underlying assumptions, 79
- disparity space image, 78
- distortion
 - correcting for, 55
 - non linear, 53, 54
 - radial, 54
 - tangential, 55
- duality, 49
- emissivity, 34
 - temperature measurements, 44
- epipolar constraint, 62
- epipolar line, 65, 76
- epipolar plane, 65
- Euler-Lagrange, 97
- exterior camera parameters, 52, 122
- feature-based stereo, 76
- field-of-view, 102
- focal length, 50
- focal plane, 50
- focal plane array, 41, 93
- foreshortened, 28
- FPA
 - focal plane array, 41
- frame grabber, 104
- fronto-parallel, 79
- fronto-parallel assumption, 80
- fundamental matrix, 62, 66
- Gaussian image pyramids, 89, 128
- geometric error, 61, 157
- grazing angles, 80

- greenhouse gases, 7
- greybody, 35
- Herschel, William, 30
- homogeneous, 47, 48
- homography, 49, 58
 - estimation of, 155
- homologous, 75, 79
- IAPSO, 25
- ideal diffuse, 29
- image coordinate system, 50
- image plane, 50
- image pyramids, 83
- imaging, 23
- index of refraction, 36
- infrared cameras, 101
- Intergovernmental Panel on Climate Change, 7
- interior camera parameters, 53, 55, 120
- irradiance, 27
- Kirchhoff's Law, 34
- Kirchhoff's law, 31
- Lambert's cosine law, 29
- Lambertian
 - assumption in matching, 80
 - surface, 29
- Laplace, 97
- membrane model, 93, 96, 97, 100
- microEnable 2, 104
- MWIR, 42
- noise equivalent temperature difference, 40
- non-uniformity correction, 93, 102
- Opacity, 81
- optical flow, 23
- outlier removal, 95
- passive, 23
 - remote sensing device, 22
- penetration depth, 38
- photon detector arrays, 41
- photon detectors, 41
- pinhole camera model, 50
- pinhole model, 47
- pixel pitch, 53
- Planck, 31
- Planck's constant, 32
- points at infinity, 49
- principal axis, 50
- quantum efficiency, 41
- QWIP, 42
- radial distortion, 54
- Radiance, 28
- radiant energy, 27
- Radiant exitance, 27
- radiant flux, 27
- Radiant intensity, 28
- radiant power, 27
- Radiometric calibration, 93
- radiometric calibration, 43, 94
- radiometric quantities, 26
- radiometry, 25
- RAID, 106
- RANSAC
 - Random sample consensus, 153
- Raytheon Amber Radiance, 101
- re-projection error, 61, 122
- rectification, 62, 67, 126
- Reflective Stereo Slope Gauge, 16
- reflectivity, 34
- regularization, 93
- remote-sensing, 23
- representation of matches, 77
- S.M.S. Hyäne, 13
- Santa Barbara Infrared Inc., 104
- Schmidt number scaling, 11
- Selective emitters, 35
- shape-from-shading, 17
- shiftable windows, 83
- skew, 53

- solid angle, 26
- spectral radiant energy, 30
- spherical coordinates, 26
- Stefan-Boltzmann law, 32
- stereo
 - area-based, 76
 - correspondence problem, 75
 - feature-based, 76
 - particle tracking velocimetry, 76
 - underlying assumptions, 79
- stereo particle tracking velocimetry, 76
- sub-pixel disparity, 87
- synchronization
 - of cameras, 106

- tangential distortion, 55
- temperature
 - apparent, 44, 95
 - measurement, 44
- texture, 130
 - assumption in matching, 82
- thermal detector arrays, 41
- Thermosensorik CMT 384, 101
- transmissivity, 34
- triangulation, 73
 - accuracy of, 138, 150, 151
 - figure, 13
- Tsukuba, 128

