Universität Rostock

Traditio et Innovatio

A Virtual Object Point Model for the Calibration of Underwater Stereo Cameras to Recover Accurate 3D Information

Dissertation
zur
Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing)
der Fakultät für Informatik und Elektrotechnik
der Universität Rostock

vorgelegt von
Tim Dolereit, geb. am 17.02.1984 in Schwerin
aus Rostock

Rostock, den 22. Dezember 2017

urn:nbn:de:gbv:28-diss2018-0152-1

Gutachter:

Prof. Dr.-Ing. Bodo Urban
Universität Rostock, Fraunhofer IGD

Prof. Dr.-Ing. Reinhard Koch
Christian-Albrechts-Unversität zu Kiel

Prof. Dr.-Ing. Uwe von Lukas
Fraunhofer IGD

Datum der Verteidigung:

26. September 2018

Zusammenfassung

Maschinelles Sehen zielt auf die Informationsverarbeitung von Daten ab, die von bildgebenden Systemen aufgezeichnet werden. Im Vergleich zum konventionellen Maschinellen Sehen in Luft befindet sich das Maschinelle Sehen im Unterwasser-Bereich noch in einem frühen Entwicklungsstadium. Das Wasser stellt eine herausfordernde Umgebung dar, die sich neben der Dämpfung des Lichts durch Absorption und Streuung insbesondere durch die Lichtbrechung an Medienübergängen auszeichnet. Brechungseffekte stellen für die Gewinnung bemaßter 3D-Strukturen aus Unterwasser-Bilddaten ein Problem dar. Der Fokus dieser Dissertation liegt auf der Rückgewinnung von präzisen 3D-Informationen aus Unterwasser-Aufnahmen.

In dieser Dissertation werden die verschiedenen Grundkonzepte der Stereo-3D-Rekonstruktion für Luft durch eine explizite Berücksichtigung von Brechungseffekten für den Unterwasser-Bereich erweitert. Insbesondere sind dies Projektionen für die Koordinaten-Transformation zwischen 3D- und 2D-Räumen, die Kalibrierung des bildgebenden Systems und die Rückgewinnung von 3D-Koordinaten. Es wurde ein neuartiges Modell, hier „*Virtual Object Point*“-Modell genannt, entwickelt, um die notwendigen Erweiterungen zu realisieren. Es bildet das Fundament dieser Dissertation, zeichnet sich durch seine Integrierbarkeit in den geometrischen Teilprozess der Unterwasser-Bildentstehung aus und ermöglicht neuartige Entwicklungen auf dieser Basis.

Im Rahmen der Unterwasser-Stereo-3D-Rekonstruktion liegt der Fokus dieser Dissertation auf der Kalibrierung einer speziellen Klasse von bildgebenden Systemen unter Berücksichtigung von Lichtbrechung. Diese Systeme, hier „*Shared Flat Refractive System*“(*SFRS*) genannt, bestehen aus einer Stereo-Kamera und einem Sichtfenster, welches einer einzelnen flachen, transparenten, lichtbrechenden Scheibe entspricht. Die Parameter des *SFRS* modellieren den geometrischen Teilprozess der Unterwasser-Bildentstehung unter expliziter Berücksichtigung von Brechungseffekten. Die Untersuchungen haben gezeigt, dass die Eigenschaften des *SFRS*-Modells für die Kalibrierung vorteilhaft sind und es sich rentiert, das grundsätzliche Design eines bildgebenden Systems dementsprechend anzupassen. Ein weiteres Ergebnis dieser Dissertation ist die Entwicklung von verschiedenen Verfahren für die Kalibrierung der Parameter eines *SFRS*. Diese zeichnen sich dadurch aus, dass ein Teil der Parameter durch lineare Optimierung bestimmt werden können. Darüber hinaus konnte ein weiteres wesentliches Ziel, Einschränkungen durch die umständliche und zeitaufwendige Prozedur der Handhabung spezieller Kalibrier-Objekt unter Wasser zu minimieren oder aufzuheben, erreicht werden. Die entwickelten Verfahren für die Kalibrierung eines *SFRS* weisen entsprechend unterschiedliche Grade der Abhängigkeit von Kalibrierungsobjekten auf und reichen bis hin zur vollständigen Unabhängigkeit von diesen.

Abstract

Computer vision aims at processing information that is recorded with the aid of imaging devices. Compared to conventional computer vision in air, underwater computer vision is at an early stage of development. Water is a challenging environment, which, apart from the attenuation of light by absorption and scattering, is characterized in particular by the refraction of light at media transitions. Refractive effects are a problem for the recovery of metric 3D structure from underwater image data. The focus of this thesis is on recovering accurate 3D information from underwater images.

In this thesis, the concepts of stereo 3D reconstruction in air will be extended for underwater environments by an explicit consideration of refractive effects. These concepts comprise projections for coordinate transformations between 3D and 2D spaces, calibration of the imaging device and the actual recovery of 3D coordinates. A novel model, named *Virtual Object Point* model, was developed to realize the necessary extensions. It is the cornerstone of this thesis, is characterized by its integrability into the geometric sub-process of underwater image formation and enables novel developments based on this.

Within underwater stereo 3D reconstruction, the focus of this thesis is on the refractive calibration of a special class of underwater imaging systems. These systems, named *Shared Flat Refractive System* (*SFRS*), consist of a stereo camera and a viewing window, which is a single flat, transparent, refractive interface. The geometric sub-process of underwater image formation is modeled by the parameters of the *SFRS* with an explicit consideration of refractive effects. The performed investigations have shown that the particularities of the *SFRS* are beneficial for refractive calibrations and worth to adjust the basic design of the imaging system accordingly. A major contribution of this thesis is the development of various approaches for the calibration of the parameters of a *SFRS*. These approaches are characterized by the fact that part of the refractive parameters are determined by linear optimization. Furthermore, another crucial goal of this thesis, to relax or even eliminate restrictions due to the cumbersome and time-consuming procedure of handling special calibration objects under water, was achieved. Thus, the developed approaches for the refractive calibration of a *SFRS* show different degrees of dependence on calibration objects culminating in a full independence from those.

Contents

1 Introduction **1**
1.1 Motivation . . . 2
1.2 Challenges . . . 3
1.3 Research Goals . . . 4
1.4 Contributions . . . 4
1.5 Structure and Overview . . . 6

2 Basics **7**
2.1 Mathematics . . . 8
2.2 Computer Vision . . . 10
2.2.1 Image Formation . . . 10
2.2.2 3D Structure Recovery . . . 13
2.3 Physics . . . 15
2.3.1 Law of Refraction . . . 16
2.3.2 Ray Tracing Formulas . . . 17
2.3.3 Apparent Depth . . . 17
2.4 Implications of Refraction . . . 21
2.5 Definitions . . . 25
2.6 Summary . . . 28

3 Related Works **29**
3.1 Disregard of Refractive Effects . . . 30
3.2 Absorption of Refractive Effects . . . 32
3.3 Explicit Modeling of Refractive Effects . . . 36
3.3.1 Refractive Image Formation . . . 37
3.3.2 Refractive Indices . . . 41
3.3.3 Layer Thicknesses . . . 42
3.3.4 Specialized Calibration Approaches . . . 44
3.3.5 Flat Refractive Systems . . . 46
3.3.6 Flat Refractive Stereo Systems . . . 51
3.3.7 Shared Flat Refractive Systems . . . 54
3.3.8 Categorization and Aptitude Analysis . . . 55
3.4 Summary . . . 58

4 Virtual Object Points **59**
4.1 Motivation and Challenges . . . 60

4.2 Position Determination . . . 62
4.2.1 Single Flat Refractive Interface . . . 63
4.2.2 Two Flat Refractive Interfaces . . . 65
4.2.3 Selection . . . 68
4.3 Model for Two Flat Refractive Interfaces . . . 69
4.4 Summary . . . 72

5 Calibration of Shared Flat Refractive Systems 73
5.1 Requirements, Prerequisites and Restrictions . . . 74
5.2 Virtual Object Point Model and Monocular Vision . . . 75
5.2.1 Refractive Forward Projection . . . 77
5.2.2 Refractive Back-Projection . . . 79
5.2.3 Virtual Camera Model . . . 82
5.3 Virtual Object Point Model and Binocular Vision . . . 86
5.3.1 Computation of Virtual Object Points . . . 87
5.3.2 Computation of Real Object Points . . . 89
5.4 Definition and Computation of Cost Functions . . . 91
5.4.1 Pattern-Based Invariants . . . 91
5.4.2 Pattern-Based Cost Functions . . . 95
5.4.3 3D Scene Cost Functions . . . 97
5.4.4 2D Image Cost Functions . . . 100
5.5 Computation of Refractive Parameters . . . 104
5.5.1 Pattern-Based System Axis Determination . . . 104
5.5.2 Computation of the Air Layer Thickness . . . 107
5.5.3 Execution Sequence . . . 108
5.6 Summary . . . 111

6 Underwater Stereo 3D Reconstruction 113
6.1 Correspondence Curves . . . 114
6.2 Recovery of 3D Coordinates . . . 116
6.3 Summary . . . 118

7 Experiments & Results 119
7.1 Refractive Calibration - Simulated Test Data . . . 120
7.1.1 Stage 1 - Full Simulation of 3D-2D Point Correspondences . . . 120
7.1.2 Stage 2 - Rendered Images . . . 136
7.2 Refractive Calibration - Real Test Data . . . 141
7.2.1 Benchmark Data Generation . . . 141
7.2.2 Stage 1 - With Calibration Object . . . 142
7.2.3 Stage 2 - Without Calibration Object . . . 147
7.3 Application: Underwater Stereo 3D Reconstruction . . . 151
7.3.1 Correspondence Curves . . . 151
7.3.2 Dense Stereo Matching & Recovery of 3D Coordinates . . . 152
7.4 Summary . . . 155

8 Conclusion **157**
8.1 Summary 157
8.2 Future Works 161

Appendices **163**

A Number of Features **165**

B Simulated Test Data Stage 1 - Supplementary Setups **169**

C Effects of Inaccurate Refractive Indices **171**

D Simulated Test Data Stage 2 - Supplementary Setups **175**

Bibliography **179**

Own Publications **191**

Thesis Statements **193**

List of Figures

2.1 Orthogonal projection of a vector onto another vector. 9
2.2 Spherical coordinates. 9
2.3 The pinhole camera model. Left: 3D view of the mapping process. Right: 2D view of the image plane. 11
2.4 Transformation from world to camera coordinate system. 13
2.5 Refraction of an incident ray i at a flat interface Φ between two transparent media with different refractive indices n_1 and n_2. 16
2.6 Ray tracing starting from point C. 17
2.7 Apparent depth in the case of a perpendicular incident ray. 18
2.8 Derivation process of the *VOP* locations in the case of a non-perpendicular incident ray. 20
2.9 Relation between the location of the *VOP* $\check{O}$ and the real object point O in the case of a single refractive interface. 21
2.10 Left: Image formation based on the pinhole camera model in air. Right: Invalidity of the pinhole camera model for underwater image formation. . 22
2.11 Distance-dependence of refraction. 23
2.12 Epipolar lines an their invalidity in underwater stereo vision, illustrated by three orthographic views. 24
2.13 Left: Schematic illustration and parametrization of a *Flat Refractive System* (*FRS*). Right: Object axis s_1. 26
2.14 Left: Schematic illustration and parametrization of a *Flat Refractive Stereo System* (*FRSS*). Right: Schematic illustration and parametrization of a *SFRS*. 27

3.1 Left: Refractive back-projection of image point I to ray w in water. Right: Refractive forward projection of object point O to image point I. 37
3.2 Left: Caustic surface. Right: Axial camera. 39
3.3 Schematic representation: Calibration object (CO) fixed to the refractive interface Φ_1. A transformation $(\boldsymbol{R}, \boldsymbol{t})$ links the CO and the *FRS*. 46
3.4 Schematic representation: Calibration object (CO) in water. A transformation $(\boldsymbol{R}, \boldsymbol{t})$ links the CO and the *FRS*. 47
3.5 Schematic representation: Single (FRS_1) or several, fixed *FRS*s (FRS_i, with $i \in \{1, .., n\}$), a calibration object (CO) and the transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$, with $i \in \{1, .., m\}$) that link them. 49

3.6 Schematic representation: *FRSS* composed of two single *FRS*s (FRS_1 and FRS_2). Left: Linking by 3D reconstruction of object points O_i (with $i \in \{0, .., n\}$) in the scene. Right: Calibration object (CO) and the linking transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$, with $i \in \{0, .., m\}$). 52
3.7 Schematic representation: *SFRS* composed of single cameras (C_i with $i \in \{1, .., n\}$) and calibration objects (CO) on the refractive interface Φ_1 or in water and the linking transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$ with $i \in \{1, .., m\}$). . . . 54

4.1 The possible *VOP* locations $\check{O}_m$, $\check{O}_s$ and $\check{O}_M$. 62
4.2 Front view of a right circular cone that bounds the possible starting directions of an unknown optical path between a real object point O and a viewpoint V. 63
4.3 Traces of the *VOP* locations in the case of one refractive interface. Left: 2D front view. Right: Perspective 3D view. 64
4.4 *VOP* locations in the case of one refractive interface. Left: 2D front view. Right: Perspective 3D view. 65
4.5 Derivation process of the *VOP* locations in the case of two flat refractive interfaces. 66
4.6 Linear shift to the new *VOP* locations due to an additional refractive interface. 67
4.7 Traces of the *VOP* locations in the case of two refractive interfaces. Left: 2D front view. Right: Perspective 3D view. 68
4.8 Relation between the virtual and the real object points. Left: In the case of a single flat refractive interface. Right: In the case of two parallel flat refractive interfaces. 69

5.1 Combination of a *FRS* and the *VOP* model. 75
5.2 Refractive forward projection: Intermediate step of the iterative processing. 77
5.3 Refractive forward projection. Left: Final step of the iterative processing. Right: Schematic illustration of the bisection strategy. 79
5.4 Refractive back-projection. Left: Ray path (a, g and w) from full ray tracing. Right: Avoid full ray tracing by application of the *VOP* model to compute the ray w expressed by Equation 5.2. 80
5.5 Refractive back-projection: Computation of the direction vector $\boldsymbol{w}$ of ray w. 81
5.6 Virtual camera model: Left: Exemplary illustration of the virtual camera that is associated to the image point I. Right: Superimposition of the principal axis of the real and the virtual camera. 83
5.7 Left: Ambiguity of the virtual principal axis. Middle: Superimposition of the real and virtual image plane. Right: Resulting coordinate system of the virtual camera. 85
5.8 Combination of a *SFRS* and the *VOP* model. 86
5.9 Left: Plane of refraction Π belonging to the image point I. Right: Plane of refraction Π' belonging to the image point I'. 87
5.10 Intersection of the planes of refraction Π and Π'. 88

5.11 Computation of real object points from *VOP*s. 90
5.12 Left: Ray paths from the calibration object to the cameras. Right: Exemplary construction of a feature plane Ψ_i. 92
5.13 Left: Intersection of the object axes s_{ij} and of the feature planes Ψ_i and Ψ_j with the refractive interface Φ. Right: Orthogonal view of the refractive interface Φ. 93
5.14 Left: Computation of the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ for a hypothetical system axis h_0. Right: Erroneous resulting plane Θ from plane fitting into the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$. 96
5.15 Left: Schematic computation of the *RTE*. Right: The shortest distance between the two rays w_i and w'_i. 99
5.16 Schematic computation of the *VIE*. 100
5.17 Left: Schematic computation of the *RRE*. Right: Top view of the image planes. 102
5.18 Left: Schematic computation of the *VRE*. Top Right: Superimposition of the virtual and the real image planes. Bottom Right: Best fitting real object point $\hat{O}_i$ from triangulation of the rays w_i and w'_i. 103
5.19 Pattern-based system axis determination. Left: Initialization of the iterative processing and generation of hypothetical system axes h_{ij}. Middle: Intermediate result. Right: Result close to the ideal solution. 105
5.20 Schematic illustration of the derivation of the linear equations for the computation of the air layer thickness. 108
5.21 Search space for the system axis, which is bounded by a defined range of the spherical coordinates (purple section), with respect to the camera coordinate system. Left: Utilized spherical coordinates. Middle: Side view. Right: Top view. 110

6.1 Invalidity of epipolar geometry and rectification (Compare with Section 2.4). 115
6.2 Sampling of the ray w to compute a correspondence curve. 115
6.3 Recovery of 3D coordinates by *Virtual Intersection*. 117
6.4 Recovery of 3D coordinates by *Refractive Triangulation* 117

7.1 Maximally overlapping field of view at the location of a 3D point cloud. . 123
7.2 ***a*** **:** Representation of the normalized direction vector $\bar{\boldsymbol{s}}$ of a system axis s_0. ***b*** **:** Valid system axes (top view). ***c*** **:** Valid system axis (side view). ***d*** **:** Invalid system axis. 124
7.3 Inaccuracies of refractive calibration with *PAE*. 125
7.4 Plotting of the *PAE*. 125
7.5 Inaccuracies of refractive calibration with *PFE*. 126
7.6 Plotting of the *PFE*. 126
7.7 Inaccuracies of refractive calibration with *RTE*. 127
7.8 Plotting of the *RTE*. 127
7.9 Inaccuracies of refractive calibration with *VIE*. 128
7.10 Plotting of the *VIE*. 128

7.11 Inaccuracies of refractive calibration with *RRE*. 129
7.12 Plotting of the *RRE*. 129
7.13 Inaccuracies of refractive calibration with *VRE*. 130
7.14 Plotting of the *VRE*. 130
7.15 Range of the valid values of the air layer thickness. 132
7.16 Comparison of the inaccuracies for the setups with $\phi = -10^\circ$ and $\theta = 0^\circ$. Top row: $t_a = 5cm$. Middle row: $t_a = 15cm$. Bottom row: $t_a = 25cm$ . . 133
7.17 Comparison of the mean run-times for the setups with $\phi = -10^\circ$ and $\theta = 0^\circ$.133
7.18 Comparison of the inaccuracies for the setups with $\phi = -10^\circ$ and $\theta = -20^\circ$. Top row: $t_a = 5cm$. Middle row: $t_a = 15cm$. Bottom row: $t_a = 25cm$ . . 135
7.19 Comparison of the mean run-times for the setups with $\phi = -10^\circ$ and $\theta = -20^\circ$. 135
7.20 Directions of the length measurements on a checker pattern. 137
7.21 Installation for the evaluations with generation of benchmark data. 142
7.22 Acquisition of calibration data. 144
7.23 Refractive calibration without a known calibration object. Left: Arbitrary underwater scene. Right: Acquisition of calibration images. 148
7.24 Computed correspondence curves for the inner corners of a checker pattern in two exemplary views. 151
7.25 Aquarium setup for underwater stereo 3D reconstruction. 152
7.26 Real and virtual LEGO® model of a shipwreck. 153
7.27 Two exemplary views of the recovered point cloud of the imaged scene. . 154
7.28 3D view. Left: Ground truth 3D point cloud. Middle: Recovered 3D point cloud after manual adjustment. Right: Registration of both 3D point clouds.154
7.29 Side view. Left: Ground truth 3D point cloud. Middle: Recovered 3D point cloud after manual adjustment. Right: Registration of both 3D point clouds. 155
7.30 Top view. Left: Ground truth 3D point cloud. Middle: Recovered 3D point cloud after manual adjustment. Right: Registration of both 3D point clouds. 155

A.1 *PAE*: Accuracy improvement versus run-time. 166
A.2 *PFE*: Accuracy improvement versus run-time. 166
A.3 *RTE*: Accuracy improvement versus run-time. 166
A.4 *VIE*: Accuracy improvement versus run-time. 167
A.5 *RRE*: Accuracy improvement versus run-time. 167
A.6 *VRE*: Accuracy improvement versus run-time. 167

B.1 Analysis of the inaccuracies of refractive calibration with *PAE*, *PFE*, *RTE*, *VIE*, *RRE* and *VRE* for supplementary setups 170

C.1 Angle error for a setup with $\phi = -10^\circ, \theta = 0^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$. 172

C.2 Angle error for a setup with $\phi = -10^\circ, \theta = -20^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$. 172
C.3 Angle error for a setup with $\phi = -10^\circ, \theta = -20^\circ, t_a = 5cm$ and varying distance to the simulated point cloud d. Left: $d = 100cm$. Middle: $d = 200cm$. Right: $d = 500cm$. 172
C.4 Air layer thickness error for a setup with $\phi = -10^\circ, \theta = 0^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$. 173
C.5 Air layer thickness error for a setup with $\phi = -10^\circ, \theta = -20^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$. 173
C.6 Air layer thickness error for a setup with $\phi = -10^\circ, \theta = -20^\circ, t_a = 5cm$ and varying distance to the simulated point cloud d. Left: $d = 100cm$. Middle: $d = 200cm$. Right: $d = 500cm$. 173
C.7 3D error for a setup with $\phi = -10^\circ, \theta = 0^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$. 174
C.8 3D error for a setup with $\phi = -10^\circ, \theta = -20^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$. 174
C.9 3D error for a setup with $\phi = -10^\circ, \theta = -20^\circ, t_a = 5cm$ and varying distance to the simulated point cloud d. Left: $d = 100cm$. Middle: $d = 200cm$. Right: $d = 500cm$. 174

List of Tables

3.1 Categorization of calibration approaches for *FRS*s. 57

5.1 Labeling of elements in illustrations. 76
5.2 Labeling of vectors and distances, which is generally omitted in illustrations. 76

7.1 Error analysis of the recovered 3D point cloud from simulated image data of a setup with perpendicularity between the image sensors and the interface normal. 140
7.2 Error analysis of the recovered 3D point cloud from simulated image data of a general setup with no perpendicularity between the image sensors and the interface normal. 140
7.3 Error analysis of the recovered 3D point cloud from real image data. The preceding refractive calibration was performed with the aid of a known calibration object. 146
7.4 Error analysis of the recovered 3D point cloud from real image data (Part 1). The preceding refractive calibration was performed without a known calibration object. 149
7.5 Error analysis of the recovered 3D point cloud from real image data (Part 2). The preceding refractive calibration was performed without a known calibration object. 150

D.1 Error analysis of the recovered 3D point cloud from simulated image data of a setup with $d = 100cm$, $b = 10cm$. 176
D.2 Error analysis of the recovered 3D point cloud from simulated image data of a setup with $d = 200cm$, $b = 20cm$. 176
D.3 Error analysis of the recovered 3D point cloud from simulated image data of a setup with $d = 500cm$, $b = 50cm$. 177

Glossary

λ	scalar
$\alpha(\boldsymbol{x}_1, \boldsymbol{x}_2)$	angle between two Euclidean vectors
$\alpha, \beta, \gamma, ..$	angles
$\|\boldsymbol{x}\|$	Euclidean norm of a vector
$\|x\|$	absolute value of a real number
π	mathematical constant, commonly approximated as 3.14159
$a, b, c, ..$	real numbers, line, line segment or ray
$c = 3 \times 10^8 ms^{-1}$	speed of light in vacuum
c_n	speed of light in medium with refractive index n
n_a, n_g, n_w	refractive indices of air, glass and water
$d(.. , ..)$	Euclidean distance between two elements
$\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, ..	matrices
A, B, C, ..	point in 2D or 3D, which will be represented by a position vector $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, ..
$\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ ..	vectors
$\boldsymbol{x} = (x_1, x_2, .. , x_n)^T$	Euclidean vector of dimension n
$\tilde{\boldsymbol{x}} = (x_1, x_2, .. , x_n, 1)^T$	homogenous vector of dimension $n + 1$
$\bar{\boldsymbol{x}} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$	normalized Euclidean vector
Π, Φ	planes

Acronyms

CY	*Chen and Yang*
DLT	*Direct Linear Transformation*
FLANN	*Fast Library for Approximate Nearest Neighbors*
FRSS	*Flat Refractive Stereo System*
FRS	*Flat Refractive System*
PAE	*Plane Angles Error*
PFE	*Plane Fitting Error*
RRE	*Refractive Reprojection Error*
RTE	*Refractive Triangulation Error*
SAD	*Sum of Absolute Differences*
SFM	*Structure from Motion*
SFRS	*Shared Flat Refractive System*
SIFT	*Scale-Invariant Feature Transform*
VIE	*Virtual Intersection Error*
VOP	*Virtual Object Point*
VRE	*Virtual Reprojection Error*

1 Introduction

Three quarters of the surface of the planet Earth is covered by water and most of it is covered by oceans. Oceans play a role in everything, from driving weather to regulating temperature to the air we breathe and ultimately support all living organisms. Despite their importance, we know very little about our oceans and only a small part of the ocean floor is known in detail. There is mostly knowledge on shallower waters and about 95 percent of the oceans remain unexplored, unseen by human eyes. Therefore, ocean exploration is still pioneer work. Exploring and understanding underwater environments is a demanding task, which requires equipping ocean explorations with the latest exploration tools. One of these tools is underwater computer vision, which aims at processing information that is recorded with the aid of imaging devices, such as cameras. However, water is a challenging environment that requires huge efforts from both the technological and data-processing points of view. Compared to traditional computer vision, underwater computer vision is at an early stage of development.
This chapter will comprise a brief motivation of underwater computer vision and especially of underwater stereo 3D reconstruction. The challenges that arise in an underwater environment will be described, the research goals of this thesis will be defined and the contributions to underwater computer vision will be listed. The chapter will conclude with the structure and overview of this thesis.

1.1 Motivation

The application of imaging devices in underwater environments has become a common practice. Today, visual sensors are almost standardly installed on autonomous underwater vehicles or remotely operated vehicles and it is also possible to equip divers with them. The biggest advantages of underwater imaging are its non-destructive behavior toward marine life and its repeatable application. These and the detectable image features make underwater imaging to an efficient sampling tool.
Underwater imaging is utilized in many fields of application. To name a few: fish monitoring and the determination of the 3D structure of fish schools in the field of aquaculture, abundance and habitat surveys of the benthos and of the seabed, seafloor mapping and biomass estimation in the field of marine biology, archaeological site and shipwreck surveys in the field of underwater archeology, the quantitative analysis of the 3D human motion in the fields of sport science and biomechanics, inspections of ship hulls or the visual support for repairs on marine equipment. There is various information encoded in the images, such as photometric data, the geometric 3D structure of the imaged scene and the pose of the camera relative to the scene. From a computer vision point of view, the possible applications for the extraction of information from underwater images comprise, for example, change detection, motion estimation, vision-based navigation and location of underwater vehicles, the creation of 2D mosaics and 3D maps of the seafloor or 3D reconstruction and 3D measurements of submerged objects.
The focus of this thesis is on *recovering accurate 3D information from underwater images*. Recovered 3D information is useful for distance-dependent underwater image processing approaches, such as image enhancement and image restoration. In contrast to imaging in air, underwater images show a characteristic color cast, which is not only dependent on the constituents of the water body, but also on the distances to the imaging system and the available light sources. Therefore, the recovery of 3D information is necessary for a distance-dependent color correction of underwater images that is following physical laws. However, besides distance determination, the major purpose of recovered 3D information is to create metric digital models.
There are various computer vision techniques for 3D reconstruction. These differ primarily by the deployed hardware of the imaging system. In this thesis, the imaging system of choice is a stereo camera. In contrast to techniques like structured light or *Structure from Motion* (*SFM*), stereo 3D reconstruction is popular because of its capabilities to work without additional light sources and without the necessity of a moving camera or a rigid scene. However, it is also possible to do stereo 3D reconstruction during movements of the stereo camera. With a calibrated stereo camera, metric 3D models can be created, length measurements in 3D space can be performed and distances can be determined. Underwater stereo 3D reconstruction faces severely different challenges than stereo 3D reconstruction in air. As will be seen in the following chapter, this makes a straightforward application of well-established computer vision algorithms impossible and requires elaborate extensions for the underwater environment. Therefore, this thesis will become part of the very recent and active research field of underwater computer vision.

1.2 Challenges

From an equipment point of view, the first challenge of underwater computer vision stems from the preparation of the imaging system for an application under water. Its electrical parts need to be protected from water penetration. Common setups are cameras looking into an aquarium, cameras looking through a viewing window or cameras being placed inside a special waterproof housing. This leads to setups with transparent bounding interfaces, which are characterized in general by a water-glass-air transition of the entering light. While the walls of an aquarium are usually flat, the protecting interfaces in the case of underwater housings can be flat or dome ports. Flat ports are plane-parallel and cheaper to produce, but introduce severe distortions due to the refraction of light. Dome ports are hemispheres and require an exact coincidence of its center and the center of projection of the camera. If this can be assured, there is theoretically no refraction, since all of the entering light rays would have an angle of incidence of 90 degrees. This requirement is hard to fulfill in practice and deviations introduce refractive distortions that are severely more complex than those of flat ports. Since the production of dome ports is more complicated, they are more expensive. Furthermore, they introduce problems on determining the correct focus distance [MNR17]. The focus of this thesis will be on stereo cameras with flat refractive interfaces. The interested reader may find further information on dome ports in [Men+16].
Besides the refractive interfaces that stem from the preparation of the imaging system, natural refractive interfaces, such as the wavy water surface, and the water body itself result in further challenges. The major differences to imaging in air, which characterize underwater imaging, are:

- light attenuation due to absorption and scattering of light,
- view disturbances, such as moving animals, floating particles or air bubbles,
- photometric artifacts, such as flickering caustics due to the wavy water surface and
- the refraction of light due to the water-glass-air transition of the light.

In this thesis, from a computer vision point of view, refraction of light is of particular importance. While the first three aspects can mostly be handled by image pre-processing, such as restoration, enhancement, lighting correction and elimination of view disturbances, the occurring magnifications due to refraction pose problems to image acquisition and image analysis. Submerged objects seem to be closer to the observer and hence bigger than they actually are. As a consequence, the focus distance of the camera needs to be adjusted. Furthermore, the field of view of a camera decreases severely. From an image analysis point of view, magnification is also problematic. This refraction-induced deviation in dimension is dependent on the distance of the imaged objects to the refractive interface and on the incidence angle of light rays entering the camera through this interface. This dependency acts in a non-linear way and poses a problem to every discipline aiming at the recovery of metric 3D structure from image data.

1.3 Research Goals

With a stereo camera as the imaging system of choice, the aim of this thesis is to *extend the concepts of stereo 3D reconstruction in air for underwater environments by an explicit consideration of refractive effects.* This comprises the determination of a suitable model for image formation, which forms the basis for calibration, stereo matching and the recovery of 3D coordinates. The imaging system will be characterized by the model for a *Shared Flat Refractive System* (*SFRS*). Due to the aforementioned challenges of underwater imaging, an extension of the concepts of stereo 3D reconstruction in air for underwater environments is indispensable. Therefore, the research goals read as follows:

R1) Develop a model for handling refractive effects in all concepts of stereo 3D reconstruction. This basic model is supposed to benefit all concepts mentioned above, particularly with regard to handling refractive effects.

R2) Develop approaches for the refractive calibration of a *SFRS*. Within the concepts of stereo 3D reconstruction the focus will be on calibration of the imaging system, which is usually the first step of the recovery of metric 3D structure from image data. Calibration approaches are highly dependent on a suitable model for image formation. The underlying image formation model will be the model for a *SFRS*.

R3) Relax or even eliminate the need to handle special calibration objects in underwater environments during the refractive calibration of a *SFRS*. This relaxation or elimination primarily concerns restrictions that are imposed by calibration objects with necessary properties, such as a certain shape (3D frame or 2D pattern), known dimensions, known 3D coordinates of their features and combinations of those. It can only benefit the entirety of a measurement campaign if the cumbersome and time-consuming handling of a calibration object under water can be avoided.

1.4 Contributions

Derived from an in-depth review of the related literature, a taxonomy and a parametrization of the image formation models of imaging systems exposed to refraction will be proposed and complemented. The analyzed imaging systems are characterized by plane-parallel interfaces, which separate the camera or cameras from water. Since the focus of this thesis is on calibration of such an imaging system, the approaches from the literature will be categorized with respect to their capabilities of fulfilling the research goals. From the developments in this thesis, the following contributions will arise. These primarily comprise the fulfillment of the defined research goals from the previous section. The major contributions of this thesis read as follows:

C1) A *Virtual Object Point* (*VOP*) model will be introduced, which enables the extension of the concepts of stereo 3D reconstruction in air for

underwater environments. The model will be the cornerstone of this thesis. It will be based on a proposed concept for *VOP*s and will be readily integrable within the image formation models of monocular and binocular vision. In the context of image formation, the projections for coordinate transformations between 3D and 2D spaces are of major importance. The combined models will enable the implementation of multiple approaches for those projections and the derivation of invariant properties, which both will be used for the definition of cost functions in the context of refractive calibration. Besides refractive calibration, the *VOP* model will be applicable for stereo matching with the aid of correspondence curves and for the actual recovery of 3D coordinates. Hence, all the necessary concepts of stereo 3D reconstruction will be covered and contribution **C1** will therefore directly solve research goal **R1**.

C2) Various approaches for the refractive calibration of a *SFRS*, which differ due to their capabilities to work with or without the necessity of a special calibration object, will be proposed, analyzed and compared. In this thesis, the imaging system of choice will be characterized by the model for a *SFRS* and the main contribution will be approaches for its refractive calibration. It will be shown how the aforementioned projections and cost functions can be utilized for the optimization of the necessary parameters. Various approaches for the refractive calibration of a *SFRS* will be proposed, analyzed and compared. Hence, with these approaches, contribution **C2** will be able to solve research goal **R2**. The approaches will be characterized by their capabilities to work with or without the necessity of a special calibration object and by their capabilities to determine the necessary parameters consecutively or simultaneously. Therefore, with the different dependencies on calibration objects of the respective approaches, contribution **C2** will furthermore address and solve research goal **R3**. The characterization of the approaches will result in two strategies, which will be implemented with interchangeable cost functions. Besides being analyzed and compared amongst themselves, the proposed calibration strategies will be compared to alternative strategies from the literature. Therefore, extensive evaluations on simulated and real test data will be performed.

C3) The determination of the refractive parameters will be partly realizable by linear optimization. This major contribution of this thesis will concern both of the previously mentioned calibration strategies. Within both of them, the determination of the refractive parameters will be partly realized by linear optimization. Thus, solving research goal **R2** will be realized efficiently.

C4) An approach for benchmark data generation for real test data will be proposed. A further contribution will be an approach for benchmark data generation for real test data. This addresses the research goals only indirectly. However, it represents a simple means for the evaluation of underwater stereo 3D reconstruction.

1.5 Structure and Overview

Chapter 2 will cover the basic principles from the fields of mathematics, computer vision and physics that are necessary for the developments in this thesis. Since it is a conventional foundation of computer vision approaches, such as the concepts of stereo 3D reconstruction, image formation based on the pinhole camera model will be introduced. From the field of physics, refraction and its implications on computer vision approaches will be described. The most severe implication is the invalidity of pinhole camera model. Furthermore, necessary definitions will be listed, a taxonomy of refractive systems will be proposed and the refractive parameters of the image formation model utilized in this thesis, namely the *SFRS*, will be defined.
Chapter 3 will cover the related work on calibration strategies for *FRS*s. These are the disregard, the absorption and the explicit modeling of refractive effects. The shortcomings of them will be described and it will be shown that the explicit modeling of refractive effects is indispensable in general. The chapter will provide a categorization of the various approaches and of those that will be developed in this thesis. All of them will be categorized with respect to defined exclusion criteria.
In the context of underwater 3D reconstruction, the focus will be set particularly on the refractive calibration of a *SFRS*. Therefore, a *VOP* model that enables meeting the arising challenges will be motivated in Chapter 4. Driven by the findings from Chapters 2 and 3, this chapter will contain a specification of the contributions of the *VOP* model. Furthermore, the proposed concept of *VOP*s will be introduced. Since there are multiple valid possibilities for the location of a *VOP*, the chapter will comprise the derivation of the most suitable one, which will be used to define the *VOP* model.
Chapter 5 will cover the refractive calibration of a *SFRS* and therefore the main contribution of this thesis. It will be shown how the *VOP* model can be readily integrated into the underwater image formation models of monocular and binocular vision for the development of two strategies for the computation of the refractive parameters, with the particularity that both strategies comprise a determination of the parameters that can be partly realized by linear optimization. Moreover, the requirements, prerequisites and restrictions of the approaches to be developed will be described.
Chapter 6 will cover how the *VOP* model can be utilized for the necessary extensions of the remaining concepts of stereo 3D reconstruction in air for underwater environments. An approach for the computation of correspondence curves will be proposed, which enables the reduction of the search space during stereo matching. Furthermore, two approaches for the recovery of 3D coordinates will be proposed, which can be used to compute the 3D structure of an imaged scene.
Chapter 7 will cover the experiments and results of the proposed approaches on simulated and real test data. It will primarily cover the evaluation of the two proposed strategies for the refractive calibration of a *SFRS*. Therefore, these strategies will be evaluated by comparisons amongst themselves, with a state-of-the-art approach and with further strategies from the literature. Furthermore, the most promising one will be applied during first experiments on underwater stereo 3D reconstruction of a real scene.
Finally, Chapter 8 will cover the conclusion and future works.

2 Basics

The purpose of this chapter is to introduce the basic principles from the fields of mathematics, computer vision and physics that are necessary to understand the given problem this thesis is confronted with and the proposed solutions. Of particular importance are the consequences of physical phenomena, such as refraction of light, for the applicability of conventional computer vision approaches that are made for aerial usage in underwater environments. Before the description of the conventional foundations of computer vision, the recurring mathematical expressions will be provided. One of the most important foundations of many computer vision approaches is the image formation based on the pinhole camera model. To be able to recover 3D structure from image data, it is necessary to know how image formation works. Hence, the perspective projection based on the pinhole camera model will be explained. In this thesis, a stereo camera is the imaging system of choice for the recovery of 3D structure. Therefore, the basic concepts of stereo 3D reconstruction will be described to introduce the challenges that arise from an extension for underwater environments. As mentioned before, the physical phenomenon of the refraction of light poses some major challenges. From the field of physics, the law of refraction, its application in the equations for ray tracing and a model for the apparent depth will be presented. While the latter two will be utilized for the developments in this thesis, the basics on refraction emphasize some major implications, such as the invalidity of the pinhole camera model for underwater image formation. Some important consequences will be shown. The chapter concludes with the definition of a model for image formation that considers refraction explicitly. The contents regarding the implications of refraction and the definitions in this chapter have been partially proposed in Dolereit et al. [DL16; Sch+15; Luk+15].

2.1 Mathematics

This section provides a brief overview on recurring mathematical expression in this thesis. These are based on a standard reference [BS79] while being expressed in the style that is proposed in the glossary.

Angle between two Planes. The angle between two planes in 3D space can be computed from the normal vectors of both planes. With the aid of the definition of the scalar product, this is expressed as the angle between two vectors by:

$$\alpha(\bar{\boldsymbol{n}}_1, \bar{\boldsymbol{n}}_2) = \arccos(\bar{\boldsymbol{n}}_1 \cdot \bar{\boldsymbol{n}}_2), \tag{2.1}$$

with $\bar{\boldsymbol{n}}_1$ and $\bar{\boldsymbol{n}}_2$ being the normalized normal vectors of the respective planes.

Distance between two Points. The Euclidean distance between two points A and B, represented by their position vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, is expressed by:

$$d(A, B) = d(\boldsymbol{a}, \boldsymbol{b}) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}, \tag{2.2}$$

with $n \in (2, 3)$.

Distance between Point and Line. The Euclidean distance between a line $l : \boldsymbol{a} = \boldsymbol{b} + \lambda \boldsymbol{l}$, with $\boldsymbol{b}$ being the position vector to a point on the line, λ being a scalar and $\boldsymbol{l}$ being the direction vector of the line, and a point C, represented by position vector $\boldsymbol{c}$, is expressed by:

$$d(C, l) = \frac{\|(\boldsymbol{c} - \boldsymbol{b}) \times \boldsymbol{l}\|}{\|\boldsymbol{l}\|}. \tag{2.3}$$

Distance between two Lines. For a line $l_1 : \boldsymbol{a} = \boldsymbol{b}_1 + \lambda_1 \boldsymbol{l}_1$ and a line $l_2 : \boldsymbol{a} = \boldsymbol{b}_2 + \lambda_2 \boldsymbol{l}_2$, with $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ being the position vectors to a point on the line, λ_1 and λ_2 being scalars and $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ being the direction vectors of both lines, their Euclidean distance can be computed by:

$$d(l_1, l_2) = \frac{|(\boldsymbol{b}_1 - \boldsymbol{b}_2) \cdot \boldsymbol{n}|}{\|\boldsymbol{n}\|}, \tag{2.4}$$

with $\boldsymbol{n} = \boldsymbol{l}_1 \times \boldsymbol{l}_2$.

Distance between Point and Plane. The Euclidean distance between a point A, represented by the position vector $\boldsymbol{a}$, and a plane Π can be computed easily if the Hesse normal form of the equation of the plane is available. Rearranging the Hesse normal form $\Pi : \bar{\boldsymbol{n}} \cdot \boldsymbol{b} = d$, with $\bar{\boldsymbol{n}}$ being the normalized normal vector of the plane, $\boldsymbol{b}$ being the position vector to a point on the plane and d being the distance in normal direction to the coordinate system origin, yields:

$$d(A, \Pi) = \bar{\boldsymbol{n}} \cdot \boldsymbol{a} - d, \tag{2.5}$$

whit substitution of $\boldsymbol{b}$ by $\boldsymbol{a}$.

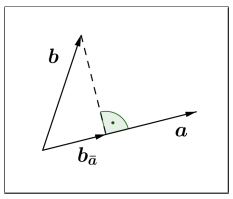

Figure 2.1: Orthogonal projection of a vector onto another vector.

Projection of a Vector onto a Vector. The orthogonal projection of a vector $\boldsymbol{b}$ onto a vector $\boldsymbol{a}$ (See Figure 2.1) is expressed by:

$$\boldsymbol{b}_{\bar{\boldsymbol{a}}} = (\boldsymbol{b} \cdot \bar{\boldsymbol{a}})\, \bar{\boldsymbol{a}}. \tag{2.6}$$

Therefore, the length of the vector $\boldsymbol{b}_{\bar{\boldsymbol{a}}}$ is:

$$\|\boldsymbol{b}_{\bar{\boldsymbol{a}}}\| = \boldsymbol{b} \cdot \bar{\boldsymbol{a}}. \tag{2.7}$$

Spherical Coordinates. The spherical coordinates can be used to express the direction of a vector in 3D space by three numbers (See Figure 2.2). Therefore, a point A on the surface of a unit sphere is utilized. The first number is the radial distance on the unit sphere. The second number is the angle φ, with a range of $-\pi < \varphi \leq \pi$, and the third number is the angle ϑ, with a range of $0 \leq \vartheta \leq \pi$. The conversion to Cartesian coordinates can be realized by:

$$\begin{aligned} x &= r \cdot \sin\vartheta \cdot \cos\varphi \\ y &= r \cdot \sin\vartheta \cdot \sin\varphi \\ z &= r \cdot \cos\vartheta. \end{aligned} \tag{2.8}$$

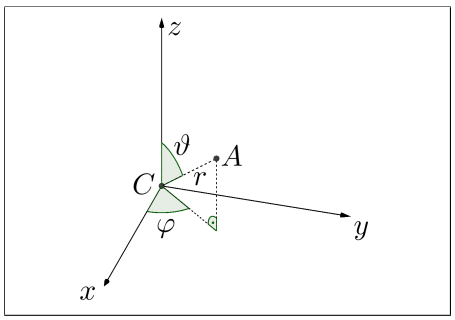

Figure 2.2: Spherical coordinates.

2.2 Computer Vision

From the point of view of computer vision, this thesis comes into contact with the fields image formation and 3D structure recovery. The system of choice is a stereo camera. To be able to perform tasks like underwater stereo 3D reconstruction, it is essential to understand its conventional foundations. One of them is the pinhole camera model, which is crucial to most conventional 3D reconstruction approaches. If the reconstructed data is supposed to be metric, the camera systems needs to be calibrated to determine the parameters of the image formation process and a scale. Most calibration approaches for a common camera system are founded on the pinhole camera model. After modeling and parametrization of the image formation process, it is possible to recover the metric 3D structure of the scene if additional cues, such as corresponding pixels in the two views of a stereo camera, are available.

2.2.1 Image Formation

The pinhole camera model describes the forward projection of 3D points to the image plane and can therefore be used to approximate the image formation of a camera. This mapping of 3D points to 2D points needs to be inverted for 3D structure recovery. Since forward projection results in a loss of information, the 3D point can not be recovered directly. However, a mapping of 2D points to 3D rays is possible. This mapping process is called back-projection. As will be seen in the following, the basic building blocks of image formation in air are rays and thus the collinearity of points.

Pinhole Camera Model

According to Hartley and Zisserman [HZ04], the pinhole camera model describes the central projection mapping of points in 3D space onto a plane as shown on the left side of Figure 2.3. The center of projection C forms the origin of a Euclidean 3D coordinate system. The principal axis p of the camera coincides with the z-axis of this camera coordinate system. The plane is called image plane and is located perpendicularly at $z = f$, with f being the focal length of the camera. The point where the principal axis p intersects this plane is called principal point P_1. On the left side of Figure 2.3, the image plane is located in positive direction of the principal axis, since this is mathematically more convenient. The physical image plane is the image sensor of the camera at $z = -f$. It is represented by the base of the right pyramid with an apex at C and with a principal point that is now P_2. Both representations can be used equivalently under correct consideration of algebraic signs. The point in space to be projected is O. It can be expressed by a position vector $\boldsymbol{o} = (x, y, z)^T$. By intersection with the image plane, this projection finally results in an image point I, which is represented by a position vector $\boldsymbol{i} = (u, v)^T$ in the 2D image coordinate system in pixel units (See right side of Figure 2.3). The object point O, the center of projection C and the associated Euclidean points I_1 and I_2 are collinear. As the points I_1 and I_2 result in the same image coordinates, point I_2 will be used in this document for beneficial illustration purposes and the subscript will

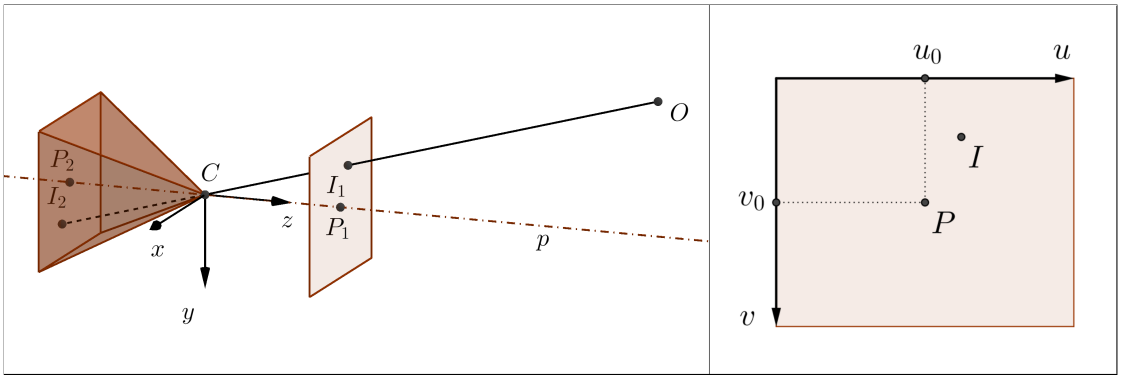

Figure 2.3: The pinhole camera model. Left: 3D view of the mapping process. Right: 2D view of the image plane.

be omitted. The same applies to the points P_1 and P_2. The central projection can be expressed as a linear mapping between the homogenous vectors of the points O and I:

$$\tilde{\boldsymbol{i}} = \boldsymbol{P}\tilde{\boldsymbol{o}}. \tag{2.9}$$

In this expression, a homogenous vector $\tilde{\boldsymbol{o}} = (x, y, z, 1)^T$ in the 3D space of the camera coordinate system is mapped to a homogenous vector $\tilde{\boldsymbol{i}} = (u, v, 1)^T$ in the 2D space of the image coordinate system in terms of a matrix multiplication with the 3×4 camera projection matrix:

$$\boldsymbol{P} = \boldsymbol{K}\left[\boldsymbol{I}|\boldsymbol{0}\right]. \tag{2.10}$$

This matrix is composed of the 3×3 identity matrix $\boldsymbol{I}$, the zero vector $\boldsymbol{0}$ and the 3×3 camera calibration matrix $\boldsymbol{K}$. The camera calibration matrix realizes the mapping from camera to image coordinates and has the form

$$\boldsymbol{K} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.11}$$

The entries $f_u = m_u f$ and $f_v = m_v f$ are the focal length of the camera scaled to pixel units. The scaling factors m_u and m_v differ if the image pixels are not square. The entries u_0 and v_0 are the coordinates of the principal point P in the image coordinate system. As the principal point is usually not the origin of this coordinate system, its coordinates determine the offset that needs to be added during the mapping as is shown on the right side of Figure 2.3. Equation 2.9 can now be written as:

$$\tilde{\boldsymbol{i}} = \boldsymbol{K}\left[\boldsymbol{I}|\boldsymbol{0}\right]\tilde{\boldsymbol{o}}, \tag{2.12}$$

which is the basic mapping of the pinhole camera model of a 3D point in the camera coordinate system. It expands to:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{2.13}$$

The result are the homogenous coordinates in the image coordinate system:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f_u x}{z} + u_0 \\ \frac{f_v y}{z} + v_0 \\ 1 \end{bmatrix}. \tag{2.14}$$

Forward Projection

In the general case, the 3D point to be projected onto the image plane is not expressed in the camera coordinate system, but in a world coordinate system. Both coordinate systems are related by a 3×3 rotation matrix $\boldsymbol{R}$ and a translation vector $\boldsymbol{t}$ as is shown in Figure 2.4. The origin of the world coordinate system is W. Let $\boldsymbol{o}'$ be the position vector to the point to be projected and let $\boldsymbol{c}'$ be the position vector to the center of projection C in the world coordinate system. The respective position vector $\boldsymbol{o}$ in camera coordinates can then be written as $\boldsymbol{o} = \boldsymbol{R}(\boldsymbol{o}' - \boldsymbol{c}')$. This can be expressed in homogenous coordinates and with that, Equation 2.12 can be written as:

$$\tilde{\boldsymbol{i}} = \boldsymbol{K}\boldsymbol{R}\left[\boldsymbol{I}| - \boldsymbol{c}'\right]\tilde{\boldsymbol{o}}'. \tag{2.15}$$

Together with $\boldsymbol{t} = -\boldsymbol{R}\boldsymbol{c}'$, the general camera projection matrix in Equation 2.10 becomes:

$$\boldsymbol{P} = \boldsymbol{K}\left[\boldsymbol{R}|\boldsymbol{t}\right]. \tag{2.16}$$

Back-Projection

The inversion of forward projection is called back-projection. Since the z-coordinate of a mapped 3D point is dropped in Equation 2.14, it is only possible to map a 2D image point to a ray in 3D space. The distance along this ray can not be recovered from image coordinates. A ray in the camera coordinate system can be represented by a normalized vector. This normalized vector $\bar{\boldsymbol{o}}$ can be computed from an homogenous vector $\tilde{\boldsymbol{i}}$ by multiplication with the inverted camera calibration matrix $\boldsymbol{K}$:

$$\bar{\boldsymbol{o}} = \frac{\boldsymbol{K}^{-1}\tilde{\boldsymbol{i}}}{\|\boldsymbol{K}^{-1}\tilde{\boldsymbol{i}}\|}. \tag{2.17}$$

The normalization is done by division by the resulting vector's length.

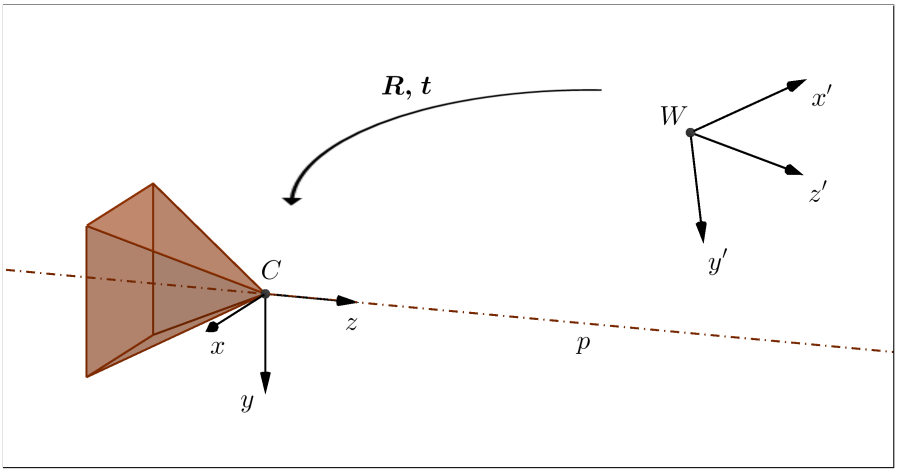

Figure 2.4: Transformation from world to camera coordinate system.

2.2.2 3D Structure Recovery

With the definition of the underlying camera model and the basic projection equations for transformations between 3D and 2D spaces, it is possible to recover the 3D structure of the imaged scene. Therefore, cues like motion, binocular disparity, shading, texture or focus are exploited [Sze11]. According approaches are *SFM*, stereo 3D reconstruction, or approaches called shape from X. In contrast to these passive methods, there are active methods like structured light. Since the research goal of this thesis is to extend the concepts of stereo 3D reconstruction in air for underwater environments, the according concepts will be described briefly in the following.

Stereo 3D Reconstruction

The purpose of 3D reconstruction is to generate a digital 3D model of the imaged scene. In the simplest case, this model can be a 3D point cloud, which contains a number of 3D points that are characterized by their 3D coordinates only. More sophisticated models represent surfaces by triangulated meshes and also with additional textures. In this thesis, the reconstruction process will end with the 3D point cloud, since surface generation can be considered as a post-process.

Cost Functions. An important part is the definition of cost functions. The values of cost functions are used for feedback on the accuracy of parameter estimation during calibration of the imaging system and can therefore be utilized as the target of an optimization problem. The most common cost function is the reprojection error, which is computed by forward projection of 3D scene points. It is based on 3D-2D point correspondences. The distance between a 2D point of interest in the image and the

projection of the corresponding 3D point onto this image is measured in some defined metric. This is usually in pixel units. The reprojection error is therefore defined by:

$$E = \sum_{i=1}^{k} d\Big(I_i, \hat{I}_i\Big)^2 + d\left(I_i', \hat{I}_i'\right)^2, \tag{2.18}$$

with k being the number of 3D-2D point correspondences, I_i and I_i' being the measured left and right image points and $\hat{I}_i$ and $\hat{I}_i'$ being the projected left and right image points [HZ04].

Calibration. The calibration of the stereo camera is usually the first step. Based on the underlying pinhole camera model, the intrinsic and extrinsic camera parameters can be estimated. The intrinsic camera parameters are described by the camera calibration matrix $\boldsymbol{K}$ in Equation 2.11. As no real lens is expected to fulfill the linear mapping of the pinhole camera model perfectly, some non-linear terms accounting for radial and tangential distortion are usually incorporated into the calibration process. These distortions can be modeled in the image space and thus are independent of the scene. The distortion parameters are commonly included by the term intrinsic camera parameters. The extrinsic camera parameters represent the pose (rotation and translation) of a camera in a reference coordinate system. If a world coordinate system is known, they can be described by the rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{t}$ in Equation 2.15. However, it is more important to estimate the relative extrinsic camera parameters of both cameras in a stereo system. In this case, one of the cameras is defined to be the master camera, which provides the reference coordinate system. The pose of the slave camera needs to be known relative to the master camera. If these relative extrinsic parameters can be determined in Cartesian coordinates, the scale for the 3D reconstruction result is known as well.
As will be shown in the following, the calibration of the stereo camera itself is unaffected by the underwater environment, since it can be realized in a pre-process. The parameters in question are usually estimated by solving an optimization problem through minimization of a cost function like the reprojection error (See Equation 2.18). A detailed discussion on calibration approaches, such as the *Direct Linear Transformation* (*DLT*) or the approach of [Zha00] can be found in [HZ04]. Readily usable algorithms can be found in the *Camera Calibration Toolbox for Matlab* [Bou08] and *OpenCV* [Bra00].

Stereo Matching. Stereo matching aims at finding corresponding pixels in the left and the right view of the stereo camera. It can be distinguished into sparse and dense stereo matching. An important step is the reduction of the search space, since it is unintended and inefficient to compare a pixel in the first view with all the others in the second view. This applies especially for dense stereo matching, which is characterized by matching every single pixel. Such a reduction can be realized with the aid of epipolar geometry and image rectification. In the case of sparse stereo matching, only distinctive pixels are matched in both views. Consequently, feature detection approaches, such as

corner detection (For example [HS88]) or the *Scale-Invariant Feature Transform* (*SIFT*) [Low99], are applied on both views and only the related pixels are matched.

Epipolar Geometry. The epipolar geometry describes the geometry in a stereo camera and is used to reduce the search space during dense stereo matching. With its aid, the search space for a corresponding pixel in the second view can be reduced to a single line [Sch05]. A further simplification can be achieved by rectification of both views. This comprises a transformation of both image planes into a common plane. Afterward, this single line coincides with a single row of pixels that is the same for both views. This makes it possible to characterize a pair of corresponding pixels by the so-called disparity. As will be seen in the following, epipolar geometry is invalid under water. Therefore, a detailed description is omitted. At this point, the interested reader may be referred to [HZ04] or [Sch05].

Recovery of 3D Coordinates. The approach for the recovery of 3D coordinates depends on the stereo matching technique. After dense stereo matching, it is possible to acquire the 3D coordinates from pixel positions and disparity. Since this approach is invalid under water, the interested reader may be referred to [Sch05] again. The usual approach to recover 3D points after sparse matching is known as triangulation [HZ04]. Therefore, two rays, which are the result of back-projection of a pair of corresponding pixels, are intersected in 3D space. Because of inaccuracies during calibration and stereo matching, these rays are not guaranteed to intersect in general. Hence, the approach has to be extended to account for these inaccuracies. One simple possibility is to compute the points on both rays with the shortest distance to each other and take the mean of both points as the recovered 3D coordinate. As will be seen in the following, the basic triangulation assumption can be reused in underwater stereo 3D reconstruction.

2.3 Physics

The previous section on computer vision has shown that the underlying pinhole camera model assumes collinearity between an object point, the center of projection and the image point. Hence, lines or rays are utilized frequently. The modeled rays correspond essentially to the light rays that are emitted from real object points. A phenomenon known from physics is the refraction of light at the transition between different media. A waterproof stereo camera, which is looking into water, is usually exposed to a water-glass-air transition of the light rays. As will be seen in the following sections, refraction has some major consequences on the applicability of the conventional computer vision approaches made for aerial environments. This section will provide the basic physical laws that need to be considered during underwater 3D reconstruction, such as the law of refraction. Furthermore, concepts like the apparent depth and formulas for physically correct tracing of rays, which will be part of the solutions to be developed in this thesis, will be introduced.

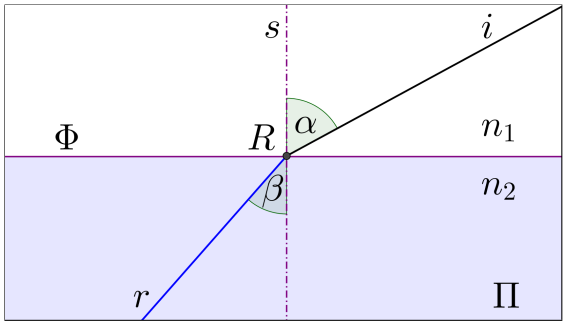

Figure 2.5: Refraction of an incident ray i at a flat interface Φ between two transparent media with different refractive indices n_1 and n_2.

2.3.1 Law of Refraction

According to Tipler et al. [TMB12], refraction is known as the effect that occurs when light hits an interface Φ between two different transparent media non-perpendicularly, as illustrated in Figure 2.5. A transparent medium is characterized by its refractive index. This is the relation of the speed of light in vacuum c to the speed of light in a medium c_n with the refractive index n:

$$n = \frac{c}{c_n}. \tag{2.19}$$

The angle α between the incident ray i and the interface normal s is called angle of incidence. The plane formed by both these lines is called plane of refraction Π. Figure 2.5 represents a perpendicular view onto the plane Π. The light ray that enters the second medium is called the refracted ray r. This ray is situated in the plane of refraction as well. Its angle β to the interface normal s is called angle of refraction. The light ray r in the second medium has a different direction than the incident ray i. If the propagation speed of light is decreased at the transition of the interface Φ, the angle of refraction β is smaller than the angle of incidence α and the refracted ray r is refracted towards the interface normal s and vice versa. The angle of incidence α and the angle of refraction β are related by the law of refraction, also known as Snell's law:

$$n_1 \cdot \sin\alpha = n_2 \cdot \sin\beta \tag{2.20}$$

It can be derived from Fermat's principle or from the Huygens-Fresnel principle [TMB12]. With Snell's law, a physical model of the path of light is formulated.
In the remainder of this thesis, the first indexed medium will be air and the second one will be water. Their refractive indices are $n_a = n_1 = 1.0$ and $n_w = n_2 = 1.33$, respectively. A useful property of the physical model of the path of light is that it is invertible. Hence, the designations incident ray and refracted ray, as well as angle of incidence and angle of refraction can switch positions. Therefore, a ray in water will from now on be designated by w and a ray in air by a. The point on the interface Φ where the refraction happens will be designated as refraction point R.

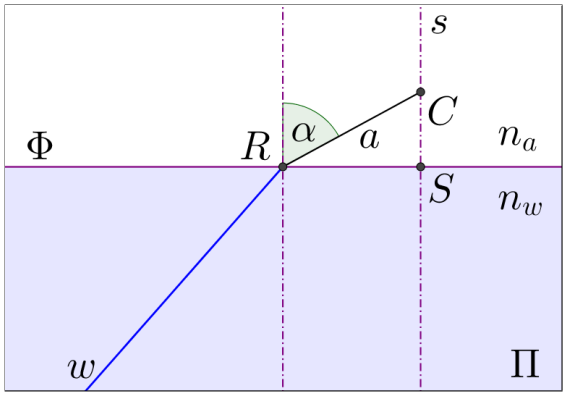

Figure 2.6: Ray tracing starting from point C.

2.3.2 Ray Tracing Formulas

Let us assume that the position vector $\boldsymbol{c}$ to the starting point C of a ray a in air and its normalized direction vector $\bar{\boldsymbol{a}}$ are known (See Figure 2.6). A common approach to trace the ray physically correct is to determine its intersection point R with the refractive interface Φ and the normalized direction vector $\bar{\boldsymbol{w}}$ of the refracted ray w in water. The refraction point R can be represented by its position vector, which can be computed according to [CY14] by:

$$\boldsymbol{r} = \boldsymbol{c} + \frac{t_a}{\cos\alpha}\bar{\boldsymbol{a}}, \tag{2.21}$$

with $t_a = d(C, S)$. According to Glassner [Gla95], the refracted ray can be computed by:

$$\bar{\boldsymbol{w}} = \frac{n_a}{n_w}\bar{\boldsymbol{a}} + \bar{\boldsymbol{s}}\left(\frac{n_a}{n_w}\cos\alpha - \sqrt{1 - \left(\frac{n_a}{n_w}\right)^2\left(1 - \cos^2\alpha\right)}\right), \tag{2.22}$$

with $\bar{\boldsymbol{s}}$ being the normalized direction vector of the interface normal s. Since the direction vectors are normalized, the following substitution can be performed:

$$\cos\alpha = \frac{\bar{\boldsymbol{a}} \cdot \bar{\boldsymbol{s}}}{\underbrace{|\bar{\boldsymbol{a}}|\,|\bar{\boldsymbol{s}}|}_{=1}} = \bar{\boldsymbol{a}} \cdot \bar{\boldsymbol{s}}. \tag{2.23}$$

If the refractive system is more complex, meaning that there is more than one refraction, the Equations 2.21 and 2.22 have to be used iteratively to determine each refraction point and each refracted ray's direction.

2.3.3 Apparent Depth

When an observer is looking at an object submerged in water, it seems to be closer to him as it actually is and hence seems to be bigger. This effect occurs due to refraction and implies, that a virtual image of the real object is seen. The question that arises is

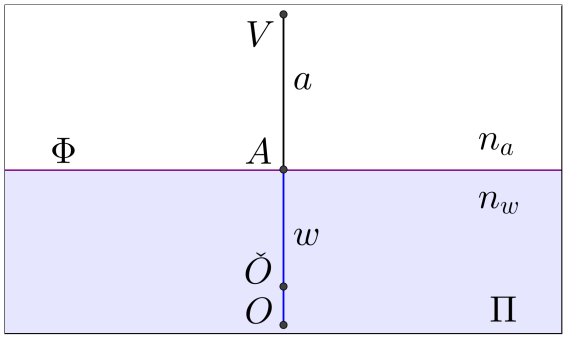

Figure 2.7: Apparent depth in the case of a perpendicular incident ray.

where this virtual image is located. This location is called apparent depth [TMB12]. The following considerations are concerning point objects, single flat refractive interfaces and monocular viewing.

Perpendicularity of Incident Rays

The case that is most commonly explained in physics textbooks is an observer at a viewpoint V with a viewing direction perpendicular to a flat refractive interface Φ between two media like air and water, as illustrated in Figure 2.7. The light ray is emitted from a point object O. The incident ray in this case is the ray w in water and the refracted ray is the ray a in air.
According to Tipler et al. [TMB12], an equation for the refraction at a single interface can be set up by utilizing Snell's law and the small-angle approximation ($\sin\alpha = \alpha$) by:

$$\frac{n_w}{x} + \frac{n_a}{\check{x}} = \frac{n_a - n_w}{r}, \tag{2.24}$$

with $x = d(A, O)$ being the object distance, $\check{x} = d(A, \check{O})$ being the image distance and r being the curvature radius of the interface. Since the interface is flat, the curvature radius amounts to $r = \infty$. As the incident ray coincides with the interface normal, the incidence angle is zero degrees and there is no refraction. After reorganization we end up with:

$$\check{x} = -\frac{n_a}{n_w} \cdot x. \tag{2.25}$$

Since $\check{x}$ is negative, the image is virtual and on the same side of the refractive interface Π as the object. This is what most physics textbooks teach for the case of an incidence angle of zero degrees. As can be seen, Equation 2.25 is independent of the distance to the viewpoint V of the observer. Hence, V could be situated anywhere on the ray a in air. In the remainder of this thesis, the point object O will be called real object point and $\check{O}$ will be called *VOP*.

Non-Perpendicularity of Incident Rays

Equation 2.25 can also be used as an approximation for small deviations of the angle of incidence from zero degrees. Larger angles of incidence are most commonly not considered explicitly in physics textbooks. Since these perpendicular incident rays do not represent the more general case that this thesis is confronted with, some further research results for non-perpendicular rays need to be considered in the following. According to Tipler et al. [TMB12], Equation 2.24 will not result in the correct location for non-perpendicular rays, since approximate perpendicularity is a prerequisite for its application. In the case of non-perpendicular rays, the apparent depth of an object is a controversial topic in physics.

Bartlett et al. [BLJ84] propose that the image of a point object, submerged in a medium with refractive index n_2 (here: water with n_w), from a viewpoint in a medium with refractive index n_1 (here: air with n_a), is in general astigmatic. This means that two virtual images can be formed, which do not coincide except for an angle of incidence of zero degrees. The schematic derivation of both locations of the virtual image can be seen in Figure 2.8. In the observations of Bartlett et al. [BLJ84], the ray w in water is the central ray of a cone representing a bundle of rays, which are emitted from a real object point O. The derivation of the locations is done by a limit value observation in two planes. According to the definition of astigmatism [TMB12], these planes are perpendicular to each other and are called the meridional plane Ψ_m and the sagittal plane Ψ_s, respectively. For the application of the limit value observation, a second ray in each of these planes (green ray in Ψ_s and red ray in Ψ_m) is chosen from the cone in Figure 2.8 $\boldsymbol{a}$ (the cone's size is exaggerated for visualization). The limit value observation is done for an infinitesimal radius of the cone, as is illustrated in Figure 2.8 $\boldsymbol{b} - \boldsymbol{d}$. The final results are the two possible locations of the virtual image at the *VOP*s $\check{O}_m$ and $\check{O}_s$ in Figure 2.8 $\boldsymbol{d}$. These *VOP*s are the points of intersection of the virtual extensions of the rays into water (dashed). Due to occlusions, the two rays in the sagittal plane can only be seen in the top views of Figure 2.8 $\boldsymbol{b} - \boldsymbol{d}$ and the two rays in the meridional plane in the side views. In the sagittal case, the planes of refraction of both ray paths obviously differ. However, both of them have to intersect in the interface normal s, which is passing through the real object point O. Hence, the virtual extensions of the rays have to intersect in the interface normal s and the virtual image is located at the sagittal *VOP* $\check{O}_s$. The limit value observation is just performed formally. The meridional plane Ψ_m and the plane of refraction Π, which belongs to the ray w, are identical. Therefore, the location of the meridional *VOP* $\check{O}_m$ has to be situated in this plane as well. It is located higher than the real object point O and closer to the observer. The viewpoint of the observer can be located arbitrarily on the ray a in air. As can be seen, both locations are situated on the backwards extended ray a.

