



TECHNISCHE UNIVERSITÄT
CHEMNITZ

A Novel Approach for Spherical Stereo Vision

of the Faculty of Electrical Engineering and Information Technology
at Chemnitz University of Technology

approved

Dissertation

to obtain the academic degree

Doktor-Ingenieur

Dr.-Ing.

submitted

by Dipl.-Ing. Michel Findeisen

born on 17 November 1979 in Rochlitz

filed on 19 February 2015

Assessors: Prof. Dr.-Ing. Gangolf Hirtz
Prof. Dr.-Ing. habil. Madhukar Chandra

Day of Conferment: 23 April 2015



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Ein Neuer Ansatz für Sphärisches Stereo Vision

von der Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Chemnitz

genehmigte

Dissertation

zur Erlangung des akademischen Grades

Doktor-Ingenieur

Dr.-Ing.

vorgelegt

von Dipl.-Ing. Michel Findeisen

geboren am 17. November 1979 in Rochlitz

eingereicht am 19. Februar 2015

Gutachter: Prof. Dr.-Ing. Gangolf Hirtz

Prof. Dr.-Ing. habil. Madhukar Chandra

Tag der Verleihung: 23. April 2015

Bibliographic Description

Author Michel Findeisen

Title A Novel Approach for Spherical Stereo Vision

Organization Chemnitz University of Technology, Faculty of Electrical Engineering and Information Technology, Professorship of Digital Signal Processing and Circuit Technology, PhD Thesis, April 2015

Keywords Stereo Vision; Spherical Stereo Vision; Omnidirectional Stereo Vision; RGB-D Sensor; Depth Sensing; Trinocular Camera Configuration; Visual Indoor Surveillance; Fisheye Camera

Abstract

The Professorship of Digital Signal Processing and Circuit Technology of Chemnitz University of Technology conducts research in the field of three-dimensional space measurement with optical sensors. In recent years this field has made major progress.

For example innovative, active techniques such as the “structured light“-principle are able to measure even homogeneous surfaces and have found their way into the consumer electronic market in terms of Microsofts Kinect[®] at the present time. Furthermore, high-resolution optical sensors establish powerful, passive stereo vision systems in the field of indoor surveillance. Thereby they induce new application domains such as security and assistance systems for domestic environments.

However, the constraint field of view can still be considered as an essential characteristic of all these technologies. For instance, in order to measure a volume in size of a living space, two to three deployed 3D sensors have to be applied nowadays. This is due to the fact that the commonly utilized perspective projection principle constrains the visible area to a field of view of approximately 120° . On the other hand, novel fish-eye lenses allow the realization of omnidirectional projection models. Therewith, the visible field of view can be enlarged up to more than 180° . In combination with a 3D measurement approach, thus, the number of required sensors for entire room coverage can be reduced considerably.

Motivated by the requirements of the field of indoor surveillance, the present work focuses on the combination of the established stereo vision principle and omnidirectional projection methods. The entire 3D measurement of a living space by means of one single sensor can be considered as major objective.

As a starting point for this thesis, Chapter 1 discusses the underlying requirement, referring to various relevant fields of application. Based on this, the distinct purpose for the present work is stated.

The necessary mathematical foundations of computer vision are reflected in Chapter 2 subsequently. Based on the geometry of the optical imaging process, the projection characteristics of relevant principles are discussed and a generic method for modelling fish-eye cameras is selected.

Chapter 3 deals with the extraction of depth information using classical (perceptively imaging) binocular stereo vision configurations. In addition to a complete recap of the processing chain, especially occurring measurement uncertainties are investigated.

In the following, Chapter 4 addresses special methods to convert different projection models. The example of mapping an omnidirectional to a perspective projection is employed in order to develop a method for accelerating this process and, hereby, for reducing the computational load associated therewith. Any errors that occur, as well as the necessary adjustment of image resolution, are an integral part of the investigation. As a practical example, an application for person tracking is utilized in order to demonstrate to which extent the usage of “virtual views“ can increase the recognition rate for people detectors in the context of omnidirectional monitoring.

Subsequently, an extensive search with respect to omnidirectional imaging stereo vision techniques is conducted in Chapter 5. It turns out that the complete 3D capture of a room is achievable by the generation of a hemispherical depth map. Therefore, three cameras have to be combined in order to form a trinocular stereo vision system. As a basis for further research, a known trinocular stereo vision method is selected. Furthermore, it is hypothesized that, by applying a modified geometric constellation of cameras, more precisely in the form of an equilateral triangle, and using an alternative method to determine the depth map, the performance can be increased considerably. A novel method is presented which shall require fewer operations to calculate the distance information and which is to avoid a computational costly step for depth map fusion as necessary in the comparative method.

In order to evaluate the presented approach as well as the hypotheses, a hemispherical depth map is generated in Chapter 6 by means of the new method. Simulation results, based on artificially generated 3D space information and realistic system parameters, are presented and subjected to a subsequent error estimate.

A demonstrator for generating real measurement information is introduced in Chapter 7. In addition, the methods that are applied for calibrating the system intrinsically as well as extrinsically are explained. It turns out that the calibration procedure utilized cannot estimate the extrinsic parameters sufficiently. Initial measurements present a hemispherical depth map and thus confirm the operativeness of the concept, but also identify the drawbacks of the calibration used. The current implementation of the algorithm shows almost real-time behaviour.

Finally, Chapter 8 summarizes the results obtained along the studies and discusses them in the context of comparable binocular and trinocular stereo vision approaches. For example the results of the simulations carried out produced a saving of up to 30% in terms of stereo correspondence operations in comparison with a referred trinocular method. Furthermore, the concept introduced allows the avoidance of a weighted averaging step

for depth map fusion based on precision values that have to be calculated in a costly manner. The achievable accuracy is still comparable for both trinocular approaches.

In summary, it can be stated that, in the context of the present thesis, a measurement system has been developed which has great potential for future application fields in industry, security in public spaces as well as home environments.

Zusammenfassung

Die Professur Digital- und Schaltungstechnik der Technischen Universität Chemnitz forscht auf dem Gebiet der dreidimensionalen Raumvermessung mittels optischer Sensorik. In den letzten Jahren konnte dieses Forschungsgebiet wesentliche Fortschritte verzeichnen.

Beispielsweise erlauben innovative, aktive Verfahren wie das „Structured Light“-Prinzip die präzise Erfassung auch homogener Oberflächen und halten gegenwärtig in Form der Microsoft Kinect[®] Einzug in die Konsumerelektronik. Des Weiteren ermöglichen hochauflösende optische Sensoren die Etablierung leistungsfähiger, passiver Stereo-Vision Systeme im Bereich der Raumüberwachung und schaffen damit neuartige Anwendungsfelder wie etwa Sicherheits- und Assistenzsysteme für das häusliche Umfeld.

Eine wesentliche Einschränkung dieser Technologien bildet der nach wie vor stark limitierte Sichtbereich der Sensorik. So sind zum Beispiel zur optischen, dreidimensionalen Erfassung eines Volumens der Größe eines Wohnraumes aktuell etwa zwei bis drei verteilte Sensoren erforderlich. Als Ursache ist hauptsächlich das zugrunde liegende perspektivische Abbildungsprinzip der 3D-Messverfahren zu nennen, welches den sichtbaren Bereich auf einen Öffnungswinkel von etwa 120° beschränkt. Neuartige Fischaugenobjektive hingegen ermöglichen die Umsetzung omnidirektionaler Projektionsmodelle und damit die Erweiterung des Erfassungsbereichs auf 180° des Sichtfeldes. In Kombination mit einem 3D-Messverfahren kann damit die Anzahl der benötigten Sensoren für eine vollständige Raumvermessung wesentlich reduziert werden.

Motiviert, insbesondere durch die anwendungsbezogenen Anforderungen der Raumüberwachung, befasst sich die vorliegende Arbeit mit der Kombination des etablierten Stereo-Vision Prinzips mit omnidirektionalen Projektionsverfahren. Das Ziel ist die vollständige, dreidimensionale Erfassung eines Raumes mit nur einem optischen 3D-Sensor.

Als Ausgangspunkt der Arbeit wird in Kapitel 1 die zugrundeliegende Problemstellung in Bezug auf verschiedene, relevante Anwendungsfelder dargelegt. Davon ausgehend wird das wesentliche Ziel der Untersuchungen formuliert.

Die notwendigen mathematischen Grundlagen des maschinellen Sehens werden anschließend in Kapitel 2 reflektiert. Ausgehend von der Geometrie des optischen Abbildungsprozesses werden die Eigenschaften relevanter Projektionsprinzipien erörtert und ein generisches Verfahren zur Modellierung von Fischaugenkameras ausgewählt.

In Kapitel 3 wird die Gewinnung von Tiefeninformationen unter Verwendung klassischer, perspektivisch abbildender binokularer Stereo-Vision Konfigurationen behandelt. Neben einer Aufarbeitung der kompletten Verarbeitungskette werden insbesondere auftretende Messungenauigkeiten untersucht.

Im Folgenden werden in Kapitel 4 spezielle Verfahren zur Konvertierung verschiedener Projektionsmodelle diskutiert. Am Beispiel der Transformation einer omnidirektionalen in eine perspektivische Abbildung wird eine Methode entwickelt, welche die damit verbundene Rechenlast reduziert und das Verfahren wesentlich beschleunigt. Auftretende Fehler sowie die notwendige Anpassung der Bildauflösung sind integraler Bestandteil der Untersuchungen. Am Beispiel einer Anwendung zur Personenlokalisierung kann gezeigt werden, dass sich die Erkennungsrate durch den Einsatz „virtueller Abbildungen“ in der omnidirektionalen Überwachung signifikant steigern lässt.

Aufbauend auf den gelegten Grundlagen wird im sich anschließenden Kapitel 5 eine ausführliche Recherche zu omnidirektional abbildenden Stereo-Vision-Verfahren dargestellt. Es zeigt sich, dass die vollständige 3D-Erfassung eines Raumes mit Hilfe einer hemisphärischen Tiefenkarte möglich ist und dazu prinzipiell drei Kameras zu einem trinokularen Messsystem kombiniert werden müssen. Als Grundlage für die weiteren Untersuchungen wird ein bekanntes trinokulares Stereo-Vision Verfahren ausgewählt. Davon ausgehend wird die Hypothese aufgestellt, dass bei Änderung der geometrischen Konstellation der Kameras zu einem gleichschenkligen Dreieck sowie Anwendung einer alternativen Methode zur Bestimmung der Tiefenkarte die Performance vergleichsweise signifikant gesteigert werden kann. Es wird ein neues Verfahren vorgestellt, welches im Vergleich weniger Operationen zur Berechnung der Entfernungsinformationen benötigt und einen rechenaufwändigen Schritt zur Fusionierung von Tiefenkarten, wie im Vergleichsverfahren notwendig, unterbinden soll.

Zur Überprüfung des dargelegten Konzeptes und der getroffenen Hypothesen wird in Kapitel 6 eine hemisphärische Tiefenkarte auf Grundlage des neuen Verfahrens generiert. Simulationsergebnisse auf Basis künstlich generierter 3D-Rauminformationen und realistischer Systemparameter werden präsentiert und einer anschließenden Fehlerabschätzung unterzogen.

Ein Demonstrator zur Erzeugung realer Messinformationen wird in Kapitel 7 vorgestellt. Die verwendeten Methoden zur intrinsischen sowie extrinsischen Kalibrierung des Systems werden dargelegt. Es stellt sich heraus, dass das verwendete Kalibrierverfahren die

extrinsischen Parameter nicht genau genug schätzen kann. Erste vorgestellte Messergebnisse in Form einer hemisphärischen Tiefenkarte bestätigen die Funktionsfähigkeit des Konzeptes, zeigen aber auch die Nachteile der verwendeten Kalibrierung. Die aktuelle Implementierung des Verfahrens zeigt nahezu Echtzeitverhalten.

Abschließend werden in Kapitel 8 die erreichten Ergebnisse der getätigten Untersuchungen zusammengefasst und im Kontext vergleichbarer binokularer und trinokularer Stereo-Vision Ansätze diskutiert. Beispielsweise zeigen die durchgeführten Simulationen eine Einsparung der Rechenoperationen bei der Bestimmung der Tiefenkarte um bis zu 30% im Vergleich zum Referenzverfahren. Weiterführend kann im vorgestellten Konzept bei der Fusionierung von Entfernungsinformationen auf eine gewichtete Mittelwertbildung unter Verwendung aufwendig zu berechnender Genauigkeitswerte verzichtet werden. Die erreichbare Präzision ist dennoch für beide trinokulare Verfahren vergleichbar.

Zusammenfassend kann gesagt werden, dass im Rahmen der vorliegenden Arbeit ein Messsystem entstanden ist, welches großes Potenzial für zukünftige Aufgabenfelder in Industrie, der Sicherheit im öffentlichen Raum sowie im häuslichen Bereich aufweist.

Contents

Abstract	7
Zusammenfassung	11
Acronyms	27
Symbols	29
Acknowledgement	33
1 Introduction	35
1.1 Visual Surveillance	35
1.2 Challenges in Visual Surveillance	38
1.3 Outline of the Thesis	41
2 Fundamentals of Computer Vision Geometry	43
2.1 Projective Geometry	43
2.1.1 Euclidean Space	43
2.1.2 Projective Space	44
2.2 Camera Geometry	45
2.2.1 Geometrical Imaging Process	45
2.2.1.1 Projection Models	46
2.2.1.2 Intrinsic Model	47
2.2.1.3 Extrinsic Model	50
2.2.1.4 Distortion Models	51
2.2.2 Pinhole Camera Model	51
2.2.2.1 Complete Forward Model	52
2.2.2.2 Back Projection	53
2.2.3 Equiangular Camera Model	54
2.2.4 Generic Camera Models	55
2.2.4.1 Complete Forward Model	56
2.2.4.2 Back Projection	58

2.3	Camera Calibration Methods	58
2.3.1	Perspective Camera Calibration	59
2.3.2	Omnidirectional Camera Calibration	59
2.4	Two-View Geometry	60
2.4.1	Epipolar Geometry	61
2.4.2	The Fundamental Matrix	63
2.4.3	Epipolar Curves	64
3	Fundamentals of Stereo Vision	67
3.1	Introduction	67
3.1.1	The Concept Stereo Vision	67
3.1.2	Overview of a Stereo Vision Processing Chain	68
3.2	Stereo Calibration	69
3.2.1	Extrinsic Stereo Calibration With Respect to the Projective Error	70
3.3	Stereo Rectification	72
3.3.1	A Compact Algorithm for Rectification of Stereo Pairs	73
3.4	Stereo Correspondence	76
3.4.1	Disparity Computation	76
3.4.2	The Correspondence Problem	77
3.5	Triangulation	79
3.5.1	Depth Measurement	79
3.5.2	Range Field of Measurement	80
3.5.3	Measurement Accuracy	80
3.5.4	Measurement Errors	81
3.5.4.1	Quantization Error	82
3.5.4.2	Statistical Distribution of Quantization Errors	83
4	Virtual Cameras	87
4.1	Introduction and Related Works	88
4.2	Omni to Perspective Vision	90
4.2.1	Forward Mapping	90
4.2.2	Backward Mapping	93
4.2.3	Fast Backward Mapping	96
4.3	Error Analysis	99
4.4	Accuracy Analysis	101
4.4.1	Intrinsics of the Source Camera	102
4.4.2	Intrinsics of the Target Camera	102
4.4.3	Marginal Virtual Pixel Size	104
4.5	Performance Measurements	109
4.6	Virtual Perspective Views for Real-Time People Detection	110

5	Omnidirectional Stereo Vision	113
5.1	Introduction and Related Works	113
5.1.1	Geometrical Configuration	116
5.1.1.1	H-Binocular Omni-Stereo with Panoramic Views	117
5.1.1.2	V-Binocular Omnistereoo with Panoramic Views	119
5.1.1.3	Binocular Omnistereoo with Hemispherical Views	120
5.1.1.4	Trinocular Omnistereoo	122
5.1.1.5	Miscellaneous Configurations	125
5.2	Epipolar Rectification	126
5.2.1	Cylindrical Rectification	127
5.2.2	Epipolar Equi-Distance Rectification	128
5.2.3	Epipolar Stereographic Rectification	128
5.2.4	Comparison of Rectification Methods	129
5.3	A Novel Spherical Stereo Vision Setup	129
5.3.1	Physical Omnidirectional Camera Configuration	131
5.3.2	Virtual Rectified Cameras	131
6	A Novel Spherical Stereo Vision Algorithm	135
6.1	Matlab Simulation Environment	135
6.2	Extrinsic Configuration	136
6.3	Physical Camera Configuration	137
6.4	Virtual Camera Configuration	137
6.4.1	The Focal Length	138
6.4.2	Prediscussion of the Field of View	138
6.4.3	Marginal Virtual Pixel Sizes	139
6.4.4	Calculation of the Field of View	142
6.4.5	Calculation of the Virtual Pixel Size Ratios	143
6.4.6	Results of the Virtual Camera Parameters	144
6.5	Spherical Depth Map Generation	147
6.5.1	Omnidirectional Imaging Process	148
6.5.2	Rectification Process	148
6.5.3	Rectified Depth Map Generation	150
6.5.4	Spherical Depth Map Generation	151
6.5.5	3D Reprojection	153
6.6	Error Analysis	154
7	Stereo Vision Demonstrator	163
7.1	Physical System Setup	163
7.2	System Calibration Strategy	165
7.2.1	Intrinsic Calibration of the Physical Cameras	165

7.2.2	Extrinsic Calibration of the Physical and the Virtual Cameras . . .	166
7.2.2.1	Extrinsic Initialization of the Physical Cameras	167
7.2.2.2	Extrinsic Initialization of the Virtual Cameras	167
7.2.2.3	Two-View Stereo Calibration and Rectification	167
7.2.2.4	Three-View Stereo Rectification	168
7.2.2.5	Extrinsic Calibration Results	169
7.3	Virtual Camera Setup	171
7.4	Software Realization	172
7.5	Experimental Results	172
7.5.1	Qualitative Assessment	172
7.5.2	Performance Measurements	174
8	Discussion and Outlook	177
8.1	Discussion of the Current Results and Further Need for Research	177
8.1.1	Assessment of the Geometrical Camera Configuration	178
8.1.2	Assessment of the Depth Map Computation	179
8.1.3	Assessment of the Depth Measurement Error	182
8.1.4	Assessment of the Spherical Stereo Vision Demonstrator	183
8.2	Review of the Different Approaches for Hemispherical Depth Map Generation	184
8.2.1	Comparison of the Equilateral and the Right-Angled Three-View Approach	184
8.2.2	Review of the Three-View Approach in Comparison with the Two-View Method	185
8.3	A Sample Algorithm for Human Behaviour Analysis	187
8.4	Closing Remarks	189
A	Relevant Mathematics	191
A.1	Cross Product by Skew Symmetric Matrix	191
A.2	Derivation of the Quantization Error	191
A.3	Derivation of the Statistical Distribution of Quantization Errors	192
A.4	Approximation of the Quantization Error for Equiangular Geometry . . .	194
B	Further Relevant Publications	197
B.1	H-Binocular Omnidirectional Stereo Vision with Panoramic Views	197
B.2	V-Binocular Omnidirectional Stereo Vision with Panoramic Views	198
B.3	Binocular Omnidirectional Stereo Vision with Hemispherical Views	200
B.4	Trinocular Omnidirectional Stereo Vision	201
B.5	Miscellaneous Configurations	202
	Bibliography	209
	List of Figures	223

List of Tables	229
Affidavit	231
Theses	233
Thesen	235
Curriculum Vitae	237

Inhaltsverzeichnis

Zusammenfassung (englisch)	7
Zusammenfassung (deutsch)	11
Abkürzungsverzeichnis	27
Symbolverzeichnis	29
Danksagung	33
1 Einleitung	35
1.1 Optische Überwachung	35
1.2 Herausforderungen bei der Optischen Überwachung	38
1.3 Überblick über die Arbeit	41
2 Geometrische Grundlagen des Maschinellen Sehens	43
2.1 Projektive Geometrie	43
2.1.1 Euklidischer Raum	43
2.1.2 Projektiver Raum	44
2.2 Kamera Geometrie	45
2.2.1 Geometrischer Abbildungsprozess	45
2.2.1.1 Projektionsmodelle	46
2.2.1.2 Intrinsisches Modell	47
2.2.1.3 Extrinsisches Modell	50
2.2.1.4 Verzerrungsmodell	51
2.2.2 Lochkameramodell	51
2.2.2.1 Vollständiges Vorwärtsmodell	52
2.2.2.2 Rückprojektion	53
2.2.3 Equiangulares Kameramodell	54
2.2.4 Generische Kameramodelle	55
2.2.4.1 Vollständiges Vorwärtsmodell	56
2.2.4.2 Rückprojektion	58

2.3	Kamerakalibrierverfahren	58
2.3.1	Perspektivische Kamerakalibrierung	59
2.3.2	Omnidirektionale Kamerakalibrierung	59
2.4	Zwei-Kamera Geometrie	60
2.4.1	Epipolargeometrie	61
2.4.2	Die Fundamentalmatrix	63
2.4.3	Epipolarkurven	64
3	Die Grundlagen des Stereo Vision Verfahrens	67
3.1	Einleitung	67
3.1.1	Das Konzept Stereo Vision	67
3.1.2	Überblick über die Stereo Vision Prozesskette	68
3.2	Stereokalibrierung	69
3.2.1	Extrinsische Stereokalibrierung	70
3.3	Stereorektifizierung	72
3.3.1	Ein Kompakter Algorithmus für die Rektifizierung von Stereopaaren	73
3.4	Stereokorrespondenz	76
3.4.1	Disparitätsberechnung	76
3.4.2	Das Korrespondenzproblem	77
3.5	Triangulation	79
3.5.1	Entfernungsbestimmung	79
3.5.2	Grenzen des Messfeldes	80
3.5.3	Messgenauigkeit	80
3.5.4	Messfehler	81
3.5.4.1	Quantisierungsfehler	82
3.5.4.2	Statistische Verteilung des Quantisierungsfehlers	83
4	Virtuelle Kameras	87
4.1	Einleitung und Relevante Veröffentlichungen	88
4.2	Von der Omnidirektionalen zur Perspektivischen Abbildung	90
4.2.1	Vorwärtsverfahren	90
4.2.2	Rückwärtsverfahren	93
4.2.3	Schnelles Rückwärtsverfahren	96
4.3	Fehleranalyse	99
4.4	Analyse der Genauigkeit	101
4.4.1	Die Intrinsischen Parameter der Physikalischen Kamera	102
4.4.2	Die Intrinsischen Parameter der Virtuellen Kamera	102
4.4.3	Grenzipixelgröße	104
4.5	Geschwindigkeitsmessungen	109
4.6	Virtuelle Perspektivische Ansichten für eine Echtzeit-Personenerkennung	110

5	Omnidirektionales Stereo Vision	113
5.1	Einleitung und Relevante Veröffentlichungen	113
5.1.1	Geometrische Konfiguration	116
5.1.1.1	H-Binokulares Omnidirektionales Stereo Vision mit Panoramischer Abbildung	117
5.1.1.2	V-Binokulares Omnidirektionales Stereo Vision mit Panoramischer Abbildung	119
5.1.1.3	Binokulares Omnidirektionales Stereo Vision mit Hemisphärischer Abbildung	120
5.1.1.4	Trinokulares Omnidirektionales Stereo Vision	122
5.1.1.5	Sonstige Konfigurationen	125
5.2	Epipolare Rektifizierungsverfahren	126
5.2.1	Zylindrische Rektifizierung	127
5.2.2	Epipolare Equidistante Rektifizierung	128
5.2.3	Epipolare Stereographische Rektifizierung	128
5.2.4	Vergleich der Rektifizierungsmethoden	129
5.3	Ein Neuartiges Stereo Vision Setup	129
5.3.1	Physikalische Omnidirektionale Kamerakonfiguration	131
5.3.2	Virtuelle Rektifizierte Kameras	131
6	Ein Neuartiger Algorithmus für Sphärisches Stereo Vision	135
6.1	Matlab Simulationsumgebung	135
6.2	Extrinsische Konfiguration	136
6.3	Physikalische Kamerakonfiguration	137
6.4	Virtuelle Kamerakonfiguration	137
6.4.1	Die Brennweite	138
6.4.2	Diskussion des Blickfeldes	138
6.4.3	Grenzpixelgröße	139
6.4.4	Berechnung des Blickfeldes	142
6.4.5	Berechnung der Virtuellen Pixelgrößen	143
6.4.6	Ergebnisse der Berechnung Virtueller Kameraparameter	144
6.5	Generierung der Sphärischen Tiefenkarte	147
6.5.1	Omnidirektionaler Abbildungsprozess	148
6.5.2	Abbildungsprozess der Rektifizierten Kameras	148
6.5.3	Generierung der Rektifizierten Tiefenkarten	150
6.5.4	Generierung der Sphärischen Tiefenkarte	151
6.5.5	3D Rückprojektion	153
6.6	Fehleranalyse	154
7	Demonstrator für Sphärisches Stereo Vision	163
7.1	Physikalische Systemkonfiguration	163