- verging angle, 103
- vignetting, 115

- Wien's displacement law, 32
- world coordinate system, 52, 122

Bibliography

- [1] L. P. Adams and J. D. Pos. *Model harbour wave form studies by short range photogrammetry*. *Photogrammetric Record*, 10(58):457–470, 1981.
- [2] L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert. *Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach*. Technical Report RR3874, INRIA, Sophia-Antipolis, France, January 2000.
- [3] G. Balschbach. *Untersuchungen statistischer und geometrischer Eigenschaften von Windwellen und ihrer Wechselwirkung*. PhD thesis, Universität Heidelberg, February 2000.
- [4] G. Balschbach, J. Klinke, and B. Jähne. *Multichannel shape from shading techniques for moving specular surfaces*. In *Proceedings of ECCV 98*, 1998.
- [5] M. L. Banner, I. S. F. Jones, and J. C. Trinder. *Wave number spectra of short gravity waves*. *J.Fluid Mech.*, 198:321–344, 1989.
- [6] E. Bock. *personal communication*, 2000.
- [7] E. Bock et al. *Overview of the CoOP experiments: Physical and chemical measurements parameterizing air-sea heat exchange*. In M. Donelan, editor, *Gas Transfer at Water Surfaces*. American Geophysical Union, 2002.
- [8] E. J. Bock and T. Hara. *Optical measurements of capillary-gravity wave spectra using a scanning laser slope gauge*. *Journal of Atmospheric and Oceanic Technology*, 12(2):395–403, April 1995.
- [9] E. J. Bock and T. Hara. *Relationship between air-sea gas transfer and short wind waves*. *J.Geophys.Res.*, 104(25821), 1999.
- [10] J.-Y. Bouguet. *Camera calibration toolbox for Matlab*. http://www.vision.caltech.edu/~bouguetj/calib_doc/index.html, , 2002.
- [11] D. C. Brown. *Decentering distortion of lenses*. *Photogrammetric Engineering*, 32(3):444–462, 1966.
- [12] D. C. Brown. *Lens distortion for close-range photogrammetry*. *Photogrammetric Engineering*, 37(8):855–866, 1971.

- [13] M. Z. Brown, D. Burschka, and G. D. Hager. *Advances in computational stereo*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, August 2003.
- [14] T. A. Clarke and J. G. Fryer. *The development of camera calibration methods and models*. *Photogrammetric Record*, (16(91)):51–66, April 1998.
- [15] L. J. Cote. *The directional spectrum of a wind generated sea as determined from data obtained by the stereo wave observation project*. *Meteorological Papers*, 2(6):1–88, 1960.
- [16] C. Cox and W. Munk. *Measurements of the roughness of the sea surface from photographs of the sun's glitter*. *Journal of the Optical Society of America*, 44(11):838–850, 1954.
- [17] C. Cox and W. Munk. *Statistics of the sea surface derived from sun glitter*. *Journal of Marine Research*, 13(2):198–227, 1954.
- [18] J. Cruset. *Photogrammetric measurement of the sea swell*. *Photogrammetria*, pages 122–125, 1953.
- [19] U. R. Dhond and J. K. Aggarwal. *Structure from stereo - a review*. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1489–1510, November/December 1989.
- [20] J. Dieter. *Analysis of small ocean wind waves by image sequence analysis of specular reflections*. PhD thesis, University of Heidelberg, Germany, May 1998.
- [21] J. Dieter, H. Lauer, and B. Jähne. *Measurements of slope statistics on a wind driven water surface*. In A. Gruen and H. Kahmen, editors, *Optical 3-D Measurement Techniques 4 - Applications in architecture, quality control, robotics, navigation, medical imaging and animation*, pages 357–364, 1997.
- [22] H. D. Downing and D. Williams. *Optical constants of water in the infrared*. *Journal of Geophysical Research*, 80(12):1656–1661, 1975.
- [23] V. A. Dulov, V. N. Kudryavtsev, and A. N. Bolshakov. *A field study of whitecap coverage and its modulations by energy containing surface waves*. In M. Donelan, E. Saltzman, and R. Wanninkhof, editors, *Gas Transfer at Water Surfaces*. American Geophysical Union, 2002.