The apparent depth of an object is still a controversial topic in physics. This is shown, for example, by the recent work of Quick et al. [QGP15]. The authors revise the findings of Bartlett et al. [BLJ84] and conclude in favor of the sagittal *VOP* $\check{O}_s$.

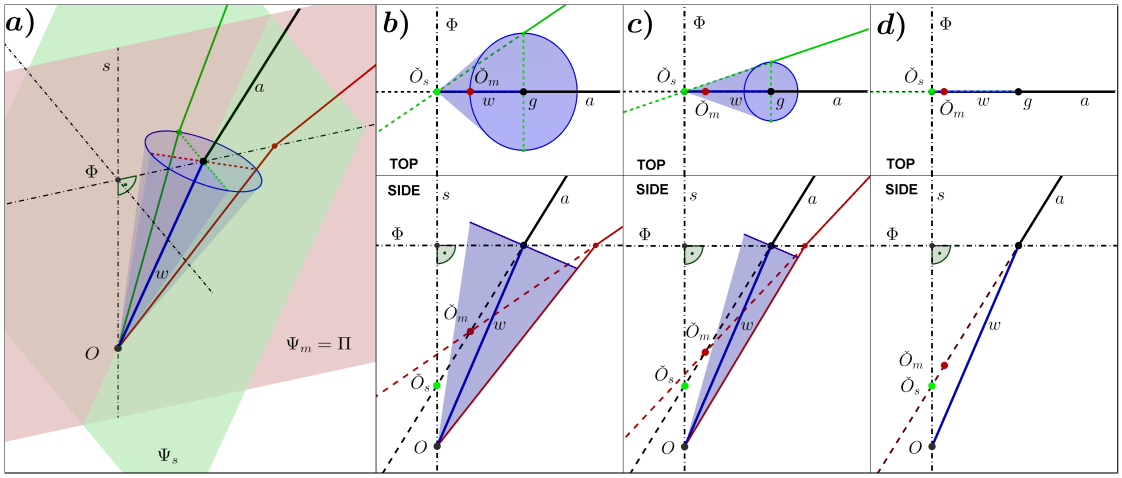

Figure 2.8: Derivation process of the *VOP* locations in the case of a non-perpendicular incident ray.

Model for a Single Flat Refractive Interface

As will be seen in Chapter 4, the physical foundation of the *VOP*s is not of primary importance for the developments in this thesis. Therefore both locations represent valid possibilities. Since the location of the *VOP* $\check{O}_m$ will not be utilized for the approaches to be developed in this thesis, as will be justified in Section 4, only the geometrical derivation from the previous section is presented. The interested reader may be referred to [BLJ84] or [QGP15] for a more detailed derivation. In contrast, the location of the *VOP* $\check{O}_s$ will be used in the following model for a single flat refractive interface.
The relation between the location of the *VOP* $\check{O} = \check{O}_s$ and the real object point O in Figure 2.9 can be derived from Snell's law and trigonometry in right-angled triangles [BLJ84]. The path of the light is inverted once again and with that, the relation can be expressed by:

$$\frac{\check{x}}{x} = \frac{\cos\alpha}{\sqrt{\left(\frac{n_w}{n_a}\right)^2 - \sin^2\alpha}}, \tag{2.26}$$

with $\check{x} = d(A, \check{O})$ being the distance to the *VOP*, $x = d(A, O)$ being the distance to the real object point and α being the angle of incidence. Note that this is also valid for media that are different from water and air. Therefore, the refractive indices n_a and n_w have to be replaced accordingly. The substitution of an angle of incidence that is equal to zero degrees into this equation and some rearranging leads to Equation 2.25, which represents the apparent depth in the case of a ray that hits the refractive interface perpendicularly. Note that the negative algebraic sign, which is used from the physics perspective to make clear that a virtual image is produced, is omitted. Equation 2.26 is independent of the distance to the viewpoint of the observer. Hence, this viewpoint can be situated anywhere on the ray a.

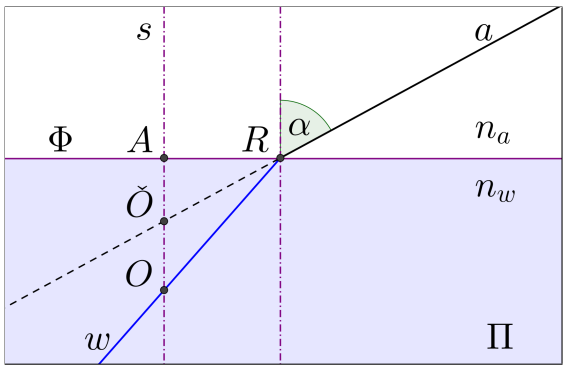

Figure 2.9: Relation between the location of the *VOP* $\check{O}$ and the real object point O in the case of a single refractive interface.

Note that some conversions of Equation 2.26 are possible. Therefore, the Pythagorean trigonometric identity [BS79]:

$$\sin^2 \alpha + \cos^2 \alpha = 1 \tag{2.27}$$

can be used. If the normalized direction vectors $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{s}}$ of the ray a in air and the interface normal s are known, the angle of incidence α can be expressed by Equation 2.23. With these substitutions, Equation 2.26 can be converted to the form:

$$\frac{\check{x}}{x} = \frac{\bar{\boldsymbol{a}} \cdot \bar{\boldsymbol{s}}}{\sqrt{\left(\frac{n_w}{n_a}\right)^2 - 1 + (\bar{\boldsymbol{a}} \cdot \bar{\boldsymbol{s}})^2}}. \tag{2.28}$$

Thereby, the need to compute trigonometric functions can be eliminated completely.

2.4 Implications of Refraction

As already indicated in the previous section, the physical phenomenon of refraction of light has some major consequences on the applicability of the conventional computer vision approaches made for aerial environments. The approaches that rely on the pinhole camera model are based on collinearity. Due to refraction, an object point in water, the center of projection of a camera and an image point are in general not collinear any more. Hence, the pinhole camera model can not be applied as it has been in aerial environments. In the following, it will be shown why the approximation of underwater image formation by the pinhole camera model is invalid and that the distortion due to refraction is distance-dependent. This distance-dependence makes clear that refractive distortions are not image space distortions, such as lens distortion. A direct consequence of the invalidity of the pinhole camera model is the invalidity of the epipolar geometry.

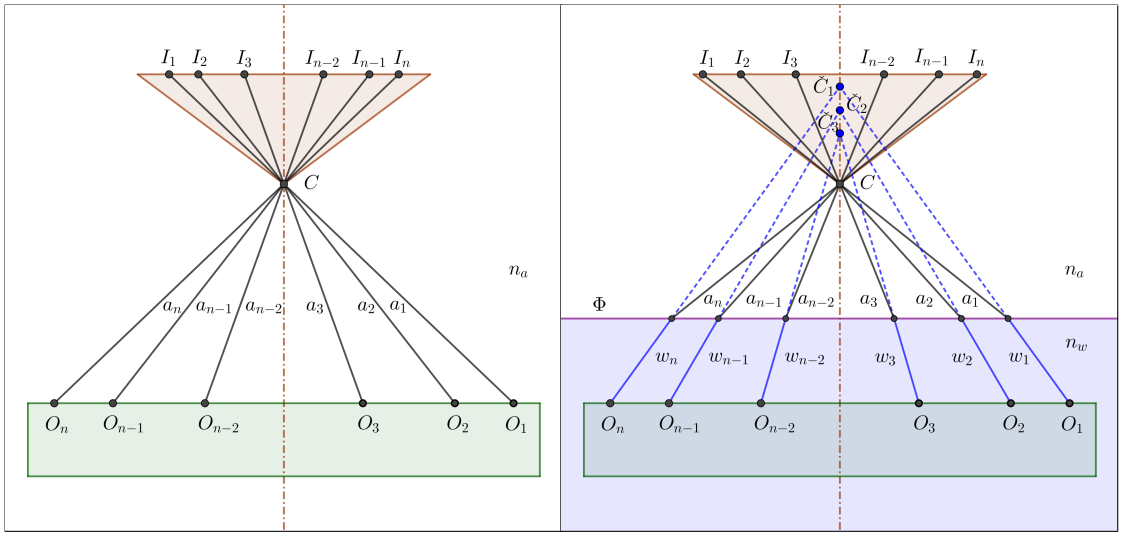

Figure 2.10: Left: Image formation based on the pinhole camera model in air. Right: Invalidity of the pinhole camera model for underwater image formation.

Invalidity of the Pinhole Camera Model. The left side of Figure 2.10 shows the image formation based on the pinhole camera model in air. Let us consider a number of real object point O_i, with $i \in \{1, .., n\}$, which are projected to image space. The imaged object points O_i on an arbitrary surface are the starting point of the rays a_i. All these rays intersect at a single point, namely the center of projection C, and end up on the image plane at the respective image points I_i. A real object point O_i, the center of projection C and an image point I_i lie on a single ray and are therefore collinear.
The right side of Figure 2.10 shows the image formation in the case of a flat refractive interface Φ, which separates air and water. This time, the object points O_i on the imaged surface are located in water. They are the starting points of the rays w_i. These rays change their direction at the interface Φ due to refraction, intersect at the center of projection C and end up on the image plane at the respective image points I_i. The difference to image formation in air is that a real object point O_i, the center of projection C and an image point I_i are not collinear any more. As can be seen the on right side of Figure 2.10, the extension of the rays w_i into air (dashed blue lines) and the pairwise intersection of rays with equal incidence angles leads to several intersection points. These represent multiple virtual centers of projection ($\check{C}_1$, $\check{C}_2$, $\check{C}_3$), which are in this case located on the principal axis p of the real camera. Hence, the pinhole camera model is not valid for the approximation of the image formation in underwater imaging setups [Tre+12] . This occurrence of multiple virtual centers of projection makes it impossible to infer a single focal length adjustment of a camera that can represent the image formation in terms of the pinhole camera model correctly.

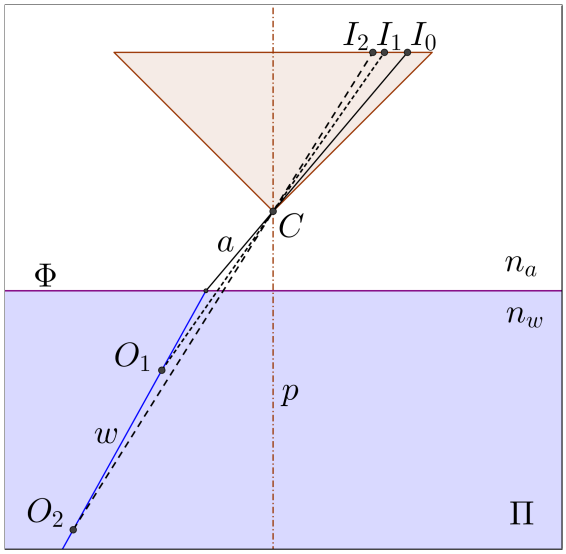

Figure 2.11: Distance-dependence of refraction.

Distance-Dependent Distortion due to Refraction. The distortions in an image that stem from refraction are dependent on the distance of the imaged object point to the refractive interface. Let us assume that refraction can be modeled as an image space distortion, such as in the case of radial and tangential distortion of a real imaging system. This would mean that the pinhole camera model with some additional non-linear terms for image distortion can be used for the approximation of the underwater image formation. Figure 2.11 shows that this is theoretically impossible. The image point I_0 is the result of the mapping of all possible object points O_i, with $i \in \{1, .., n\}$, on a ray w in water. This ray gets refracted at the interface Φ, which results in the ray a. Two of the infinitely many possible locations of the object points O_i on the ray w are illustrated by O_1 and O_2. In reality, both end up in image point I_0. As can be seen, the linear mapping of the point O_1 and O_2 by the pinhole camera model would result in the two different image points I_1 and I_2. Hence, refractive distortion is clearly dependent on the 3D coordinates of the mapped object points and thus not an image space distortion, which would be independent of the scene. Therefore, refractive distortion can theoretically not be compensated by additional non-linear terms for image distortion. This is also shown by Treibitz et al. [TSS08; Tre+12].

Invalidity of the Epipolar Geometry. The epipolar geometry makes it possible to transform both views of a stereo camera into a common plane to reduce the search space of corresponding pixels in both views. Let us assume that this transformation is unnecessary due to the already perfect orientation of the stereo camera. Figure 2.12 shows such a perfectly oriented stereo camera from three orthographic views. Both cameras are have identical intrinsic parameters and their principal axes p and p' are parallel.

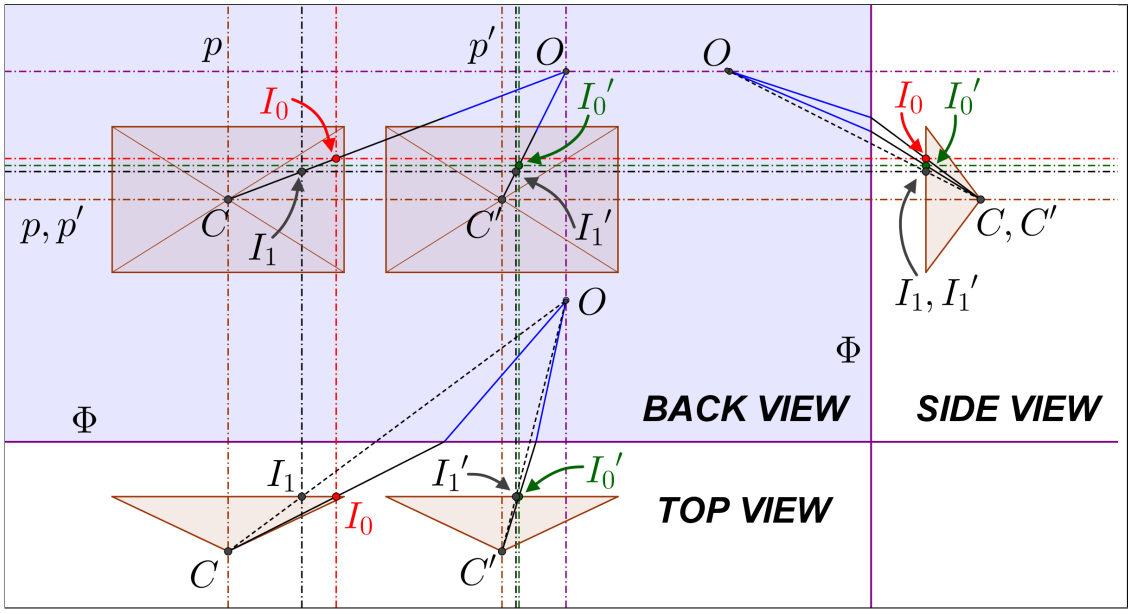

Figure 2.12: Epipolar lines an their invalidity in underwater stereo vision, illustrated by three orthographic views.

The back view, the side view and the top view of this setup can then be stitched in the presented way. The back view is essentially constructed by the pairwise intersections of parallel lines from the side and the top view. Note that the representation of the cameras with a flipped and therefore virtual image plane is necessary for this kind of visualization (See Section 2.2.1).

Let us consider the mapping of a real object point O by this stereo camera. This is illustrated once based on the pinhole camera model (dashed black) and once with explicit modeling of refraction at an interface Φ (solid blue turning solid black). The dashed black rays and the image points I_1 and I_1' represent the mapping if water would not exist. The horizontal line through the points I_1 and I_1' in the back view represents the epipolar line, which can be used as the search space for stereo matching of these two points. In an aerial environment, the assumption of epipolar geometry is that the rays connecting an object point O with the two centers of projection C and C' form a plane [HZ04]. This assumption does not hold if refraction occurs [Tre+12]. If the epipolar geometry would be valid in the case of the mapping with explicit modeling of refraction, the dashed horizontal lines in red and green would coincide. The illustration clearly shows that the two image points I_0 and I_0' do not share an epipolar line in general. The only exception are object points that are located in the plane that is formed by the interface normal and the two centers of projection C and C'. In the illustrated case, this would mean that the object point O needs to lie in the plane that is formed by the two centers of projection C and C' and the principal axes p and p' to get a valid epipolar line. However, in the general case, the epipolar geometry is invalid for underwater image formation [Tre+12].

2.5 Definitions

The previous sections have shown that the pinhole camera model is not suitable for modeling underwater image formation. In this section, a more appropriate model will be defined for the target system. It consists of a stereo camera, which is protected from water penetration by a single transparent medium, and it explicitly considers the refraction of light. Therefore, the defined model belongs to the category of refractive systems. In this section, to put it into context, a taxonomy and a parametrization will be presented for the addressed refractive systems in this thesis. It will be characterized by the introduced model for a *SFRS*, which is a specialization of the more general model for a *FRS*.

Flat Refractive System. A camera needs to be protected by a transparent medium for underwater usage. The transparent medium that will be considered in this thesis is a plane-parallel plate. A single camera in combination with such a plate is termed *FRS* [TSS08; Tre+12] and is usually characterized by a water-glass-air transition of the light rays. It can be described by a system axis, layers and interfaces [Agr+12], which will be defined in the following. These components are usually fixed to the camera. Some special cases of *FRS*s arise in the less frequent setup where a camera is moved with respect to a fixed plane-parallel plate. The left side of Figure 2.13 shows an exemplary *FRS*. The plane-parallel plate needs to be parametrized with respect to the camera, which is described by its center of projection C, its camera coordinate system (x-axis, z-axis and perpendicular y-axis, which is omitted in the illustration), its principal axis p and its principal point P. This entities of this parametrization will be summarized under the term refractive parameters, which will be defined in the following as well.

System Axis. The system axis is a special form of the surface normal of the plane-parallel plate. Therefore, it has the same direction vector as the surface normal, which up to now was labeled by s. The system axis' main characteristic is that it passes through the center of projection C of the camera. This is illustrated on the left side of Figure 2.13. In what follows, the system axis will be labeled by s_0.

Object Axis. Similarly to the system axis, the so-called object axis is a special form of the surface normal of the plane-parallel plate as well. The object axis' main characteristic is that it passes through a real object point O in water. This is illustrated on the right side of Figure 2.13. In what follows, the system axis will be labeled by s_1.

Interface. Interfaces are the planes bounding a layer. The term interface will be utilized in this thesis synonymously with the term flat refractive interface. An interface has no thickness. On the left side of Figure 2.13, Φ_0, Φ_1 and Φ_2 represent flat refractive interfaces. Note that Φ_0 is a virtual interface that passes through the center of projection C of the camera. In this thesis, virtual interfaces, such as Φ_0, will be generally omitted in the following illustrations.

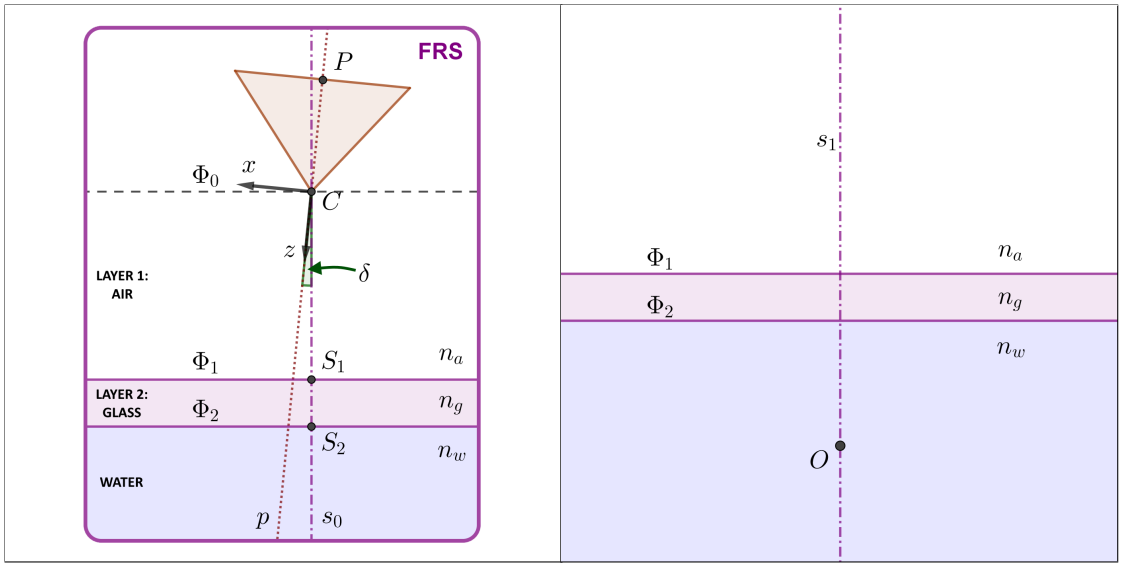

Figure 2.13: Left: Schematic illustration and parametrization of a *FRS*. Right: Object axis s_1.

Layer. Layers are fixed with respect to a camera. They are plane-parallel, have a thickness and a refractive index. Hence, the water area on the left side of Figure 2.13 is not regarded as a layer. The first layer is starting at the center of projection C of the camera and it is characterized by the refractive index n_a. The planes bounding this first layer are the virtual interface Φ_0 and the real interface Φ_1. The second layer is the plane-parallel, transparent plate, which is characterized by the refractive index n_g. The planes bounding the second layer are the two interfaces Φ_1 and Φ_2. Note that the model of layers and interfaces can be easily extended if additional plane-parallel media would exist. In this thesis, there will be two layers of interest, namely the air layer and the glass layer. A layer will be characterized by its thickness, which can be described by distances along the system axis s_0, and by its refractive index.

Refractive Parameters. The refractive parameters are basically adopted from the parametrization of Agrawal et al. [Agr+12]. First of all, the orientation of the system axis s_0 has to be determined with respect to the camera. In the 2D illustration on the left side of Figure 2.13, this is described by the angle δ. In 3D, the parametrization has to be realized by a 3D vector or an equivalent representation, such as spherical coordinates (See Figure 2.2). Besides the determination of the orientation of the system axis s_0, the layer thicknesses need to be determined. In this thesis, these will be the air layer thickness t_a and the glass layer thickness t_g. As can be seen on the left side of Figure 2.13, the air layer thickness t_a amounts to the Euclidean distance between the center of projection C of the camera and point S_1 on interface Φ_1 and the glass layer thickness amounts to the Euclidean distance between point S_1 and point S_2 on interface

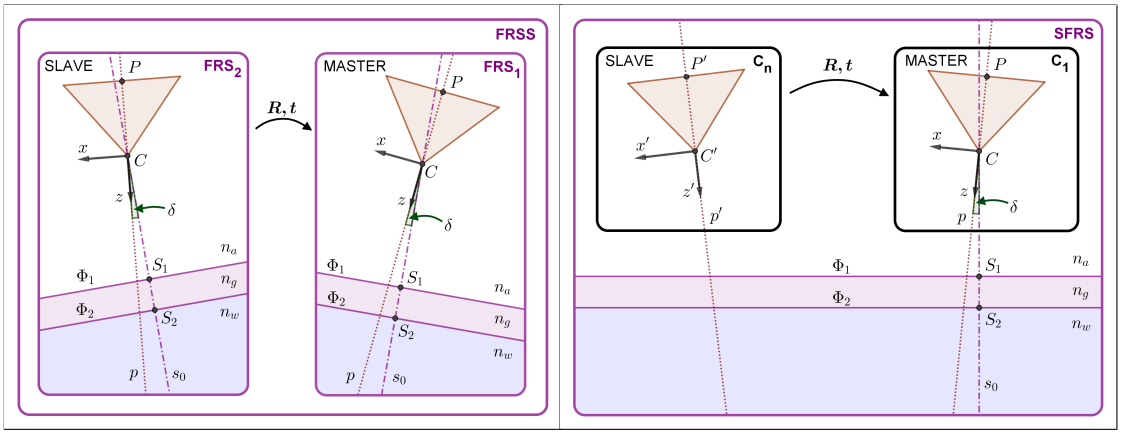

Figure 2.14: Left: Schematic illustration and parametrization of a *FRSS*. Right: Schematic illustration and parametrization of a *SFRS*.

Φ_s. Therefore, $t_a = d(C, S_1)$ and $t_g = d(S_1, S_2)$. Further refractive parameters are the refractive indices of all participating media (n_a, n_g, n_w).

Flat Refractive Stereo System. *FRSS* is the proposed term for the concatenation of two individual *FRS*s, as can be seen on the left side of Figure 2.14. In this case, a stereo camera forms the basis. As will be described in Chapter 3, this case is distinguished from further concatenations of several individual *FRS*s. The main characteristic of the model for a *FRSS* is that two cameras are mandatory for the specially developed calibration approaches from the literature. The layers and interfaces are fixed with respect to the single cameras and may differ. Therefore, the refractive parameters have to be determined for each individual *FRS*. Additionally, the pose (rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$) of the slave camera needs to be determined with respect to the master camera.

Shared Flat Refractive System. *SFRS* is the proposed term for another specialization of the previously defined *FRS*. The schematic illustration on the right side of Figure 2.14 shows a representative of this category. It contains a master camera and one or multiple slave cameras. A slave camera can be described by its center of projection C', its camera coordinate system (x'-axis, z'-axis and perpendicular y'-axis, which is omitted in the illustration), its principal axis p' and its principal point P'. The main characteristic of the model for a *SFRS* is that the layers and interfaces are shared by the cameras. Therefore, the refractive parameters need to be determined only with respect to the master camera. The refractive parameters with respect to the slave camera, or cameras, can be implicitly established by their relative poses (rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$) with respect to the master camera. Hence, the amount of parameters gets reduced significantly. The target of this thesis is a *SFRS* with exactly two cameras.

2.6 Summary

This chapter has shown that the physical phenomenon known as refraction of light implicates some major consequences on the applicability of conventional computer vision approaches that are made for aerial usage in underwater environments. The brief descriptions of the basic concepts of stereo 3D reconstruction and the basics on refraction have shown that an extension for the recovery of 3D structure from underwater images is challenging and not straightforward. Implications, such as that refractive distortion is due to its dependence on the imaged scene not an image space distortion, lead to the fact that an approximation of underwater image formation by the pinhole camera model with non-linear distortion terms is theoretically not feasible. Nevertheless, researchers actively apply such approximation strategies, as will be seen in the following chapter. The pinhole camera model is the very foundation of many conventional computer vision approaches. Its invalidity under water has the direct consequence that, for example, the naive combination of underwater imaging systems and stereo 3D reconstruction made for aerial applications has to fail. Another important consequence is that epipolar geometry is invalid as well. This makes a very efficient tool for increasing the performance of stereo matching unavailable.

Consequently, starting with the calibration of an underwater imaging system, the refraction of light needs to be considered explicitly. The law of refraction, its application in the equations for ray tracing and a model for the apparent depth form the foundation of the developments in this thesis. The proposed image formation model for this thesis is the *SFRS*, with a water-glass-air transition of the light rays. It is characterized by distinct layers and its parametrization contains an air layer thickness, a glass layer thickness, an orientation of the system axis and the refractive indices of the participating media. These parameters will be called refractive parameters in the remainder of this thesis. Within this image formation model, refraction is considered explicitly. It has been defined within the proposed taxonomy, which will be utilized in the following analysis of related works. In the following chapter it will be shown that such an explicit modeling of refractive effects is indispensable in general.

3 Related Works

Several methods for 3D reconstruction have been adapted to underwater environments in recent years. A detailed survey can be found in the work of Massot-Campos and Oliver-Codina [MO15]. An important first step is the calibration of the imaging system, without which it would not be possible to regain metric information from image data. In this chapter, the focus will be on calibration approaches from the related literature, which are used to estimate the parameters of recording systems that are exposed to refractive effects. It will become apparent that the main calibration strategies are to either neglect these effects, to absorb them or to model them explicitly by additional parameters. In the following, these three strategies will be analyzed with regard to their major shortcomings. The amount of the visible refractive effects in an image is, amongst other things, dependent on the shape of the refractive interface. In this thesis, the shape of interest is a flat refractive interface and the recording system is a stereo camera. Therefore, this chapter will primarily comprise publications that can be related to this target system.

Calibration approaches made for aerial environments are usually based on image formation following the pinhole camera model. An overview of image formation models used in underwater imaging can be found in the work of Sedlazeck and Koch [SK12]. The major part of this chapter will deal with the explicit modeling of refractive effects. As will be seen in the following, this is indispensable if refraction is supposed to be handled properly. Ultimately, the goal of this thesis is the estimation of the refractive parameters originating from the model for a *SFRS*. At the end of this chapter, the approaches to be developed in this thesis will be located in a categorization scheme together with relevant related approaches.

3.1 Disregard of Refractive Effects

The first strategy to handle refractive effects is to simply ignore them. This can be realized in two ways, which are both founded on the pinhole camera model. The first is to calibrate the recording system with a conventional approach in air before underwater usage. The second is the application of a conventional in air calibration approach on underwater images. Admittedly, the second way and the calibration strategy with an absorption of refractive effects, which will be described in Section 3.2, are most of the times difficult to distinguish. The weak criteria that are used for the following categorization are that it is either not clear if the authors intentionally try to solve the problems originating from refractive effects or that the authors do not mention refraction at all.

Single Camera

Throughout their works on large area 3D reconstruction [PES04; PES09] and relative pose estimation [PES03], the only information provided by Pizarro et al. concerning camera calibration is that high quality calibration is assumed. Carreras et al. [Car+03] present an approach for vision-based localization of an underwater robot in a structured environment. The calibration of the intrinsic parameters of the camera is said to be done off-line. The authors do not give any details. Rzhanov et al. [RHB14] propose a method for close-range Euclidean 3D reconstruction of underwater scenes with a single free moving camera. The camera is calibrated with the aid of the *Camera Calibration Toolbox for Matlab* [Bou08], while refractive effects are apparently ignored. The reconstructed results are not evaluated. Pessel et al. [POA03a; POA03b] describe a method to determine the intrinsic parameters of a single camera by self-calibration in an underwater environment. Apparently, the authors ignore the refraction of light. Although they evaluate their results with simulated data, it has to be assumed that refraction has not been considered during simulation as well, since there is no detailed explanation given. This would result in a systematical error.

Stereo Camera

Addressing Calibration Indirectly. There is a number of works that deal with stereo camera calibration indirectly, since it obviously is integrated into their workflows, but apparently it is not the primary concern of the author's developments. In [Joh+10] the authors deal with large-scale 3D reconstruction from underwater robotic surveys. Stereo calibration is done in a pool prior to the deployment. However, the authors do not provide their intentions on why they are doing this. The results are said to be of reasonable quality, although the resulting errors are in centimeter range. Beall et al. [Bea+11] deal with the same application. The only provided information concerning stereo camera calibration is that high quality calibration is assumed. In [Jas+08] the authors deal with underwater 3D mapping and pose estimation for ROV Operations. They just mention the use of a calibrated stereo camera and make the statement that the

resulting 3D model shows a slight bend. Brandou et al. [Bra+07] do 3D reconstruction of small-scale scenes with a stereo camera. The camera parameters are calculated off-line from underwater images using the *Camera Calibration Toolbox for Matlab* [Bou08]. Since there is no evaluation of the results, it can not be analyzed if there are any discrepancies due to ignoring refractive effects.

Addressing Calibration Directly. The following works deal with camera calibration more or less directly. Kunz & Singh [KS10] aim at self-calibration of the extrinsic parameters of an underwater stereo camera and a scale in the context of producing metric maps of the seafloor. Visual information and navigation data acquired by an AUV are used for calibration. The question whether or not it is possible to do a full self-calibration (intrinsic and extrinsic parameters) the authors have not yet answered. The authors remark that even with high quality lenses, distortion is an issue underwater, since refraction renders the pinhole camera model incorrect (Referencing [TSS08; Tre+12]) and that it probably should not be ignored. In [SR12] the authors present preliminary results of the measurement of the 3D topography of the seafloor with a consumer stereo camera. Its calibration is done with the *Camera Calibration Toolbox for Matlab* [Bou08]. The authors claim that calibration is performed under water for compensation of lens distortion. The refractive distortion is not mentioned. After calibration, a surface is generated by stereo 3D reconstruction. However, only a qualitative analysis of the results is performed by the authors. Wehkamp & Fischer [WF14] present a workflow for stereoscopic measurements for marine biologists. The work is a step-by-step guide for non-experts. The described procedure is said to be sufficiently precise to meet the authors goals, which are that it should be practical. As in the previous approach, the calibration is done with the *Camera Calibration Toolbox for Matlab* [Bou08]. Refractive effects are ignored throughout the paper. Silvatti et al. [Sil+12; Sil+13] compare three different underwater camera calibration approaches and analyze their accuracy in the context of swimming kinematics. The authors experiment with static control points and nonlinear DLT ([Hat88]), with a moving wand, a nonlinear camera model and bundle adjustment ([Tri+00]) and with a moving 2D calibration object and Zhang's method ([Zha00]). The refraction of light is not mentioned in the paper. Hence, it can not be determined if the authors intentionally aim at absorption of the refractive effects by their approach. Only an acquisition volume of a fixed size is calibrated and evaluated. A common problem with these approaches is the increase of errors outside of this volume.

Findings

The various underwater applications that are relying on camera calibration demand different requirements in terms of accuracy. Sometimes the authors prioritize practicality of the approaches over accuracy. On large-scale 3D reconstruction, the common goal is to generate a visualization of large structures. The focus does not lie on details and hence, accuracy is not the priority. In these cases, refractive effects are mostly ignored. Since many approaches are often tested immediately in the field, there is no ground truth data to evaluate the results. A more careful strategy would be a previous evaluation in a

controlled environment to identify sources and amounts of error. Besides often missing evaluations, the approaches mentioned above contain only qualitative evaluations or evaluations that might be corrupted and the details on camera calibration are mostly very unspecific. Sometimes it is not even clear, if the cameras are calibrated in air or in water. As demonstrated in the previous chapter (Section 2.4), refractive effects can be severe and should not be ignored, which has been indicated in this section by Kunz & Singh [KS10] as well.

3.2 Absorption of Refractive Effects

The second strategy to handle refractive effects is to let the conventional intrinsic camera parameters absorb them. Hence, it is also based on the pinhole camera model. The absorption is realized in form of an application of a conventional calibration approach for in air usage on underwater images. The refractive effects are expected to be absorbed by the focal length and radial distortion terms. In contrast to the strategy before, the authors perform this absorption strategy intentionally and analyze its capabilities to compensate the refractive effects.

Relation between the Focal Length in Air and Water

In an early analysis of the calibration of underwater cameras, Fryer and Fraser [FF86] regard refractive effects in *FRS*s as radial distortion. They claim that the total effect of radial distortion due to refractive effects can be obtained by subtracting the calibrated radial distortion in water from those in air. Additionally, they claim that refractive effects cause the focal length in water to increase by a factor equal to the common value of the refractive index of water. Lavest et al. [LRL00; LRL03] propose that the intrinsic parameters of a submerged camera can be computed from a calibration in air. Therefore, the refractive surface between water and air has to be flat and the refractive index of both media has to be known. Like Fryer and Fraser [FF86], they propose that the focal length of the camera is equivalent to the value measured in air multiplied with the refractive index of water. Concerning distortion, they have not been able to determine the relationship between air and water and claim that it is doubtful if such a relationship exists. According to their results, the image principal point remains unchanged. The authors conclude that the fit between their theory and measured data is almost completely fulfilled. In a recent publication, Agrafiotis & Georgopoulos [AG15] disagree with the proposal of Lavest et al. [LRL00; LRL03]. Their experiments show that the effective ratio between the focal length in air and water is proportional to the percentages of air and water in the total camera-to-object distance. Rahman et al. [Rah+13] also report that the factor between the focal length of a camera calibrated in air and in water, respectively, is not exactly the refractive index of water and that the image principal point is significantly different, which is in contradiction with [LRL00; LRL03].

Indication for Sufficient Absorption Capabilities

Kang et al. [KWY12a] argue that the pinhole camera model with radial distortion correction and a focal length adjustment can compensate for refraction in underwater 3D reconstruction. Despite knowing that the conventional pinhole camera model is invalid and that refractive distortion cannot be modeled by an image space distortion model in theory, the authors claim that refractive distortion can be well approximated if a proper similarity transformation is applied to the scene structure. The authors perform experiments with simulated and real data to quantitatively evaluate the influence of refraction on the accuracy of their proposed model in underwater 3D reconstruction. In later publications [KWY12b; Kan+15] (See Sections 3.3.5 and 3.3.6) the authors depart from this strategy and tend to model refraction explicitly.
Sedlazeck et al. [SKK09] present a structure from motion system for 3D reconstruction from underwater images and videos. The authors are aware of the refractive effects and conclude that the camera needs to be calibrated in-situ or that the camera calibration above water needs to be converted with an approach like the one of Lavest et al. [LRL00]. The pinhole camera model with radial distortion terms is used and a calibration in a water tank is said to be sufficiently accurate for the authors purposes. Since there is no ground truth data available, the 3D reconstructed results are not evaluated for accuracy. In a later publication [JK12] (See Section 3.3.5) the authors rely on a strategy that models refraction explicitly.

Indication of the Incompleteness of Absorption Capabilities

Harvey and Shortis [HS95] deal with the calibration of an underwater stereo camera in the context of length measurement of marine biological specimens. In later publications of the same authors, they provide insight into their concepts [SH98] and into stability issues of the calibration approach [HS98]. They explicitly mention the complications of underwater length estimation due to refractive effects and that calibration must incorporate these effects at the various air-glass and glass-water interfaces. Their underlying model does not contain explicit terms for refractive effects, since the authors conclude from related works [LTZ96; FF86] that the principal component of the refractive effect is radial. The assumption is that refractive surfaces are in general perpendicular to the image plane. As a consequence, their approach is to allow the refractive effects to be absorbed by the conventional camera parameters. Their reasoning is that the calibration of the complete system under prevailing conditions is likely to be more accurate and reliable. Therefore, they use a purpose built 3D control frame under water for a so-called self-calibrating multi-station bundle solution [Gra80]. Their results imply a consistent under-estimation of length measurements [SH98]. The authors conclude in [HS98] that the disadvantage of the algorithm is the incompleteness of the model concerning refractive effects, since it is unlikely that these can be completely modeled by the conventional camera parameters. In a later publication of the same authors [Sho+00], they conclude that a significant scale bias remains and a potential source amongst others is the incomplete refraction model. In a recent publication based on the same concepts [SHA09], the authors conclude that

there is the need for an additional modeling of light paths and that any imaging system in water is unlikely to match the precision and accuracy of the equivalent configuration in air. In a more recent publication, Shortis [Sho15] mentions systematic errors in the absorption approach. Nevertheless, these errors are said to be generally less than the precision that can be achieved by the measurements when the ranges for calibration and measurement are commensurate.
The authors of [FM11] present results of a study on camera calibration for swimming kinematics. They make use of the *DLT* method [Abd71]. Target for the calibration are large calibration volumes made from aluminum, which were filmed by cameras under and above water. The authors report higher errors during underwater analysis and speculate that this is possibly due to refraction.
Kwon et al. [KC06] pursue various strategies to minimize errors of camera calibration approaches due to refractive effects. Their field of research is underwater motion analysis. They argue that it is impossible to remove/absorb the refraction error completely with conventional camera calibration methods [KC06]. They review the shortcomings of the *DLT* method [Abd71] for camera calibration. An example is that it can not correct nonlinear errors caused by refraction because it simply provides the best fit for a given experimental setup. They claim that although image deformation caused by refraction appears to be similar to pin-cushion distortion, there is a fundamental difference [KL00]. Therefore, the extended *DLT* method with additional DLT parameters to account for lens distortion [MK75] is said to be also not able to correct the error due to refraction. The authors conclude that a decreased overall reconstruction accuracy in the control space occupied by the calibration object, a large error at the boundary of the control space and an even larger error outside the control space is to be expected for *DLT* methods [KL00]. The need for a large control frame is said to be a disadvantage. For error reduction, the authors propose the localized *DLT* method [KL00] and the double plane strategy [KAP02], which both lead to an improvement, but can not correct the errors completely. In their most recent work [KC06], the authors conclude that to correct image deformation caused by refraction, calibration must include a refraction correction model like the one in their theoretical work [Kwo99], which has been tested only by simulation.

Assessment of Absorption Capabilities

The required setup for the absorption strategy to work is a flat refractive interface with its normal being perpendicular to the image plane. In such a setup, the refractive effects are said to behave like a radial distortion. Concerning this, Treibitz et al. [Tre+12] show that the refractive distortion is not merely radial and that it depends on the object distance in addition to the radial coordinate. Telem and Filin [TF10] also propose that the assumption to have aerial-like radial properties is inaccurate. It is confirmed by the results from synthetically rendered images of Kang et al. [KWY12b] that an image-space distortion model for radial distortion is incapable of compensating for the refractive effects. Jordt-Sedlazeck and Koch [JK12] tested the absorption strategy by calibration of a stereo camera. They found in their experiments that part of the refractive effects

is absorbed by erroneous camera poses, especially by translation. Their results of a computed 3D point cloud differ strongly from the ground truth data. The authors propose that the resulting errors grow with increasing interface distance and with stronger tilt of the interface normal with respect to the image plane. Similarly, a remark of the analyzes of Kang et al. [KWY12a] reads: "Under general configuration, the refractive distortion in an image is neither symmetric with respect to the image center nor monotonic along the radial direction". Hence, in a more general setup with non-perpendicularity between the interface normal and the image plane, the absorption strategy becomes even more unreliable.
Nevertheless, some researchers propose requirements where absorption leads to an acceptable approximation. By referring to [SGN03], Kang et al. [KWY12a] propose that if the locus of the caustic surface (See Section 3.3.1) is small, it can be approximated by the pinhole camera model. Similarly, Treibitz et al. [Tre+12] propose that one should place the camera such that the distance to the refractive interface is as small as possible in order to make the pinhole camera model more valid. They further propose, that the pinhole camera model can be used for approximation when working at a constant range, which should be the range in which the calibration was performed. Concerning the field of view, they say that the approximation should probably not be used if it is wide. Similarly, Kang et al. [KWY12a] propose that if the field of view is small, it can be approximated using the pinhole camera model.
The absorption strategy is also dependent on the used calibration object. Boutros et al. [BSH15] compare approaches utilizing 3D and 2D calibration objects with the outcome that calibration with a 3D object improves the measurement accuracy. Kunz and Singh [KS08] comment on the attempt to build a large 3D model from several overlapping sets of images, without modeling refractive effects, that the accumulated error would quickly lead to inconsistencies in the reconstruction and conclude: "For a typical setup, 2D distortion can be largely compensated for with nonlinear lens distortion terms in the perspective projection model, and 3D error is small compared to that induced by errors in camera motion estimation and pixel correspondence estimation. But these attempts to minimize refraction-induced distortion will ultimately fail when several images are used together for large-scale photomosaicking and structure from motion modeling.". Maas concludes in his work [Maa15b]: "neglecting the effects caused by refraction in the imaging process or trying to absorb it by standard lens distortion compensation parameters contributes to a degradation of the quality of results".

Findings

The absorption of the refractive effects by the parameters of the pinhole camera model with radial distortion is a quite common strategy, since it involves the simple application of algorithms made for usage in air and no adjustments need to be performed for the underwater environment. However, this strategy shows an incomplete elimination of refractive effects. In general setups with non-perpendicularity between interface normal and image plane the resulting errors get severe. This shows that refractive effects need to be modeled explicitly.

3.3 Explicit Modeling of Refractive Effects

The third and most important strategy to handle refractive effects is to model them explicitly. Huang et al. [HL14] propose and evaluate that even the presence of a flat refractive interface in front of a camera, without being submerged in water, induces distortions that can not be modeled by radial distortion. Together with the findings from the previous section, this shows that modeling of refractive effects is indispensable. As described in Section 2.5, this introduces additional parameters to the image formation model, which need to be incorporated into the calibration procedure. After calibration of the refractive system, it is possible to trace rays from the image sensor physically correct back into the scene. Besides 3D reconstruction with the aid of ray tracing, it is also possible, for example, to do a distance dependent correction for refractive effects in the image space when the system is calibrated like in [KFO03]. Somewhere in between the strategy to model refractive effects explicitly and the absorption strategy, the work of Ferreira at al. [FCS05] can be located. They propose an approximation approach to handle the refractive effects, which is restricted to small incidence angles and hence is not universally valid.

The first part of this section will be a review of the basics that are needed for the calibration, such as image formation models and their corresponding projection properties with regard to refraction. This will be followed by an analysis of the additional refractive properties of the whole system. These are the refractive indices of the participating media and the layer thicknesses as well as the approaches for their determination. The major part of this section will comprise the actual calibration of *FRS*s and their specializations. Another aspect will be the review of necessary calibration objects, since it is a goal of this thesis to reduce or even eliminate dependencies on calibration objects. Finally, the relevant reviewed approaches will be categorized according to their restrictions to generate a meaningful tabular representation.

In this section, the focus will be on *FRS*s. Besides the approaches that will be analyzed in the following, there are approaches that are restricted to a different kind of interface or that are able to handle further interfaces besides flat refractive ones. Therefore, some interesting representatives of these approaches will be mentioned here in passing for the interested reader. In [Yam+03] the authors propose an approach for calibration of a system for 3D measurement, composed of a cylindrical interface placed on a turntable, a camera and a laser range finder. Another system developed by the same authors is presented in [Yam+11]. Two omni-directional cameras placed in an acrylic waterproof tube are calibrated. In [YSK10] the authors calibrate a monocular stereo vision system. The system is composed of a single camera and a tank that essentially works as a beam splitter. Yano et al. [YNM13] use a double calibration target with two separated checker patterns. Only one of those patterns is submerged. This setups is used for the calibration of cameras that are imaging into a fish bowl. In [Bos+15a; Bos+15b] the authors calibrate a system composed of six cameras in an underwater housing made out of a cylinder and a hemisphere on top of it. Another scenario is the reconstruction of the 3D position and the surface normal of arbitrarily shaped refractive interfaces like the wavy surface of a liquid [MK11].

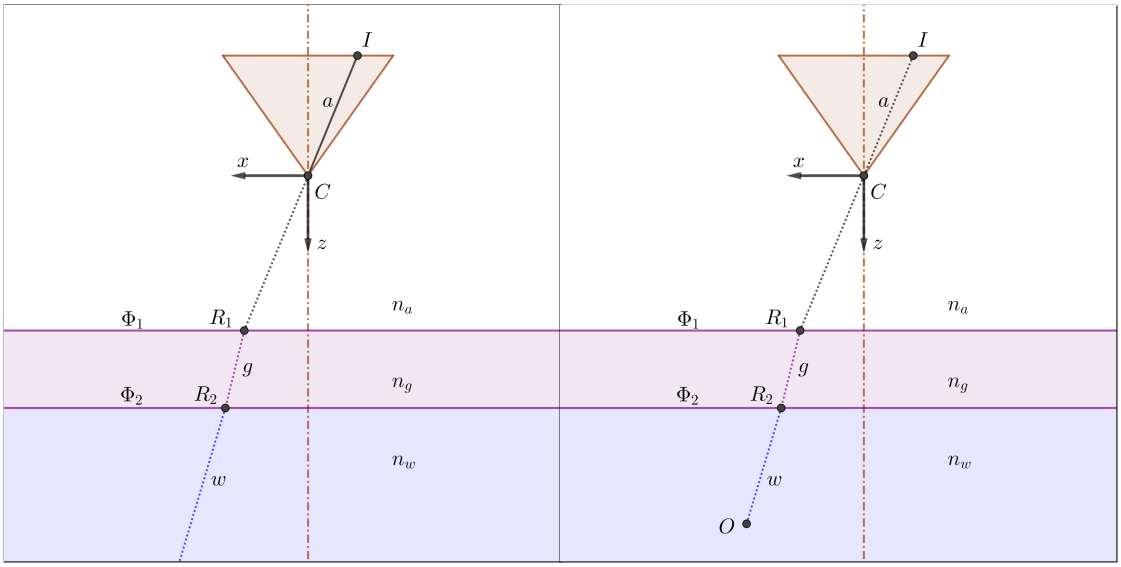

Figure 3.1: Left: Refractive back-projection of image point I to ray w in water. Right: Refractive forward projection of object point O to image point I.

3.3.1 Refractive Image Formation

The calibration of a refractive system is supposed to determine the parameters that are relevant to the process of the image formation. Its foundation is a camera model with a corresponding projection scheme that is suitable with regard to refraction. Closely associated with this is the work of Chari and Strum [CS09]. The authors present multi-view relationships between cameras in the case of a single flat refractive interface. In their theoretical work, they show the existence of structures like the fundamental matrix and the homography.

Refractive Back-Projection

If refraction occurs, the inverse process to image formation is refractive back-projection. The purpose of refractive back-projection is a transformation from a 2D image point I to its corresponding 3D ray w in the scene (See left side of Figure 3.1). This can be partly realized by the principle for back-projection described in Section 2.2 to calculate the starting direction of ray a in the camera coordinate system. If a refractive interfaces exists, the difference is that each ray has to be refracted at each interface it has to pass (Φ_1 and Φ_2). Hence, the intersection points (R_1 and R_2) have to be computed and the ray's direction has to be changed at these points according to Snell's law (See Section 2.3.2). The whole process is essentially a ray tracing approach and is straightforward in a calibrated system. However, its computational expense is severely higher than the one of back-projection based on the pinhole camera model.

Refractive Forward Projection

In this case, a 3D point O in the scene has to be transformed to an image point I on the image plane (See right side of Figure 3.1). Because of the refractive interfaces, this amounts to a ray tracing approach as well and hence, the principles from Section 2.2 can not be applied. In contrast to the refractive back-projection, no starting direction of the ray w in water is given and just the 3D coordinate of O is known. Hence, the direct computation of image point I by using Snell's law is not possible.
In [GS00] the authors define an equation for the forward projection in the case of a single flat refractive interface. To solve the equation, the roots of a polynomial of degree four need to be determined. Chang and Chen [CC11] build upon the so-called refrax [GS00], which is the point on the refractive interface where the refraction happens (e.g. R_1 and R_2 in Figure 3.1). They define the so-called refrax ratio and incorporate it into a forward projection equation. A simplification is expected by separating the process in a refractive projection to get the refrax, followed by a conventional perspective projection. For computation of the refrax, solving an equation of degree four is necessary. Agrawal et al. [Agr+12] derive analytical forward projection equations for refractions at flat interfaces. They derive an equation of degree four for the case of an air-water and an air-glass-air transition and an equation of degree twelve for the case of an air-glass-water transition.
Alternatives to solving equations of degree four or twelve are approaches that prefer approximations or iterative solutions. Duda and Gaudig [DG16] propose and approximation of the refractive forward projection by first-order Taylor expansion. Yau et al. [YGY13] perform a 1D bisection search for the angle of a camera ray in the plane of refraction for a given 3D point. This approach is said to be accurate and reasonably fast. While the authors of [Han+15] propose a forward projection similar to [YGY13], Mulsow [Mul10] proposes the so-called alternating forward ray tracing (noteworthy: forward means in this case back-projection, since the definitions between computer vision and photogrammetry are contrary). Another iterative solution is developed by Maas [Maa92; Maa95; Maa15a; Maa15b]. It is suited for *FRS*s and is said to be easily extendable to an arbitrary number of interfaces [Maa15b]. Instead of real object points, *VOP*s are used. This approach will be briefly reviewed in Section 4.2.

Refractive Forward Projection by Virtual Cameras

As has been already shown in Section 2.4, the pinhole camera model is invalid if a flat refractive interface between media with different refractive indices exists. In contrast to the conventional perspective projection in air, there is no single center of projection, which would allow an approximation of the image formation by the pinhole camera model. Nevertheless, the pinhole camera model can be useful to derive virtual camera models that represent an alternative to refractive forward projection. These virtual camera models usually have in common that the locus of the virtual centers of projection of a *FRS* has a 3D extent.

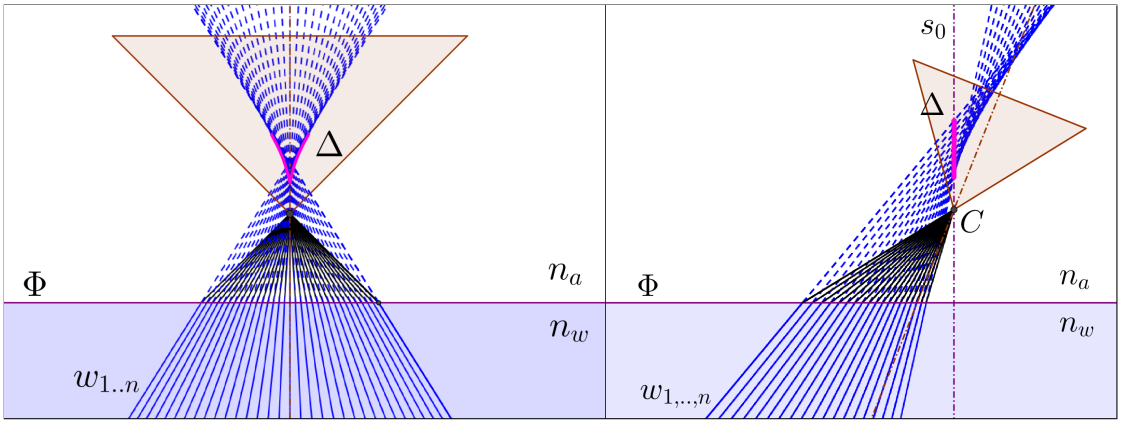

Figure 3.2: Left: Caustic surface. Right: Axial camera.

Caustic Surface. A way to characterize a *FRS* is by the term non-single viewpoint model. Treibitz et al. [TSS08; Tre+12] show that the system has indeed no single center of projection and instead propose that the locus of viewpoints is a caustic surface. A 2D visualization of such a 3D caustic surface, to which all rays $w_{i,..,n}$ are tangent, is illustrated by Δ on the left side of Figure 3.2. Similarly, [SK11] propose that all the rays meet in a singularity of the bundle of rays $w_{i,..,n}$ with its locus Δ being a caustic surface. Such a caustic surface describes a number of virtual centers of projection.

Axial Camera. A recent and important finding regarding the center of projection in *FRS*s is proposed in the work of Agrawal et al. [Agr+12]. The *FRS* is proposed to be an axial camera. A visualization of such an axial camera is illustrated by Δ on the right side of Figure 3.2. The axial camera is characterized by a system axis s_0, which is the interface normal that passes through the center of projection C of the camera. All planes of refraction and hence all refracted (solid black) and non-refracted (solid blue turning dashed blue) rays intersect in this system axis s_0, which therefore contains the locus Δ of the virtual centers of projection. In terms of computational complexity, the determination of this axis is considerably more efficient than the determination of the caustic surface.

Virtual Camera Models. Telem and Filin [TF10] propose a virtual camera model. It is derived for a *FRS* with perpendicularity between image sensor and interface normal and gets extended for the case of non-perpendicularity later on. The thickness of the interface is neglected. Their concept is based on the assumption that the principal axis of the camera, an incident ray and the related refracted ray lie in the same plane. In [SK11; SK12] the authors comment on this publication that they found that the ray coming from water not necessarily intersects the principal axis if the interface is not parallel to the image sensor. Hence, Sedlazeck and Koch [SK11] propose their own virtual camera model based on the assumption that the locus of virtual centers of projection in a *FRS*

is a caustic surface. For every 3D point that needs to be projected, a virtual camera with the caustic point as the center of projection needs to be computed. The caustic points are computed numerically. Jordt-Sedlazeck and Koch [JK13] build upon their ideas in [SK11], but do not rely on the caustic surface any more. Following [Agr+12], the virtual centers of projection are defined with the aid of an axial camera. With the proposed virtual camera models in [SK11; JK13] the refractive forward projection can be replaced by perspective projection of a 3D point onto the image plane of its virtual camera. However, the necessary computation of a virtual camera per 2D image point results in a computational overhead. Furthermore, the known 2D point on the real image needs to be transformed onto the virtual image in an additional processing step. The coordinates of both 2D points can then be used in the context of calibration by minimization of the reprojection error.

Cost Functions

A vital step for calibration approaches is the definition of cost functions that can be minimized during an optimization problem to determine the necessary parameters. The cost functions that can be found in the literature are mainly 2D image errors or 3D scene errors. The 2D cost functions that is most commonly used is the reprojection error (See Section 2.2.2). As will be seen later on, this is also the case for underwater calibration approaches. If refraction occurs, the necessary refractive forward projection of 3D points is realized as described in previous sections.
The ability of stereo cameras to compute 3D scene geometry from both images by refractive back-projection and subsequent intersection of the resulting rays makes it possible to utilize 3D scene cost function. An example is the measurement of known 3D objects, as proposed in [Brä+15a; BKN15; Brä+15b] or [TSS08; Tre+12]. Another example is proposed by Gedge et al. [GGY11]. There, the distance between the refracted rays belonging to a pair of corresponding pixels is used. Similarly, Li et al. [Li+97] and Servos et al. [SSW13] use the triangulation error of corresponding refracted rays by determination of their shortest distance. Henrion et al. [Hen+15] use the so-called residual, which is defined to be the shortest distance between a known 3D calibration point and its corresponding ray.

Correspondence Curves

Since the pinhole camera model is invalid in a *FRS*, the conventional epipolar geometry can not be used to determine epipolar lines (See Section 2.4). Nevertheless, Chari and Sturm [CS09] show the theoretical existence of structures like the fundamental matrix. They also show that a 3D line is mapped to a quartic curve in a *FRS* with a single interface. The computation of these curves is closely connected to refractive forward projection. These curves allow a reduction of the search space for corresponding pixels in the two views of a stereo camera from the whole image to a single curve, the so-called correspondence curve. Gedge et al. [GGY11] propose a sampling strategy for the computation of correspondence curves in a *FRS* with a single interface (glass thickness

is neglected). Therefore, they sample a 3D ray in space. For every sample point, an equation for refractive forward projection of degree four is solved. The resulting 2D image points are connected to a curve by piecewise linear approximation. This approach is also adapted by [SSW13]. A similar approach, which is based on ray tracing, is proposed by [Maa92; Maa15b]. Another ray tracing approach for the calculation of correspondence curves is mentioned by Yamashita et al. [YKK06; YSK10; Yam+11; Nar+12], but not explained in detail.

Findings

The refractive back-projection requires the computation of the ray direction and the point of refraction at every interface and is hence severely more complex than in air. The refractive forward projection is even more computationally expensive, as no direct computation is possible. Instead, equations of degree four (single interface), twelve (two interfaces), or iterative solutions are necessary. Common cost functions that are used in air can essentially be adapted with the aid of the refractive projections. The computation of correspondence curves as counterpart to the epipolar lines in air is also severely more complex, since it directly depends on refractive forward projection. Virtual camera models are an alternative to refractive forward projection, but require the computation of a virtual camera and two projections per 3D-2D point correspondence.

3.3.2 Refractive Indices

In this section, a brief overview on the determination of refractive indices of the participating media will be presented. In the most common case, these are air, water and some kind of glass or perspex. The values for air and water are generally close to 1.0 and 1.33, respectively.

Determination in a Non-submerged System

The first case to be considered is a camera with its protective medium in front of it, but without being submerged in water. There are only few approaches to determine the refractive index of a transparent and flat medium from image data. Shimizu and Okutomi [SO08] use the concept of reflection stereo to calibrate a double-sided half-mirror plate in a complex setup. Chen et al. [Che+11; Che+12] determine the refractive index of the medium from three different images of the scene. The first image is taken without and the other two with the medium in front of the camera. It is necessary that the medium is positioned in two different poses. Agrawal et al. [Agr+12] derive equations that can be used to incorporate unknown refractive indices into their calibration approach. For the case of air-water and air-water-air, this results in solving an equation of degree six and good initial guesses are required. For the case of air-glass-water, the authors could not obtain an analytical equation.

Determination in a Submerged System

In the case of a submerged *FRS*, there are only few approaches that incorporate unknown refractive indices into their optimization problems for calibration. Telem and Filin [TF10] calibrate the refractive index of water, while neglecting the interface thickness. Servos et al. [SSW13] also neglect the interface thickness and determine the refractive index of air and water, but do not evaluate their results in detail. Maas [Maa92] mentions that the incorporation of the refractive indices into his bundle adjustment solution raises some problems that result in questionable estimates. In a later publication however, Maas [Maa95] presents practical tests. According to the author, the refractive index of water could be determined with an accuracy that is better as the one of most commercial refractometers. Similarly, Mulsow [Mul10; MM14] determines the refractive index of water. In his approach, the solution would become singular if all refractive indices would be incorporated. Maas and Mulsow both use a calibration setup with restrictions, as will be briefly presented later on. A completely different strategy is proposed by Yamashita et al. in [YIK05]. The estimation is done in form of an exploratory search by changing the refractive index gradually while the consistency of a 3D measurements is checked. A different approach of the same authors is proposed in [YFK08]. For this approach to work, images of objects that are partly submerged and partly in air are necessary and the system needs to be calibrated. There are also approaches, where the authors report that it is theoretically possible to calibrate the refractive indices, such as in [Ke+08], [Hen+15] or [Li+97], but instead decide to use known values or direct measurements.

Findings

In summary, most of the described approaches require either a special configuration of the imaging system or a special calibration setup. Most of the approaches are too restricted or even too complex to compute. The best reason to omit the determination of the refractive indices during calibration is that they are generally known in the case of air and water. A common referencing of authors (For example [SK11], [Tre+12]) is that according to Mobley [Mob94], the index of refraction of water changes only about 2% in the whole relevant range of water bodies. The changes depend on the optical wavelength as well as water temperature, salinity and depth. A possibility to obtain this value, which is also used by Maas [Maa15b], is an empirical formula proposed in [Höh71]. The refractive index of the protective medium is generally known from the manufacturer. Besides this, the refractive indices can also often be measured directly.

3.3.3 Layer Thicknesses

Besides the refractive indices of the participating media in a refractive system, there are further properties that are important. Following the parametrization described in Section 2.5, these are the thicknesses of the parallel layers that are fixed to the refractive system. Subject to this section are the thicknesses of the fixed layers after the air layer. The number of layers can be arbitrary, but the most common case is the water-glass-air transition with a fixed air and glass layer. In this case, this is only the glass layer.

Disregard

A common strategy, especially for systems comprising a water-glass-air transition, is to simply neglect the thickness of the glass layer. Representatives of this strategy are, for example, [Yam+04], [KYK09] and [CC11]. Telem and Filin [TF10] analyze the effects up to a layer thickness of 4mm and conclude that this quantity can still be considered negligible. Treibitz et al. [Tre+12] propose that the shift caused by the glass layer is smaller in magnitude than the shift caused by angular refraction. Hence, they neglect its thickness, but also mention that as the glass thickness increases, the effect of the shift may become significant. Relying on [Tre+12], Gedge et al. [GGY11] and Servos et al. [SSW13] also neglect the glass layer thickness.

Explicit Determination

Each layer thickness can be part of refractive calibration. Agrawal et al. [Agr+12] present an approach where the thickness of multiple layers in a *FRS* can be obtained linearly when the system axis and the refractive indices are known. Within simulations, the authors show that the thicknesses of multiple layers can be approximated by simpler cases, such as a single layer or two layers. Yau et al. [YGY13] propose an approach that is based on [Agr+12] and hence similar. Chen and Yang [CY14] also incorporate the estimation of layer thicknesses into their calibration approach. After extensive testing with a setup comprising two refractions, the authors strongly recommend that the thickness of the layers should be excluded from the calibration. Yamashita et al. [YIK05] propose a system that can determine the full shape of a refractive surface. For their experiments they use a flat refractive surface, a camera and a point laser on a manipulator. The refractive surface is reconstructed by observation of the laser reflections on the surface, followed by triangulation. The relations between camera, laser and manipulator are assumed to be calibrated beforehand.

Direct Measurement

The most comfortable case is when the layer thicknesses are known. They can most of the times be measured directly or are known from manufacturer data [Tre+12]. Li et al. [Li+97] mention that the thickness could be solved for, but decide to use direct measurements. Belden [Bel13] simulates the thickness estimation. His results are inaccurate estimates and hence, he proposes that the thickness should be fixed and measured. It is very common to assume known layer thicknesses, as can be seen by the vast amount of representatives: [Yam+03], [KS08], [Ke+08],[Mul10], [YSK10], [SK11], [JK12], [Nar+12], [Maa15b], [Han+15] and [Hen+15].

Findings

In summary, the glass layer thickness has to be modeled explicitly, since the amount of its effects can be significant as the thickness increases. Especially in deep sea exploration, cameras need to be protected by housings with a thickness in the range of centimeters.

An aquarium can also reach this range quickly. The glass layer thickness should not be calibrated because of reported inaccuracies and the computational overhead. The best reason to omit the determination of the thickness during calibration is that it often can be measured easily or is even already known from manufacturer data.

3.3.4 Specialized Calibration Approaches

This section is supposed to represent a brief overview on calibration approaches that are characterized by some specializations, which usually require certain setups. The described examples are systems, where the camera needs to be in motion relative to the refractive interface, non-submerged systems or approaches that depend on special calibration devices.

Camera Motion Relative to the Refractive Interface. The focus of this thesis is on optical imaging systems that are characterized by flat refractive interfaces and cameras, which form a fixed unit. On the contrary, there are also systems that do not form fixed units and which are characterized by cameras in motion relative to a refractive interface. An exemplary calibration approach for such a system is proposed by Chang and Chen [CC11]. The authors use a single camera moving around a scene with the water surface in an aquarium being the refractive interface. The camera is equipped with an inertial measurement unit (IMU) to provide necessary information, which is incorporated within a structure from motion framework.
Another exemplary approach is proposed by Mulsow et al. [Mul10; MM14]. The authors make use of calibration objects, which are fixed to the refractive interfaces. They experiment with a flat refractive interface [Mul10] and a more complex one [MM14]. The known calibration objects are fixed to the interface between water and glass as well as to the interface between air and glass. Multiple images of the system need to be taken by moving the camera in front of it.

Calibration of Non-Submerged Systems. Chen et al. [Che+11; Che+12] propose an approach to recover the orientation of a flat refractive interface between a camera and an object in air. Therefore, the scene is captured twice, once with and once without the flat refractive interface. Since the setup contains only an air-glass-air transition, there is just a depth-independent shift due to the refraction in glass. Therefore, the distance to the interface can not be estimated.
Closely related is the calibration approach for so-called reflection stereo systems of Shimizu and Okutomi [SO08]. In such a system, a single camera is imaging through a transparent flat refractive medium (air-medium-air transition). The authors perform experiments with a double-sided half-mirror plate with a 75% reflection ratio to generate the needed, so-called complex images, which contain stereo information from multiple reflections.
Huang et al. [HL14] develop an approach for the calibration of the intrinsic camera parameters in the case of an existing perpendicular flat refractive interface, but not for the refractive parameters.

Special Calibration Devices. Narasimhan and Nayar [Nar+05; NN05] propose a calibration approach in the context of light stripe range scanning. Therefore, they need a camera and a DLP projector outside a water filled aquarium and two known planes with a number of a priori measured 3D points on them placed vertically and at known distances in the aquarium. Their approach is based on the raxel model of [GN05], which means that a pixel on the image is connected to its corresponding ray in water. Therefore, no explicit calibration of refractive parameters is performed. The projector is used to sweep a light stripe over the planes to connect the image points with the 3D points. The remaining points have to be interpolated.
Yau et al. [YGY13] extend the work of Agrawal et al. [Agr+12] by considering the dispersion of light. Therefore, they use a special calibration device for the calibration of a *FRS* with an arbitrary number of layers. The device is described as a watertight acrylic enclosure that contains a number of LEDs. The LEDs are arranged in a grid pattern with known coordinates. Point light sources with a measurable amount of dispersion are generated by one half of a LED emitting light of blue color and one half emitting red color simultaneously. The so-called dispersion constraint is used for estimation of the system axis. For the estimation of the layer thicknesses, the so-called wavelength triangulation constraint is defined. The estimated parameters are refined by non-linear optimization by minimizing the reprojection error. For the approach to work, the refractive indices of the participating media have to be known, the intrinsic parameters of the camera have to be calibrated, the images have to be corrected for chromatic aberrations and the image noise must be reduced in a pre-processing step.
Motivated by the approach of Yau et al. [YGY13], Chen and Yang [CY17] propose an approach for the calibration of a *FRS*, which makes use of triple wavelength dispersion. They determine the system axis in the same way as Yau et al. [YGY13]. Their major contribution is the derivation and proof of a solution for the determination of the air layer thickness that can be applied after the system axis is determined. They do not need any calibration objects. However, to generate the required triple wavelength dispersion, they need a device to illuminate the scene consecutively with red, green and blue light.