7.2	Strategie für die Kalibrierung	165
7.2.1	Intrinsische Kalibrierung der Physikalischen Kameras	165
7.2.2	Extrinsische Kalibrierung der Physikalischen und Virtuellen Kameras	166
7.2.2.1	Extrinsische Initialisierung der Physikalischen Kameras	167
7.2.2.2	Extrinsische Initialisierung der Virtuellen Kameras	167
7.2.2.3	Stereokalibrierung und -Rektifizierung eines Zweikamera- systems	167
7.2.2.4	Rektifizierung eines Dreikamerasystems	168
7.2.2.5	Ergebnisse der Extrinsischen Kalibrierung	169
7.3	Virtuelle Kamerakonfiguration	171
7.4	Softwarerealisierung	172
7.5	Experimentelle Ergebnisse	172
7.5.1	Qualitative Begutachtung	172
7.5.2	Geschwindigkeitsmessungen	174
8	Diskussion und Ausblick	177
8.1	Diskussion der Ergebnisse und Weiterer Forschungsbedarf	177
8.1.1	Begutachtung der Geometrischen Kamerakonfiguration	178
8.1.2	Begutachtung der Berechnung der Tiefenkarte	179
8.1.3	Begutachtung der Messfehler	182
8.1.4	Begutachtung des Demonstratos	183
8.2	Review der Verschiedenen Ansätze zur Berechnung einer Hemisphärischen Tiefenkarte	184
8.2.1	Vergleich der Verfahren zur Bestimmung einer Tiefenkarte mit Drei Kameras	184
8.2.2	Vergleich der Zwei- und Dreikamerasysteme	185
8.3	Ein Beispielverfahren zur Analyse Menschlichen Verhaltens	187
8.4	Schlußbemerkungen	189
A	Mathematische Herleitungen	191
A.1	Kreuzprodukt einer Schiefsymmetrischen Matrix	191
A.2	Ableitung des Quantisierungsfehlers	191
A.3	Ableitung der Statistischen Verteilung des Quantisierungsfehlers	192
A.4	Approximation des Quantisierungsfehlers für Equiangulare Geometrie	194
B	Weitere Relevante Veröffentlichungen	197
B.1	H-Binokulares Omnidirektionales Stereo Vision mit Panoramischer Abbil- dung	197
B.2	V-Binokulares Omnidirektionales Stereo Vision mit Panoramischer Abbil- dung	198
B.3	Binokulares Omnidirektionales Stereo Vision mit Hemisphärischer Abbildung	200

B.4 Trinokulares Omnidirektionales Stereo Vision	201
B.5 Sonstige Konfigurationen	202
Quellenverzeichnis	209
Abbildungsverzeichnis	223
Tabellenverzeichnis	229
Versicherung	231
Thesen (englisch)	233
Thesen (deutsch)	235
Lebenslauf	237

Acronyms

AAL	Ambient Assisted Living.
ADL	Activities of Daily Living.
CCTV	Closed Circuit Television.
COTS	Commercial Off-The-Shelf.
DOF	Degrees of Freedom.
FOV	Field of View.
FPGA	Field Programmable Gate Array.
FPS	Frames Per Second.
GMM	Gaussian Mixture Model.
HMD	Head Mounted Display.
HMI	Human Machine Interface.
LUT	Look-Up Table.
PVS	Panoramic Virtual Stereo Vision.
RGB-D	Red Green Blue-Depth.
SNR	Signal to Noise Ratio.
SSD	Sum of Squared Differences.
vPTZ	Virtual Pan-Tilt-Zoom.

Acronyms

WCS	World Coordinate System.
WTA	Winner Takes it All.

Symbols

\mathcal{A}	Affine transformation.
a	Physical pixel width.
a^*	Image resolvability.
α_x	Focal length normalized to the physical horizontal pixel length.
α_y	Focal length normalized to the physical vertical pixel length.
b	Physical baselength.
\mathbf{C}	3x1 origin vector.
c_x	Horizontal pixel density.
c_y	Vertical pixel density.
\mathcal{D}	Distortion component.
d	Disparity.
\mathbf{D}	Disparity map.
d_{\max}	Maximal measurable disparity.
d_{\min}	Minimal measurable disparity.
δa	Derivative of physical pixel width.
δb	Derivative of baselength.
δd	Derivative of disparity.
Δ_r	Radial distortion component.
Δ_t	Tangential distortion component.
δf	Derivative of focal length.
$\delta\phi$	Physical disparity error.
δZ	Derivative of depth.

\mathcal{F}	Radially symmetric projection function.
f	Focal length.
\mathbf{F}	Fundamental matrix.
\mathcal{F}_x	Function that transforms target pixel positions into a source pixel column.
\mathcal{F}_y	Function that transforms target pixel positions into a source pixel row.
\mathbf{H}	4x4 homography.
$\hat{\mathbf{H}}$	4x4 estimated homography.
height	Image height.
\mathbf{K}	Camera calibration matrix.
\mathcal{K}	Polynomial projection function.
\mathbf{k}	Vector of unit length.
i	Radial distortion coefficients of Kannala polynomial.
j	Tangential distortion coefficients of Kannala polynomial.
k	Symmetric coefficients of Kannala polynomial.
l	Radial distortion coefficients of Kannala polynomial.
m	Tangential distortion coefficients of Kannala polynomial.
λ	Arbitrary scale factor.
l_e	Epipolar line.
\mathbf{M}	3x3 sub matrix of 3x4 camera matrix.
m_x	Horizontal pixel density.
m_y	Vertical pixel density.
\mathbf{P}	3x4 projection matrix.
p	Component of camera matrix.
\mathcal{P}	Complete Kannala forward model.
\mathbf{p}	Kannala coefficients vector.
\mathbf{p}_{1C}	Column no. one of camera matrix.
\mathbf{p}_{1R}	Row no. one of camera matrix.
\mathbf{p}_{2C}	Column no. two of camera matrix.
\mathbf{p}_{2R}	Row no. two of camera matrix.
\mathbf{p}_{3C}	Column no. three of camera matrix.

\mathbf{p}_{3R}	Row no. three of camera matrix.
\mathbf{p}_{4C}	Column no. four of camera matrix.
p_x	Horizontal pixel length.
p_y	Vertical pixel length.
p	Probability density function.
ϕ_l	Angle of the left incoming ray on the principle plane with respect to the baseline.
ϕ_r	Angle of the right incoming ray on the principle plane with respect to the baseline.
p_n	Probability density function.
e	Epipole.
Φ	Incoming light ray.
ϕ	Physical disparity.
\mathbf{R}	3x3 rotation matrix.
R	Residuum.
ρ	Radius of the projected normalized sensor point with respect to the centre of the image.
\mathbf{T}	3x1 translation vector.
θ	Angle between the incoming light ray and the optical axis.
\mathbf{u}_r	Unit vector in radial direction.
\mathbf{u}_φ	Unit vector in tangential direction.
φ	Azimuth angle of the incoming light ray.
φ_x	Euler angle.
φ_y	Euler angle.
φ_z	Euler angle.
w	W-component.
w_{cam}	W-component, camera coordinates.
w_{img}	W-component, image coordinates.
w_{world}	W-component, world coordinates.
width	Image width.

x	X-component.
\mathbf{X}_{2D}	Point in 2D space.
$\tilde{\mathbf{X}}_{2D}$	Point in 2D space (inhomogeneous representation).
\mathbf{X}_{3D}	Point in 3D space.
$\tilde{\mathbf{X}}_{3D}$	Point in 3D space (inhomogeneous representation).
\mathbf{x}	X-Axis of camera coordinate system.
x_{cam}	X-component, camera coordinates.
x_{img}	X-component, image coordinates.
x_{norm}	X-component, normalized sensor coordinates.
x_{world}	X-component, world coordinates.
\mathbf{X}_{cam}	Camera point in 3D space.
$\tilde{\mathbf{X}}_{\text{cam}}$	Camera point in 3D space (inhomogeneous representation).
\mathbf{X}_{img}	Image point in 2D space.
$\tilde{\mathbf{X}}_{\text{img}}$	Image point in 2D space (inhomogeneous representation).
\mathbf{X}_{norm}	Normalized sensor point in 2D space.
$\tilde{\mathbf{X}}_{\text{norm}}$	Normalized sensor point in 2D Space (inhomogeneous representation).
$\mathbf{X}_{\text{world}}$	World point in 3D space.
$\tilde{\mathbf{X}}_{\text{world}}$	World point in 3D space (inhomogeneous representation).
y	Y-component.
\mathbf{y}	Y-Axis of camera coordinate system.
y_{cam}	Y-component, camera coordinates.
y_{img}	Y-component, image coordinates.
y_{norm}	Y-component, normalized sensor coordinates.
y_{world}	Y-component, world coordinates.
Z	Physical depth.
z	Z-component.
\mathbf{z}	Z-Axis of camera coordinate system.
z_{cam}	Z-component, camera coordinates.
Z_{max}	Maximal measurable physical depth.
Z_{min}	Minimal measurable physical depth.
Z^*	Depth resolvability.
z_{world}	Z-component, world coordinates.

Acknowledgement

The present thesis summarizes my current state of research related to omnidirectional stereo vision and has been accepted by the Faculty of Electrical Engineering and Information Technology at Chemnitz University of Technology as dissertation.

The examination of this research field has been a challenging and leisure time consuming endeavour. However, I think the achievements of this work vitally contribute to progressive innovation in this area and also generate new issues for further investigation.

Above all I would like to thank Prof. Dr.-Ing. Gangolf Hirtz for the extensive freedom he granted for elaborating my topic of choice. Furthermore, I offer my sincere thanks to the colleagues of the Professorship of Digital Signal Processing and Circuit Technology for their indefatigable assistance. Last but not least, I wish to express my gratitude to my family which guided me through that laborious time with patience and moral support.

Chemnitz, April 2015

Michel Findeisen

Chapter 1

Introduction

The research and development activities of the Professorship of Digital Signal Processing and Circuit Technology (Chemnitz University of Technology) focus on human behaviour analysis by means of optical sensors. The underlying motivation originates from the recent progress that has been achieved in the field of *visual surveillance* and, associated therewith, the emerge of new promising applications and markets as will be briefly presented in Section 1.1.

However, a rising spread of this technology amongst new (consumer) application fields gives rise to a multitude of associated requirements and hence generates new challenges for this field of research. Section 1.2 sketches important needs for development based on state-of-the-art indoor surveillance systems and thereby specifies the distinct topic of the thesis.

Finally Section 1.3 outlines the structure of the present document and hence reveals the proceeding for achieving the objective constituted in Section 1.2.

1.1 Visual Surveillance

The technology of Closed Circuit Television (CCTV) has been of common usage since the last century above all for applications like security and surveillance. Networks of cameras record non-stop video data, file it or directly communicate it to a control room where dedicated personnel assesses the material. By now modern machine vision based algorithms carry out this process by analysing an observed scene automatically. Furthermore, advances in optical sensor technologies allow a 3D measurement of a certain scene in the meantime rather than only perceiving 2D information. Augmenting image data by depth information principally generates the possibility to develop more powerful computer vision algorithms. For example the detection of a person becomes more stable because it can be reliably separated from the images background. In addition, by utilizing

3D information the localization of humans becomes a measurement rather than an estimation. Figure 1.1 provides a brief schematic of the variety of state-of-the-art 3D measurement techniques. However, for surveillance tasks the passive stereo vision method has turned out to be the most commonly used method. This is due to the fact that it can

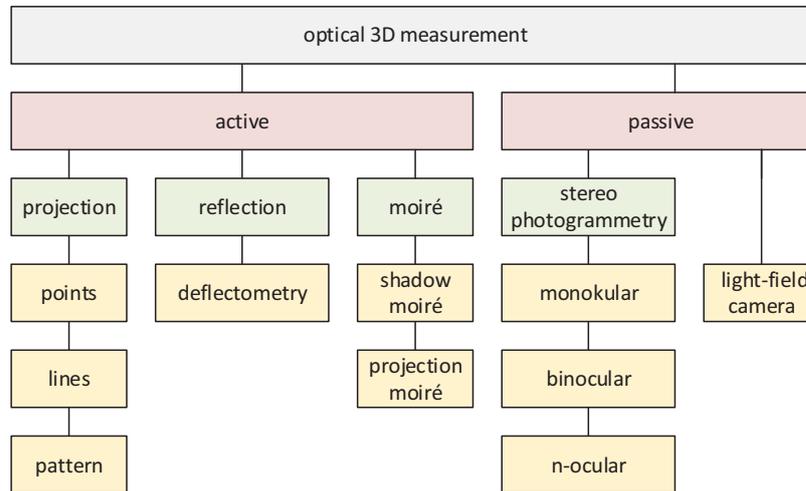


Figure 1.1: Optical 3D measurement techniques

be used freely without restriction, it is cost-efficient and suitable for monitoring people since it belongs to the *passive* approaches. It does not impact its vicinity.

The combination of 3D measurement capability and powerful miniaturized processing platforms with intelligent image processing algorithms makes sensors nowadays somehow *smart*: They detect and track people independently and communicate events like "*2 people at position x y*" rather than the raw video material. This saves technical bandwidth, personnel costs and is more sensitive to privacy concerns than ordinary cameras could ever be. Figure 1.2 exemplarily outlines such an optical smart sensor that is used for surveillance purposes.



Figure 1.2: Optical smart sensor based on passive stereo vision (source: Intenta GmbH, 2014)

In order to cover a complex environment like a public place or a whole apartment those sensors are combined to so-called *smart sensor networks* each analysing a part of the scene for itself and messaging any occasion to a human operator or computer. Figure 1.3 demonstrates the deployment of such a system in an artificial living environment.



Figure 1.3: Apartment equipped with a smart sensor network

For example the new upcoming Ambient Assisted Living (AAL) market claims safety and assistance functionality for elderly people that live at home alone. A smart sensor based system can provide functionalities like *fall detection*, *emergency alerting* and *home automation services* while paying attention to the inhabitants privacy concerns. Figure 1.4 shows a scenario where an elderly person that has fallen on the floor is guarded by an optical sensor system.



Figure 1.4: Emergency detection in a domestic environment (source: Vitracom AG, 2015)

A few more examples for relevant markets shall be stated in Table 1.1.

Table 1.1: Markets and application fields

Market	Explanation	Exemplary Applications
AAL	Age-based homes equipped with smart sensors take care of the elderly inhabitants and assist with their everyday live.	Fall detection, emergency alerting, home automation services, assistance for carrying out ADLs, care optimization by evaluation of statistical data of the inhabitants behaviour and health status
Security	Private homes as well as public buildings are equipped with intelligent sensors that message suspicious or violent behaviour.	Intrusion detection, access control, crowd density monitoring abandoned object detection, theft detection
Building Automation	Smart sensors detect and count people to make buildings somehow intelligent by automatically performing certain services.	Air conditioning control, access control, energy management, HMI for established building automation systems e.g. of Honeywell International Inc.
Retail	Smart sensor systems for <i>retail business intelligence</i> collect data about the number of customers and their purchasing behaviour.	Managing queues of customers at cash boxes, analysing the shop performance
Industry	Perspectively, manual manufacturing processes will be guided by automatic systems that take care of production policies, safety and logistical concerns.	Safety in production by intrusion or tiredness detection, process monitoring

1.2 Challenges in Visual Surveillance

Reviewing state-of-the art visual surveillance technology and its markets, the system costs turn out to be an everlasting crucial issue. One of the key cost drivers is the number of applied sensors for observing a certain scene. Hence, it is obvious that a reduction of necessary sensors (and the attached infrastructure) can have an essential impact on the systems price.

Complex scenes as for example a living environment usually require 2 to 3 ordinary perspective sensors per room for almost full coverage. Due to their nature of perspective projection they are restricted to a Field of View (FOV) of much less than 180° . Special devices called omnidirectional sensors work either with catadioptric mirror configurations or are equipped with fish-eye lenses in order to generate a FOV of more than 180° . Figure 1.5 shows an exemplary living environment observed by a perspective, and for comparison, an omnidirectional imaging device. The latter generates a much larger visible area and is hence appropriate to replace multiple perspective sensors.



Figure 1.5: Standard perspective camera versus omnidirectional camera in a home environment

While in earlier times image processing algorithms almost exclusively utilized color information, nowadays a growing number of algorithms claim 3D data. As an example the robust Z-Map tracking algorithm published by Richter et al [99] shall be mentioned, which utilizes 3D point clouds generated by standard stereo vision devices.

The combination of omnidirectional sight with stereo vision based depth measurement hence could merge the advantages of a large FOV together with 3D information. Since accuracy of the most depth sensing principles depends on the distance from sensor to target object, the propagated idea has another advantage: An omnidirectional Red Green Blue-Depth (RGB-D) sensor mounted in the central of a rooms ceiling would provide on average a lower distance to the scene points under observation in comparison to a common configuration that applies a network of perspective stereo sensors. This issue is sketched in Figure 1.6 where the floor plan of an exemplary living environment is sketched with a standard sensor configuration on the left and an omnidirectional sensor on the right. The sensors are indicated by their individual frame. The shortest distance of each scene point to the adjacent sensor that observes it is highlighted color-coded. White areas are covered by neither camera.

It can be seen that the (artificial) omnidirectional depth measuring device can replace three standard perspective sensors approximately, while optimizing the sensor to target distance in addition.

Motivated by this consideration, the objective for this thesis will be the investigation of a stereo vision based omnidirectional measurement device that is able to cover one

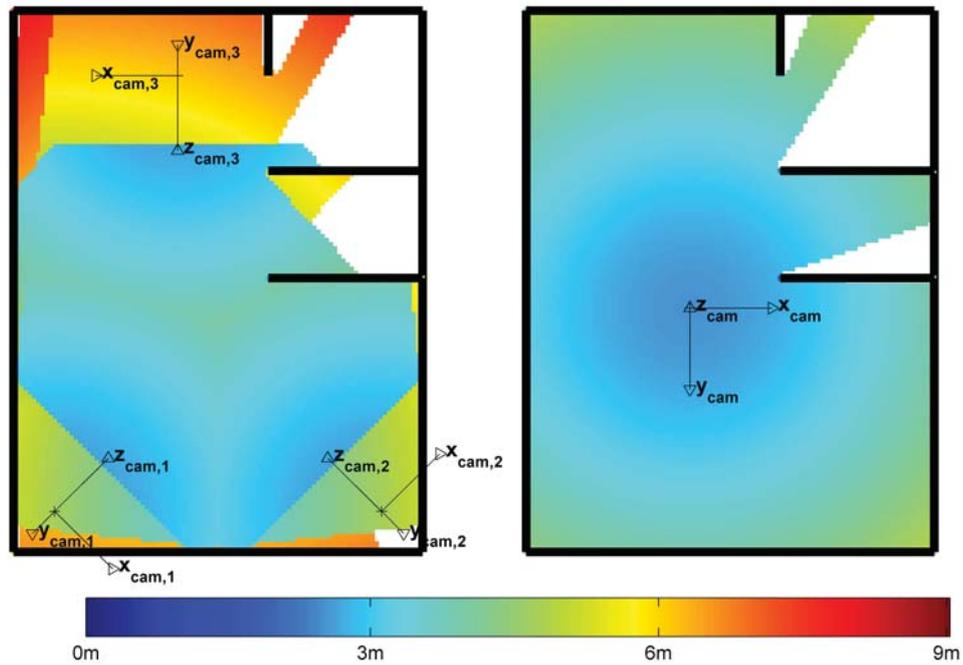


Figure 1.6: Floor plan of a single apartment equipped with optical stereo vision sensors (indicated by their camera frames)

complete room from floor to ceiling with almost homogeneous measurement accuracy and preferably low computational effort.



Figure 1.7: Exemplary application fields (upper left: home automation, upper right: industry, lower left: security in public transportation, lower right: security in public buildings)

Figure 1.7 projects this demand onto different target application fields.

1.3 Outline of the Thesis

The present script is organized as follows:

Chapter 2 - Fundamentals of Computer Vision Geometry

As a starting point relevant basics of computer vision are structured and elaborated for certain aspects that are important for this work. This is done in Chapter 2.

Chapter 3 - Fundamentals of Stereo Vision

Stereo vision fundamentals for standard perspective camera geometry are presented successively in Chapter 3. This technology plays a key role in this work and will be transferred to other camera models subsequently.

Chapter 4 - Virtual Perspective Cameras

For the simultaneous usage of different camera models and their conversion, it is important to understand the concept of virtual cameras. In Chapter 4 several conversion strategies from omnidirectional to perspective vision are investigated exemplarily.

Chapter 5 - Spherical Stereo Vision

The research field of stereo vision with respect to multiple omnidirectional geometries is revised in Chapter 5. Different approaches are compared and assessed with respect to the application of visual surveillance.

Chapter 6 - A Novel Spherical Stereo Vision Algorithm

Based on the knowledge presented in Chapter 2, Chapter 3, Chapter 4 and Chapter 5 a new method for omnidirectional stereo processing is presented and simulated in Chapter 6.

Chapter 7 - Spherical Stereo Vision Demonstrator

A real demonstrator compiled of standard industry cameras is presented in Chapter 7.

Chapter 8 - Discussion and Outlook

Finally, in Chapter 8 the so far achieved results of this thesis are discussed and open points collected. In addition, possible improvements for the presented concept are suggested. A sample application that employs omnidirectional RGB-D data concludes this thesis practically.

Chapter 2

Fundamentals of Computer Vision Geometry

This chapter discusses fundamental topics related to computer vision geometry. As starting point vital basics for homogeneous and inhomogeneous coordinate representations are briefly reviewed in Section 2.1. After this the principles of transferring a 3D point located in an observed scene to a 2D image point is treated in Section 2.2 while distinguishing different camera models. Subsequently a brief overview of current state-of-the-art methods for estimating the parameters of different camera models for the usage in real camera devices is presented in Section 2.3. Finally an insight into the fundamental mechanics of epipolar geometry is given in Section 2.4. The constraints that two adjacent cameras introduce on the relationships between their images are explained.

2.1 Projective Geometry

In the following sections the most important geometrical point representations for two- and three-dimensional space are explained - more precisely *homogeneous* and *inhomogeneous* representations.

2.1.1 Euclidean Space

The most common form of representing a point is the expression of *cartesian coordinates*. Let \mathbf{X}_{2D} and \mathbf{X}_{3D} be points in 2D and 3D space, than one can represent them as shown

in Equation 2.1

$$\mathbf{X}_{2D} = \begin{pmatrix} x \\ y \end{pmatrix}, \mathbf{X}_{3D} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (2.1)$$

where x , y and z are the vector components in the direction of the x-, y- and z-axis, respectively. This model is commonly used in the *euclidean space*. Different scalar values of the components represent different points.

2.1.2 Projective Space

In the *projective space* the distinct values of the vectors components are not significant for determining a unique point, but so are the ratios between the values. For example in this space a point $\mathbf{X}_1 = (1, 2, 3)^T$ is equivalent to a point $\mathbf{X}_2 = (2, 4, 6)^T$, since both head for the same direction. Although a point $\mathbf{X}_3 = (\infty, \infty, \infty)^T$ resides at infinity it can be equivalent to \mathbf{X}_1 and \mathbf{X}_2 as well if their direction is identical.

In order to represent finite and infinite points *homogeneously*, the *homogeneous representation* was introduced, by adding an additional scale factor w . This new parameter encodes the scale of the vector while the others represent its direction. Consequently \mathbf{X}_{2D} and \mathbf{X}_{3D} are now given as 3- and 4-dimensional vectors in 2- and 3-dimensional space and are called the *homogeneous representation*:

$$\mathbf{X}_{2D} = \begin{pmatrix} x \\ y \\ w \end{pmatrix}, \mathbf{X}_{3D} = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \quad (2.2)$$

The *inhomogeneous representation* (without w) is marked in the following with a *tilde* as $\tilde{\mathbf{X}}_{2D}$ and $\tilde{\mathbf{X}}_{3D}$:

$$\tilde{\mathbf{X}}_{2D} = \begin{pmatrix} x/w \\ y/w \end{pmatrix}, \tilde{\mathbf{X}}_{3D} = \begin{pmatrix} x/w \\ y/w \\ z/w \end{pmatrix} \quad (2.3)$$

The afore-mentioned example points have the following homogeneous and inhomogeneous representations:

$$\begin{aligned} \mathbf{X}_1 = (1, 2, 3, 1)^T &\rightarrow \tilde{\mathbf{X}}_1 = (1, 2, 3)^T \\ \mathbf{X}_2 = (1, 2, 3, 0.5)^T &\rightarrow \tilde{\mathbf{X}}_2 = (2, 4, 6)^T \\ \mathbf{X}_3 = (1, 2, 3, 0)^T &\rightarrow \tilde{\mathbf{X}}_3 = (\infty, \infty, \infty)^T \end{aligned} \quad (2.4)$$

From now on the distinction between both models is made.