- [24] D. Engelmann, C. S. Garbe, M. Stöhr, and B. Jähne. *Three dimensional flow dynamics beneath the air-water interface*. In *Proc. of the Symposium on the Wind-Driven Air-Sea Interface*, page 181, Sydney, Australia, 1999.
- [25] O. Faugeras. *Three-dimensional computer vision: A geometric viewpoint*. MIT Press, 1993.
- [26] O. Faugeras, B. Hotz, et al. *Real time correlation-based stereo: algorithm, implementations and applications*. Technical Report 2013, INRIA, August 1993.
- [27] O. Faugeras and Q.-T. Luong. *The geometry of multiple images*. MIT Press, 2001.
- [28] R. Feynman, R. B. Leighton, and M. Sands. *The Feynman lectures on physics*. Addison Wesley, 1964.
- [29] M. Fischler and R. Bolles. *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. *Communications of the ACM*, 24(6):381–395, 1981.
- [30] N. M. Frew. *The role of organic films in air-sea gas exchange*. In P. S. Liss and R. A. Duce, editors, *The Sea Surface and Global Change*, chapter 5, pages 121–171. Cambridge University Press, Cambridge, UK, 1997.
- [31] N. M. Frew, E. J. Bock, W. R. McGillis, A. V. Karachintsev, T. Hara, T. Münsterer, and B. Jähne. *Variation of air-water gas transfer with wind stress and surface viscoelasticity*. In B. Jähne and E. C. Monahan, editors, *Air-water Gas Transfer, selected papers from the Third International Symposium on Air-Water Gas Transfer*, Hanau, 1995.
- [32] A. Fusiello, A. Trucco, and A. Verri. *A compact algorithm for rectification of stereo pairs*. *Machine Vision and Applications*, (12):16–22, 2000.
- [33] D. Fuß. *Kombinierte Höhen- und Neigungsmessung von winderzeugten Wasserwellen (in preparation)*. PhD thesis, University of Heidelberg, Germany, 2004.
- [34] C. Garbe. *Entwicklung eines Systems zur dreidimensionalen Particle Tracking Velocimetry mit Genauigkeitsuntersuchungen und Anwendung bei Messungen in einem Wind-Wellen Kanal*. Master's thesis, University of Heidelberg, Heidelberg, Germany, 1998.

- [35] C. S. Garbe. *Measuring heat exchange processes at the air-water interface from thermographic image sequence analysis*. PhD thesis, University of Heidelberg, Heidelberg, Germany, December 2001.
- [36] C. S. Garbe, H. Haußecker, and B. Jähne. *Measuring the sea surface heat flux and probability distribution of surface renewal events*. In E. Saltzman, M. Donelan, W. Drennan, and R. Wanninkhof, editors, *Gas Transfer at Water Surfaces*, Geophysical Monograph. American Geophysical Union, 2001.
- [37] C. S. Garbe, B. Jähne, and H. Haußecker. *Measuring the sea surface heat flux and probability distribution of surface renewal events*. In M. Donelan, E. Saltzman, and R. Wanninkhof, editors, *Gas Transfer at Water Surfaces*, pages 109–114. American Geophysical Union, 2002.
- [38] J. Gluckman and S. K. Nayar. *Rectifying transformations that minimize resampling effects*. In *Proc. of IEEE Computer Vision and Pattern Recognition*, Kauai, Hawaii, December 2001.
- [39] G. H. Golub and C. F. van Loan. *Matrix computations*. John Hopkins University Press, Baltimore and London, 3rd edition, 1996.
- [40] V. Haltrin. *Fresnel reflection coefficient of very turbid waters*. In *Proceedings of the 5th International Conference: Remote Sensing for Marine and Coastal environments, San Diego*, volume II, 1998.
- [41] G. C. Harris and M. J. Stevens. *A combined corner and edge detector*. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
- [42] R. I. Hartley. *In defence of the 8-point algorithm*. In *ICCV*, Boston, June 1995.
- [43] R. I. Hartley and P. F. Sturm. *Triangulation*. *Computer Vision and Image Understanding*, 68(2):146–157, November 1997.
- [44] R. I. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, U.K., 2000.
- [45] H. Haußecker. *Radiation*. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications*, volume 1, chapter 2, pages 7–35. Academic Press, San Diego, CA, 1999.
- [46] H. Haußecker, U. Schimpf, C. S. Garbe, and B. Jähne. *Physics from IR image sequences: Quantitative analysis of transport models and parameters of air-sea*