Findings

All of the described approaches require special setups. Cameras that are not fixed to the refractive interface can be seen as a special form of *SFRS*. Equipping the camera with an IMU, however, depends on an actual movement to gather the required information. Approaches that require to fix calibration objects to the refractive interface are a common alternative for the calibration of refractive systems, as will be seen in the following sections. However, this is not always realizable. Non-submerged systems could be used to calibrate a refractive system before underwater deployment, but the presented approaches are too restricted by their necessary setups. The utilization of special calibration devices represents a similarly severe restriction.

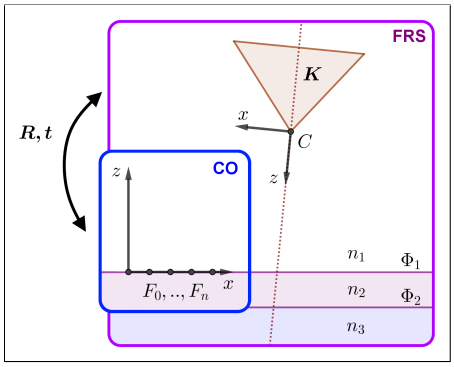

Figure 3.3: Schematic representation: Calibration object (CO) fixed to the refractive interface Φ_1. A transformation $(\boldsymbol{R}, \boldsymbol{t})$ links the CO and the *FRS*.

3.3.5 Flat Refractive Systems

Due to their different parametrization and naming conventions, it is difficult to produce a uniform representation of the various calibration approaches of *FRS* from the literature. Therefore, it is the intention to map the essential properties of the different approaches to schematic representations of the current setup, which will be shown at the beginning of each category. In these schematic representations, the relevant entities and their linkage will be illustrated.

As proposed in the previous section, fixing calibration objects to the refractive interface is a common alternative for the calibration of refractive systems. In the case of *FRS*s, the most convenient way is to use the air-sided interface to fix a calibration object (See Figure 3.3). The relevant entities are the calibration object and the *FRS*, which have their own coordinate system. Hence, only the transformation (rotation $\boldsymbol{R}$, translation $\boldsymbol{t}$) between the two coordinate systems need to be determined. This approach is independent of the layer structure and the refractive indices of the participating media. Note that this and the following schematic representations in this chapter could be extended to an arbitrary number of participating media (n_i, with $i \in \{1, .., n\}$) and interfaces (Φ_i, with $i \in \{1, .., (n-1)\}$).
Yamashita et al. [Yam+04] propose a system for 3D measurement composed of a *FRS* and a slit light laser. The camera and the laser are calibrated in air. A checker pattern is fixed to the refractive interface. Since there is no detailed description, it has to be assumed that some kind of pose from 3D-2D point correspondences algorithm is used together with the calibrated intrinsic camera parameters $\boldsymbol{K}$ to determine the transformation. In [KYK09], the authors replace the slit light laser of the previous approach by a monochromatic light source that is able to project binary patterns. The rest of the setup and the calibration strategy stays the same. Such a pose estimation from 3D-2D point correspondences algorithm is only possible if the camera is able to focus on the interface to acquire a sharp image of the calibration object. Since cameras often are placed very close the refractive

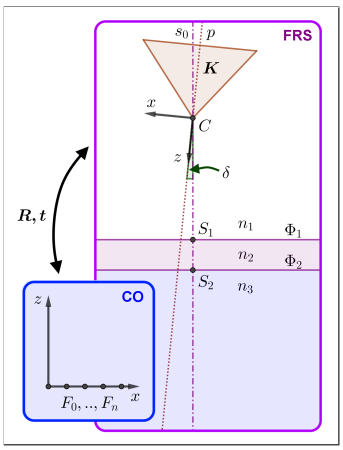

Figure 3.4: Schematic representation: Calibration object (CO) in water. A transformation $(\boldsymbol{R}, \boldsymbol{t})$ links the CO and the *FRS*.

interface, this is generally not the case. The following approaches do not depend on calibration objects fixed to the interface.

Model Deficiencies

Some calibration approaches are based on an underlying model for refractive image formation that is in some way restricted. This can be a necessary perpendicularity between the interface normal and the image plane, dependencies on a necessary amount of layers or incomplete models. Figure 3.4 shows a schematic representation of a *FRS* based on a general model for refractive image formation with three participating media. As a consequence of perpendicularity between the interface normal and the image plane, the system axis s_0 and the principal axis p would coincide ($\delta = 0$). Incomplete models are mostly characterized by neglecting the glass layer thickness. Hence the physical glass layer gets reduced to a single interface ($\Phi_1 = \Phi_2$) and its refractive index gets ignored. Such model deficiencies can reduce the computational complexity if the image formation of the *FRS* can still be approximated sufficiently. However, they pose a problem if the values of the neglected parameters become significant.

Restriction. Treibitz et al. [TSS08; Tre+12] present a calibration method for a *FRS* with an interface normal that has to be perpendicular to the image plane and a neglected glass layer thickness. The authors characterize the system through a ray map of the pixels. Hence, every pixel is connected to a ray in water (raxel model). A caustic surface (See Section 3.3.1) is utilized as the locus of the starting points of the rays. For its calculation, the air layer thickness is needed. Therefore, the authors develop a calibration scheme to determine the intrinsic camera parameters and the air layer thickness. It is required that more than one object of a known length at a known range is imaged and that the objects need to be positioned parallel to the interface. Therefore, the authors use

a checker pattern as calibration object with its lines being the objects of known length. Compared to the schematic illustration in Figure 3.4, this means that the transformation $(\boldsymbol{R}, \boldsymbol{t})$, which links the coordinate system of the *FRS* and the calibration object, is known. Good initial estimates of the parameters are said to be advantageous.
Kang et al. [Kan+15] propose a method for the calibration of a *FRS* with more than two layers. Their theory is derived for the case of perpendicularity between image plane and interface normal, but is extended to arbitrary orientations by a rectification procedure. The estimated parameters are the distances from the center of projection of the camera to every single interface (not each layer thickness) and the orientation of the system axis. A special calibration object is not needed but the refractive indices of all layers need to be known. A restriction of the approach is that at least three refractive interfaces are needed and that the layer thicknesses, except the air layer, need to be measured in advance to generate necessary, additional constraints. It is proposed that using more layers is increasing the accuracy.

Incompleteness. Telem and Filin [TF10] implement a calibration approach for a *FRS*, which is based on a model that neglects the glass layer thickness. The intrinsic camera parameters are estimated in air. The parameters that are determined are the orientation of the system axis, the air layer thickness and the relationship between the refractive indices of the two participating media. These parameters are incorporated into collinearity equations with the aid of the virtual camera model described earlier (See Section 3.3.1). By simulation, the authors evaluate that the calibration can be done using a single image of a planar calibration object. Nevertheless, the preferred strategy is to acquire and use a set of images. In real experiments, the authors use planar calibration objects and additional objects with known dimensions.

The Most General Approach

Models for underwater image formation that are not restricted in the above described ways are more flexible. The following approaches will be further distinguished by their capability to calibrate a single *FRS* or even relate the poses of several, fixed *FRS*s to each other. Figure 3.5 shows a schematic representation of the possible setups. The approaches in this section are in general characterized by a single (FRS_1) or several *FRS*s (FRS_i, with $i \in \{1, .., n\}$) and a calibration object. Transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$, with $i \in \{1, .., m\}$) link the coordinate system of the respective *FRS* and the coordinate system of the calibration object. With these transformations, all the *FRS*s that see the calibration object simultaneously can be linked by relative transformations to each other.

Single Flat Refractive System. Similarly to Treibitz et al. [TSS08; Tre+12], Kunz and Singh [KS08] address the calibration of a raxel model that accounts for refractive effects. The camera is calibrated in air and can be used in a flat or hemispherical refractive system. The glass layer thickness, the refractive indices and optionally the radius of the hemisphere need to be known or measured directly. The parameters for the interface pose are estimated with the aid of a calibration object of known dimensions.

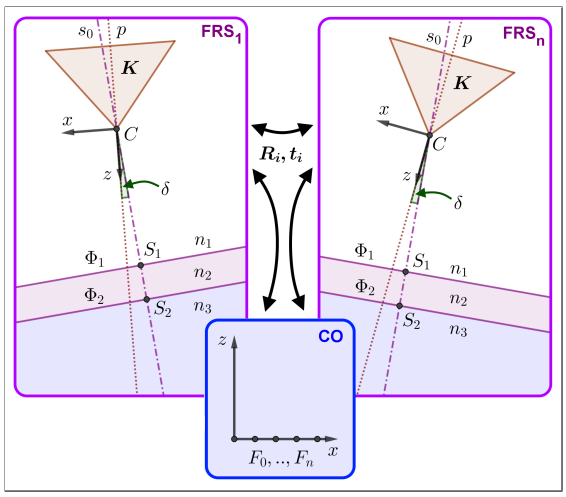

Figure 3.5: Schematic representation: Single (FRS_1) or several, fixed FRSs (FRS_i, with $i \in \{1, .., n\}$), a calibration object (CO) and the transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$, with $i \in \{1, .., m\}$) that link them.

The authors claim that one image of the calibration object in water is necessary and that multiple images are likely to produce more accurate results. The proposed strategy is an optimization of the reprojection error that needs good starting guesses for the parameters. This strategy is only formulated theoretically by the authors.
Agrawal et al. [Agr+12] present an approach for the calibration of a FRS with an arbitrary number of layers. The intrinsic camera parameters need to be calibrated in advance. The calibration consists of the estimation of the orientation of the system axis, all layer thicknesses and the refractive indices. Therefore, the authors derive two constraints. Firstly, the flat refractive constraint, which incorporates all unknowns. Solving it directly is said to be quite difficult because of its complexity. Secondly, the coplanarity constraint, which is independent of the number of layers, their thicknesses and refractive indices. This constraint is used for the estimation of the orientation of the system axis. The authors show how it can be mapped to the 5-point algorithm [Nis03]. The correct solution can be determined after estimation of the layer thicknesses. The layer thicknesses can be obtained linearly with the aid of the flat refractive constraint when the refractive indices are known and analytically when they are unknown. A non-linear refinement of the initial solution is performed using an analytical forward projection (See Section 3.3.1) to minimize the reprojection error. In the presented experiment, three calibration objects with a checker pattern are placed in air behind an aquarium. The approach would also work with a single image of a checker pattern with known coordinates, but it is proposed to use multiple calibration objects or multiple images to increase the accuracy.

Single and Multiple Flat Refractive Systems. Henrion et al. [Hen+15] develop a so-called refraction corrected ray tracing calibration algorithm for a *FRS* (optionally *FRSS*). A single image of a 3D calibration object with 50 known marker locationsis utilized. An optimization to minimize the so-called residual (See Section 3.3.1) is performed by tracing light rays from the camera to the calibration object. The intrinsic and extrinsic camera parameters as well as the interface pose relative to the calibration object are optimized. The glass layer thickness needs to be known. The algorithm needs the 3D coordinates of the marker points and the refractive indices of all participating media as input. Initial values as well as a lower and an upper bound for all of the parameters need to be provided by the user.
Jordt-Sedlazeck and Koch [JK12] propose an approach for the calibration of a single or of multiple *FRS*s. Therefore, besides calibrating each *FRS* separately, the absolute and relative pose between the single systems is estimated. The cameras need to be calibrated for their intrinsic parameters in advance. The thicknesses of the glass layers and the refractive indices of the participating media are known. Multiple images of a calibration object with a checker pattern with known corner coordinates are taken in water. These are used within the approach of [Agr+12] for individual initialization of all parameters per *FRS*. As the authors want to avoid the optimization of the reprojection error due to the complexity of refractive forward projection, an evolutionary optimization-algorithm (Covariance Matrix Adaptation Evolution-Strategy) is coupled with an analysis-by-synthesis (AS) approach. The AS approach consist of rendering the images of the checker pattern, comparing these to the real ones and minimizing the differences. This is intended to eliminate errors introduced by inaccuracies in corner detection and has the advantage that only refractive back-projection is necessary. The whole approach is said to be time-consuming.
Haner et al. [Han+15] propose a calibration approach for a special multi-camera setup containing a pool and cameras in air and in water. Every camera is calibrated for its intrinsic parameters in air before installation. The underwater cameras are *FRS*s. All refractive indices and the thickness of the glass are known. A special calibration object is used that is composed of a straight rod with eight markers at known locations. Half of this calibration object is submerged. The camera poses are initialized by measured information relative to the pool and the interface poses are initialized by assuming them to be at the pool walls. Each pose of the calibration object is determined for the camera with the best view of it. Bundle adjustment is used to refine the initial values for the interface poses, camera poses and calibration object poses by minimizing the reprojection error.

Findings

The calibration of a *FRS*, if it is not based on an underlying model for underwater image formation that has one of the described deficiencies, is the most general approach. This is because all the other refractive systems that will be covered in the following sections are specializations of the *FRS*. Therefore, besides calibration of the *FRS* itself, extensions to calibrate the relative poses between two or more *FRS*s are advantageous. All of the

presented approaches have in common that a calibration object with known feature point coordinates is necessary.

3.3.6 Flat Refractive Stereo Systems

*FRSS*s are one example of refractive systems that are a specialization of the *FRS*. In contrast to the approaches from the previous section that are able to calibrate the relative pose between several *FRS*s additionally to the single *FRS* itself, it is now mandatory that two *FRS*s exist. A single *FRS* can not be calibrated with these approaches.

Similarly to Section 3.3.5, calibration approaches that are based on an underlying model for underwater image formation that is restricted occur in this category as well. Kang et al. [KWY12b] propose a calibration approach for a *FRSS* with both image planes being perpendicular to the respective interface normal. Additionally, the layer thickness of the housing is neglected. The intrinsic camera parameters and the refractive indices of the participating media need to be known in advance. The authors use the concept of the refrax (See Section 3.3.1) and develop two new concepts, namely the ellipse of refrax and the refractive depth. These are used in an algorithm for the estimation of the relative translation of the stereo camera, air layer thicknesses and 3D structure. This algorithm is incorporated into a differential evolution approach to search for the best rotation between the two cameras. Afterward, the parameters and the 3D structure are refined by bundle adjustment to minimize the reprojection error. In a secondary version of their work, Kang et al. [Kan+17] extend their bundle adjustment approach by a local sliding procedure for an improvement of the convergence behavior of their optimization framework.
The authors in [Brä+15a; BKN15; Brä+15b] develop an underwater 3D-scanner, which is composed of a projection unit and a *FRSS*. The authors propose an approach for the calibration of the air layer thickness for each camera. A requirement is that the interfaces are perpendicular to the image planes and that the refractive indices of the participating media are known. The layer thicknesses of the housing need to be known as well and the stereo camera system needs to be calibrated for its intrinsic and extrinsic parameters in advance. The calibration of the air layer thicknesses is realized with the aid of measurements of a plane and an object of known length. Therefore, a systematic search is performed in the range of minimally and maximally possible air layer thicknesses with a certain step width in order to calculate 3D points for the objects to be measured. The measurement error is minimized.

Dependencies on Calibration Objects

Underlying models for underwater image formation that are not restricted by the above described perpendicularity are the more flexible ones. Figure 3.6 shows schematic representations of possible setups for *FRSS*s. These are characterized by two single *FRS*s (FRS_1 and FRS_2) that are based on general models for refractive image formation, respectively. A calibrated stereo camera allows the reconstruction of 3D coordinates of

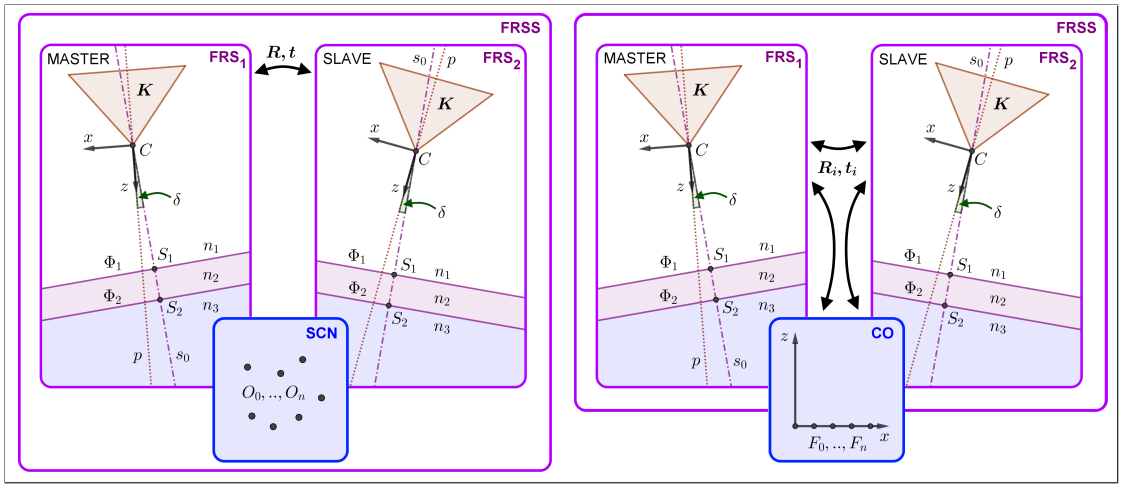

Figure 3.6: Schematic representation: *FRSS* composed of two single *FRS*s (FRS_1 and FRS_2). Left: Linking by 3D reconstruction of object points O_i (with $i \in \{0, .., n\}$) in the scene. Right: Calibration object (CO) and the linking transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$, with $i \in \{0, .., m\}$).

object points in the scene with respect to the coordinate system of the master camera. This is also the case for *FRSS*s (See left side of Figure 3.6). The advantage is that this relaxes the dependency on calibration objects with known feature coordinates (See right side of Figure 3.6). This in turn avoids the need to know or to determine the transformation between the coordinate systems of the calibration object and of the *FRSS*. As will be seen in the following, not many approaches make use of this possibility.

Known Coordinates of Features on Calibration Objects. Li et al. [Li+97] propose an underwater photogrammetric model based on a 3D ray tracing technique. The authors formulate a set of equations for the intersection of corresponding left and right rays in water. The stereo camera is calibrated for its intrinsic and relative extrinsic parameters in advance. Each glass layer thickness and the refractive indices of the participating media are known. The unknown parameters are estimated by linearization of the formulated equations, which then are solved using the least squares principle [Li95]. For calibration, a 3D frame with known 3D coordinates is necessary and a set of stereo images of it needs to be acquired under water.

No Calibration Object. In the approach of Sedlazeck and Koch [SK11] the intrinsic parameters of both cameras are calibrated in advance and the glass layer thicknesses need to be known. The estimated parameters are the interface poses and the absolute as well as the relative extrinsic parameters of the stereo camera. There is no need for a calibration object, since stereo correspondences, which are matched by using *SIFT* features, are sufficient for the estimation. Initial values for the camera poses and 3D scene points are necessary. An approximate intrinsic camera calibration by the method

of [FF86], [LRL00; LRL03] (approximation of refractive effects by focal length and radial distortion) is used to refine the initial camera poses and the 3D scene points by bundle adjustment. This results in an approximation of the scene, which is used to find the interface poses by a further bundle adjustment step. Together with the true intrinsic parameters of the cameras, this optimization is done in nested loops. An inner loop minimizes the reprojection error to estimate the air layer thickness and an outer loop minimizes a virtual error function for estimation/refinement of all remaining parameters. This virtual error function is derived from the authors virtual camera model described in Section 3.3.1, which is based on a caustic surface.
Chen and Yang [CY14] propose an approach for the calibration of a *FRSS* with an arbitrary number of layers. The stereo camera is calibrated for its intrinsic and relative extrinsic parameters in a pre-processing step and the number of layers as well as their refractive indices need to be known. Estimated are the interface orientations and the layer thicknesses. No calibration object is needed, since stereo correspondences generated by a projector, which projects gray code structured light patterns onto the scene, are utilized. The important finding of the authors is that the layer thicknesses can be estimated if the orientations of the system axes are known. Therefore, they formulate a constraint, which makes the determination of the layer thicknesses possible by solving a linear set of equations. To determine the axes, a binary search over a defined search space is performed for both cameras. For every combination of axes during this search, the layer thicknesses need to be determined, which is followed by a refractive back-projection to compute 3D points for every stereo correspondence. These 3D points are used to compute the reprojection error as a measure of parameter accuracy. Therefore, by referring to [Agr+12], an equation of degree four for a single refraction and of degree twelve for two refractions needs to be solved. Although the approach is said to be able to handle an arbitrary number of layers, it is not mentioned what needs to be done if there are more than two interfaces. All the determined parameters are refined by sparse bundle adjustment. After experiments with this approach, the authors remark that one should try to avoid the cases when both the axes and the glass layer thicknesses are unknown. They recommend that the glass layer thicknesses should be measured instead.

Findings

As within the calibration approaches for *FRS*s, a provided perpendicularity between the interface normals and the image plane simplifies the calibration problem, but restricts the setup severely. If the underlying image formation model is not restricted in this way and if a stereo system is available, which allows a refractive calibration from stereo correspondences, suitable calibration approaches amount to a powerful tool. The disadvantage is that the calibration of a single *FRS* is missing. On the contrary, they have the advantage that no calibration object with known feature coordinates is necessary. This advantage outweighs in particular if otherwise a 3D calibration object has to be handled underwater, which can be quite cumbersome.

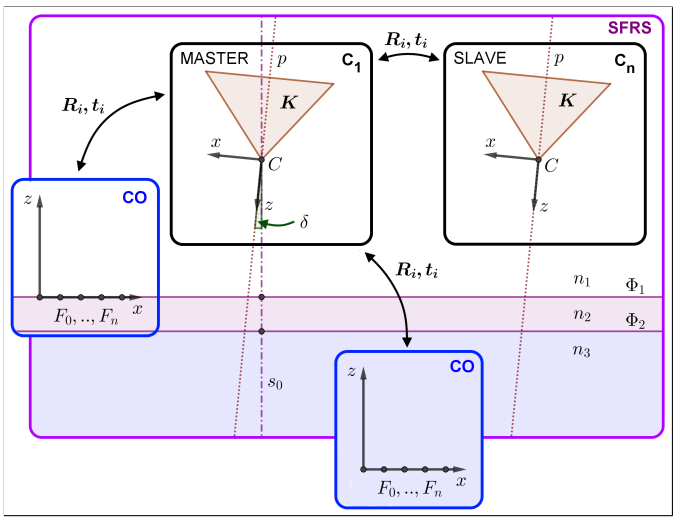

Figure 3.7: Schematic representation: *SFRS* composed of single cameras (C_i with $i \in \{1, .., n\}$) and calibration objects (CO) on the refractive interface Φ_1 or in water and the linking transformations ($\boldsymbol{R}_i, \boldsymbol{t}_i$ with $i \in \{1, .., m\}$).

3.3.7 Shared Flat Refractive Systems

*SFRS*s are another specialization of *FRS*s. Figure 3.7 shows a schematic representation of possible setups for *SFRS*s. The number of cameras amounts to at least two and all of them (C_i with $i \in \{1, .., n\}$) have in common that they are imaging through the same flat refractive interfaces (here: Φ_1 and Φ_2). Since the refractive parameters have to be determined only with respect to the master camera, this has the advantage that fewer parameters need to be determined. In this case, the refractive parameters of the slave cameras are encoded implicitly within the relative extrinsic camera parameters. As in both categories before, approaches with an underlying, incomplete image formation model exist as well as approaches that depend on known coordinates of features on calibration objects, which are either fixed to the refractive interface, have to be handled underwater or both (See Figure 3.7).

Gedge et al. [GGY11] present a method for the calibration of a *SFRS*, whereby the glass layer thickness is neglected. A stereo camera, which is calibrated for its intrinsic and extrinsic parameters in advance, is utilized. The refractive index of water is needs to be known. All the parameters have to be initialized. A calibration object with a checker pattern is imaged multiple times underwater. The feature correspondences are used in a non-linear optimization procedure. Therefore, the authors derive a cost function that represents the distance between the refracted rays of a feature correspondence.

Ke et al. [Ke+08] present an approach a method for the calibration of a *SFRS*. The intrinsic and extrinsic parameters of the stereo camera are calibrated in advance. The refractive indices of the participating media and the glass layer thickness need to be known. The interface is parametrized as an equation of a plane in the master camera coordinate

systems. The proposed optimization is based on a bundle adjustment procedure that minimizes the reprojection error for the feature points on a submerged, planar calibration object. To ensure that the optimization converges, it is said that it is essential to determine reasonable initial values for the interface parameters. This is done by reconstruction of another planar calibration object, which is placed parallel and close to the air-sided interface, and which is therefore essentially fixed to this interface. The approach is evaluated by numerical simulations.
Maas incorporates his multimedia module (See Section 3.3.1) into a self-calibrating bundle adjustment for two [Maa15b] or three [Maa92] cameras forming a *SFRS*. Therefore, a complex linearization of observation equations needs to be performed. The refractive index of water and the glass layer thickness are known during the executed experiments. The author uses a custom made calibration object with 85 known point coordinates on two levels [Maa92], which can be fixed in water at different distances from the interface with the aid of spacers of known length. Hence, the pose of the refractive interface is already known with respect to the calibration object and the extrinsic and intrinsic camera parameters are estimated.

Findings

A *SFRS* leads to the advantage that the number of parameters can be reduced significantly. However, this is accompanied by a restriction of the flexibility in system design. Besides reducing the number of parameters, the approaches above do not provide any major findings that have not already been described in the two previous sections.

3.3.8 Categorization and Aptitude Analysis

In this section, the previously listed calibration approaches for *FRS*s and for their specializations will be analyzed to generate a tabular representation for their comparison. This representation will be used to identify suitable approaches for the target system of this thesis. The calibration approaches to be developed in this thesis will be located in this representation as well.

Categorization and Representation

The results can be seen in Table 3.1. It is organized in three major columns, which include the authors of the respective approach, the underlying image formation model of the system the approach is designed for and its restrictions. All these systems and restrictions have been described in detail in the previous three sections and will be summarized briefly in the following.
The order of the subdivisions of the 'Models' and 'Restrictions' columns was created as follows. The order of the 'Models' columns from left to right represents an increase in specialization. As described earlier, the *FRS* is the most general one. Theoretically, it is possible to calibrate multiple *FRS*s if there is a mechanism to relate the pose of the single systems to each other. Since it is most of the times not the case that such a mechanism is presented, the markings in the tabular representation are placed according

to the systems the approach is explicitly designed for by the authors. Possible extensions to more specialized systems are not considered. If a marking is accompanied by the character '+' in brackets, this stands for multiple instances of a *FRS*. The *FRSS* is the first specialization. In this case, a combination of two *FRS*s is mandatory. The second specialization is the *SFRS*. It is characterized by multiple cameras in front of a single, shared refractive interface. Markings in the column of the *SFRS*s are accompanied by the number of cameras, which can either be specified (For example 2) or arbitrary (n). The order of the 'Restrictions' columns from left to right is at most a proposal. There is no measure for the significance of the single restrictions. The order could be exchanged. The 'Restricted Model' and the 'Incomplete Model' have been described earlier as model deficiencies. The 'Restricted Model' is primarily characterized by perpendicularity between the image plane and the interface normal and the 'Incomplete Model' by neglecting the thickness of an existing layer between air and water. Both of them pose particular problems if values of the neglected parameters become significant. The 'Features on Interface' stands for calibration objects that need to be fixed to an interface. This can be on the air-sided or on the water-sided interface, which is not realizable in general. These calibration objects contain feature points, which are easily detectable and which are usually, but not necessarily, characterized by known 3D coordinates in the coordinate system of the respective calibration object. Therefore, 'Features on Interface' is closely connected to 'Known Feature Coordinates'. If the known feature coordinates are necessary, this is usually accompanied by a cumbersome handling of calibration objects under water. The final category is the 'Initialization of Parameters'. This means that all the parameters that get optimized within the calibration approach need to have initial values, which can range from default values to educated guesses close to the real values.

Aptitude Analysis

The target system in this thesis is a stereo camera that is imaging through a single flat refractive interface of known thickness. The most suitable parametrization originates from the earlier proposed model for a *SFRS*. With the research goals to realize a calibration approach for the determination of the refractive parameters and to reduce the dependence on calibration objects, Table 3.1 can be analyzed as follows:

1) It is sufficient to be located in the 'Models' column 'Shared Flat Refractive System'.

2) The markings in the 'Restrictions' columns are zero.

Concerning point 2, an exception could be made for the 'Initialization of Parameters' column. The other four columns represent exclusion criteria. Such an initialization does not contradict the research goals of this thesis. However, regarding this relaxation, the only approach becoming eligible is the one of Sedlazeck and Koch [SK11] and the only approach that is fulfilling both criteria completely is the one of Chen and Yang [CY14]. All the other approaches in the table meet at least one of the exclusion criteria.
Concerning the underlying image formation model, the approaches of Sedlazeck and Koch [SK11] and Chen and Yang [CY14] are a little more general than the approaches for the

Authors	Models			Restrictions				
	Flat Refractive System (*FRS*)	Flat Refractive Stereo System (*FRSS*)	Shared Flat Refractive System (*SFRS*)	Restricted Model	Incomplete Model	Features on Interface	Known Feature Coordinates	Initialization of Parameters
Yamashita et al. [Yam+04]	x				x	x	x	
Treibitz et al. [TSS08; Tre+12]	x			x	x		x	x
Telem and Filin [TF10]	x				x		x	
Kang et al. [Kan+15]	x			x				
Kunz and Singh [KS08]	x						x	
Agrawal et al. [Agr+12]	x						x	x
Henrion et al. [Hen+15]	x (+)						x	x
Jordt-Sedlazeck and Koch [JK12]	x (+)						x	x
Haner et al. [Han+15]	x (+)						x	x
Kang et al. [KWY12b; Kan+17]		x		x	x			
Bräuer-Burchard et al. [Brä+15a; BKN15; Brä+15b]		x		x			x	
Li et al. [Li+97]		x					x	
Sedlazeck and Koch [SK11]		x						x
Chen and Yang [CY14]		x						
Gedge et al. [GGY11]			x (2)		x			x
Ke et al. [Ke+08]			x (2)			x	x	x
Maas [Maa15b; Maa92]			x (n)			x	x	
Own Approaches			x (2)					

Table 3.1: Categorization of calibration approaches for *FRS*s.

calibration of a *SFRS* to be developed in this thesis. The approach of Sedlazeck and Koch [SK11] is characterized by nested optimization loops to minimize multiple cost functions and by the utilization of the caustic surface for the locus of the viewpoints. Compared to this, the approaches to be developed will be more straightforward, simpler to apply, more efficient and will not need an initialization. The more recent approach of Chen and Yang [CY14] will be used as the reference approach for the refractive calibration of a *SFRS*. The approaches to be developed in this thesis will be compared to an adapted reimplementation of this reference approach in Chapter 7. While Chen and Yang [CY14] concentrate on the development of an optimization problem for calibration, this thesis will comprise the development of a model whose integration into the underlying image formation models benefits the computation of refractive projections, the computation of adapted and novel cost functions, the definition of optimization problems and therefore all aspects of refractive calibrations. Furthermore, this model will enable the necessary extensions of all the concepts of stereo 3D reconstruction in air for underwater environments.

3.4 Summary

Refractive effects, such as those in a *FRS*, must not be ignored in the field of underwater computer vision. Nevertheless, there are approaches in underwater 3D reconstruction where this is done. These are mostly approaches where accuracy is not of major concern and ground truth data does not exist. However, ignoring refractive effects at all must be avoided.

A different strategy is the absorption of refractive effects by the conventional camera parameters. The advantage of this strategy is the simple application of algorithms made for in air usage. The biggest disadvantage is the requirement of an image plane that is perpendicular to the interface normal. Besides an absorption of refractive effects by the parameters for focal length and radial distortion, there is an absorption by camera pose parameters in multi-camera setups. The resulting absorption leads to erroneous parameters and is incomplete, since refractive distortion is not an image space distortion, but depends on the object distance in 3D space.

This chapter has shown that refractive effects need to be modeled explicitly. As a consequence, refractive back-projection and forward projection become complex and computationally expensive operations. There is no approach to compute the refractive forward projection of 3D points onto the image plane directly. The computational intensity is especially inconvenient since both projections are basic operations on which the concepts of stereo 3D reconstruction depend on. One example are cost functions, which are used during calibration for feedback on the accuracy of the estimated parameters and their optimization. The literature has shown that not all of the refractive parameters need to be part of the calibration process. The most common physical setup is a water-glass-air transition, where the glass layer thickness can be measured physically. The refractive indices of the participating media are usually measurable or even commonly known. This leaves the air layer thickness and the orientation of the system axis as the requested refractive parameters. Some of the presented models for underwater image formation with consideration of refraction suffer from deficiencies. These can be restrictions, such as perpendicularity between image plane and interface normal, or these can be incomplete models, which, for example, neglect the glass layer thickness. The analysis has shown that all of the considered approaches for the calibration of *FRS*s, without any of these model deficiencies, depend on known coordinates of features points on the calibration object. This usually requires the estimation of additional transformation parameters. Although the approaches for the calibration of a *FRS* are the most general ones, the more specialized stereo and multi-camera systems provide the possibility of a calibration utilizing only stereo correspondences. Thereby, the cumbersome handling of calibration objects under water and the determination of additional transformation parameters can be avoided.

The approach to be developed in this thesis has been located in a categorization scheme together with the related approaches. It has been shown that almost all of the considered approaches meet one or more of the defined exclusion criteria. With the approach of Chen and Yang [CY14], a reference approach for the refractive calibration of a *SFRS* has been identified.

4 Virtual Object Points

In this chapter, the challenges that have to be met in this thesis will be specified more precisely and the solutions to be developed will be motivated. Since it is obvious from the previous chapters that refractive effects must not be ignored, the research goal to extend the concepts of stereo 3D reconstruction in air for underwater environments by an explicit consideration of refractive effects can be confirmed to be appropriate. As mentioned before, the underwater imaging system will be characterized by the model for a *SFRS*. In the following, it will be motivated that the extension of this model with a *VOP* model amounts to an efficient tool for the achievement of the research goals from Section 1.3. In this chapter, it will be the goal to acquire a basic model for *VOP*s, which can be readily integrated into the presented underwater image formation models. Therefore, the previously presented findings on *VOP*s will be analyzed. In the following, the *VOP* location that will be utilized in the remainder of this thesis will be determined. This will be done based on the physically founded apparent depth in the case of a single flat refractive interface, which has been presented in Section 2.3.3, and on the findings from Section 3.3.1. Since there are multiple valid possibilities for the location of *VOP*s, it is necessary to make a selection amongst them. Therefore, the impacts of two flat refractive interfaces on the validity of all the considered possibilities will be analyzed. Before making a decision, it is most important to assign the *VOP*s to the real object points and vice versa. As will be seen, an obvious selection can be made. The final step in this chapter will be the formulation of a model on how to relate *VOP*s to real object points and vice versa in the case of two flat refractive interfaces. This *VOP* model represents one of the major contributions of this thesis. This chapter is partially based on Dolereit et al. [DK14; DL16].

4.1 Motivation and Challenges

The basic building block of any 3D reconstruction approach is the underlying image formation model. In this case, this is the *SFRS*, as defined in Section 2.5. The major challenge is the development of particularities of this model that can be utilized for the extension of the concepts of stereo 3D reconstruction in air for underwater environments, especially refractive calibration. Up to now, a detailed analysis of the *SFRS* is missing in the literature. The primary argument in favor of it is that there are properties that are beneficial and worth to adjust the basic design of the imaging system accordingly. It will be shown that the extension of this model with the developed *VOP* model amounts to an efficient tool for the achievement of the research goals from Section 1.3.
As mentioned before, the most important concepts of stereo 3D reconstruction are the projections for coordinate transformations between 3D and 2D spaces. Considering an underwater environment, these are refractive forward projection and refractive back-projection, which are both based on the underlying image formation model and its parametrization. These two projection directions are essential for all the remaining concepts, such as refractive calibration and recovery of 3D coordinates. However, a big challenge is that both transformations are computationally expensive, especially the refractive forward projection (See Section 3.3.1). Since the focus of this thesis is on calibration of the refractive parameters of a *SFRS*, the implementation of some kind of refractive projection is unavoidable. An example are cost functions, which are a vital part of calibration approaches, since they reflect the quality of the determined parameters. The most common cost function is the reprojection error, which is based on refractive forward projection. An alternative is some kind of 3D scene error. This can for example be the result of the triangulation of two rays, which would therefore be based on refractive back-projection.

Contributions for Refractive Projections. The *VOP* model to be developed will be the foundation of most of the developments and can be utilized within monocular and binocular vision. It will be shown that together with this model, the challenge to reduce computational expenses can be met. For example, simplifications in both projection directions will be proposed. These will be realized by the integration of the *VOP* model into the more general model for a *FRS*. Hence, these simplifications are readily usable in stereo and non-stereo systems. Additionally, it will be shown that the integration of the *VOP* model enables alternative coordinate transformations between 3D and 2D spaces. One example is to avoid refractive forward projection at all by definition of a virtual camera model that enables conventional forward projection. The virtual camera model that will be defined in this thesis will also be usable in stereo and non-stereo systems. Besides the above mentioned simplification, an efficient alternative to refractive back-projection will be proposed. It is based on the integration of the *VOP* model into the model for a *SFRS* and it is especially useful for the determination of the refractive parameters.

Contributions for Refractive Calibration. The developments of the coordinate transformations between 3D and 2D spaces will be accompanied by the development of novel cost functions for refractive calibration. Another vital part of calibration approaches is the definition of an optimization problem. It will be shown that the proposed *VOP* model enables the determination of the refractive parameters partly by linear optimization. An important finding that will be presented is that the refractive parameters can be determined independently of each other. A question that will be answered in this context is if this is advisable.
Altogether, the development of various approaches for refractive calibration and hence the development and the efficient computation of cost functions and of suitable optimization problems will be presented. The most preferable result is an approach for refractive calibration that is independent of calibration objects. Therefore, the developments will be accompanied by an analysis of a refractive calibration with or without calibration object, with simultaneous or consecutive estimation of the refractive parameters and by a comparison of the utilized projection methods and cost functions. Since known feature coordinates on a calibration object were defined to be an exclusion criteria in Section 3.3.8, the analysis with calibration object comprises testing if the refractive parameters can be determined from arrangements of feature points only.

Contributions for the Recovery of 3D Coordinates. It will be shown that the *VOP* model can furthermore be utilized to realize valuable contributions on the remaining concepts of stereo 3D reconstruction. These will cover the actual recovery of 3D coordinates of the imaged scene with the aid of stereo matching. Due to the 3D-dependence of refractive effects, epipolarlines are not straight lines any more. Therefore, an alternative in form of the computation of correspondence curves, will be proposed.

Overview. For the extension of the concepts of stereo 3D reconstruction in air for underwater environments, a *VOP* model will be developed, which enables:

- the utilization within monocular and binocular vision,
- the simplification of refractive forward projection,
- the simplification of refractive back-projection,
- the definition and implementation of a virtual camera model to avoid refractive forward projection and hence to enable conventional forward projection,
- the implementation of an efficient alternative to refractive back-projection,
- the definition of novel cost functions,
- the consecutive and the simultaneous estimation of refractive parameters, with and without calibration object, which can be party realized by linear optimization,
- the computation of correspondence curves and
- the recovery of 3D coordinates.

4.2 Position Determination

Let us consider the case of a real object point O in water, as illustrated in Figure 4.1. The ray w in water, which is emitted from O, is refracted at the interface Φ. The result is the ray a in air. Up to now, three possibilities for the *VOP* location were introduced. These are the two physically founded ones from Section 2.3.3 and the one proposed by Maas [Maa95] in Section 3.3.1. All three are situated on the virtual extension of the ray a. Summarized, these are

- the point $\check{O}_m$ (m - meridional, see Section 2.3.3) being the result of the intersection of two rays with an infinitesimal different angle of incidence in the same plane of refraction,
- the point $\check{O}_s$ (s - sagittal, see Section 2.3.3) being the result of the intersection of two rays with the same angle of incidence in different planes of refraction and
- the point $\check{O}_M$ (M - Maas [Maa95]) being the intersection of the ray a with a plane Γ passing through the real object point O, parallel to the interface Φ.

In this thesis, a *VOP* model will be utilized to reproduce the underwater image formation process with the aid of the pinhole camera model, which has been shown to be invalid. To make the pinhole camera model valid for underwater image formation, every real object point needs to be related to a *VOP* on ray a. An important circumstance is that 3D point objects are projected to 2D point image features. A direct consequence is that it theoretically does not matter which one of the *VOP* locations is used for image formation. As long as it is situated on the same ray, it will result in the same image point. Therefore, the choice of the location becomes a question of convenience in terms of relating it to the real object point. The final choice has been proposed in Dolereit and Kuijper [DK14]. In the following, the properties of this relation, which were leading to this choice, will be analyzed in detail.

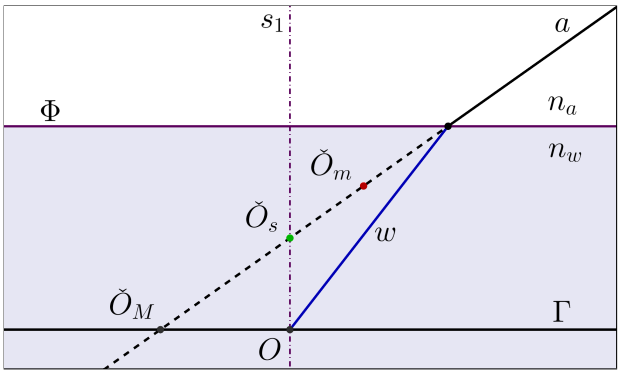

Figure 4.1: The possible *VOP* locations $\check{O}_m$, $\check{O}_s$ and $\check{O}_M$.

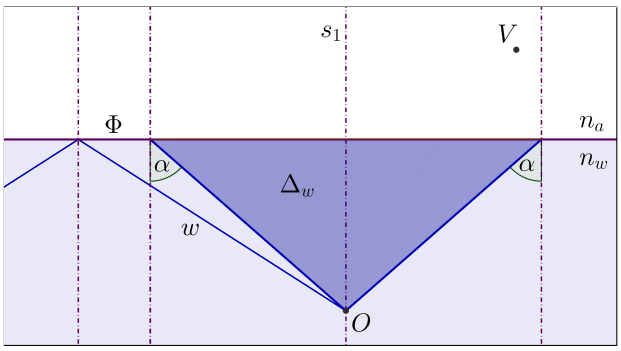

Figure 4.2: Front view of a right circular cone that bounds the possible starting directions of an unknown optical path between a real object point O and a viewpoint V.

4.2.1 Single Flat Refractive Interface

Let us assume that only a real object point O in water and the viewpoint V of an observer in air are known, as illustrated in Figure 4.2. According to the law of refraction, there is only one correct optical path that connects both points. Due to the unknown starting direction of the ray in water, the optical path can not be determined directly, making its recovery computationally expensive. This is equivalent to the problem of refractive forward projection. In contrast to the real object point O, the related VOP lies on the up to now unknown ray in air, which is passing through the viewpoint V. To make the best possible use of this later on, the relation between the virtual and the real object point should be as convenient as possible. To analyze the properties of this relation and due to the unknown starting direction of the ray in water, the VOPs will be determined and traced for a significant amount of all the possible directions of the rays that are emitted from the real object point O.

Traces of Virtual Object Points

The entirety of the possible directions of the emitted rays are limited by a right circular cone Δ_w, with the object axis s_1 as its axis and a radius that is limited by the occurring total reflection of the rays in water, as illustrated in Figure 4.2. For the transition from water to air, the maximal angle of incidence α is about 49 degrees. Rays outside this cone, such as w, exceed this maximal angle and get reflected at the interface Φ. The ray in water, which forms the beginning of the optical path in question, is part of Δ_w. Note that the cone Δ_w from Figure 4.2 is reduced to its circular outline on the left side of Figure 4.3. Both images in Figure 4.3 are visualizations of the traces of the three possible VOP locations. These traces are the result of a representative movement of the point of refraction R of the ray w on the refractive interface Φ. This movement consists of the circular outline of the base area of the cone Δ_w and a straight line passing through the object axis s_1 on the base area. It represents a significant amount of all the possible directions of the ray w. The traces are regulated by the previously presented

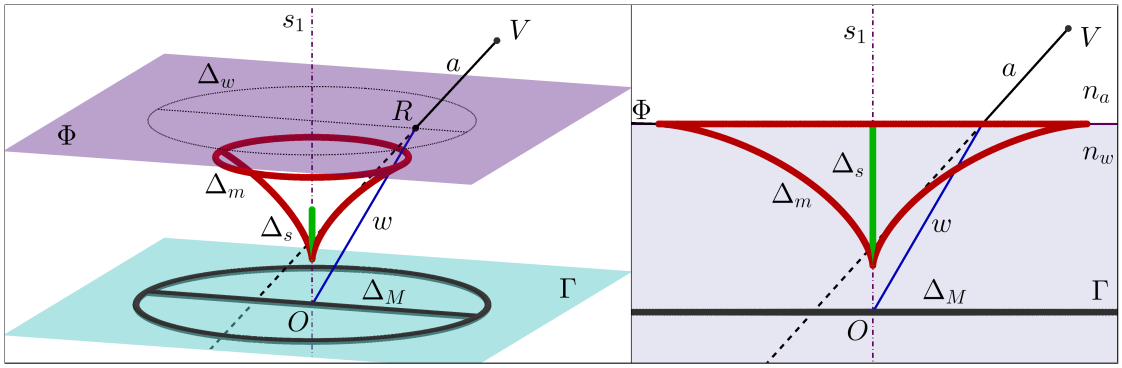

Figure 4.3: Traces of the *VOP* locations in the case of one refractive interface. Left: 2D front view. Right: Perspective 3D view.

determination processes of the three proposed locations $\check{O}_m$, $\check{O}_s$ or $\check{O}_M$, respectively. In each case, the result is a 3D structure.
The left side of Figure 4.3 illustrates the resulting traces of the *VOP* locations up to an optical path close to total reflection and the right side shows the resulting traces up to total reflection. This differentiation is done for visualization purposes of the trace Δ_M. Both sides show the one optical path in question that truly connects O and V.

- In the case of $\check{O}_M$, the result is the circular area Δ_M on the plane Γ. If the angle of incidence of the ray w in water tends to total reflection, the radius of Δ_M tends towards infinity.

- The result in the case of $\check{O}_m$ is the 3D surface Δ_m. Its extent along the object axis s_1 is limited at the bottom by a ray w with an angle of incidence that amounts to zero degrees and can therefore be computed by Equation 2.25. The upper end is limited by the rays whose angles of incidence tend to total reflection and therefore, by the flat refractive interface Φ. On Φ, the base area of Δ_m coincides with the base area of the cone Δ_w. At the bottom, the apex of the structure intersects the object axis s_1.

- In the case of $\check{O}_s$, the result is the straight line segment Δ_s along the object axis s_1. The limits are the same as for Δ_m. Therefore, the line segment reaches the refractive interface Φ at the upper end and coincides with the apex of Δ_m at the bottom.

These traces make clear that the assignment of *VOP*s to a single real object point is ambiguous if no further information on the optical path in question is known. The 3D structures Δ_M, Δ_m and Δ_s must contain the *VOP*s that lie on the virtual part of the ray a in question. The most convenient case is indisputably Δ_s. It has the least spacial extent and the simplest 3D structure. Furthermore, its range can be determined easily. An argument against Δ_M is its infinite extent and an argument against Δ_m is a resulting

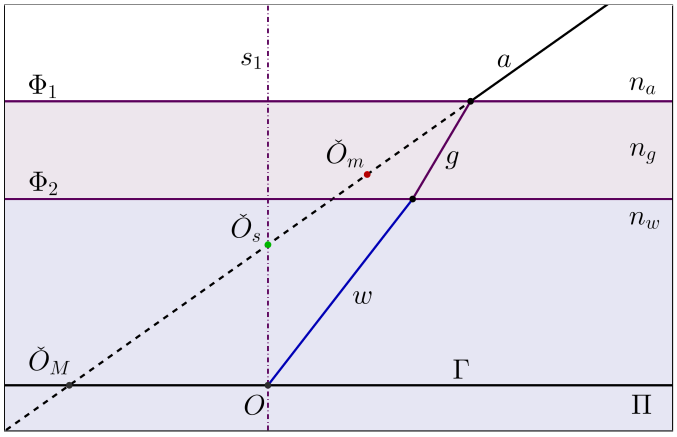

Figure 4.4: *VOP* locations in the case of one refractive interface. Left: 2D front view. Right: Perspective 3D view.

ambiguity. As can be seen on the right side of Figure 4.3, the trace Δ_m intersects the ray a in question twice due to its symmetry. The respective traces can be considered as a search space for the determination of the true optical path between a given real object point O and a viewpoint V. Hence, the reduction of Δ_s to being a single straight line is quite advantageous. Now let us take a look at the assignment of real object points to *VOP*s. With a known viewpoint one can assign a real object point non-ambiguously to a *VOP*. This is possible since the ray passing through a viewpoint and a *VOP* allows the direct computation of its angle of incidence. In the case of $\check{O}_s$, this in turn allows solving Equation 2.26 directly, which has been previously used to express the relation between the location of the virtual and the real object point.

4.2.2 Two Flat Refractive Interfaces

All of the previous considerations have been performed for the case of a single flat refractive interface. Since the housing of an underwater camera or a side of an aquarium has a thickness, there are two parallel refractive interfaces and the rays are refracted twice. This is illustrated in Figure 4.4. The two flat refractive interfaces Φ_1 and Φ_2 bound the additional layer, which is usually some kind of protective glass with a refractive index n_g. The proposed possibilities for the *VOP* locations $\check{O}_m$, $\check{O}_s$ and $\check{O}_M$ essentially remain, but it has to be analyzed if these locations stay valid in this setup, especially in the case of the physically founded locations based on the concept of the apparent depth. As illustrated in the case of $\check{O}_m$, it is possible that the *VOP*s are situated inside the glass layer as well.

Adapted Derivation of the Apparent Depth

The parallelism between the two flat refractive interfaces Φ_1 and Φ_2 implies that both interfaces share the same interface normal (See Figure 4.4). As is known from physics,

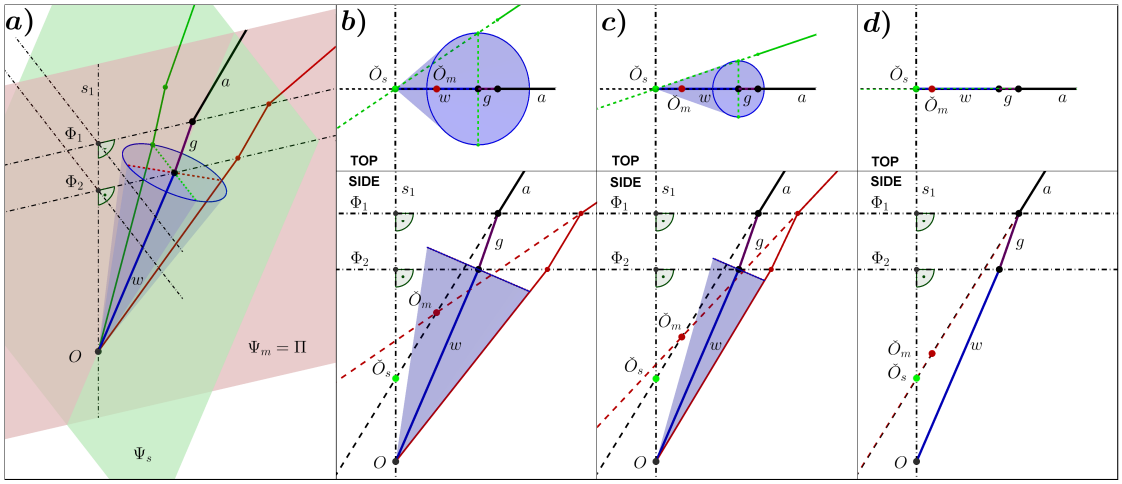

Figure 4.5: Derivation process of the *VOP*s locations in the case of two flat refractive interfaces.

refraction happens in a plane. Hence, the ray a in air, the ray g in glass and the ray w in water share the same plane of refraction Π. The existence of two flat refractive interfaces requires to check the impacts on the *VOP* locations $\check{O}_m$ and $\check{O}_s$, which are formed following Bartlett et al. [BLJ84]. Therefore, the derivation of the apparent depth in Section 2.3.3 needs to be adapted.

Similarly to Section 2.3.3, the ray w in water is the central ray of a cone representing a bundle of rays, which are emitted from a real object point O, as can be seen in Figure 4.5. A limit value observation is performed with the aid of a second ray in the cone's meridional Ψ_m and sagittal plane Ψ_s, respectively (the cone's size is exaggerated for visualization). The limit value observation is illustrated in Figure 4.5 $\boldsymbol{b} - \boldsymbol{d}$. Due to the parallelism of the two refractive interfaces, similar prerequisites as before exist and the two possible locations of the virtual images at the *VOP*s $\check{O}_m$ and $\check{O}_s$ in Figure 4.5 $\boldsymbol{d}$ differ only slightly from the ones in Figure 2.8. Due to occlusions, the two rays in the sagittal plane can only be seen in the top views of Figure 4.5 $\boldsymbol{b} - \boldsymbol{d}$ and the two rays in the meridional plane in the side views. As before, the limit value observation is done for an infinitesimal radius of the cone. In the sagittal case, the planes of refraction of both ray paths obviously differ. Due to the parallelism of the interfaces and their shared interface normal, both still have to intersect in the object axis s_1. Therefore, the sagittal *VOP* is located on the object axis s_1 at point $\check{O}_s$ and the limit value observation is just performed formally, as in the case with a single flat refractive interface. In the meridional case, the whole derivation takes place in the plane of refraction Π of the ray w in water, which is identical to the meridional plane Ψ_m. This is not altered by additional parallel interfaces and the virtual extensions of both rays still intersect in the meridional *VOP* $\check{O}_m$, which is located higher than the real object point O and closer to the observer. As before, both locations are situated on the backwards extended ray a.

The direct finding is that the essential derivation processes of both *VOP* locations are

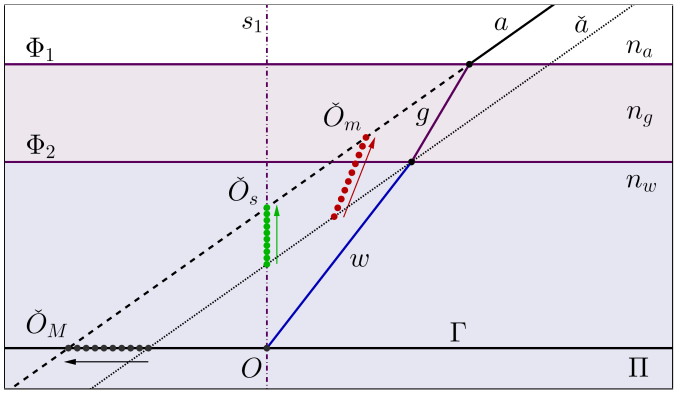

Figure 4.6: Linear shift to the new *VOP* locations due to an additional refractive interface.

still valid, since the planes of refraction are not influenced by an additional refractive layer. However, the final locations are only similar and their mathematical computation differs. The same applies for several additional, parallel interfaces.

Linear Shift of Locations

Although the basic results of the adapted derivation of the two physically founded locations are similar, there is still a change compared to their final locations in the case of a single refractive interface. The derivation process of $\check{O}_M$ is not altered and still contains the intersection of the ray a with a plane Γ passing through the real object point O, parallel to the interface Φ. However, its final location is changed as well. These changes will be analyzed briefly in the following. In contrast to the previous case of a water-air transition of the rays, the additional glass layer results in a water-glass-air transition. As is known from physics, an additional layer between two media does not alter the final direction of the refracted ray. Since the current viewing direction starts at the real object point O, this is the ray a in air. It can be constructed geometrically by a shift from the ray $\check{a}$ in the case of a single interface to the ray a in the case of two interfaces. Therefore, the direction of the ray g in glass does not even need to be known. Note that the ray $\check{a}$ is illustrated in a virtual version. As can be seen in Figure 4.6, the change of all the locations of the *VOP*s simply amounts to a linear shift, respectively. All the segments of the optical path (w, g and a) and all the *VOP*s ($\check{O}_s$, $\check{O}_m$ and $\check{O}_M$) lie in a single plane of refraction Π, which is formed by the ray w in water and the object axis s_1.

Traces of Virtual Object Points

Figure 4.7 illustrates the traces of the *VOP* locations in the case of two flat refractive interfaces. It was generated in the same way as Figure 4.3. As can be seen, the shapes of the basic 3D structures, which are resulting from tracing the respective locations of the *VOP*s, essentially stay the same. However, there are also some differences.

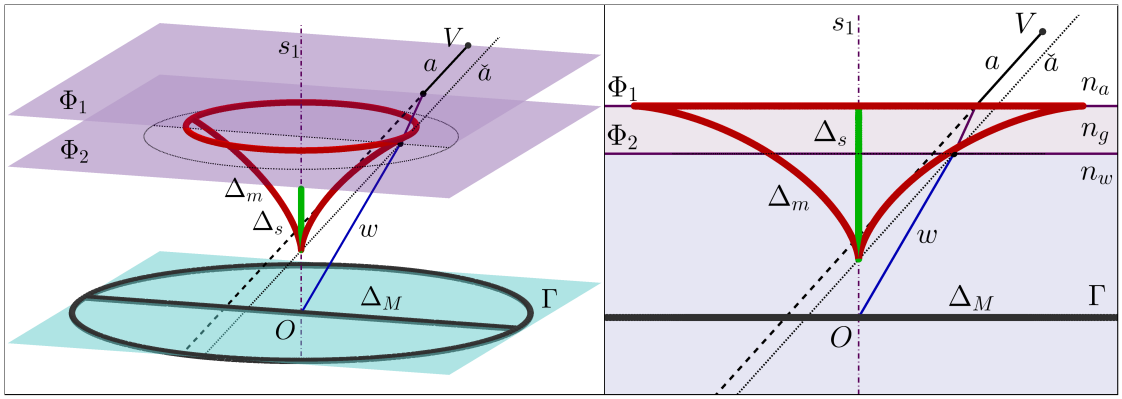

Figure 4.7: Traces of the *VOP* locations in the case of two refractive interfaces. Left: 2D front view. Right: Perspective 3D view.

- The circular area Δ_M on the plane Γ is not concerned, since its radius already tends towards infinity in the case of a single refractive interface.

- The extent of Δ_m is altered along the object axis s_1. The upper end is extended up to the interface Φ_1 and the apex at the bottom is shifted upwards. Due to the additional medium, the bottom end can not be computed by Equation 2.25 any more.

- The extent of Δ_s is altered along the object axis s_1 by the same amount as Δ_m is.

4.2.3 Selection

As in the case of a single flat refractive interface, the determined traces of the *VOP* locations in the case of two flat refractive interfaces make clear that *VOP*s can be assigned ambiguously but regulated to a single real object point. The analysis of the resulting 3D structures shows that the earlier proposed arguments against Δ_M and Δ_m remain. Summarized, these are the infinite extent of Δ_M and the ambiguous intersections of the ray a in question with Δ_m. The most convenient case is still indisputably Δ_s. By being a single straight line, it has the least spacial extent, the simplest 3D structure and is the most suitable search space for the determination of the true optical path between a given real object point O and a viewpoint V. In the case of two interfaces, the extent of this line along the object axis s_1 is influenced by the shift of the *VOP* $\check{O}_s$. The direct consequence is that the relation between the virtual and the real object point can not be expressed by Equation 2.26 any more. It gets invalid and needs to be adapted. The derivation of the new relation will be performed in the following section. The chosen *VOP* location $\check{O}_s$ will be referred to as *VOP* $\check{O}$ in the remainder of this thesis.

4.3 Model for Two Flat Refractive Interfaces

The proposed *VOP* model represents the relation between the virtual and the real object points in the case of two parallel flat refractive interfaces and is based on Equation 2.26, which represents this relation in the case of a single flat refractive interface. However, as mentioned in the previous section, this equation is insufficient and needs to be adapted. Compared to the previous setup, which can be seen on the left side of Figure 4.8, the current setup on the right side shows some similarities. The prerequisites resemble each other to a certain degree and hence, there is the potential for the adaption of Equation 2.26. Let us consider the optical path on the right side of Figure 4.8 that contains the ray a in air, the ray g in glass and the ray w in water. From a formal point of view, the viewing direction is from air to water and the angle α is the angle of incidence. Therefore, the ray a gets refracted at the interfaces Φ_1 and Φ_2. This results in the ray w in water that is passing through the real object point O. Let us assume that the glass layer and hence the refractive interface Φ_1 does not exist. This would result in the ray $\check{w}$. Since it is already known from physics that the final direction of the refracted ray is not influenced by an additional layer, this ray can also be constructed by shifting the ray w in parallel to pass through the auxiliary point A_3. The points A_1, A_2, A_3 and $\check{O}$ on the right side of Figure 4.8 represent the exact same setup as is illustrated on the left side. This shows that the relation between the virtual and the real object point in the case of two parallel flat refractive interfaces can be reduced to this relation in case of a single flat refractive interface. Therefore, the formulation by Equation 2.26 is able to serve as a starting point. The necessary adaptions for the new relation will be formulated in the following. With the aid of the selected *VOP* location, the derivation of the new relation can be realized by trigonometry in right-angled triangles, similarly as in Section 2.3.3. The proposed *VOP* model is partially based on Dolereit and von Lukas [DL16].

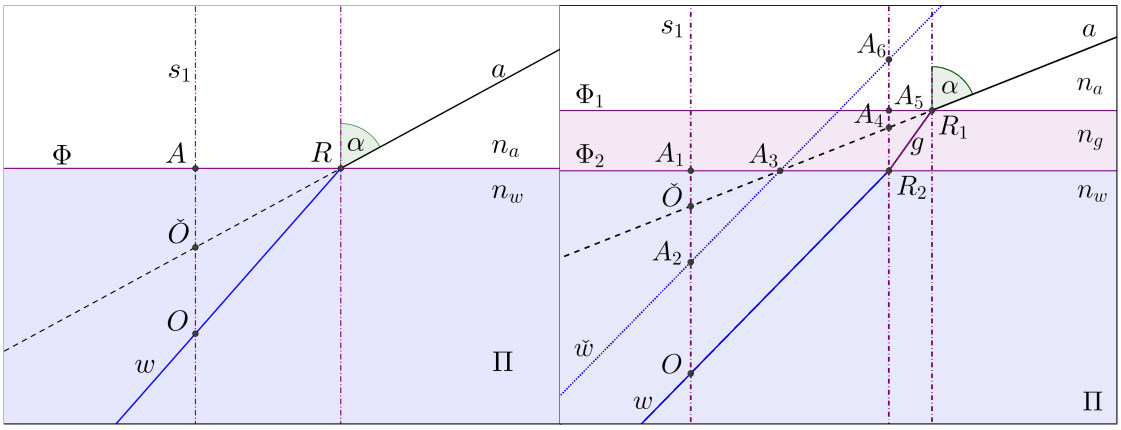

Figure 4.8: Relation between the virtual and the real object points. Left: In the case of a single flat refractive interface. Right: In the case of two parallel flat refractive interfaces.

Derivations. The two right-angled triangles formed by the auxiliary points A_1, A_2, A_3 and the *VOP* $\check{O}$ form the starting point. By utilization of Equation 2.26, the distance $d(A_1, A_2)$ can be related to the distance to the *VOP* $d(A_1, \check{O})$ by

$$d(A_1, A_2) = \lambda_1(\alpha, n_a, n_w) \cdot d(A_1, \check{O}), \tag{4.1}$$

with λ_1 being a scalar, which is dependent on the angle of incidence α and the refractive indices n_a and n_w:

$$\lambda_1(\alpha, n_a, n_w) = \frac{\sqrt{\left(\frac{n_w}{n_a}\right)^2 - \sin^2\alpha}}{\cos\alpha}. \tag{4.2}$$

As can be seen, the auxiliary point A_2 does not coincide with the real object point O. This results in an offset, which amounts to the distance $d(A_2, O)$. Since the points A_2, O, R_2 and A_6 form a parallelogram, the distance $d(A_2, O)$ is equal to the distance $d(R_2, A_6)$. Note that the two right-angled triangles formed by the points A_3, A_4, A_6 and R_2 and the previously utilized triangles, which are formed by the points A_1, A_2, A_3 and $\check{O}$, are geometrically similar. Therefore, Equation 2.26 can be utilized again to relate the offset $d(A_2, O)$ to the distance $d(R_2, A_4)$:

$$d(A_2, O) = d(R_2, A_6) = \lambda_1(\alpha, n_a, n_w) \cdot d(R_2, A_4). \tag{4.3}$$

The necessary distance $d(R_2, A_4)$ is dependent on the properties of the glass layer. It is computed by:

$$d(R_2, A_4) = t_g - d(A_4, A_5), \tag{4.4}$$

with $t_g = d(A_5, R_2)$ being the known glass layer thickness. Let us take a look at the two right-angled triangles formed by the points A_4, A_5, R_1 and R_2. Due to the refractive index of glass, these are not geometrically similar to the previously used triangles any more. However, it is still valid to compute the distance $d(A_4, A_5)$ by utilization of Equation 2.26:

$$d(A_4, A_5) = \lambda_2(\alpha, n_a, n_g) \cdot t_g, \tag{4.5}$$

with λ_2 being a scalar, which is dependent on the angle of incidence α and the refractive indices n_a and n_g:

$$\lambda_2(\alpha, n_a, n_g) = \frac{\cos\alpha}{\sqrt{\left(\frac{n_g}{n_a}\right)^2 - \sin^2\alpha}}. \tag{4.6}$$

Note that in contrast to Equations 4.1 and 4.3 the rearrangement differs. However, all the necessary relations can be derived straightforward from the same basic principle.

Real-from-Virtual. The application of Equations 4.1 to 4.6 results in the final formulation of the model for the relation between the virtual and the real object point. The distance to the real object point is represented by $x = d(A_1, O)$ and the distance to the *VOP* by $\check{x} = d(A_1, \check{O})$. The distance to the real object point can be computed by

$$\begin{aligned} x &= d(A_1, A_2) + d(A_2, O) \\ &= \lambda_1(\alpha, n_a, n_w) \cdot \check{x} + \lambda_1(\alpha, n_a, n_w) \cdot (t_g - \lambda_2(\alpha, n_a, n_g) \cdot t_g) \\ &= \lambda_1(\alpha, n_a, n_w) \cdot (\check{x} + t_g \cdot (1 - \lambda_2(\alpha, n_a, n_g))) \\ &= \lambda_1(\alpha, n_a, n_w) \cdot (\check{x} + t_g \cdot (1 - \lambda_2(\alpha, n_a, n_g))) \\ &= \frac{\sqrt{\left(\frac{n_w}{n_a}\right)^2 - \sin^2\alpha}}{\cos\alpha} \cdot \left(\check{x} + t_g \cdot \left(1 - \frac{\cos\alpha}{\sqrt{\left(\frac{n_g}{n_a}\right)^2 - \sin^2\alpha}}\right)\right), \end{aligned} \tag{4.7}$$

if the distance to the *VOP* $\check{x}$, the glass layer thickness t_g, the angle of incidence α and the refractive indices n_a, n_g and n_w are known.

Virtual-from-Real. Vice versa, the distance to the *VOP* can be computed by

$$\begin{aligned} \check{x} &= \frac{x}{\lambda_1(\alpha, n_a, n_w)} - t_g \cdot (1 - \lambda_2(\alpha, n_a, n_g)) \\ \check{x} &= \frac{x \cdot \cos\alpha}{\sqrt{\left(\frac{n_w}{n_a}\right)^2 - \sin^2\alpha}} - t_g \cdot \left(1 - \frac{\cos\alpha}{\sqrt{\left(\frac{n_g}{n_a}\right)^2 - \sin^2\alpha}}\right), \end{aligned} \tag{4.8}$$

if distance to the real point x, the glass layer thickness t_g, the angle of incidence α and the refractive indices n_a, n_g and n_w are known.