2.2 Camera Geometry

A camera usually maps an observed 3D scene to a 2D image. The geometrical relationship between these 3D scene points $\tilde{\mathbf{X}}_{3D}$ and the resulting 2D image points $\tilde{\mathbf{X}}_{2D}$ is often quite complex due to the physical construction of the imaging device.

This complex issue can be approximated by an appropriate mathematical model which is commonly known as *camera model* or *projection model*. By employing a finite amount of numerical parameters, the actual process of projection can be replaced by an appropriate mathematical relationship of bearable computational cost.

Since the original scene points refer to a certain 3D world coordinate system, they are from now on called $\mathbf{X}_{\text{wrlld}}$. Since the target 2D points refer to an image coordinate system, they are from now on called \mathbf{X}_{img} :

$$\mathbf{X}_{\text{wrlld}} = \begin{pmatrix} x_{\text{wrlld}} \\ y_{\text{wrlld}} \\ z_{\text{wrlld}} \\ w_{\text{wrlld}} \end{pmatrix}, \mathbf{X}_{\text{img}} = \begin{pmatrix} x_{\text{img}} \\ y_{\text{img}} \\ w_{\text{img}} \end{pmatrix} \quad (2.5)$$

2.2.1 Geometrical Imaging Process

The overall imaging process that transfers a scene point $\mathbf{X}_{\text{wrlld}}$ to an image point \mathbf{X}_{img} successively utilizes an *extrinsic* and an *intrinsic model* as can be seen in Figure 2.1:

- The former describes the geometrical relationship of a global reference, the World Coordinate System (WCS), with respect to the camera itself, the camera coordinate system. It considers $\mathbf{X}_{\text{wrlld}}$ with respect to a world reference as a *world point* and transfers it to a *camera point* \mathbf{X}_{cam} . See Section 2.2.1.2.
- The latter describes the projection process from camera points \mathbf{X}_{cam} to image points \mathbf{X}_{img} in the image coordinate system. See Section 2.2.1.1.

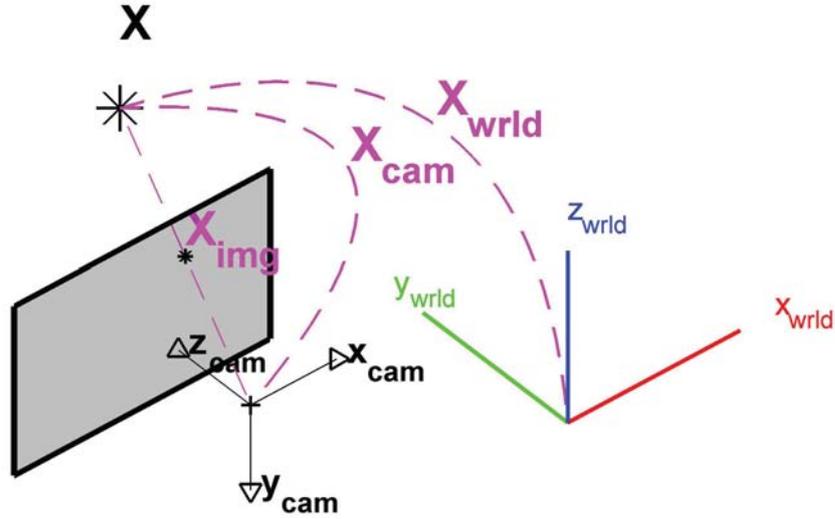


Figure 2.1: Coordinate transformations in the geometrical imaging process

2.2.1.1 Projection Models

The projection model describes the mapping from incoming light rays, denoted by \mathbf{X}_{cam} , to normalized sensor coordinates, denoted by \mathbf{X}_{norm} , whereby

$$\mathbf{X}_{\text{cam}} = \begin{pmatrix} x_{\text{cam}} \\ y_{\text{cam}} \\ z_{\text{cam}} \\ w_{\text{cam}} \end{pmatrix} \quad \text{and} \quad \mathbf{X}_{\text{norm}} = \begin{pmatrix} x_{\text{norm}} \\ y_{\text{norm}} \\ 1 \end{pmatrix} \quad (2.6)$$

are the corresponding homogeneous Cartesian representations in camera and normalized sensor coordinate system respectively.

A special group of projection principles are so called *radially symmetric models*. Therefore it may be suitable to note the spherical and the polar representation respectively:

$$\mathbf{X}_{\text{cam}} \hat{=} \begin{pmatrix} \theta \\ \varphi \\ w_{\text{cam}} \end{pmatrix} \quad \text{and} \quad \mathbf{X}_{\text{norm}} \hat{=} \begin{pmatrix} \rho \\ \varphi \\ 1 \end{pmatrix} \quad (2.7)$$

The following conventions apply for these mathematical expressions:

- The variable θ is the angle between the incoming light ray and the optical axis.
- The variable φ forms the azimuth angle of the incoming light ray.
- The variable ρ forms the radius of the projected normalized sensor point with respect to the centre of the image.

In Figure 2.2 the relationship of the coordinates of \mathbf{X}_{cam} before and of \mathbf{X}_{norm} after the projection by a radially symmetric model can be seen.

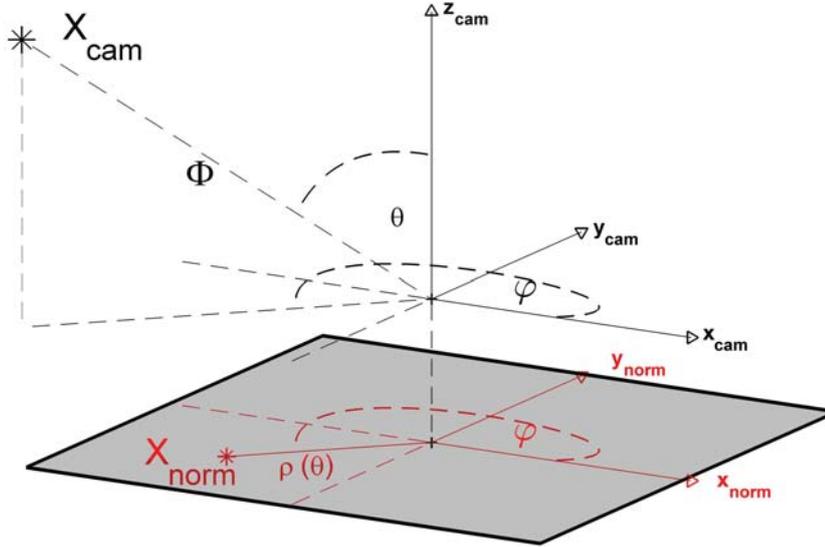


Figure 2.2: Projection process

According to [56] the radial projection of an incoming light ray $\Phi = (\theta, \varphi)^T$ onto a virtual image plane with a distance of $f = 1$ to the projection center can be modelled with the radial projection function

$$\tilde{\mathbf{X}}_{\text{norm}} = \mathcal{F}(\Phi) = \rho(\theta)\mathbf{u}_r(\varphi) \text{ with } \mathbf{u}_r(\varphi) = \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \quad (2.8)$$

where $\rho(\theta)$ is the projection radius as a function of θ and \mathbf{u}_r is the unit vector in radial direction.

In Table 2.1 a variety of the most important radially symmetric projection models are summarized as presented in [56]. Asymmetric effects of projection are not covered by those models.

In Figure 2.3 the projection principles of Table 2.1 are visualized.

For projection principles called *non-radially symmetric models* it is more appropriate to model the projection by employing Cartesian coordinates:

$$\tilde{\mathbf{X}}_{\text{norm}} = \mathcal{F}(\tilde{\mathbf{X}}_{\text{cam}}) \quad (2.14)$$

2.2.1.2 Intrinsic Model

Modern cameras employ a discrete number of sensor or picture elements on the light sensitive chip. Image information is thereby described by image or pixel coordinates

\mathbf{X}_{img} resulting from an affine transformation \mathbf{K} on the normalized sensor coordinates:

$$\mathbf{X}_{\text{img}} = \mathbf{K} \cdot \mathbf{X}_{\text{norm}} \quad (2.15)$$

where \mathbf{K} , called *camera calibration matrix*, is given by

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & c_x \\ & \alpha_y & c_y \\ & & 1 \end{bmatrix} \text{ with } \alpha_x = f \cdot m_x, \alpha_y = f \cdot m_y \quad (2.16)$$

containing the five intrinsic parameters of a CCD camera. These parameters are briefly explained further down as a list.

Table 2.1: Radially symmetric projection models

Projection Type	Function	Properties and Application
Perspective projection	$\rho = f \tan \theta$ (2.9)	Applied in conventional cameras that follow the pinhole principle with $\theta < \pi/2$ [rad]. See Section 2.2.2.
Stereographic projection	$\rho = 2f \tan \frac{\theta}{2}$ (2.10)	Commonly used for the generation of stellar maps and photography. Angles are preserved (equal angles) and circles are imaged without distortions (equal circles).
Equidistance projection	$\rho = f\theta$ (2.11)	It is the most popular model for describing fisheye projection lenses. Incidence angles of rays are imaged linearly. See Section 2.2.3.
Equisolid angle projection	$\rho = 2f \sin \frac{\theta}{2}$ (2.12)	Each pixel corresponds to one distinct solid angle which makes the image look like the reflection by a specular ball.
Orthogonal projection	$\rho = f \sin \theta$ (2.13)	This camera principle models the projection onto a sphere. The projection radius ρ is increasing with increasing incidence angle θ until θ reaches $\pi/2$. Then it is declining again - cf. Figure 2.3.

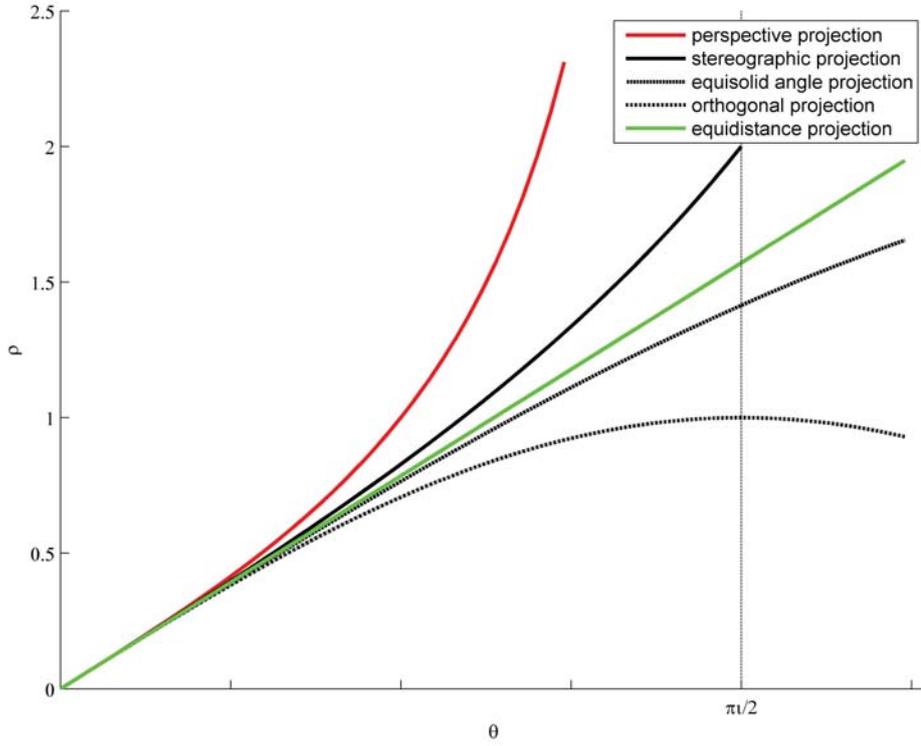


Figure 2.3: Radially symmetric projections

- α_x is the focal length f normalized to the physical horizontal pixel length p_x . $[\alpha_x] = \frac{px}{mm}$. The inverse of p_x is called horizontal pixel density m_x . $[m_x] = \frac{1}{mm}$
- α_y is the focal length f normalized to the physical vertical pixel length p_y . $[\alpha_y] = \frac{px}{mm}$. The inverse of p_y is called vertical pixel density m_y . $[m_y] = \frac{1}{mm}$
- c_x is the horizontal offset value of the principle point from the pixel frame. $[c_x] = px$.
- c_y is the vertical offset value of the principle point from the pixel frame. $[c_y] = px$.
- s is the affine distortion of the image in horizontal direction, called skewness factor. $[s] = mm \cdot rad$.

For an ideal camera the offset values c_x and c_y can be directly exploited to calculate the image size {height,width}. The relationship is outlined in Equation 2.17.

$$\{c_x, c_y\} = \left\{ \frac{\text{width} + 1}{2}, \frac{\text{height} + 1}{2} \right\} \quad (2.17)$$

2.2.1.3 Extrinsic Model

Generally the scene point is expressed in a world coordinate system $\mathbf{X}_{\text{world}}$ which is different from the camera coordinate system and which is useful especially when there are multiple cameras in the scene - see Figure 2.4. The geometrical relationship between

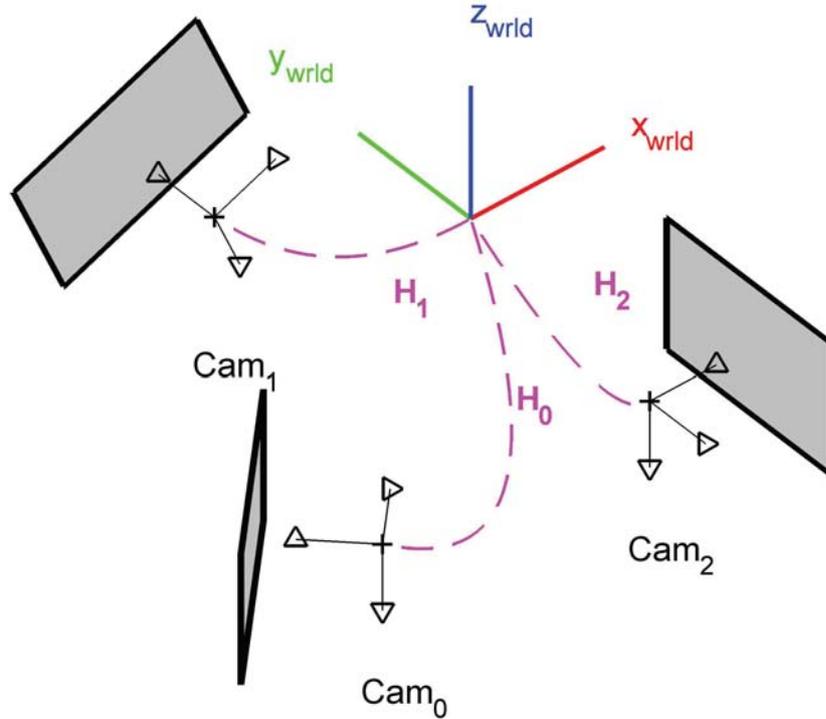


Figure 2.4: World coordinate system

$\mathbf{X}_{\text{world}}$ the coordinate systems $\mathbf{X}_{\text{cam},i}$ (with i being the index of the distinct camera) is composed of a 3×3 rotation matrix \mathbf{R}_i and a 3×1 translation vector \mathbf{T}_i . A point given in world coordinates can hence be transferred into camera coordinates by applying Equation 2.18:

$$\tilde{\mathbf{X}}_{\text{cam},i} = \mathbf{R}_i \cdot \tilde{\mathbf{X}}_{\text{world}} + \mathbf{T}_i \quad (2.18)$$

Using homogeneous coordinates, this procedure can be expressed as matrix operation:

$$\mathbf{X}_{\text{cam},i} = \mathbf{H}_i \cdot \mathbf{X}_{\text{world}} \quad (2.19)$$

where \mathbf{H}_i denotes the 4×4 *homography matrix* describing the world coordinate system with respect to the camera coordinate system. \mathbf{H}_i can be decomposed as follows:

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{R}_i & \mathbf{T}_i \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_i & -\mathbf{R}_i \mathbf{C}_i \\ \mathbf{0} & 1 \end{bmatrix} \quad (2.20)$$

where \mathbf{C}_i is a 3×1 vector which encodes the position of the camera in world coordinates.

Furthermore \mathbf{R} can be decomposed into the angles φ_x , φ_y and φ_z . They represent the rotation of the camera with respect to the WCS as a consecutive turn of the camera around its own x-, y- and z-axis respectively. These angles are called *Euler rotations*.

2.2.1.4 Distortion Models

Every projection model has the limitation that it approximates real Commercial Off-The-Shelf (COTS) cameras at its best due to manufacturing issues. Furthermore even lenses of the same type differ in their imaging behaviour. These properties are denoted as distortions. It is commonly distinguished between:

- **Radial distortions:** Especially wide-angle lenses of cheap production commonly give rise to deviations of the projection result from the invoked model in radial direction.
- **Tangential distortions:** Especially skew assembling of the lens and die in a camera lead to tangential deviations of the projection result.

There are several approaches available to extend the camera models with a description of these lens imperfections. See Section 2.2.2 and Section 2.2.4.

2.2.2 Pinhole Camera Model

Most conventional cameras which follow the perspective camera model use the pinhole camera model as an approximation. Essentially it is a perspective projection of the spatial coordinates \mathbf{X}_{cam} given in the camera coordinate system denoted by

$$\mathbf{X}_{\text{cam}} = \begin{pmatrix} x_{\text{cam}} \\ y_{\text{cam}} \\ z_{\text{cam}} \\ 1 \end{pmatrix} \quad (2.21)$$

into normalized sensor plane coordinates \mathbf{X}_{norm}

$$\mathbf{X}_{\text{norm}} = \begin{pmatrix} x_{\text{norm}} \\ y_{\text{norm}} \\ 1 \end{pmatrix} = \frac{1}{z_{\text{cam}}} \begin{pmatrix} f \cdot x_{\text{cam}} \\ f \cdot y_{\text{cam}} \\ z_{\text{cam}} \end{pmatrix} = \frac{1}{z_{\text{cam}}} \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 \end{bmatrix} \mathbf{X}_{\text{cam}} \quad (2.22)$$

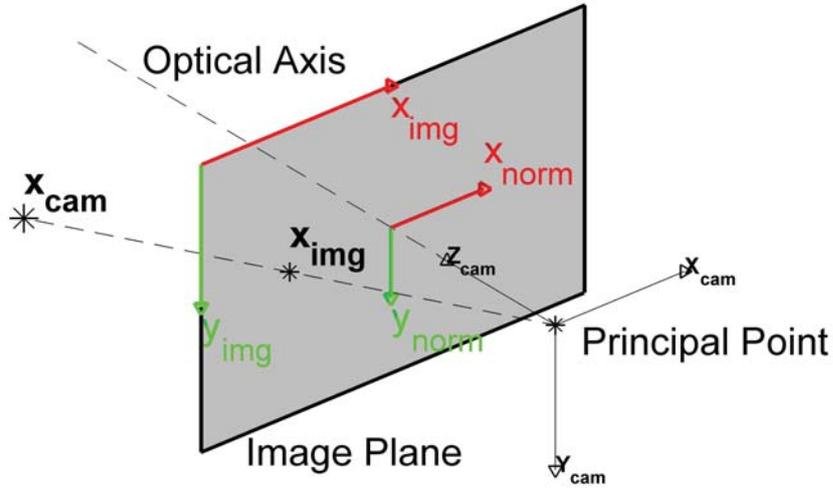


Figure 2.5: Perspective camera model

whereby f forms the focal length of the camera and hence the distance between the principal point and the sensor plane of the camera. In Figure 2.5 the camera model presented so far is visualized.

Similarly a radially symmetric projection function \mathcal{F}_p can describe the projection process by combining Equation 2.8 with Equation 2.9:

$$\tilde{\mathbf{X}}_{\text{norm}} = \mathcal{F}_p(\Phi) = f \tan \theta \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \hat{=} \left(f \frac{x_{\text{cam}}}{z_{\text{cam}}}, f \frac{y_{\text{cam}}}{z_{\text{cam}}} \right)^T \quad (2.23)$$

2.2.2.1 Complete Forward Model

Perspective projection model

Integrating all the previous relations, a 3D point $\mathbf{X}_{\text{world}}$ can be directly transferred into image coordinates by setting up a 3×4 camera projection matrix \mathbf{P} :

$$\mathbf{X}_{\text{img}} = \mathbf{P} \cdot \mathbf{X}_{\text{world}} \quad (2.24)$$

This matrix is composed of the *intrinsic parameters* \mathbf{K} as well as *extrinsic parameters* \mathbf{R} and \mathbf{T} :

$$\mathbf{P} = \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \quad (2.25)$$

Altogether this model provides 11 Degrees of Freedom (DOF) for modelling a perspective camera.

In a mathematical sense, the matrix \mathbf{P} is the projection of a four dimensional space (3D homogeneous points) to a three dimensional space (2D homogeneous points). Amongst other possibilities, \mathbf{P} (with the 12 scalar elements p_{11} to p_{34}) can be represented as

- a column vector of components \mathbf{p}_{1R} , \mathbf{p}_{2R} and \mathbf{p}_{3R} where each element is a 1×4 row vector.
- a row vector of components \mathbf{p}_{1C} , \mathbf{p}_{2C} , \mathbf{p}_{3C} and \mathbf{p}_{4C} where each element is a 3×1 column vector.

This relationship is highlighted in Equation 2.26.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_{1R} \\ \mathbf{p}_{2R} \\ \mathbf{p}_{3R} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_{1C} & \mathbf{p}_{2C} & \mathbf{p}_{3C} & \mathbf{p}_{4C} \end{bmatrix} \quad (2.26)$$

Decomposing \mathbf{P} into row vectors means to represent it by a number of special planes:

- \mathbf{p}_{1R} is the plane orthogonal to the camera's x-axis containing the principal point and origin of the image coordinate system. It is visualized in Figure 2.6 as red plane.
- \mathbf{p}_{2R} is the plane orthogonal to the camera's y-axis containing the principal point and origin of the image coordinate system. It is visualized in Figure 2.6 as green plane.
- \mathbf{p}_{3R} is the principal plane. It is visualized in Figure 2.6 as blue plane.

Decomposing \mathbf{P} into column vectors means to represent it by a number of special points:

- \mathbf{p}_{1C} , \mathbf{p}_{2C} and \mathbf{p}_{3C} are the projections of the x-, y- and z-axis directions of the world frame.
- \mathbf{p}_{4C} is the projection of the origin of the world frame.

2.2.2.2 Back Projection

The operation of projecting a 3D point $\mathbf{X}_{\text{wrlld}}$ to an image point \mathbf{X}_{img} (see Equation 2.24) is not reversible since \mathbf{P} is not invertible. In other words, a 2D image point cannot be used to recover a 3D world point. Rather than this, it can be used to recover a ray of points by employing the pseudo inverse of \mathbf{P} , denoted by \mathbf{P}^+ :

$$\mathbf{X}_{\text{wrlld}}(\lambda) = \lambda \cdot \mathbf{P}^+ \mathbf{X}_{\text{img}} + \mathbf{C} \text{ with } \mathbf{P}^+ = \mathbf{P}^T (\mathbf{P}\mathbf{P}^T)^{-1} \quad (2.27)$$

At this point λ represents an arbitrary scale factor and hence the uncertainty of the world point recovery.