- gas transfer*. In M. Donelan, E. Saltzman, W. Drennan, and R. Wanninkhof, editors, *Gas Transfer at Water Surfaces*, Geophysical Monograph. American Geophysical Union, 2001.
- [47] J. Heikkilä and O. Silven. *A four-step camera calibration procedure with implicit image correction*. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1106–1112, 1997.
- [48] J. Hewett. *Infrared cameras near consumer applications. Interview with Earl Lewis of FLIR systems*. *Opto and Laser Europe*, pages p15–17, January 2004.
- [49] N. K. Hojerslev. *Optical properties of sea water*. In *Oceanography*, volume 3 of *Landolt-Börnstein: Numerical data and functional relationships in science and technology (New Series)*, page 383. Springer, 1986.
- [50] G. C. Holst. *Testing and evaluation of infrared imaging systems*. SPIE Optical Engineering Press, 2nd edition, 1998.
- [51] G. C. Holst. *Common sense approach to thermal imaging*. SPIE Optical Engineering Press, Bellingham, WA, 2000.
- [52] B. K. P. Horn. *Robot vision*. MIT Press, Cambridge, MA, 1986.
- [53] IPCC. *Climate Change 2001: Synthesis report. A contribution of working groups I, II, and III to the third assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2001.
- [54] B. Jähne. *The Heidelberg Aeolotron air-sea interaction facility*. CD-ROM Aeon Verlag, Hanau, 2000.
- [55] B. Jähne. *personal communication*, 2000.
- [56] B. Jähne. *Digital image processing*. Springer-Verlag, Heidelberg, Germany, 5th edition, 2002.
- [57] B. Jähne, P. Geißler, and H. Haußecker. *Handbook of computer vision and applications*. Academic Press, San Diego, 1999.
- [58] B. Jähne, H. Haußecker, U. Schimpf, and G. Balschbach. *The Heidelberg Aeolotron - a new facility for laboratory investigations of small scale air-sea interaction*. In M. L. Banner, editor, *The Wind-Driven Air-Sea Interface: Electromagnetic and Acoustic Sensing, Wave Dynamics and Turbulent Fluxes*, Sydney, Australia, 1999.

- [59] B. Jähne, J. Klinke, and S. Waas. *Imaging of short ocean wind waves: a critical theoretical review*. *Journal of the Optical Society of America*, 11(8): 2197–2209, 1994.
- [60] B. Jähne, K. O. Münnich, R. Bösinger, A. Dutzi, W. Huber, and P. Libner. *On the parameters influencing air-water gas exchange*. *Journal of Geophysical Research*, 92(C2):1937–1949, 1987.
- [61] B. Jähne and K. Riemer. *Two-dimensional wave number spectra of small-scale water surface waves*. *J. Geophys. Res.*, 95:11531–11646, 1990.
- [62] C. Janßen. *Ein miniaturisiertes Endoskop-Stereomesssystem zur Strömungsvisualisierung in Kiesbetten*. Master's thesis, University of Heidelberg, Germany, July 2000.
- [63] B. Kinsman. *Wind waves*. Prentice Hall, Englewood Cliffs, 1965.
- [64] C. Kittel. *Introduction to solid state physics*. Wiley & Sons, New York, 1971.
- [65] M. Klar. *3-D particle-tracking velocimetry applied to turbulent open-channel flow over a gravel layer*. Master's thesis, University of Heidelberg, Germany, May 2001.
- [66] J. Klinke. *Optical measurements of small-scale wind generated water surface waves in the laboratory and the field*. PhD thesis, University of Heidelberg, Germany, 1996.
- [67] J. Klinke and B. Jähne. *Measurements of short ocean waves during the MBL ARI West Coast experiment*. In B. J. Monahan, editor, *Air-Water Gas Transfer - Selected papers from the Third International Symposium on Air-Water Gas Transfer*, pages 165–173, 1995.
- [68] E. Kohlschütter. *Die Forschungsreise S.M.S. Planet II. Stereophotogrammetrische Aufnahmen*. *Annalen der Hydrographie und Maritimen Meteorologie*, 34:220–227, 1906.
- [69] P. S. Liss and et.al., editors. *Solas science plan and implementation strategy*, April 2003.
- [70] P. S. Liss and L. Merlivat. *Air-sea gas exchange rates: Introduction and synthesis*. In P. Buat-Menard, editor, *The role of air-sea exchange in geochemical cycling*, pages 113–129. Reidel, Boston, MA, 1986.