Particularities. Note that in both cases similar conversions as in Section 2.3.3 are possible. For example, with the substitution of Equations 2.27 and 2.23, it is possible to eliminate the need to compute the trigonometric functions in these two equations explicitly. The represented relations can be readily introduced into the underwater image formation models with explicit consideration of refractive effects. However, the proposed model has some limitations. The point R_2 represents the closest location the real object point O can be at. It corresponds to a distance equal to zero. Since the real object can not penetrate the glass layer, it coincides with the physical limitation. However, a *VOP* $\check{O}$ can be situated inside the glass layer. Therefore, a point symmetry in A_3 is exploited for Equations 4.7 and 4.8 to be valid in this case. This means that the distances to *VOP*s inside the glass layer, which are measured with respect to the auxiliary point A_1 on the interface Φ_2, need to get a negative algebraic sign before being inserted into these equations.

4.4 Summary

Since the focus of this thesis is on calibration of the refractive parameters of a *SFRS*, basic operations like the computationally expensive refractive projections for coordinate transformations between 3D and 2D spaces are unavoidable. In this chapter, a *VOP* model, which has the capabilities to affect these refractive projections in a beneficial way, has been motivated and the previously presented findings on *VOP*s have been analyzed. The result of the analysis of the impacts of two flat refractive interfaces on the three previously proposed *VOP* locations is that they are still generally valid. This is especially relevant for the two locations that are physically founded by the concept of the apparent depth. In each case, the only difference is a linear shift due to the additional glass layer. Since all three possibilities are valid locations, a selection based on convenience needed to be made. With an unknown optical path between a real object point and a viewpoint, one can assign *VOP*s ambiguously but regulated to this single object point. This regulation is based on the derivation processes of the respective locations. Therefore, the 3D structures that are formed by the computed traces of the *VOP* locations for a significant amount of all possible directions of the rays emitted from a real object point have been analyzed. These 3D structures represent a search space for the determination of the true optical path. The criteria for the selection of the most suitable location are the properties of these 3D structures, which finally enabled an obvious selection of $\check{O}_s$. Its trace has the least spacial extent and, by being a straight line, it has the simplest 3D structure. The *VOP* $\check{O}_s$ lies exactly on the object axis. This is the *VOP* location that will be utilized in the remainder of this thesis and it will be referred to by $\check{O}$ from now. The model that has been formulated is based on simple trigonometric rules and describes how to non-ambiguously relate *VOP*s to real object points and vice versa in the case of two flat refractive interfaces. As will be seen in the following chapters, this model can be readily integrated into the underwater image formation models and thereby provides an ideal basis to meet the challenges that arise during the extensions of the concepts of stereo 3D reconstruction in air for underwater environments. The previously utilized viewpoint will be replaced in the following by the center of projection of the camera. Thereby, collinearity between the *VOP*, the center of projection and an image point occurs and the requirements that make an approximation by the pinhole camera model valid are met.

5 Calibration of Shared Flat Refractive Systems

In this chapter, the proposed approaches for the determination of the refractive parameters of a *SFRS* will be presented. These represent one of the major contributions of this thesis. The requested refractive parameters are the orientation of the system axis and the air layer thickness. The obvious advantage of a shared flat refractive interface is that the refractive parameters have to be estimated just with respect to the master camera. Although being more specialized than single camera systems, the model for a *SFRS* can be designated as a representative of a common viewing condition. There are multiple practical configurations of imaging setups that can be represented by this model. A first exemplary configuration consists of two cameras in front of one side of an aquarium. A second configuration consists of two cameras in a single underwater housing.
The refractive calibration of the imaging system is an important first step of underwater stereo 3D reconstruction. A justification of the requirements for the approaches to be developed, of the prerequisites for the approaches to work and its resulting restrictions will be done in this chapter. Unavoidable for calibration are basic operations like the computationally expensive refractive projections between 3D and 2D spaces. It will be shown how these basic operations can be affected by the *VOP* model from Section 4.3 in a beneficial way. Therefore, this model needs to be integrated into the utilized models for underwater image formation. The integration within monocular and binocular vision will enable four alternative approaches for the coordinate transformation between 3D and 2D spaces. All of them will be used for the definition and computation of cost functions. These cost functions reflect the quality of the determined parameters and are a vital part of calibration approaches. Another vital part of calibration approaches are the definitions of optimization problems, which regulate the interaction between parameter determination and cost functions. Within the proposed optimization approaches, it will be shown how the refractive parameters can be determined consecutively and simultaneously. Common cost functions, such as the reprojection error or 3D scene errors, and one that is a novel development will be utilized for the simultaneous optimization of the refractive parameters. In contrast, pattern-based invariants will be derived from the *VOP* model for the novel development of pattern-based cost functions. Their application will allow the determination of the system axis, independently of the air layer thickness. Thereby, the consecutive optimization of the refractive parameters will be possible. As will be seen, the proposed calibration approaches will show a progressive independence from calibration objects and will allow the determination of the air layer thickness by linear optimization, which are major contributions of this thesis. This chapter is partially based on Dolereit et al. [DK14; Dol15; DLK15a; DLK15b; DL16].

5.1 Requirements, Prerequisites and Restrictions

Since the focus of this thesis is on the calibration of the refractive parameters of a *SFRS*, the requirement for the approaches to be developed is to be able to handle such a system. The approaches should serve the calibration of the refractive parameters defined in Section 2.5 and the developments should contain a progressive independence from calibration objects. The following prerequisites have to be satisfied for the current implementation:

- The imaging setup can be characterized by the *SFRS*.
- There are two interfaces and the thickness of the glass layer is known.
- The refractive indices of the participating media (air, glass and water) are known.
- The stereo camera with constant intrinsic and extrinsic parameters is calibrated.

These prerequisites simultaneously impose restrictions on the approach, which will be justified in the following. The major restriction is the *SFRS* itself. As mentioned before, having a shared interface is not an unusual setup. Its biggest advantage are some major computational simplifications that will be presented in the following chapter. A second restriction is that only the air layer thickness will be estimated and the glass layer thickness has to be known. This is not that severe, since most of the time it is possible to measure the thickness of, for example, a glass or acrylic plate of an underwater housing or of an aquarium directly. It is even possible that it is known from the manufacturer. To assume a known glass layer thickness is a common practice and it is also recommended to prefer measurement over estimation (See Section 3.3.3). The same common practice is the assumption of known refractive indices of all participating media (See Section 3.3). In the case of a water-glass-air transition, there is the refractive index of air which can be approximated reliably by 1. The refractive index of water can be approximated by 1.33, since it is said to change only about 2% in the whole relevant range of water bodies [Mob94]. The refractive index of glass is known from the material properties, which are usually available from the manufacturer. The last restriction is the calibration of the stereo camera in advance. A major assumption is that camera lens distortion and refractive distortion do not influence each other. Lens distortion is independent of the 3D scene, whereas refractive distortion is highly dependent on it. Since the literature on this subject does not state otherwise, it is assumed that both are separable. This enables the calibration of the stereo parameters individually. Camera lens distortion can then be eliminated from underwater images as a pre-process.
Compared to approaches from the literature, the developed ones will not be restricted as follows:

- The image plane does not need to be perpendicular to the interface normal.
- There is no need for known coordinates of feature points on a calibration object.
- The number of interfaces and layers is extendable as long as the additional layer thicknesses and the refractive indices are known.

5.2 Virtual Object Point Model and Monocular Vision

An advantageous property of image formation in air is the approximability by the pinhole camera model. Unfortunately, due to the refraction of light rays, the pinhole camera model is invalid under water. A valid starting point is the model for a *FRS*. Since it contains only one camera, this is a generalization of the model for the *SFRS*. Note that the previously used viewpoint of the observer is replaced by the center of projection C of the camera. Assuming that all the refractive parameters are already known, the underwater image formation process is still computationally expensive. The integration of the *VOP* model from Section 4.3 is illustrated in Figure 5.1. The advantage of the combined model is that the *VOP* $\check{O}$ is situated collinear to the center of projection C of the camera and the image point I. In the following, it will be shown how this can be utilized for transformations between 3D and 2D spaces.

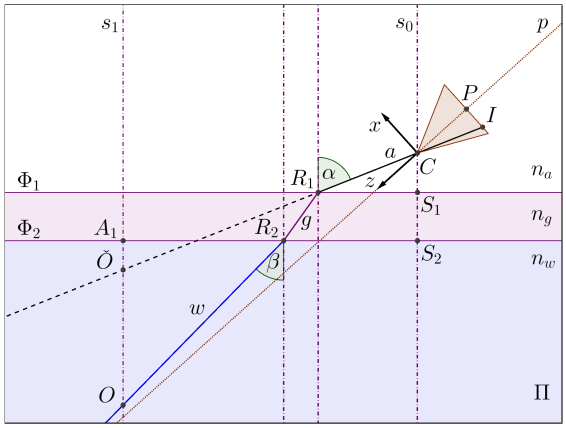

Figure 5.1: Combination of a *FRS* and the *VOP* model.

Labeling. A detailed description of the labeling in Figure 5.1 is presented in Table 5.1. This labeling can be adapted for all following figures in this chapter. Note that the illustration is a top view of the plane of refraction Π that contains the full ray path between an image point I and a real object point O. In contrast to the depicted case, in which the x-axis and the z-axis of the camera coordinate system coincide with the plane of refraction Π and the omitted y-axis is orthogonal, the principal axis p does not necessarily lie this plane. Note that the ray path would not change if the camera rotates around its center of projection C, or even if it additionally translates along the ray a. Only its endpoint, which is captured on the image plane at image point I, would be affected. The ray path contains the ray a in air, the ray g in glass and the ray w in water. As before, the ray a in air is not distinguished by the labeling from its virtual part in water (dashed). The two interface normals passing through the points of refraction R_1 and R_2 are only necessary for the illustration of the angle of incidence α and the angle of refraction β. Their labeling is omitted. Since the computation of the angle of refraction

of the ray g in glass is not necessary for the approaches to be developed, its illustration is omitted as well.

label	description
Π	plane of refraction
Φ_1	flat refractive interface between air and glass
Φ_2	flat refractive interface between glass and water
n_a	refractive index of air
n_g	refractive index of glass
n_w	refractive index of water
s_0	system axis
s_1	object axis
S_1	intersection point of s_0 and Φ_1
S_2	intersection point of s_0 and Φ_2
O	real object point
$\check{O}$	virtual object point
A_1	auxiliary point on Φ_2
R_1, R_2	points of refraction
C	center of projection of the camera and origin of the camera coordinate system
x	x-axis of the camera coordinate system
z	z-axis of the camera coordinate system
y	y-axis is not illustrated as is perpendicular to the plane of refraction Π
p	principal axis of the camera
P	principal point of the camera
I	image point
a	ray in air
g	ray in glass
w	ray in water
α	angle of incidence
β	angle of refraction

Table 5.1: Labeling of elements in illustrations.

In the figures, the labeling of vectors and distances is usually omitted for the sake of clarity. However, the labeling of vectors in the text follows the scheme of matching characters, which is presented exemplary at the top section of Table 5.2. Note that there is no difference between the labeling of a position vector and a direction vector. As can be seen at the bottom section of Table 5.2, some recurring distances have special labels. In the texts, the labeling of distances will usually be accompanied by the notation for the Euclidean distance $d(.., ..)$, which was introduced in Section 2.1.

label	description
$\boldsymbol{a}$	direction vector of the ray a
$\bar{\boldsymbol{a}}$	normalized direction vector of the ray a
$\boldsymbol{o}$	position vector of the point O
t_a	air layer thickness, which is the distance $d(C, S_1)$
t_g	glass layer thickness, which is the distance $d(S_1, S_2)$
$\check{x}$	distance between a virtual point and the closest interface along the interface normal, e.g. $d(A_1, \check{O})$
x	distance between a real point and the closest interface along the interface normal, e.g. $d(A_1, O)$
f	focal length, which is the distance $d(C, P)$

Table 5.2: Labeling of vectors and distances, which is generally omitted in illustrations.

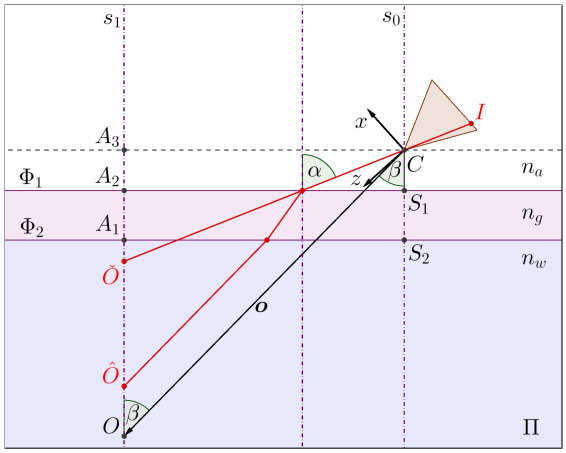

Figure 5.2: Refractive forward projection: Intermediate step of the iterative processing.

5.2.1 Refractive Forward Projection

Let us assume that a real object point O is known in the camera coordinate system illustrated in Figure 5.2. As shown in Section 3.3.1, the refractive forward projection of this point onto the image plane is computationally expensive. In contrast to perspective projection, the starting direction of the ray that is emitted from the point O is unknown. For the depicted case with two flat refractive interfaces, a solution from the literature is to solve higher-order polynomials up to degree twelve. The proposed approach in this thesis is an iterative computation of the distance $d(A_1, \check{O})$ by a simple bisection strategy. The advantage is that the associated *VOP* $\check{O}$ can be mapped to the image plane by perspective projection.

Requirements. The required quantities are the refractive indices n_a, n_w and n_g, the glass layer thickness $t_g = d(S_1, S_2) = d(A_1, A_2)$, the air layer thickness $t_a = d(C, S_1) = d(A_2, A_3)$, the position vector $\boldsymbol{o}$ of the real object point O and the normalized direction vector $\bar{\boldsymbol{s}}$, which is identical for the system axis s_0 and the object axis s_1.

Constraints and Definitions. In Figure 5.2, the *VOP* candidate $\check{O}$ and its related real object point candidate $\hat{O}$ of an intermediate step of the iterative computation are shown. As can be seen, the real object point candidate $\hat{O}$ and the true real object point O do not coincide. Ultimately, the constraint $d(O, \hat{O}) = 0$ needs to be fulfilled. Since every real object point has to be located in water, the distance $x = d(A_1, O)$ needs to be equal to or greater than zero. With $\hat{x} = d(A_1, \hat{O})$, the cost function that has to be minimized during the iterative processing is defined to be:

$$E = x - \hat{x}. \tag{5.1}$$

The distance $\hat{x}$ can be computed with the aid of the *VOP* model from Section 4.3. From this model, it is known that the true *VOP* $\check{O}_0$ has to be located between the

points A_2 and O on the object axis s_1. Let us define that the auxiliary point A_1 is the origin, the direction to O is positive and the direction to A_2 is negative, as shown in the depicted coordinate system on the left side of Figure 5.3. These definitions ensure that the algebraic sign of the value of E is negative in section I and positive in sections II and III, with a point of inflection at $\check{O}_0$. This point of inflection is the *VOP* in question. The proposed bisection strategy is used to optimize the distance $\check{x} = d(A_1, \check{O})$ to get the *VOP* $\check{O}$ closest to $\check{O}_0$. It is composed of an initialization phase and an iteration phase.

Initialization. The initialization phase comprises the following definitions and initial computations:

- Compute the distance $d(A_3, O)$ by orthogonal projection of vector $\boldsymbol{o}$ onto the normalized vector $\bar{\boldsymbol{s}}$, with the aid of Equation 2.7.
- Compute the distance x by: $x = d(A_1, O) = d(A_3, O) - t_g - t_a$.
- Compute the angle β between the vectors $\bar{\boldsymbol{s}}$ and $\boldsymbol{o}$ and with that, compute the distance $d(A_3, C)$ by: $d(A_3, C) = \tan\beta \cdot d(A_3, O)$.
- Define a maximal deviation value ε for the termination of the iteration.
- Initialize the minimum value by $min = -d(A_1, A_2)$, the maximum value by $max = d(A_1, O)$ and the middle value for the bisection by $mid = \frac{min+max}{2}$.

Iteration. The iteration phase comprises the following steps for the optimization of $\check{x}$:

1) Set the current distance $\check{x}$ to: $\check{x} = d(A_1, \check{O}) = mid$.

2) Compute the angle of incidence α by:

$$\alpha = \arctan\left(\frac{d(A_3, C)}{t_a + t_g + \check{x}}\right).$$

3) Compute the distance $\hat{x} = d(A_1, \hat{O})$ by application of Equation 4.7 from $\check{x}$, α, t_g, n_a, n_w, n_g.

4) Compute the function value E of Equation 5.1: $E = x - \hat{x}$.

5) If $E < 0$ then $max = mid$, if $E > 0$ then $min = mid$.

6) Compute the new middle value by bisection: $mid = \frac{min+max}{2}$.

7) Terminate the optimization of $\check{x}$ if $|E| \leq \varepsilon$.

8) Repeat from **1)**.

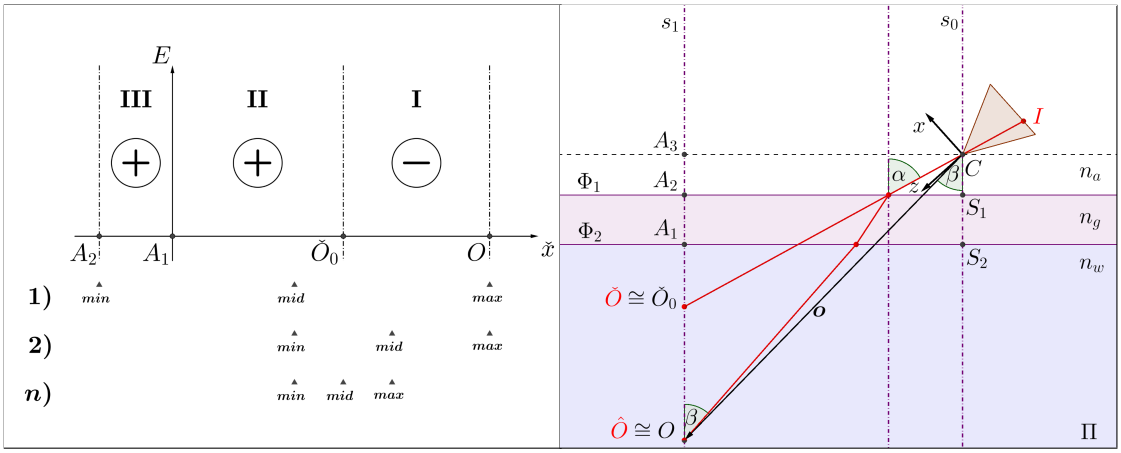

Figure 5.3: Refractive forward projection. Left: Final step of the iterative processing. Right: Schematic illustration of the bisection strategy.

This iterative processing is illustrated exemplary in Figure 5.3. The row 1) marks the initialization phase. The row n) marks the termination of the optimization of distance $\check{x}$, since the point of inflection at $\check{O}_0$ and hence the *VOP* in question is found. The result is the *VOP* $\check{O}$ that is closest to $\check{O}_0$ in terms of ε (See right side of Figure 5.3). Since the normalized direction vector $\bar{\boldsymbol{s}}$ of the object axis s_1 is known, the position vector $\boldsymbol{o_0}$ of the *VOP* $\check{O}$ can be determined straightforward from the final distance $\check{x}$ by: $\boldsymbol{o_0} = \boldsymbol{o} - (x - \check{x}) \cdot \bar{\boldsymbol{s}}$. Afterward, the image point I can be determined by perspective projection of the *VOP* $\check{O}$ with the aid of the known intrinsic camera parameters as described in Section 2.2.1.

5.2.2 Refractive Back-Projection

The goal of refractive back-projection is the computation of the ray w in water from a given image point I, as illustrated on the left side of Figure 5.4. The proposed approach is an alternative to the explicit 3D ray tracing commonly used in the literature (See Section 3.3.1). In the depicted situation, ray tracing would comprise the computation of the direction vectors of the rays a, g and w, as well as the computation of the points of refraction R_1 and R_2. On the contrary, the proposed approach comprises only the computation of two ray directions and one point, as will be described in the following.

Requirements. The required quantities are the refractive indices n_a, n_w, n_g, the glass layer thickness $t_g = d(S_1, S_2)$, the air layer thickness $t_a = d(C, S_1)$ and the system axis s_0 with a normalized direction vector $\bar{\boldsymbol{s}}$.

Definitions. The ray w in question can be expressed by:

$$w : \boldsymbol{o} = \boldsymbol{c} + \lambda \cdot \boldsymbol{w}. \tag{5.2}$$

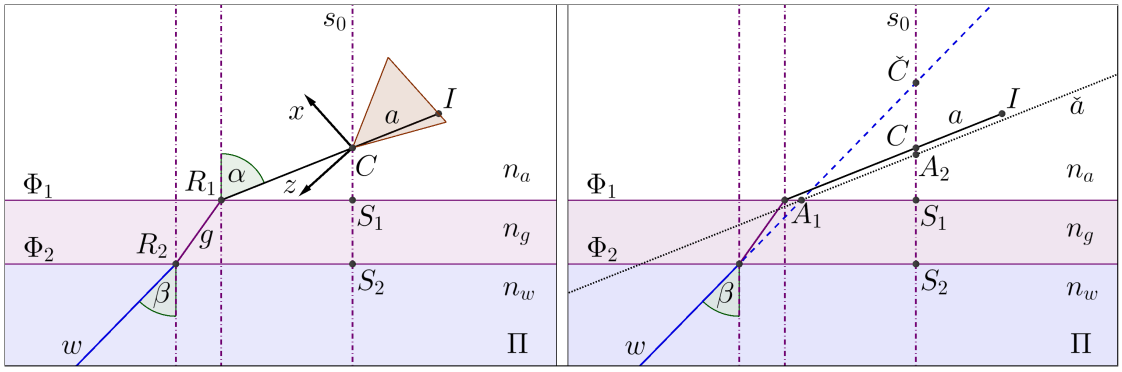

Figure 5.4: Refractive back-projection. Left: Ray path (a, g and w) from full ray tracing. Right: Avoid full ray tracing by application of the *VOP* model to compute the ray w expressed by Equation 5.2.

This comprises a position vector $\boldsymbol{c}$ and a direction vector $\boldsymbol{w}$. With an according scalar λ, every real object point O on this ray can be expressed by its position vector $\boldsymbol{o}$. The position vector $\boldsymbol{c}$ belongs to the point of intersection of the system axis s_0 and the virtual part of ray w. This point of intersection is a *VOP*. On the right side of Figure 5.4 it is labeled by $\check{C}$. Note that the illustration of the camera and its coordinate system is omitted for the sake of clarity.

Computation of the Position Vector. As can be seen on the right side of Figure 5.4, the setup resembles the one of Section 4.3, which has been used to formulate the *VOP* model for two flat refractive interfaces. The only difference is that the direction of consideration of the ray path is inverted. Here, the ray w in water is considered as the incident ray and its virtual part (dashed) intersects the system axis s_0 at the *VOP* $\check{C}$. This *VOP* $\check{C}$ belongs to the real object point C and both can be related by the previously proposed *VOP* model.
The first preparatory step is the computation of the normalized direction vector $\bar{\boldsymbol{a}}$ of the ray a in air, with respect to the camera coordinate system. With the aid of Equation 2.17, it can be computed from image point I and the known intrinsic camera parameters. Subsequently, the angle of incidence α, which is the angle between the two normalized direction vectors $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{s}}$, can be computed straightforward. Under these conditions, the position vector $\boldsymbol{c}$ can be computed as follows:

- Compute the distance $\check{x} = d(S_1, \check{C})$ from distance $x = t_a$, with t_g, α, n_a, n_g and n_w by application of Equation 4.8 under consideration that the direction of the ray path is inverted: replace angle α by angle β and switch n_a and n_w.

- Compute the angle of refraction β according to Snell's law from the angle of incidence α by Equation 2.20.

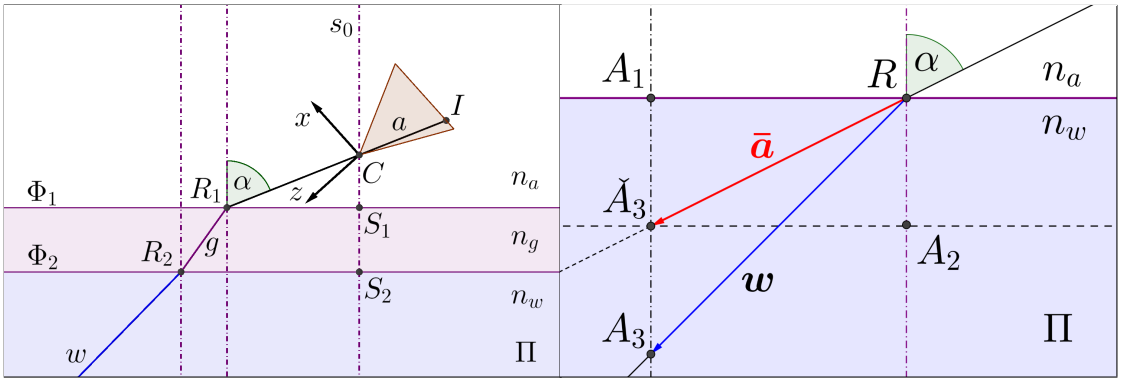

Figure 5.5: Refractive back-projection: Computation of the direction vector $\boldsymbol{w}$ of ray w.

- Since the normalized direction vector $\bar{\boldsymbol{s}}$ of the system axis s_0 is known, the position vector $\boldsymbol{c}$ can be determined straightforward by: $\boldsymbol{c} = \check{x} \cdot \bar{\boldsymbol{s}}$.

Computation of the Direction Vector. The computation of the direction vector $\boldsymbol{w}$ of the ray w in water can be realized by the basic *VOP* model for a single flat refractive interface from Section 2.3.3. The actual presence of two flat refractive interfaces does not influence this computation.
The ray a, with the normalized direction vector $\bar{\boldsymbol{a}}$, and the ray w, with direction vector $\boldsymbol{w}$, share the same plane of refraction Π. Only the known, normalized direction vector $\bar{\boldsymbol{a}}$ is necessary for the computation of $\boldsymbol{w}$. Based on the basic *VOP* model for a single flat refractive interface from Section 2.3.3, the exploited relations can be constructed from the setup on the left side of Figure 5.5. This construction can be seen on the right side of Figure 5.5. Note that the glass layer can be ignored and that the ray w is represented by its direction vector $\boldsymbol{w}$. Let vector $\boldsymbol{w}$ point to an arbitrary real object point A_3, which is chosen in a way that the normalized direction vector $\bar{\boldsymbol{a}}$ points exactly to a *VOP* $\check{A}_3$. With this, all the prerequisites for the application of the *VOP* model are fulfilled. Therefore, the computation of $\boldsymbol{w}$ consists of the following steps:

- Compute the distance $\check{x} = d(A_1, \check{A}_3) = d(R, A_2)$ by orthogonal projection of the normalized vector $\bar{\boldsymbol{a}}$ onto the normalized vector $\bar{\boldsymbol{s}}$, with the aid of Equation 2.7.

- Compute the distance $x = d(A_1, A_3)$ from the known angle of incidence α, distance $\check{x}$ and the refractive indices n_a and n_w with the aid of an appropriately rearranged Equation 2.26.

- Since the normalized direction vector $\bar{\boldsymbol{s}}$ of the system axis s_0 is known, the direction vector $\boldsymbol{w}$ can be determined straightforward by: $\boldsymbol{w} = \bar{\boldsymbol{a}} + (x - \check{x})\bar{\boldsymbol{s}}$

5.2.3 Virtual Camera Model

In the context of the computation of the reprojection error, the utilization of a virtual camera represents an alternative to refractive forward projection. This means that a known 3D point needs to be projected to 2D image space, while its measured 2D image coordinates are known. By comparison of the projected and the measured image point, the accuracy of the recovered 3D point can be evaluated. The proposed virtual camera model is based on Dolereit and Kuijper [DK14].
The starting point of the computation of the virtual camera is a real image point I, which belongs to an arbitrary real object point O on the ray w. As can be seen on the left side of Figure 5.6, the real object point O, the center of projection $\check{C}$ of the virtual camera and the virtual image point $\check{I}$ are collinear. Hence, refraction is avoided and conventional forward projection based on the pinhole camera model becomes an option. The virtual image point $\check{I}$ can be determined by perspective projection of the real object point O with the aid of a set of virtual camera parameters. These have to be computed per image point I, since they differ as the angle of incidence α changes. Note that a set of parameters of a virtual camera, such as the one on the left side of Figure 5.6, is valid for all the possible real object points on the ray w in water.
The proposed approach differs from the approaches proposed in the literature. The first difference is that every resulting virtual camera has an image sensor that has the same size the as real camera. Furthermore, for every image point I, the virtual image point $\check{I}$ is located at the exact same image coordinates. Hence, every pair of these two images is identical for this one image point.

Requirements. The required quantities are the refractive indices n_a, n_w, n_g, the glass layer thickness $t_g = d(S_1, S_2)$, the air layer thickness $t_a = d(C, S_1)$ and the axis s_0 with a normalized direction vector $\bar{\boldsymbol{s}}$.

Preparatory Computations. This part of the computation comprises the steps of the previously proposed refractive back-projection. The normalized direction vector $\bar{\boldsymbol{a}}$ of the ray in air a, the angle of incidence α and the ray in water w, with direction vector $\boldsymbol{w}$ and position vector $\boldsymbol{c}$, are computed as described in Section 5.2.2. The position vector $\boldsymbol{c}$ belongs to the *VOP* $\check{C}$, which will be the center of projection of the virtual camera. Additionally, the computation of the angle γ between the direction vector $\bar{\boldsymbol{a}}$ and the direction vector $\boldsymbol{p}_1$ of the principal axis p and the known focal length $f = d(C, P)$ of the real camera are necessary for the following derivations.

Computation of the Intrinsic Parameters of the Virtual Camera. This part of the computations comprises the determination of the virtual focal length $\check{f} = d(\check{C}, \check{P})$ and the angle $\check{\gamma}$ between the direction vector $\boldsymbol{p}_2$ of the virtual principal axis $\check{p}$ and the direction vector $\boldsymbol{w}$ of the ray w. The necessary steps are:

- Since the real object point O can be located arbitrarily on the ray w, choose the related *VOP* $\check{O}$ to be the point of intersection between the refractive interface Φ_2 and the ray a (the coordinates do not need to be calculated explicitly).

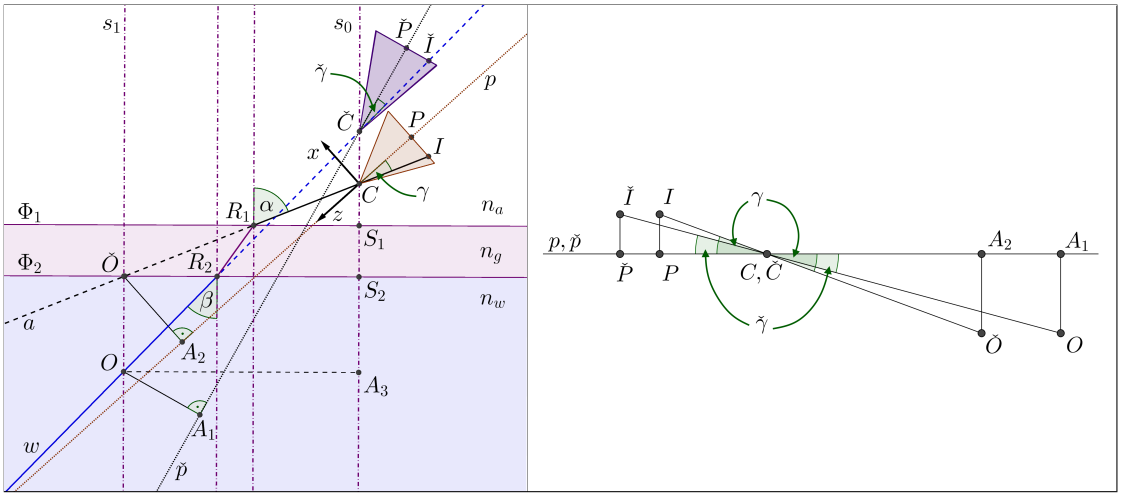

Figure 5.6: Virtual camera model: Left: Exemplary illustration of the virtual camera that is associated to the image point I. Right: Superimposition of the principal axis of the real and the virtual camera.

- Compute the distances $d(C, \check{O})$ and $d(\check{O}, S_2)$ in the right-angled triangle formed by the points $C\check{O}S_2$ by simple trigonometric rules utilizing the known angle α and distance $d(C, S_2) = t_a + t_g$.
- Compute the angle of refraction β according to Snell's law from the angle of incidence α by Equation 2.20.
- Compute the distance $d(\check{C}, O)$ in the right-angled triangle formed by the points $\check{C}A_3O$ by simple trigonometric rules utilizing angle β and distance $d(O, A_3) = d(\check{O}, S_2)$.

Let us extract some of the quantities from the left side of Figure 5.6 and illustrate them as shown on the right side. One can easily derive that if $d(P, I) = d(\check{P}, \check{I})$, then

$$d(A_2, \check{O}) = d(A_1, O). \tag{5.3}$$

With the previously computed distances $d(\check{C}, O)$ and $d(C, \check{O})$ the virtual parameters $\check{\gamma}$ and $\check{f}$ can be computed as follows:

- Compute the distance $d(A_2, \check{O})$ in the right-angled triangle formed by the points $A_2C\check{O}$ by: $d(A_2, \check{O}) = \sin\gamma \cdot d(\check{O}, C)$.
- Compute the angle $\check{\gamma}$ in the right-angled triangle formed by the points $A_1\check{C}O$ with the aid of Equation 5.3 by:

$$\check{\gamma} = \arcsin\left(\frac{d(A_1, O)}{d(O, \check{C})}\right) = \arcsin\left(\frac{\sin\gamma \cdot d(\check{O}, C)}{d(O, \check{C})}\right).$$

- Compute the distance $d(I, P)$ in the right-angled triangle formed by the points CIP by simple trigonometric rules utilizing the known angle γ and focal length f.
- Compute the virtual focal length $\check{f} = d(\check{C}, \check{P})$ in the right-angled triangle formed by the points $\check{C}\check{I}\check{P}$ by simple trigonometric rules utilizing angle $\check{\gamma}$ and distance $d(I, P)$.

With the virtual focal length $\check{f}$, it is possible to compose a virtual camera calibration matrix. As already described in Section 2.2.1, it has the form:

$$\boldsymbol{K} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The entries $f_u = m_u \check{f}$ and $f_v = m_v \check{f}$ are the virtual focal length of the virtual camera scaled to pixel units. The scaling factors m_u and m_v, as well as the coordinates of the principal point $P = (u_0, v_0)$ in the image coordinate system are adopted from the real camera.

Computation of the Orientation of the Virtual Camera. To complete the parameters of the virtual camera, this part of the computation comprises the determination of the orientation of the virtual camera, which can be represented by the axes of its coordinate system. The virtual principal axis $\check{p}$ is to a certain degree arbitrary as it is up to now only parametrized by the angle $\check{\gamma}$. In 3D space, as can be seen on the left side of Figure 5.7, this leads to a number of possible locations arranged in a circle Ω around w. The obviously choice is one of the two possibilities $\check{p}_1$ and $\check{p}_2$ in the plane of refraction Π. Both form an angle equal to $\check{\gamma}$ with the ray w. The following steps are necessary:

- Compute the direction vector $\boldsymbol{p}_{21}$ of the virtual principal axis $\check{p}_1$ or the direction vector $\boldsymbol{p}_{22}$ of the virtual principal axis $\check{p}_2$ with the aid of angle $\check{\gamma}$.
- Choose $\boldsymbol{p}_{21}$ or $\boldsymbol{p}_{22}$ to simultaneously be the direction vector of the virtual principal axis $\check{p}$ and of the $\check{z}$-axis of the coordinate system in question.
- Compute the $\check{x}$-axis utilizing the superimposition of the image plane of the virtual camera and the image plane of the real camera illustrated in the middle of Figure 5.7:
  - The direction vector of a line connecting image point I and principal point P forms an angle ε to the x-axis of the real camera. Consequently, the direction vector of a line connecting virtual image point $\check{I}$ and virtual principal point $\check{P}$ has to form the same angle to the $\check{x}$-axis in question.
  - Compute the $\check{x}$-axis in question from this angle ε and the fact that the $\check{x}$-axis needs to be orthogonal to the $\check{z}$-axis.
- Compute the $\check{y}$-axis with the cross product of direction vectors of the $\check{z}$- and the $\check{x}$-axis, since it needs to be orthogonal to both of them.

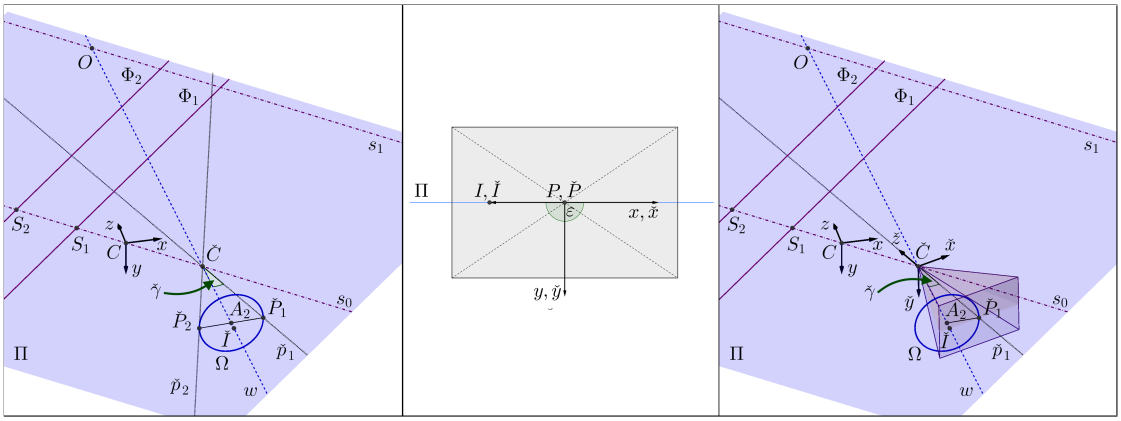

Figure 5.7: Left: Ambiguity of the virtual principal axis. Middle: Superimposition of the real and virtual image plane. Right: Resulting coordinate system of the virtual camera.

The vectors $\bar{\boldsymbol{x}}$, $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{z}}$ are the normalized direction vectors of the axes of the coordinate system of the virtual camera in the coordinate system of their computation (e.g. coordinate system of real camera), as illustrated on the right side of Figure 5.7. The transformation from the coordinate system of the real camera into the coordinate system of the virtual camera can be expressed by the rotation matrix:

$$\boldsymbol{R} = (\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{z}})^{-1}$$

and by the translation vector:

$$\boldsymbol{t} = \boldsymbol{R} \cdot (-\boldsymbol{c}),$$

with $\boldsymbol{c}$ being the position vector of the virtual center of projection $\check{C}$.

Projection Matrix of the Virtual Camera. The final projection matrix of the virtual camera for one image point I and hence for one ray w is composed of:

$$\boldsymbol{P} = \boldsymbol{K}\,[\boldsymbol{R}|\boldsymbol{t}]\,.$$

With this projection matrix, any given real object point O on the ray w in the 3D coordinate space of the real camera can be projected to the point $\check{I}$ in the 2D image space of the virtual camera. Since the coordinates of the virtual image point $\check{I}$ and its corresponding real image point I are equal by definition of the proposed virtual camera model, this direct projection is equivalent to refractive forward projection, which usually needs to be computed iteratively. However, although a direct projection is possible, a disadvantageous overhead arises out of the need for the computation of the virtual camera parameters per image point. The proposed virtual camera model shows how underwater imaging can be converted to imaging in air. In other words, an underwater image can be equivalently expressed by image formation based on the pinhole camera model. Its usefulness will be shown in context of the computation of the reprojection error in Section 5.4.4.

5.3 Virtual Object Point Model and Binocular Vision

The integration of the *VOP* model from Section 4.3 into the model for a *SFRS* is illustrated in Figure 5.8. Note that the illustration is not a top view of the plane of refraction any more. The ray paths between I and O and between I' and O are in general located in different planes of refraction. Concerning the labeling, most of the one that was proposed in Table 5.1 for the *VOP* model and monocular vision can be adopted. Due to the second camera, the major difference is that an additional right version of each quantity exists, which is labeled with an additional quotation mark. As should be clear, the object axis s_1, the auxiliary point A_1 and the real object point O are excluded, since they are the same for both cameras. Additional quantities are the rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{t}$, which represent the transformation between the two cameras. In a *SFRS*, two *VOP*s $\check{O}$ and $\check{O}'$ are associated to one real object point O. Each of them is collinear to the center of refraction and the image point of the respective camera. Both *VOP*s do not coincide in general. The only exception is the case of equal angles of incidence. In this section, it will be shown how to easily compute these *VOP*s from pairs of corresponding image points in a *SFRS*. This in turn enables the straightforward computation of the related real object point, which has already been proposed in Section 4.3. Therefore, the computation of the *VOP*s in combination with the computation of the real object point amounts to an alternative for refractive back-projection and even for the recovery of 3D coordinates. With this approach, ray tracing can be avoided completely.

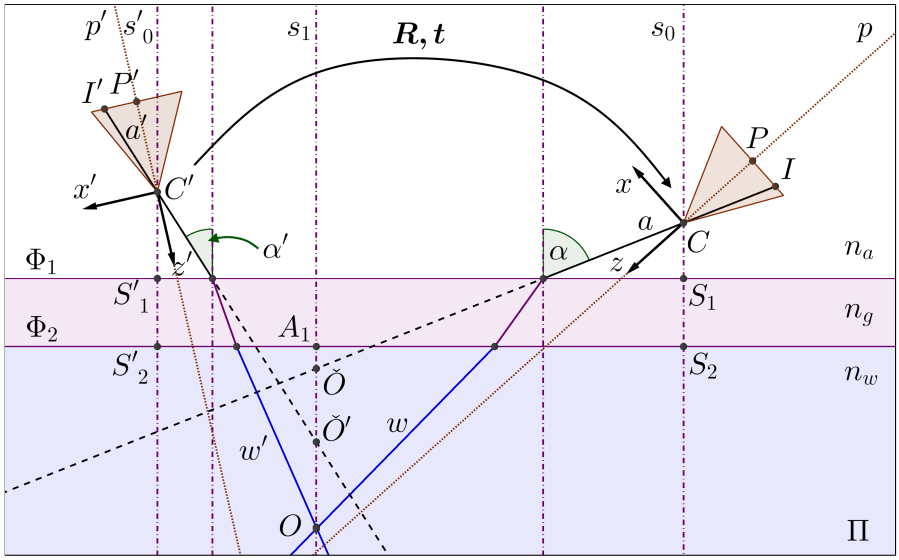

Figure 5.8: Combination of a *SFRS* and the *VOP* model.

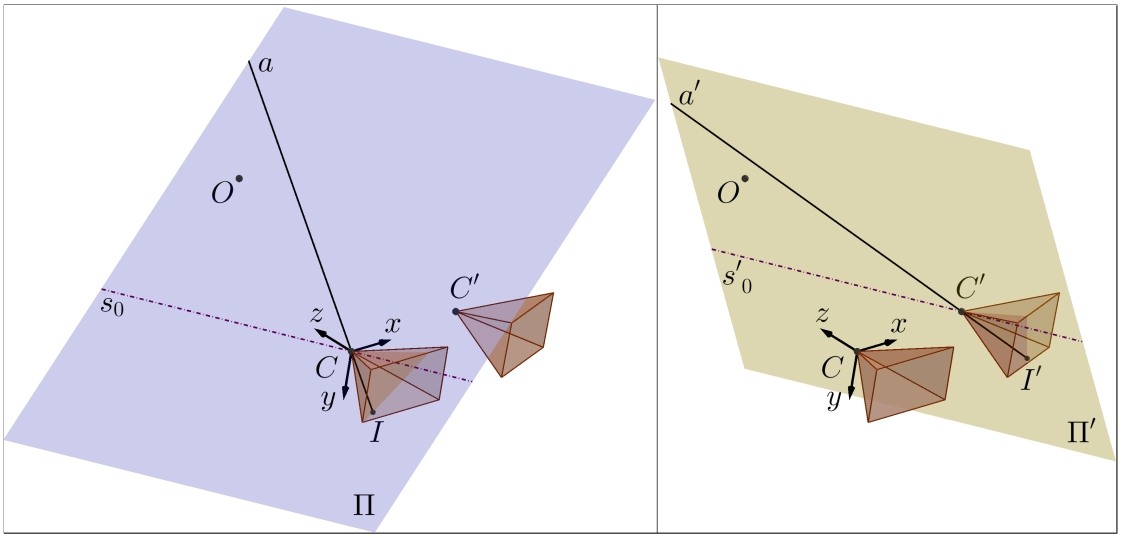

Figure 5.9: Left: Plane of refraction Π belonging to the image point I. Right: Plane of refraction Π' belonging to the image point I'.

5.3.1 Computation of Virtual Object Points

The computation of the *VOP*s in a *SFRS* is closely related to the process of intersecting the planes of refraction Π and Π' belonging to a pair of corresponding image points I and I', as illustrated in Figure 5.9. To maintain clarity, the illustration is done separately for both planes.

Requirements. The required quantity is the direction vector $\bar{\boldsymbol{s}}$ of the system axis s_0.

Preparatory Computations. The first step is the computation of the normalized direction vectors $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{a}}'$ of the rays a and a' in their respective camera coordinate systems. This can be realized with the aid of Equation 2.17 from a pair of corresponding image points I and I' and the known intrinsic camera parameters. Since both direction vectors are computed in the camera coordinate system of their respective camera, one of them needs to be transformed. In this case, the direction vector $\bar{\boldsymbol{a}}'$ is transformed into the camera coordinate system of the left camera, since the direction vector $\bar{\boldsymbol{s}}$ of the system axis s_0 is also know with respect to it.
The ray a and the system axis s_0 form the plane of refraction Π, as shown on the left side of Figure 5.9. The plane of refraction Π' on the right side of Figure 5.9 is formed by the ray a' and the ray s_0', which passes through C'. The ray s_0' has the same normalized direction vector $\bar{\boldsymbol{s}}$ as the system axis s_0. Since both planes of refraction originate from the same real object point O, they can either be entirely coincident or they can intersect in a line, but they can not be non-intersecting. If both planes Π and Π' intersect in a line, this line always has the same direction vector $\bar{\boldsymbol{s}}$ as the system axis s_0. It is straightforward

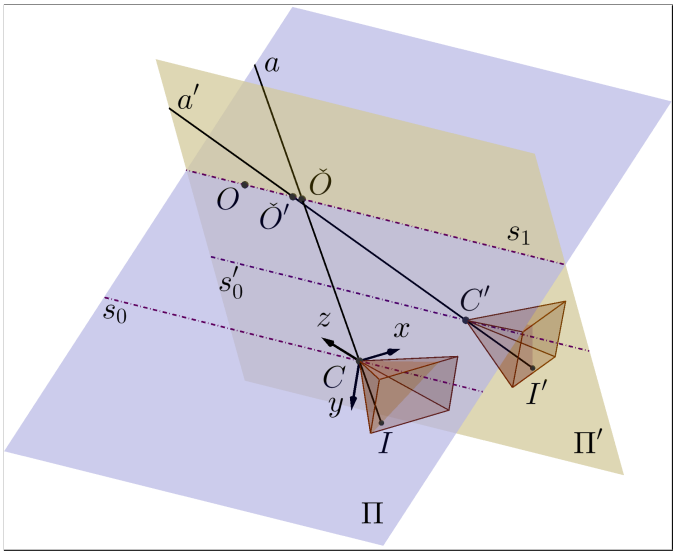

Figure 5.10: Intersection of the planes of refraction Π and Π'.

that this is always true. Mathematically, this can be derived from the definition of the double cross product [BS79]:

$$\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b} \cdot (\boldsymbol{a} \cdot \boldsymbol{c}) - \boldsymbol{c} \cdot (\boldsymbol{a} \cdot \boldsymbol{b}). \tag{5.4}$$

By utilization of the cross product, the normal vector of plane Π is $\boldsymbol{n} = \bar{\boldsymbol{s}} \times \bar{\boldsymbol{a}}$ and that of Π' is $\boldsymbol{n}' = \bar{\boldsymbol{s}} \times \bar{\boldsymbol{a}}'$. Since the line of intersection in question has to be orthogonal to both these normal vectors, it can be computed by the cross product $\boldsymbol{n} \times \boldsymbol{n}'$. Then, the application of Equation 5.4 leads to:

$$(\bar{\boldsymbol{s}} \times \bar{\boldsymbol{a}}) \times (\bar{\boldsymbol{s}} \times \bar{\boldsymbol{a}}') = \bar{\boldsymbol{s}} \cdot \overbrace{((\bar{\boldsymbol{s}} \times \bar{\boldsymbol{a}}) \cdot \bar{\boldsymbol{a}}')}^{\lambda} - \bar{\boldsymbol{a}}' \cdot \overbrace{((\bar{\boldsymbol{s}} \times \bar{\boldsymbol{a}}) \cdot \bar{\boldsymbol{s}})}^{0} \tag{5.5}$$

$$= \lambda \bar{\boldsymbol{s}}. \tag{5.6}$$

The result of the scalar triple product $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$ is always a scalar value. Therefore, the first term of Equation 5.5 amounts to some scalar value λ and the second term amounts to zero. This shows that the final direction vector of the line of intersection of the planes Π and Π' becomes $\bar{\boldsymbol{s}}$ scaled by λ. This line of intersection is the object axis s_1, as can be seen in Figure 5.10.

Virtual Object Points. The *VOP*s $\check{O}$ and $\check{O}'$ need to be situated on the object axis s_1. Since the real object point O is unknown, both *VOP*s are computed by intersecting ray a with plane of refraction Π' and ray a' with plane Π. This is done exemplary for the first case. The plane is expressed in the Hesse normal form of the equation of a plane:

$$\Pi' : \bar{\boldsymbol{n}}' \cdot \boldsymbol{o} = d', \tag{5.7}$$

with $\bar{\boldsymbol{n}}'$ being the normalized normal vector of the plane Π', $\boldsymbol{o}$ being the position vector of a point on the plane and d' being the distance of the plane to the origin of the coordinate system in normal direction. This distance d' is computed by $d' = \bar{\boldsymbol{n}}' \cdot \boldsymbol{c}'$, with $\boldsymbol{c}'$ being the position vector of C'. The ray is expressed as a line in vector form by:

$$a : \boldsymbol{o} = \boldsymbol{c} + \lambda \cdot \bar{\boldsymbol{a}}, \tag{5.8}$$

with $\boldsymbol{o}$ being the position vector of a point on the line, $\boldsymbol{c}$ being the position vector of C (note: $\boldsymbol{c} = \boldsymbol{0}$), λ being a scalar and $\bar{\boldsymbol{a}}$ being the direction vector of the line. Substitution of Equation 5.8 into Equation 5.7 leads to:

$$\bar{\boldsymbol{n}}' \cdot \boldsymbol{c} + \lambda \cdot \bar{\boldsymbol{n}}' \cdot \bar{\boldsymbol{a}} = d'. \tag{5.9}$$

Rearranging Equation 5.7 for λ and substitution into Equation 5.8 results in the position vector $\boldsymbol{o}$ of the point of intersection:

$$\boldsymbol{o} = \boldsymbol{c} + \frac{d' - \bar{\boldsymbol{n}}' \cdot \boldsymbol{c}}{\bar{\boldsymbol{n}}' \cdot \bar{\boldsymbol{a}}} \cdot \bar{\boldsymbol{a}}, \tag{5.10}$$

which is exactly the *VOP* $\check{O}$. By substitution of plane Π and ray a', the position vector $\boldsymbol{o}'$ of the *VOP* $\check{O}'$ is computed equivalently by Equations 5.7 to 5.10 (note: $\boldsymbol{c}' \neq \boldsymbol{0}$ and $d = 0$).
As can be seen, the computation of the *VOP*s in a *SFRS* is completely independent of the refractive indices and the layer thicknesses. Hence, these refractive parameters do not need to be considered. A limitation of this approach is that the *VOP*s can only be computed in the presented way if the planes of refraction Π and Π' intersect in a line. If the planes coincide, the rays a and a' obviously coincide with these planes and no point of intersection can be computed.

5.3.2 Computation of Real Object Points

Figure 5.11 shows that the rays a and a' are refracted at each of the two parallel interfaces Φ_1 and Φ_2 and that the resulting rays w and w' intersect at the real object point O. This real object point can be computed from each of the *VOP*s $\check{O}$ and $\check{O}'$, independently of the other. In contrast to the computation of the *VOP*s, the remaining refractive parameters need to be considered now.

Requirements. In addition to the already known normalized direction vector $\bar{\boldsymbol{s}}$ of the system axis s_0, the required quantities are the refractive indices n_a, n_w, n_g, the glass layer thickness $t_g = d(S_1, S_2)$ and the air layer thickness $t_a = d(C, S_1)$.

Real Object Points. The necessary computation steps are as follows:

- Compute the angle of incidence α between the two direction vectors $\bar{\boldsymbol{s}}$ and $\bar{\boldsymbol{a}}$ and/or compute the angle of incidence α' between the two direction vectors $\bar{\boldsymbol{s}}$ and $\bar{\boldsymbol{a}}'$.

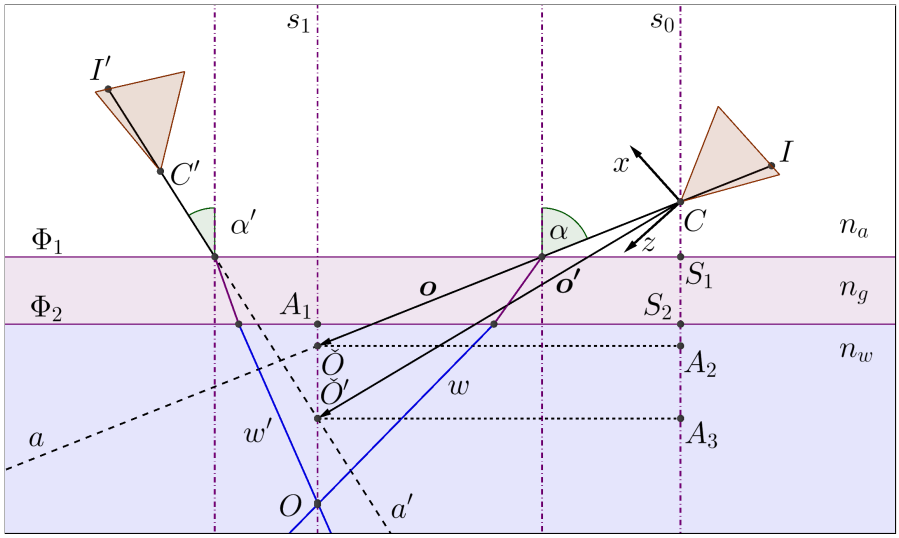

Figure 5.11: Computation of real object points from *VOP*s.

- Compute the distance $d(C, A_2)$ by orthogonal projection of the position vector $\boldsymbol{o}$ of *VOP* $\check{O}$ onto vector $\bar{\boldsymbol{s}}$ and/or
  compute the distance $d(C, A_3)$ by orthogonal projection of the position vector $\boldsymbol{o}'$ of *VOP* $\check{O}'$ onto vector $\bar{\boldsymbol{s}}$ with the aid of Equation 2.7.
- Compute $\check{x} = d(A_1, \check{O}) = d(C, A_2) - t_a - t_g$ and/or
  compute $\check{x}' = d(A_1, \check{O}') = d(C, A_3) - t_a - t_g$.
- Compute the distance $x = d(A_1, O)$ from distance $\check{x}$ and angle of incidence α and/or from distance $\check{x}'$ and angle of incidence α' by application of Equation 4.7 from the known quantities n_a, n_w, n_g, t_a and t_g.
- Since the normalized direction vector $\bar{\boldsymbol{s}}$ of object axis s_1 is known, compute the position vector $\boldsymbol{o_1}$ of the real object point O by: $\boldsymbol{o_1} = \boldsymbol{o} + (x - \check{x})\bar{\boldsymbol{s}}$ and/or
  by: $\boldsymbol{o_1} = \boldsymbol{o}' + (x - \check{x}')\bar{\boldsymbol{s}}$.

As can be seen, the computation of the *VOP*s in combination with the computation of the real object point amounts to an efficient alternative for refractive back-projection. It additionally allows the direct recovery of the 3D coordinates of a real object point. Due to the previously mentioned limitation, it is not universally applicable. However, it is ideally suited for the purposes of this thesis, which will be described in the following sections.

5.4 Definition and Computation of Cost Functions

Let us assume that an approach for the determination of the refractive parameters already exists and that the results of this approach are hypothetical refractive parameters with no information about their quality. The function value of a cost functions has the purpose to reflect the accuracy of the determined parameters. The minimization of such a cost function leads to the optimization of the parameters. This section comprises the definition and computation of novel cost function, as well as the adapted computation of common ones. In every case, the basic operation from Sections 5.2 and 5.3 are utilized for the computations. The single cost functions show a progressive independence from calibration objects.
The distinction of two stages serves to proof that the refractive parameters can be determined to a certain degree independently of each other. The first stage includes the derivation of pattern-based invariants by utilization of the *VOP* model. These invariants are used for the definition of two pattern-based cost functions. Both of them are only dependent on the system axis of the *SFRS*. Therefore, they enable the optimization of the orientation of the system axis without considering the remaining refractive parameters. The restrictions on the necessary pattern on the calibration object decrease from invariant one to invariant two.
At the second stage, the cost functions are independent of calibration objects with special patterns. Instead, they make use of the recovery of 3D scene points or of the projection of these 3D scene points onto the image plane. In contrast to the first stage, these cost functions depend on all the refractive parameters and therefore serve their simultaneous optimization. The associated common cost functions are the reprojection error and the triangulation error, which will be adapted to refractive projections. Furthermore, a new development of a cost function will be presented, which is based on the *VOP* model itself. This section is partially based on Dolereit et al. [DLK15a; DLK15b].

5.4.1 Pattern-Based Invariants

In this section, it will be shown how invariants can be derived that only depend on the system axis and some properties of the pattern on the calibration object. The calibration object does not need to be measured completely. The only requirement is that the features of the pattern are arranged in straight lines. A checker pattern provides the necessary properties for both proposed invariants, while the single invariants do not depend on all of them. This makes it interchangeable. The restrictions on the used calibration object decrease between the first and the second invariant. The first invariant depends on a known arrangement between multiple straight lines on the pattern and the second depends only on feature points arranged in straight lines.

Elementary Components. Figure 5.12 shows the elements that will be needed for the derivation of the invariants. Note that only one of the flat refractive interfaces is illustrated to improve clarity. This is sufficient, since the second interface does not influence the basics of the invariants. Let us consider a set of real object points O_{ij}

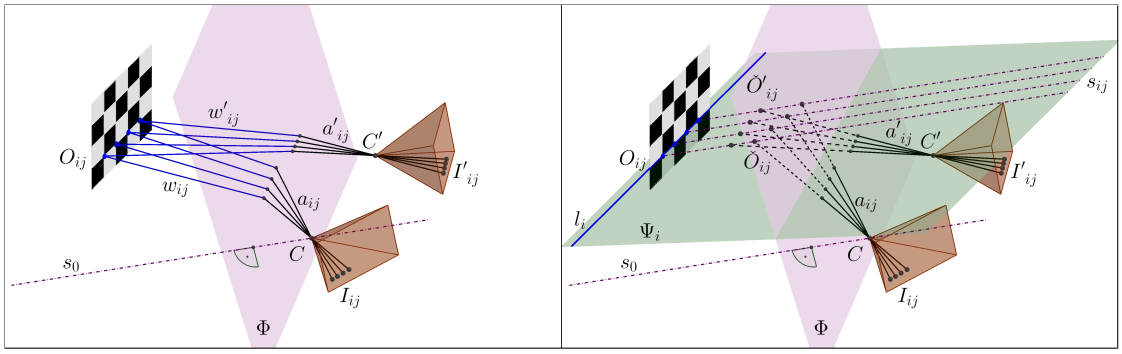

Figure 5.12: Left: Ray paths from the calibration object to the cameras. Right: Exemplary construction of a feature plane Ψ_i.

at the inner corners of the squares on a checker pattern arranged in $i \in \{1, .., m\}$ rows and $j \in \{1, .., n\}$ columns. On the left side of Figure 5.12, the ray paths for a subset of these points are shown. The emitted rays w_{ij} and w'_{ij} are refracted at the interface Φ. After passing through the respective centers of projection C or C', they end up at the left and right image points I_{ij} and I'_{ij}. The depicted points O_{ij} on the right side of Figure 5.12 form a line l_i. According to the utilized *VOP* model, the *VOP*s are situated on the object axes s_{ij}. All these object axes have the same normalized direction vector $\bar{\boldsymbol{s}}$, which is known from the system axis s_0. Hence, a line on the calibration object and the respective subset of the object axes s_{ij} form a plane. Such a plane will be called feature plane in the following. This is shown exemplary for the feature plane Ψ_i with $i = 1$, which is formed by any point O_{ij} on the line l_i, the normalized direction vector $\bar{\boldsymbol{l}}_i$ of this line and the normalized direction vector $\bar{\boldsymbol{s}}$ of the system axis s_0. The normalization of both direction vectors is not mandatory.
As proposed in Sections 4.2 and 5.3.1, the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$, seen by the left and the right camera, are located on the object axes s_{ij}, at the intersection points with the respective left and right rays a_{ij} and a'_{ij}. The basic constraint is that a feature plane, such as Ψ_i, contains all the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ that are related to the respective real object points O_{ij} on the line l_i. This is true for every feature plane in horizontal and vertical direction. In the following, it will be presented that this basic constraint implies invariants that can be utilized later on to define novel cost functions. All that is necessary is the assignment of *VOP*s to straight lines on the pattern of the calibration object.

1. Invariant Property

The first invariant is constrained the most, since it additionally makes use of an arrangement of the straight lines on the pattern. On a checker pattern, these lines run horizontally and vertically, distinguished in their direction by a right angle. On the left side of Figure 5.13, all the object axes s_{ij} are shown, while the cameras and the rays are neglected for better visibility. All the feature planes Ψ_i and Ψ_j are formed by these

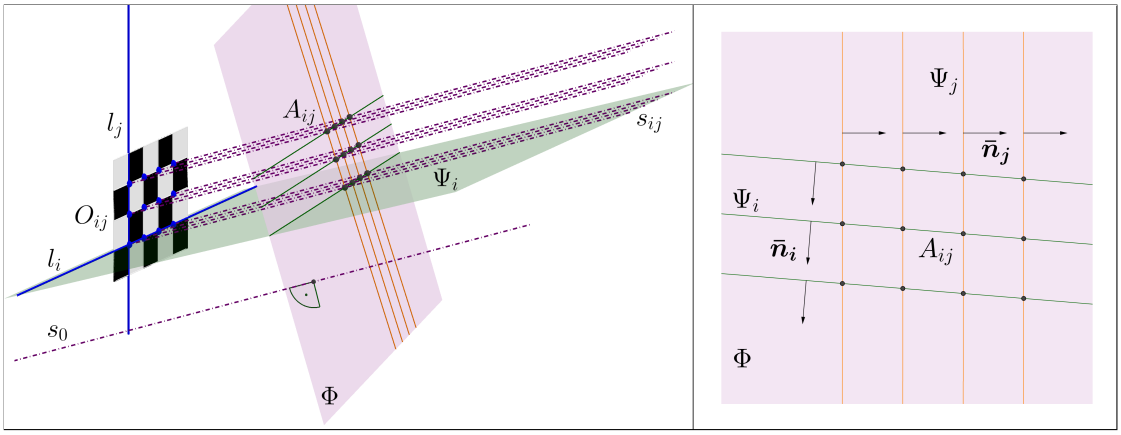

Figure 5.13: Left: Intersection of the object axes s_{ij} and of the feature planes Ψ_i and Ψ_j with the refractive interface Φ. Right: Orthogonal view of the refractive interface Φ.

object axes and the horizontal lines l_i or the vertical lines l_j on the checker pattern. This is shown exemplary for the feature plane Ψ_i with $i = 1$.
Let us consider the intersections of the feature planes with the refractive interface Φ. If the refractive interface is considered as a virtual image plane, the mapping of the real object points O_{ij} to the points A_{ij} has the same properties as an orthographic projection. The orthographic projection is an affine transformation that preserves parallel lines [HZ04]. This makes it evident that the feature planes originating from a checker pattern have to be parallel in horizontal as well as in vertical direction. The right side of Figure 5.13 shows an orthogonal view of the refractive interface Φ. Since the viewing direction is exactly the same as the direction of the object axes s_{ij}, these are seen as the points A_{ij} and the feature planes Ψ_i and Ψ_j are seen as parallel lines in two different directions. The shear of the rectangles is a direct consequence of a checker pattern that is not parallel to the refractive interface. As should be clear by now, an additional parallel interface does not change these properties.
The first pattern-based invariant property is based on the constraint that the feature planes are parallel in horizontal as well as in vertical direction. The orientation of a feature plane can be represented by its normalized normal vector.

- The first part of the invariant property can be defined as that the angles between the normal vectors of parallel feature planes have to be zero.

Besides the parallel feature planes, the ones that intersect can be considered as well.

- The second part of the invariant can be defined as that the angles between the normal vectors of all the intersecting feature planes have to be equal.

Note that the second part of the invariant is in this form only valid for feature planes that run parallel in only two different directions, as in the case of a checker pattern. If feature planes in more than two directions are considered, it needs to be modified

accordingly. The angle between two feature planes $\alpha(\bar{\boldsymbol{n}}_1, \bar{\boldsymbol{n}}_2)$, with $\bar{\boldsymbol{n}}_1$ and $\bar{\boldsymbol{n}}_2$ being the normalized normal vectors of the feature planes, can be computed with the aid of Equation 2.1. The arithmetic mean of the all the angles between the parallel feature planes and likewise between the intersecting feature planes can be computed by:

$$\mu_p = \frac{1}{\binom{m}{2}} \left(\sum_{i=1}^{m-1} \sum_{k=1}^{m-i} \alpha(\bar{\boldsymbol{n}}_i, \bar{\boldsymbol{n}}_{i+k}) \right) + \frac{1}{\binom{n}{2}} \left(\sum_{j=1}^{n-1} \sum_{k=1}^{n-j} \alpha(\bar{\boldsymbol{n}}_j, \bar{\boldsymbol{n}}_{j+k}) \right) = 0, \tag{5.11}$$

$$\mu_c = \frac{1}{m \cdot n} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \alpha(\bar{\boldsymbol{n}}_i, \bar{\boldsymbol{n}}_j) \right). \tag{5.12}$$

In the case of parallel feature planes, the mean μ_p is computed in horizontal and vertical direction. Since it is supposed to be equal to zero, it is used directly in the expression for the invariant. In the case of intersecting feature planes, the mean μ_c has to be subtracted from each angle. Overall, the invariant has to result in zero and can be expressed by:

$$\mu_p + \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \alpha(\bar{\boldsymbol{n}}_i, \bar{\boldsymbol{n}}_j) - \mu_c \right) = 0. \tag{5.13}$$

It is worth mentioning that this invariant could be extended to all possible straight lines on the pattern. This includes the diagonals in both directions. Furthermore, any calibration object with features arranged in straight lines could be used instead, if comparable relations between the feature planes can be defined.

2. Invariant Property

The second invariant property does not depend on a known arrangement of the straight lines on the calibration object. The only prerequisite is the arrangement of the points O_{ij} on the calibration object in straight lines. Let us take the feature plane Ψ_i on the right side of Figure 5.12 as an example. The invariant is based on the fact that all the object axes s_{ij} that originate from line l_i lie in the same feature plane Ψ_i. The direct consequence is that the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ that are related to the real object points O_{ij} on the line l_i also have to lie in this plane.
The second pattern-based invariant property is based on the constraint that the feature planes contain *VOP*s.

- The invariant property can be defined as that the distance between a feature plane and all of its associated *VOP*s has to be zero.