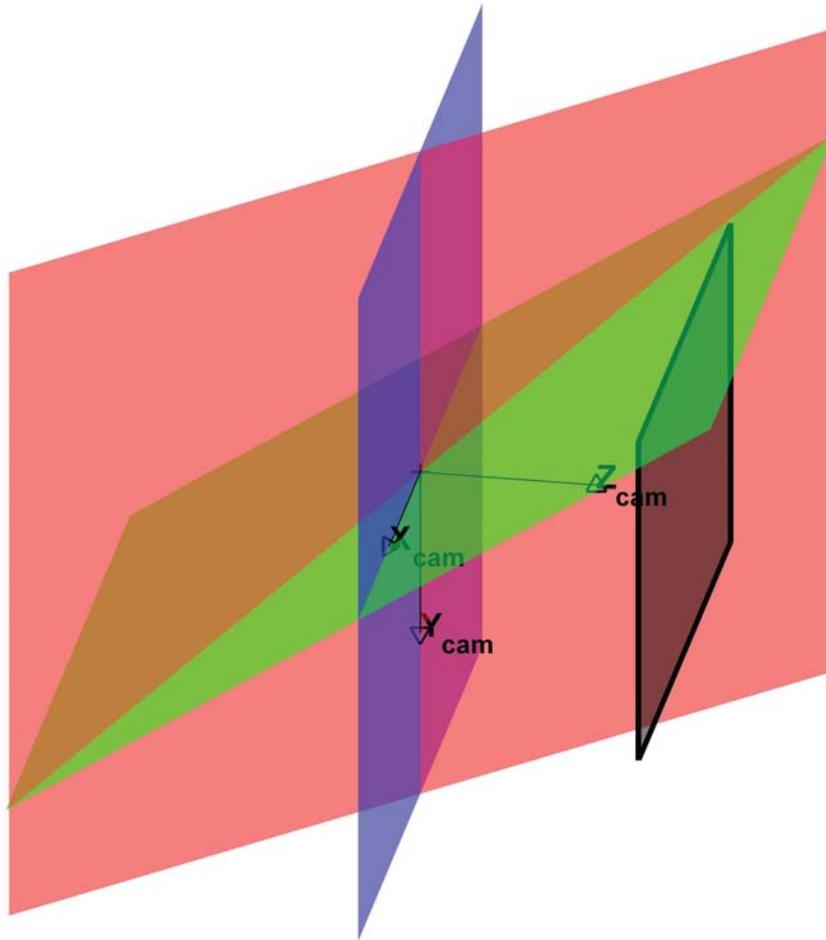


Figure 2.6: Decomposition of P (row vectors)

2.2.3 Equiangular Camera Model

A further relevant projection principle is called equiangular camera model. It has the inherent property that each pixel covers the same solid angle. The full projection process \mathcal{F}_e of this model can be obtained by combining Equation 2.11 with Equation 2.8:

$$\tilde{\mathbf{X}}_{\text{norm}} = \mathcal{F}_e(\Phi) = f\theta \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \quad (2.28)$$

Since the norm of the unit vector \mathbf{u}_r is linearly dependent on θ , the projection behaves as if the sensor or image plane would have been bend spherically. See [110, 111] and Figure 2.7.

This equiangular model overcomes the limitation of the perspective model in terms of FOV. It is at least theoretically not limited anymore. Figure 2.8 shows an exemplary image that approximates the projection of Equation 2.28.

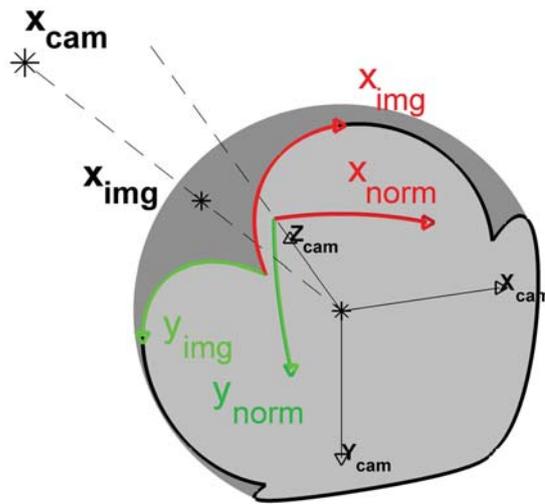


Figure 2.7: Equiangular camera model

In the remainder of the document each model that obeys or approximates the equiangular projection model will also be denoted as *omnidirectional camera model*.

2.2.4 Generic Camera Models

An alternative approach compared to standard projection models as presented in Section 2.2.2 and Section 2.2.3 is to employ a generic model. It has the advantage to be able to imitate almost any other model *generically*. In recent years, two methods have turned out to be most important: The approaches of *Davide Scaramuzza* and *Juho Kannala*. Relevant publications have been summarized in Table 2.2.



Figure 2.8: Omnidirectional image

Table 2.2: Generic projection models

Projection Model	Properties and Application	Publications
Kannala's Camera Model	Describes a variety of common as well as omnidirectional camera types. Claims to sufficiently approximate all types of Table 2.1.	[53, 55–57]
Scaramuzza's Camera Model	Describes a variety of common as well as omnidirectional camera types. It puts emphasis on catadioptric imaging systems.	[102–104]

The model proposed by Juho Kannala shall be examined here in more detail, while the method of Davide Scaramuzza has been mentioned as an alternative way. In [56] Juho Kannala describes his method to comprise three components: a *radially symmetric projection model*, an *asymmetric distortion model* and a *transformation to image coordinates*.

Together these components describe the so called *complete forward model* to transfer a 3D point \mathbf{X}_{cam} into an image point \mathbf{X}_{img} . The inverse way is called *back projection* - cf. Section 2.2.1.1 and Section 2.2.1.2. In the following both are outlined briefly.

2.2.4.1 Complete Forward Model

Radial symmetric projection model

The radial symmetric model proposed by [56] describes the projection (Section 2.2.1) in a general polynomial form

$$\rho(\theta) = \mathcal{K}(\theta) = k_1\theta + k_2\theta^3 + k_3\theta^5 + k_4\theta^7 + k_4\theta^9 + \dots \quad (2.29)$$

where the even powers are neglectable, as the projection is radially symmetric related to the projection center. By parametrizing the polynomial appropriately, all common projection types can be approximated sufficiently. According to [56] the first five coefficients are sufficient. Exemplarily the Taylor series of the pinhole camera model with $\rho(\theta) = \tan \theta$, and hence its approximation can be parametrized as follows:

$$\rho(\theta) = \theta + \frac{1}{3}\theta^3 + \frac{2}{15}\theta^5 + \frac{17}{315}\theta^7 + \frac{62}{2835}\theta^9 + R(\theta) \quad (2.30)$$

with $R(\theta)$ being the residuum. For the perfect equiangular projection it reduces just to

$$\rho(\theta) = \theta \quad (2.31)$$

With Equation 2.8 and Equation 2.29 a generic projection function \mathcal{F}_g can be established:

$$\tilde{\mathbf{X}}_{\text{norm}} = \mathcal{F}_g(\Phi) = \mathcal{K}(\theta) \mathbf{u}_r(\varphi) \text{ with } \mathbf{u}_r(\varphi) = \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \quad (2.32)$$

Asymmetric distortion model

Kannala proposed two terms $\Delta_r(\theta, \varphi)$ and $\Delta_t(\theta, \varphi)$ which extend Equation 2.29 in order to model distortion effects:

$$\begin{aligned} \tilde{\mathbf{X}}_{\text{norm}} = & \rho(\theta) \mathbf{u}_r(\varphi) \\ & + \Delta_r(\theta, \varphi) \mathbf{u}_r(\varphi) + \Delta_t(\theta, \varphi) \mathbf{u}_\varphi(\varphi) \end{aligned} \quad (2.33)$$

where $\mathbf{u}_r(\varphi)$ is the unit vector in radial and $\mathbf{u}_\varphi(\varphi)$ is the unit vector in tangential direction. Furthermore $\Delta_r(\theta, \varphi)$ and $\Delta_t(\theta, \varphi)$ are the radial and tangential distortion components given by

$$\begin{aligned} \Delta_r(\theta, \varphi) = & (l_1\theta + l_2\theta^3 + l_3\theta^5) \\ & \cdot (i_1 \cos \varphi + i_2 \sin \varphi + i_3 \cos 2\varphi + i_4 \sin 2\varphi) \end{aligned} \quad (2.34)$$

and

$$\begin{aligned} \Delta_t(\theta, \varphi) = & (m_1\theta + m_2\theta^3 + m_3\theta^5) \\ & \cdot (j_1 \cos \varphi + j_2 \sin \varphi + j_3 \cos 2\varphi + j_4 \sin 2\varphi) \end{aligned} \quad (2.35)$$

where $\mathbf{p}_a = (l_1, \dots, l_3, m_1, \dots, m_3, i_1, \dots, i_4, j_1, \dots, j_4)$ are the 14 asymmetric distortion parameters.

The following properties shall be mentioned:

- The distortion functions are separable in θ and φ , which makes the estimation process much less complex.
- The left term denotes a Taylor series employing an odd polynomial.
- The right term denotes Fourier series that is periodic in 2π . Since every periodic function $f : [-\pi, +\pi]$ that converges in the L^2 -space ($\int_{-\pi}^{+\pi} f(x)dx < \infty$) can be represented by a Fourier series, in principle every kind of distortion can be modelled.

Hence Equation 2.8 can be extended by the afore-mentioned distortion terms, now denoted as distortion component \mathcal{D} :

$$\tilde{\mathbf{X}}_{\text{norm}} = (\mathcal{D} \circ \mathcal{F}_g)(\Phi) \quad (2.36)$$

Transformation to image coordinates

To finally transfer normalized sensor coordinates into image coordinates, a further affine transformation \mathcal{A} is applied in accordance with Section 2.2.1.2:

$$\tilde{\mathbf{X}}_{\text{img}} = \begin{pmatrix} x_{\text{img}} \\ y_{\text{img}} \end{pmatrix} = \mathcal{A}(\tilde{\mathbf{X}}_{\text{norm}}) = \begin{bmatrix} \alpha_x & 0 \\ 0 & \alpha_y \end{bmatrix} \tilde{\mathbf{X}}_{\text{norm}} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} \quad (2.37)$$

Complete forward model

Altogether the camera model defined above contains 23 camera parameters. Therefore it is denoted by \mathbf{p}_{23} . Leaving out the asymmetric part it is reduced to \mathbf{p}_9 . Using only two instead of five parameters for the radial component we obtain \mathbf{p}_6 . The resulting parameter vectors are given by

$$\begin{aligned} \mathbf{p}_6 &= (k_1, k_2, \alpha_x, \alpha_y, c_x, c_y)^T \\ \mathbf{p}_9 &= (k_1, k_2, \alpha_x, \alpha_y, c_x, c_y, k_3, k_4, k_5)^T \\ \mathbf{p}_{23} &= (k_1, k_2, \alpha_x, \alpha_y, c_x, c_y, k_3, k_4, k_5, l_1 \dots l_3, i_1 \dots i_4, j_1 \dots j_4)^T \end{aligned} \quad (2.38)$$

Summarizing all the operations, the whole forward camera model $\mathcal{P}(\Phi)$ can be denoted as:

$$\tilde{\mathbf{X}}_{\text{img}} = \mathcal{P}(\Phi) = (\mathcal{A} \circ \mathcal{D} \circ \mathcal{F}_g)(\Phi) \quad (2.39)$$

2.2.4.2 Back Projection

Back projection is the process of recovering a ray \mathbf{X}_{cam} parametrized by Φ (compare with Equation 2.7 and Equation 2.8) from an image point \mathbf{X}_{img} .

$$\mathbf{X}_{\text{cam}} \hat{=} \Phi = (\mathcal{F}_g^{-1} \circ \mathcal{D}^{-1} \circ \mathcal{A}^{-1}) \tilde{\mathbf{X}}_{\text{img}} \quad (2.40)$$

2.3 Camera Calibration Methods

Since each projection model that shall be applied needs to be parametrized appropriately, in the following common methods for the process denoted as camera calibration will be presented at a glance. Provided that, the physical projection can be roughly reproduced by an ideal mathematical model as seen in Section 2.2, at this stage the estimation of all parameters of the distinct model is within the focus.

Since there exist a vast range of approaches for calibration in literature for not less than a multitude of camera models, this section can merely provide a brief overview of the most important procedures for the *pinhole camera model* and the *omnidirectional camera model*. A more detailed illustration can be found in [79].

2.3.1 Perspective Camera Calibration

The principles of calibrating a perspective camera have been investigated since the mid of the last century. According to [128] known approaches can be categorized into three types, namely *direct nonlinear minimization*, *closed-form solution* and *two-step methods*.

Table 2.3 provides a summary of publications that primarily contributed to the development of calibration algorithms for perspective camera devices.

Table 2.3: Summary of common methods for perspective camera calibration

Year	Author(s)	Remark	Publications
1966	D.C. Brown	Direct Nonlinear Minimization	[16]
1971	Y.I. Abdel-Aziz,	Closed-Form Solution	[2]
	H.M. Karara	Direct Linear Transformation (DLT)	
1975	K.W. Wong	Closed-Form Solution	[129]
1984	S. Ganapathy	Closed-Form Solution	[29]
1987	R.Y. Tsai,	Two-Step Method	[125], [64]
	R.K. Lenz		
1987	O.D. Faugeras,	Closed-Form Solution	[22]
	G. Toscani		
1992	J. Weng	Two-Step Method	[128]
1994	T. Melen		[83]
1996			
1997	J. Heikkila	3D, Circular Features	[39], [38]
2000			
1998	Z. Zhang	2D, Chessboard Pattern	[142]

2.3.2 Omnidirectional Camera Calibration

In order to calibrate an omnidirectional camera device (e.g. a fish-eye lens or a catadioptric camera) in principle the pinhole camera model with an appropriate distortion model can be applied. For example [15] undistorts lens distortions in that way but reports a residuum error of multiple pixels. Furthermore those methods underlie the constraints imposed by the pinhole model. For example the FOV has to be limited to a value much smaller than 180 degrees.

Better results can be expected from new procedures employing generic camera models e.g. those published by J. Kannala and D. Scaramuzza as presented in Section 2.2.4. Kannala reports residuum errors of much less than a tenth of a pixel, while being able to cover FOVs wider than 180 degrees.

Table 2.4 provides a summary of publications that primarily contributed to the development of calibration algorithms for omnidirectional camera devices.

Table 2.4: Summary of common methods for omnidirectional camera calibration

Year	Author(s)	Remark	Publications
2001	C. Braeuer-Burchardt, K. Voss	Simple method for correction of fisheye lens distortion based on pinhole camera model	[15]
2001	F. Devernay, O. Faugeras	‘Straight lines have to be straight’: based on pinhole camera model; optimization of straight lines; may be applied for different calibration targets	[19]
2004-2010	J. Kannala	Generic camera model: applicable for omnidirectional as well as perspective cameras	[55], [56], [57], [54]
2006-2008	D. Scaramuzza	Generic camera model: primarily for catadioptric cameras	[103], [104] [102]

2.4 Two-View Geometry

Imaging is a process that maps 3D world points to 2D image points going hand in hand with an information loss because there exist infinitely many world points that could be the reason for the same image point.

In order to recover a distinct 3D world point from image point information, two or multiple views can be exploited. They can be either static (not moving, multiple isochronal images at different places) or dynamic (moving, multiple successive images at different places). Examples are:

- **Static setup:** Two-view or n-view stereo configuration.
- **Dynamic setup:** Structure from motion.

Both configurations obey the same principles, namely the *epipolar geometry* which is extensively described in [37]. A survey of publications that discuss this subject can be reviewed in Table 2.5.

For simplification, further explanations with respect to *epipolar geometry* in Section 2.4.1 and Section 2.4.2 refer to the ideal pinhole camera model from Section 2.2.2. Subsequently Section 2.4.3 generalizes certain aspects to omnidirectional imaging devices.

Table 2.5: Summary of publications that treat common two-view geometry

Year	Author(s)	Remark	Publications
1987	H. C. Longuet-Higgins	A computer algorithm for reconstructing a scene from two projections	[72]
1993	Faugeras, Oliver D.	Three dimensional computer vision: A geometric viewpoint	[21]
1996	Luong, Quan-Tuan and Faugeras, Olivier D.	The fundamental matrix: Theory, algorithms, and stability analysis	[75]
1997	Gabriella Csurka and Cyril Zeller and Zhengyou Zhang and Olivier D. Faugeras	Characterizing the Uncertainty of the Fundamental Matrix	[18]
1998	Zhang, Zhengyou	Determining the epipolar geometry and its uncertainty: A review	[141]
1999	Gang Xu and Zhengyou Zhang	Epipolar Geometry in Stereo, Motion and Object Recognition	[131]
2000, 2003	Hartley, Richard and Zisserman, Andrew	Multiple view geometry in computer vision	[36, 37]

2.4.1 Epipolar Geometry

We consider two views in arbitrary geometrical configuration as can be seen in Figure 2.9. Both views are individually described by projection matrices \mathbf{P}_l (left) and \mathbf{P}_r (right). However, their mutual geometrical relationship can be described by epipolar geometry. While the structure of the observed scene itself does not influence this geometry, it depends on the camera's intrinsics and relative extrinsics. In Figure 2.9 a 3D world point $\mathbf{X}_{\text{world}}$ projects to the image points $\mathbf{X}_{\text{img},l}$ and $\mathbf{X}_{\text{img},r}$.

Important properties of this configuration are:

- **Baseline:** The line that connects the camera centres C_1 and C_r is called baseline. Figure 2.9 shows this as a red solid line.
- **Epipoles:** The points where the baseline intersects the image planes are called epipoles e_l and e_r .
- **Epipolar lines:** The lines that connect the epipoles of the left and the right view with an image point, denoted as $X_{img,l}$ and $X_{img,r}$ are called epipolar lines $l_{e,l}$ and $l_{e,r}$. One of these is indicated as a green solid line.
- **Epipolar plane:** The plane comprising the baseline and the world point, denoted as $X_{wrl,d}$ is called epipolar plane.

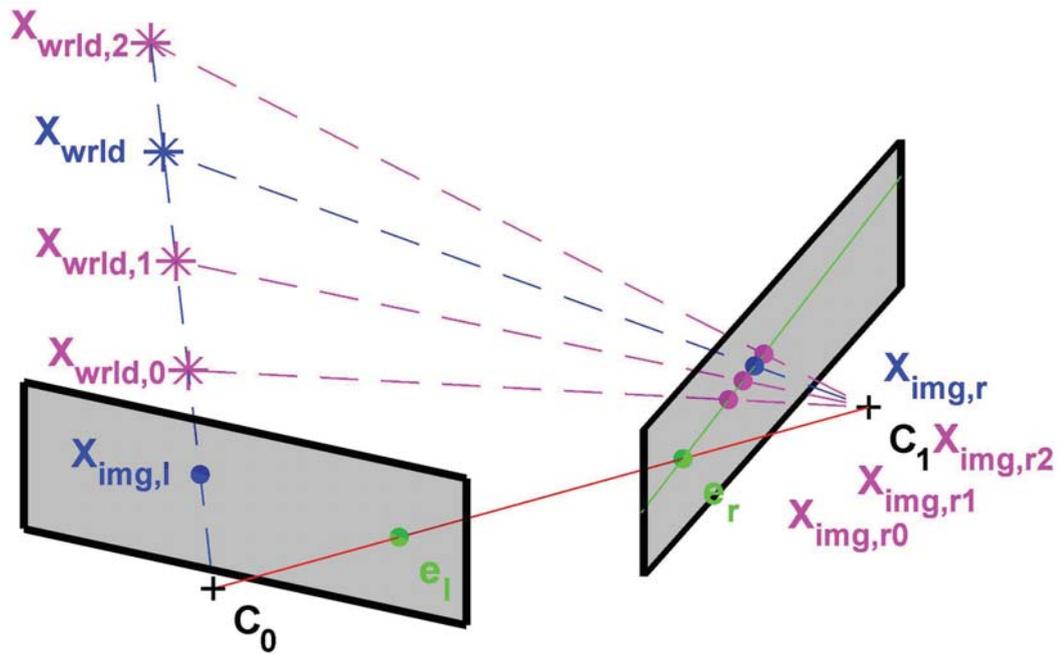


Figure 2.9: Epipolar geometry

In order to recover $X_{wrl,d}$, the image point $X_{img,l}$ can be re-projected by use of Equation 2.27. It results in a ray of 3D points, here exemplary represented by $X_{wrl,d}$, $X_{wrl,0}$, $X_{wrl,1}$ and $X_{wrl,2}$. As can be seen in Figure 2.9 all the projections $X_{img,r}$, $X_{img,r0}$, $X_{img,r1}$ and $X_{img,r2}$ reside on the epipolar line $l_{e,r}$ (green solid line).

As a conclusion, it can be stated that a 3D reconstruction is possible when the corresponding image points are identified. The search can be performed along the epipolar line in order to speed up the process.

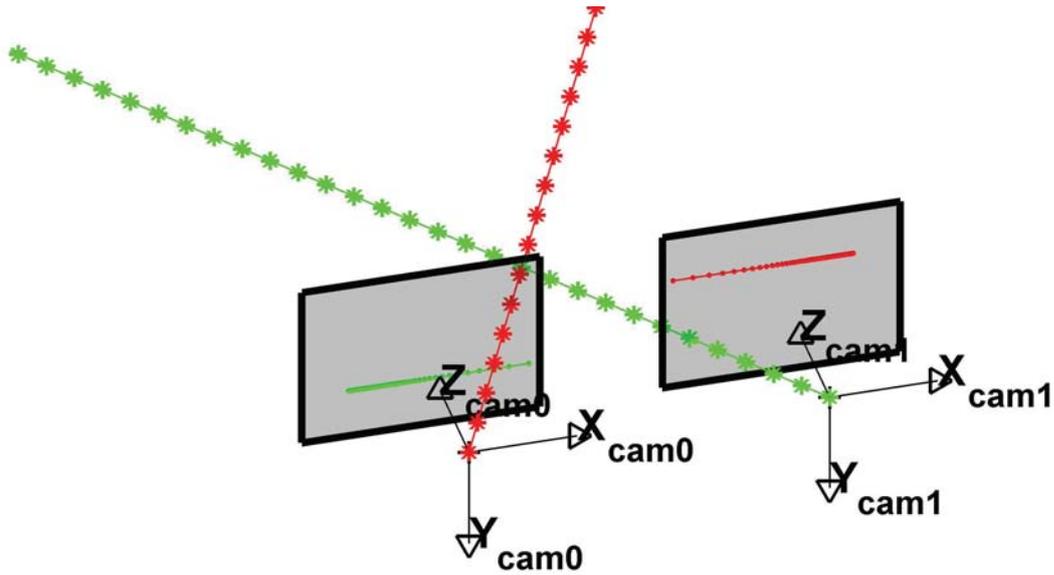


Figure 2.10: Parallel camera configuration

In Figure 2.10 it can be seen that as soon as the orientation of the camera pair and hence their principle axes get parallel, the epipolar lines will be parallel as well. The latter are indicated as red and green parallel lines in the image planes.

Obtaining a parallel camera configuration and hence parallel epipolar lines is very beneficial for the search of image point correspondences as will be seen in Chapter 3.

2.4.2 The Fundamental Matrix

Assume the left camera \mathbf{P}_l to coincide with the world coordinate system and the right one not. According to Equation 2.25, both cameras can be described as follows:

$$\mathbf{P}_l = \mathbf{K}_l \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \quad \mathbf{P}_r = \mathbf{K}_r \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \quad (2.41)$$

With both being *pure pinhole cameras in homogeneous representation* a so-called *fundamental matrix* \mathbf{F} can be computed according to [33] as follows:

$$\mathbf{F} = (\mathbf{K}_r^{-1})^T \cdot [\mathbf{T}_\times] \cdot \mathbf{R} \cdot \mathbf{K}_l \quad (2.42)$$

Hereby $[\mathbf{T}_\times]$ is the *skew-symmetric matrix* that describes the translation vector \mathbf{T} . See Section A.1.

As described in Section 2.4.1, a corresponding image point $\mathbf{X}_{\text{img},r}$ with respect to the image point $\mathbf{X}_{\text{img},l}$ can be found on the epipolar line $\mathbf{l}_{e,r}$. In other words, epipolar geometry provides for every image point in the left view a corresponding line in the right one - and vice versa. This mapping is performed by the *fundamental matrix* \mathbf{F} and is called *epipolar constraint*. A correspondence condition can be formulated as follows:

$$\mathbf{X}_{\text{img},r}^T \mathbf{F} \mathbf{X}_{\text{img},l} = 0 \text{ with } \mathbf{l}_{e,r} = \mathbf{F} \mathbf{X}_{\text{img},l} \quad (2.43)$$

\mathbf{F} is a 3x3 matrix that is not of full rank ($\text{rank}(\mathbf{F}) = 2$) since it maps points to lines. The operation is not invertible. From corresponding views, \mathbf{F} can be estimated with the so-called *8-Point Algorithm* as extensively described in [37].