- [71] C. Loop and Z. Zhang. *Computing rectifying homographies for stereo vision*. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 125–131, Fort Collins, CO, June 1999.
- [72] T. Luhmann and W. Tecklenburg. *Optische Messung der Wellentopographie*. Technical report, Fachhochschule Oldenburg, Institut für angewandte Photogrammetrie, 2000.
- [73] A. Mucha and B. Szczechowski. *Photogrammetrische Vermessung des Wellengangs an hydrotechnischen Modellen*. *Jenaer Rundschau*, (4), 1983.
- [74] M. Planck. *Distribution of energy in the spectrum*. *Annalen der Physik*, 4(3): 553–563, 1901.
- [75] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical recipes in C*. Cambridge University Press, Cambridge, MA, 2nd edition, 1992.
- [76] G. Redweik. *Untersuchung zur Eignung der digitalen Bildzuordnung für die Ableitung von Seegangsparemtern*. PhD thesis, Fachbereich Vermessungswesen der Universität Hannover, 1993.
- [77] Rottok. *Meereswellen-Beobachtungen*. *Annalen der Hydrographie und Maritimen Meteorologie*, (8):329–341, 1903.
- [78] F. Santel, C. Heipke, K. S., and H. Wegmann. *Image sequence matching for the determination of three-dimensional wave surfaces*. In *Proceedings of the ISPRS Commission V Symposium*, volume XXXIV, pages 596–600, Corfu, Greece, September 2002.
- [79] P. M. Saunders. *The temperature at the ocean-air interface*. *Journal of Atmospheric Sciences*, 24(3):269–273, 1967.
- [80] H. Scharr and J. Weickert. *An anisotropic diffusion algorithm with optimized rotation invariance*. In *DAGM*, pages 460–467, Kiel, Germany, September 2000.
- [81] D. Scharstein and R. Szeliski. *Stereo vision research page*. <http://www.middlebury.edu/stereo>, Middlebury College, Middlebury, VT, USA, 2002.
- [82] D. Scharstein and R. Szeliski. *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. *International Journal of Computer Vision*, (47(1)):7–42, 2002.