If the Hesse normal form of the equation of the feature plane is available:

$$\Psi_i : \bar{\boldsymbol{n}}_i \cdot \boldsymbol{x} = d_i,$$

then Equation 2.5 can be utilized. It directly yields the Euclidean distance $d(A, \Psi_i)$ between an arbitrary point A and the feature plane Ψ_i. Note that the point A needs

to be substituted by the *VOP*s $\check{O}_{ij}$ or $\check{O}'_{ij}$. Since this distance is supposed to be zero between all the *VOP*s and their associated feature plane, the invariant can be expressed exemplary for the line l_i with $i \in \{1, .., m\}$ on the calibration object by:

$$\sum_{j=1}^{n} d(\check{O}_{ij}, \Psi_i) + \sum_{j=1}^{n} d(\check{O}'_{ij}, \Psi_i) = 0. \tag{5.14}$$

Every single line on the pattern of the calibration object is independent of the remaining straight lines. Therefore, the utilized checker pattern can be replaced accordingly.

5.4.2 Pattern-Based Cost Functions

In this section, the previously derived pattern-based invariants are used to define cost functions. Since the restrictions on the used calibration object decrease between the invariants, this is same for the defined cost functions. Due to the independence of the refractive indices and the layer thicknesses, these cost functions can be used to optimize the orientation of the system axis individually (See Section 5.5.1). Both pattern-based cost functions are able to reflect the accuracy of the determined axis.

Plane Angles Error

For the definition of the *Plane Angles Error* (*PAE*), the first invariant from Section 5.4.1 is utilized, which is based on the arrangement of feature points in straight lines on the pattern of a calibration object. Summarized, the underlying constraint is that the formed feature planes run parallel in their respective directions. The first part of the invariant property is that the angles between the normal vectors of parallel feature planes have to be zero and the second part is that the angles between the normal vectors of intersecting feature planes have to be equal.
Figure 5.14 illustrates the approach for the computations. Note that it is generally valid for both pattern-based cost functions. Since the coordinates of the real object points are usually unknown, their corresponding image points I_{ij} and I'_{ij} form the starting point of any calculation. The indices $i \in \{1, .., m\}$ and $j \in \{1, .., n\}$ represent the lines in horizontal and in vertical direction on the utilized calibration object. Let us assume that a hypothetical system axis h_0 is known and the true system axis s_0 is searched for. The necessary steps for the computation of a function value of the cost function are as follows:

- Compute the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ from pairs of corresponding image points I_{ij} and I'_{ij} for a hypothetical system axis h_0, as described in Section 5.3.1.
- Arrange these *VOP*s according to the lines in horizontal (l_i) and vertical (l_j) direction on the checker pattern.
- Fit a plane to the set of *VOP*s per line to get the feature planes in horizontal direction $\Theta_i : \bar{\boldsymbol{n}}_i \cdot \boldsymbol{x} = d_i$ and the feature planes in vertical direction $\Theta_j : \bar{\boldsymbol{n}}_j \cdot \boldsymbol{x} = d_j$, with $\bar{\boldsymbol{n}}_i$ and $\bar{\boldsymbol{n}}_j$ being the normalized normal vectors, $\boldsymbol{x}$ being the position vector

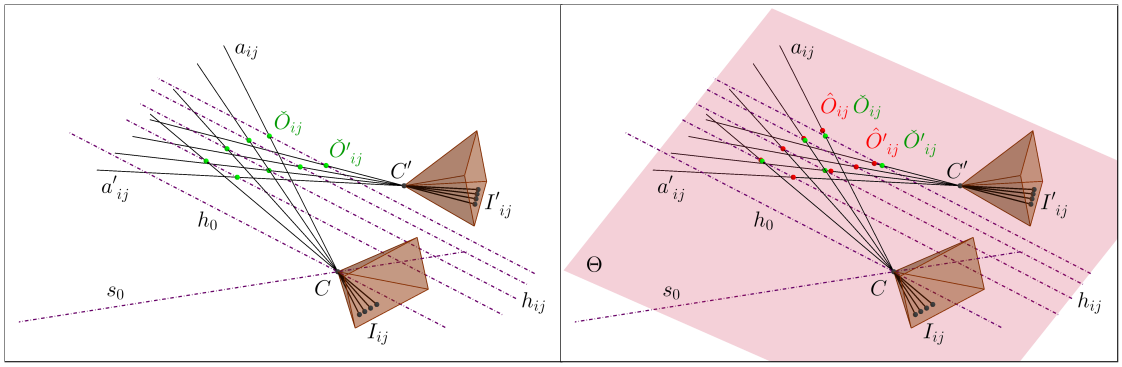

Figure 5.14: Left: Computation of the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ for a hypothetical system axis h_0. Right: Erroneous resulting plane Θ from plane fitting into the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$.

to a point on the plane and d_i as well as d_j being the distances of the respective plane to the origin of the coordinate system.

The fitting of a plane Θ into the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$, which are known to belong to a single straight line on the calibration object, is illustrated exemplary on the right side of Figure 5.14.
By means of Equations 5.11 to 5.13 and with $\alpha(\bar{\boldsymbol{n}}_1, \bar{\boldsymbol{n}}_2)$ being the angle between the normalized normal vectors of two fitted feature planes (Equation 2.1), the *PAE* is defined by:

$$
\begin{aligned}
E &= E_p + E_c, \qquad (5.15)\\
E_p &= \frac{1}{\binom{m}{2}} \left(\sum_{i=1}^{m-1} \sum_{k=1}^{m-i} \alpha(\bar{\boldsymbol{n}}_i, \bar{\boldsymbol{n}}_{i+k})^2 \right) + \frac{1}{\binom{n}{2}} \left(\sum_{j=1}^{n-1} \sum_{k=1}^{n-j} \alpha(\bar{\boldsymbol{n}}_j, \bar{\boldsymbol{n}}_{j+k})^2 \right),\\
E_c &= \frac{1}{m \cdot n} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \left(\alpha(\bar{\boldsymbol{n}}_i, \bar{\boldsymbol{n}}_j) - \mu_c\right)^2 \right),\\
\mu_c &= \frac{1}{m \cdot n} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \alpha(\bar{\boldsymbol{n}}_i, \bar{\boldsymbol{n}}_j) \right).
\end{aligned}
$$

In this set of equations E_p denotes the error resulting from the parallel feature planes in horizontal as well as in vertical direction, while E_c denotes the error resulting from the intersecting feature planes. The mean of the angles between the intersecting feature planes is denoted by μ_c. Note that a square function is used to penalize larger deviations more severely.

Plane Fitting Error

For the definition of the *Plane Fitting Error* (*PFE*), the second invariant from Section 5.4.1 is utilized, which is based on the arrangement of feature points in straight lines on

a calibration object. Summarized, the underlying constraint is that the feature planes contain *VOP*s and the invariant property derived from this is that all the distances between a feature plane and its associated *VOP*s has to be zero.
The necessary steps for the computation of the error measure are basically the same as for the *PAE* (See Section 5.4.2):

- Compute the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ from pairs of corresponding image points I_{ij} and I'_{ij} for a hypothetical system axis h_0, as described in Section 5.3.1.
- Arrange these *VOP*s according to the lines in horizontal (l_i) and vertical (l_j) direction on the checker pattern.
- Fit a plane to the set of *VOP*s per line to get the feature planes in horizontal direction $\Theta_i : \bar{\boldsymbol{n}}_i \cdot \boldsymbol{x} = d_i$ and the feature planes in vertical direction $\Theta_j : \bar{\boldsymbol{n}}_j \cdot \boldsymbol{x} = d_j$, with $\bar{\boldsymbol{n}}_i$ and $\bar{\boldsymbol{n}}_j$ being the normalized normal vectors, $\boldsymbol{x}$ being the position vector to a point on the plane and d_i as well as d_j being the distances of the respective plane to the origin of the coordinate system.

Let us once again consider the exemplary plane fitting on the right side of Figure 5.14. For an erroneous hypothetical system axis h_0, this fitted plane Θ_i is most likely to deviate from the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$. This is illustrated by the points $\hat{O}_{ij}$ and $\hat{O}'_{ij}$, which are the new intersection points of the rays a_{ij} and a'_{ij} with plane Θ_i. The cost function is a measure of the distances between the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ and their associated fitted feature planes Θ_i and Θ_j. By means of Equation 5.14, the *PFE* is defined by:

$$
\begin{aligned}
E &= E_h + E_v, \qquad (5.16)\\
E_h &= \frac{1}{m \cdot n}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} d(\check{O}_{ij}, \Theta_i)^2\right) + \frac{1}{m \cdot n}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} d(\check{O}'_{ij}, \Theta_i)^2\right),\\
E_v &= \frac{1}{m \cdot n}\left(\sum_{j=1}^{n}\sum_{i=1}^{m} d(\check{O}_{ij}, \Theta_j)^2\right) + \frac{1}{m \cdot n}\left(\sum_{j=1}^{n}\sum_{i=1}^{m} d(\check{O}'_{ij}, \Theta_j)^2\right).
\end{aligned}
$$

In this set of equations E_h denotes the error resulting from the lines on the checker pattern in horizontal direction and E_v denotes the error resulting from the lines in vertical direction. Note that a square function is used to penalize larger deviations more severely.

5.4.3 3D Scene Cost Functions

In contrast to the two pattern-based cost functions from the previous section, the following two do not depend on a special pattern on a calibration object, nor do they depend on a calibration object at all. All that is necessary are stereo correspondences. A set of left and right image points I_i and I'_i, with $i \in (1, .., k)$, forms the starting point of the calculations. They correspond pairwise and every pair is related to a single real object point. Assuming that hypothetical refractive parameters are known, this fact can be exploited. The calibrated stereo camera together with these refractive parameters enables

the recovery of the 3D coordinates of a real object point by intersection of the associated left and the right ray path. This can be realized, for example, by triangulation of the rays. For erroneous hypothetical refractive parameters, these rays are most likely to intersect not at all. In that case, the cost function indicates how far they are apart. An alternative to the triangulation-based cost function is the proposed second cost function. Its definition and its computation is a result of the application of the computation of virtual and real object points from Sections 5.3.1 and 5.3.2.
Both these cost functions make use of the recovery of the 3D scene and are able to reflect the quality of all the determined refractive parameters. These cost functions can be used to optimize all the refractive parameters in question simultaneously (See Section 5.5.3).

Refractive Triangulation Error

The *Refractive Triangulation Error* (RTE) makes use of the constraint that the left ray w in water and the right ray w' in water, which are computed from a pair of corresponding image points I and I', need to intersect in a single real object point O. This leads to the invariant property that the shortest distance between two corresponding rays needs to be equal to zero. For the computation of the rays, refractive back-projection is necessary. The RTE reflects if both rays intersect and if not, how far they are apart. The necessary steps for the computation of the RTE are:

- Compute the left rays w_i and the right rays w_i' from pairs of corresponding image points I_i and I_i' for a hypothetical axis h_0 and a hypothetical air layer thickness $h_a = d(C, H_1)$ by refractive back-projection, as described in Section 5.2.2.

- Transform either w_i or w_i' to the coordinate system of the opposite camera with the known transformation $\boldsymbol{R}, \boldsymbol{t}$ between both camera coordinate systems.

On the left side of Figure 5.15, the computation of the RTE is illustrated exemplary for one stereo correspondence. For an erroneous hypothetical axis and/or air layer thickness, the left ray w and the right ray w' are most likely to be non-intersecting and with that their shortest distance will not be zero. How far they are apart can be computed by the shortest distance between two lines, which is the distance between the points O and O'. The magnified section on the right side of Figure 5.15 shows this shortest distance, which is characterized by the length of a line segment that is orthogonal to both rays w and w'. The endpoints of this line segment are exactly the points O and O'. With $d(x_1, x_2)$ being the shortest distance between two lines (See Equation 2.4), the RTE is defined by:

$$E = \frac{1}{k} \left(\sum_{i=1}^{k} d\left(w_i, w_i'\right)^2 \right). \tag{5.17}$$

Note that the explicit computation of the real object points O_i and O_i' is not necessary for the computation of the RTE. Once again, a square function is used to penalize larger deviations more severely.

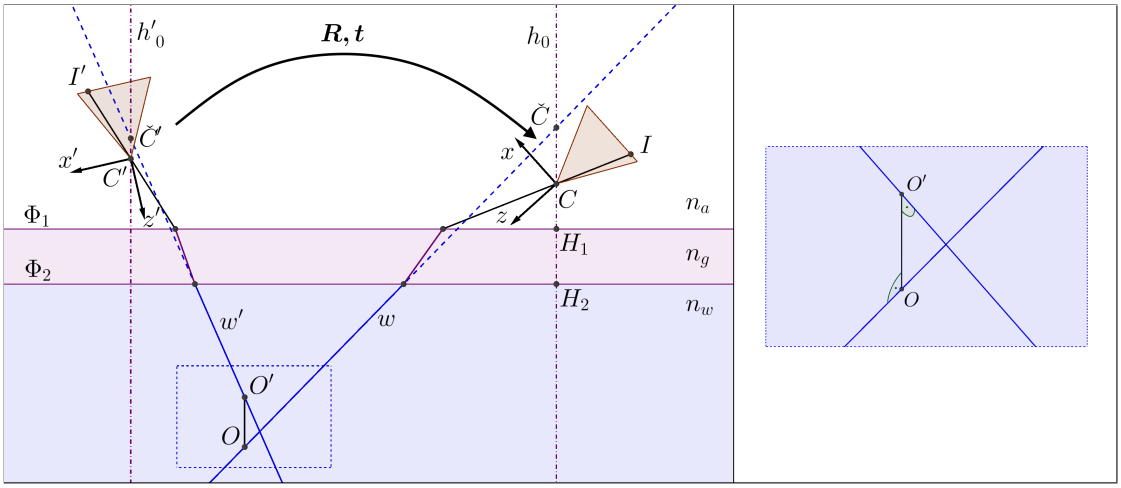

Figure 5.15: Left: Schematic computation of the *RTE*. Right: The shortest distance between the two rays w_i and w_i'.

Virtual Intersection Error

Like the *RTE*, the *Virtual Intersection Error* (*VIE*) makes use of the constraint that the left ray w in water and the right ray w' in water, which are computed from a pair of corresponding image points I and I', need to intersect in a single real object point O. From the *VOP* model from Section 4.3, a novel invariant property can be derived. As is known by now, a left *VOP* and a right *VOP* can be computed from a pair of corresponding left and right image points. Both these *VOP*s differ in general and they can coincide if both angles of incidence are equal. However, both related real object points need to coincide always. Therefore, the distance between the real object point O that is computed from a left *VOP* $\check{O}$ and the real object point O' that is computed from the right *VOP* $\check{O}'$ needs to be equal to zero. The *VIE* reflects if both real object points coincide and if not, how far they are apart. The necessary steps for the computation of the *VIE* are essentially the same as for the computation of the virtual and real object points:

- Compute the *VOP*s $\check{O}_i$ and $\check{O}_i'$ from pairs of corresponding image points I_i and I_i' for a hypothetical axis h_0, as described in Section 5.3.1.

- Compute the distances to the real object points $x_i = d(A_{1i}, O_i)$ and $x_i' = d(A_{1i}, O_i')$ for a hypothetical air layer thickness $h_a = d(C, H_1)$, as described in Section 5.3.2.

Note that the explicit computation of the coordinates of the real object points O_i and O_i' is not necessary. On the left side of Figure 5.16, the computation of the *VIE* is illustrated exemplary for one stereo correspondence. For an erroneous hypothetical axis and/or air

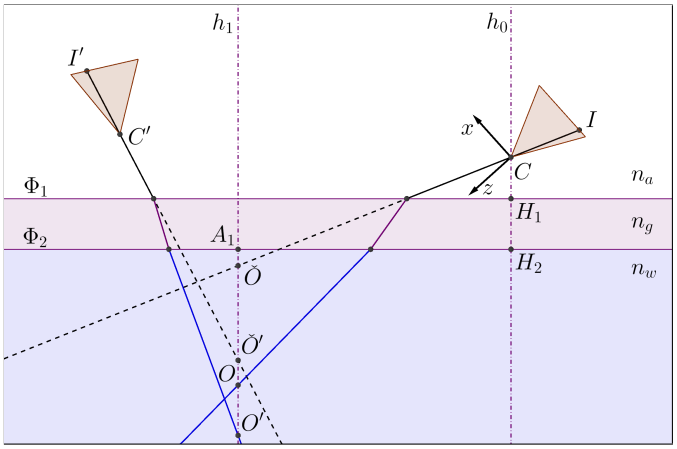

Figure 5.16: Schematic computation of the *VIE*.

layer thickness, the distances x_i and x_i' are most likely to differ. How far they are apart can be computed straightforward. Therefore, the *VIE* is defined by:

$$E = \frac{1}{k}\left(\sum_{i=1}^{k}\left(x_i - x_i'\right)^2\right). \tag{5.18}$$

Once again, a square function is used to penalize larger deviations more severely. As has been shown earlier, the necessary computations are an efficient alternative to refractive-back projection, which avoid ray tracing completely. Therefore, the *VIE* is expected to be computationally more efficient than the *RTE*.

5.4.4 2D Image Cost Functions

In contrast to the two cost functions from the previous section, the following two cost functions go one step further. While the *RTE* and the *VIE* stopped in 3D space, the cost functions in this section additionally require a forward projection onto the 2D image plane. Similarly to the previous section, no special pattern on a calibration object, nor a calibration object at all is necessary. All that is necessary are stereo correspondences. One again, the image points I_i and I_i' with $i \in (1, .., k)$ form the starting point of the calculations. Assuming that hypothetical refractive parameters are known, the stereo setup enables the recovery of 3D coordinates, as already described in the previous section. For erroneous hypothetical refractive parameters, the recovered 3D scene points are most likely to be erroneous as well. The accuracy of these points can be indicated by the common reprojection error. However, its computation needs to be adapted for explicit consideration of refractive effects. This can be supported by the approach for refractive forward projection from Section 5.2.1. An alternative implementation is the computation of the reprojection error by utilization of the virtual camera model from Section 5.2.3. Both approaches are presented in the following.

Refractive Reprojection Error

The *Refractive Reprojection Error* (*RRE*) makes use of the constraint that the image points are projections of the real object points. Forward projection of a recovered 3D scene point $\hat{O}$ results in a projected image point $\hat{I}$. This leads to the invariant property that the distance between a projected image point $\hat{I}$ and a measured image point I needs to be equal to zero. The *RRE* is supposed to reflect if the left and the right projected image points coincide with the according measured image points and if not, how far they are apart. The first two steps for the computation of a function value of the cost function are essentially the same as the ones for the *VIE* (See Section 5.4.3):

- Compute the *VOP*s $\check{O}_i$ and $\check{O}'_i$ from pairs of corresponding image points I_i and I'_i for a hypothetical axis h_0, as described in Section 5.3.1.
- Compute the distances to the real object points $x_i = d(A_{1i}, O_i)$ and $x'_i = d(A_{1i}, O'_i)$ for a hypothetical air layer thickness $h_a = d(C, H_1)$, as described in Section 5.3.2.

On the left side of Figure 5.17, the computation of the *RRE* is illustrated exemplary for one stereo correspondence. For an erroneous hypothetical axis and/or air layer thickness, the distances x_i and x'_i are most likely to differ. Before application of refractive forward projection, the necessary input needs to be computed first. Note that the explicit recovery of the coordinates of the real object points O_i is not necessary for the proposed approach for refractive forward projection in Section 5.2.1. Therefore, the computation of the necessary input reads as follows:

- Compute the mean values x_{0i} of the distances x_i and x'_i to get the distances to the recovered 3D scene points $\hat{O}_i$:

$$x_{0i} = d(A_{1i}, \hat{O}_i) = \frac{x_i + x'_i}{2}.$$

The distances x_{0i} are sufficient to perform refractive forward projection (Illustrated by red elements in Figure 5.17) for both cameras, as described in Section 5.2.1, comprising:

- the iterative computation of the distances $\check{x}_i = d(A_{1i}, \check{O}_{0i})$ and $\check{x}'_i = d(A_{1i}, \check{O}'_{0i})$,
- the computation of the *VOP*s $\check{O}_{0i}$ from the distances $\check{x}_i$,
- the computation of the *VOP*s $\check{O}'_{0i}$ from the distances $\check{x}'_i$ and
- the perspective projection of the *VOP*s $\check{O}_{0i}$ and $\check{O}'_{0i}$ to the image points $\hat{I}_i$ and $\hat{I}'_i$.

The projected image points $\hat{I}_i$ and $\hat{I}'_i$ and the measured image points I_i and I'_i are illustrated exemplary on the right side of Figure 5.17. By means of the reprojection error in Equation 2.18 and with $d(X_1, X_2)$ being the distance between two 2D points, the *RRE* is defined by:

$$E = \frac{1}{k}\left(\sum_{i=1}^{k} d\left(I_i, \hat{I}_i\right)^2 + d\left(I'_i, \hat{I}'_i\right)^2\right). \tag{5.19}$$

Once again, a square function is used to penalize larger deviations more severely.

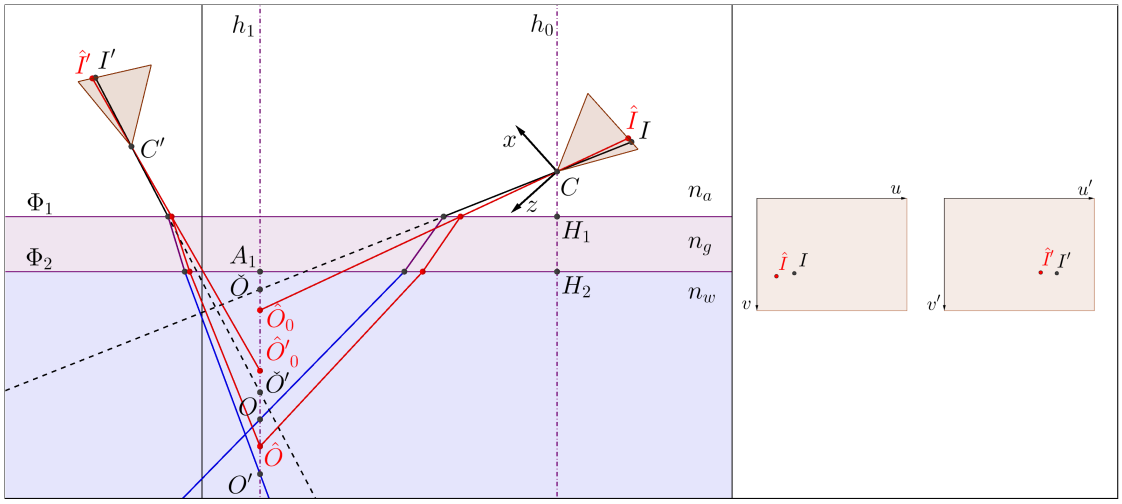

Figure 5.17: Left: Schematic computation of the *RRE*. Right: Top view of the image planes.

Virtual Reprojection Error

Like the *RRE*, the *Virtual Reprojection Error* (*VRE*) makes use of the constraint that the image points are projections of the real object points. Therefore, the invariant property is the same as for the *RRE* (Compare with Section 5.4.4), which is that the distance between a projected image point $\hat{I}$ and a measured image point I needs to be equal to zero. The *VRE* is also supposed to reflect if the left and the right projected image points coincide with the according measured image points and if not, how far they are apart. Although the basics and the purpose of the *RRE* and the *VRE* are identical, the computation of a function value of the cost function is considerably different. The first two steps are to a certain degree equal to the ones of the *RTE* in Section 5.4.3:

- Compute the left rays w_i in water and the right rays w_i' in water from pairs of corresponding image points I_i and I_i' for a hypothetical axis h_0 and a hypothetical air layer thickness $h_a = d(C, H_1)$ by refractive back-projection, as described in Section 5.2.2.

- Transform either w_i or w_i' to the coordinate system of the opposite camera with the known transformation $\boldsymbol{R}, \boldsymbol{t}$ between both camera coordinate systems.

On the left side of Figure 5.18, the computation of the *VRE* is illustrated exemplary for one stereo correspondence. For an erroneous hypothetical axis and/or air layer thickness, the corresponding left ray w_i in water and the right ray w_i' in water are most likely to be non-intersecting. Therefore, the best fitting real object points $\hat{O}_i$ are computed by triangulation of non-intersecting lines. The magnified section on the bottom right side of Figure 5.18 shows the shortest distance between the rays w_i and w_i', which is characterized by the length of a line segment that is orthogonal to both rays. The endpoints of this line segment are exactly the points O_i and O_i' and their midpoints are

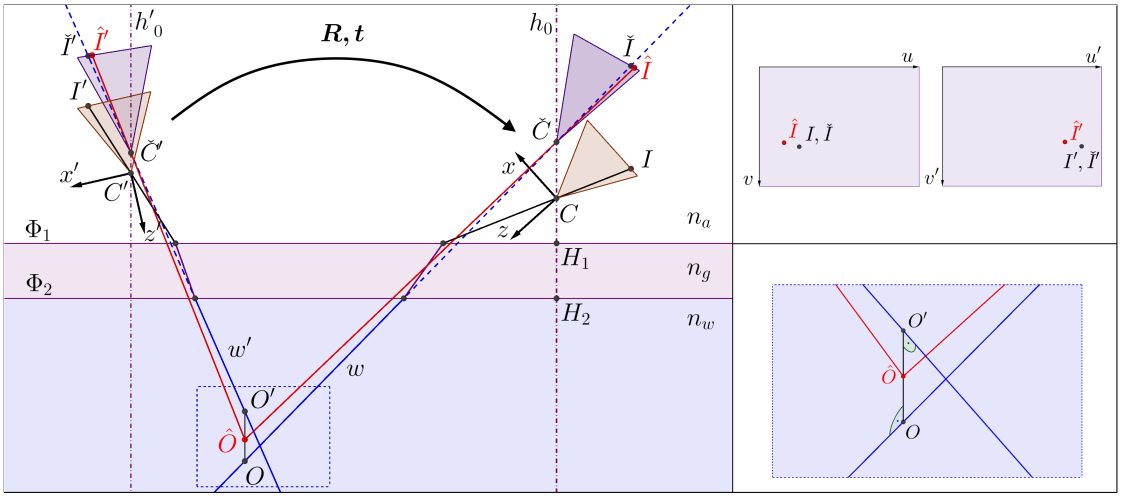

Figure 5.18: Left: Schematic computation of the *VRE*. Top Right: Superimposition of the virtual and the real image planes. Bottom Right: Best fitting real object point $\hat{O}_i$ from triangulation of the rays w_i and w'_i.

considered as being the best fitting real object points $\hat{O}_i$. To avoid an iterative approach for refractive forward projection, these points $\hat{O}_i$ can alternatively be projected into a virtual camera. As described earlier, a virtual camera needs to be determined per image point. Therefore, the following steps need to be performed:

- Compute the virtual camera parameters and hence the virtual projection matrix for each image point I_i and I'_i, as described in Section 5.2.3.
- Compute the image points $\hat{I}_i$ and $\hat{I}'_i$ by perspective projection of the triangulated real object points $\hat{O}_i$ with the according projection matrices.

The resulting projected image points $\hat{I}_i$ and $\hat{I}'_i$, the measured image points I_i and I'_i and the virtual image points $\check{I}_i$ and $\check{I}'_i$ are illustrated exemplary on the top right side of Figure 5.18. The advantage of the proposed virtual camera model is that the measured image points have the same coordinates in the coordinate systems of the virtual and of the real camera by definition. Therefore, by avoiding refractive forward projection, the function values of the cost function can be computed directly. By means of the reprojection error in Equation 2.18, with $d(X_1, X_2)$ being the distance between two 2D points and with $I_i = \check{I}_i$ and $I'_i = \check{I}'_i$, the *VRE* is defined by:

$$E = \frac{1}{k}\left(\sum_{i=1}^{k} d\left(I_i, \hat{I}_i\right)^2 + d\left(I'_i, \hat{I}'_i\right)^2\right) \tag{5.20}$$

Once again, a square function is used to penalize larger deviations more severely.

5.5 Computation of Refractive Parameters

In the previous sections, it has been assumed that the refractive parameters are either already known or that hypothetical ones exist. In this section, it will be shown how the approaches from the previous sections can be combined for the computation of the unknown refractive parameters of a *SFRS*. The refractive parameters in question are the orientation of the system axis and air layer thickness. The proposed calibration strategies were developed with regard to the conditions described in Section 5.1. Summarized, these comprise calibrated intrinsic and extrinsic camera parameters of the stereo system, known refractive indices of all participating media and a known glass layer thickness. A calibrated stereo system enables the computation of *VOP*s by the approach described in Section 5.3.1. These *VOP*s are the essential quantities for the proposed calibration strategies. In the following sections, optimization problems will be presented that will be used to regulate the interaction of parameter determination and cost functions.
With the variety of different cost functions, the calibration can be achieved by two different strategies. The first one is the consecutive determination of refractive parameters and the second one is the simultaneous one. Both contain a system axis determination by iterative optimization, which is performed either explicitly of implicitly. The defined pattern-based cost functions can be utilized to determine the system axis explicitly and also independently of the air layer thickness. The air layer thickness itself can not be computed individually. However, it can be computed directly by solving a set of linear equations for a determined system axis. A major difference between the two calibration strategies is that the consecutive one depends on a calibration object and the simultaneous one does not. This section is based on Dolereit and von Lukas [Dol15; DL16].

5.5.1 Pattern-Based System Axis Determination

The pattern-based system axis determination makes use of the pattern-based invariants and cost functions from Sections 5.4.1 and 5.4.2. Since these are independent of the remaining refractive parameters, the orientation of the system axis can be optimized individually. The optimization problem that needs to be solved can be defined as a search for the best system axis by testing hypothetical system axes. This is realized by a self-directed testing strategy. One of the two proposed pattern-based cost functions needs to be used to indicate the accuracy of the tested hypothetical axes. Therefore, a calibration object is needed with a pattern that contains feature points arranged in straight lines. This can be a checker pattern with its squares arranged in $i \in \{1, .., m\}$ rows and $j \in \{1, .., n\}$ columns. The optimization is about finding the hypothetical axis with the minimal function value of the chosen cost function. A vital part is the iterative generation of new hypothetical axes. This generation is self-directed. As will be shown in the following, it is driven by the accuracy of the currently tested candidate axis.

Generation of Hypothetical System Axes. The first steps for the generation of hypothetical system axes are basically the same as for the computation of the *PAE* and the *PFE* in Section 5.4.2. Once again, the image points I_{ij} and I'_{ij} form the starting

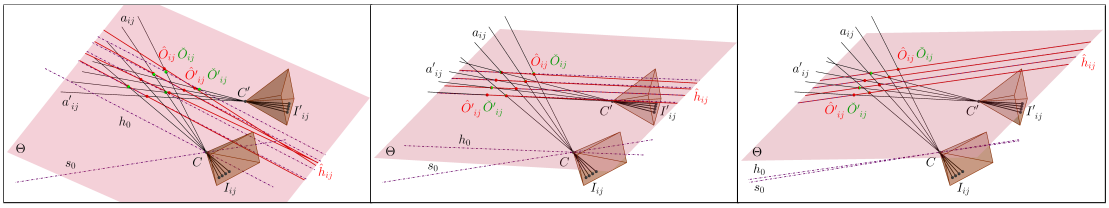

Figure 5.19: Pattern-based system axis determination. Left: Initialization of the iterative processing and generation of hypothetical system axes h_{ij}. Middle: Intermediate result. Right: Result close to the ideal solution.

point of the computations. The generation of hypothetical system axes is illustrated exemplary on the left side of Figure 5.19 and comprises the following steps:

- Compute the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ from pairs of corresponding image points I_{ij} and I'_{ij} for a hypothetical system axis h_0, as described in Section 5.3.1.
- Arrange these *VOP*s according to the lines in horizontal (l_i) and vertical (l_j) direction on the checker pattern.
- Fit a plane to the set of *VOP*s per line to get the feature planes in horizontal direction $\Theta_i : \bar{\boldsymbol{n}}_i \cdot \boldsymbol{x} = d_i$ and the feature planes in vertical direction $\Theta_j : \bar{\boldsymbol{n}}_j \cdot \boldsymbol{x} = d_j$, with $\bar{\boldsymbol{n}}_i$ and $\bar{\boldsymbol{n}}_j$ being the normalized normal vectors, $\boldsymbol{x}$ being the position vector to a point on the plane and d_i as well as d_j being the distances of the respective plane to the origin of the coordinate system.

For an erroneous hypothetical system axis h_0, these fitted planes are most likely to deviate from their associated *VOP*s. This can be utilized for the computation of new hypothetical axes as follows:

- The intersection of the fitted planes Θ_i and Θ_j with the corresponding rays a_{ij} and a'_{ij} (the same intersection process as described in Section 5.3.1) results in the position vectors of the points of intersection $\hat{O}_{ij}$ and $\hat{O}'_{ij}$.
- Paired by matching indices, the points $\hat{O}_{ij}$ and $\hat{O}'_{ij}$ form the hypothetical object axes $\hat{h}_{ij}$ with the normalized direction vectors $\bar{\boldsymbol{h}}_{ij}$.

The hypothetical object axes $\hat{h}_{ij}$ with the normalized direction vectors $\bar{\boldsymbol{h}}_{ij}$ represent the generated hypothetical system axes. The direction vectors computed in this way differ as long as the *VOP*s $\check{O}_{ij}$ and $\check{O}'_{ij}$ do not coincide with the points $\hat{O}_{ij}$ and $\hat{O}'_{ij}$ ideally.

Initialization. The initialization phase is illustrated exemplary on the left side of Figure 5.19 and comprises the following definitions and initial computations:

- Choose the initial hypothetical system axis h_0 to be the one with a direction vector $\bar{\boldsymbol{h}} = (0, 0, 1)^T$.

- Compute the value E of the PAE or PFE for system axis h_0 as described in Section 5.4.2 and initialize the current global minimal error value E_{gmin} with: $E_{gmin} = E$.

- Generate the new hypothetical system axes $\hat{h}_{ij}$ for system axis h_0 as described previously.

- Define the termination conditions:
  - Provide a stability counter that terminates the iterative processing if there is no improvement for n iterations.
  - Provide a significance counter that terminates the iterative processing if there is no significant amount of improvement ω, compared to a threshold ε, for n iterations.

Iteration. The iteration phase is illustrated exemplary in the middle and on the right side of Figure 5.19 and comprises the following steps:

1) Compute the error values E_{ij} for all new hypothetical system axes $\hat{h}_{ij}$ as described in Section 5.4.2.

2) Find the local minimum $E_{lmin} = E_{ij}$ of this iteration and set $h_0 = \hat{h}_{ij}$ accordingly.

3) If a new global minimum is found ($E_{lmin} < E_{gmin}$),

3.1) update the current global minimum ($E_{gmin} = E_{lmin}$),

3.2) reset the stability counter and

3.3) compute the improvement ($\omega = E_{gmin} - E_{lmin}$).

3.4) If the improvement is smaller than the significance threshold ($\omega < \varepsilon$), increase the significance counter.

3.5) If the improvement is equal or bigger than the significance threshold ($\omega \geq \varepsilon$), reset the significance counter.

4) If no new global minimum is found ($E_{lmin} \geq E_{min}$), increase the stability counter.

5) If the stability or the significance counter is exceeded, terminate.

6) Generate the new hypothetical system axes for h_0 and repeat from **1)**.

On the right side of Figure 5.19, a result of one iteration is illustrated. It comes close to the ideal result. The hypothetical system axis h_0 almost coincides with the true system axis s_0 and the VOPs $\check{O}_{ij}$ and $\check{O}'_{ij}$ almost coincide with the points $\hat{O}_{ij}$ and $\hat{O}'_{ij}$. The final result of this self-directed testing strategy is the best fitting system axis.

5.5.2 Computation of the Air Layer Thickness

The proposed approach for the computation of the air layer thickness makes use of the *VOP* model from Section 4.3. In contrast to the pattern-based system axis determination, no iterative processing is necessary, since the optimization problem that needs to be solved can expressed by a overdetermined set of linear equations. With the aid of the *VOP* model, these linear equations can be derived for each pair of corresponding image points and their solution can be realized by standard methods of linear optimization.

Requirements. The required quantities are the refractive indices n_a, n_w, n_g, the glass layer thickness $t_g = d(H_1, H_2)$ and the direction vector $\bar{\boldsymbol{h}}$ of a hypothetical system axis h_0.

Derivation of the Set of Linear Equations. The computation of the air layer thickness $t_a = d(C, H_1)$ is illustrated exemplary in Figure 5.20. It can not be determined independently of the system axis. The starting point of the computations are pairs of corresponding image points I_i and I_i', with $i \in (1, .., k))$. The first necessary step comprises the computation of the *VOP*s. Summarized, these are:

- Compute the *VOP*s $\check{O}_i$ and $\check{O}_i'$ from pairs of corresponding image points I_i and I_i' for a hypothetical system axis h_0 as described in Section 5.3.1. Note that this includes the computation of the angles of incidence α_i and α_i'.
- Compute the distances $\check{x}_i = d(A_{1i}, \check{O}_i) = d(C, A_{2i})$ by orthogonal projection of the position vector $\boldsymbol{o}$ of *VOP* $\check{O}_i$ onto vector $\bar{\boldsymbol{h}}$ (See Equation 2.7).
- Compute the distances $\check{x}_i' = d(A_{1i}, \check{O}_i') = d(C, A_{3i})$ by orthogonal projection of the position vector $\boldsymbol{o}'$ of *VOP* $\check{O}_i'$ onto vector $\bar{\boldsymbol{h}}$ (See Equation 2.7).

The set of linear equations can be derived by means of Equation 4.7, which reads:

$$x = \lambda_1 \left(\alpha, n_a, n_w\right) \cdot \left(\check{x} + t_g \cdot \left(1 - \lambda_2 \left(\alpha, n_a, n_g\right)\right)\right),$$

with:

$$\lambda_1(\alpha, n_a, n_w) = \frac{\sqrt{\left(\frac{n_w}{n_a}\right)^2 - \sin^2\alpha}}{\cos\alpha}, \quad \lambda_2(\alpha, n_a, n_g) = \frac{\cos\alpha}{\sqrt{\left(\frac{n_g}{n_a}\right)^2 - \sin^2\alpha}}.$$

Together with $t = t_a + t_g$, the respective distances to the real object points $x_i = d(A_{4i}, O_i)$ and $x_i' = d(A_{4i}, O_i')$ can be computed by (See Figure 5.20):

$$x_i = \lambda_1(\alpha_i, n_a, n_w) \cdot (\check{x}_i - t + t_g \cdot (1 - \lambda_2(\alpha_i, n_a, n_g))) \tag{5.21}$$

$$x_i' = \lambda_1(\alpha_i', n_a, n_w) \cdot (\check{x}_i' - t + t_g \cdot (1 - \lambda_2(\alpha_i', n_a, n_g))). \tag{5.22}$$

The distances $x_i = d(A_{4i}, O_i)$ and $x_i' = d(A_{4i}, O_i')$ have to be equal, since the points O_i and O_i' are expected to coincide. Therefore, Equation 5.21 and Equation 5.22, with some

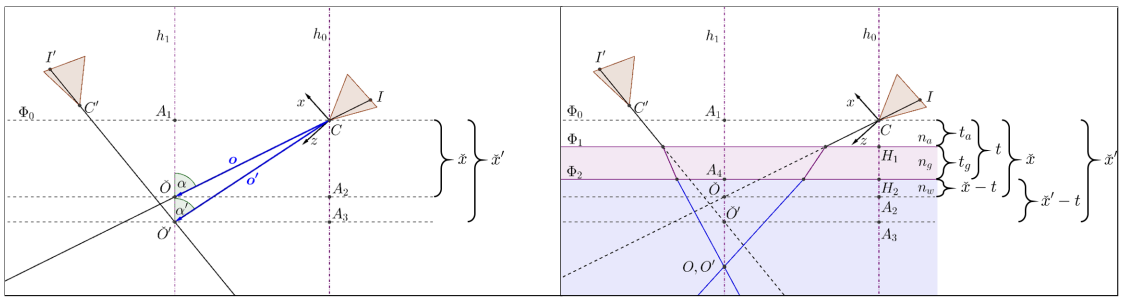

Figure 5.20: Schematic illustration of the derivation of the linear equations for the computation of the air layer thickness.

additional rearranging (note that the expressions for λ_1 and λ_2 are shortened due to a lack of space), produce:

$$u_i - \lambda_1(\alpha_i, n_a, n_w) \cdot t = v_i - \lambda_1(\alpha_i', n_a, n_w) \cdot t, \tag{5.23}$$

with:

$$u_i = \lambda_1\left(\alpha_i, n_a, n_w\right) \cdot \left(\check{x}_i + t_g \cdot \left(1 - \lambda_2\left(\alpha_i, n_a, n_g\right)\right)\right),$$
$$v_i = \lambda_1\left(\alpha_i', n_a, n_w\right) \cdot \left(\check{x}_i' + t_g \cdot \left(1 - \lambda_2\left(\alpha_i', n_a, n_g\right)\right)\right).$$

Some further rearranging of Equation 5.23 results in the following equation for every stereo correspondence:

$$\underbrace{\left(\lambda_1(\alpha_i, n_a, n_w) - \lambda_1(\alpha_i', n_a, n_w)\right)}_{\boldsymbol{A}} \cdot \underbrace{t}_{\boldsymbol{x}} = \underbrace{v_i - u_i}_{\boldsymbol{b}}. \tag{5.24}$$

The application of Equation 5.24 on every stereo correspondence results in a system of linear equations with the shape $\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b}$, with $\boldsymbol{A}$ being a $i \times 1$ matrix containing the elements $(\lambda_1(\alpha_i, ..) - \lambda_1(\alpha_i', ..))$ and $\boldsymbol{b}$ being a $i \times 1$ vector containing the elements $v_i - u_i$. A possibility to solve this overdetermined linear system for $\boldsymbol{x}$, which is a 1×1 vector with t as its only element, is singular value decomposition [BS79]. The result is the best fitting value for t. By subtraction of the known glass layer thickness t_g, the final value of the air layer thickness t_a for a hypothetical system axis h_0 can be determined straightforward: $t_a = t - t_g$. Note that the computation of the air layer thickness does only depend on stereo correspondences. Hence, no pattern on a calibration object, nor the calibration object itself are necessary.

5.5.3 Execution Sequence

The proposed pattern-based system axis determination and the computation of the air layer thickness enable the determination of the refractive parameters in question. Together with the proposed approaches from Sections 5.2, 5.3 and 5.4, all the necessary tools are available to realize the refractive calibration of a *SFRS* by two different strategies. The first one is the consecutive determination and the second one is the simultaneous determination of the refractive parameters.

Consecutive Determination of Refractive Parameters

The consecutive calibration strategy makes use of a pattern on a calibration object and the necessary computation steps are the following:

1) Perform the pattern-based system axis determination from Section 5.5.1 to get a single system axis.

2) Compute the air layer thickness for this system axis as described in Section 5.5.2.

Note that with these steps, the refractive parameters in question can already be determined. However, a calibration object that matches the requirements is needed for step one of this strategy. Furthermore, the resulting system axis is fixed during the computation of the air layer thickness. This has the disadvantage that inaccuracies of the system axis determination can be incorporated into the computation of the air layer thickness. An alternative is the simultaneous determination of the refractive parameters.

Simultaneous Determination of Refractive Parameters

The simultaneous calibration strategy is independent of calibration objects. Therefore, the explicit pattern-based determination of the system axis needs to be bypassed. The essential solution of the optimization problem, which is a search for the best system axis by testing hypothetical system axes, can be retained. However, an approach for the generation of these hypothetical system axes is necessary. Since the approach for the generation of hypothetical system axes from the self-directed testing strategy of the pattern-based system axis determination is not applicable, an alternative is necessary.

Generation of Hypothetical System Axes. The alternative approach is illustrated in Figure 5.21. It is a simple sampling of the possible direction vectors. Therefore, a normalized direction vector $\bar{\boldsymbol{h}}$ is represented with the aid of a unit sphere and the two angles θ and ϕ. This representation is shown on the left side of the figure. It is similar to spherical coordinates and well suited for the intended sampling. The depicted coordinate system is the one of the master camera. The sampling space is located only on the positive z-axis and is bounded by a defined range of the two angles θ and ϕ. In this thesis, the range was defined to be between -45° and $+45^\circ$ for both angles. This will be justified in Section 7.1.1. The generation of hypothetical system axes is an iterative sampling, which starts on a sparse grid in full 1° steps. After determination of the most appropriate direction of the system axis on this grid level, the search space gets reduced and the grid gets refined. The new search space surrounds the determined spherical coordinates at a range between -1° and $+1^\circ$ for both angles and the grid is refined into 0.1° steps to determine the next decimal. This iterative process is continued up to the desired decimal.

Initialization. The simultaneous determination of the refractive parameters makes use of the cost functions proposed in Sections 5.4.3 and 5.4.4. All four of them indicate the

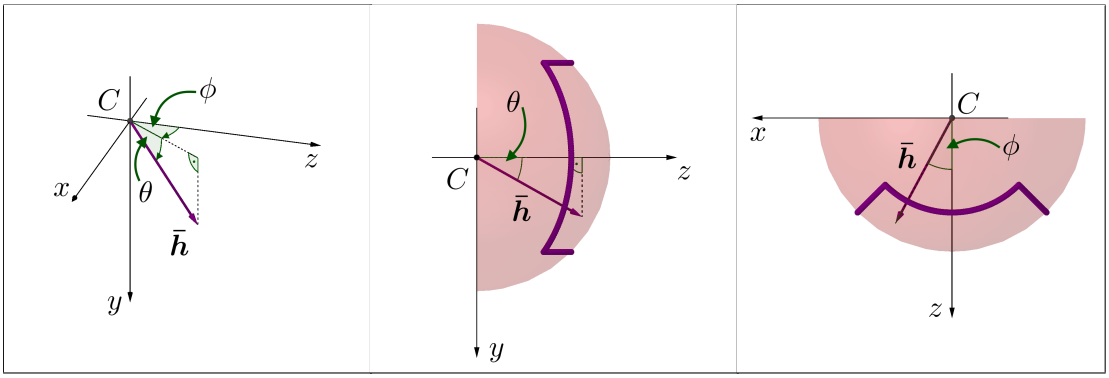

Figure 5.21: Search space for the system axis, which is bounded by a defined range of the spherical coordinates (purple section), with respect to the camera coordinate system. Left: Utilized spherical coordinates. Middle: Side view. Right: Top view.

accuracy of the determined air layer thickness and system axis. Therefore, the necessary initialization steps are:

- Choose one of the four cost functions RTE, VIE, RRE or VRE.
- Initialize the global minimal function value E with infinity.
- Generate a set of n hypothetical system axes on the initial grid level by the proposed sampling strategy.

Iteration. The iteration phase comprises the following steps:

1) Compute the air layer thickness for each of the n hypothetical system axes by the linear optimization approach described in Section 5.5.2.

2) Compute the function value E of the chosen cost function for each of the n pairs of air layer thickness and system axis.

3) Find the minimal function value E. If a new global minimum is found ($E < E_{min}$), set the new minimum ($E_{min} = E$).

4) Refine the search space for the generation of new hypothetical axes around the coordinates that correspond to the current global minimum E_{min} and repeat the process on the next grid level up to the desired decimal.

Since the system axis is chosen according to the best air layer thickness in each iteration, both of them get optimized simultaneously. In contrast to the consecutive strategy, possible inaccuracies during of the parameter determination are distributed to the system axis and the air layer thickness. Both strategies will be compared in detail in Chapter 7.

5.6 Summary

In this chapter, the proposed approaches for the calibration of a *SFRS* have been presented. The refractive parameters comprise the orientation of the system axis and air layer thickness. The approaches work, if the following conditions are met: The setup can be modeled by a *SFRS* and the intrinsic and extrinsic camera parameters, the refractive indices of all participating media and the glass layer thickness are known. The proposed approaches have been distinguished by two underlying strategies. On the one hand, this is the consecutive determination of the refractive parameters and on the other hand, the simultaneous one. A vital part of both of them are cost functions that reflect the accuracy of the determined parameters and the definition of optimization problems that regulate the interaction between parameter determination and cost functions.
The basic operations that are necessary for the computation of the function values of cost functions are transformations between 3D and 2D spaces, which are usually the computationally expensive refractive forward projection and refractive back-projection. It has been demonstrated that the *VOP* model from Section 4.3 can be readily integrated into the model for a *SFRS*, as well as into the more general model for a *FRS*. With this, four possibilities for the transformation between 3D and 2D spaces have been proposed. By using the *VOP* model within monocular vision, a refractive forward projection avoiding higher-order polynomials and a refractive back-projection avoiding full 3D ray tracing have been realized, thereby reducing the computational expenses. The third one is the definition of a virtual camera model to avoid refractive forward projection at all and hence to enable conventional forward projection. As has been demonstrated, this is especially useful in the context of the computation of the reprojection error. The fourth one has been arising from using the *VOP* model within binocular vision. It represents an efficient alternative to refractive back-projection. This is the computation of the *VOP*s from stereo correspondences, which in turn enables the computation of the related real object points.
The transformations between 3D and 2D spaces have been utilized for the computation of four different cost functions. The first category makes use of the transformation direction from 2D to 3D. Its first representative is an underwater adaptation of the triangulation error, named *RTE*, which is based on refractive back-projection. The second representative is a newly defined cost function, named *VIE*, which is based on the computation of the *VOP*s. It is expected to be the most computationally efficient one of all four cost functions. While the processing in the first category stops at the recovery of 3D coordinates, the second category goes one step further and makes use of the transformation direction from 3D to 2D. The two proposed representatives are underwater adaptations of the common reprojection error, named *RRE* and *VRE*. They are based on refractive forward projection and the virtual camera model, respectively. All four cost functions of both categories do neither depend on a known pattern on a calibration object, nor on a calibration object at all. In contrast to these pattern-independent cost functions, the third category comprises newly developed pattern-based cost functions that are independent of refractive indices and layer thicknesses. For their definition, pattern-based invariants have been derived by exploiting feature points that

are arranged in straight lines on a calibration object. The two representatives, named *PAE* and *PFE*, are based on the computation of the *VOP*s.
All these cost functions have been either utilized during the consecutive strategy for the determination of the refractive parameters, or during the simultaneous one. Both strategies have in common that the air layer thickness is optimized by solving a set of linear equations, which is one of the main developments of this thesis. However, with the proposed developments, an optimization problem that can be solved linearly could not be defined for system axis determination. Therefore, during both strategies, the system axis needs to be optimized iteratively by utilization of the proposed cost functions. Since the cost functions show a progressive independence from calibration objects, both strategies do so as well. The consecutive determination of the refractive parameters needs a known calibration object and utilizes the pattern-based cost functions. Therefore, the pattern-based system axis determination is performed explicitly. The simultaneous strategy needs no calibration object at all. It comprises an implicit system axis determination and therefore utilizes the cost functions from the first two categories. The differentiations that are reflected by these various approaches for the refractive calibration of a *SFRS*, enable the following comparisons:

- Implicit (simultaneous strategy) versus explicit system axis determination (consecutive strategy) and simultaneously,
- pattern-based cost functions (with calibration object) versus pattern-independent cost functions (without calibration object).
- 3D scene cost functions versus 2D image cost functions in the case of pattern-independent cost functions.

Subject of the following chapters will be a detailed analysis of the proposed calibration approaches. Furthermore, further applications of the proposed transformations between 3D and 2D spaces will be presented briefly. These are beneficial approaches, such as the computation of correspondence curves and the recovery of 3D coordinates of scene points.

6 Underwater Stereo 3D Reconstruction

The purpose of stereo 3D reconstruction is to generate a 3D model of the imaged scene. In the simplest case, this model is a 3D point cloud. Besides calibration of the imaging system, the necessary steps of stereo 3D reconstructions comprise some kind of stereo matching and the recovery of 3D coordinates. A pursuing step for the generation of a more sophisticated model would be to represent surfaces by triangulated meshes. However, the actual recovery of 3D coordinates from pairs of corresponding image points in the left an right view of the stereo camera is most important.
Similarly to the approaches for the calibration of a *SFRS* in the previous chapter, stereo matching and the recovery of 3D coordinates need to be adapted to the prevailing conditions of underwater imaging. Stereo matching can be distinguished into sparse and dense stereo matching. Sparse stereo matching is usually feature-based and therefore does not get affected by refractive effects. On the contrary, dense stereo matching is usually geometry-based and therefore does get affected severely by refractive effects. In this chapter, it will be shown how to compute correspondence curves for the reduction of the search space for pairs of corresponding image points during dense stereo matching. Their computation is based on the proposed transformations between 3D and 2D spaces from the previous chapter. Furthermore, it will be shown how 3D coordinates of points in the scene can be recovered with explicit consideration of refractive effects. Some of the various approaches of the previous chapter already came into contact with recovering 3D coordinates, but this has been done mostly implicitly. In this chapter, it will be shown how to perform it explicitly, based on the earlier proposed transformations between 3D and 2D spaces. This chapter is partially based on the concepts of Dolereit [Dol15].

6.1 Correspondence Curves

Up to now it has been assumed that a set of pairwise corresponding feature points in the two views of the *SFRS* is already established. Therefore, easily detectable and matchable checker patterns can be used. Conventional algorithms are capable to detect the checker corners at sub-pixel level. However, when it comes to full stereo 3D reconstruction of an imaged scene, usually a matching at pixel level is performed between the left and the right view. This one-to-one matching process is called dense stereo matching. In air, epipolar geometry and rectification are a means for dense stereo matching to find the pairs of corresponding pixels. The main purpose is the reduction of the search space to a straight line. However, both of them are invalid under water. It is known that, due to the 3D-dependence of refractive effects, these lines are not straight any more, with one exception (Compare with Section 2.4 and Figure 6.1). Note that the line connecting the left and right centers of projection C and C' and the system axis (Here: Coinciding with the left and right principal axes p and p') form a plane. Therefore, the only exception is the case where the left ray path lies in this plane. In this case, the left and the right plane of refraction would coincide with this plane. Hence, the right ray path needs to lie in this shared plane as well. The resulting search space in the second view is a line that can only be straight. Regarding Figure 6.1, this would mean that all the dot-dashed lines running horizontally coincide.
For the remaining cases, the search spaces are not straight, but curved lines. The computation of these correspondence curves, based on the concepts of Dolereit [Dol15], can be realized by utilizing the transformations between 3D and 2D spaces, which have been presented in the previous chapter. Therefore, the imaging system needs to be calibrated for its intrinsic, extrinsic and refractive parameters. These parameters can either be known in advance, or a calibration needs to be performed explicitly. For example, by the approach proposed in this thesis:

- Compute the refractive parameters by application of the calibration of a *SFRS*, as proposed in Section 5.5.

After calibration of the imaging system, a sampling strategy is performed. This is illustrated exemplary in Figure 6.2. The following steps are necessary:

- Compute the ray w for the master camera by application of the refractive back-projection from Section 5.2.2.
- Specify a working range and step width to sample the ray w at the points O_i.
- Map the points O_i onto the image plane of the slave camera by application of the refractive forward projection from Section 5.2.1.

Since the working distance is usually known, the necessary range in which the ray w will be sampled can be determined by the user. It starts at the water-sided interface Φ_2 and its maximal value is arbitrary. Since the underwater visibility is mostly limited to a few meters, this user-defined value is appropriate. The step width can be user-defined as

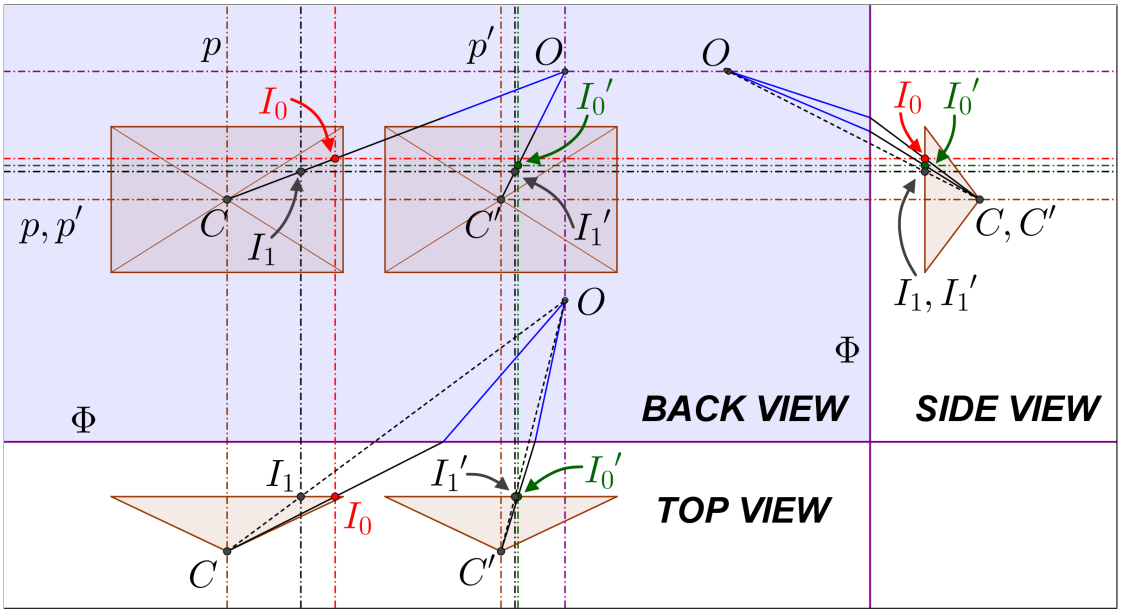

Figure 6.1: Invalidity of epipolar geometry and rectification (Compare with Section 2.4).

well. While a fine step width most likely already results in a connected curve, a coarse step width may need some interpolation between the image points I_i'.
Dense stereo matching is associated with a high computational effort. The proposed approach is computationally expensive, since one correspondence curve needs to be computed per image point I. Note that the approach is not limited to SFRS*s*. It can be applied for a stereo system composed of two *FRS*s in arbitrary orientation, which is calibrated for its intrinsic, extrinsic and refractive parameters in advance, as well.

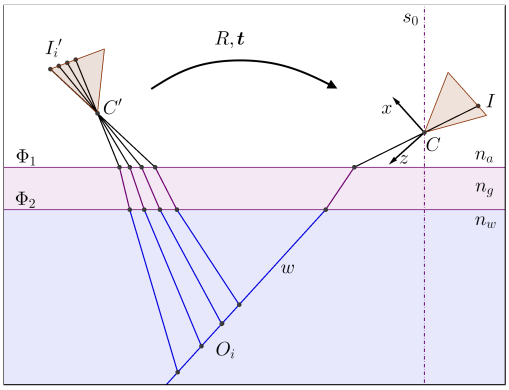

Figure 6.2: Sampling of the ray w to compute a correspondence curve.

6.2 Recovery of 3D Coordinates

Since epipolar geometry and disparity are not valid concepts in underwater environments, as has been shown in Section 2.4, they can not be utilized for underwater stereo 3D reconstruction. Hence, the recovery of the 3D coordinates associated to a pair of corresponding image points has to be realized in an alternative way.
As a prerequisite, some kind of stereo matching has to be performed to establish the pairs of corresponding image points. This can be realized at pixel or sub-pixel level. Pixel level usually means geometry-based dense stereo matching, as described in the previous section. For every pixel in one stereo view, a correspondence curve in the second view needs to be computed. The matching itself can be realized, for example, by any kind of algorithm for the comparison of image intensity values. Sub-pixel level usually means feature-based sparse stereo matching. Therefore, image features need to be detected, described and matched by a defined metric. Sparse stereo matching is only dependent on the performance of these three steps. The computation of correspondence curves is not essential to it and it therefore is not affected by refractive effects. After stereo matching, the 3D coordinates are ready to be recovered.
The proposed two approaches have already been used implicitly in Chapter 5 as part of refractive calibration. Nevertheless, both of them can also be utilized for explicit recovery of 3D coordinates. Due to possible inaccuracies of the determined intrinsic, extrinsic and/or refractive parameters of the imaging system, it is most likely that there is no ideal recovery. Therefore, a mechanism to deal with these small deviations is necessary for both of the following approaches.

Virtual Intersection. The first one is named *Virtual Intersection* and comprises the following steps (See Figure 6.3):

- Compute the *VOP*s $\check{O}$ and $\check{O}'$ as described in Section 5.3.1 followed by the computation of the related real object points O and O' from Section 5.3.2.
- Since O and O' may not coincide ideally, compute their mean $\hat{O}$ on object axis s_1.

Since this approach depends on the line of intersection of a pair of corresponding planes of refraction, the shortcoming of this approach is the inability of the necessary computations in the case of coinciding planes of refraction. This can occur when the left and right centers of projection, the system axis and the left and right ray paths coincide in a single plane. Besides this case of an ideal coincidence, cases close to coincidence should be avoided as well, as they can result in inaccuracies.

Refractive Triangulation. The second approach for the recovery of 3D coordinates is named *Refractive Triangulation* and does not suffer from the aforementioned shortcomings. It comprises the following steps (See Figure 6.4):

- Compute the rays w and w' by application of refractive back-projection from Section 5.2.2 and transform one of them to the coordinate system of the opposite camera with the known transformation $\boldsymbol{R}, \boldsymbol{t}$ between both camera coordinate systems.

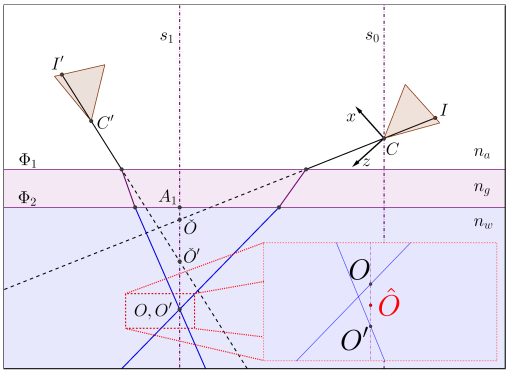

Figure 6.3: Recovery of 3D coordinates by *Virtual Intersection*.

- Compute the point of intersection O by conventional triangulation of the rays w and w' (Section 2.2.2). Since the rays w and w' may not coincide at all, the midpoint $\hat{O}$ of the shortest distance $d(O, O')$ between both rays is considered to be the final point of the recovery.

Both approaches have their right to exist. As has been shown in the previous chapter, *Virtual Intersection* is the foundation of the determination of the air layer thickness by linear optimization and is hence utilized in each of the proposed approaches for the refractive calibration of a *SFRS*. Its computational effort is low and the condition of coinciding planes of refraction is less likely to occur for a set of sparse pairs of corresponding image points. If it nevertheless occurs, the processing can be skipped without any problems, since not all pairs are required for refractive calibration. Note that *Virtual Intersection* is only applicable for *SFRS*s. If it is necessary to recover the 3D coordinates of all pairs of corresponding feature points, *Refractive Triangulation* is better suited. Furthermore, *Refractive Triangulation* is not limited to *SFRS*s.

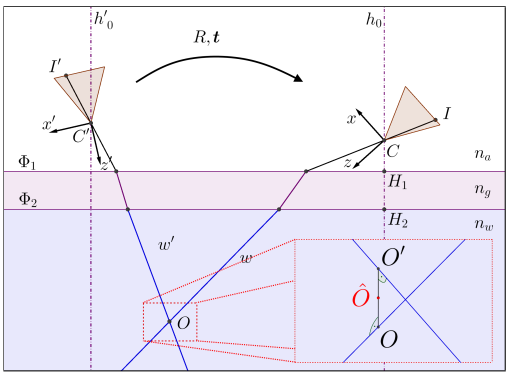

Figure 6.4: Recovery of 3D coordinates by *Refractive Triangulation*

6.3 Summary

In this chapter, beneficial applications of the proposed transformations between 3D and 2D spaces from Chapter 5 have been presented in the context of underwater stereo 3D reconstruction. These are the computation of correspondence curves and the recovery of 3D coordinates. Together with the earlier proposed approaches for refractive calibration, all the necessary extensions of the concepts of stereo 3D reconstruction for the prevailing conditions of underwater imaging have been covered. All of them are essentially based on the *VOP* model from Section 4.3. Therefore, the underlying principles of the necessary computation steps can be reused frequently between refractive calibration, correspondence curve computation and recovery of 3D coordinates. The final result are 3D point clouds of the imaged scene with a metric scale, which are created by explicitly considering refractive effects in every computation step. Therefore, refractive calibration is a prerequisite for correspondence curve computation and recovery of 3D coordinates. In the context of dense stereo matching, the computation of correspondence curves represents an alternative to the invalid epipolar geometry and rectification. However, it is computationally expensive. The two proposed approaches for the recovery of 3D coordinates from pairs of corresponding image points in the left and right view of the stereo camera have been connected to their most suitable application area. Due to its low computational effort, *Virtual Intersection* is suitable within the refractive calibration approaches proposed previously. Since the *Virtual Intersection* shows some shortcomings, *Refractive Triangulation* is better suited for recovering 3D coordinates of a full scene. In the following chapter, all the proposed approaches for underwater stereo 3D reconstruction will be analyzed in a series of experiments.

7 Experiments & Results

In this chapter, the results of the performed experiments will be presented. These experiments will be performed on simulated and real test data. Their purpose will be the evaluation of the proposed two strategies, namely the consecutive and the simultaneous determination of refractive parameters, and of all six cost functions (PAE, PFE, RTE, VIE, RRE and VRE). Besides a comparison amongst themselves, a comparison to a current state-of-the-art approach for refractive calibration and to further calibration strategies from the literature will be preformed.
The evaluation will be realized in multiple stages. *Simulated Test Data Stage 1* will comprise a full simulation of 3D-2D point correspondences. Thereby, ground truth data is available for the refractive parameters and for the 3D point cloud to be recovered. Since ideal pinhole cameras with known parameters will be used, an augmentation of the 2D feature points with Gaussian noise will be performed to evaluate the stability of the proposed approaches. Therefore, evaluation criteria will be defined that are suitable for the available ground truth data. Within the evaluation, a variation of the requested refractive parameters will be performed. This stage will show if an approach works and how well it works in the presence of noise. *Simulated Test Data Stage 2* will comprise rendered images of a 3D scene. The processed ground truth data will be reduced to set up a workflow that can be adapted to real world setups. Therefore, additional evaluation criteria will be defined that are suitable for the available ground truth data. In contrast to the previous stage, the calibration of the stereo camera and the feature detection will be performed algorithmically by common approaches as a pre-processing step of refractive calibration. Besides establishing a workflow, this stage will be used to evaluate if refractive calibration in combination with the pre-processing step works.
Real Test Data Stage 1 will comprise a workflow and a setup for evaluation, which are essentially the same as within *Simulated Test Data Stage 2*. Within this stage, an approach for benchmark data generation in real world setups, based on Dolereit [Dol15], will be presented. This stage will be used to evaluate the accumulation of errors of the pre-process comprising stereo calibration and feature detection and of the process of refractive calibration. *Real Test Data Stage 2* will differ from the previous stages by avoiding known calibration objects.
The last part of this chapter will present the results from first experiments on underwater stereo 3D reconstruction by application of the proposed approaches for refractive calibration. It will comprise a visual examination of the proposed approach for the computation of correspondence curves and a qualitative evaluation of the proposed approaches for dense stereo matching and the recovery of 3D coordinates of a real submerged object.

7.1 Refractive Calibration - Simulated Test Data

Simulated test data is especially advantageous, since it makes ground truth data available. It offers the highest level of controlled conditions and arbitrary setups can be realized. A preparatory step of the proposed refractive calibration approaches is the calibration of the stereo camera to determine its intrinsic and extrinsic camera parameters. Since this was not the object of this thesis and as these parameters need to be known in advance, the camera parameters were considered to be constant and ideal. By simulation of the cameras, stereo calibration can be skipped and further potential sources of error, such as radial lens distortion, can be eliminated. This provides a reliable means to evaluate if an approach for refractive calibration is correct in general.
Both of the previously presented strategies for the calibration of a *SFRS*, namely the consecutive and the simultaneous determination of refractive parameters, were evaluated. To ensure the comparability of the evaluations, the utilized feature points had to be the same for both strategies. Consequently, since the consecutive strategy is dependent on feature points on a calibration object that are arranged in straight lines, the corners of a simulated checker pattern were chosen. The checker pattern was generated to result in a format-filling arrangement of the feature points in the stereo image pairs. In this way, the image regions that are affected most obviously by refractive effects were included. It was assumed that erroneous feature coordinates in these regions have the greatest negative influence and should not be ignored in an evaluation.

7.1.1 Stage 1 - Full Simulation of 3D-2D Point Correspondences

Simulated Test Data Stage 1 comprises the addition of Gaussian noise to the simulated data to get closer to reality. Hence, Gaussian noise was added during the first processing step relevant to this thesis. This is the detection of pairs of corresponding feature points in a stereo image pair. A checker pattern can be represented by the 3D point cloud of its inner corners. Within simulation, the ideal 2D coordinates of the feature points were generated from such a 3D point cloud. The noise was added to these ideal coordinates. It was increased gradually to simulate erroneous image coordinates for the feature points.