2.4.3 Epipolar Curves

If we consider a pair of cameras employing the pinhole camera model with distortion or as can be seen in Figure 2.11 applying a completely different camera model, the epipolar constraint no longer produces straight lines. Furthermore it gives rise to *epipolar curves*.

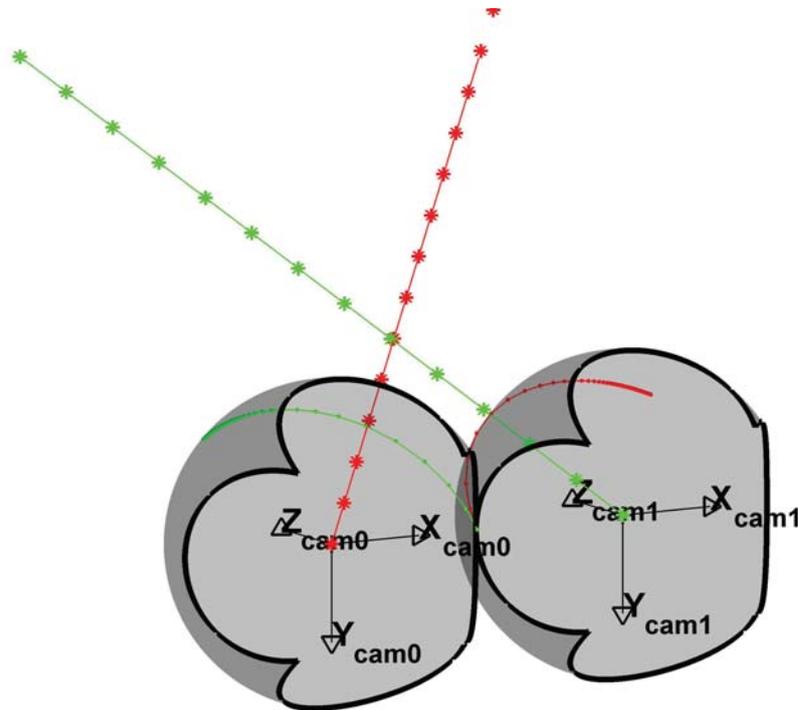


Figure 2.11: Parallel camera configuration

According to [33] the mathematical relationships of Equation 2.42 and Equation 2.43 hence have to be generalized to:

$$\begin{aligned} \text{dist}\left(\mathbf{X}_{\text{img},l}, \overline{\mathbf{P}_l\left(\mathbf{P}_r^{-1}[\mathbf{X}_{\text{img},r}] \setminus z=0\right)}\right) &= 0 \text{ and} \\ \text{dist}\left(\mathbf{X}_{\text{img},r}, \overline{\mathbf{P}_r\left(\mathbf{P}_l^{-1}[\mathbf{X}_{\text{img},l}] \setminus z=0\right)}\right) &= 0 \end{aligned} \tag{2.44}$$

Depending on the distinct camera model, the epipolar curves can be described mathematically. For example if the pinhole camera model is applied with a simple radial distortion only described by one parameter, the resulting curves are cubic. See [139].

If the equiangular model is used as done in Figure 2.11, the resulting curves are so-called *great circles*, which is the intersection curve of the sphere and a plane passing through the cameras principle point. We will later come back to this fact.

Chapter 3

Fundamentals of Stereo Vision

This chapter discusses fundamentals related to the estimation of depth information from two images, so called stereo vision. It starts with a brief overview about the whole computation chain for stereo processing in Section 3.1. Successively the distinct steps for calculating a depth map are outlined consecutively in Section 3.2 (stereo calibration), Section 3.3 (stereo rectification), Section 3.4 (stereo correspondence) and finally Section 3.5 (triangulation). The latter also provides an insight into relevant error estimation methods that go hand in hand with the principle of stereo vision.

Later on the content of Chapter 3 is used and enhanced in order to calculate depth values for the novel omnidirectional stereo vision setup.

3.1 Introduction

The notion "stereo" has its origin in the Greek language: "stereos" can be translated into "solid" and implies that two-dimensional vision is enhanced to the third dimension. Objects can be perceived "*as solid in three spacial dimensions - width, height and depth [...]. It is the added perception of the depth dimension that makes stereo vision so rich and special.*" [1].

3.1.1 The Concept Stereo Vision

Stereo vision is a natural concept to infer depth information from two or more views. Modern machine vision configurations usually apply two imaging devices (e.g. industrial cameras) with an attached computational part in order to copy the concept of human vision. See Figure 3.1. Stereo vision based on two views is called *binocular stereo vision*.



Figure 3.1: Human and machine based stereo vision [1]

Remark: 3D information can also be obtained from a single static camera assumed that the camera is intrinsically calibrated and the geometry of the observed scene or object is known.

3.1.2 Overview of a Stereo Vision Processing Chain

Processing steps that are part of a basic technical stereo vision system can be distinguished into:

- **Offline computation:** The computations are performed once and are not part of the runtime environment like camera- and stereo calibration.
- **Online computation:** The computations take place continuously during runtime like stereo correspondence operations and the process of triangulation. Real-time capability is a crucial issue.

Table 3.1 summarizes the processes that are part of the stereo computation principle.

[77] by Stefano Mattoccia (University of Bologna) gives a comprehensive introduction into all important issues and the latest developments related to stereo vision.

Table 3.1: Overview of a stereo vision system

Process	Illustration	Remark
Camera- and stereo calibration <i>offline</i>		Intrinsic camera parameters and mutual extrinsic parameters are estimated. See Section 3.2.
Rectification <i>offline and online</i>		Input images are transformed in order to rectify the camera setup. See Section 3.3.
Stereo correspondence <i>online</i>		Rectified images are correlated in order to find correspondences and to generate a disparity map. See Section 3.4.
Triangulation <i>online</i>		Based on the disparity map the physical depth can be determined. See Section 3.5.

3.2 Stereo Calibration

Each pair of cameras that observes the same scene can be considered as a *stereo camera system*. Each physical setup of that kind has to be calibrated initially.

While in Section 2.3 the procedure of *calibration* describes the estimation of *intrinsic parameters* of a single camera, *stereo calibration* means finding the mutual *extrinsic* relationship $\hat{\mathbf{H}}_{\text{calib}}$ between two cameras. A straight forward approach for stereo calibration is to estimate the orientation and location for each camera separately with respect to a global reference (e.g. a single calibration target), hence estimating $\hat{\mathbf{H}}_{\text{cam,l}}$ and $\hat{\mathbf{H}}_{\text{cam,r}}$ [33]:

$$\hat{\mathbf{H}}_{\text{calib}} = \hat{\mathbf{H}}_{\text{cam,r}} \cdot \hat{\mathbf{H}}_{\text{cam,l}}^{-1} \quad (3.1)$$

This principle is presented in Figure 3.2. A calibration target represents N world points $\mathbf{X}_{\text{world}}^n$ which can be projected to N point correspondences $\{\mathbf{X}_{\text{norm,l}}^n, \mathbf{X}_{\text{norm,r}}^n\}$ in both views.

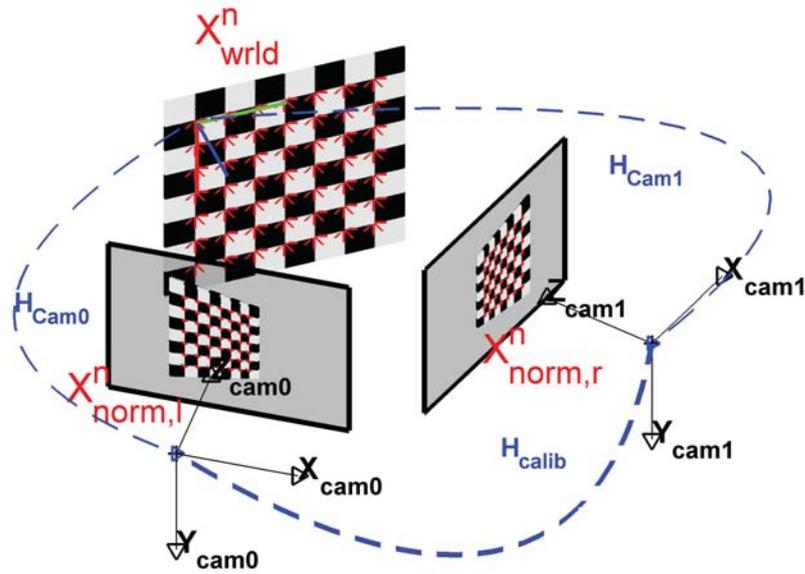


Figure 3.2: The principle of stereo calibration

Hanning [33] states that this basic approach is not accurate enough for stereo vision and must be enhanced by using multiple target positions. Commonly the new global reference then coincides with the left camera coordinate system as already mentioned in Equation 2.41.

According to Hanning [33] calibration approaches for stereo configurations can be categorized as follows:

- Extrinsic stereo camera calibration with generalized epipolar constraints
- Extrinsic stereo camera calibration with respect to the projective error
- Extrinsic and intrinsic stereo camera calibration

Table 3.2 summarizes related work with regards to stereo calibration. In the following a procedure based on minimizing the projective error is considered.

3.2.1 Extrinsic Stereo Calibration With Respect to the Projective Error

If we assume both cameras to be distortion-free (all intrinsic parameters are known), a possible calibration process can be described as follows: A setup as seen in Figure 3.2 is used to generate N point correspondences $\{\mathbf{X}_{\text{norm}_l}^{mn}, \mathbf{X}_{\text{norm}_r}^{mn}\}$ (left and right camera) for each of M arbitrary positioned calibration targets. The key points of the target represent the world points $\mathbf{X}_{\text{world}}$.

Table 3.2: Summary of publications that consider stereo calibration

Year	Author(s)	Remark	Publications
1990	Juyang Weng et al.	Two-step calibration method: (1) parameter estimation based on distortion-free camera model. (2) Non-linear optimization step.	[127]
1996	Fuxing Li et al.	Computes the fundamental matrix with an active moving stereo camera pair with common elevation platform.	[65]
1996	Zhengyou Zhang	Overview paper of different methods, that compute the fundamental matrix.	[138–140]
1999	Richard Hartley	Overview paper of different methods, that calibrate stereo vision systems. Compares strong, weak and non-calibrated geometries.	[35]
2004, 2006, 2011	Tobias Hanning et al.	Treatment of a generalized approach for the epipolar connection between two images and derivation of a special calibration method.	[32–34]
2005	Niklas Pettersson and Lars Petersson	Suggests a method of online calibration for moving camera pairs.	[98]
2006	Chuan Zhou et al.	A stereo calibration method specially for wide-angled lenses is investigated.	[143]
2007	Hyukseong Kwon et al.	A closed-form solution for online calibration for moving (panning, tilting) camera pairs.	[63]

The world points can be projected to both cameras. This can be modelled as follows:

$$\mathbf{X}_{\text{norm}_l}^{\text{mn}} = \mathbf{R}_l \mathbf{X}_{\text{world}}^{\text{mn}} + \mathbf{T}_l \text{ and } \mathbf{X}_{\text{norm}_r}^{\text{mn}} = \mathbf{R}_r \mathbf{X}_{\text{world}}^{\text{mn}} + \mathbf{T}_r \quad (3.2)$$

where \mathbf{R}_l , and \mathbf{R}_r are the unknown rotation matrices for the left and the right camera. \mathbf{T}_l , and \mathbf{T}_r represent unknown translations for both views. Both equations of Equation 3.2 can be combined in order to remove the world point $\mathbf{X}_{\text{world}}^{\text{mn}}$ and establish the relationship between left and right view:

$$\mathbf{X}_{\text{norm}_r}^{\text{mn}} = \mathbf{R}_r \mathbf{R}_l^T \cdot \mathbf{X}_{\text{norm}_l}^{\text{mn}} - \mathbf{R}_r \mathbf{R}_l^T \cdot \mathbf{T}_l + \mathbf{T}_r \quad (3.3)$$

Further simplification leads to:

$$\mathbf{X}_{\text{norm}_r}^{\text{mn}} = \mathbf{R} \cdot \mathbf{X}_{\text{norm}_l}^{\text{mn}} + \mathbf{T} \text{ with } \mathbf{R} = \mathbf{R}_r \mathbf{R}_l^T \text{ and } \mathbf{T} = -\mathbf{R}_r \mathbf{R}_l^T \cdot \mathbf{T}_l + \mathbf{T}_r \quad (3.4)$$

The average values of the computed parameters \mathbf{R} and \mathbf{T} represent the complete extrinsic relationship $\mathbf{H}_{\text{calib}}$ between both cameras. They can be further refined by a non-linear optimization step and form the precondition for the process of *stereo rectification*.

3.3 Stereo Rectification

After calibrating each camera intrinsically ($\mathbf{K}_l, \mathbf{K}_r$) as well as estimating the mutual extrinsic relationship ($\mathbf{H}_{\text{calib}}$), the stereo system needs to be rectified. In Figure 3.3 the *rectification process* is illustrated. Two cameras in general position with the mutual extrinsic relationship of $\mathbf{H}_{\text{calib}}$ need to be rotated to fulfil two requirements:

1. x-axes are collinear and z-axes (optical axes) are parallel.
2. Intrinsic parameters are identical.

The positions of the cameras cannot be changed since they are connected rigidly. The actual computation takes place offline as part of the stereo calibration process.

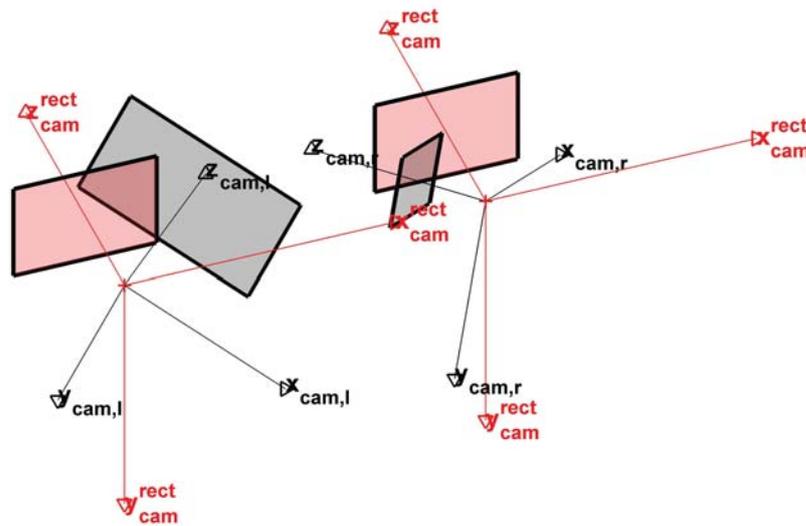


Figure 3.3: The principle of stereo rectification

Since one cannot rotate the cameras physically, a transformation has to be applied to the images itself. This is done by appropriate image transformation mappings.

Table 3.3: Summary of publications that consider stereo rectification

Year	Author(s)	Remark	Publications
1999	Charles Loop and Zhengyou Zhang	A technical solution for rectification of image pairs with known epipolar geometry is proposed.	[73]
2000	Andrea Fusiello et al.	A linear rectification method for general, unconstrained stereo rigs is presented.	[28]
2004	Martin Matousek et al.	A rectification procedure is proposed that minimizes the loss of local image neighbourhood discriminability in rectified images.	[76]
2010	Pascal Monasse et al.	A robust geometric stereo rectification method by a three-step camera rotation method is proposed.	[85]
2011	Huihuang Su and Bingwei He	A two-step method is presented: (1) Setup of two-virtual cameras with parallel optical axes. (2) Projection of the unrectified cameras into the virtual ones.	[120]

In Table 3.3 some major contributions to the range of available algorithmic approaches are mentioned.

Rectified stereo vision setups are called *standard* or *canonical stereo setups*. Since the epipolar lines are parallel (see Section 2.4), the search space for image point correspondences is reduced from two to one dimension(s). From the computational point of view, this leads to a tremendous speed-up. Now it becomes sufficient to search along a single image row for finding pair of points instead of scanning the whole image.

As an example a rectification shall be performed by applying the algorithm of Andrea Fusiello [28].

3.3.1 A Compact Algorithm for Rectification of Stereo Pairs

Andrea Fusiello [28] describes a compact algorithm that is easy to implement and powerful at the same time.

The starting point is a calibrated camera pair described by \mathbf{P}_l and \mathbf{P}_r as seen in Equation 2.41. The target are two new rectified camera matrices, denoted $\mathbf{P}_l^{\text{rect}}$ and $\mathbf{P}_r^{\text{rect}}$.

Computation of the new Rotation Matrices

In order to obtain new rectified rotation matrices $\mathbf{R}_l^{\text{rect}}$ and $\mathbf{R}_r^{\text{rect}}$, new camera frame vectors \mathbf{x}^{rect} , \mathbf{y}^{rect} and \mathbf{z}^{rect} describing the coordinate system of the rectified cameras are computed. They will compose the new matrices $\mathbf{R}_l^{\text{rect}}$ and $\mathbf{R}_r^{\text{rect}}$.

The new x-axis \mathbf{x}^{rect} (see Figure 3.3) points from the origin of the left camera to the origin of the right one:

$$\mathbf{x}^{\text{rect}} = \frac{\mathbf{C}_l - \mathbf{C}_r}{\|\mathbf{C}_l - \mathbf{C}_r\|} = \frac{-\mathbf{C}_r}{\|-\mathbf{C}_r\|} \quad (3.5)$$

The new y-axis \mathbf{y}^{rect} is orthogonal to \mathbf{x}^{rect} and \mathbf{k} .

$$\mathbf{y}^{\text{rect}} = \mathbf{x}^{\text{rect}} \times \mathbf{k} \quad (3.6)$$

where \mathbf{k} is a unit vector that fixes the position of the new y-axis in the plane orthogonal to \mathbf{x}^{rect} . \mathbf{k} can be successively computed from both z-axes as follows:

$$\mathbf{k} = \frac{1}{2}(\mathbf{z}_l + \mathbf{z}_r) \quad (3.7)$$

The algorithm fails when the optical axis is parallel to the base line (e.g. when there is a pure forward motion), but this is not relevant for stereo vision setups.

The new z-axis is orthogonal to both x- and y-axis:

$$\mathbf{z}^{\text{rect}} = \mathbf{x}^{\text{rect}} \times \mathbf{y}^{\text{rect}} \quad (3.8)$$

The new rotation matrix \mathbf{R}^{rect} that must be identical for both cameras in order to obtain the same orientation for both views as can be seen in Figure 3.3 can be composed as follows:

$$\mathbf{R}^{\text{rect}} = \begin{bmatrix} \mathbf{x}^{\text{rect}T} \\ \mathbf{y}^{\text{rect}T} \\ \mathbf{z}^{\text{rect}T} \end{bmatrix} \quad (3.9)$$

Computation of the new Camera Calibration Matrix

In principle the new camera calibration matrix \mathbf{K}^{rect} can be chosen freely, but should meet some requirements:

- The same matrix \mathbf{K}^{rect} must be assigned to both cameras.
- \mathbf{K}^{rect} should be adopted from the camera that provides the largest FOV.

Computation Rectifying Transformations

The computation of the camera rectification establishes new camera matrices $\mathbf{P}_1^{\text{rect}}$ and $\mathbf{P}_r^{\text{rect}}$:

$$\mathbf{P}_1^{\text{rect}} = \mathbf{K}^{\text{rect}} \begin{bmatrix} \mathbf{R}^{\text{rect}} & \mathbf{0} \end{bmatrix} \quad \mathbf{P}_r^{\text{rect}} = \mathbf{K}^{\text{rect}} \begin{bmatrix} \mathbf{R}^{\text{rect}} & \mathbf{T}^{\text{rect}} \end{bmatrix} \quad (3.10)$$

Due to the rigid connection of stereo rigs all camera parameters \mathbf{P} are fixed. Therefore all intrinsic and extrinsic changes have to be done virtually. The computation of *rectifying transformations* enables the generation of new *rectified images* from the *non-rectified images*.

As a result of Section 2.2.2 the new camera matrices can be decomposed as:

$$\mathbf{P}_1^{\text{rect}} = \begin{bmatrix} \mathbf{M}_1^{\text{rect}} & \mathbf{P}_{4C,l} \end{bmatrix} \quad \mathbf{P}_r^{\text{rect}} = \begin{bmatrix} \mathbf{M}_r^{\text{rect}} & \mathbf{P}_{4C,r} \end{bmatrix} \quad (3.11)$$

[28] shows that the rectifying transformation for example $\mathbf{T}_1^{\text{rect}}$ for the left image can be computed as follows:

$$\mathbf{T}_1^{\text{rect}} = \mathbf{M}_1^{\text{rect}} \cdot \mathbf{M}_1^{-1} \quad (3.12)$$

with

$$\mathbf{X}_{\text{img},l}^{\text{rect}} = \mathbf{T}_1^{\text{rect}} \cdot \mathbf{X}_{\text{img},l} \quad (3.13)$$

The next chapter will discuss strategies to resolve non-integer pixel positions as appearing in Equation 3.13 by interpolation.



Figure 3.4: Rectified images with exemplary epipolar lines

Figure 3.4 shows a sample rectified image pair with a number of epipolar lines. It can be seen that those lines are collinear.

3.4 Stereo Correspondence

In order to exploit two images for depth computation, the next step has to find pairs of points that belong to identical scene details. This aspect is known as *stereo correspondence problem* and describes the challenge to find for each pixel $\mathbf{X}_{\text{img},l}$ from the left image a corresponding pixel $\mathbf{X}_{\text{img},r}$ in the right one.

3.4.1 Disparity Computation

After aligning the epipolar lines collinearly, the actual correspondence search for a point $\mathbf{X}_{\text{img},l}$ in the left image can take place on the same row in the right image - cf. Section 2.4.



Figure 3.5: The principle of disparity computation

Figure 3.5 illustrates the correspondence search for a point $\mathbf{X}_{\text{img},l}$ in the left image (intersection of red and blue line). The search takes place in the right image on the red line and leads to the corresponding point $\mathbf{X}_{\text{img},r}$ (intersection of red and magenta line) and hence the disparity d :

$$d = x_{\text{img},l} - x_{\text{img},r} \quad (3.14)$$

- The disparity d is the horizontal displacement of conjugate image points. $[d] = px$

A two-dimensional matrix \mathbf{D} of disparity values d assigned to each pixel $\mathbf{X}_{\text{img},l}$ in the left image is called a *disparity map*. A color coded example is outlined in Figure 3.6.

Black areas represent regions in the left image for which no corresponding points in the right image could be found.

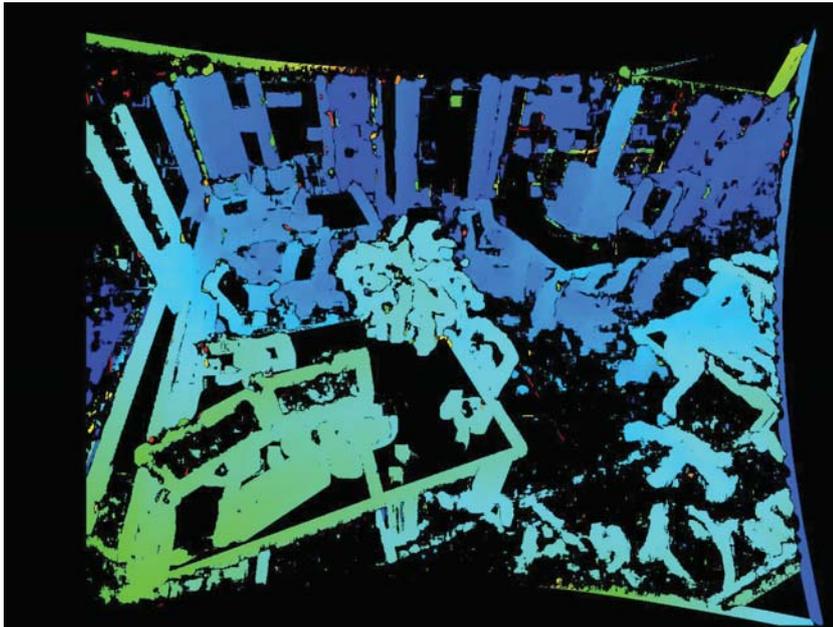


Figure 3.6: Colour coded disparity map

3.4.2 The Correspondence Problem

Finding corresponding pixels is the crucial point for stereo vision and has recently attracted considerable interest amongst many researchers. On the *Middlebury website* [108] modern algorithms are evaluated and ranked competitively. For benchmarking they provide a framework and an appropriate dataset as can be seen exemplary in Table 3.4.