- [83] U. Schimpf. *personal communication*, 2003.
- [84] U. Schimpf, C. S. Garbe, and B. Jähne. *Investigation of the transport processes across the sea-surface microlayer by infrared imagery*. *Journal of Geophysical Research*, 2004. to appear.
- [85] H. Schultz. *Specular surface stereo: a new method for retrieving the shape of a water surface*. In L. Estep, editor, *Optics of the air-sea interface: theory and measurements*, number 1749, pages 283–294, 1992.
- [86] H. Schultz. *Shape reconstruction from multiple images of the ocean surface*. Technical report, Department of Computer Science, University of Massachusetts, Amherst, 1994.
- [87] A. Schumacher. *Die stereophotogrammetrischen Wellenaufnahmen der Deutschen Atlantischen Expedition*. *Ergänzungsheft 3 zur Zeitschrift der Gesellschaft für Erdkunde*, pages 105–120, 1928.
- [88] A. Schumacher. *Stereophotogrammetrische Wellenaufnahme mit schneller Bildfolge*. *Deutsche Hydrographische Zeitschrift*, 3(1/3):78–82, 1950.
- [89] C. E. Shannon. *The mathematical theory of information*. University of Illinois Press, Urbana, IL, 1949.
- [90] O. H. Shemdin and H. M. Tran. *Measuring short surface waves with stereophotography*. *Photogrammetric Eng. and Remote Sens.*, 93:311–316, 1992.
- [91] O. H. Shemdin, H. M. Tran, and S. C. Wu. *Directional measurements of short ocean waves with stereophotography*. *J. Geophys. Res.*, 93:13891–13901, 1988.
- [92] H. Spies. *Bewegungsdetektion und Geschwindigkeitsanalyse zur Untersuchung von Sedimentverlagerungen und Porenströmungen*. Master's thesis, University of Heidelberg, Germany, 1998.
- [93] H. Spies. *Analysing dynamic processes in range data sequences*. PhD thesis, University of Heidelberg, Heidelberg, Germany, July 2001.
- [94] H. Spies, H. Haussecker, and H. J. Köhler. *Material transport and structure changes at soil-water interfaces*. In *Geofilters*, pages 91–97, Warsaw, Poland, 2000.
- [95] R. H. Stewart. *Methods of satellite oceanography*. University of California Press, Berkeley, 1985.

- [96] D. J. Stilwell. *Directional energy spectra of the sea from photographs*. *J. Geophys. Res.*, 74:1974–1986, 1969.
- [97] T. Takahashi. *Net sea-air CO₂ flux over the global oceans: An improved estimate based on the sea-air pCO₂ difference*. In *Proceedings of the 2nd International Symposium CO₂ in the Oceans*, Tsukuba, Japan, 1999. Center for Global Environmental Research, National Institute for environmental Studies.
- [98] T. Takahashi and S. C. Sutherland. *Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects*. *Deep Sea Research II*, 49(1601), 2002.
- [99] D. Tschumperle. *PDEs based regularization of multivalued images and applications*. PhD thesis, Universite de Nice, Sophia Antipolis, France, 2002.
- [100] H. Vogel. *Gerthsen Physik*. Springer, 18. edition, 1995.
- [101] S. Waas. *Entwicklung eines Verfahrens zur Messung kombinierter Höhen- u. Neigungsverteilungen von Wasserwellen*. Master's thesis, University of Heidelberg, 1988.
- [102] S. Waas. *Combined slope-height measurements of short wind waves: First results from field and laboratory measurements*. In *Optics of the Air-Sea Interface: Theory and Measurements*, San Diego, 1992. SPIE's 1992 International Symposium.
- [103] S. Waas. *Entwicklung eines feldgängigen optischen Meßsystems zur stereoskopischen Messung von Wasseroberflächenwellen*. PhD thesis, Universität Heidelberg, 1992.
- [104] R. Wanninkhof. *Relationship between gas exchange and wind speed over the ocean*. *Journal of Geophysical Research*, 97(C5):7373–7382, 1992.
- [105] D. Wierzimok and B. Jähne. *3-dimensional particle tracking beneath a wind-stressed wavy water surface with image processing*. In D. Wierzimok and B. Jähne, editors, *5th Int. Symp. Flow Visualization*, Prague, August 1989.
- [106] C. J. Zappa, W. Asher, A. T. Jessup, J. Klinke, and S. Long. *Effect of microscale wave breaking on air-water gas transfer*. In M. Donelan, E. Saltzman, W. Drennan, and R. Wanninkhof, editors, *Gas Transfer at Water Surfaces*, Geophysical Monograph. American Geophysical Union, 2001.

- [107] Z. Zhang. *Flexible camera calibration by viewing a plane from unknown orientations*. In *Proceedings of the ICCV*, 1999.
- [108] Z. Zhang. *A flexible new technique for camera calibration*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000.
- [109] C. L. Zitnick and T. Kanade. *A cooperative algorithm for stereo matching and occlusion detection*. *PAMI*, 22(7):675–684, July 2000.