Preliminaries

Noise. To check the stability of the proposed calibration approaches, Gaussian noise was generated. Gaussian noise has a probability density function equal to that of the normal distribution. In both images of a stereo image pair, the noise was added to the ideal values of the image coordinates of every feature point in x- and y-direction. The parameters were an overall mean of $\mu = \mu_x = \mu_y = 0$ and an overall standard deviation of:

$$\sigma = \sqrt{\sigma_x^2 + \sigma_y^2},\ \sigma_x = \sigma_y = \frac{\sigma}{\sqrt{2}}. \tag{7.1}$$

Both were scaled in pixel units.

Number of Features. In most optimization approaches, a higher number of utilized feature points leads to more accurate results. In this thesis, a certain number of distinct setups were analyzed with regard to a joint number of feature points, which offers a reasonable compromise between run-time and accuracy improvement of the approaches for the refractive calibration of a *SFRS*. The execution of both calibration strategies in combination with the six cost functions for the determination of the number of feature points itself amounted to a time-consuming process. This number is important under several aspects. The following evaluation with simulated data was realized by a variety of repetitions, which were in their entirety very time-consuming. Therefore, a lower number would save time. However, representative results with regard to accuracy were supposed to be achieved during these repetitions. Therefore, a too low number should be avoided. The goal was to find a compromise that keeps the necessary time for evaluation within days and not weeks. Another aspect was that the determined number of feature points should match a feasible number of feature points to be utilized in the following experiments with real data.
Three distinct setups for a *SFRS* were examined. The refractive parameters were determined repeatedly. One stereo image pair with format-filling feature points was generated in every repetition. 100 repetitions were performed for every tested number of feature points. Each repetition included randomly generated Gaussian noise that was added to the feature points. The noise level that was assumed to be appropriate here is the mean value of the applied noise range during the evaluation in the next section. Hence, the value that was applied had a standard deviation of $\sigma = 0.5$. The 3D point cloud, which was used to generate the image feature points, was slightly rotated randomly in every repetition. By doubling the number of tested points, a range from 50 to 1600 feature points was covered.
Based on the results of the full evaluation in Appendix A the chosen number of features was 400. This number has been utilized in the following experiments. These experiments represent a major part of the evaluation and comprise a variation of the refractive parameters in question.

Scene. After the determination of the number of feature points on the checker pattern, a scene containing the 3D point cloud of this checker pattern needed to be simulated. Additionally, this scene had to contain a water-glass-air transition of the light rays. The simulated setup was chosen to resemble the test tank setup of *Real Test Data Stage 1* and *2* (Compare with Section 7.2.2 and 7.2.3). The simulated tank was made of glass, with a wall thickness of $1cm$ and with a refractive index of 1.6. As before, the resulting flat refractive interfaces were plane parallel. The refractive indices of water and air were 1.33 and 1, respectively. The real test tank was limited in its depth. Therefore, the 3D point cloud of the checker pattern was placed to be in general at a distance of about 25cm to the water-sided interface. The two cameras were placed outside of this tank. Both of them had a resolution of 1920×1080 pixels. The focal length was chosen to result in a wide field of view. The according angles of view were about 75° in diagonal direction, about 67° in horizontal direction and about 41° in vertical direction.

Evaluation Criteria. Since ground truth data was available, the following three criteria were utilized for the evaluation with simulated test data:

- Angle Error: The difference between the true system axis and the calibrated system axis in degrees. The angle was measured between the normalized direction vectors of both axes.
- Air Layer Thickness Error: The difference between the true air layer thickness and the calibrated air layer thickness.
- 3D Error: The mean Euclidean distance between the coordinates of the ground truth 3D point cloud and the 3D coordinates that had been recovered after refractive calibration.

The *Refractive Triangulation* from Section 6.2 was utilized to recover the 3D coordinates of the corresponding feature points and therefore needed to be evaluated as well. Provided that the refractive parameters were determined accurately, the evaluation of the *Refractive Triangulation* itself can be realized by utilization of noise level zero. If it works correctly, the recovered 3D coordinates need to be error-free at this noise level. Thereby, this evaluation was performed simultaneously.

Variation of Refractive Parameters

In *Simulated Test Data Stage 1*, the two strategies for the calibration of a *SFRS* in combination with the six presented cost functions were analyzed. Therefore, each cost function was considered for itself.

Angle Error, Air Layer Thickness Error and 3D Error. The experiments regarding the three evaluation criteria were conducted with underlying, simulated 3D point clouds of a checker pattern, which were generated to result in format-filling feature points in a stereo image pair. To form a maximally overlapping field of view at the location of the 3D point cloud, cameras with converging principal axes p and p' were necessary. In Figure 7.1, the area of this location is marked with a blue rectangle. The generated converging stereo camera setup was realized by defining its relative extrinsic parameters accordingly. The system axis s_0 is represented with respect to the coordinate system of the master camera. As can be seen in Figure 7.2 $\boldsymbol{a}$, its normalized direction vector $\bar{\boldsymbol{s}}$ can be represented with the aid of a unit sphere and the two angles ϕ and θ. By definition of $\phi' = -\phi$ and $\theta' = \theta$, with $\{\phi, \theta\}$ representing the orientation of the system axis s_0 with respect to the master camera and $\{\phi', \theta'\}$ representing the orientation of the system axis s_0' with respect to the slave camera, the basic setup for the conducted experiments was generated. In the 2D illustration in Figure 7.1, the angle θ amounts to $\theta = 0^\circ$. To evaluate the proposed approaches for refractive calibration, the refractive parameters were varied. During the conducted experiments, the air layer thickness was always equal for both cameras. The variation of the refractive parameters was realized as follows:

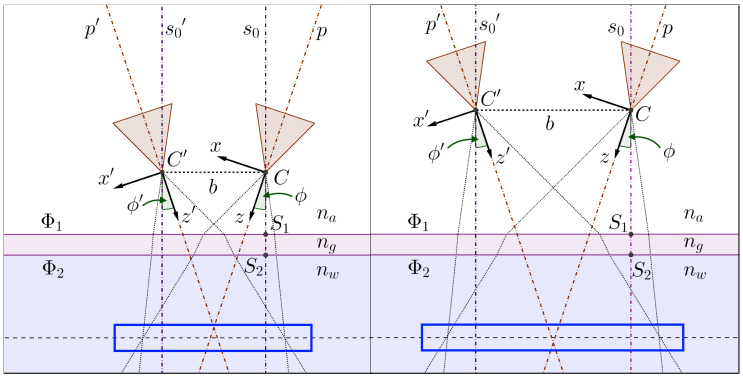

Figure 7.1: Maximally overlapping field of view at the location of a 3D point cloud.

V1) Increase the air layer thickness $t_a = d(C, S_1)$, while the orientation of the system axis s_0 stays the same. The fields of view (dotted black lines) of the two cameras limit a certain range of angles of incidence on the refractive interface Φ_1. To ensure a comparability between the calibration results of the varied setups, this range was attempted to be kept constant, while the air layer thickness was varied between $t_a = 5cm$, $t_a = 15cm$ and $t_a = 25cm$. As can be seen exemplary in Figure 7.1, the consequence was that the angles ϕ and ϕ' needed to be kept fixed, while the stereo baseline b had to be increased to provide a maximally overlapping field of view at the same distance from the interface Φ_2 and therefore, a similar range of angles of incidence.

V2) Vary the orientation of the system axis s_0, while the air layer thickness t_a stays the same. Two cases were part of the evaluation. To get the necessary converging setup in both cases, $\phi = -10^\circ$ and $\phi' = 10^\circ$ were chosen. However, while the angels θ and θ' were equal and amounted to $\theta = \theta' = 0^\circ$ in the first case, they amounted to $\theta = \theta' = -20^\circ$ in the second case.

These variations resulted in six distinct setups. Note that the symmetric setups are only a matter of the design of the chosen evaluation scenario. Such a symmetry is not a requirement for the approaches. A brief analysis of further setups that do not show symmetric properties can be found in Appendix B.
The noise levels ranged from a standard deviation of $\sigma = 0$ to $\sigma = 1$, with a 0.2 step width. The evaluation was realized in 100 iterations per noise level, each with randomly generated Gaussian noise and a slight, random rotation of the underlying 3D point cloud. The resulting values of the three chosen evaluation criteria for the described variation of the refractive parameters are illustrated in a collection of diagrams. Without going into details, this can be seen exemplary in Figure 7.3. Besides an error analysis, these diagrams will show that the single approaches for refractive calibration work in general.

Error Plotting. As described in Chapter 5, the orientation of the system axis needs to be optimized iteratively within the proposed approaches. These approaches are the

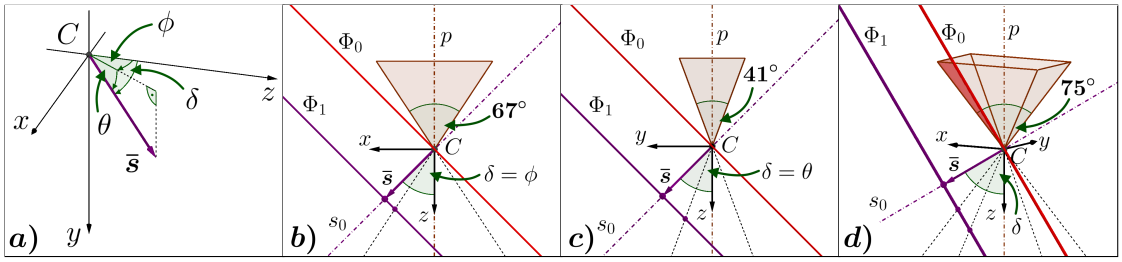

Figure 7.2: ***a*** **:** Representation of the normalized direction vector $\bar{s}$ of a system axis s_0. ***b*** **:** Valid system axes (top view). ***c*** **:** Valid system axis (side view). ***d*** **:** Invalid system axis.

explicit, pattern-based system axis determination with *PAE* or *PFE* from the consecutive strategy and the implicit system axis determination with *RTE*, *VIE*, *RRE* or *VRE* from the simultaneous strategy (See Section 5.5.1). Besides their implicit evaluation within the parameter variation of the previous section, an additional error plotting for a certain search space was chosen as a means for their explicit evaluation. The search space of the iterative optimization was defined by the angles ϕ and θ, which ranged from -45° to $+45^\circ$, respectively. This chosen range can be justified by a brief analysis of invalid system axes. Therefore, let us consider the angle δ between a system axis s_0 and the principal axis p of the camera in Figure 7.2. The field of view of the camera amounted to 75° in diagonal, 67° in horizontal and 41° in vertical direction. Not every possible combination of the field of view and a specific system axis makes sense. In Figure 7.2 ***b*** and ***c*** only one of the angles ϕ and θ is not equal to zero. Hence, the angle δ is identical. If one of those angles reaches the maximum or, equally, the minimum of the range of values, the full field of view still intersects the flat refractive interface Φ_1. However, if both angles reach a maximum or minimum, the angle δ amounts to $+/-60^\circ$. This means that some rays in the field of view can not intersect Φ_1 any more. This is illustrated by the plane Φ_0, which is parallel to Φ_1. While Φ_0 does not intersect the field of view in Figure 7.2 ***b*** and ***c***, it clearly does so in Figure 7.2 ***d***. Therefore, the section of the field of view that will never reach Φ_1 is marked red. To keep the number of these non-intersecting rays small, the chosen range is reasonable.

Any remaining non-intersecting rays can be beneficial during system axis optimization. Therefore, a criterion to mark a system axis to be invalid was defined to be that not all of the rays in both camera's fields of view intersect the flat refractive interface. The reasoning behind this is that rays that are parallel to, or that are pointing away from the refractive interface are usually unintended, since they can not contribute underwater information. One could argue if such a setup could make sense in some special cases. However, in this thesis, these cases were marked as invalid.

The generated error plots (See exemplary Figure 7.4) are surface diagrams with a 1° step width. These diagrams are colored according to the size of the error. Invalid system axes are colored in red. Without going into details, the diagrams can be seen exemplary in Figure 7.4. The error plots are presented for two distinct configurations.

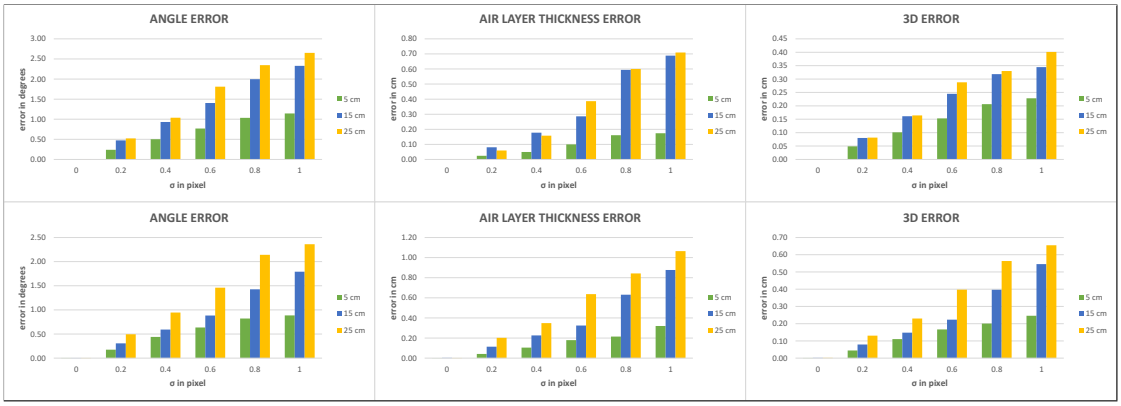

Figure 7.3: Inaccuracies of refractive calibration with PAE. Top row: $\phi = -10^\circ$, $\theta = 0^\circ$. Bottom row: $\phi = -10^\circ$, $\theta = -20^\circ$.

Plane Angles Error (PAE). The first two columns in Figure 7.3 represent the mean differences between the determined refractive parameters and the ground truth. The determined parameters were used for the recovery of the 3D point cloud that was used for refractive calibration. In the last column, the respective inaccuracies of the 3D recovery, which therefore accumulate the inaccuracies of parameter determination, are displayed. The error values almost equal to zero in all three categories at noise level $\sigma = 0$ show that the PAE is valid and that the approach was working correctly. With a few exceptions, the diagrams show very clearly, that an increase of the noise level σ, an increase of the air layer thickness t_a and an increase of the absolute value of the angle θ resulted in increased error values, respectively. Figure 7.4 shows that despite of the coarse sampling grid, the area where the overall minimal error value can be found was unambiguously detectable in both setups and that this area contained the ground truth system axis.

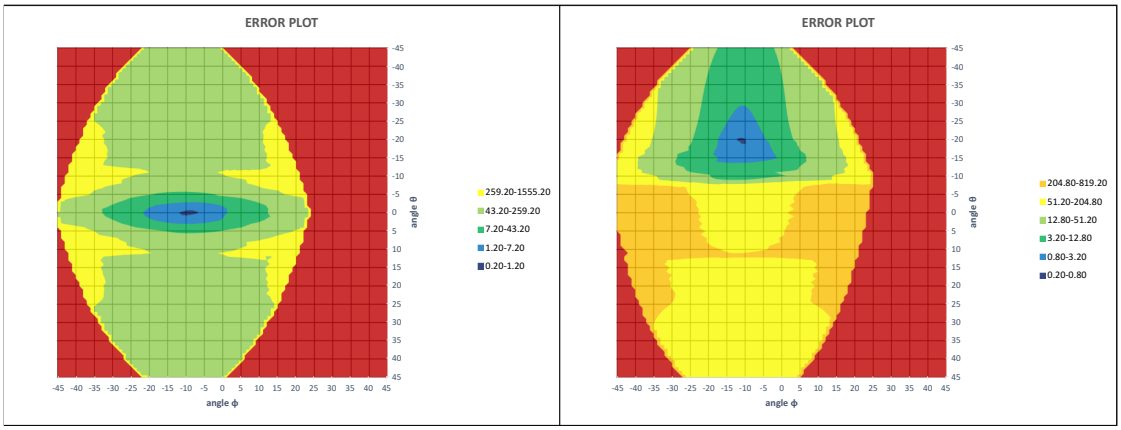

Figure 7.4: Plotting of the PAE. Left: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = 0^\circ$, $\sigma = 1$. Right: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = -20^\circ$, $\sigma = 1$.

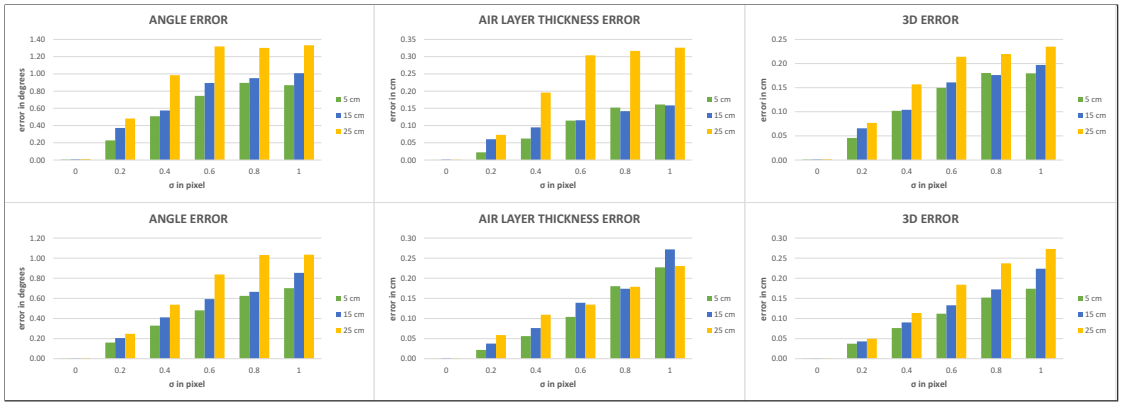

Figure 7.5: Inaccuracies of refractive calibration with *PFE*. Top row: $\phi = -10^\circ$, $\theta = 0^\circ$. Bottom row: $\phi = -10^\circ$, $\theta = -20^\circ$.

Plane Fitting Error (PFE). The error values almost equal to zero in all three categories at noise level $\sigma = 0$ show that the *PFE* is valid and that the approach was working correctly (See Figure 7.5). An increase of the noise level σ led to increased inaccuracies in almost every case. Regarding the error values for the variation of the air layer thicknesses t_a, it became apparent that from noise level $\sigma = 0.6$ on, an increase of the thickness did not necessarily increase them. The approach was the most error prone at the setup with $\phi = -10^\circ$, $\theta = 0^\circ$ and $t_a = 25cm$. However, the accuracy of the recovered 3D point clouds showed no significant anomalies. An increase of the air layer thickness t_a led to an increase of the error values in almost every case. In contrast, an increase of the absolute value of the angle θ did not lead to a consistent increase. Figure 7.6 shows that the area with the overall minimal error value could be detected unambiguously in both setups and that it contained the ground truth system axis.

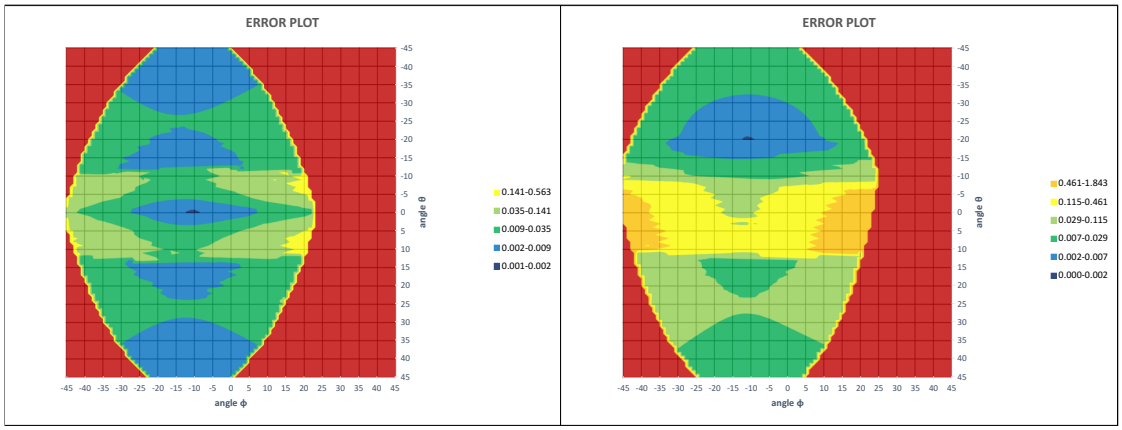

Figure 7.6: Plotting of the *PFE*. Left: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = 0^\circ$, $\sigma = 1$. Right: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = -20^\circ$, $\sigma = 1$.

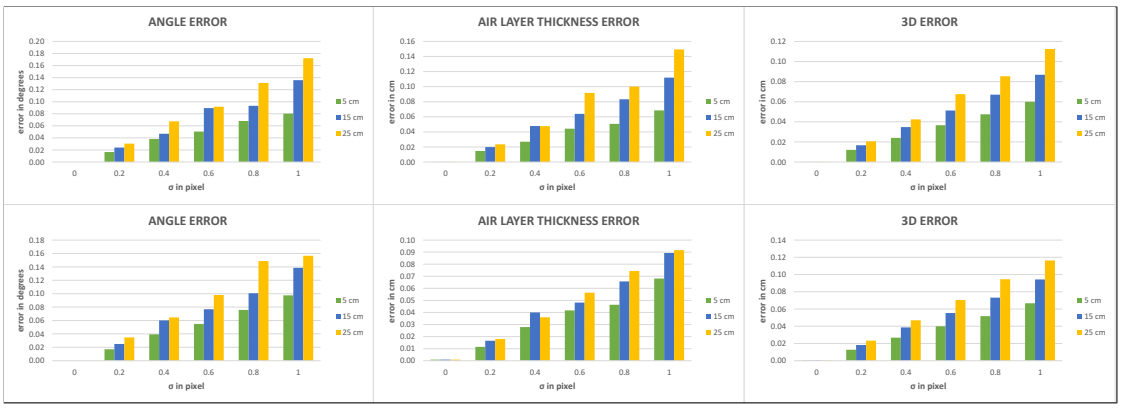

Figure 7.7: Inaccuracies of refractive calibration with RTE. Top row: $\phi = -10^\circ$, $\theta = 0^\circ$. Bottom row: $\phi = -10^\circ$, $\theta = -20^\circ$.

Refractive Triangulation Error (RTE). The error values almost equal to zero in all three categories at noise level $\sigma = 0$ show that the RTE is valid and that the approach was working correctly (See Figure 7.7). An increase of the noise level σ led to increased inaccuracies in each case. With few exceptions, an increase of the air layer thickness t_a led to increases at the single noise levels in almost every case. Significant differences between the varied orientations of the system axis by angle θ did not become apparent consistently. The inaccuracies of the 3D recovery were similar in particular. Since only two different orientations were tested, the indication at this point is that the effects may not be severe in general. Hence, the obvious growth of the error values was only consistent concerning an increased noise level σ and an increased air layer thickness t_a. Figure 7.8 shows that the area with the overall minimal error value could be detected unambiguously in both setups and that it contained the ground truth system axis.

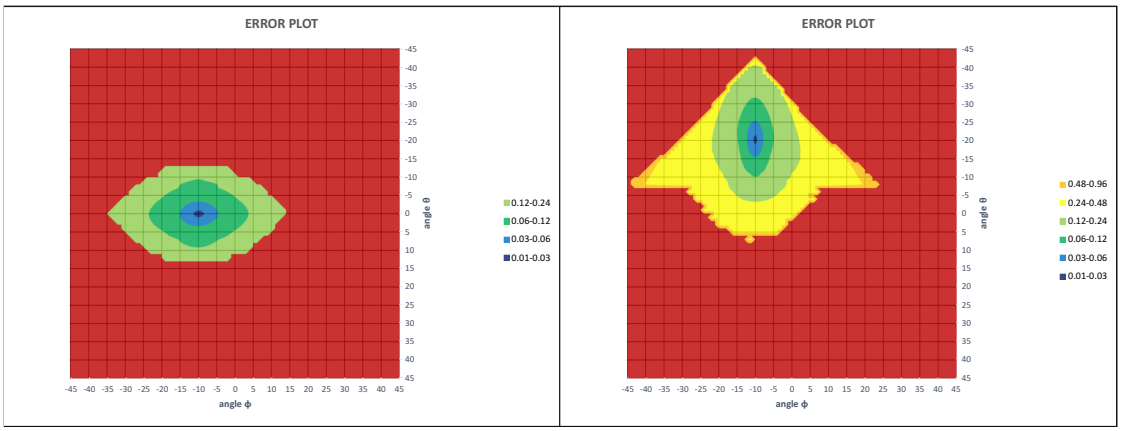

Figure 7.8: Plotting of the RTE. Left: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = 0^\circ$, $\sigma = 1$. Right: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = -20^\circ$, $\sigma = 1$.

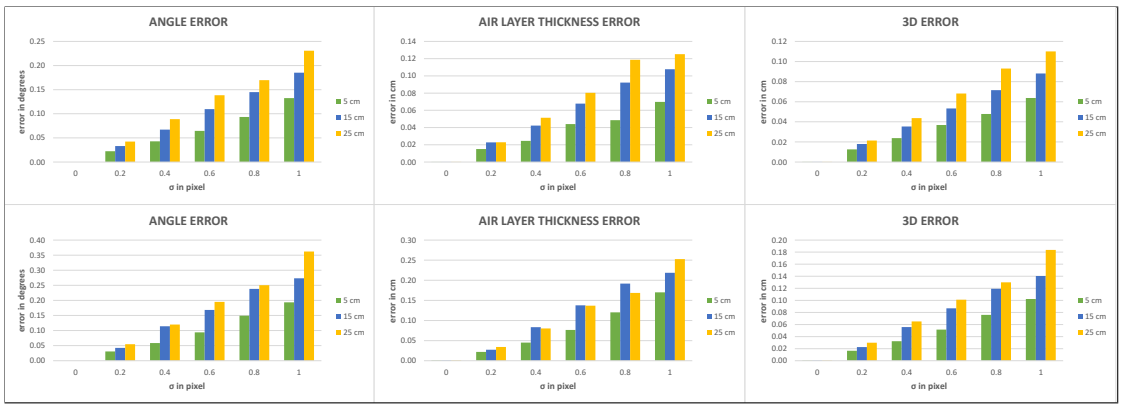

Figure 7.9: Inaccuracies of refractive calibration with *VIE*. Top row: $\phi = -10^\circ$, $\theta = 0^\circ$. Bottom row: $\phi = -10^\circ$, $\theta = -20^\circ$.

Virtual Intersection Error (VIE). The error values almost equal to zero in all three categories at noise level $\sigma = 0$ show that the *VIE* is valid and that the approach was working correctly (See Figure 7.9). An increase of the noise level σ led to increased inaccuracies in every case. With few exceptions, an increase of the air layer thickness t_a led to increases at the single noise levels in almost every case. In contrast to the *RTE*, significant differences between the varied orientations of the system axis by the angle θ did become apparent. Since only two different orientations were tested, the indication at this point is that the effects may be severe in general. Summarized, the growth of the error values was consistent concerning an increased noise level σ and an increased air layer thickness t_a and it was severe concerning an increased absolute value of the angle θ. Figure 7.10 shows that the area with the overall minimal error value could be detected unambiguously in both setups and that it contained the ground truth system axis.

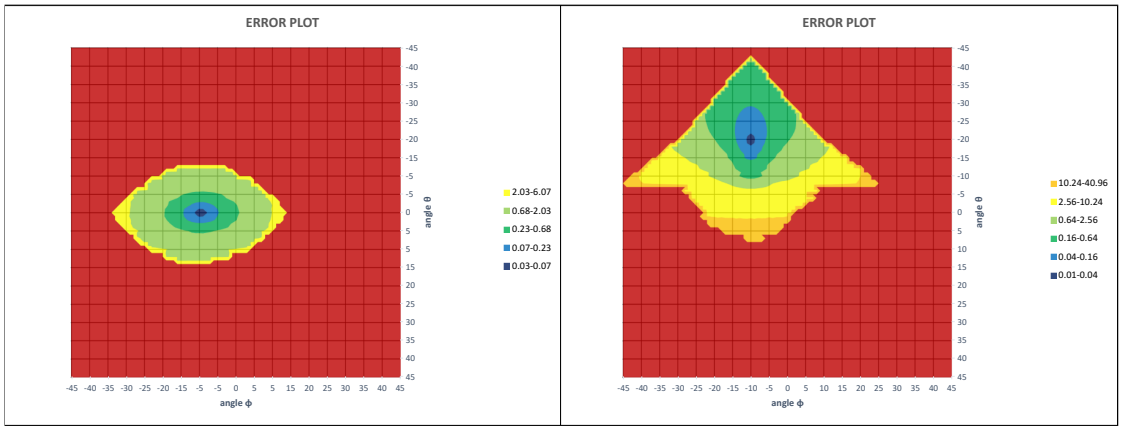

Figure 7.10: Plotting of the *VIE*. Left: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = 0^\circ$, $\sigma = 1$. Right: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = -20^\circ$, $\sigma = 1$.

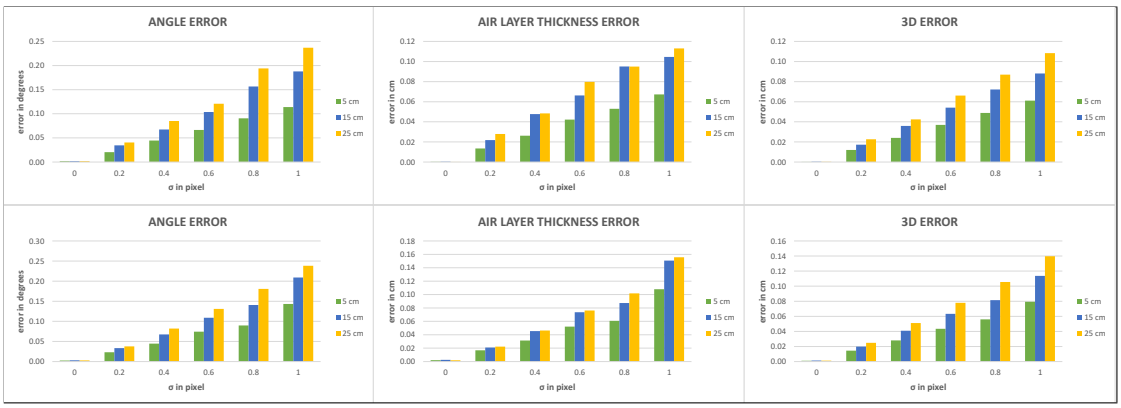

Figure 7.11: Inaccuracies of refractive calibration with *RRE*. Top row: $\phi = -10^\circ$, $\theta = 0^\circ$. Bottom row: $\phi = -10^\circ$, $\theta = -20^\circ$.

Refractive Reprojection Error (RRE). The error values almost equal to zero in all three categories at noise level $\sigma = 0$ show that the *RRE* is valid and that the approach was working correctly (See Figure 7.11). The further observations matched the ones of the *VIE*. These are that an increase of the noise level led to increased inaccuracies in every case and that, with few exceptions, an increase of the air layer thickness t_a led to increases at the single noise levels in almost every case. Significant differences between the varied orientations of the system axis by the angle θ did become apparent. Similarly, the obvious growth of the values was consistent concerning an increased noise level θ and an increased air layer thickness t_a and the growth was severe concerning an increased absolute value of the angle θ. Figure 7.12 shows that the area with the overall minimal error value could be detected unambiguously in both setups and that it contained the ground truth system axis.

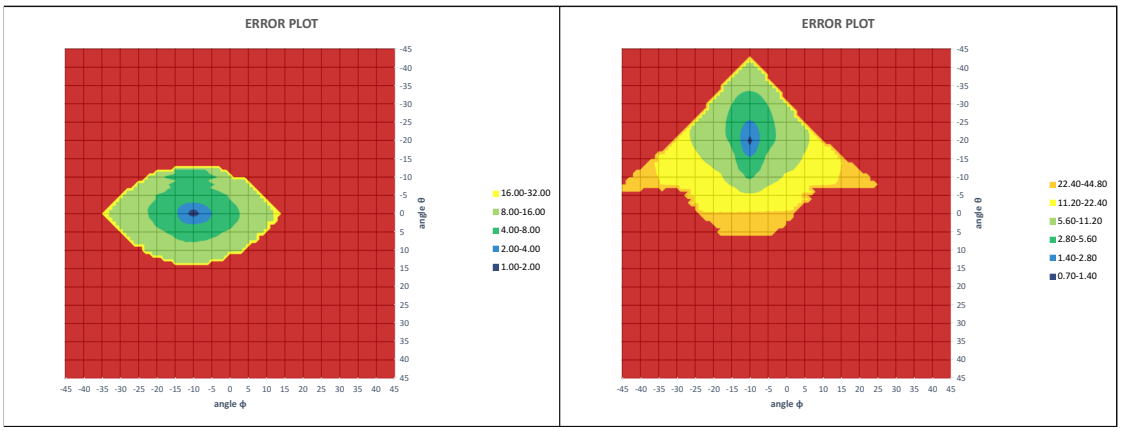

Figure 7.12: Plotting of the *RRE*. Left: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = 0^\circ$, $\sigma = 1$. Right: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = -20^\circ$, $\sigma = 1$.

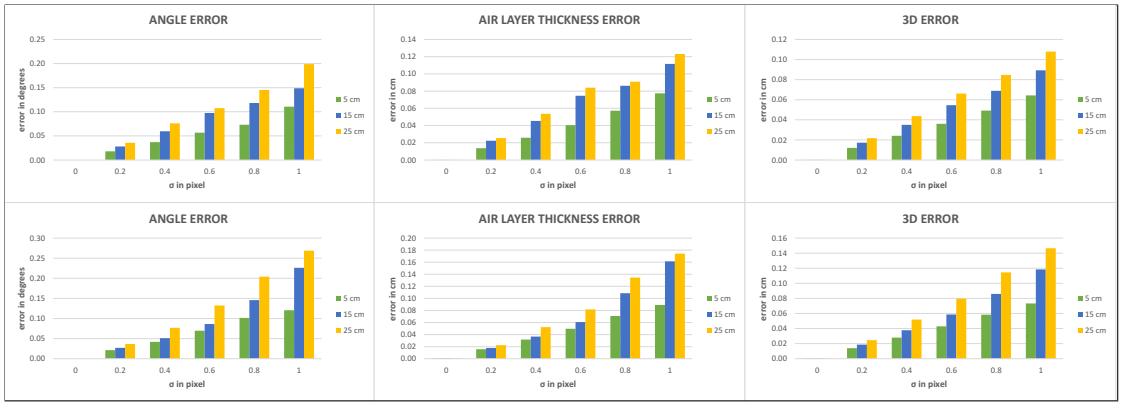

Figure 7.13: Inaccuracies of refractive calibration with *VRE*. Top row: $\phi = -10^\circ$, $\theta = 0^\circ$. Bottom row: $\phi = -10^\circ$, $\theta = -20^\circ$.

Virtual Reprojection Error (VRE). The error values almost equal to zero in all three categories at noise level $\sigma = 0$ show that the *VRE* is valid and that the approach was working correctly (See Figure 7.13). The further observations nearly matched the ones of the *VIE* and the *RRE*. In every case, an increase of the noise level σ led to increased inaccuracies. The only difference was that, with no exceptions, an increase of the air layer thickness t_a led to increases at the single noise levels. As before, significant differences between the varied orientations of the system axis did become apparent. Hence, the obvious growth of the error values was consistent concerning an increased noise level σ and an increased air layer thickness t_a and the growth was severe concerning an increased absolute value of the angle θ. Figure 7.14 shows that the area with the overall minimal error value could be detected unambiguously in both setups and that it contained the ground truth system axis.

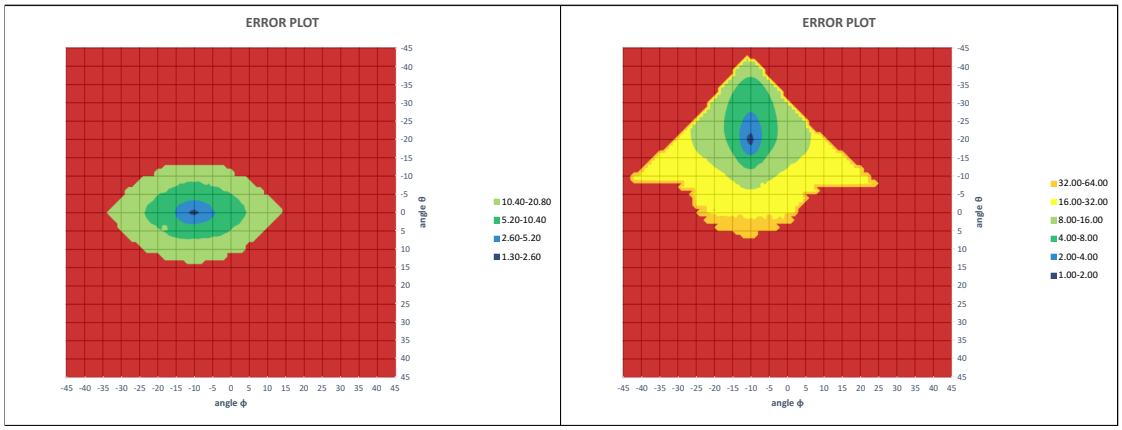

Figure 7.14: Plotting of the *VRE*. Left: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = 0^\circ$, $\sigma = 1$. Right: $t_a = 25cm$, $\phi = -10^\circ$, $\theta = -20^\circ$, $\sigma = 1$.

Comparison of Cost Functions and Calibration Strategies

After consideration of each cost function for itself, this section will comprise a comparison amongst themselves and hence of the two proposed calibration strategies for the calibration of a *SFRS*, namely the consecutive and the simultaneous determination of refractive parameters. A very obvious difference can be seen in the shapes of the areas in the previously shown error plots that contain the invalid system axes. The ones of the pattern-based cost functions *PAE* (Figure 7.4) and *PFE* (Figure 7.6) can be clearly distinguished from the rest (Figure 7.8, 7.10, 7.12 and 7.14). The reason is the difference between the approaches for the optimization of the system axis. Both cost functions *PAE* and *PFE* are independent of the air layer thickness and can therefore only be used for pattern-based system axis determination within the consecutive strategy. The best system axis is determined explicitly and the air layer thickness is only computed for this single axis. Hence, invalid axes can only be identified according to the criterion that not all of the rays in both camera's fields of view intersect the flat refractive interface, which has been defined earlier. In contrast, the simultaneous strategy comprises an implicit system axis determination with *RTE*, *VIE*, *RRE* or *VRE*. Therefore, further criteria for invalid axes can be applied within the determination of the air layer thickness. Since the air layer thickness is determined for every single hypothetical system axis, further axes can be invalidated. An example of these further criteria are resulting air layer thicknesses that fall below the minimally possible distance or that exceed the maximally possible distance. Falling below the minimally possible distance means that the resulting air layer thickness would lead to a flat refractive interface above Φ_0 in Figure 7.15, which is impossible due to the physical presence of the closest camera. Exceeding the maximally possible distance means that the resulting flat refractive interface would lead to *VOP*s being situated in air. As has been shown earlier, the *VOP*s $\check{O}$ and $\check{O}'$ are usually situated in water, can be at most inside glass, but not in air. Therefore, in theory, the valid range of values of the air layer thickness t_a is between $t_a = d(C, C) = 0$ and $t_a = d(C, A_3)$. Note that for multiple pairs of corresponding image points I and I', the maximal distance gets determined by the *VOP* that is the farthest away. In practice, the minimal distance is not realizable, since the center of projection is usually situated somewhere inside the optics of a camera.

The comparison of the results of the approaches for refractive calibration will contain the six distinct setups from the previous section, which were chosen according to the subsequent experiments with real data. The diagrams in Figure 7.16 and 7.18 are another representation of the results from Figure 7.3, 7.5, 7.7, 7.9, 7.11 and 7.13, which is better suited for comparison of the various cost functions amongst themselves. Furthermore, Figure 7.17 and 7.19 illustrate a comparison of the run-times of the applied approaches. Within these diagrams, the six distinct setups will be grouped according to their differences in the amount of the angle θ, which will result in two sets. Besides the six proposed refractive calibration approaches, the current state-of-the-art approach of *Chen and Yang* (*CY*) [CY14] will be included into the comparison. To be comparable, it was reimplemented and slightly adapted for the calibration of a *SFRS*.

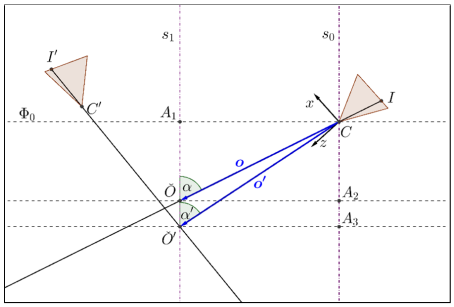

Figure 7.15: Range of the valid values of the air layer thickness.

Comparisons in the First Set. The first set of diagrams in Figure 7.16 and 7.17 is characterized by constant angles $\phi = -10^\circ$ and $\theta = 0^\circ$, and by a varying air layer thickness of $t_a = 5cm$, $t_a = 15cm$ and $t_a = 25cm$. Of the two proposed calibration strategies, the simultaneous determination of the refractive parameters outperformed the consecutive one. Hence, the two pattern-based cost functions *PAE* and *PFE* performed worse than the four pattern-independent cost functions *RTE*, *VIE*, *RRE* and *VRE*. As can be seen, *PAE* and *PFE* clearly produced higher inaccuracies in each case. Since both of them can only be used to determine the orientation of the system axis, the resulting inaccuracies get incorporated into the subsequent computation of the air layer thickness. This becomes obvious in the second column. The direct consequence is that the inaccuracies of the recovered 3D coordinates in column three are similarly high.
The first column of Figure 7.16 shows that the simultaneous strategy with the pattern-independent cost functions *RTE*, *VIE*, *RRE* and *VRE*, and the reference approach *CY* produced similar results, with a maximal inaccuracy of the determined system axis of about 0.3°. Regarding the air layer thickness, the second column shows that *CY* produced slightly lower inaccuracies. Since the approach for the computation of the air layer thickness was the same for the consecutive and the simultaneous strategy, the second column shows that it produced better results in almost every case within the simultaneous strategy. The consequence for the simultaneous strategy is that the overall inaccuracies are distributed between the inaccuracies of the systems axis and the air layer thickness. Therefore, since the air layer thickness could be computed more accurately, the implicit determination of the system axis within the simultaneous strategy resulted in more accurate results as well. The maximal inaccuracy of the pattern-independent cost functions at the maximally tested noise level $\sigma = 1$ amounted to about $0.15cm$. Regarding the inaccuracies of the recovery of the 3D point cloud in column three, *RTE*, *VIE*, *RRE*, *VRE* and *CY* produced similar results, once again. The maximal inaccuracy amounted to about $0.11cm$. The depicted run-times in Figure 7.17 show a clear order. The two pattern based approaches utilizing *PAE* and *PFE* and the one utilizing *VIE* were the fastest. These were closely followed by the approach utilizing *RRE*. The approaches utilizing *RTE*, *VRE* and also the reference approach *CY* were severely slower.

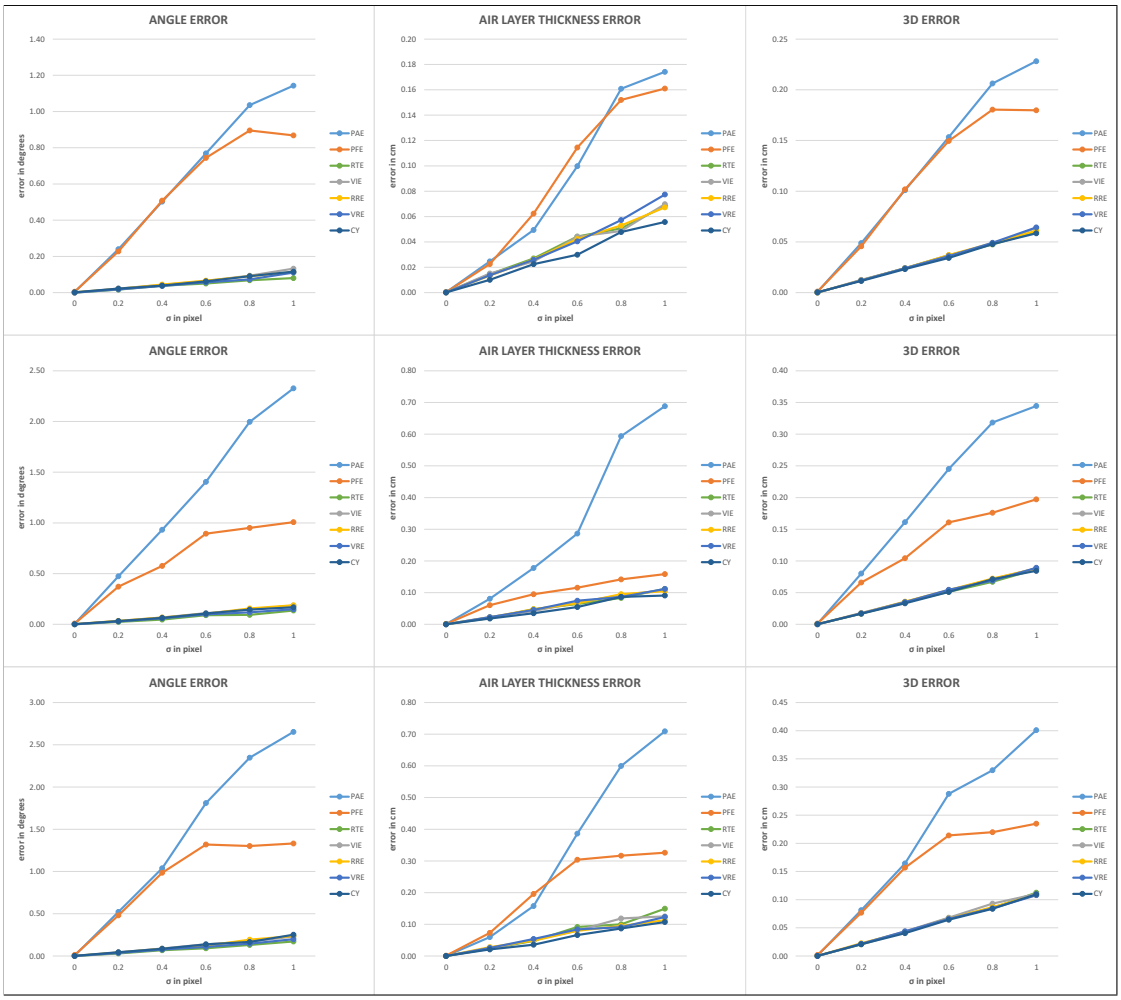

Figure 7.16: Comparison of the inaccuracies for the setups with $\phi = -10^\circ$ and $\theta = 0^\circ$. Top row: $t_a = 5cm$. Middle row: $t_a = 15cm$. Bottom row: $t_a = 25cm$

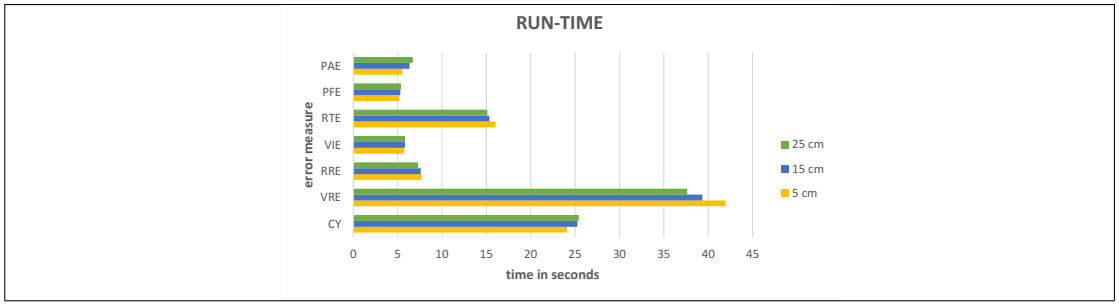

Figure 7.17: Comparison of the mean run-times for the setups with $\phi = -10^\circ$ and $\theta = 0^\circ$.

Comparisons in the Second Set. The second set of diagrams in Figure 7.18 and 7.19 is characterized by constant angles $\phi = -10^\circ$ and $\theta = -20^\circ$, and by a varying air layer thickness of $t_a = 5cm$, $t_a = 15cm$ and $t_a = 25cm$. Due to an angle θ that was differing from zero, the system axis in question was additionally tilted. Compared to Figure 7.16, Figure 7.18 shows that an increase of the angle θ led to results that were more scattered. A difference was that the results of the pattern-based cost function *PFE* approached the results of the four pattern-independent cost functions *RTE*, *VIE*, *RRE* and *VRE* in some cases, while the *PAE* still performed worst. Of these four pattern-independent cost functions, the *VIE* performed worst in almost every case. This time, the reference approach *CY* had no advantages, while the *RTE* seemed to handle the increased tilt the best. Compared to the previous setups with no tilt, the maximal inaccuracy of the determined system axis increased from about 0.3° to about 0.4°, that of the determined air layer thickness from about $0.15cm$ to about $0.25cm$ and that of the recovered 3D point cloud from about $0.11cm$ to about $0.19cm$. Each of the depicted run-times in Figure 7.19 shows a slight increase, but the general order remained.

Findings

The observations were that all six cost functions are valid and that their implementations were working correctly. The observations at noise level $\sigma = 0$ have been further used as a means to evaluate the approach for the recovery of 3D coordinates by *Refractive Triangulation* from Section 6.2. The conformity of the recovered and the ground truth 3D point clouds shows that it was working correctly as well.
An obvious growth of the inaccuracies was consistent concerning an increased noise level σ and an increased air layer thickness t_a. Regarding the differing orientations of the system axes by the angles ϕ and θ, an increase of the absolute value of θ was performed to increase the range of angles of incidence in the field's of view of the stereo system. Therefore, in almost all cases, such a tilted setup led to a severe increase of the inaccuracies. The worst results were produced by the consecutive determination of the refractive parameters by utilizing the pattern-based *PAE*. Utilizing the *PFE* performed better. Therefore, from a simulation point of view, it was easy to decide in favor of the calibration strategy with simultaneous determination of the refractive parameters. In contrast, it was hard to decide in favor of one of the four pattern-independent cost functions *RTE*, *VIE*, *RRE* and *VRE*, since they performed similarly well within the setups with no tilt of the system axis and only differed slightly within the setups with tilt. The reimplementation of the reference approach of Chen and Yang [CY14] (*CY*) ranked among them, without performing significantly better or worse. Regarding the run-times of the single approaches, a clear order became apparent. However, this order could not be mapped to the produced inaccuracies of the respective approaches.
The refractive indices of the participating media are not determined within the proposed approaches, since they are usually measurable or known. The refractive index of water is commonly known and therefore usually assumed to be 1.33, however, it can vary slightly. An outcome of a simulation of the effects of inaccurate values (See Appendix C) is that refractive calibration is preferably performed with image features of close objects.

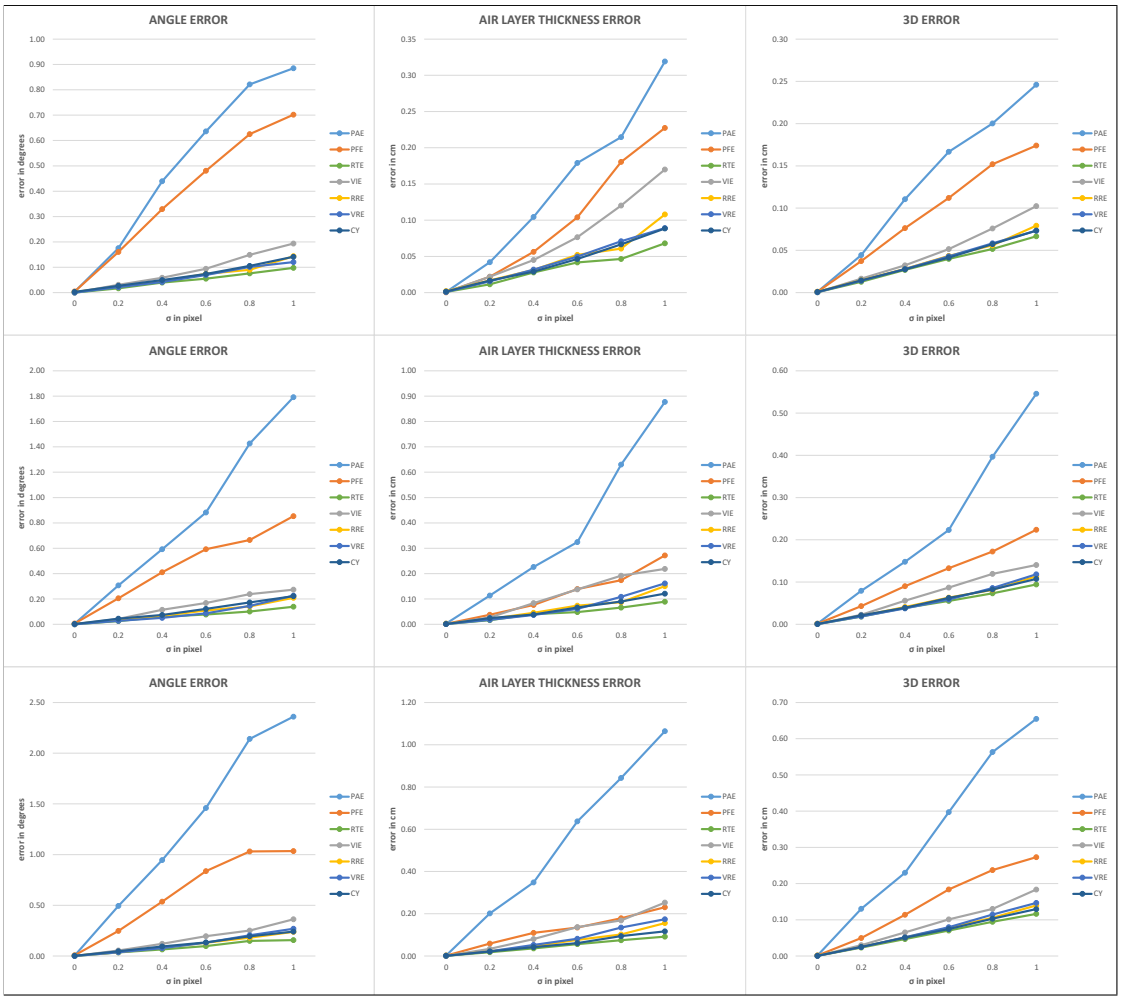

Figure 7.18: Comparison of the inaccuracies for the setups with $\phi = -10^\circ$ and $\theta = -20^\circ$. Top row: $t_a = 5cm$. Middle row: $t_a = 15cm$. Bottom row: $t_a = 25cm$

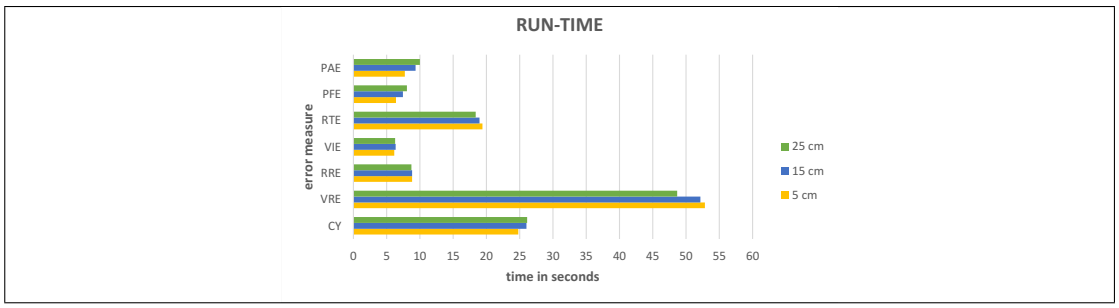

Figure 7.19: Comparison of the mean run-times for the setups with $\phi = -10^\circ$ and $\theta = -20^\circ$.

7.1.2 Stage 2 - Rendered Images

Simulated Test Data Stage 2 made use of rendered image data, which was generated with *Blender*[1]. In this case, the calibration of the stereo camera and the feature detection needed to be performed algorithmically as a pre-processing step of refractive calibration. As before, a checker pattern was used to generate the feature points in the stereo images for refractive calibration. Therefore, a conventional feature detector for finding the corners of this pattern was introduced into the processing. Additionally to evaluating the proposed refractive calibration approaches, this was a good opportunity to collect quantitative data to point out the extent of the errors that arise from further calibration strategies from the literature. Two representative setup configurations for a *SFRS* were generated and analyzed. In accordance with *Simulated Test Data Stage 1*, a brief analysis of further setups can be found in Appendix D.

Preliminaries

Further Calibration Strategies from the Literature. As described in Section 3.2, the absorption strategy (*ABS*) is supposed to be an alternative to handle refractive effects in special viewing configurations. For a general viewing configuration, the fundamental finding is that refractive effects should be modeled explicitly. This was analyzed briefly. A common alternative to estimate explicitly modeled refractive parameters, which can be found in the literature (See Section 3.3.5), is to fix a calibration object with a known pattern onto the refractive interface. Afterward, some kind of pose from 3D-2D point correspondences algorithm can be used to estimate (*EST*) the pose of the master camera with respect to the calibration object. This strategy was analyzed briefly as well. For the sake of completeness, the strategy to disregard (*DIS*) refractive effects was also analyzed briefly. On the contrary to the weak definition in Section 3.1, this rather stands for a calibration of the imaging system in air and ignoring refractive effects completely.

Reference Approaches for Calibration in Air. Since both the disregard and the absorption of refractive effects rest upon the application of in air calibration approaches without any adjustments to the underwater environment, a matching method had to be determined. The selected methods were the ones of Zhang [Zha00] and Bouguet [Bou08]. All that they need is a set of simultaneously taken views of a planar calibration object, such as a checker pattern of known dimensions. The advantage is that both approaches are readily usable in form of available algorithms in the *Camera Calibration Toolbox for Matlab* [Bou08] and *OpenCV* [Bra00]. In this thesis, the *OpenCV* implementation was chosen, which is based on both Zhang [Zha00] and Bouguet [Bou08].
Since the intrinsic and extrinsic camera parameters need to be known in advance in all the following experiments, this was also the method of choice for the calibration of the stereo camera itself (in air). The necessary sets of stereo views of the checker pattern on the calibration object were acquired in air and under water. A subset of the underwater images was simultaneously used for refractive calibration.

[1] Blender: `https://www.blender.org/`

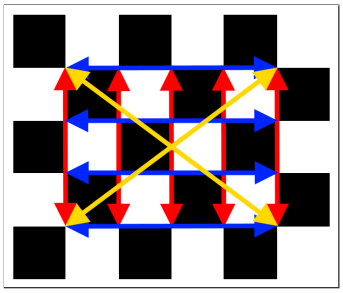

Figure 7.20: Directions of the length measurements on a checker pattern.

Evaluation Criteria. In most real world setups there is no ground truth data available for the refractive parameters. However, it is possible to evaluate the computed refractive parameters implicitly, by explicitly evaluating the 3D coordinates of points in the scene, which were recovered with the aid of these refractive parameters. Therefore, some additional criteria are necessary, which represent valid means for evaluations in real world setups. The following criteria were utilized from here on:

- Reprojection Error (*DIS*, *ABS*) or Refractive Reprojection Error (Others).
- 3D Error (Equal to *Simulated Test Data Stage 1*).
- The error resulting from plane fitting into the recovered 3D points.
- Length measurements on the known checker pattern.

From the earlier proposed criteria, only the 3D Error could be reused. The necessary benchmark data was made available in the following way. Starting point was a plane in the scene, which was textured with a checker pattern. A rendered stereo image pair of this plane was acquired once with and once without the flat refractive interface and water. A 3D point cloud of the checker pattern corners was recovered from the rendered stereo image pair in air by means of conventional triangulation. This 3D point cloud was used as benchmark data. The rendered underwater stereo image pair was used to recover the 3D point cloud of the checker pattern corners, by utilizing the refractive parameters from the respectively tested calibration strategy. The comparison of both 3D point clouds resulted in the overall 3D Error. Additionally, the mean Euclidean distances in each of the three coordinate directions were computed separately to determine the one with the largest extent.
Since the recovered 3D point cloud had to be planar, plane fitting was used for evaluation as well. The best fitting plane was computed and the mean Euclidean distance of the 3D points to this plane characterized the accuracy of the fitting.
For the length measurements, some of the inner corners of the checker pattern were used. As can be seen exemplary in Figure 7.20, target of the measurements were all the horizontal lines, all the vertical lines and the two diagonals. All the lines in one of the three directions had the same length. The 3D coordinates of the respective starting and ending points were used to calculate the Euclidean distances between them. Afterward,

the mean Euclidean distances in all three directions were computed and their absolute differences to the known lengths were used for evaluation.
Note that this planar shape with a checker pattern was chosen for the sake of producing simple, additional means for evaluation. It is not a requirement for the proposed refractive calibration approaches and any other object could have been used, but without it, the straightforward computation of the benchmark data, comprising a 3D point cloud, known lengths and a target for plane fitting, would not be applicable.

Comparison of Cost Functions and Calibration Strategies

For each of the two simulated setups, all of the described calibration strategies were evaluated. These comprised the disregard (*DIS*) and the absorption (*ABS*) of refractive effects as well as the estimation (*EST*) of the master camera's pose by a pose from 3D-2D point correspondences algorithm. Furthermore, the proposed consecutive strategy utilizing *PAE* or *PFE* and the simultaneous strategy utilizing *RTE*, *VIE*, *RRE* or *VRE* to determine the refractive parameters were evaluated and compared to the reference approach *CY*. Both setups were characterized by cameras with a focal length of $8.5mm$ and an image sensor with a width of $11.34mm$ and a height of $7.13mm$, resulting in a wide field of view. These parameters matched the parameters of the cameras of the subsequent evaluations with real data. The air layer thickness amounted to $11cm$ per camera and the stereo baseline amounted to $10cm$. The glass layer thickness amounted to $1cm$ and the checker pattern was located at a distance of $40cm$ to the water-sided interface. The images that were needed for all calibration procedures comprised:

- one image of a checker pattern on the refractive interface for *EST*,
- a set of stereo images of a checker pattern in air for stereo calibration itself, which is equal to *DIS*,
- a set of underwater stereo images of a checker pattern for *ABS* and
- a subset of the underwater stereo images for refractive calibration (*PAE*, *PFE*, *RTE*, *VIE*, *RRE*, *VRE*, *CY*).

The subset of the underwater stereo images was chosen to contain about 400 feature points in total, to match the earlier determined number.

First Setup. In the first simulated setup, the image sensors of the two cameras were perpendicular to the interface normal. Hence, both principal axes were parallel and the system axis was characterized by $\phi = 0^\circ$ and $\theta = 0^\circ$. This setup matches the conditions that are expected to make the absorption strategy (*ABS*) applicable. The results for this setup can be seen in Table 7.1. In contrast to the previous evaluation without explicit stereo calibration and feature detection, the values represent the accumulation of the inaccuracies from stereo calibration, feature detection and refractive calibration. The first eight columns show that all the proposed approaches of this thesis (*PAE*, *PFE*, *RTE*, *VIE*, *RRE* and *VRE*) performed very well, as was expected with simulated data. As

before, *PAE* performed worst, but still very well. The reference approach *CY* performed equally well and the pose estimation (*EST*) similarly well. *DIS* produced unacceptable results in all categories, while *ABS* produced results that were at least more tolerable, except for the inaccuracies of the 3D coordinates in z-direction. Although the setup is supposed to match *ABS*, it was clearly outperformed by the explicit modeling of refractive effects together with refractive calibration.

Second Setup. The second simulated setup differed from the first by the orientation of the two cameras. Both principal axes were not parallel any more, but converged, and the perpendicularity between the image sensors and interface normal did no longer exist. Furthermore, the cameras were rolled by 1° and 2° around their principal axes. The system axis was characterized by $\phi = -6^\circ$ and $\theta = -10^\circ$. Compared to the first setup, this increased the range of angles of incidence in the field's of view of the stereo system. The rest of the setup stayed the same. The resulting error values are shown in Table 7.2. Again, the first eight columns show that all of the proposed approaches of this thesis performed very well. However, compared to the first setup, a slight degradation was evident in every case. This was most likely happening due to the increased range of angles of incidence, which matches the findings from *Simulated Test Data Stage 1* in Section 7.1.1. The overall tendencies remain similar to the first setup. The two approaches *DIS* and *ABS* were clearly outperformed. *DIS* even suffers from additional inaccuracies of the 3D coordinates in x- and y-direction and *ABS* shows increased inaccuracies at all length measurements. Consequently, it can be confirmed that *ABS* is better suited for setups with perpendicularity between the image sensors and the interface normal.

Findings

In contrast to *Simulated Test Data Stage 1* in Section 7.1.1, *Simulated Test Data Stage 2* did not contain an augmentation with Gaussian noise. Therefore, its major task was to show that the accumulation of errors of the pre-process containing common approaches for stereo calibration and feature detection from *OpenCV* and of the process of refractive calibration is low, as would be expected from simulated data. Nevertheless, the texture mapping during the rendering of the utilized checker patterns is not guaranteed to be ideal. Therefore, a source of error exists. However, all of the proposed refractive calibration approaches from this thesis performed well on rendered images. The introduced evaluation criteria allowed comparisons of the proposed approaches amongst themselves and to alternative approaches from the literature. The estimation (*EST*) of the master camera's pose by a pose from 3D-2D point correspondences algorithm performed equally well on simulated data. The analysis of the absorption strategy (*ABS*) showed that, in special viewing conditions, it can lead to length measurements that are reasonable to a certain degree, but the inaccuracies of the recovered 3D coordinates were always significant. It could be confirmed that the absorption capabilities decreased with non-perpendicularity between image sensor and interface normal, as was already stated in the literature (Section 3.2). Due to its severe inaccuracies, the strategy to disregard (*DIS*) refractive effects should definitely be avoided.

ERROR MEASURE		EST	PAE	PFE	RTE	VIE	RRE	VRE	CY	DIS	ABS
reprojection	$left(pixel)$	0.028	0.066	0.026	0.026	0.026	0.027	0.026	0.026	2.495	0.155
	$right(pixel)$	0.028	0.066	0.026	0.026	0.026	0.026	0.026	0.027	2.495	0.155
3D	$x(cm)$	0.003	0.049	0.005	0.004	0.004	0.004	0.004	0.004	0.284	0.032
	$y(cm)$	0.003	0.013	0.004	0.004	0.004	0.003	0.004	0.003	0.159	0.016
	$z(cm)$	0.039	0.059	0.018	0.018	0.017	0.017	0.018	0.017	12.038	3.709
	$3D(cm)$	0.04	0.084	0.02	0.019	0.019	0.019	0.019	0.018	12.047	3.710
length measurement	$horizontal(cm)$	0.014	0.024	0.006	0.006	0.006	0.006	0.006	0.006	2.030	0.135
	$horizontal(\%)$	0.04	0.069	0.017	0.016	0.017	0.017	0.016	0.017	5.638	0.374
	$vertical(cm)$	0.007	0.019	0.005	0.005	0.005	0.005	0.005	0.005	0.573	0.032
	$vertical(\%)$	0.028	0.070	0.019	0.019	0.019	0.019	0.019	0.019	2.204	0.122
	$diagonal(cm)$	0.012	0.008	0.002	0.002	0.002	0.002	0.002	0.002	2.504	0.212
	$diagonal(\%)$	0.028	0.016	0.004	0.003	0.005	0.004	0.003	0.005	5.638	0.477
plane fitting	(cm)	0.01	0.009	0.008	0.008	0.008	0.008	0.008	0.008	1.071	0.093

Table 7.1: Error analysis of the recovered 3D point cloud from simulated image data of a setup with perpendicularity between the image sensors and the interface normal.