Challenges for Stereo Correspondence Approaches

Algorithms that correlate two images have to face remaining *image distortions* and *pixel noise*, *reflecting*, *transparent* and *specular surfaces*, *perspective distortions* due to different camera positions, *homogeneous* and *ambiguous regions* as well as *repetitive patterns*. Furthermore *occlusions* and *discontinuities* create problems for the stereo methods.

In contrast microcomputer architectures for possible target platforms of embedded solutions limit the amount of available computational power.

A Taxonomy of Stereo Algorithms

Scharstein et al. [107] sketch a taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Generally, these approaches can be distinguished into:

Table 3.4: Evaluation dataset of Middlebury website

Name	Left Image	Right Image	Ground Truth of Left Image	Reference
Cones				[106]
Teddy				[40, 105]
Aloe				[40, 105]

- **Local algorithms:** They perform a so-called Winner Takes it All (WTA) selection strategy for image correspondences. Pixel value ambiguities are reduced by aggregating matching costs over a support window (also known as kernel or correlation window). This increases the Signal to Noise Ratio (SNR). [77].
- **(Semi-) global algorithms:** These algorithms search for disparity assignments that minimize an energy function over the whole stereo pair using a pixel-based matching cost (sometimes the matching cost is aggregated over a support). [77].

According to Scharstein et al. [107], common stereo algorithms consist of the following steps: (1.) *matching cost computation*, (2.) *cost aggregation*, (3.) *disparity computation and optimization*, (4.) *disparity refinement*.

One of the most commonly used algorithms are the local (kernel window based) methods. Using the standard Sum of Squared Differences (SSD) algorithm, the following steps are performed:

1. **Matching cost computation:** The matching cost is the squared difference of intensity values at a given disparity.
2. **Cost aggregation:** Aggregation is done by summing matching cost over square windows with constant disparity.

3. **Disparity computation and optimization:** Disparities are computed by selecting the minimal (winning) aggregated value at each pixel.
4. **Disparity refinement:** A refinement can be performed e.g. by *subpixel estimation*.

Since stereo correspondence is not a focus of this thesis, it shall be considered from now on as an abstract operation \mathcal{D} that generates a disparity map D from the image coordinates $\mathbf{X}_{\text{img},l}$ and $\mathbf{X}_{\text{img},r}$ respectively.

$$D = \mathcal{D}(\mathbf{X}_{\text{img},l}, \mathbf{X}_{\text{img},r}) \quad (3.15)$$

3.5 Triangulation

Applying the systems parameter of a stereo vision setup, a physical depth Z can be computed from a disparity value d by similar triangles. This process of *stereo triangulation* bases on camera geometry with parallel optical axes.

The range a stereo camera can measure and its accuracy is constrained by the physical parameters of a stereo camera as well.

3.5.1 Depth Measurement

In Figure 3.7 a two view setup in canonical configuration is outlined.

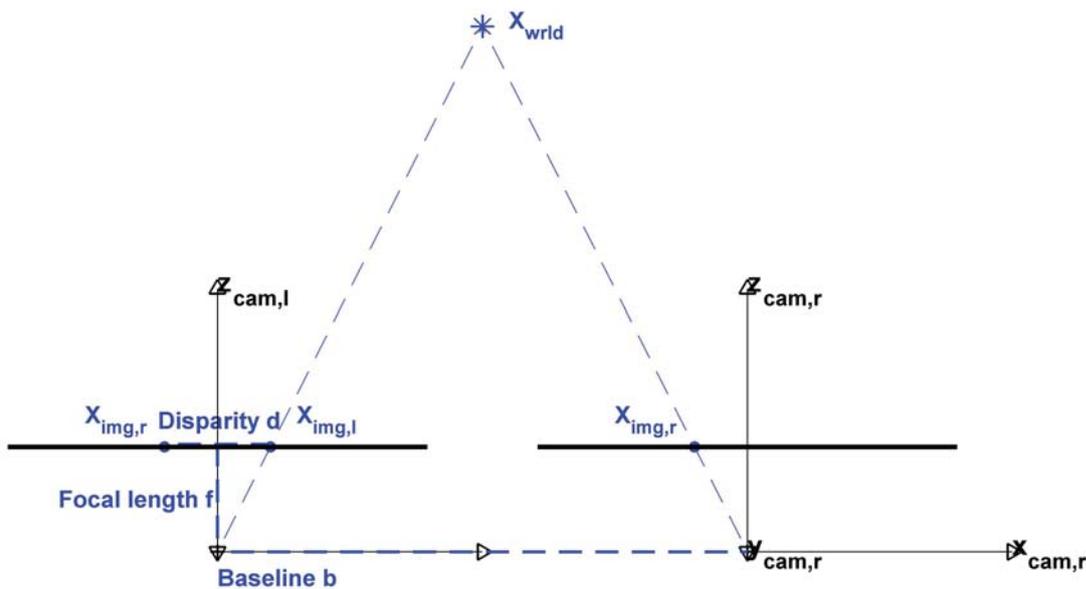


Figure 3.7: Triangulation

According to Chang et al. [17] a depth value Z can be estimated from Equation 3.16.

$$Z = b \cdot \frac{f}{ad} \quad (3.16)$$

where

- a is the physical pixel width and hence the image sampling interval. It is identical with p_x (see Section 2.2.1.2). $[a] = mm$.

By using the intrinsic parameters of the camera calibration matrix \mathbf{K} , the equation can be re-expressed as follows:

$$Z = b \cdot \frac{\alpha_x}{x_{\text{img},l} - x_{\text{img},r}} = b \cdot \frac{\alpha_x}{d} \quad (3.17)$$

Each depth value Z represents the z -component of a 3D point given in camera coordinates.

3.5.2 Range Field of Measurement

According to Mattoccia [77] the range of operation $[Z_{\text{max}}, Z_{\text{min}}]$, so-called *horopter*, is constrained by the disparity range $[d_{\text{min}}, d_{\text{max}}]$.

- d_{min} : The minimal measurable disparity is 0.5. This concludes that the maximum distance that can be measured directly depends on the spacial resolution of the working camera. $[d_{\text{min}}] = px$.
- d_{max} : The maximal measurable disparity depends on the parametrization. This has to be balanced between a required closest area that should be measurable and the desired observable horizontal field of view FOV_h . The latter is directly reduced by the maximum disparity. This effect can be observed in Figure 3.6 as a black (blind) left margin. $[d_{\text{max}}] = px$.

3.5.3 Measurement Accuracy

Measurement accuracy arises from the limited image resolution given by a . The accuracy of depth measurements of a stereo vision system is denoted as *depth resolvability* Z^* and can be calculated as follows:

$$Z^* = \frac{bf}{a^*}, \quad a^* = \frac{a}{2} \quad (3.18)$$

where

- Z^* is the *depth resolvability*. $[Z^*] = mm$.

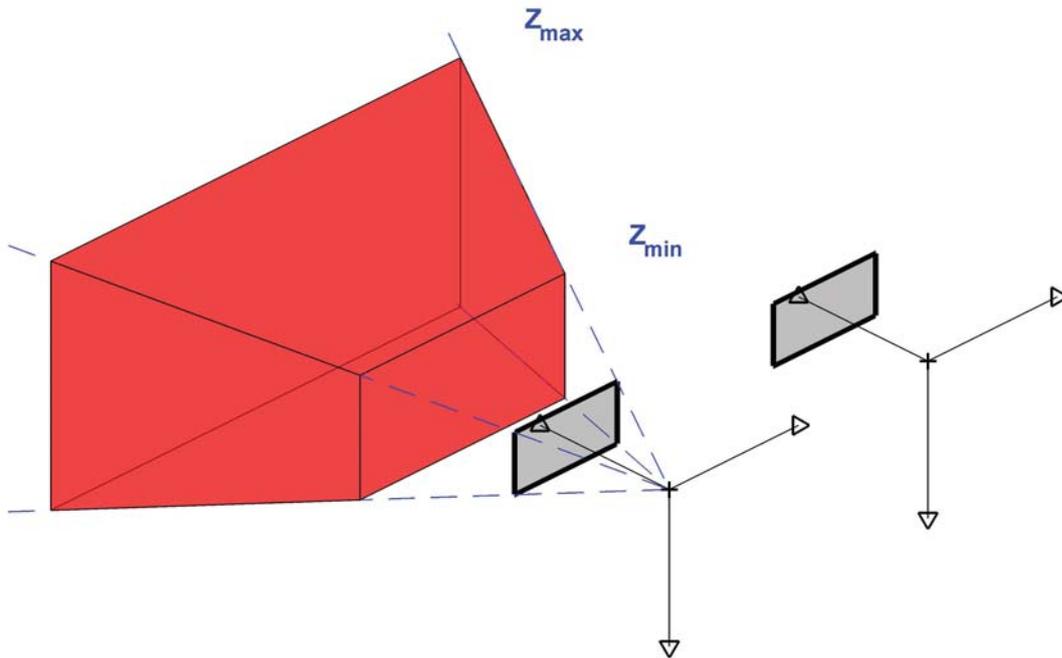


Figure 3.8: Horopter

- a^* is the *image resolution* with $a^* = \frac{a}{2}$. $[a^*] = mm$.

According to Chang [17] the depth resolvability "[...] represents the maximum depth resolvable by the given stereo imaging system."

If the disparities (to measure) are less than a^* , and hence $Z > Z^*$, the origin points have to be considered as points with Z at infinity:

$$\{a^* \rightarrow 0\} \Rightarrow \{Z^* \rightarrow \infty\} \quad (3.19)$$

3.5.4 Measurement Errors

Following Chang [17] the reasons of erroneous measurements in stereo vision are due to a vast amount of factors:

- **Angular errors:** They result from the fact that the image plane is not exactly perpendicular to the camera axis.
- **Rotation error:** There is a certain rotational offset between camera and image coordinate system.
- **Position error:** The imager is not sufficiently flat.

Additional errors are focal length error, quantization error and mismatching errors, to name a few more. For further procedure it is assumed that a depth estimation error only arises due to spacial quantization effects in the image plane. Following the explanations of [27] and [4] an analysis of errors for stereo vision can be regarded as an *geometrical issue*, an *statistical issue* as well as from the viewpoint of *geometrical qualities*. Section 3.5.4.1 and Section 3.5.2 review the error computation from the aspects of geometry and statistics in more detail.

3.5.4.1 Quantization Error

A depth measurement error can be specified as *absolute depth error* δZ as well as *percentage depth error* $\delta Z/Z$:

$$\delta Z = \delta Z_b + \delta Z_f + \delta Z_d = \frac{f}{d}\delta b + \frac{b}{d}\delta f - \frac{bf}{d^2}\delta d \quad (3.20)$$

and

$$\frac{\delta Z}{Z} = \frac{\delta b}{b} + \frac{\delta f}{f} - \frac{\delta ad}{ad} \quad (3.21)$$

where

- δZ_b denotes the partial derivative of Z with respect to b . $[\delta Z_b] = mm$.
- δZ_f denotes the partial derivative of Z with respect to f . $[\delta Z_f] = mm$.
- δZ_d denotes the partial derivative of Z with respect to d . $[\delta Z_d] = mm$.

As a consequence of Equation 3.20 and Equation 3.21, the major contributions to δZ are drawn by the *baseline inaccuracy* δb , *focal length error* δf and *disparity error* δd . According to Chang [17] the contributions δb and δf to the overall error value can be neglected for accurate devices. Hence the afore-mentioned error quantities can be simplified to:

$$\delta Z \approx -\frac{Z^2\delta d}{bf} \quad (3.22)$$

and

$$\frac{\|\delta Z\|}{Z} \approx \frac{Z\|\delta d\|}{bf} \quad (3.23)$$

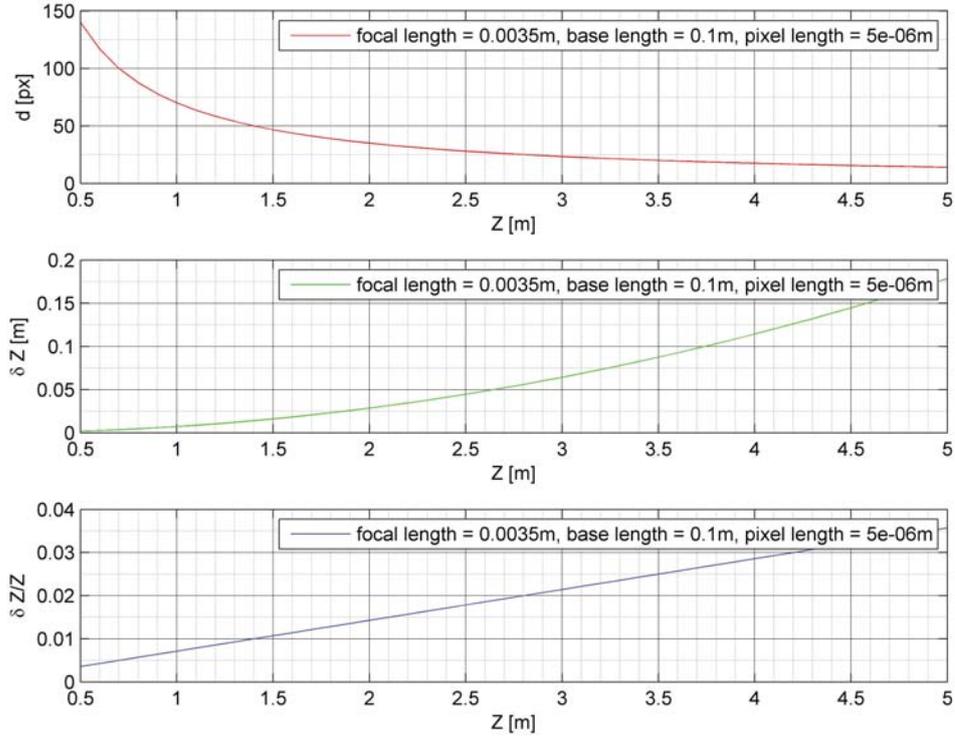


Figure 3.9: Example quantization error

Figure 3.9 shows an exemplary quantization error plot for an artificial stereo vision setup with the parameters $f = 3.5\text{mm}$, $b = 10\text{cm}$ and $p_x = 5\mu\text{m}$.

From Equation 3.22 and Figure 3.9 it can be concluded that the depth error is quadratically dependent on Z and proportional to δd . Furthermore from Equation 3.23 and Figure 3.9 it can be concluded that the percentage depth error depends linearly on Z .

A deviation of δZ can be reviewed in Section A.2.

3.5.4.2 Statistical Distribution of Quantization Errors

A further way to examine error occurrences is to describe a probability distribution of the range error and the expected value of range error magnitude.

Following Chang [17] a probability density function p for the likelihood of δZ in a stereo vision geometry can be stated as follows:

$$p(\delta Z) = \begin{cases} \frac{Z^*}{2Z_{\max}Z_{\min}} - \frac{(Z^*)^2(Z_{\min}^2 + Z_{\min}Z_{\max} + Z_{\max}^2)}{12(Z_{\min}Z_{\max})^3} \cdot |\delta Z| & \text{if } 0 \leq |\delta Z| \leq \frac{2Z_{\min}^2}{Z^*} \\ \frac{1}{Z_{\max} - Z_{\min}} \left[\frac{2}{3} \sqrt{\frac{Z^*}{2|\delta Z|}} + \frac{(Z^*)^2}{12Z_{\max}^3} \cdot |\delta Z| - \frac{Z^*}{2Z_{\max}} \right] & \text{if } \frac{2Z_{\min}^2}{Z^*} \leq |\delta Z| \leq \frac{2Z_{\max}^2}{Z^*} \end{cases} \quad (3.24)$$

A deviation of $p(\delta Z)$ can be reviewed in Section A.3. There it can be studied that the depth Z to measure can be considered as a random variable between a minimal depth Z_{\min} and a maximum depth Z_{\max} .

The information of the probabilistic distribution $p(\delta Z)$ can be further conditioned to get a statement about the expected magnitude of the absolute depth error $E(|\delta Z|)$ which is defined as:

$$E(|\delta Z|) = \int_{-\infty}^{+\infty} f(Z)E(|\delta Z||Z) dZ \quad (3.25)$$

Rodriguez et al. [100] derive the expected absolute depth error in magnitude for the case of uniformly distributed depth over the interval $[Z_{\min}, Z_{\max}]$ to be:

$$E(|\delta Z|) = \frac{2}{9Z^*} (Z_{\max}^2 + Z_{\min}Z_{\max} + Z_{\min}^2) \quad (3.26)$$

The complete derivation from Equation 3.25 to Equation 3.26 can be reviewed in [100].

For an exemplary stereo vision setup with $f = 1.0mm$, $b = 10cm$ and $p_x = 5\mu m$ the graphic Figure 3.10 demonstrates the normalized probability density function $p_n(\delta Z)$ and the expected absolute depth error $E(|\delta Z|)$ at a target depth $Z = 2.5m$, with:

$$p_n(\delta Z) = \frac{p(\delta Z)}{\max(p(\delta Z))} \quad (3.27)$$

Figure 3.11 presents a selection of the left cameras pixels re-projected to their appropriate camera rays. The probability distribution of the measured depth Z has been illustrated as color coded plot. It is visible that the measurement of euclidean distances of camera points \mathbf{X}_{cam} for a constant target depth Z is most accurate at the (perspective) cameras center.

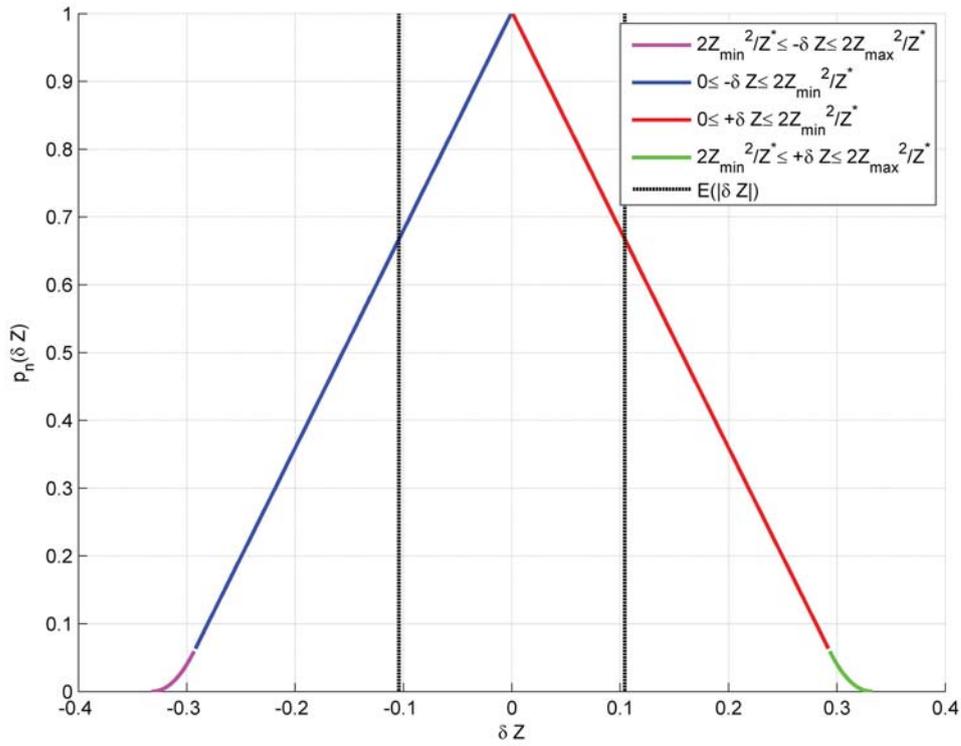


Figure 3.10: Sample probability distribution of δZ at $Z = 2.5m$ (2D)

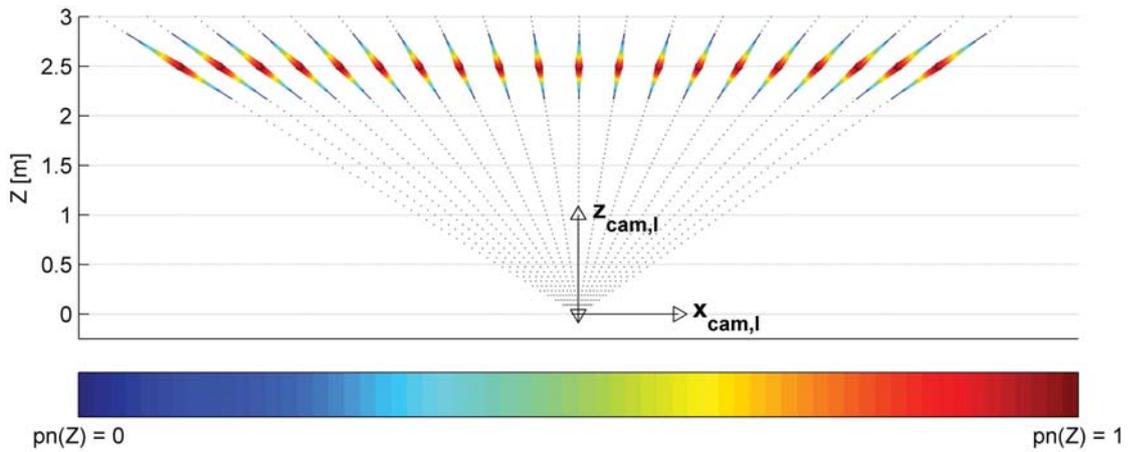


Figure 3.11: Sample probability distribution of δZ at $Z = 2.5m$ (3D)

Chapter 4

Virtual Cameras

Chapter 2 presented a selection of available camera models, e.g. the perspective and the omnidirectional projection method. Depending on the application it may be necessary to transfer image data from one to another camera model. As these cameras are not necessarily real cameras, they are called *virtual cameras*. Related scientific work on that principle is summarized in Section 4.1.

The most prominent method of projection model transition is called Virtual Pan-Tilt-Zoom (vPTZ) camera and describes the transfer from a physical omnidirectional projection to a virtual perspective camera that can be configured freely in terms of rotation and intrinsic parameters. The theory behind this principle is examined in Section 4.2, whereby a new technique is presented in order to speed up this process. Section 4.3 investigates the mapping error that occurs while Section 4.4 proposes a novel method for calculating appropriate pixel sizes for the virtual cameras. Section 4.5 outlines performance measurements in order to compare the different vPTZ algorithms with respect to their computational efficiency. Finally Section 4.6 concludes this chapter by evaluating the practical relevance for the vPTZ technology with respect to machine vision based people recognition.

Later on the content of Chapter 4 is revisited in Chapter 6 in order to parametrize the omnidirectional stereo vision setup and in Chapter 8 where a vPTZ is employed in an advanced system for human behaviour analysis.

The results of the investigations presented in Chapter 4 have been published by the author in [25], [81] and [82].

4.1 Introduction and Related Works

The generation of perspective views from an omnidirectional image has already been picked up in previous scientific work.

Peri et al. [97] developed a software system which is able to generate multiple perspective and panoramic views from an omnidirectional image. As an input source of the omnidirectional image serves a catadioptric camera system consisting of a conventional camera and a parabolic mirror. Virtual cameras, placed and parametrized by the user, generate the different views. The pixel intensity on the imaging surface of such a virtual camera is determined by reprojection on the parabolic mirror. To process in real-time, a geometric map is generated for each view only once. As long as the viewing parameters of the virtual camera are not changed, the reprojection can be sped up by a Look-Up Table (LUT). However, the omnidirectional projection of the catadioptric camera is only estimated with a relatively simple parabolic equation.

Onoe et al. [93] improved the algorithm with the aim of a faster perspective image generation. They divided the destination image into equal parts using a rectangular grid. Each grid point is reprojected in the omnidirectional image, whereas four grid points always form a quadrilateral. Each quadrilateral in the omnidirectional image is warped to the corresponding position in the perspective image applying a bi-linear interpolation. In this way the quality of the perspective image is reduced in favour of a faster calculation. Similar to Peri et al., they used a catadioptric camera and model the image formation by applying an equation for the hyperboloidal surface of the mirror.

Zhang et al. [136] used a camera with a fisheye lens. They modelled the lens with a sphere on whose surface the omnidirectional image is projected. The computational expensive rectification algorithm was implemented on a Field Programmable Gate Array (FPGA) for the purpose of real-time processing.

The approaches presented above share a common disadvantage. They lack a precise and generic camera model to describe various off-the-shelf cameras - cf. Section 2.2.4.

An mentionable application for the generation of a panoramic image from an omnidirectional image was presented by Ardouin et al. [5]. There, a catadioptric camera system is mounted on a helmet. A captured omnidirectional image is transformed into a panoramic image. The result is displayed on an HMD. So a person wearing that helmet experiences a 360° surround view.

Table 4.1 provides an overview of relevant scientific contributions to this topic.

The following sections detail the afore-mentioned technology.

Table 4.1: Summary of scientific contributions to methods of camera model conversion

Year	Author(s)	Remark	Publications
1997	Peri et al.	Generation of perspective and panoramic views from a catadioptric camera system	[97]
1998	Onoe et al.	Generation of perspective views from a catadioptric camera system by applying an LUT	[93]
2002	Peter Sturm	Analysis of relations between multiple views of different camera types (paracatadioptric, perspective or affine cameras)	[119]
2003	Tarak Gandhi and Mohan M. Trivedi	Parametric ego-motion compensation with an omnidirectional camera	[30]
2003	Chiharu Ishii et al.	Consideration of omni-directional vision using an equidistant projection fish eye lens; a partial domain of an original image in fish eye image is selected, and this is converted to a perspective image	[45]
2004, 2007, 2008	Sheng-Wen Jeng and Wen-Hsiang Tsai	Construction of perspective and panoramic images from omnidirectional images taken from hypercatadioptric cameras	[47–49]
2007	Simon Thibault	Surveillance system featuring a panoramic panomorph lens for event detection, recognition and identification over a 360-degree area with 100 percent coverage	[124]
2008	Zhang et al.	Rectification of a fish-eye lens based image by modelling the projection with a sphere on a FPGA architecture	[136]
2012	Ardouin et al.	Generation of a panoramic image from an omnidirectional image for a Head Mounted Display (HMD)	[5]
2013	Michel Findeisen et al.	A fast approach for omni- to perspective image conversion and error analysis	[25]

4.2 Omni to Perspective Vision

In order to transform an omnidirectional image into one or multiple perspective images, several approaches are available as presented in Section 4.1.

The method introduced now treats the perspective images as the result of n virtually distributed perspective cameras $\text{Cam}_0^{\text{per}}, \dots, \text{Cam}_i^{\text{per}}, \dots, \text{Cam}_n^{\text{per}}$ that are completely described by their extrinsic parameters $\mathbf{C}_i^{\text{per}}, \mathbf{R}_i^{\text{per}}$ and intrinsic parameters $\mathbf{K}_i^{\text{per}}$ as discussed in Section 2.2.

The omnidirectional camera is described by its extrinsic parameters $\mathbf{C}^{\text{omni}}, \mathbf{R}^{\text{omni}}$ and its intrinsic parameters \mathbf{K}^{omni} , which are fixed and provided by the physical mount and construction of the fisheye camera. These parameters are determined during the camera calibration procedure. Furthermore the projection behaviour of the omnidirectional camera is described by its projection model as seen in Section 2.2.

For the virtual perspective views the extrinsic calibration of the omnidirectional camera forms the world reference. Hence the parameters $\mathbf{C}_i^{\text{per}}$ and $\mathbf{R}_i^{\text{per}}$ refer to \mathbf{C}^{omni} and \mathbf{R}^{omni} . Since the location of every $\text{Cam}_i^{\text{per}}$ is identical with \mathbf{C}^{omni} , $\mathbf{C}_i^{\text{per}} = \mathbf{0}$ (3 DOF are fixed), while its rotation $\mathbf{R}_i^{\text{per}}$ can be chosen freely with reference to \mathbf{R}^{omni} (3 DOF are variable). Furthermore, $\mathbf{K}_i^{\text{per}}$ can be chosen freely.

Figure 4.1 shows a sample setup comprising an omnidirectional camera (with the image from Figure 2.8 mapped onto a sphere - cf. Section 2.2.3) and a single perspective camera (with a plane sensor surface - cf. Section 2.2.2).

After having set up the omnidirectional camera as well as at least one virtual perspective camera, perspective images can be generated using one of two possible mapping strategies: Forward Mapping or Backward Mapping. Both strategies are explained in Section 4.2.1 and Section 4.2.2. For the latter an enhanced process is presented in Section 4.2.3 denoted as Fast Backward Mapping.

4.2.1 Forward Mapping

The principle of *forward mapping* shapes the algorithm in a way that the computation sequence starts at the omnidirectional image source and ends at the perspective image target. The mapping algorithm can be summarized as outlined in Figure 4.2.

The remainder of this subsection treats each step in more detail.

Back-Projection (Source): The color information as seen in Figure 4.1 relates to a grid of two-dimensional pixel coordinates $\mathbf{X}_{\text{img}}^{\text{omni}}$. The full grid $\mathbf{X}_{\text{img}}^{\text{omni}}$ is back-projected to their appropriate 3D rays, described as Cartesian expression $\mathbf{X}_{\text{cam}}^{\text{omni}}$ with respect to the omnidirectional camera coordinate system.

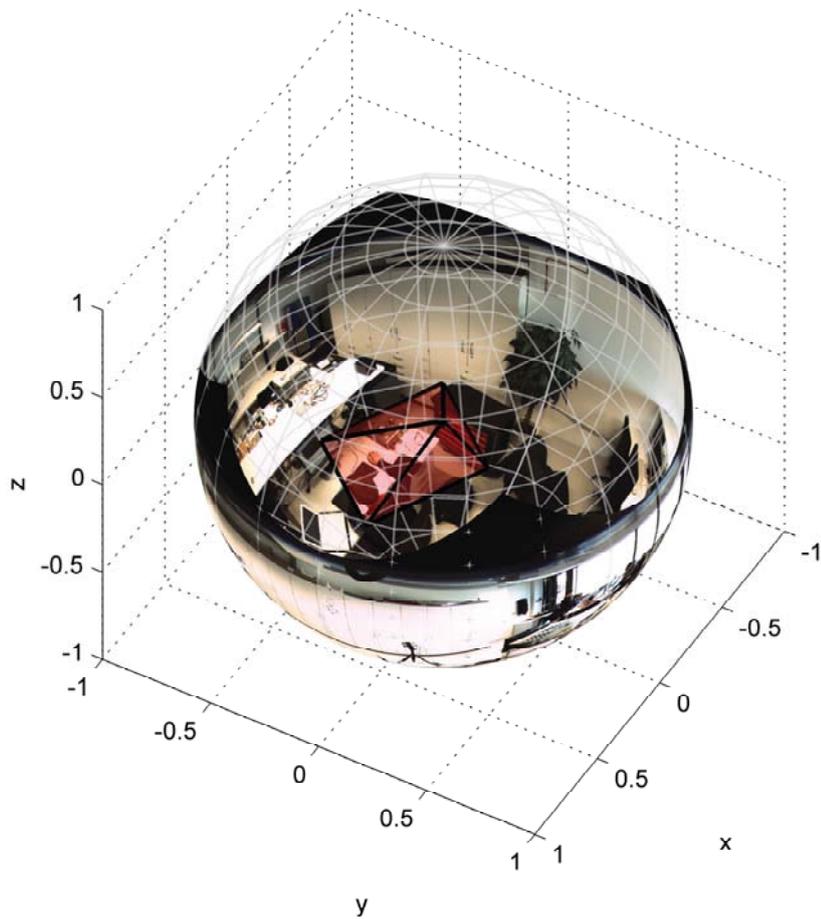


Figure 4.1: A virtual perspective camera

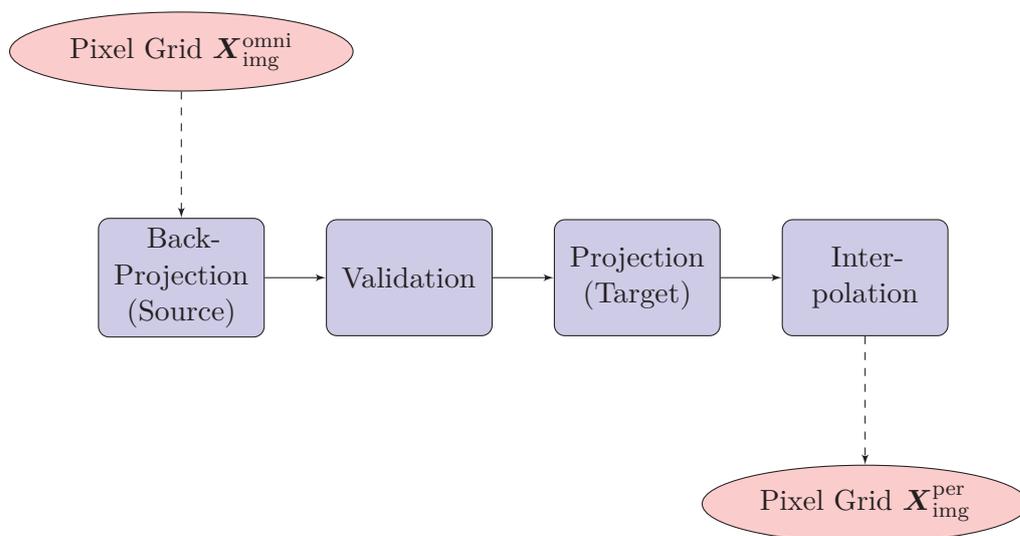


Figure 4.2: Summary of the forward transformation algorithm

Therefore $\mathbf{X}_{\text{img}}^{\text{omni}}$ is transferred to its normalized sensor coordinates $\mathbf{X}_{\text{norm}}^{\text{omni}}$ by employing Equation 2.15 inversely:

$$\mathbf{X}_{\text{norm}}^{\text{omni}} = \mathbf{K}^{-1} \cdot \mathbf{X}_{\text{img}}^{\text{omni}} \quad (4.1)$$

In the successive stage, $\mathbf{X}_{\text{norm}}^{\text{omni}}$ is represented by appropriate polar coordinates - cf. Equation 2.6 and Equation 2.7:

$$\mathbf{X}_{\text{norm}}^{\text{omni}} = \begin{pmatrix} x_{\text{norm}}^{\text{omni}} \\ y_{\text{norm}}^{\text{omni}} \\ 1 \end{pmatrix} \hat{=} \begin{pmatrix} \rho_{\text{norm}}^{\text{omni}} \\ \varphi_{\text{norm}}^{\text{omni}} \\ 1 \end{pmatrix} \quad (4.2)$$

Using Equation 2.8 and Equation 2.29, hence employing generic back-projection, the appropriate camera points can be set up by applying the following relationship:

$$\mathbf{X}_{\text{cam}}^{\text{omni}} = \begin{pmatrix} x_{\text{cam}}^{\text{omni}} \\ y_{\text{cam}}^{\text{omni}} \\ z_{\text{cam}}^{\text{omni}} \\ w_{\text{cam}}^{\text{omni}} \end{pmatrix} \hat{=} \begin{pmatrix} \theta_{\text{cam}}^{\text{omni}} \\ \varphi_{\text{cam}}^{\text{omni}} \\ \rho_{\text{cam}}^{\text{omni}} \end{pmatrix} = \begin{pmatrix} \mathcal{K}^{-1}(\rho_{\text{norm}}^{\text{omni}}) \\ \varphi_{\text{norm}}^{\text{omni}} \\ 1 \end{pmatrix} \quad (4.3)$$

where $\mathbf{X}_{\text{cam}}^{\text{omni}}$ is expressed as spherical coordinates $(\theta_{\text{cam}}^{\text{omni}}, \varphi_{\text{cam}}^{\text{omni}}, \rho_{\text{cam}}^{\text{omni}})^T$ with $\rho_{\text{cam}}^{\text{omni}}$ being the distance of the 3D point from the origin \mathbf{C}^{omni} . To overcome the uncertainty of the distance between the point in camera coordinates and the camera center, choose $\rho_{\text{cam}}^{\text{omni}} = \|\mathbf{X}_{\text{cam}}^{\text{omni}}\| \stackrel{!}{=} 1$ to normalize the points to a unity sphere.

Validation: The following stage treats the imaging process of $\mathbf{X}_{\text{cam}}^{\text{omni}}$ into the virtual perspective camera Cam^{per} . An omnidirectional camera covers a considerably larger FOV than a perspective one. Back-projected image points $\mathbf{X}_{\text{cam}}^{\text{omni}}$ might partially have negative z^{per} -components which, if projected, lead to non-valid image points. Those points have to be neglected. Referring to Figure 2.6, valid 3D points have to reside rightwards of \mathbf{p}_{3R} (principle plane). Hence those points comply with the following equation:

$$\mathbf{p}_{3R} \cdot \mathbf{X}_{\text{cam}}^{\text{omni}} \geq 0 \quad (4.4)$$

Projection (Target): Now $\mathbf{X}_{\text{cam}}^{\text{omni}}$ can be projected into the perspective camera model by applying Equation 2.24:

$$\hat{\mathbf{X}}_{\text{img}}^{\text{per}} = \mathbf{P} \cdot \mathbf{X}_{\text{cam}}^{\text{omni}} \quad (4.5)$$

This can be done since $\mathbf{X}_{\text{cam}}^{\text{omni}} \hat{=} \mathbf{X}_{\text{world}}^{\text{per}}$. Performing Equation 4.5 leads to real-number target image pixel positions $\hat{\mathbf{X}}_{\text{img}}^{\text{per}}$. To distinguish between integer and real-number coordinates, the latter is indicated by $\hat{\mathbf{X}}$.

Interpolation: In order to obtain pixel intensities I^{per} of the integer coordinates $\mathbf{X}_{\text{img}}^{\text{per}} = (x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}, 1)^T$ from the real-number pixel positions $\hat{\mathbf{X}}_{\text{img}}^{\text{per}}$ and its appropriate pixel intensities I^{omni} a standard image interpolation procedure (e.g. bilinear or bi-cubic) is employed.

The forward procedure has significant disadvantages. Regardless of the specification of the virtual target camera, the complete set of source image coordinates is transformed despite not all coordinates being within the visible range of the target image. Furthermore, if applying an accurate generic camera model as presented in Section 2.2.4, the mathematical relationship has to be solved inversely by finding the roots of Kannala's polynomial. This leads to a tremendous lack of performance. Another important aspect is the fact that the use of low order interpolation filters, like bilinear or bi-cubic filter masks, can produce gaps in the target images if the source pixel positions map onto the target pixel positions in a sparse distribution.

An exemplary perspective result can be reviewed in Figure 4.3.

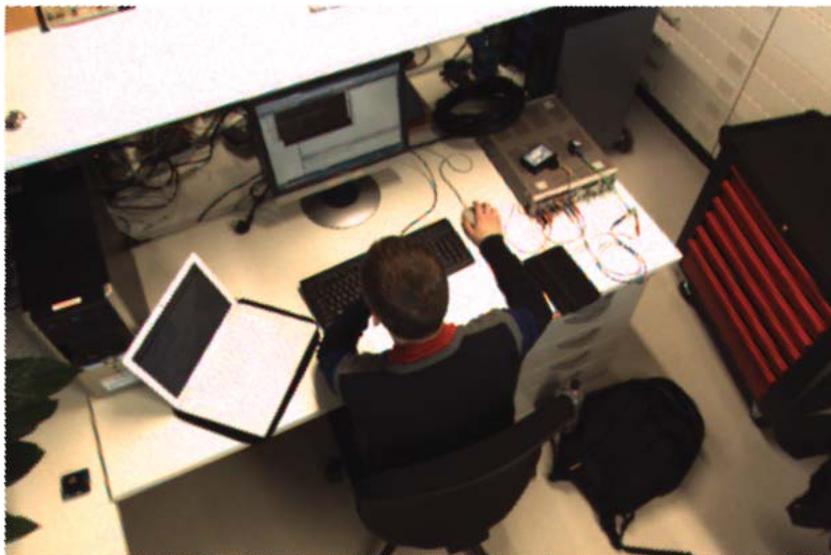


Figure 4.3: Result forward projection

4.2.2 Backward Mapping

An alternative approach can overcome the disadvantages mentioned in the previous section while offering further possibilities to improve performance. *Backward mapping* inverses the problem - from target to source. The mapping algorithm can be summarized as outlined in Figure 4.4.

The remainder of this subsection treats each step in detail.

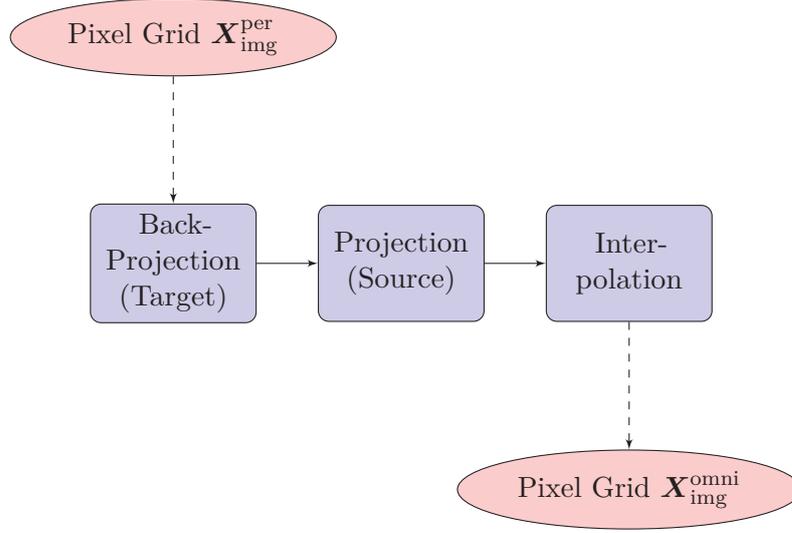


Figure 4.4: Summary of the backward transformation algorithm

Back-Projection (Target): Equivalently to Section 4.2.1, the target image mesh of pixel positions $\mathbf{X}_{\text{img}}^{\text{per}}$ is set up and back-projected to its appropriate world points $\mathbf{X}_{\text{wrlld}}^{\text{per}}$ using Equation 2.24 inversely:

$$\mathbf{X}_{\text{wrlld}}^{\text{per}}(\lambda) = \lambda \cdot \mathbf{P}^+ \cdot \mathbf{X}_{\text{img}}^{\text{per}} \quad (4.6)$$

where \mathbf{P}^+ forms the pseudo-inverse of \mathbf{P} and λ is an arbitrary scale factor that expresses the one dimensional space of points on the ray. To overcome the uncertainty of distance from the camera center, choose $\|\mathbf{X}_{\text{wrlld}}^{\text{per}}\| \stackrel{!}{=} 1$. See Figure 4.5.

Projection (Source): Project all the normalized world points $\mathbf{X}_{\text{wrlld}}^{\text{per}}$ into the omnidirectional camera model using Equations 2.6, 2.15 and 2.29 to obtain real-number source image pixel positions $\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$. See Figure 4.6.

Interpolation: In order to obtain pixel intensities I^{per} of the integer coordinates $\mathbf{X}_{\text{img}}^{\text{per}} = (x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}, 1)^T$ from the real-number pixel positions $\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ and its appropriate pixel intensities I^{omni} a standard image interpolation procedure (e.g. bilinear or bi-cubic) is employed.

An exemplary perspective result can be reviewed in Figure 4.7.

Lookup-Table Operation: The nonlinear transformation between $\mathbf{X}_{\text{img}}^{\text{per}}$ and $\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ can be performed using an LUT for faster video processing. Two LUTs \mathbf{map}_x and \mathbf{map}_y describe the transformation for x-coordinates and y-coordinates from $\mathbf{X}_{\text{img}}^{\text{per}}$ to $\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ separately. An exemplary LUT \mathbf{map}_x for the mapping of x-coordinates in a target image of 640×480 is shown in Figure 4.8.

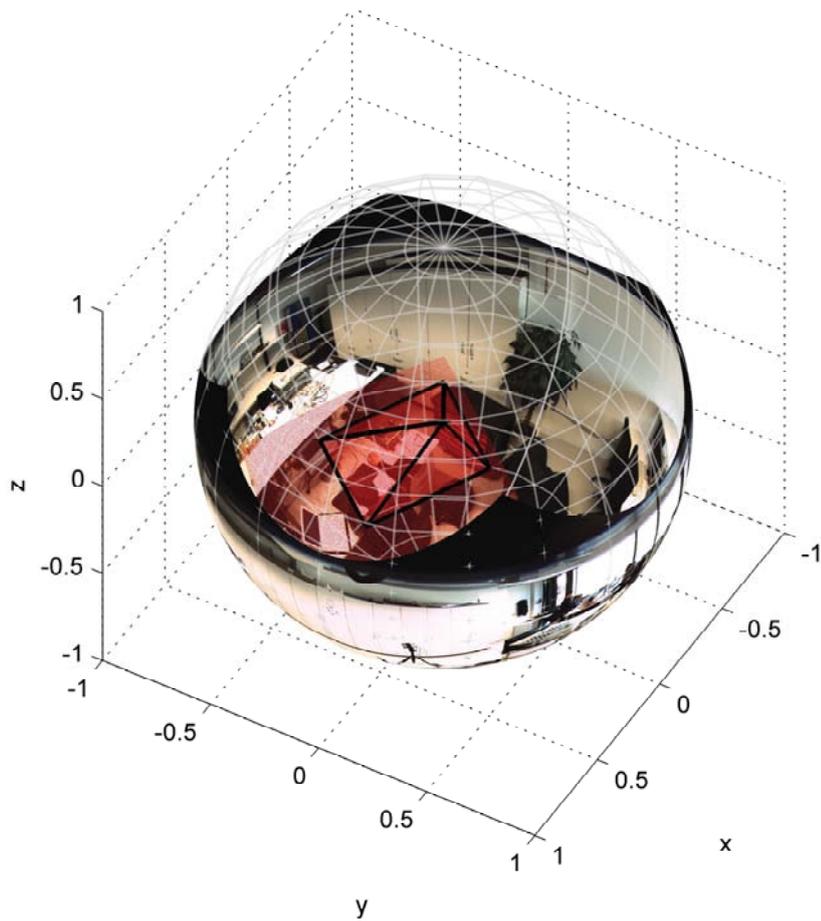


Figure 4.5: Reprojection of 2D target image coordinates to 3D camera coordinates onto unit sphere



Figure 4.6: Projection of 3D camera coordinates to 2D source image coordinates



Figure 4.7: Result backward projection

4.2.3 Fast Backward Mapping

In order to accelerate the process of LUT computation, a change of the scale space can be employed by using a *Gaussian pyramid* [46] and sub-sampling the space grid that forms the LUT. The mapping algorithm can be summarized as outlined in Figure 4.9.

Since the LUT (as seen in Figure 4.8) predominantly comprises low order frequency components, a compression should not lead to significant quality loss as will be seen in Section 4.3. To obey the rules of the sampling theorem, it is not sufficient just to

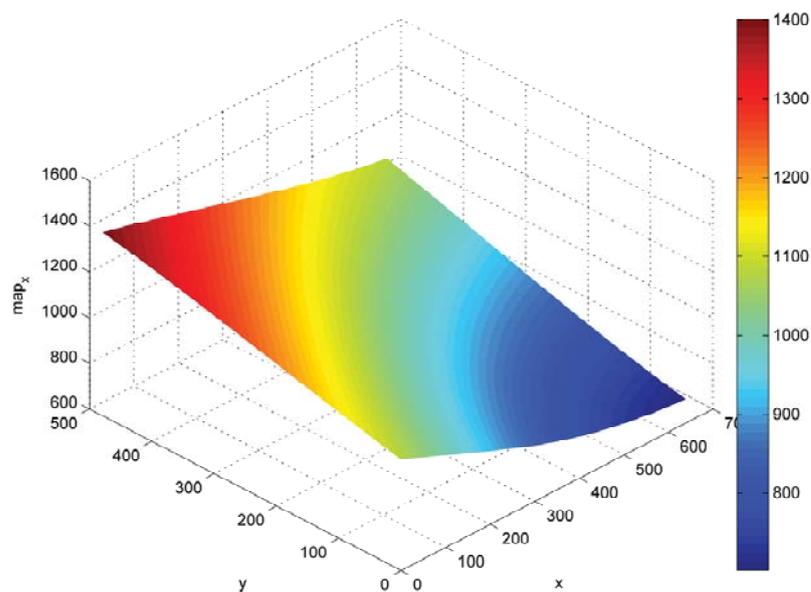


Figure 4.8: Exemplary lookup-table for the backward mapping of target pixel x-coordinates

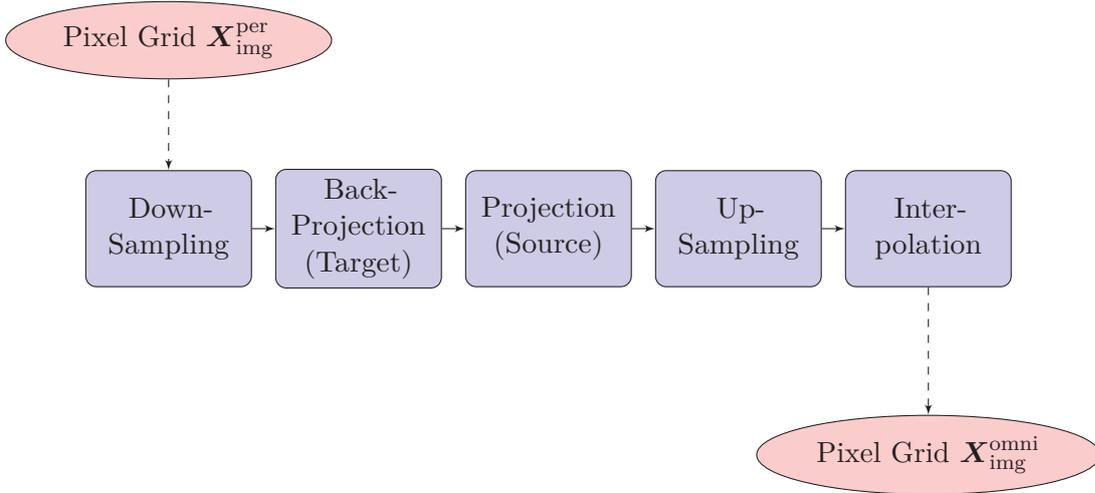


Figure 4.9: Summary of the fast backward transformation algorithm

select every n^{th} coordinate of the grid. This would result in an aliased pattern and unsatisfactory results respectively. That is why all structures that are sampled less than four times per wavelength are suppressed by a smoothing filter. The remainder of this subsection treats each step in detail.