ERROR MEASURE		EST	PAE	PFE	RTE	VIE	RRE	VRE	CY	DIS	ABS
reprojection	$left(pixel)$	0.036	0.065	0.035	0.034	0.034	0.033	0.034	0.033	3.022	1.185
	$right(pixel)$	0.036	0.065	0.034	0.034	0.035	0.033	0.034	0.033	3.022	1.18
3D	$x(cm)$	0.006	0.028	0.005	0.006	0.007	0.006	0.006	0.005	1.286	1.394
	$y(cm)$	0.008	0.035	0.004	0.005	0.004	0.003	0.005	0.004	1.995	2.589
	$z(cm)$	0.039	0.067	0.019	0.02	0.021	0.02	0.02	0.02	12.024	0.771
	$3D(cm)$	0.041	0.083	0.021	0.023	0.024	0.022	0.023	0.022	12.266	3.071
length measurement	$horizontal(cm)$	0.014	0.024	0.008	0.009	0.008	0.008	0.009	0.006	1.998	0.453
	$horizontal(\%)$	0.04	0.069	0.023	0.025	0.021	0.022	0.025	0.018	5.55	1.259
	$vertical(cm)$	0.008	0.020	0.007	0.007	0.006	0.006	0.007	0.005	0.503	0.535
	$vertical(\%)$	0.03	0.074	0.026	0.027	0.022	0.023	0.027	0.02	1.934	2.057
	$diagonal(cm)$	0.025	0.054	0.019	0.019	0.017	0.017	0.019	0.015	2.409	0.953
	$diagonal(\%)$	0.056	0.123	0.042	0.043	0.039	0.039	0.043	0.035	5.425	2.145
plane fitting	(cm)	0.01	0.012	0.009	0.009	0.009	0.009	0.009	0.009	1.076	0.301

Table 7.2: Error analysis of the recovered 3D point cloud from simulated image data of a general setup with no perpendicularity between the image sensors and the interface normal.

7.2 Refractive Calibration - Real Test Data

A major difference between the evaluation with simulated data and real data is that the deployment of real cameras implicates lens distortion. While simulated cameras are modeled according to the pinhole camera model, real cameras can only be approximated by it. In contrast to the utilization of ideal pinhole cameras during simulation, the utilization of real cameras most likely introduces additional inaccuracies, such as lens distortions, which consequently influence the calibration of the stereo camera. However, regarding lens distortion, the assumption is that it neither is influenced by, nor does it influence refractive effects. Therefore, all the processed images were undistorted after stereo camera calibration in a pre-processing step. Image quality in terms of sharpness and blurring are further factors that can introduce additional inaccuracies. Stereo calibration, as well as refractive calibration depend on corresponding features in the two views of the imaging system. In this thesis, stereo calibration is always performed with the aid of planar calibration objects with checker patterns, while refractive calibration is only sometimes. The necessary corner detection algorithms are most likely to be affected by the image quality.
The evaluation with real test data was realized in two stages. Since the evaluation with simulated test data already pointed out some tendencies regarding the performance differences between the consecutive and the simultaneous determination of the refractive parameters of a *SFRS*, it was analyzed, if these tendencies apply in real world setups as well. In *Real Test Data Stage 1*, all six of the proposed approaches were compared by utilizing images of planar calibration objects with a checker pattern, while in *Real Test Data Stage 2* only the four pattern-independent approaches were tested without using calibration objects at all. In general, ground truth data for the system axis and the air layer thickness are not available in real world setups. If these are not available, the previously introduced reprojection error, or rather refractive reprojection error, the 3D Error, the measurements of known lengths and plane fitting can be applied as criteria for the implicit evaluation of the refractive parameters by explicit evaluation of a recovered 3D point cloud. However, usually there is no ground truth data for the pose of a 3D point cloud in real world setups as well, which makes the 3D Error unavailable. In the following section, an alternative way to produce benchmark data for a 3D point cloud will be described, which enables the utilization of the 3D Error.

7.2.1 Benchmark Data Generation

The approach for benchmark data generation is based on Dolereit [Dol15] and consists of the determination of a 3D point cloud and its pose with respect to the master camera of a stereo camera. An aquarium setup was used and the target of the recovery of 3D coordinates were the inner corners of a checker pattern on a submerged planar object (benchmark object). The proposed approach is to fix this benchmark object rigidly to the stereo camera with the aid of profile rails, as can be seen on the left side of Figure 7.21. Therefore, a mount that surrounds the aquarium was constructed. It allowed the fixation of the rigidly coupled camera-object system on two levels. On level one (Left side of

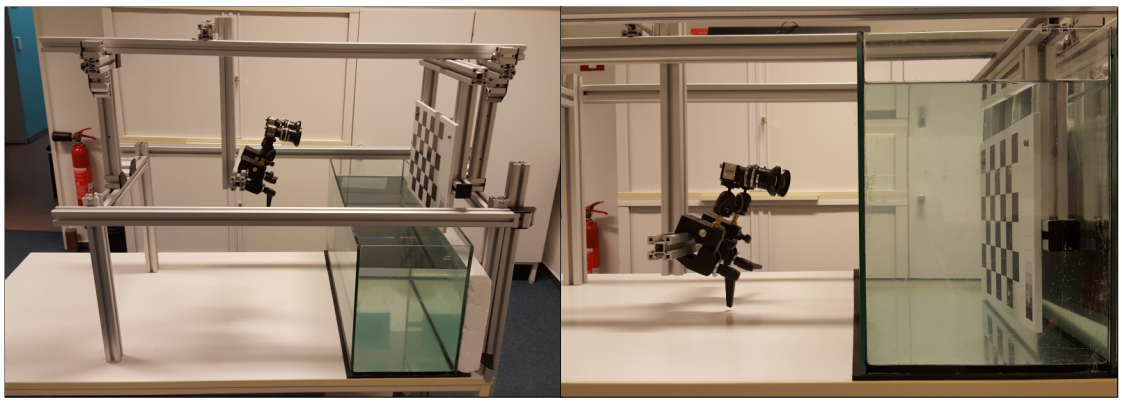

Figure 7.21: Installation for the evaluations with generation of benchmark data.

Figure 7.21), the stereo camera and the object were situated above water. On level two (Right side of Figure 7.21), the benchmark object got submerged, while the stereo camera was situated in front of one long side of the aquarium. The profile rails allowed a flexible adjustment of the camera-to-object distances and simultaneously an adjustment of the air layer thickness. The adjustment of the stereo baseline was possible as well. The used ball tripod heads allowed a flexible adjustment of the camera orientations and simultaneously an adjustment of the orientation of the system axis. With this setup, it was easily possible to calibrate the stereo camera in advance. A calibrated stereo camera makes the recovery of 3D coordinates in air possible. Therefore, the fixation of the camera-object system on the upper level was used to generate the benchmark 3D point cloud data of the checker corners by means of a conventional triangulation approach. This 3D point cloud was generated with respect to the master camera and its accuracy depended on the quality of the stereo calibration and feature detection algorithms. Since the same camera parameters were used for the subsequent refractive calibration approaches and since the same feature detection algorithm was used to specify the target points for 3D recovery, the way to generate the benchmark data was an obvious choice.

7.2.2 Stage 1 - With Calibration Object

Comparative Evaluation

Evaluated Setups. The parameters of the utilized stereo camera resembled the ones from simulation. Both cameras consisted of a lens with a focal length of $8.5mm$ and an image sensor of dimensions $11.34mm \times 7.13mm$, resulting in a wide field of view. Both had a resolution of 1936×1216 pixels. The principal axes of both cameras converged to generate a maximally overlapping field of view at the target object. Since the intrinsic and extrinsic parameters of the stereo camera were not varied at the distinct setups and therefore stayed constant, they were determined with the aid of *OpenCV* in advance once. The calibrated stereo baseline amounted to $10,143cm$. The reprojection error can be regarded as a quality measure of the stereo calibration. The resulting reprojection error

amounted to an average of 0.117 pixels. This error is mostly due to the underlying pinhole camera model, which can only approximate the image formation of real cameras. To remove lens distortion, all the processed images were undistorted based on the results of the conducted stereo camera calibration. Despite the relatively low value of the resulting reprojection error, it induced inaccuracies into the subsequent refractive calibration.
As mentioned before, the implicit evaluation of the refractive parameters by explicit evaluation of a recovered 3D point cloud was the method of choice. Target of the recovery of 3D coordinates was the benchmark object with a checker pattern of dimension 6×5, which was placed at about $30cm$ to the water-sided interface, as can be seen on the right side of Figure 7.21). This interface was plane-parallel, with a measured glass layer thickness that amounted to $t_g = 0.6cm$ and a refractive index of $n_g = 1.6$. The refractive indices of air and water were assumed to be $n_a = 1$ and $n_w = 1.33$.
Three distinct setups, which were realized with the aid of the construction on the right side of Figure 7.21, were part of the evaluation. The parameters of the stereo camera stayed fixed, while the air layer thickness and the system axis were varied slightly at each of the three setups. These slight variations were realized by translating the stereo camera along the horizontal profile rail and by rotating it with the aid of the mounting of the vertical profile rail. This was conducted arbitrarily in the range of some millimeters and degrees, respectively, and therefore, no ground truth data was available. By the point of view of an visual assessment, the air layer thickness amounted to approximately $15cm$ and the system axis was not perpendicular to the image sensor of the master camera in each of the three setups. Hence, ϕ and θ differed significantly from zero.

Benchmark Data. In *Real Test Data Stage 1*, the previously introduced approach for benchmark data generation (*BD*) form Section 7.2.1 was used to generate a 3D point cloud of the fixed checker pattern, with respect to the master camera. This benchmark 3D point cloud was evaluated first. Therefore, the same evaluation criteria as in *Simulated Test Data Stage 2* (See Section 7.1.2) were used, namely the reprojection error, the 3D Error, the measurements of known lengths and plane fitting. The results of the evaluation of the benchmark 3D point cloud are displayed next to the results of the evaluation of the refractive calibration in Table 7.3. However, the focus is only on the columns within the heading *BD* for now. Obviously, the 3D Errors (marked with X) can not be collected, since this would require the generation of another benchmark 3D point cloud.
The benchmark 3D point cloud was generated once for each of the three setups. It only depends on stereo camera calibration, which did not change between setups. In Table 7.3, the inaccuracies from the three setups are pooled together and the mean values and standard deviations are displayed. The reprojection error amounted to an average of about 0.17 pixels for the left and the right image. The length measurements were accurate and the plane fitting error indicated that the resulting 3D point cloud was very close to an ideal plane. Therefore, the generated benchmark data were considered to be valid. They represent the capability of the calibrated stereo camera for 3D recovery in air and therefore represent a valid means for the evaluation of the capability of the calibrated *SFRS* for 3D recovery under water.

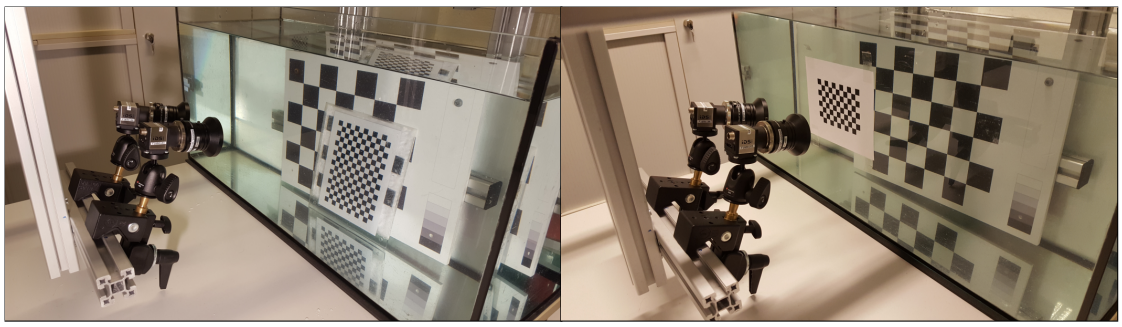

Figure 7.22: Acquisition of calibration data for: Left: *ABS*, *PAE*, *PFE*, *RTE*, *VIE*, *RRE*, *VRE* and *CY*, with a calibration object in water. Right: *EST*, with a calibration object fixed to the air-sided interface.

Further Calibration Approaches. Further evaluated approaches were the disregard (*DIS*) and the absorption (*ABS*) of refractive effects, as well as the estimation (*EST*) of the master camera's pose by a pose from 3D-2D point correspondences algorithm. The results can be seen in the respective columns of Table 7.3. Since *DIS* is equal to stereo camera calibration in air, it was only performed once overall. However, the subsequent recovery of the 3D point cloud of the benchmark object and its evaluation needed to be performed once for each of the three setups. As with *BD*, the results of these three setups are pooled together and the mean values and standard deviations are displayed. The results confirm that this strategy performs worst overall, as has been already indicated during simulation.
The calibration with *ABS* was performed once for every setup. Therefore, several underwater images of a plane-parallel calibration object with a checker pattern were used, as can be seen on the left side of Figure 7.22 (Small calibration object in front of benchmark object). As with *BD* and *DIS*, the subsequent recovery of the 3D point cloud of the benchmark object and its evaluation were performed once for each of the three setups and the results are displayed in the same way. Again, the findings from the simulations were confirmed.
The process for the estimation (*EST*) of the master camera's pose by a pose from 3D-2D point correspondences algorithm is shown exemplary on the right side of Figure 7.22. This can be realized by a single view of a suitable calibration object. Therefore, a checker pattern printed on paper was fixed to the air-sided interface. *EST* was executed one time for each of the three setups, as was the subsequent recovery of the 3D point cloud of the benchmark object and its evaluation. The displayed results are the mean values and standard deviations of the pooled configurations. As can be seen, the results were close to those of *BD* and partly even better. However, its greatest disadvantage is that it can only be performed in situations where the master camera can focus the checker pattern, but not in general. If the camera is in close proximity to the refractive interface, this becomes impossible.

Proposed Refractive Calibration Approaches. Similarly to the evaluation with simulated data, the proposed consecutive strategy utilizing *PAE* or *PFE* and the simultaneous strategy utilizing *RTE*, *VIE*, *RRE* or *VRE* to determine the refractive parameters were evaluated and compared to the reference approach *CY* and to the approaches *DIS*, *ABS* and *EST*. To evaluate both strategies, underwater images of a calibration object were used. To satisfy the earlier defined number of feature points of about 400, a subset of the underwater calibration images, which were already used for *ABS* (See left side of Figure 7.22), were chosen. Since more images than necessary were available, all of the proposed approaches were executed three times for each of the three setups, each time with a different subset of calibration images. Thus, the inaccuracies from nine distinct configurations were pooled and their mean values and standard deviations are displayed in Table 7.3. As can be seen, the overall tendencies do not differ significantly from the simulation results for a similar setup, which can be seen in Table 7.2. The consecutive strategy utilizing *PAE* or *PFE* performed worse than all the simultaneous ones. In contrast to the simulations, *PAE* outperformed *PFE*. The results of *PAE* came closer to the ones of *BD* and *EST*. *PFE* seemed to be more error-prone to the additional sources of error, such as inaccuracies of the stereo camera calibration. *RTE*, *VIE*, *RRE* and *VRE* performed equally well. There were hardly any differences. All four showed the lowest reprojection errors of all tested approaches, even lower than with *BD*. The reference approach *CY* performed slightly worse, just as *EST* did. However, compared to the simulation results the resulting inaccuracies of the proposed approaches in Table 7.3 increased, which indicates the accumulation of the inaccuracies of stereo calibration, checker corner detection and refractive calibration.

Findings

The approach for the generation of benchmark data (*BD*) worked reliably and marked the capabilities of the calibrated stereo camera for 3D recovery in air. The quality of the determined camera parameters is crucial, since inaccuracies are incorporated into the 3D recovery. Since these inaccuracies are also incorporated into the refractive calibration approaches, it is an valid means to generate benchmark data.
The simulation results matched the real world results quite well. The insufficient capabilities of *ABS* and *DIS* to handle refractive effects were substantiated. The evaluation also substantiated that the simultaneous strategy utilizing the pattern-independent *RTE*, *VIE*, *RRE* or *VRE* is preferable over the consecutive strategy utilizing the pattern-based *PAE* or *PFE* and even over the reference approach *CY*. Of these four pattern-independent cost functions none is standing out, since all performed equally well. Differences that had been evident during the evaluation with simulated data were not evident any more. The additional influencing factors like lens distortion and blurring had visible effects. As expected, inaccuracies of stereo calibration, checker corner detection and refractive calibration obviously accumulated.

ERROR MEASURE		BD		ABS		DIS		EST		PAE		PFE	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
reprojection	$left(pixel)$	0.168	0.062	0.382	0.196	2.004	0.006	0.106	0.003	0.193	0.071	0.401	0.295
	$right(pixel)$	0.170	0.064	0.387	0.191	2.085	0.029	0.109	0.004	0.197	0.071	0.408	0.297
3D	$x(cm)$	X	X	3.153	0.917	1.264	0.207	0.065	0.029	0.095	0.052	0.146	0.090
	$y(cm)$	X	X	3.207	0.460	1.539	0.018	0.016	0.003	0.088	0.037	0.115	0.094
	$z(cm)$	X	X	4.687	1.363	6.769	0.048	0.116	0.008	0.268	0.155	0.283	0.332
	$3D(cm)$	X	X	6.626	1.133	7.061	0.019	0.139	0.008	0.314	0.141	0.373	0.338
length measurement	$horizontal(cm)$	0.053	0.025	0.027	0.016	0.399	0.039	0.017	0.023	0.044	0.020	0.039	0.024
	$horizontal(\%)$	0.240	0.111	0.123	0.072	1.798	0.174	0.078	0.102	0.197	0.089	0.174	0.109
	$vertical(cm)$	0.024	0.011	0.035	0.022	0.195	0.021	0.054	0.017	0.026	0.021	0.033	0.015
	$vertical(\%)$	0.135	0.060	0.197	0.124	1.099	0.115	0.306	0.095	0.149	0.117	0.188	0.083
	$diagonal(cm)$	0.029	0.034	0.062	0.042	0.541	0.041	0.034	0.011	0.038	0.021	0.022	0.018
	$diagonal(\%)$	0.103	0.119	0.219	0.147	1.903	0.145	0.120	0.037	0.132	0.075	0.078	0.063
plane fitting	(cm)	0.017	0.002	0.199	0.035	0.433	0.002	0.027	0.004	0.022	0.011	0.038	0.037

ERROR MEASURE		BD		RTE		VIE		RRE		VTE		CY	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
reprojection	$left(pixel)$	0.168	0.062	0.065	0.008	0.074	0.014	0.067	0.008	0.070	0.017	0.123	0.066
	$right(pixel)$	0.170	0.064	0.066	0.009	0.076	0.015	0.069	0.008	0.071	0.018	0.127	0.069
3D	$x(cm)$	X	X	0.077	0.030	0.074	0.035	0.077	0.032	0.075	0.028	0.070	0.061
	$y(cm)$	X	X	0.053	0.008	0.054	0.015	0.053	0.006	0.050	0.014	0.059	0.026
	$z(cm)$	X	X	0.041	0.016	0.062	0.054	0.046	0.025	0.045	0.025	0.286	0.284
	$3D(cm)$	X	X	0.108	0.025	0.122	0.052	0.111	0.030	0.109	0.026	0.324	0.267
length measurement	$horizontal(cm)$	0.053	0.025	0.019	0.023	0.019	0.023	0.019	0.023	0.018	0.023	0.036	0.018
	$horizontal(\%)$	0.240	0.111	0.083	0.103	0.087	0.103	0.085	0.105	0.081	0.105	0.163	0.083
	$vertical(cm)$	0.024	0.011	0.044	0.015	0.044	0.016	0.044	0.016	0.046	0.017	0.053	0.027
	$vertical(\%)$	0.135	0.060	0.250	0.087	0.248	0.093	0.248	0.089	0.257	0.093	0.296	0.154
	$diagonal(cm)$	0.029	0.034	0.029	0.004	0.029	0.010	0.029	0.005	0.030	0.004	0.051	0.036
	$diagonal(\%)$	0.103	0.119	0.101	0.014	0.101	0.036	0.101	0.017	0.106	0.015	0.178	0.126
plane fitting	(cm)	0.017	0.002	0.018	0.002	0.018	0.004	0.018	0.003	0.019	0.004	0.038	0.031

Table 7.3: Error analysis of the recovered 3D point cloud from real image data. The preceding refractive calibration was performed with the aid of a known calibration object.

7.2.3 Stage 2 - Without Calibration Object

Comparative Evaluation

Evaluated Setups. In *Real Test Data Stage 2*, the evaluated setups were exactly the same three setups as in *Real Test Data Stage 1* (See Section 7.2.2). The evaluation criteria were also the same, namely the reprojection error, the 3D Error, the measurements of known lengths and plane fitting. The major difference was that no calibration object was used and that refractive calibration was performed solely with the aid of stereo correspondences in some additionally acquired calibration images. Therefore, it was possible to treat both stages in the same workflow.
Since the utilized aquarium setup did not provide many detectable features, a scene needed to be generated. The method of choice was to use waterproof images of a real underwater scene, which supposedly contained many detectable features. An example of the image and of its utilization can be seen in Figure 7.23. The image was printed on three sheets of paper in A4 format and laminated. The three parts were loosely coupled. Thus, due to its non-rigidity and its floating behavior, it was deviating from being a plane and gained significant 3D structure. It was placed arbitrarily in the aquarium and only one stereo image pair of it was used for refractive calibration.

Benchmark Data. Equally to *Real Test Data Stage 1*, the approach for benchmark data generation (*BD*) form Section 7.2.1 was used.

Further Calibration Approaches. In *Real Test Data Stage 1*, the further evaluated approaches were the disregard (*DIS*) and the absorption (*ABS*) of refractive effects, as well as the estimation (*EST*) of the master camera's pose by a pose from 3D-2D point correspondences algorithm. As should be clear, all of them can not be performed without calibration objects.

Proposed Refractive Calibration Approaches. Of the proposed consecutive and simultaneous strategies to determine the refractive parameters, the evaluation without calibration objects is only possible for the simultaneous strategy utilizing the pattern-independent cost functions *RTE*, *VIE*, *RRE* or *VRE*. As should be clear, the consecutive strategy utilizing the pattern-based cost functions *PAE* or *PFE* does not have the necessary capabilities.
In contrast to this stage, the previous evaluation stages utilized the corners of checker patterns as image features. These are easily detectable by common corner detection algorithms and easily matchable in two views of a stereo system, due to the known pattern structure. Without such a pattern, images features need to be handled differently. Therefore, sparse stereo matching was applied. The method of choice for feature detection and feature description was the *SIFT* [Low99]. It was used to calculate feature points in the stereo image pair. Afterward, a quick and efficient matching between left and right feature points was performed by using the *Fast Library for Approximate Nearest Neighbors* (*FLANN*) [ML09]. Their popularity and their performance were decisive

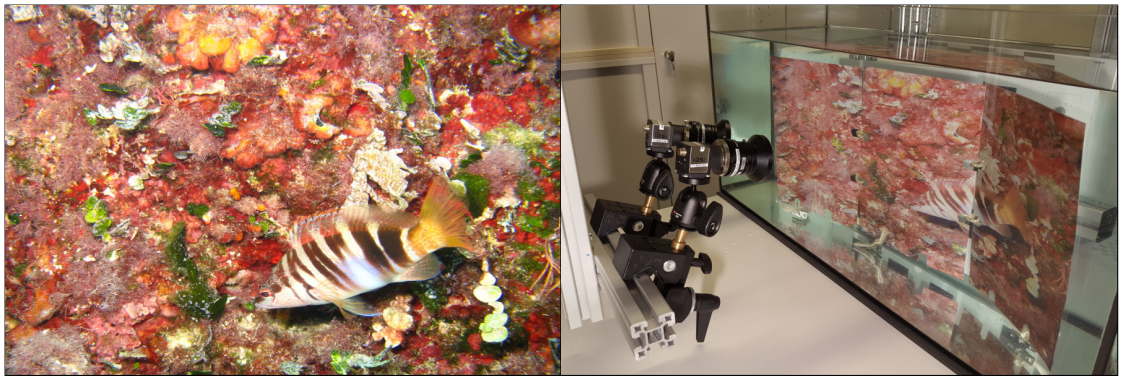

Figure 7.23: Refractive calibration without a known calibration object. Left: Arbitrary underwater scene. Right: Acquisition of calibration images.

for their choice. Note that any other combination of algorithms for feature detection, description and matching could have been used.
The applied testing scheme of the refractive calibration approaches was rather simple. To avoid erroneously matched feature pairs, a defined number of random feature pairs was chosen from the full set of available feature pairs. With these, refractive calibration was performed. It returned the refractive parameters and an error value for the utilized cost function. The refractive calibration was applied for a defined number of 100 repetitions, each time with a new set of random feature pairs. Finally, the refractive parameters with the lowest error value were taken as the result. The resulting inaccuracies for the three setups were pooled together and their mean values and standard deviations are displayed in Tables 7.4 and 7.5. The tested numbers of the random feature pairs amounted to 100, 200, 300 and 400.
The four proposed approaches utilizing RTE, VIE, RRE and VRE were tested and compared to the reference approach CY. As in the previous evaluation, all four proposed approaches performed equally well. The utilized number of feature points influenced the results only marginally. The reference approach CY was outperformed by the four proposed approaches in almost every case. Compared amongst itself, the reference approach CY performed best with the lowest number of feature points.

Findings

Although the testing scheme for feature pair selection was rather simple, all four proposed approaches performed equally well as during the evaluation with a calibration object (Compare with Table 7.3). This shows that it is possible to perform an accurate refractive calibration by the proposed approaches without the need of a known calibration object. At this point, since the quality of all four was nearly equivalent, the run-times of the single approaches become a factor. Since the analysis of the run-times during simulation has shown a clear order (Compare with Figure 7.17 and Figure 7.19), the following order can be accepted as the suggested order of application: VIE, RRE, RTE and VRE.

ERROR MEASURE		RTE 100		VIE 100		RRE 100		VRE 100		CY 100	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
reprojection	$left(pixel)$	0.062	0.010	0.068	0.008	0.058	0.005	0.060	0.008	0.117	0.092
	$right(pixel)$	0.063	0.009	0.070	0.010	0.059	0.006	0.061	0.009	0.124	0.099
3D	$x(cm)$	0.077	0.034	0.070	0.041	0.074	0.028	0.072	0.036	0.039	0.015
	$y(cm)$	0.045	0.008	0.047	0.007	0.030	0.006	0.031	0.005	0.056	0.037
	$z(cm)$	0.040	0.010	0.038	0.006	0.039	0.010	0.038	0.007	0.085	0.063
	$3D(cm)$	0.106	0.026	0.101	0.028	0.093	0.029	0.094	0.028	0.120	0.062
length measurement	$horizontal(cm)$	0.017	0.023	0.019	0.024	0.019	0.023	0.019	0.024	0.025	0.020
	$horizontal(\%)$	0.079	0.104	0.087	0.109	0.085	0.104	0.084	0.107	0.112	0.090
	$vertical(cm)$	0.047	0.017	0.047	0.018	0.048	0.016	0.049	0.018	0.051	0.017
	$vertical(\%)$	0.265	0.095	0.262	0.103	0.268	0.088	0.276	0.101	0.288	0.097
	$diagonal(cm)$	0.031	0.003	0.035	0.004	0.030	0.006	0.034	0.006	0.026	0.010
	$diagonal(\%)$	0.111	0.012	0.123	0.015	0.106	0.021	0.121	0.021	0.090	0.036
plane fitting	(cm)	0.020	0.004	0.020	0.005	0.018	0.002	0.020	0.004	0.017	0.003

ERROR MEASURE		RTE 200		VIE 200		RRE 200		VRE 200		CY 200	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
reprojection	$left(pixel)$	0.060	0.012	0.062	0.013	0.058	0.009	0.059	0.010	0.152	0.094
	$right(pixel)$	0.061	0.013	0.064	0.014	0.060	0.010	0.060	0.009	0.161	0.102
3D	$x(cm)$	0.071	0.029	0.082	0.037	0.081	0.029	0.068	0.032	0.093	0.044
	$y(cm)$	0.034	0.015	0.041	0.004	0.042	0.007	0.036	0.009	0.042	0.027
	$z(cm)$	0.053	0.014	0.040	0.012	0.048	0.006	0.042	0.003	0.336	0.196
	$3D(cm)$	0.103	0.023	0.106	0.032	0.109	0.018	0.094	0.023	0.361	0.186
length measurement	$horizontal(cm)$	0.019	0.022	0.018	0.023	0.018	0.022	0.019	0.023	0.046	0.011
	$horizontal(\%)$	0.084	0.098	0.080	0.103	0.083	0.101	0.087	0.103	0.207	0.050
	$vertical(cm)$	0.048	0.016	0.048	0.017	0.046	0.015	0.048	0.017	0.035	0.006
	$vertical(\%)$	0.272	0.090	0.270	0.095	0.259	0.087	0.269	0.095	0.197	0.034
	$diagonal(cm)$	0.030	0.006	0.032	0.004	0.028	0.002	0.032	0.006	0.023	0.013
	$diagonal(\%)$	0.105	0.023	0.114	0.016	0.099	0.006	0.112	0.021	0.082	0.044
plane fitting	(cm)	0.020	0.002	0.020	0.003	0.018	0.002	0.020	0.003	0.025	0.007

Table 7.4: Error analysis of the recovered 3D point cloud from real image data (Part 1). The preceding refractive calibration was performed without a known calibration object.

ERROR MEASURE		RTE 300		VIE 300		RRE 300		VRE 300		CY 300	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
reprojection	$left(pixel)$	0.057	0.009	0.062	0.012	0.058	0.013	0.057	0.008	0.160	0.081
	$right(pixel)$	0.058	0.010	0.065	0.013	0.059	0.013	0.058	0.010	0.169	0.087
3D	$x(cm)$	0.074	0.031	0.078	0.036	0.078	0.036	0.072	0.032	0.095	0.066
	$y(cm)$	0.041	0.009	0.031	0.013	0.036	0.002	0.038	0.007	0.038	0.019
	$z(cm)$	0.032	0.008	0.066	0.021	0.036	0.003	0.042	0.005	0.324	0.208
	$3D(cm)$	0.097	0.024	0.115	0.034	0.099	0.029	0.098	0.024	0.356	0.193
length measurement	$horizontal(cm)$	0.018	0.024	0.019	0.021	0.018	0.024	0.018	0.023	0.048	0.007
	$horizontal(\%)$	0.080	0.107	0.084	0.094	0.083	0.107	0.083	0.102	0.217	0.030
	$vertical(cm)$	0.047	0.017	0.052	0.017	0.047	0.016	0.048	0.017	0.033	0.008
	$vertical(\%)$	0.265	0.097	0.290	0.097	0.264	0.092	0.270	0.094	0.184	0.045
	$diagonal(cm)$	0.031	0.003	0.034	0.009	0.031	0.002	0.032	0.006	0.027	0.008
	$diagonal(\%)$	0.110	0.010	0.121	0.031	0.108	0.006	0.111	0.021	0.095	0.027
plane fitting	(cm)	0.019	0.003	0.024	0.004	0.018	0.003	0.020	0.003	0.024	0.007

ERROR MEASURE		RTE 400		VIE 400		RRE 400		VRE 400		CY 400	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
reprojection	$left(pixel)$	0.056	0.009	0.059	0.006	0.056	0.011	0.057	0.009	0.195	0.108
	$right(pixel)$	0.057	0.010	0.060	0.007	0.058	0.012	0.059	0.010	0.205	0.116
3D	$x(cm)$	0.073	0.029	0.070	0.029	0.076	0.032	0.074	0.030	0.122	0.068
	$y(cm)$	0.038	0.008	0.042	0.001	0.036	0.006	0.040	0.009	0.062	0.016
	$z(cm)$	0.041	0.015	0.058	0.012	0.036	0.013	0.037	0.013	0.392	0.216
	$3D(cm)$	0.099	0.021	0.105	0.027	0.099	0.024	0.098	0.022	0.424	0.213
length measurement	$horizontal(cm)$	0.018	0.023	0.020	0.021	0.017	0.023	0.017	0.023	0.052	0.008
	$horizontal(\%)$	0.080	0.102	0.088	0.095	0.077	0.106	0.078	0.104	0.236	0.036
	$vertical(cm)$	0.048	0.016	0.047	0.015	0.048	0.016	0.048	0.017	0.034	0.006
	$vertical(\%)$	0.269	0.088	0.264	0.087	0.271	0.089	0.269	0.094	0.193	0.032
	$diagonal(cm)$	0.030	0.005	0.029	0.008	0.031	0.004	0.031	0.004	0.030	0.011
	$diagonal(\%)$	0.105	0.018	0.102	0.029	0.108	0.014	0.110	0.016	0.105	0.037
plane fitting	(cm)	0.020	0.002	0.020	0.003	0.020	0.003	0.020	0.002	0.030	0.007

Table 7.5: Error analysis of the recovered 3D point cloud from real image data (Part 2). The preceding refractive calibration was performed without a known calibration object.

7.3 Application: Underwater Stereo 3D Reconstruction

In this section, the results from first experiments on underwater stereo 3D reconstruction will be presented. Prerequisite was a *SFRS* that was calibrated for its intrinsic, extrinsic and refractive parameters. Its refractive calibration was performed without calibration object, as described in Section 7.2.3. The following experiments were performed on real data, which were acquired with the aid of an aquarium setup that was similar to the one used before. The experiments were accompanied by qualitative evaluations and comprised the computation of correspondence curves, dense stereo matching and the recovery of 3D coordinates.

7.3.1 Correspondence Curves

In the presence of the prevailing conditions of underwater imaging, the computation of correspondence curves serves the reduction of the search space for pairs of corresponding image points during dense stereo matching. A qualitative evaluation of the proposed approach for the computation of correspondence curves was realized by visual examination. Therefore, the inner corners of a planar checker pattern were used as easily detectable image feature points. For each image feature point in the master camera, a correspondence curve in the image of the slave camera was computed by following the steps described in Section 6.1. The visual examination comprised to check if these correspondence curves hit the according inner corners of the checker pattern. Figure 7.24 shows two exemplary images of a checker pattern from the view of the slave camera and the computed correspondence curves. Since the computation of correspondence curves is a sampling of rays in water, a working range of depth values and a step width need to be defined. With its starting point at the water-sided interface, the user-defined working range between $0cm$ and $100cm$ covered the existing body of water in the aquarium very generously. The sampling of each ray was realized with a user-defined step width of $0.1cm$. As can be seen in Figure 7.24, all the corner points were hit by a correspondence curve. As was expected, the correspondence curves were no straight lines.

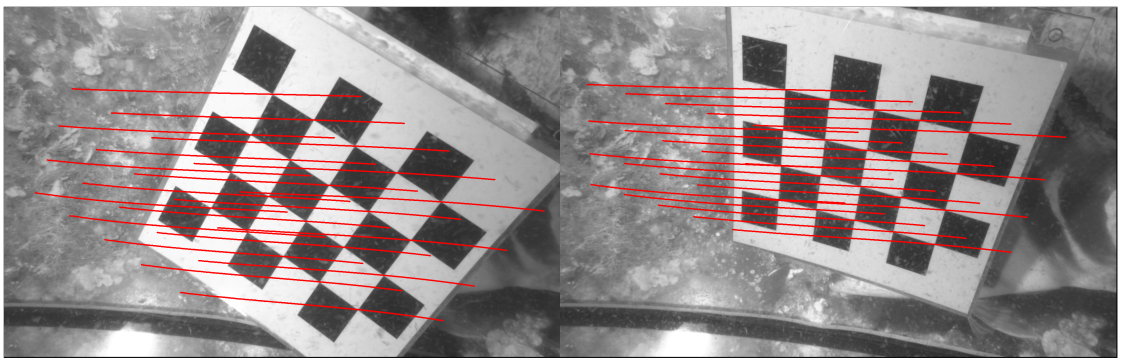

Figure 7.24: Computed correspondence curves for the inner corners of a checker pattern in two exemplary views.

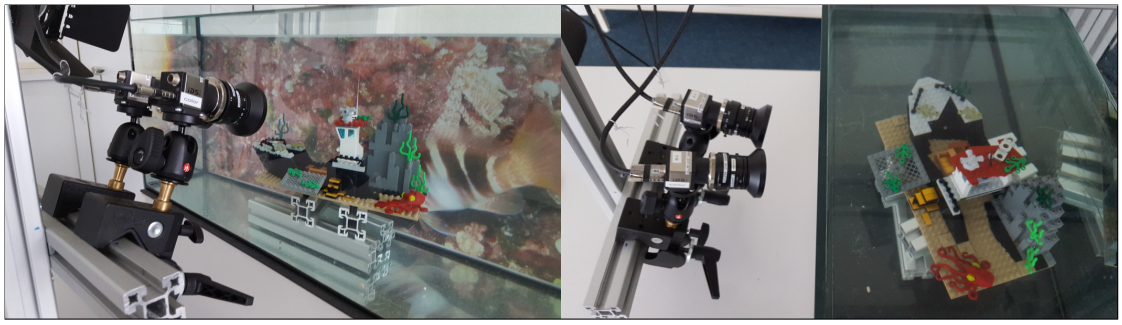

Figure 7.25: Aquarium setup for underwater stereo 3D reconstruction.

7.3.2 Dense Stereo Matching & Recovery of 3D Coordinates

Setup. The aquarium setup that was used to generate the image data for underwater stereo 3D reconstruction is shown in Figure 7.25. The aquarium was equipped with an image of an underwater scene as background. Besides this background, the scene contained a submerged *LEGO*® model of a shipwreck. The advantage of a *LEGO*® model is that it is easily possible to create an equivalent digital model. Figure 7.26 shows an image of the scene from the view of the master camera and a rendered view of the associated digital *LEGO*® model. This digital model is based on *LDraw*™[2], which is an open standard for *LEGO*® CAD programs, such as *MLCad*[3]. These *LEGO*® CAD programs enable the user to create digital *LEGO*® models and scenes. Therefore, *LDraw*™ provides a part library, which is utilized for the creation of the digital model with *MLCad*. This digital model can be imported into the program *LDView*[4] for being rendered, or being exported to common CAD file formats like STL.

3D Point Cloud Computation. After the calibration of the *SFRS*, it is possible to realize a physically correct underwater stereo 3D reconstruction of an imaged scene. Within the conducted experiments, refractive calibration utilizing *VIE* was utilized. After refractive calibration, the approach for the computation of correspondence curves, which has been proposed in Section 6.1, was performed per pixel of the image of the master camera. In contrast to Section 7.3.1, the working range of depth values was adjusted to the known dimensions of the aquarium and thus was between $0cm$ and $40cm$, starting at the water-sided interface. The step width amounted to $0.1cm$.
With the aid of the computed correspondence curves, dense stereo matching was performed. The method of choice was an implementation of block matching with the *Sum of Absolute Differences* (*SAD*). Block matching with *SAD* is a common practice for dense stereo matching. Summarized, a square-shaped neighborhood around a pixel in the image of the master camera forms a block that gets compared to all the blocks that are formed

[2] *LDraw*: `http://www.ldraw.org/`
[3] *MLCad*: `http://mlcad.lm-software.com/`
[4] *LDView*: `http://ldview.sourceforge.net/`

Figure 7.26: Real and virtual LEGO® model of a shipwreck.

around the pixels on the respective correspondence curve in the image of the slave camera. The absolute differences of the intensity values are computed pixel-wise and their sum is used as a metric of block similarity. Therefore, gray-scale images of the scene are used for the computations. The implemented dense stereo matching was realized with a block size of 21×21. Within the implemented approach, common problems of stereo matching, such as outlier removal or occlusions, were ignored for now. For each pixel in the image of the master camera, the center pixel of the block in the image of the slave camera with the lowest *SAD* value was chosen to create a pair of corresponding pixels. Only a user-defined threshold for the maximal *SAD* value was used, which decided if a matching was valid. If the lowest *SAD* value exceeded the used-defined maximal *SAD* value, the pair of corresponding pixels was simply ignored during further processing. This led to a few holes in the 3D point cloud, but it represents a first step to exclude obviously invalid matches. The pairs of corresponding pixels were used to recover the 3D coordinates of the scene with respect to the master camera's coordinate system. Therefore, the *Refractive Triangulation* form Section 6.2 was used.
The result of the performed underwater stereo 3D reconstruction was a colored 3D point cloud of the imaged scene. The program *CloudCompare*[5] was used for its visualization. Figure 7.27 shows two views of the recovered 3D point cloud. Compared to the background, the *LEGO*® model posed the problem that it did not provide a lot of texture. As was expected, many outliers from faulty matches occurred. Nevertheless, the top view on the left side of Figure 7.27 shows that the background was recovered very well and that it was planar as expected. In both views, it can be seen that the general structure of the scene could be recovered. The shipwreck is recognizable and can be clearly distinguished from the background. Overall, the results of these first experiments were satisfying.

3D Point Cloud Evaluation. The recovered 3D point cloud of the shipwreck model was compared to a ground truth 3D point cloud that was generated from the digital *LEGO*® model. Therefore, with the use of *CloudCompare*, the STL file of the digital

[5] *CloudCompare*: `http://www.danielgm.net/cc/`

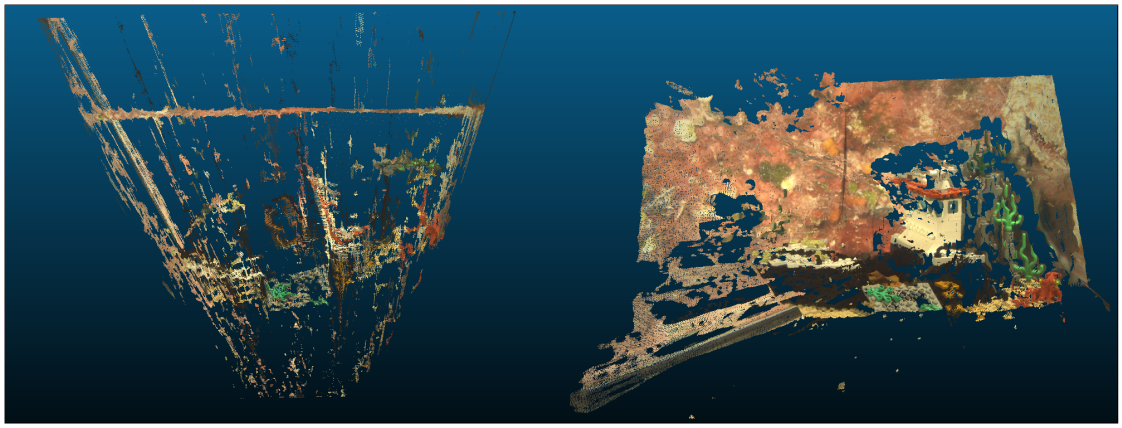

Figure 7.27: Two exemplary views of the recovered point cloud of the imaged scene.

model was imported and sampled to generate the ground truth 3D point cloud. For comparison with the ground truth 3D point cloud, the recovered 3D point cloud of the full scene was adjusted manually. Therefore, the background and the most obvious outliers were cropped by manual selection. Figures 7.28, 7.29 and 7.30 show the ground truth 3D point cloud, the recovered 3D point cloud after manual adjustment and an alignment of both of them from an arbitrary 3D view and an orthographic side and top view. The process of aligning 3D point clouds is known as registration. This registration was realized with the aid of *CloudCompare*. It enables the optimization of the registration of two 3D point clouds after manual selection of three corresponding pairs of 3D points. In terms of a qualitative evaluation by visual examination, the alignment of the 3D point clouds was satisfying. A quantitative evaluation was omitted due to the prematurity of the implemented approach.

Findings. The results of the implemented approaches for the computation of correspondence curves, for dense stereo matching and for the recovery of the 3D coordinates show that the reconstruction approach works in general. The results were plausible and

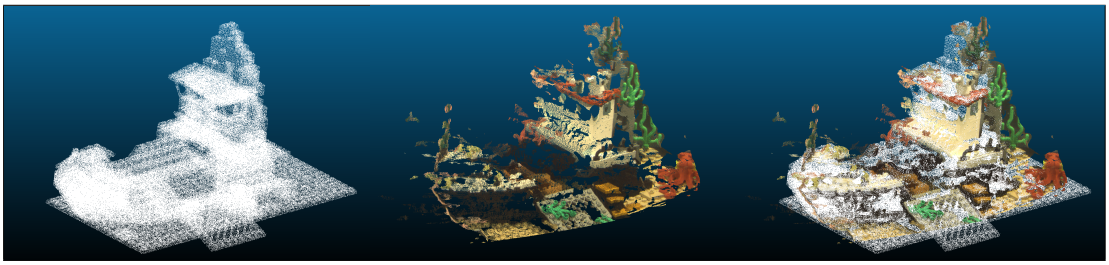

Figure 7.28: 3D view. Left: Ground truth 3D point cloud. Middle: Recovered 3D point cloud after manual adjustment. Right: Registration of both 3D point clouds.

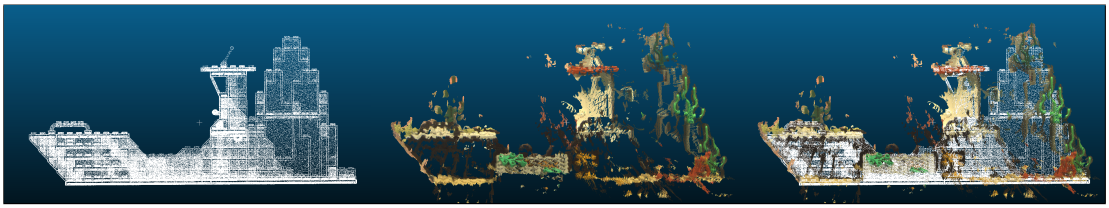

Figure 7.29: Side view. Left: Ground truth 3D point cloud. Middle: Recovered 3D point cloud after manual adjustment. Right: Registration of both 3D point clouds.

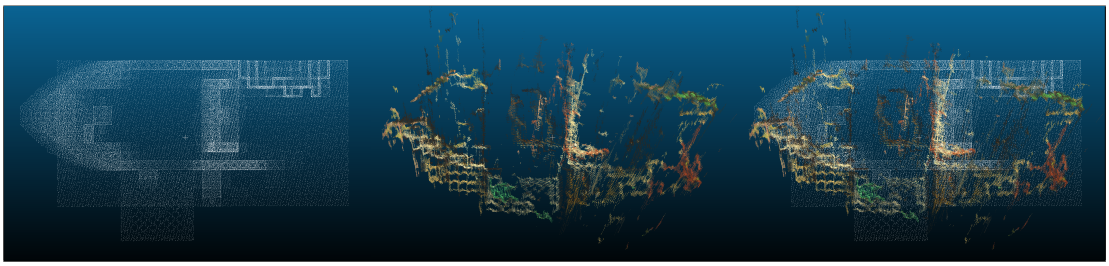

Figure 7.30: Top view. Left: Ground truth 3D point cloud. Middle: Recovered 3D point cloud after manual adjustment. Right: Registration of both 3D point clouds.

also satisfying in terms of the performed qualitative evaluation. The run-times of the computations are dependent on the user-defined working range, on the step width and most of all on the image resolution. Overall, the computations of these first experiments were time-consuming.

7.4 Summary

In this chapter, the evaluation of the consecutive and the simultaneous determination of refractive parameters and of all six cost functions (PAE, PFE, RTE, VIE, RRE and VRE) has been performed in multiple stages. *Simulated Test Data Stage 1* has shown that all six cost functions are valid and that their application in the proposed calibration strategies is working correctly. Simultaneously, it has been shown that the recovery of 3D coordinates by *Refractive Triangulation* from Section 6.2 is working correctly as well. The experiments with augmentation of the simulated test data with Gaussian noise and with variation of the refractive parameters pointed out the stability of the proposed approaches and that none of them failed completely. Based on the results, it has been easy to decide in favor of the calibration strategy with simultaneous determination of the refractive parameters. In contrast, it has been hard to decide in favor of one of the four pattern-independent cost functions RTE, VIE, RRE and VRE, since they performed similarly well. The reimplementation of the reference approach of Chen and Yang [CY14] (CY) ranked among them, without performing significantly better or worse. Regarding the run-times of the single approaches, a clear order became apparent.

Besides establishing a workflow and evaluation criteria that can be adapted to real world setups, *Simulated Test Data Stage 2* has shown that refractive calibration in combination with the pre-process comprising stereo calibration and feature detection works. All of the proposed refractive calibration approaches performed well on rendered images. This stage has been used for comparisons to alternative calibration strategies from the literature. The estimation (*EST*) of the pose of the master camera by a pose from 3D-2D point correspondences algorithm performed equally well as the proposed approaches. The analysis of the absorption strategy (*ABS*) has shown that the inaccuracies of the recovered 3D coordinates are always significant. It could be confirmed that the absorption capabilities decrease with non-perpendicularity between image sensor and interface normal, as has been already stated in the literature (Section 3.2). Due to its severe inaccuracies, the strategy to disregard (*DIS*) refractive effects performed worst overall and should therefore be avoided.
It has been shown that the approach for the generation of benchmark data (*BD*), which has been presented within *Real Test Data Stage 1*, is valid. The results form *Real Test Data Stage 1* matched the tendencies of the simulation results quite well. The insufficient capabilities of *ABS* and *DIS* to handle refractive effects were substantiated. The evaluation also substantiated that the simultaneous strategy utilizing the pattern-independent *RTE*, *VIE*, *RRE* or *VRE* is preferable over the consecutive strategy utilizing the pattern-based *PAE* or *PFE* and even over the reference approach *CY*. Of these four pattern-independent cost functions, none is standing out, since all performed equally well. Differences that had been evident during the evaluation with simulated data were not evident any more. However, compared to the simulation results for a similar setup, the resulting inaccuracies of the proposed approaches increased. Therefore, the evaluation has shown that there is an accumulation of errors. The overall error consists of errors from the pre-process containing common approaches for stereo calibration and feature detection and from the process of refractive calibration.
The evaluation in *Real Test Data Stage 2* has shown that it is possible to perform an accurate refractive calibration by the proposed approaches without the need of a known calibration object. Only the simultaneous determination of refractive parameters with the four pattern-independent cost functions *RTE*, *VIE*, *RRE* and *VRE* and the reference approach *CY* have the necessary capabilities. All four performed equally well as during the evaluation with a calibration object and the reference approach *CY* was outperformed by them in almost every case. Due to their nearly equivalent performance, the run-times of the single approaches become a factor. Based on the run-times that were measured during simulation, the following order has been accepted as the suggested order of application: *VIE*, *RRE*, *RTE* and *VRE*.
In the last part of this chapter, the results from first experiments on underwater stereo 3D reconstruction have been presented. These consisted of the application of the proposed approaches for refractive calibration. It has been shown that the reconstruction approach comprising the proposed computation of correspondence curves, of dense stereo matching and of the recovery of 3D coordinates works in general. The results were plausible and also satisfying in terms of the performed qualitative evaluation.

8 Conclusion

8.1 Summary

The aim of this thesis was to recover accurate 3D information from underwater images, with a stereo camera being the imaging system of choice. In air, the pinhole camera model is the very foundation of stereo 3D reconstruction. However, the refraction of light invalidates the pinhole camera model for underwater image formation. This leads to the requirement of suitable extensions for the recovery of 3D structure from underwater images. The underlying model for image formation, which has been used in this thesis, is the *SFRS*. Derived from an in-depth review of the related literature, a taxonomy and a parametrization of this and further image formation models for imaging systems exposed to refraction have been proposed. Although being more specialized than single camera systems, the *SFRS* it can be designated as a representative of a common viewing condition. A characteristic of the *SFRS*, which has been taken advantage of in this thesis, is that the refractive parameters have to be estimated just with respect to the master camera. It has been shown that the particularities of this model are beneficial and worth to adjust the basic design of the imaging system accordingly. To achieve the aim of this thesis, three research goals had been identified. These read as follows:

R1) Develop a model for handling refractive effects in all concepts of stereo 3D reconstruction.

R2) Develop approaches for the refractive calibration of a *SFRS*s.

R3) Relax or even eliminate the need to handle special calibration objects in underwater environments during the refractive calibration of a *SFRS*.

These research goals have been addressed by four contributions:

C1) A *Virtual Object Point (VOP)* model was introduced, which enables the extension of the concepts of stereo 3D reconstruction in air for underwater environments.

C2) Various approaches for the refractive calibration of a *SFRS*, which differ due to their capabilities to work with or without the necessity of a special calibration object, were proposed, analyzed and compared.

C3) The proposed approaches for the refractive calibration of a *SFRS* were designed so that the determination of the refractive parameters is partly realizable by linear optimization, which increases their efficiency.

C4) An approach for benchmark data generation for real test data was proposed, which can be used to evaluate approaches for the refractive calibration of a *SFRS*.

With these contributions, the research goals have been solved. In the following, with the research goals as the starting point, the results of this thesis will summarized briefly.

R1. To solve the first research goal, a *VOP* model (**C1**) was developed, which enables handling refractive effects properly. The introduced *VOP* model is based on simple trigonometric rules and describes how to non-ambiguously relate *VOP*s to real object points and vice versa in the case of two flat refractive interfaces. The various scopes of application of the *VOP* model make it the cornerstone for the fulfillment of the research goals.
The most important concepts of stereo 3D reconstruction are refractive projections for coordinate transformations between 3D and 2D spaces. Within a refractive system, these are refractive forward projection and refractive back-projection. Compared to projections in air, both of them become more complex and computationally expensive operations, especially the refractive forward projection. There is no approach to compute the refractive forward projection of 3D points onto the image plane directly. The computational intensity is especially inconvenient since both projections are basic operations on which the concepts of stereo 3D reconstruction, such as refractive calibration and recovery of 3D coordinates, depend on. Since it is readily integrable within the image formation models of monocular and binocular vision, the *VOP* model could be used to implement four possibilities for coordinate transformations between 3D and 2D spaces, namely:

- a simplification of refractive forward projection by means of an iterative solution,
- a simplification of refractive back-projection by means of avoiding a full ray tracing,
- an alternative to refractive forward projection by definition and implementation of a virtual camera model that enables conventional forward projection, and
- an alternative to refractive back-projection, which is based on the integration of the *VOP* model into the model for a *SFRS*.

The integration of the *VOP* model into the underlying image formation models not only benefits the computation of refractive projections, but also the computation of adapted and novel cost functions, the definition of optimization problems and therefore, all aspects of refractive calibrations. Besides refractive calibration, the *VOP* model enables the necessary extensions of all the concepts of stereo 3D reconstruction in air for underwater environments. Similarly to the pinhole camera model, epipolar geometry becomes invalid under water and due to the 3D-dependence of refractive effects, epipolarlines are not straight lines any more. In the context of dense stereo matching, the proposed approach for the computation of correspondence curves represents a reduction of the search space and is an alternative to epipolar geometry and rectification. However, it is computationally expensive. Furthermore, two approaches have been proposed for the actual recovery of 3D coordinates from pairs of corresponding image points in the left an right view of the stereo camera. Due to its low computational effort, *Virtual Intersection* is suitable during the proposed refractive calibration approaches. Since the *Virtual Intersection* shows some shortcomings, *Refractive Triangulation* is better suited for recovering 3D coordinates of a full scene.

R2. To solve the second research goal, various approaches for the refractive calibration of a *SFRS* (**C2**) were developed and analyzed in depth. Furthermore, to make the solution to **R2** efficient, the determination of the refractive parameters was partly realized by linear optimization (**C3**). The following prerequisites have to be satisfied for the current implementation:

- The imaging setup can be characterized by the model for a *SFRS*.
- There have to be two interfaces and the thickness of the glass layer is known.
- The refractive indices of the participating media (air, glass and water) are known.
- The stereo camera with constant intrinsic and extrinsic parameters is calibrated.

The requested refractive parameters are the air layer thickness and the orientation of the system axis. An initialization of these parameters by approximate values is not necessary. Compared to the related approaches from the literature, the developed ones are not restricted as follows:

- The image plane does not need to be perpendicular to the interface normal.
- There is no need for known coordinates of feature points on a calibration object.
- The number of interfaces and layers is extendable as long as the additional layer thicknesses and the refractive indices are known.

Two strategies for the determination of the refractive parameters of a *SFRS* have been proposed in this thesis. They are characterized primarily by their capabilities to determine the refractive parameters consecutively or simultaneously. Both strategies have in common that the air layer thickness is optimized by solving a set of linear equations. An optimization problem that can be solved linearly could not be found for the determination of the orientation of the system axis. Nevertheless, within both strategies, the system axis can be optimized iteratively by utilization of the proposed cost functions.

R3. To solve the third research goal, the various approaches for the refractive calibration of a *SFRS* (**C2**) were developed with different demands on the necessary calibration objects. These different demands are reflected in the two aforementioned strategies. Since the developed cost functions show a progressive independence from calibration objects, both calibration strategies do so as well. The consecutive determination of the refractive parameters needs a calibration object with its features arranged in straight lines and utilizes pattern-based cost functions. Within this strategy, the system axis determination is performed explicitly. The simultaneous strategy needs no calibration object at all. It comprises an implicit system axis determination and utilizes the proposed pattern-independent cost functions. Thereby, the restrictions due to the need of special calibration objects have been eliminated and the cumbersome handling of calibration objects under water can be avoided.

R2 & R3. Another contribution of this thesis (**C4**) has been the development of an approach for benchmark data generation. Since its main purpose has been to support

the evaluations with real test data, it addresses the research goals only indirectly. With the aid of a calibration object that was fixed to a stereo camera in a solid construction, a simple means for the evaluation of underwater stereo 3D reconstruction could be created. The overall results of the extensive evaluations of the various approaches for the refractive calibration of a *SFRS* on simulated and real test data is a suggested order of their application, which reads:

1.) *Virtual Intersection Error* (*VIE*)
2.) *Refractive Reprojection Error* (*RRE*)
3.) *Refractive Triangulation Error* (*RTE*)
4.) *Virtual Reprojection Error* (*VRE*)

This has been determined based on the following findings:

- The implicit system axis determination (simultaneous strategy) performs better than the explicit one (consecutive strategy) and simultaneously,
- the pattern-independent cost functions (without calibration object) perform better than the pattern-based cost functions *PAE* and *PFE* (with calibration object).
- In the case of pattern-independent cost functions, the 3D scene cost functions *RTE* and *VIE* and the 2D image cost functions *RRE* and *VRE* perform equally well.

An important finding had been that the orientation of the system axis of a *SFRS* can be determined independently of the layer thicknesses of the refractive system. However, the performed experiments have shown that the simultaneous determination works more accurately. Since the four pattern-independent cost functions and therefore the four possibilities for coordinate transformations between 3D and 2D spaces performed equally well during the conducted experiments, the run-times of the single approaches were used to determine the suggested order of application. Therefore, as was expected, the newly defined cost function *VIE* has been the most computationally efficient.
Besides having been analyzed and compared amongst themselves, the proposed strategies have been compared to alternative strategies from the literature. During the conducted experiments with real test data, these four performed better than the adapted reimplementation of the chosen reference approach of Chen and Yang [CY14]. The major difference between this thesis and the reference approach is the range of the developments. While the authors of the reference approach concentrate on one aspect of refractive calibration, namely the development of an optimization problem, the developments in this thesis are more extensive, since they comprise all the aspects of refractive calibration and also the remaining concepts of stereo 3D reconstruction.

Concluding Remarks. The final result of the developed approaches for underwater stereo 3D reconstruction are 3D point clouds of the imaged scene with a metric scale, which have been created by explicitly considering refractive effects in every computation step. As has been shown throughout this thesis, the proposed *VOP* model integrates very well in each of these steps and benefits all of them. Furthermore, the underlying principles of the necessary computation steps can be reused frequently between refractive

calibration, correspondence curve computation and recovery of 3D coordinates. The integration of the *VOP* model can be considered as a unique feature, which distinguishes the developments in this thesis from the related works in the literature.
From an application point of view, the developed approaches can be utilized within the creation of accurate, metric 3D models of underwater objects. The entirety of such a measurement campaign gets favored by the developed approaches due to avoiding the cumbersome and time-consuming handling of a calibration object under water.

8.2 Future Works

A major assumption for the experiments with real test data has been that camera lens distortion, which is independent of the imaged scene, and refractive distortion, which is highly dependent on it, do not influence each other and hence are separable. Therefore, camera lens distortion has been eliminated from the underwater images throughout the conducted experiments immediately after calibration of the stereo parameters. Since no significant anomalies occurred, no contradiction has been noticed. However, this topic should be considered in a more in-depth analysis with real test data and with a variety of different camera lenses.
Within the proposed approaches for underwater stereo 3D reconstruction, there is also room for improvements. One example is the applied testing scheme for the refractive calibration approaches with feature detection and feature description by *SIFT* [Low99] and feature matching using the *FLANN* [ML09], which is a rather simple, repeatedly performed random selection of a defined number of feature pairs. Although the utilized combination delivered satisfactory results, any other combination of algorithms for feature detection, description, matching and selection could have been used. Therefore, faster feature detectors that *SIFT* would be a reasonable alternative to increase the performance, as would be a more sophisticated selection approach. Other examples are the computation of correspondence curves and the method for dense stereo matching. Since one correspondence curve needs to be computed per image point, the proposed approach is already computationally expensive. This is due to the refractive forward projection that needs to be performed repeatedly in an interval that is defined by a sampling step width. Although the computational expense of the refractive forward projection has been relaxed in this thesis, the uniform step width becomes a problem. The depth resolution of a stereo system decreases with an increasing object distance. Therefore, on the one hand, when object point are further away, uniform sampling can lead to a mapping of multiple object points to the same pixel on the correspondence curve and on the other hand, when object points are close, it can lead to pixels on the correspondence curve that are skipped, which in turn requires interpolations. Therefore, a non-uniform sampling or a sampling that starts in the pixel space of the resulting correspondence curves could be beneficial. The method of choice for dense stereo matching has been block matching with the *SAD*. Within the implemented approach, common problems of stereo matching, such as outlier removal or occlusions, were ignored. Although these first experiments on underwater stereo 3D reconstruction delivered plausible results, correspondence curve computation

and stereo matching leave the most room for improvements. Besides this, the conducted qualitative evaluation should be accompanied by a quantitative evaluation.
As described at the beginning of this thesis, there are various computer vision techniques for 3D reconstruction. In this thesis, it has been shown how to integrate the proposed *VOP* model into the concepts of stereo 3D reconstruction. Thereby, the necessary extensions for the prevailing conditions of underwater imaging have been realized. Future works could comprise the analysis of its integrability into further techniques for 3D reconstruction. These could be passive methods, such as underwater *SFM*, or active methods, such as underwater structured light, which replace the second camera with a laser or a projector.

Appendices

A Number of Features

Accuracy Improvement versus Run-Time. In this thesis, three distinct setups were analyzed with regard to a joint number of feature points, which offers a reasonable compromise between run-time and accuracy improvement of the approaches for the refractive calibration of a *SFRS*. Another aspect was that the determined number of feature points should match a feasible number of feature points to be utilized during the experiments with real data. Therefore, both of the proposed calibration strategies in combination with the six cost functions were executed for the determination of the number of feature points. The refractive parameters were determined repeatedly. One stereo image pair with format-filling feature points was generated in every repetition. 100 repetitions were performed for every tested number of feature points. Each repetition included randomly generated Gaussian noise that was added to the feature points. The noise level had a standard deviation of $\sigma = 0.5$. The 3D point cloud, which was used to generate the image feature points, was slightly rotated randomly in every repetition. By doubling the number of tested points, a range from 50 to 1600 feature points was covered. This amounted to a time-consuming process.
The following two criteria were utilized for the determination of the number of feature points:

- Angle Error: The difference between the true system axis and the calibrated system axis in degrees. The angle was measured between the normalized direction vectors of both axes.
- Air Layer Thickness Error: The difference between the true air layer thickness and the calibrated air layer thickness in centimeters.

For each of the six cost functions, an error curve was produced for both criteria. Each visualization contains a mapping of the respective error values with regard to the conducted mean run-time of the respective algorithm. These visualizations can be seen in Figures A.1, A.2, A.3, A.4, A.5 and A.6. None of them shows a linear relationship between accuracy improvement and run-time. The doubling of the amount of feature points does not imply a doubling of the accuracy improvement. As can be seen, the largest accuracy improvements happen during the early steps between 50 and 400 feature points. Afterward, the curve progression becomes significantly flatter in most cases. Regarding the run-times, the doubling of the amount of feature points implies roughly a doubling of the run-times. Therefore, and since it was a goal to find a compromise that keeps the necessary time for performed evaluations in this thesis within days and not weeks, the chosen number of features was 400.

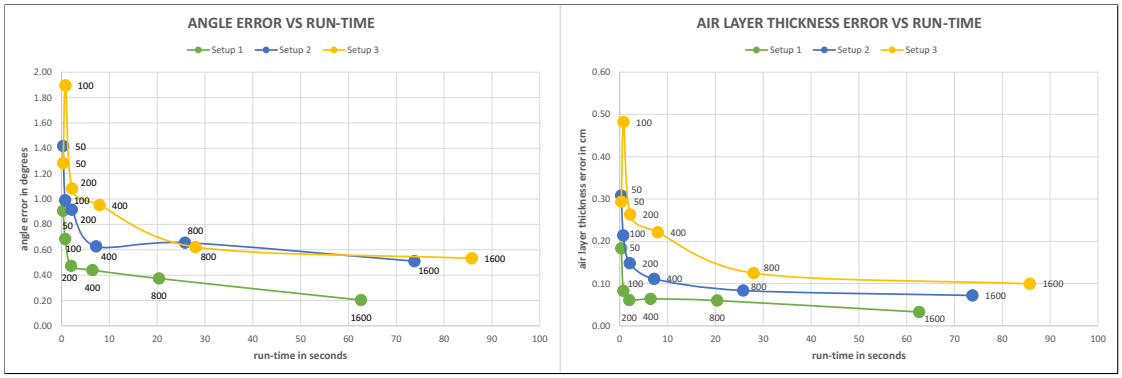

Figure A.1: *PAE*: Accuracy improvement versus run-time.

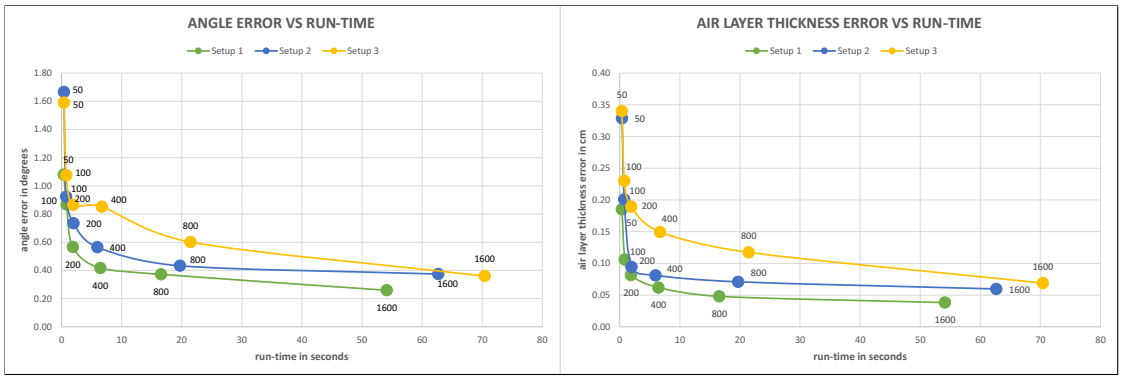

Figure A.2: *PFE*: Accuracy improvement versus run-time.

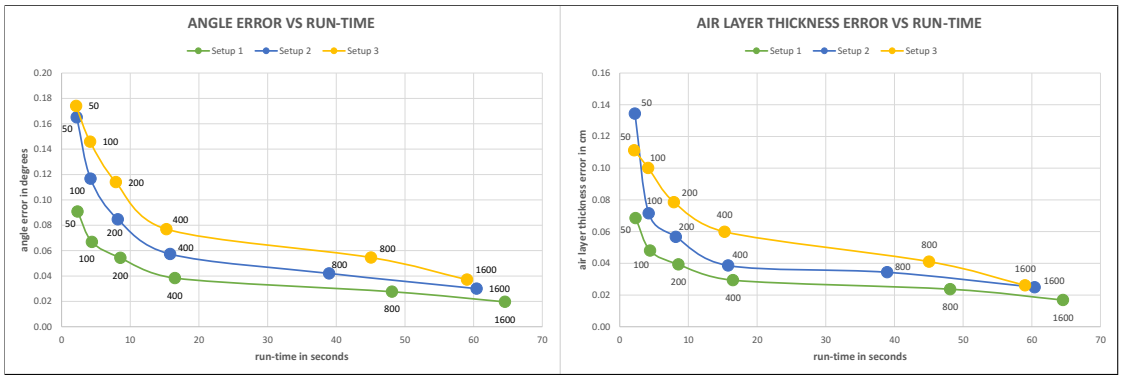

Figure A.3: *RTE*: Accuracy improvement versus run-time.

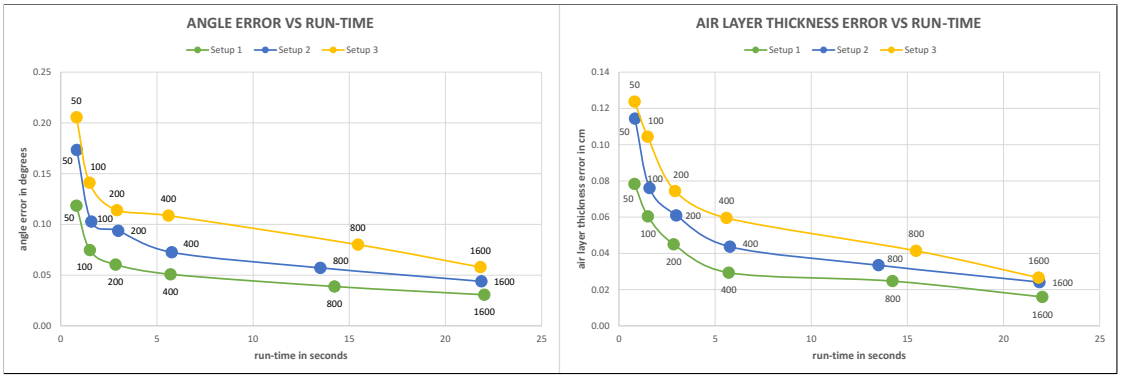

Figure A.4: *VIE*: Accuracy improvement versus run-time.

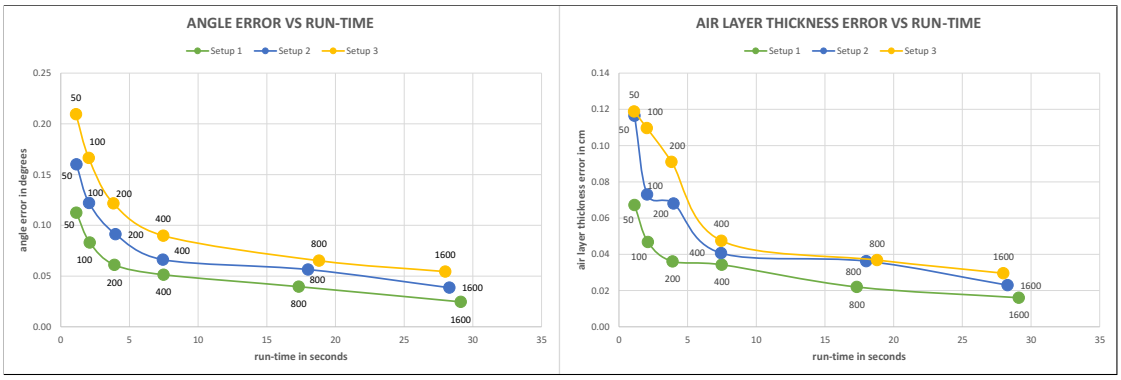

Figure A.5: *RRE*: Accuracy improvement versus run-time.

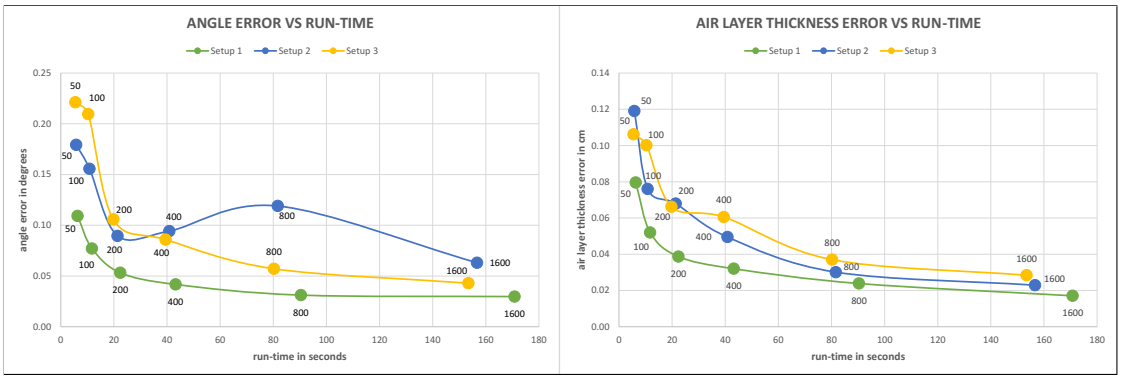

Figure A.6: *VRE*: Accuracy improvement versus run-time.

B Simulated Test Data Stage 1 - Supplementary Setups

The conducted experiments in *Simulated Test Data Stage 1* in Section 7.1.1 have been comprising six distinct setups, which were chosen according to the subsequent experiments with real data. These setups have been characterized by simulated 3D point clouds of a checker pattern at a distance of about $25cm$ to the water-sided interface. This section will cover a brief analysis of some further setups. In contrast to the previous setups, these do not show symmetric properties. The main purpose of this evaluation is to show the validity of the proposed approaches in the case of simulated 3D point clouds of a checker pattern at distances significantly greater than $25cm$.

Setups. Three setups were generated for a constant air layer thickness, which amounted to $t_a = d(C, S_1) = 5cm$ (Compare with Figure 7.1). To form a maximally overlapping field of view at the location of the 3D point cloud, cameras with converging principal axes p and p' were necessary. By definition of $\phi' = -\phi$ and $\theta' = \theta$, the basic setup for the conducted experiments was generated with $\phi = -3.5^\circ$ and $\theta = 0^\circ$. To avoid perfect symmetry, random offsets up to one degree were added to the values of ϕ', θ' and θ in each iteration. As before, the noise levels ranged from a standard deviation of $\sigma = 0$ to $\sigma = 1$, with a step width of 0.2 and the same evaluation criteria were utilized. The evaluation was realized in 100 iterations per noise level, each with randomly generated Gaussian noise and a slight, random rotation of the underlying 3D point cloud. To evaluate the proposed approaches, the distance d to the simulated 3D point cloud of the checker pattern and the stereo baseline b were varied. The variations read as follows:

- 1. Setup: $d = 100cm$, $b = 10cm$,
- 2. Setup: $d = 200cm$, $b = 20cm$,
- 3. Setup. $d = 500cm$, $b = 50cm$.

Findings. The error values almost equal to zero in all three categories at noise level $\sigma = 0$ in Figure B.1 show that all of the proposed cost functions continue to be valid and that the approaches were working correctly. However, the *PFE* failed at noise levels greater than $\sigma = 0.2$, which makes it the most error-prone. Apart from the *PFE*, the remaining cost functions showed a similar behavior. The values of the air layer thickness error and of the 3D error increased constantly with increasing noise and increasing distances of the 3D point cloud. The values of the angle error, however, only increased constantly with increasing noise and vary with increasing distances. Thus, the air layer thickness error seemes to be more decisive for the overall 3D error.

Figure B.1: Analysis of the inaccuracies of refractive calibration with *PAE*, *PFE*, *RTE*, *VIE*, *RRE* and *VRE* for supplementary setups.

The *PAE* was more stable than the *PAE*, but the error values were significantly higher than the ones of the pattern-independent cost functions *RTE*, *VIE*, *RRE* and *VRE*. All four pattern-independent cost functions showed similar values for their air layer thickness and 3D error. An outcome of this brief analysis is that refractive calibration is preferably performed with image features of close objects. The resulting refractive parameters can in turn be used for the recovery of 3D coordinates of distant objects. This approach is supposed to reduce the 3D error and there is no principled reason against it. However, a well-founded evaluation has not been conducted in this thesis.

C Effects of Inaccurate Refractive Indices

All of the described underwater imaging systems are characterized by a transition of the light rays through media with differing refractive indices. In general, this is a water-glass-air transition, of which the refractive index of the protective glass medium is generally known from the manufacturer. As has been concluded in Section 3.3.2, the best reason to omit the determination of the refractive indices of air and water during refractive calibration is that they are commonly known. Nevertheless, according to Mobley [Mob94], the index of refraction of water changes about 2% in the whole relevant range of water bodies. Therefore, the commonly presumed refractive index of water, with a value of 1.33, can be inaccurate. The changes depend on the optical wavelength as well as water temperature, salinity and depth. Significant changes are most likely to occur between locally separate water bodies, but may also happen rapidly in a single water body on a thermocline. A device for the measurement of the refractive index is the refractometer. It can provide reliable values for water bodies with constant properties and presumed values for water bodies where changes may occur.
This section will comprise a simulation of the effects of inaccurate values of the presumed refractive index of water. Therefore, the same evaluation criteria as in *Simulated Test Data Stage 1* in Section 7.1.1 were utilized. The simulations were performed for nine distinct setups. These were the six from Section 7.1.1 and the three from Appendix B. As before, the parametrization comprised the angles $\{\phi, \theta\}$, which represent the orientation of the system axis s_0 with respect to the master camera, the air layer thickness t_a and the distance d to the simulated point cloud. Within the simulations, the presumed refractive index of water amounted in each case to 1.33 and the true value ranged from 1.32 to 1.34. The results can be seen in Figures C.1, C.2, C.3, C.4, C.5, C.6, C.7, C.8 and C.9.

Findings. In terms of the accuracy of the recovered 3D point clouds, the results of the simulations showed that all of the proposed approaches for the refractive calibration of a *SFRS* and the reference approach behaved very similar. An outcome of this brief analysis is that refractive calibration is preferably performed with image features of close objects, if the presumed refractive index of water can be expected to be very inaccurate. The resulting refractive parameters can in turn be used for the recovery of 3D coordinates of distant objects. As in Appendix B, this approach is supposed to reduce the 3D error and there is no principled reason against it. However, a well-founded evaluation has not been conducted in this thesis.

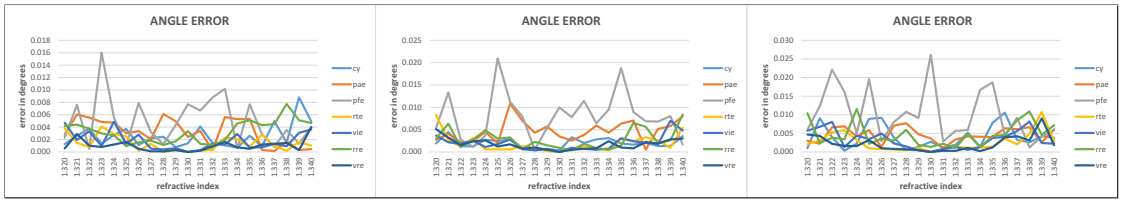

Figure C.1: Angle error for a setup with $\phi = -10^\circ, \theta = 0^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$.

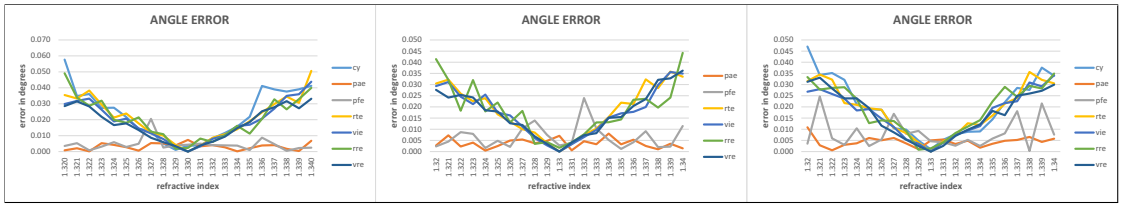

Figure C.2: Angle error for a setup with $\phi = -10^\circ, \theta = -20^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$.

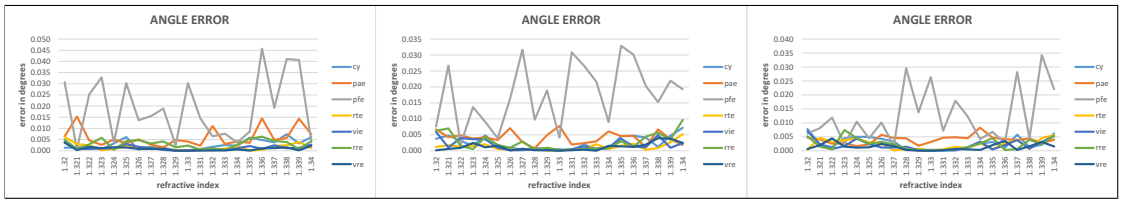

Figure C.3: Angle error for a setup with $\phi = -10^\circ, \theta = -20^\circ, t_a = 5cm$ and varying distance to the simulated point cloud d. Left: $d = 100cm$. Middle: $d = 200cm$. Right: $d = 500cm$.

Angle Error. With regard to this evaluation criterion, the results showed the greatest variability. However, the error values were very small overall (See Figures C.1, C.2, C.3). They only showed significant effects of the varying refractive index in the case of the strongly tilted setups in Figure C.2. In contrast to the others, the *PAE* and the *PFE* are a part of the consecutive strategy for the determination of the refractive parameters. The difference of the strategies is reflected in the diagrams. This can be most clearly seen in Figure C.2, since the *PAE* and the *PFE* seemingly got affected the least. Apart from the *PAE* and the *PFE*, the diagrams indicate a symmetric curve characteristic.

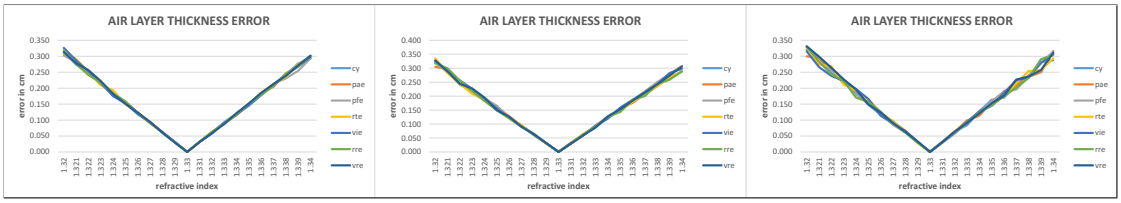

Figure C.4: Air layer thickness error for a setup with $\phi = -10^\circ, \theta = 0^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$.

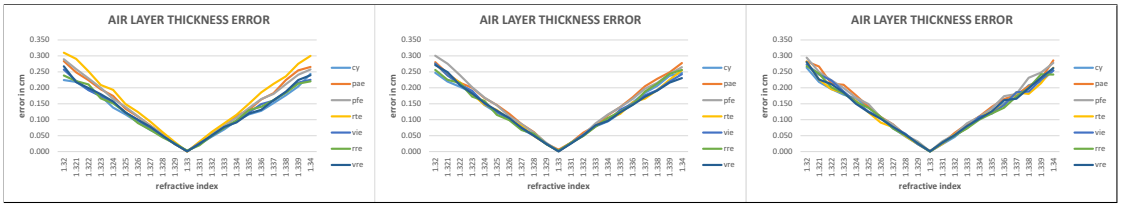

Figure C.5: Air layer thickness error for a setup with $\phi = -10^\circ, \theta = -20^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$.

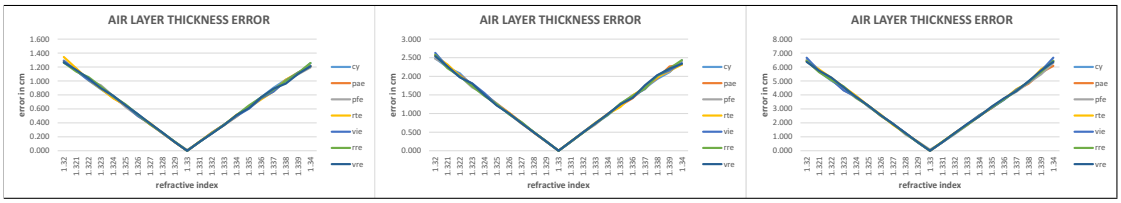

Figure C.6: Air layer thickness error for a setup with $\phi = -10^\circ, \theta = -20^\circ, t_a = 5cm$ and varying distance to the simulated point cloud d. Left: $d = 100cm$. Middle: $d = 200cm$. Right: $d = 500cm$.

Air Layer Thickness Error. With regard to this evaluation criterion, the results showed very little variances. In all the diagrams (See Figures C.4, C.5, C.6), a symmetric curve characteristic becomes apparent, as was expected. With an increasing distance to the simulated point cloud d, the error values became significant. This indicates that refractive calibration should be preventively performed with image features of close objects, if the presumed refractive index of water can be expected to be very inaccurate.

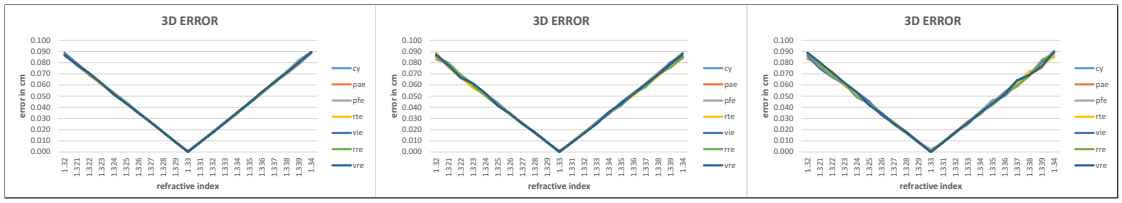

Figure C.7: 3D error for a setup with $\phi = -10^\circ, \theta = 0^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$.

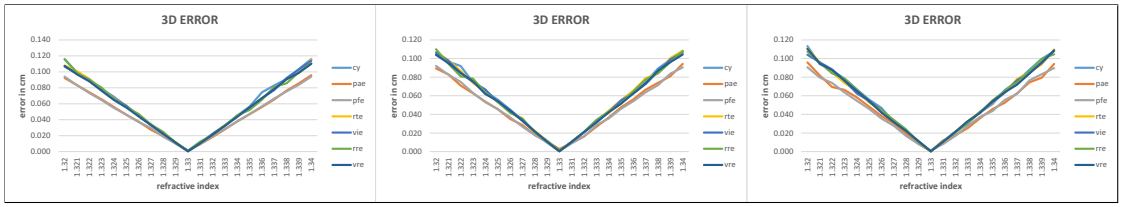

Figure C.8: 3D error for a setup with $\phi = -10^\circ, \theta = -20^\circ, d = 25cm$ and varying air layer thickness t_a. Left: $t_a = 5cm$. Middle: $t_a = 15cm$. Right: $t_a = 25cm$.

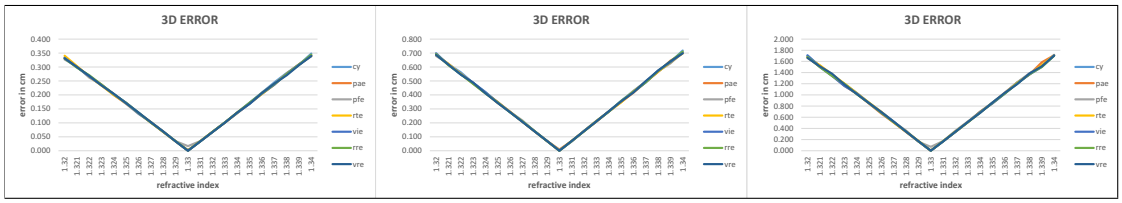

Figure C.9: 3D error for a setup with $\phi = -10^\circ, \theta = -20^\circ, t_a = 5cm$ and varying distance to the simulated point cloud d. Left: $d = 100cm$. Middle: $d = 200cm$. Right: $d = 500cm$.

3D Error. With regard to this evaluation criterion, the results showed very little variances as well. In all the diagrams (See Figures C.7, C.8, C.9), a symmetric curve characteristic becomes apparent, as was expected. As a consequence of the lower angle errors of the *PAE* and the *PFE* in the case of the strongly tilted setups in Figure C.2, the diagrams in Figure C.8 show the same trend. The results of the simulations support that refractive calibration should be preventively performed with image features of close objects, if the presumed refractive index of water can be expected to be very inaccurate.

D Simulated Test Data Stage 2 - Supplementary Setups

For the sake of completeness and in accordance with the supplementary setups for *Simulated Test Data Stage 1* in Appendix B, supplementary setups for *Simulated Test Data Stage 2* will be analyzed briefly in this section. Therefore, rendered image data, which was generated with *Blender*, was used. In accordance with appendix B, three setups were generated for a constant air layer thickness, which amounted to $t_a = 5cm$. To form a maximally overlapping field of view at the location of the 3D point cloud, cameras with converging principal axes p and p' were necessary. By definition of $\phi' = -\phi$ and $\theta' = \theta$, the basic setup for the conducted experiments was generated with $\phi = -3.5^\circ$ and $\theta = 0^\circ$. To evaluate the proposed approaches, the distance d to the simulated 3D point cloud of the checker pattern and the stereo baseline b were varied. The results can be seen in Tables D.1, D.2 and D.3.

ERROR MEASURE		EST	PAE	PFE	RTE	VIE	RRE	VRE	CY	DIS	ABS
reprojection	$left(pixel)$	0.018	0.019	0.050	0.018	0.019	0.018	0.018	0.018	0.827	0.216
	$right(pixel)$	0.018	0.019	0.050	0.018	0.019	0.019	0.018	0.018	0.841	0.218
3D	$x(cm)$	0.018	0.035	0.302	0.033	0.026	0.028	0.033	0.022	1.656	2.204
	$y(cm)$	0.012	0.016	0.044	0.012	0.014	0.014	0.012	0.013	0.439	0.175
	$z(cm)$	0.078	0.098	0.446	0.093	0.087	0.088	0.093	0.091	29.598	1.498
	$3D(cm)$	0.083	0.111	0.583	0.104	0.096	0.098	0.104	0.097	29.663	2.837
length measurement	$horizontal(cm)$	0.025	0.040	0.193	0.037	0.033	0.034	0.037	0.008	5.759	1.193
	$horizontal(\%)$	0.034	0.056	0.268	0.052	0.046	0.048	0.052	0.011	7.999	1.657
	$vertical(cm)$	0.014	0.020	0.074	0.019	0.017	0.018	0.019	0.010	1.594	0.976
	$vertical(\%)$	0.028	0.038	0.142	0.036	0.033	0.034	0.036	0.020	3.065	1.878
	$diagonal(cm)$	0.040	0.059	0.238	0.055	0.050	0.052	0.055	0.018	7.067	2.544
	$diagonal(\%)$	0.046	0.066	0.268	0.062	0.057	0.058	0.062	0.021	7.957	2.864
plane fitting	(cm)	0.044	0.046	0.066	0.045	0.045	0.045	0.045	0.044	2.748	0.705

Table D.1: Error analysis of the recovered 3D point cloud from simulated image data of a setup with $d = 100cm$, $b = 10cm$.

ERROR MEASURE		EST	PAE	PFE	RTE	VIE	RRE	VRE	CY	DIS	ABS
reprojection	$left(pixel)$	0.021	0.021	0.033	0.021	0.021	0.022	0.021	0.023	0.700	0.197
	$right(pixel)$	0.021	0.021	0.033	0.021	0.021	0.022	0.021	0.022	0.700	0.197
3D	$x(cm)$	0.032	0.090	0.308	0.087	0.049	0.108	0.087	0.122	3.517	4.786
	$y(cm)$	0.026	0.036	0.118	0.035	0.035	0.052	0.035	0.042	0.630	0.242
	$z(cm)$	0.190	0.181	0.566	0.181	0.201	0.177	0.181	0.206	57.094	2.940
	$3D(cm)$	0.197	0.217	0.706	0.215	0.217	0.229	0.215	0.261	57.235	5.798
length measurement	$horizontal(cm)$	0.049	0.063	0.120	0.063	0.069	0.059	0.063	0.036	8.327	1.223
	$horizontal(\%)$	0.039	0.050	0.095	0.050	0.055	0.047	0.050	0.028	6.608	0.971
	$vertical(cm)$	0.008	0.012	0.028	0.012	0.014	0.011	0.012	0.005	2.274	1.218
	$vertical(\%)$	0.009	0.013	0.030	0.013	0.015	0.012	0.013	0.005	2.499	1.338
	$diagonal(cm)$	0.009	0.027	0.098	0.027	0.035	0.022	0.027	0.008	10.343	3.242
	$diagonal(\%)$	0.006	0.017	0.063	0.018	0.023	0.014	0.018	0.005	6.655	2.086
plane fitting	(cm)	0.092	0.091	0.097	0.091	0.091	0.091	0.091	0.093	4.534	1.045

Table D.2: Error analysis of the recovered 3D point cloud from simulated image data of a setup with $d = 200cm$, $b = 20cm$.

ERROR MEASURE		EST	PAE	PFE	RTE	VIE	RRE	VRE	CY	DIS	ABS
reprojection	$left(pixel)$	0.020	0.022	0.470	0.020	0.020	0.020	0.020	0.020	0.853	0.210
	$right(pixel)$	0.020	0.022	0.618	0.020	0.020	0.020	0.020	0.020	0.839	0.209
3D	$x(cm)$	0.085	0.387	1.864	0.124	0.092	0.138	0.123	0.174	10.957	12.019
	$y(cm)$	0.052	0.121	0.185	0.056	0.052	0.058	0.056	0.059	1.916	0.577
	$z(cm)$	0.385	1.296	4.726	0.676	0.590	0.698	0.672	0.759	144.449	12.671
	$3D(cm)$	0.406	1.373	5.126	0.704	0.610	0.729	0.699	0.796	144.997	17.678
length measurement	$horizontal(cm)$	0.040	0.136	0.551	0.062	0.049	0.064	0.061	0.069	25.034	3.607
	$horizontal(\%)$	0.012	0.042	0.170	0.019	0.015	0.020	0.019	0.021	7.727	1.113
	$vertical(cm)$	0.041	0.007	0.050	0.015	0.016	0.015	0.015	0.016	7.105	3.317
	$vertical(\%)$	0.017	0.003	0.021	0.006	0.007	0.006	0.006	0.007	3.036	1.418
	$diagonal(cm)$	0.542	0.306	0.270	0.411	0.429	0.408	0.412	0.400	31.449	9.376
	$diagonal(\%)$	0.136	0.077	0.068	0.103	0.107	0.102	0.103	0.100	7.869	2.346
plane fitting	(cm)	0.196	0.244	0.636	0.204	0.200	0.205	0.203	0.208	12.632	2.883

Table D.3: Error analysis of the recovered 3D point cloud from simulated image data of a setup with $d = 500cm$, $b = 50cm$.

Bibliography

[Abd71] Y. I. Abdel-Aziz. „Direct linear transformation from comparator coordinates in close-range photogrammetry“. In: *American Society of Photogrammetry Symposium on Close-Range Photogrammetry. Falls Church (VA)., 1971*. 1971, pp. 1–19.

[AG15] P. Agrafiotis and A. Georgopoulos. „Camera Constant in the Case of Two Media Photogrammetry“. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40.5 (2015), p. 1.

[Agr+12] A. K. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari. „A theory of multi-layer flat refractive geometry“. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 3346–3353. DOI: 10.1109/CVPR.2012.6248073.

[Bea+11] C. Beall, F. Dellaert, I. Mahon, and S. B. Williams. „Bundle adjustment in large-scale 3D reconstructions based on underwater robotic surveys“. English. In: *OCEANS, 2011 IEEE - Spain*. 2011, pp. 1–6. ISBN: 978-1-4577-0086-6. DOI: 10.1109/Oceans-Spain.2011.6003631.

[Bel13] J. Belden. „Calibration of multi-camera systems with refractive interfaces“. en. In: *Experiments in Fluids* 54.2 (2013), pp. 1–18. ISSN: 0723-4864, 1432-1114. DOI: 10.1007/s00348-013-1463-0.

[BKN15] C. Bräuer-Burchardt, P. Kühmstedt, and G. Notni. „Combination of air-and water-calibration for a fringe projection based underwater 3D-scanner“. In: *Computer Analysis of Images and Patterns*. 2015, pp. 49–60. DOI: 10.1007/978-3-319-23117-4_5.

[BLJ84] A. A. Bartlett, R. Lucero, and G. O. Johnson. „Note on a common virtual image“. In: *American Journal of Physics* 52.7 (1984), pp. 640–643.

[Bos+15a] J. Bosch, N. Gracias, P. Ridao, and D. Ribas. „Omnidirectional Underwater Camera Design and Calibration“. In: *Sensors (Basel, Switzerland)* 15.3 (2015), pp. 6033–6065. ISSN: 1424-8220. DOI: 10.3390/s150306033.

[Bos+15b] J. Bosch, P. Ridao, D. Ribas, and N. Gracias. „Creating 360° Underwater Virtual Tours Using an Omnidirectional Camera Integrated in an AUV“. In: *OCEANS 2015 - Genova*. 2015, pp. 1–7. DOI: 10.1109/OCEANS-Genova.2015.7271525.

[Bou08] J. Y. Bouguet. *Camera Calibration Toolbox for Matlab*. 2008. URL: http://www.vision.caltech.edu/bouguetj/calib_doc/..

[Bra+07] V. Brandou, A. G. Allais, M. Perrier, E. Malis, P. Rives, J. Sarrazin, and P. M. Sarradin. „3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS“. English. In: *OCEANS 2007 - Europe*. 2007, pp. 1–6. ISBN: 978-1-4244-0635-7. DOI: 10.1109/OCEANSE.2007.4302315.

[Brä+15a] C. Bräuer-Burchardt, M. Heinze, I. Schmidt, P. Kühmstedt, and G. Notni. „Compact Handheld Fringe Projection Based Underwater 3D-SCANNER“. In: *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1 (2015), pp. 33–39.

[Brä+15b] C. Bräuer-Burchardt, M. Heinze, I. Schmidt, P. Kühmstedt, and G. Notni. „Underwater 3D Surface Measurement Using Fringe Projection Based Scanning Devices“. In: *Sensors* 16.1 (2015), p. 13. DOI: 10.3390/s16010013.

[Bra00] G. Bradski. „The OpenCV Library“. In: *Dr. Dobb's Journal of Software Tools* (2000).

[BS79] I. N. Bronstein and K. A. Semendjajew. *Taschenbuch der Mathematik*. Ed. by G. Grosche and V. Ziegler. BSB B. G. Teubner Verlagsgesellschaft, Nauka-Verlag, 1979. ISBN: 3871444928.

[BSH15] N. Boutros, M. R. Shortis, and E. S. Harvey. „A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology“. English. In: *Limnology and Oceanography: Methods* 13.5 (2015), pp. 224–236. ISSN: 1541-5856. DOI: 10.1002/lom3.10020.

[Car+03] M. Carreras, P. Ridao, R. Garcia, and T. Nicosevici. „Vision-based localization of an underwater robot in a structured environment“. In: *IEEE International Conference on Robotics and Automation, 2003. Proceedings. ICRA '03*. Vol. 1. 2003, pp. 971–976. DOI: 10.1109/ROBOT.2003.1241718.

[CC11] Y.-J. Chang and T. Chen. „Multi-View 3D Reconstruction for Scenes under the Refractive Plane with Known Vertical Direction“. In: *ICCV'11*. 2011, pp. 351–358. DOI: 10.1109/ICCV.2011.6126262.

[Che+11] Z. Chen, K.-Y. K. Wong, Y. Matsushita, X. Zhu, and M. Liu. „Self-Calibrating Depth from Refraction“. In: *2011 IEEE International Conference on Computer Vision (ICCV)*. 2011, pp. 635–642. DOI: 10.1109/ICCV.2011.6126298.

[Che+12] Z. Chen, K.-Y. K. Wong, Y. Matsushita, and X. Zhu. „Depth from Refraction Using a Transparent Medium with Unknown Pose and Refractive Index“. en. In: *International Journal of Computer Vision* 102.1-3 (2012), pp. 3–17. ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-012-0590-z.

[CS09] V. Chari and P. Sturm. „Multi-View Geometry of the Refractive Plane“. In: *Proceedings of the British Machine Vision Conference*. 2009, pp. 56.1–56.11. ISBN: 1-901725-39-1. DOI: 10.5244/C.23.56.

[CY14] X. Chen and Y.-H. Yang. „Two-View Camera Housing Parameters Calibration for Multi-layer Flat Refractive Interface“. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014, pp. 524–531. DOI: 10.1109/CVPR.2014.74.

[CY17] X. Chen and Y.-H. Yang. „A Closed-Form Solution to Single Underwater Camera Calibration Using Triple Wavelength Dispersion and Its Application to Single Camera 3D Reconstruction“. In: *IEEE Transactions on Image Processing* 26.9 (Sept. 2017), pp. 4553–4561. ISSN: 1057-7149. DOI: 10.1109/TIP.2017.2716194.

[DG16] A. Duda and C. Gaudig. „Refractive forward projection for underwater flat port cameras“. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Oct. 2016, pp. 2022–2027. DOI: 10.1109/IROS.2016.7759318.

[FCS05] R. Ferreira, J. P. Costeira, and J. A. Santos. „Stereo Reconstruction of a Submerged Scene“. English. In: *Pattern Recognition and Image Analysis*. Ed. by Jorge S. Marques, Nicolás Pérez de la Blanca, and Pedro Pina. Vol. 3522. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, pp. 102–109. ISBN: 978-3-540-26153-7. DOI: 10.1007/11492429_13.

[FF86] J. G. Fryer and C. S. Fraser. „On the Calibration of Underwater Cameras“. en. In: *The Photogrammetric Record* 12.67 (1986), pp. 73–85. ISSN: 1477-9730. DOI: 10.1111/j.1477-9730.1986.tb00539.x.

[FM11] P. Figueiredo and L. Machado. „Reconstruction Error of Calibration Volume's Coordinates for 3D Swimming Kinematics“. In: *Journal of Human Kinetics* 29 (2011), pp. 35–40. ISSN: 1640-5544. DOI: 10.2478/v10078-011-0037-6.

[GGY11] J. Gedge, M. Gong, and Y.-H. Yang. „Refractive Epipolar Geometry for Underwater Stereo Matching“. English. In: *2011 Canadian Conference on Computer and Robot Vision (CRV)*. 2011, pp. 146–152. ISBN: 978-1-61284-430-5. DOI: 10.1109/CRV.2011.26.

[Gla95] A. S. Glassner. *Principles of Digital Image Synthesis*. English. Morgan Kaufmann, 1995. ISBN: 978-0-08-051475-8.

[GN05] M. D. Grossberg and S. K. Nayar. „The Raxel Imaging Model and Ray-Based Calibration“. In: *International Journal of Computer Vision* 61.2 (2005), pp. 119–137. ISSN: 0920-5691. DOI: 10.1023/B:VISI.0000043754.56350.10.

[Gra80] S. I. Granshaw. „Bundle adjustment methods in engineering photogrammetry“. In: *The Photogrammetric Record* 10.56 (1980), pp. 181–207. DOI: 10.1111/j.1477-9730.1980.tb00020.x.

[GS00] G. Glaeser and H.-P. Schröcker. „Reflections on Refractions“. In: *Journal for Geometry and Graphics* Vol. 4.No. 1 (2000), pp. 1–18.

[Han+15] S. Haner, L. Svärm, E. Ask, and A. Heyden. „Joint Under and Over Water Calibration of a Swimmer Tracking System“. eng. In: *Proceedings of the 4th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2015)*. 2015, pp. 142–149. DOI: 10.5220/0005183701420149.

[Hat88] H. Hatze. „High-precision three-dimensional photogrammetric calibration and object space reconstruction using a modified DLT-approach“. In: *Journal of biomechanics* 21.7 (1988), pp. 533–538. DOI: 10.1016/0021-9290(88)90216-3.

[Hen+15] S. Henrion, C. W. Spoor, R. P. M. Pieters, U. K. Müller, and J. L. van Leeuwen. „Refraction corrected calibration for aquatic locomotion research: application of Snell's law improves spatial accuracy“. en. In: *Bioinspiration & Biomimetics* 10.4 (2015), p. 046009. ISSN: 1748-3190. DOI: 10.1088/1748-3190/10/4/046009.

[HL14] L. Huang and Y. Liu. „Camera Calibration for Plate Refractive Imaging System“. In: *2014 22nd International Conference on Pattern Recognition (ICPR)*. 2014, pp. 4068–4073. DOI: 10.1109/ICPR.2014.697.

[Höh71] J. Höhle. „Zur Theorie und Praxis der Unterwasser-Photogrammetrie“. PhD thesis. 1971.

[HS88] C. Harris and M. Stephens. „A combined corner and edge detector.“ In: *Alvey vision conference*. Vol. 15. 1988, p. 50. DOI: 10.5244/C.2.23.

[HS95] E. S. Harvey and M. R. Shortis. „A system for stereo-video measurement of sub-tidal organisms“. In: *Marine Technology Society Journal* 29.4 (1995), pp. 10–22.

[HS98] E. S. Harvey and M. R. Shortis. „Calibration Stability of an Underwater Stereo-Video System: Implications for Measurement Accuracy and Precision“. In: *Marine Technology Society Journal* 32.2 (1998), pp. 3–17.

[HZ04] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, 2004.

[Jas+08] P. Jasiobedzki, S. Se, M. Bondy, and R. Jakola. „Underwater 3D mapping and pose estimation for ROV operations“. English. In: *OCEANS 2008*. 2008, pp. 1–6. ISBN: 978-1-4244-2619-5. DOI: 10.1109/OCEANS.2008.5152076.

[JK12] A. Jordt-Sedlazeck and R. Koch. „Refractive Calibration of Underwater Cameras“. In: *Proceedings of the 12th European Conference on Computer Vision - Volume Part V*. ECCV'12. 2012, pp. 846–859. ISBN: 978-3-642-33714-7. DOI: 10.1007/978-3-642-33715-4_61.

[JK13] A. Jordt-Sedlazeck and R. Koch. „Refractive Structure-from-Motion on Underwater Images“. In: *2013 IEEE International Conference on Computer Vision (ICCV)*. 2013, pp. 57–64. DOI: 10.1109/ICCV.2013.14.

[Joh+10] M. Johnson-Roberson, O. Pizarro, S. B. Williams, and I. Mahon. „Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys“. en. In: *Journal of Field Robotics* 27.1 (2010), pp. 21–51. ISSN: 1556-4967. DOI: 10.1002/rob.20324.

[Kan+15] L. Kang, L. Wu, Y. Wei, and Z. Yang. „Theory of multi-level refractive geometry“. In: *Electronics Letters* 51.9 (2015), pp. 688–690. ISSN: 0013-5194. DOI: 10.1049/el.2014.4369.

[Kan+17] L. Kang, L. Wu, Y. Wei, S. Lao, and Y.-H. Yang. „Two-view underwater 3D reconstruction for cameras with unknown poses under flat refractive interfaces“. In: *Pattern Recognition* 69.Supplement C (Sept. 2017), pp. 251–269. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2017.04.006. (Visited on 10/10/2017).

[KAP02] Y.-H. Kwon, A. Ables, and P. G. Pope. „Examination of different double-plane camera calibration strategies for underwater motion analysis“. en. In: *ISBS - Conference Proceedings Archive* 1.1 (2002), pp. 329–332. ISSN: 1999-4168.

[KC06] Y.-H. Kwon and J. B. Casebolt. „Effects of light refraction on the accuracy of camera calibration and reconstruction in underwater motion analysis“. In: *Sports Biomechanics / International Society of Biomechanics in Sports* 5.2 (2006), pp. 315–340. ISSN: 1476-3141. DOI: 10.1080/14763140608522881.

[Ke+08] X. Ke, M. A. Sutton, S. M. Lessner, and M. Yost. „Robust stereo vision and calibration methodology for accurate three-dimensional digital image correlation measurements on submerged objects“. In: *Journal of Strain Analysis for Engineering Design* 43.8 (2008), pp. 689–704. DOI: 10.1243/03093247JSA425.

[KFO03] P Korduan, T Förster, and R Obst. „Unterwasser-Photogrammetrie zur 3D-Rekonstruktion des Schiffswracks Darsser Kogge“. In: *Photogrammetrie Fernerkundung Geoinformation* (2003), pp. 373–382.

[KL00] Y.-H. Kwon and S. L. Lindley. „Applicability of four Localized-Calibration Methods in Underwater Motion Analysis“. In: *ISBS - Conference Proceedings Archive* 1.1 (2000), pp. 48–55. ISSN: 1999-4168.

[KS08] C. Kunz and H. Singh. „Hemispherical Refraction and Camera Calibration in Underwater Vision“. English. In: *OCEANS 2008*. 2008, pp. 1–7. ISBN: 978-1-4244-2619-5. DOI: 10.1109/OCEANS.2008.5151967.

[KS10] C. Kunz and H. Singh. „Stereo Self-Calibration for Seafloor Mapping using AUVs“. English. In: *Autonomous Underwater Vehicles (AUV), 2010 IEEE/OES*. 2010, pp. 1–7. ISBN: 978-1-61284-980-5. DOI: 10.1109/AUV.2010.5779655.

[Kwo99] Y.-H. Kwon. „A Camera Calibration Algorithm for the Underwater Motion Analysis“. In: *ISBS - Conference Proceedings Archive* 1.1 (1999), pp. 257–260. ISSN: 1999-4168.

[KWY12a] L. Kang, L. Wu, and Y.-H. Yang. „Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the SVP camera model“. In: *Applied Optics* 51.31 (2012), pp. 7591–7603. DOI: 10.1364/AO.51.007591.

[KWY12b] L. Kang, L. Wu, and Y.-H. Yang. „Two-View Underwater Structure and Motion for Cameras under Flat Refractive Interfaces“. English. In: *Computer Vision - ECCV 2012*. Ed. by Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid. Vol. 7575. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 303–316. ISBN: 978-3-642-33764-2. DOI: 10.1007/978-3-642-33765-9_22.

[KYK09] R. Kawai, A. Yamashita, and T. Kaneko. „Three-Dimensional Measurement of Objects in Water by Using Space Encoding Method“. English. In: *IEEE International Conference on Robotics and Automation, 2009. ICRA '09*. 2009, pp. 2830–2835. ISBN: 978-1-4244-2788-8. DOI: 10.1109/ROBOT.2009.5152430.

[Li+97] R. Li, H. Li, W. Zou, R. G. Smith, and T. A. Curran. „Quantitative Photogrammetric Analysis of Digital Underwater Video Imagery“. In: *IEEE Journal of Oceanic Engineering* 22.2 (1997), pp. 364–375. ISSN: 0364-9059. DOI: 10.1109/48.585955.

[Li95] H. Li. *Quantitative analysis of underwater stereo video images*. Geomatics Engineering, University of Calgary, 1995.

[Low99] D. G. Lowe. „Object recognition from local scale-invariant features“. In: *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*. Vol. 2. 1999, pp. 1150–1157. DOI: 10.1109/ICCV.1999.790410.

[LRL00] J. M. Lavest, G. Rives, and J. T. Lapresté. „Underwater Camera Calibration“. In: *Computer Vision - ECCV 2000*. Ed. by Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, and David Vernon. Vol. 1843. Springer Berlin Heidelberg, 2000, pp. 654–668. ISBN: 978-3-540-67686-7 978-3-540-45053-5. DOI: 10.1007/3-540-45053-X_42.

[LRL03] J. M. Lavest, G. Rives, and J. T. Lapresté. „Dry camera calibration for underwater applications“. In: *Mach. Vision Appl.* 13.5-6 (2003), pp. 245–253. ISSN: 0932-8092. DOI: 10.1007/s00138-002-0112-z.

[LTZ96] R. Li, C. Tao, and W. Zou. „An Underwater Digital Photogrammetric System for Fishery Geomatics“. In: *XVIIIth ISPRS Congress*. Vienna, Austria, 1996, pp. 319–323.

[Maa15a] H.-G. Maas. „A Modular Geometric Model for Underwater Photogrammetry“. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 55 (2015), pp. 139–141. DOI: 10.5194/isprsarchives-XL-5-W5-139-2015.

[Maa15b] H.-G. Maas. „On the Accuracy Potential in Underwater/Multimedia Photogrammetry“. English. In: *Sensors (Switzerland)* 15.8 (2015), pp. 18140–18152. ISSN: 1424-8220. DOI: 10.3390/s150818140.

[Maa92] H.-G. Maas. „Digitale Photogrammetrie in der dreidimensionalen Strömungsmesstechnik“. theses. Diss. Techn. Wiss. ETH Zürich, Nr. 9665, 1992. Ref.: A. Grün ; Korref.: T. Dracos, 1992. DOI: 10.3929/ethz-a-000627387.

[Maa95] H.-G. Maas. „New developments in Multimedia Photogrammetry“. In: *Symposium A Quarterly Journal In Modern Foreign Literatures* 8.3 (1995), pp. 150–5.

[Men+16] F. Menna, E. Nocerino, F. Fassi, and F. Remondino. „Geometric and Optic Characterization of a Hemispherical Dome Port for Underwater Photogrammetry“. en. In: *Sensors* 16.1 (2016), p. 48. ISSN: 1424-8220. DOI: 10.3390/s16010048.

[MK11] N. J. W. Morris and K. N. Kutulakos. „Dynamic Refraction Stereo“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.8 (2011), pp. 1518–1531. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.24.

[MK75] G. T. Marzan and H. M. Karara. „A computer program for direct linear transformation solution of the collinearity condition, and some applications of it“. In: *Proceedings of the symposium on close-range photogrammetric systems*. 1975, pp. 420–476.

[ML09] M. Muja and D. G. Lowe. „Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration“. In: *International Conference on Computer Vision Theory and Application VISSAPP'09)*. 2009, pp. 331–340.

[MM14] C. Mulsow and H.-G. Maas. „A Universal Approach for Geometric Modelling in Underwater Stereo Image Processing“. In: *2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*. 2014, pp. 49–56. DOI: 10.1109/CVAUI.2014.14.

[MNR17] F. Menna, E. Nocerino, and F. Remondino. „Flat versus Hemispherical Dome Ports in Underwater Photogrammetry“. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W3 (2017), pp. 481–487. DOI: 10.5194/isprs-archives-XLII-2-W3-481-2017.

[MO15] M. Massot-Campos and G. Oliver-Codina. „Optical sensors and methods for underwater 3D reconstruction“. English. In: *Sensors (Switzerland)* 15.12 (2015), pp. 31525–31557. ISSN: 1424-8220. DOI: 10.3390/s151229864.

[Mob94] C. D. Mobley. *Light and water: Radiative transfer in natural waters*. Academic press, 1994.

[Mul10] C. Mulsow. „A Flexible Multi-Media Bundle Approach“. In: *International archives of photogrammetry, remote sensing and spatial information sciences* 38.5 (2010), pp. 472–477.

[Nar+05] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal. „Structured light in scattering media“. English. In: *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005*. Vol. 1. 2005, pp. 420–427. ISBN: 0-7695-2334-X. DOI: 10.1109/ICCV.2005.232.

[Nar+12] T. Naruse, T. Kaneko, A. Yamashita, and H. Asama. „3-D Measurement of Objects in Water Using Fish-Eye Stereo Camera“. In: *2012 19th IEEE International Conference on Image Processing (ICIP)*. 2012, pp. 2773–2776. DOI: 10.1109/ICIP.2012.6467474.

[Nis03] D. Nister. „An efficient solution to the five-point relative pose problem“. In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*. Vol. 2. 2003, pp. 195–202. DOI: 10.1109/CVPR.2003.1211470.

[NN05] S. G. Narasimhan and S. K. Nayar. „Structured light methods for underwater imaging: light stripe scanning and photometric stereo“. In: *Proceedings of MTS/IEEE OCEANS, 2005*. Vol. 3. 2005, pp. 2610–2617. DOI: 10.1109/OCEANS.2005.1640165.

[PES03] O. Pizarro, R. M. Eustice, and H. Singh. „Relative Pose Estimation for Instrumented, Calibrated Imaging Platforms.“ In: *DICTA*. 2003, pp. 601–612.

[PES04] O. Pizarro, R. M. Eustice, and H. Singh. „Large area 3D reconstructions from underwater surveys“. English. In: *OCEANS '04. MTTS/IEEE TECHNO-OCEAN '04*. Vol. 2. 2004, pp. 678–687. ISBN: 0-7803-8669-8. DOI: 10.1109/OCEANS.2004.1405509.

[PES09] O. Pizarro, R. M. Eustice, and H. Singh. „Large Area 3-D Reconstructions From Underwater Optical Surveys“. English. In: *IEEE Journal of Oceanic Engineering* 34.2 (2009), pp. 150–169. ISSN: 0364-9059. DOI: 10.1109/JOE.2009.2016071.

[POA03a] N. Pessel, J. Opderbecke, and M.-J. Aldon. „An Experimental Study of a Robust Self-Calibration Method for a Single Camera“. English. In: *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003*. Vol. 1. 2003, pp. 522–527. ISBN: 953-184-061-X. DOI: 10.1109/ISPA.2003.1296952.

[POA03b] N. Pessel, J. Opderbecke, and M.-J. Aldon. „Camera Self-Calibration in Underwater Environment“. In: *WSCG'03*. 2003.

[QGP15] T. Quick, J. Grebe-Ellis, and O. Passon. „Ein genauer Blick auf die optische Hebung“. de. In: *PhyDid A - Physik und Didaktik in Schule und Hochschule* 1.14 (2015), pp. 26–44. ISSN: 1865-5521.

[Rah+13] T. Rahman, J. Anderson, P. Winger, and N. Krouglicof. „Calibration of an Underwater Stereoscopic Vision System“. In: *Oceans - San Diego, 2013*. 2013, pp. 1–6. DOI: 10.23919/OCEANS.2013.6741133.

[RHB14] Y. Rzhanov, H. Hu, and T. Boyer. „Dense Reconstruction of Underwater Scenes from Monocular Sequences of Images“. In: *OCEANS 2014 - TAIPEI*. 2014, pp. 1–5. DOI: 10.1109/OCEANS-TAIPEI.2014.6964337.

[Sch05] O. Schreer. *Stereoanalyse und Bildsynthese*. German. Springer, 2005. ISBN: 3-540-23439-X 978-3-540-23439-5.

[SGN03] R. Swaninathan, M. D. Grossberg, and S. K. Nayar. „A perspective on distortions“. In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. Vol. 2. 2003, pp. 594–601. DOI: 10.1109/CVPR.2003.1211521.

[SH98] M. R. Shortis and E. S. Harvey. „Design and calibration of an underwater stereo-video system for the monitoring of marine fauna populations“. In: *International Archives Photogrammetry and Remote Sensing* 32.5 (1998), pp. 792–799.

[SHA09] M. R. Shortis, E. S. Harvey, and D. Abdo. „A Review Of Underwater Stereo-image Measurement For Marine Biology And Ecology Applications“. In: *Oceanography and Marine Biology*. CRC Press, 2009, pp. 257–292. ISBN: 978-1-4200-9421-3 978-1-4200-9422-0.

[Sho+00] M. R. Shortis, S. Miller, E. S. Harvey, and S. Robson. „An analysis of the calibration stability and measurement accuracy of an underwater stereo-video system used for shellfish surveys“. In: *Geomatics Research Australasia* 1.73 (2000), pp. 1–24.

[Sho15] M. R. Shortis. „Calibration Techniques for Accurate Measurements by Underwater Camera Systems“. en. In: *Sensors* 15.12 (2015), pp. 30810–30826. ISSN: 1424-8220. DOI: 10.3390/s151229831.

[Sil+12] A. P. Silvatti, F. A. S. Dias, P. Cerveri, and R. M. L. Barros. „Comparison of different camera calibration approaches for underwater applications“. In: *Journal of Biomechanics* 45.6 (2012), pp. 1112–1116. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2012.01.004.

[Sil+13] A. P. Silvatti, P. Cerveri, T. Telles, F. A. S. Dias, G. Baroni, and R. M. L. Barros. „Quantitative underwater 3D motion analysis using submerged video cameras: Accuracy analysis and trajectory reconstruction“. English. In: *Computer Methods in Biomechanics and Biomedical Engineering* 16.11 (2013), pp. 1240–1248. ISSN: 1025-5842. DOI: 10.1080/10255842.2012.664637.

[SK11] A. Sedlazeck and R. Koch. „Calibration of Housing Parameters for Underwater Stereo-Camera Rigs“. In: *Proceedings of the British Machine Vision Conference*. 2011, pp. 118.1–118.11. ISBN: 1-901725-43-X. DOI: 10.5244/C.25.118.

[SK12] A. Sedlazeck and R. Koch. „Perspective and Non-perspective Camera Models in Underwater Imaging - Overview and Error Analysis“. In: *Outdoor and Large-Scale Real-World Scene Analysis*. Ed. by Frank Dellaert, Jan-Michael Frahm, Marc Pollefeys, Laura Leal-Taixé, and Bodo Rosenhahn. Vol. 7474. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 212–242. ISBN: 978-3-642-34090-1. DOI: 10.1007/978-3-642-34091-8_10.

[SKK09] A. Sedlazeck, K. Köser, and R. Koch. „3D Reconstruction Based on Underwater Video from ROV Kiel 6000 Considering Underwater Imaging Conditions“. English. In: *OCEANS 2009 - EUROPE*. 2009, pp. 1–10. ISBN: 978-1-4244-2522-8. DOI: 10.1109/OCEANSE.2009.5278305.

[SO08] M. Shimizu and M. Okutomi. „Calibration and rectification for reflection stereo“. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*. 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587690.

[SR12] V. E. Schmidt and Y. Rzhanov. „Measurement of Micro-bathymetry with a GOPRO Underwater Stereo Camera Pair“. English. In: *OCEANS 2012 MTS/IEEE: Harnessing the Power of the Ocean*. 2012. ISBN: 978-1-4673-0829-8. DOI: 10.1109/OCEANS.2012.6404786.

[SSW13] J. Servos, M. Smart, and S. L. Waslander. „Underwater Stereo SLAM with Refraction Correction“. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2013, pp. 3350–3355. DOI: 10.1109/IROS.2013.6696833.

[Sze11] R. Szeliski. *Computer Vision: Algorithms and Applications*. eng. Springer, 2011. ISBN: 978-1-84882-935-0 978-1-84882-934-3.

[TF10] G. Telem and S. Filin. „Photogrammetric modeling of underwater environments“. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.5 (2010), pp. 433–444. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2010.05.004.

[TMB12] P. A. Tipler, G. Mosca, and M. Basler. *Physik für Wissenschaftler und Ingenieure*. ger. Springer Spektrum, 2012. ISBN: 978-3-8274-1945-3.

[Tre+12] T. Treibitz, Y. Y. Schechner, C. Kunz, and H. Singh. „Flat Refractive Geometry“. English. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.1 (2012), pp. 51–65. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.105.

[Tri+00] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon. „Bundle adjustment - A modern synthesis“. In: *Vision Algorithms: Theory and Practice, LNCS*. 2000, pp. 298–375. DOI: 10.1007/3-540-44480-7_21.

[TSS08] T. Treibitz, Y. Y. Schechner, and H. Singh. „Flat Refractive Geometry“. English. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*. 2008, pp. 1–8. ISBN: 978-1-4244-2242-5. DOI: 10.1109/CVPR.2008.4587844.

[WF14] M. Wehkamp and P. Fischer. „A practical guide to the use of consumer-level digital still cameras for precise stereogrammetric in situ assessments in aquatic environments“. In: *Underwater Technology: International Journal of the Society for Underwater* 32.2 (2014), pp. 111–128. DOI: 10.3723/ut.32.111.

[Yam+03] A. Yamashita, E. Hayashimoto, T. Kaneko, and Y. Kawata. „3-D Measurement of Objects in a Cylindrical Glass Water Tank with a Laser Range Finder“. English. In: *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings*. Vol. 2. 2003, pp. 1578–1583. ISBN: 0-7803-7860-1. DOI: 10.1109/IROS.2003.1248869.

[Yam+04] A. Yamashita, H. Higuchi, T. Kaneko, and Y. Kawata. „Three Dimensional Measurement of Object's Surface in Water Using the Light Stripe Projection Method“. English. In: *2004 IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04*. Vol. 3. 2004, pp. 2736–2741. ISBN: 0-7803-8232-3. DOI: 10.1109/ROBOT.2004.1307474.

[Yam+11] A. Yamashita, R. Kawanishi, T. Koketsu, T. Kaneko, and H. Asama. „Underwater Sensing with Omni-Directional Stereo Camera“. English. In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 2011, pp. 304–311. ISBN: 978-1-4673-0062-9. DOI: 10.1109/ICCVW.2011.6130257.

[YFK08] A. Yamashita, A. Fujii, and T. Kaneko. „Three dimensional measurement of objects in liquid and estimation of refractive index of liquid by using images of water surface with a stereo vision system“. English. In: *IEEE International Conference on Robotics and Automation, 2008. ICRA 2008*. 2008, pp. 974–979. ISBN: 978-1-4244-1646-2. DOI: 10.1109/ROBOT.2008.4543331.

[YGY13] T. Yau, M. Gong, and Y.-H. Yang. „Underwater Camera Calibration Using Wavelength Triangulation“. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013, pp. 2499–2506. DOI: 10.1109/CVPR.2013.323.

[YIK05] A. Yamashita, S. Ikeda, and T. Kaneko. „3-D Measurement of Objects in Unknown Aquatic Environments with a Laser Range Finder“. English. In: *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005. ICRA 2005*. 2005, pp. 3912–3917. ISBN: 0-7803-8914-X. DOI: 10.1109/ROBOT.2005.1570718.

[YKK06] A. Yamashita, S. Kato, and T. Kaneko. „Robust Sensing Against Bubble Noises in Aquatic Environments with a Stereo Vision System“. English. In: *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*. 2006, pp. 928–933. ISBN: 0-7803-9505-0. DOI: 10.1109/ROBOT.2006.1641828.

[YNM13] T. Yano, S. Nobuhara, and T. Matsuyama. „3D Shape from Silhouettes in Water for Online Novel-view Synthesis“. In: *IPSJ Transactions on Computer Vision and Applications* 5 (2013), pp. 65–69. DOI: 10.2197/ipsjtcva.5.65.

[YSK10] A. Yamashita, Y. Shirane, and T. Kaneko. „Monocular Underwater Stereo - 3D Measurement Using Difference of Appearance Depending on Optical Paths -“. English. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2010, pp. 3652–3657. ISBN: 978-1-4244-6674-0. DOI: 10.1109/IROS.2010.5649444.

[Zha00] Z. Zhang. „A flexible new technique for camera calibration“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), pp. 1330–1334. ISSN: 0162-8828. DOI: 10.1109/34.888718.

Own Publications

[DK14] T. Dolereit and A. Kuijper. „Converting Underwater Imaging into Imaging in Air“. In: *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Volume 1, Lisbon, Portugal, 5-8 January, 2014*. Ed. by Sebastiano Battiato and José Braz. 2014, pp. 96–103. ISBN: 978-989-758-003-1. DOI: 10.5220/0004685600960103.

[DL16] T. Dolereit and U. F. von Lukas. „Calibration of Shared Flat Refractive Stereo Systems“. In: *Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings*. Ed. by Aurélio Campilho and Fakhri Karray. Springer International Publishing, 2016, pp. 433–442. ISBN: 978-3-319-41501-7. DOI: 10.1007/978-3-319-41501-7_49.

[DLK15a] T. Dolereit, U. F. von Lukas, and A. Kuijper. „New Constraints for Underwater Stereo Calibration“. In: *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*. 2015, pp. 176–181. DOI: 10.1109/ISPA.2015.7306054.

[DLK15b] T. Dolereit, U. F. von Lukas, and A. Kuijper. „Underwater Stereo Calibration Utilizing Virtual Object Points“. In: *OCEANS 2015 - Genova*. 2015, pp. 1–7. DOI: 10.1109/OCEANS-Genova.2015.7271593.

[Dol15] T. Dolereit. „Concepts for Underwater Stereo Calibration, Stereo 3D-Reconstruction and Evaluation“. In: *Proceedings of the International Summer School on Visual Computing 2015*. 2015, pp. 71–80. ISBN: 978-3-8396-0960-6. URL: http://publica.fraunhofer.de/documents/N-374915.html.

[Luk+15] U. F. von von Lukas, J. Quarles, P. Kaklis, and T. Dolereit. „Underwater Mixed Environments“. en. In: *Virtual Realities*. Ed. by Guido Brunnett, Sabine Coquillart, Robert van Liere, Gregory Welch, and Libor Vasa. Lecture Notes in Computer Science 8844. Springer International Publishing, 2015, pp. 56–76. ISBN: 978-3-319-17042-8 978-3-319-17043-5. DOI: 10.1007/978-3-319-17043-5_4.

[Sch+15] J. Schulz, K. O. Möller, A. Bracher, M. Hieronymi, B. Cisewski, O. Zielinski, D. Voss, E. Gutzeit, T. Dolereit, G. Niedzwiedz, G. Kohlberg, D. Schories, R. Kiko, A. Körtzinger, C. Falldorf, P. Fischer, N. Nowald, K. Beisiegel, P. Martinez-Arbizu, N. Rüssmeier, R. Röttgers, J. Büdenbender, A. Jordt-Sedlazeck, R. Koch, U. Riebesell, M. Iversen, K. Köser, T. Kwasnitschka, J. Wellhausen, C. Thoma, K. Barz, S. Rhode, T. W. Nat-

tkemper, T. Schoening, F. Peeters, H. Hofmann, J. Busch, H.-J. Hirche, B. Niehoff, N. Hildebrandt, E. Stohr, C. Winter, G. Herbst, C. Konrad, M. Schmidt, P. Linke, T. Brey, H. W. Bange, L. Nolle, S. Krägefsky, J. Gröger, E. Sauter, M. Schulz, J. Müller, G. Rehder, D. Stepputtis, B. Beszteri, M. Kloster, G. Kauer, A. Göritz, P. Gege, U. F. von Lukas, and U. V. Bathmann. „Aquatische Optische Technologien in Deutschland“. In: *Marine Science Reports - Meereswissenschaftliche Berichte* 97 (2015), pp. 1–83. DOI: 10.12754/msr-2015-97.

Thesis Statements

1.) Underwater imaging devices are characterized by a transition of the entering light through different media, such as the very common water-glass-air transition in a *FRS*, and therefore get affected by the physical phenomenon of the refraction of light. Refraction poses a problem to every discipline aiming at the recovery of metric 3D structure from underwater image data. In contrast to lens distortion, distortion due to refraction is distance-dependent and therefore not an image space distortion.

2.) The common strategies to handle refractive effects are their disregard, their absorption or their explicit modeling. The first two essentially consider refractive effects to be image space distortions and are based on image formation following the pinhole camera model, which is invalid for underwater image formation due to refraction. Consequently, the disregard needs to be avoided and the absorption should be avoided, since the required setups represent a severe restriction and the elimination of refractive effects is incomplete. With both of these strategies, an accurate 3D reconstruction is not possible. The only way to handle refractive effects physically correct is their explicit modeling, which in turn requires the calibration of the modeled refractive parameters.

3.) An underwater image formation model that represents the parametrization of a stereo camera in front of a single flat refractive interface is the model for a *SFRS*. There are multiple practical configurations of imaging setups that can be represented by it and although being more specialized than single camera systems, it can be designated as a representative of a common viewing condition. The model for a *SFRS* comprises a master and a slave camera and shows beneficial properties that make it worth to adjust the basic design of the imaging system accordingly. An exemplary benefit is the reduction of the refractive parameters, since these have to be estimated just with respect to the master camera.

4.) The refractive parameters modeled by the *SFRS* have to be determined by refractive calibration. These are the refractive indices of the participating media, and the so-called system axis and air layer thickness, which essentially represent the orientation and the distance of the master camera to the flat refractive interface, respectively. However, since the refractive indices are in general measurable, commonly known (air and water) or known from manufacturer data (underwater housing), they can be excluded from algorithmic determination. The system axis can be determined independently of the air layer thickness.

5.) The refractive calibration of a *SFRS* can be realized by two different strategies, which can be distinguished by an implicit (simultaneous strategy) and an explicit

system axis determination (consecutive strategy). The consecutive strategy can be realized by determining the system axis with the aid of calibration objects with feature points arranged in straight lines and a following determination of the air layer thickness for that system axis. In contrast, the simultaneous strategy can be realized without any calibration objects.

6.) A *VOP* model is a beneficial extension for underwater image formation models. The *VOP* model can be readily integrated into the underwater image formation models of monocular and binocular vision. The combined models amount to an efficient tool for the extension of the concepts of stereo 3D reconstruction in air for underwater environments with an explicit consideration of refractive effects.

6.1.) The simultaneous and the consecutive strategy for the refractive calibration of a *SFRS* both benefit from the integrated *VOP* model. Based on the *VOP* model, an optimization problem can be defined for the determination of the air layer thickness of a *SFRS*. This optimization problem can be solved efficiently by linear optimization and can be realized within both calibration strategies.

6.2.) Reoccurring operations within the concepts of stereo 3D reconstruction are refractive projections for coordinate transformations between 3D and 2D spaces. Within a refractive system, these are the computationally intensive refractive forward projections and refractive back-projections. The integration of the *VOP* model serves the development of alternative strategies and the reduction of the computational expense. Together with these refractive projections, the *VOP* model furthermore benefits the computation of adapted and novel cost functions and the definition of optimization problems for the determination of the system axis of a *SFRS* in an iterative fashion.

6.3.) Refractive calibration is the first necessary step of underwater stereo 3D reconstruction. For the actual recovery of 3D coordinates with an underwater stereo camera, corresponding pixels in the left and the right view of the stereo camera need to be computed by stereo matching. Due the invalidity of the pinhole camera model for underwater image formation, epipolar geometry gets invalid as well. An alternative is the computation of correspondence curves based on refractive projections. Finally, the actual recovery of 3D coordinates can be realized with the aid of refractive projections as well.

7.) A valid approach for benchmark data generation for the evaluation of the developed approaches with real test data can be realized with the aid of a calibration object that is fixed to the stereo camera in a solid construction. With such a construction, it is possible to compute a benchmark 3D point cloud with respect to the stereo camera in air. This benchmark 3D point cloud can in turn be utilized for the implicit evaluation of the refractive parameters by explicit evaluation of a recovered 3D point cloud. It represents a simple and efficient means for the evaluation of underwater stereo 3D reconstruction.

Lebenslauf

Tim Dolereit
geb. am 17.02.1984 in Schwerin

Beruflicher Werdegang

seit 06/2016	Wissenschaftlicher Mitarbeiter Fraunhofer IGD

Promotion

seit 10/2011	Universität Rostock Fakultät für Informatik und Elektrotechnik
10/2014 - 03/2016	Wissenschaftlicher Mitarbeiter Lehrstuhl Multimediale Kommunikation Universität Rostock
10/2011 - 09/2014	Promotionsstipendium Visual Computing Research and Innovation Center am Fraunhofer IGD

Studium

04/2006 - 09/2011	Diplominformatik Universität Rostock Vertiefungsrichtung Computergraphik

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die das Entstehen dieser Dissertation ermöglicht und zur Fertigstellung beigetragen haben.
Dieser Dank gilt Professor Bodo Urban für die wissenschaftliche und methodische Unterstützung, sowie für die konstruktiven Gespräche und Anmerkungen. Weiterhin gilt mein Dank Professor Uwe von Lukas für die vielen zielführenden Diskussionen und die anhaltende Hilfestellung und Motivierung. Ich danke Professor Reinhard Koch für die fachlichen Gespräche, die mich in der Abhandlung dieser Arbeit sehr unterstützten, und Professor Arjan Kuijper für die wissenschaftliche Begleitung.
Allen Kollegen vom Visual Computing Research and Innovation Center am Fraunhofer IGD in Rostock, sowie allen Kollegen vom Fraunhofer IGD selber danke ich für die vielen wissenschaftlichen Gespräche, für Denkanstöße und Hilfestellungen. Gleichermaßen gilt mein Dank allen Kollegen für die nicht weniger wichtigen nichtwissenschaftlichen und motivierenden Gespräche, die meine Arbeit sehr unterstützt haben. Auch den Projektpartnern möchte ich für die angenehme Zusammenarbeit in den für diese Dissertation relevanten Projekten danken.
Dankeschön an alle Freunde und Verwandte, die mein Jammern über die Jahre ertragen haben, und mich durch Rücksichtnahme, notwendige Ablenkung in der Freizeit, sowie fachliche Kompetenz fürs Korrekturlesen unterstützt haben. Mein weiterer Dank gilt meinen Eltern für die seelische und moralische Unterstützung. Besonders möchte ich mich bei meiner Mutter für den notwendigen Rückhalt und die vielen Entlastungen im Alltag bedanken.

Selbstständigkeitserklärung

Ich erkläre, dass ich die eingereichte Promotionsschrift selbstständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Ich versichere weiterhin, dass ich bisher weder die vorliegende Promotionsschrift noch Teile von ihr als Prüfungsarbeit oder zum Zweck der Promotion eingereicht bzw. verwendet habe.

Rostock, den 22. Dezember 2017

Tim Dolereit