Down-Sampling: Equivalently to Section 4.2.2, the target image mesh of pixel positions X_{img}^{per} is set up. X_{img}^{per} is down-sampled by building up a Gaussian pyramid according to [46]. The reduced mesh is denoted as $X_{img}^{per\downarrow}$.

For instance, a target image of size 640×480 can be scaled down by 2^4 (in each direction) leading to a mesh size of 40×30 .

Back-Projection (Target): The algorithm for back-projection in Section 4.2.2 is used in order to transfer $X_{img}^{per\downarrow}$ to $X_{wld}^{per\downarrow}$. See Figure 4.10.

Projection (Source): The algorithm for projection in Section 4.2.2 is used in order to transfer $X_{wld}^{per\downarrow}$ to $\hat{X}_{img}^{omni\downarrow}$. $\hat{X}_{img}^{omni\downarrow}$ denotes the reduced real-number image coordinates on the omnidirectional image. See Figure 4.11.

Up-Sampling: The reduced LUT $\hat{X}_{img}^{omni\downarrow}$ is up-sampled according to the down-sampling process in order to obtain the full real-number source image pixel mesh \hat{X}_{img}^{omni} .

Since the up-sampling process introduces a large error in the margins of the LUT because of zero-padding, it is appropriate to initially enlarge the original mesh X_{img}^{per} before down-sampling. The necessary size of this initial padding depends on the used smoothing filter as can be seen in Section 4.3.

Interpolation: Get the image intensity information at the positions \hat{X}_{img}^{omni} from X_{img}^{omni} by interpolation as done in Section 4.2.2.

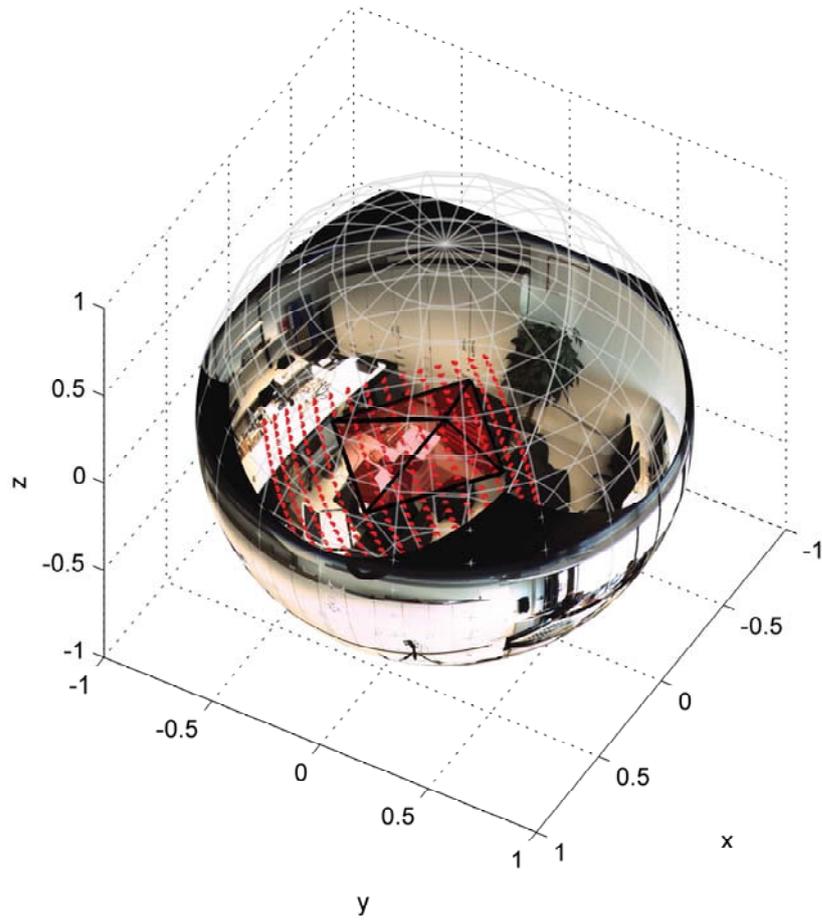


Figure 4.10: Reprojection of sub-sampled 2D target image coordinates to 3D camera coordinates onto unit sphere



Figure 4.11: Projection of sub-sampled 3D camera coordinates to 2D source image coordinates

Lookup-Table Operation: According to Figure 4.8 an exemplary reduced LUT \mathbf{map}_{xr} connecting the x-coordinates of $\mathbf{X}_{img}^{per\downarrow}$ and $\hat{\mathbf{X}}_{img}^{omni\downarrow}$ can be seen in Figure 4.12.

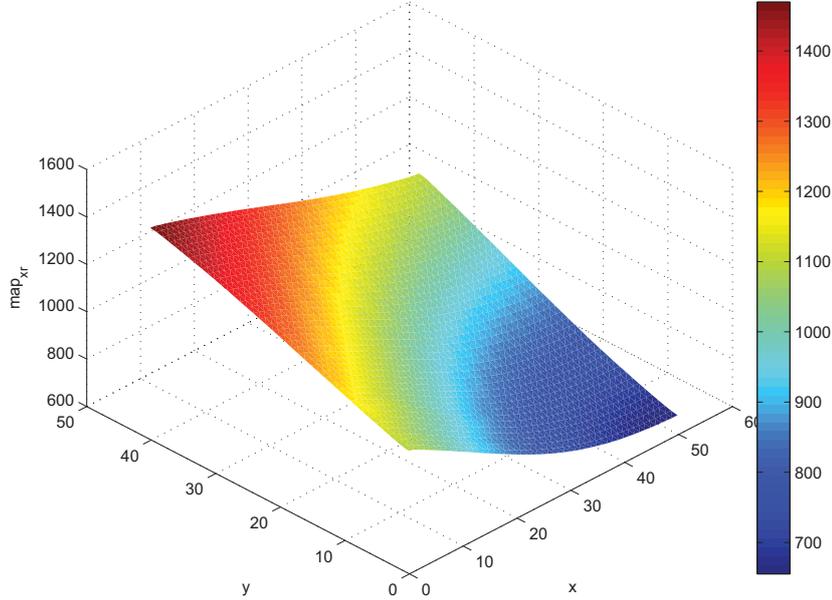


Figure 4.12: Exemplary reduced lookup-table for the fast backward mapping of target pixel x-coordinates

Having up-sampled the data again, an exemplary interpolated LUT connecting the x-coordinates of \mathbf{X}_{img}^{per} and $\hat{\mathbf{X}}_{img}^{omni}$ can be called \mathbf{map}_{xi} .

4.3 Error Analysis

Since the fast backward mapping approach leads to a loss of information, a trade-off in accuracy has to be accepted to the benefit of a performance improvement. This section considers the de facto arising error.

Let n be the power of reduction and $p \times q$ the target image size. In order to avoid large errors at the target image borders as mentioned in Section 4.2, one has to initially pad the original LUTs appropriately. So a convolution of \mathbf{map}_x and \mathbf{map}_y with the 5×5 Gaussian kernel

$$\frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (4.7)$$

requires a necessary padding of 2^n additional pixels on each side. This convolution is done to smooth the LUTs. In this way we can avoid alias during up-sampling. An error is introduced during the down- and up-sampling of the LUTs, caused by the dump of frequency components. So it is increasing with the power of reduction n .

This error can be expressed as a residuum e_x and e_y in terms of pixel displacements using the following equations:

$$e_x = \mathbf{map}_x - \mathbf{map}_{xi} \quad (4.8)$$

$$e_y = \mathbf{map}_y - \mathbf{map}_{yi} \quad (4.9)$$

An example e_x for reducing the LUT by the power of 4 can be seen in Figure 4.13.

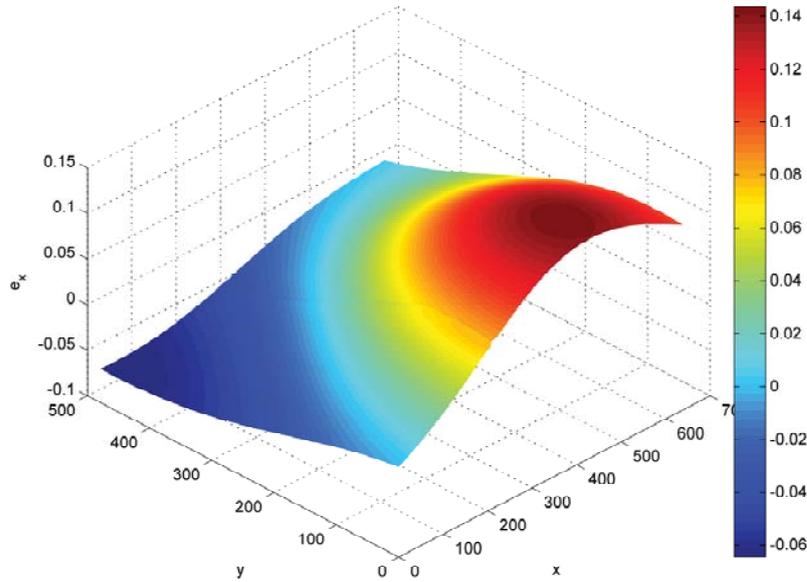


Figure 4.13: Residuum error for the fast mapping of pixel x-coordinates

The total error e_{xy} can be calculated by Equation 4.10.

$$e_{xy} = \frac{1}{pq} \sum_{i,j} (e_{x_{i,j}}^2 + e_{y_{i,j}}^2). \quad (4.10)$$

Figure 4.14 shows the error e_{xy} depending on the elevation of the pinhole camera for different powers of reduction.

Because of symmetrical reasons the error does not depend on the azimuth of the pinhole camera.

As it can be seen, up to a reduction power of $n = 4$ no significant error can be measured due to the dominating low order frequency components of the non-linear pixel transformation. For further reduction ($n > 4$) a noticeable pixel displacement trade-off has to be accepted. Further investigations can be performed in order to determine the distinct mathematical relationship between a given acceptable pixel displacement error and the achievable maximum reduction size.

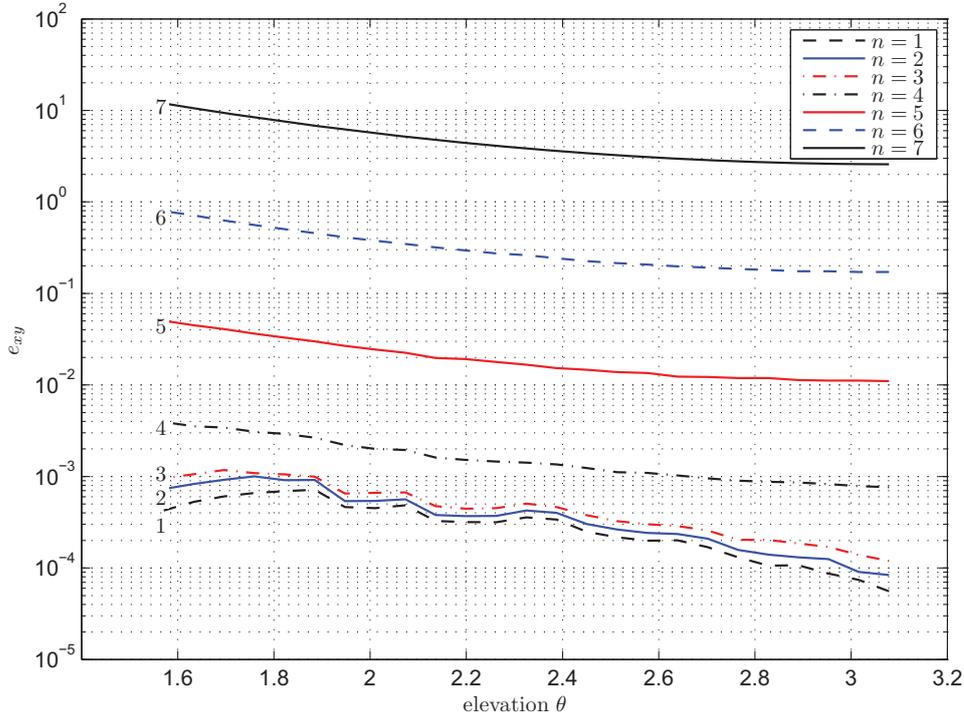


Figure 4.14: The total pixel error depending on elevation from equator to south pole for each power of reduction

4.4 Accuracy Analysis

We have seen so far that a conversion from one to another projection model is feasible in an efficient way. The question that shall be answered now is how one can parametrize the resolution of the virtual target camera based on the physical source camera. In other words: how accurate can the target camera be?

The basis for this accuracy discussion for virtual cameras is provided by the intrinsic parameters of the source as well as the target camera. The parameters α_x and α_y (of both cameras) define the physical pixel density (scaled by f) on the (physical and virtual) imaging die in x-direction, and y-direction respectively - cf. Section 2.2.1.2. These values define the physical pixel width p_x and p_y , while each pixel covers a certain solid angle - depending on the underlying camera model. Every physical step p_x and p_y represents a pixel step $\Delta x_{\text{img}} = 1$ and $\Delta y_{\text{img}} = 1$ in the image. These parameters represent the physical sampling interval of the image(er).

4.4.1 Intrinsic of the Source Camera

Our test system consists of a Fujinon FE185C086HA-1 fisheye lens with a focal length of $f = 2.7\text{mm}$ and a field of view of 185° mounted to a Basler acA2040-25gc industrial camera containing a 1-inch CMOSIS CMV4000 sensor with a (cropped) resolution of 1680×1680 pixels. This camera has been calibrated by using the generic camera model of Kannala [56] which has been summarized in Section 2.2.4. Hereby the radial symmetric model applies four polynomial coefficients as outlined in Equation 2.29.

The estimated intrinsic parameters can be found in Table 4.2, while successively computed values are presented in Table 4.3.

Table 4.2: Results of the intrinsic generic camera models estimation

Camera	k_1	k_2	α_x	α_y	c_x	c_y	k_3	k_4	k_5
Cam ^{omni}	2.40	-0.01	200.25	200.58	838.85	851.67	0.08	-0.05	0.01

Table 4.3: Physical pixel dimensions

Camera	p_x	p_y
Cam ^{omni}	$1.35 \times 10^{-5} [\text{m}]$	$1.35 \times 10^{-5} [\text{m}]$

For further distinction of source and target parameters, the pixel densities and pixel width values of the source camera are now declared as α_x^{omni} and α_y^{omni} as well as p_x^{omni} and p_y^{omni} .

4.4.2 Intrinsic of the Target Camera

In theory, the intrinsic parameters of the virtual (perspective) camera Cam^{per} can be chosen freely. A suitable starting point is to define an observable FOV in horizontal and vertical direction: $\text{FOV}_h[\text{rad}]$ and $\text{FOV}_v[\text{rad}]$. For a distortion-free perspective view this is determined as follows:

$$\begin{aligned} \text{FOV}_h &= 2 \cdot \arctan\left(\frac{\text{width}}{2 \cdot \alpha_x}\right) = 2 \cdot \arctan\left(\frac{\text{width} \cdot p_x}{2 \cdot f}\right) \\ \text{FOV}_v &= 2 \cdot \arctan\left(\frac{\text{height}}{2 \cdot \alpha_y}\right) = 2 \cdot \arctan\left(\frac{\text{height} \cdot p_y}{2 \cdot f}\right) \end{aligned} \quad (4.11)$$

From Equation 4.11 it becomes obvious that the FOV depends on a bunch of certain intrinsic values:

- If the resolution of the image {width, height} doubles, the size of each virtual pixel $\{p_x, p_y\}$ has to be divided by two for the same FOV.

- If the resolution of the image {width, height} doubles, the normalized focal length $\{\alpha_x, \alpha_y\}$ has to double as well for the same FOV.

One way to overcome this uncertainty is to manually define the resolution {width, height} and successively calculate the normalized focal length $\{\alpha_x, \alpha_y\}$. The focal length f of the virtual camera can be set to the value of the physical one, so the virtual pixel size can be computed successively. The crucial issue is now: What is an appropriate resolution for the target camera? It is quite natural to expect a larger resolution to be more useful. Since Cam^{per} directly depends on the parameters of Cam^{omni} , there should be a certain sensible limit.

For further distinction of source and target parameters, the pixel densities and pixel dimensions of the target camera are now declared as α_x^{per} and α_y^{per} as well as p_x^{per} and p_y^{per} . For further explanation, the following values are exemplarily set for the virtual perspective camera:

Table 4.4: Exemplary FOV for the target camera

Camera	FOV _h [rad]	FOV _v [rad]
Cam ^{per}	$\pi/2$	$\pi/2$

In Figure 4.15 the sample area of the target image in the source image is illustrated.

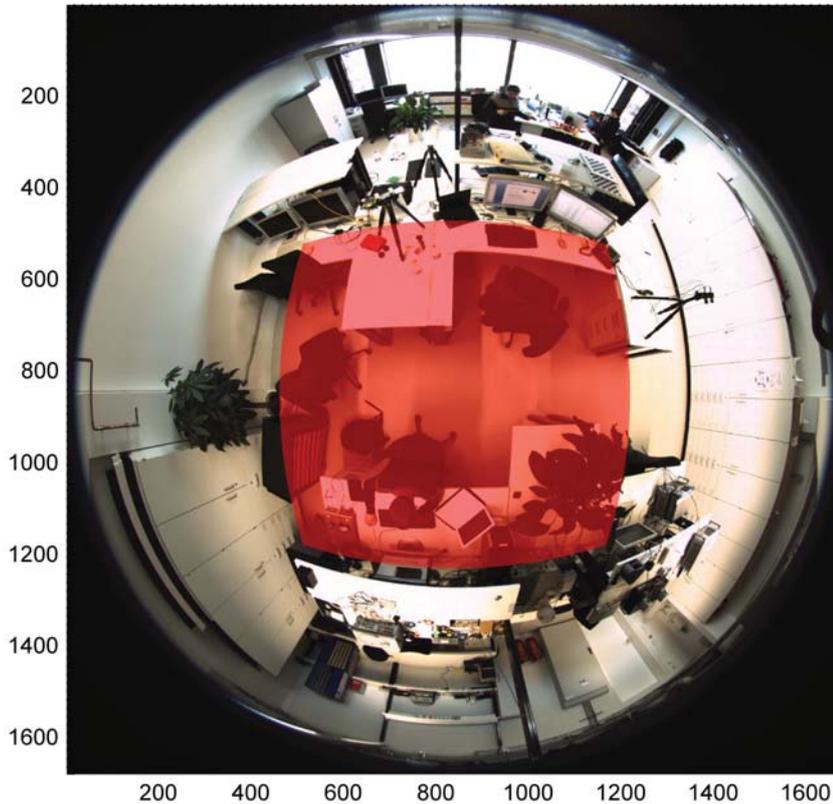


Figure 4.15: Sample area of the target image in the source image

4.4.3 Marginal Virtual Pixel Size

From Section 4.4.1 and Section 4.4.2 it becomes obvious that there must be a relationship between the virtual and physical pixel size:

$$\begin{aligned} p_x^{\text{per}} &\leftarrow (p_x^{\text{omni}}, p_y^{\text{omni}}) \\ p_y^{\text{per}} &\leftarrow (p_x^{\text{omni}}, p_y^{\text{omni}}) \end{aligned} \quad (4.12)$$

When applying the backward mapping method for virtual image generation a certain source image position $\hat{\mathbf{X}}_{\text{img}}^{\text{omni}} = (x_{\text{img}}^{\text{omni}}, y_{\text{img}}^{\text{omni}}, 1)$ is calculated from a given target image position $\mathbf{X}_{\text{img}}^{\text{per}} = (x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}, 1)$. The computed horizontal image position $x_{\text{img}}^{\text{omni}}$ and vertical image position $y_{\text{img}}^{\text{omni}}$ individually depend on the dedicated target pixel position $x_{\text{img}}^{\text{per}}$ and $y_{\text{img}}^{\text{per}}$. The functions that establish this relationship are parametrized by the pixel sizes p_x^{omni} , p_y^{omni} , p_x^{per} and p_y^{per} :

$$\begin{aligned} x_{\text{img}}^{\text{omni}} &= \mathcal{F}_x(x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}; p_x^{\text{omni}}, p_y^{\text{omni}}, p_x^{\text{per}}, p_y^{\text{per}}) \\ y_{\text{img}}^{\text{omni}} &= \mathcal{F}_y(x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}; p_x^{\text{omni}}, p_y^{\text{omni}}, p_x^{\text{per}}, p_y^{\text{per}}) \end{aligned} \quad (4.13)$$

with

- \mathcal{F}_x is a function that transforms target pixel positions into a source pixel column.
- \mathcal{F}_y is a function that transforms target pixel positions into a source pixel row.

Remark: Equation 4.13 is further parametrized by the complete set of intrinsic and extrinsic parameters for the source as well as the target camera. Since they do not contribute to the comprehension of the current problem, they are neglected in this discussion. However, for the actual calculation they are involved.

In order to investigate how the target image samples the source image, the gradient of both functions can be applied:

$$\begin{aligned} \nabla \mathcal{F}_x &= \frac{\delta \mathcal{F}_x}{\delta x_{\text{img}}^{\text{per}}} \cdot \begin{pmatrix} \Delta x_{\text{img}}^{\text{per}} \\ 0 \end{pmatrix} + \frac{\delta \mathcal{F}_x}{\delta y_{\text{img}}^{\text{per}}} \cdot \begin{pmatrix} 0 \\ \Delta y_{\text{img}}^{\text{per}} \end{pmatrix} \\ \nabla \mathcal{F}_y &= \frac{\delta \mathcal{F}_y}{\delta x_{\text{img}}^{\text{per}}} \cdot \begin{pmatrix} \Delta x_{\text{img}}^{\text{per}} \\ 0 \end{pmatrix} + \frac{\delta \mathcal{F}_y}{\delta y_{\text{img}}^{\text{per}}} \cdot \begin{pmatrix} 0 \\ \Delta y_{\text{img}}^{\text{per}} \end{pmatrix} \end{aligned} \quad (4.14)$$

Since we are interested in the magnitude of sampling intervals ($|\Delta x_{\text{img}}^{\text{omni}}|, |\Delta y_{\text{img}}^{\text{omni}}|$) and having in mind that the virtual image is sampled with unit steps ($|\Delta x_{\text{img}}^{\text{per}}| = 1, |\Delta y_{\text{img}}^{\text{per}}| = 1$)

Equation 4.14 simplifies to:

$$\begin{aligned} |\Delta x_{\text{img}}^{\text{omni}}| &\hat{=} |\nabla \mathcal{F}_x| = \left\| \left(\frac{\delta \mathcal{F}_x}{\delta x_{\text{img}}^{\text{per}}}, \frac{\delta \mathcal{F}_x}{\delta y_{\text{img}}^{\text{per}}} \right) \right\| \\ |\Delta y_{\text{img}}^{\text{omni}}| &\hat{=} |\nabla \mathcal{F}_y| = \left\| \left(\frac{\delta \mathcal{F}_y}{\delta x_{\text{img}}^{\text{per}}}, \frac{\delta \mathcal{F}_y}{\delta y_{\text{img}}^{\text{per}}} \right) \right\| \end{aligned} \quad (4.15)$$

The relationships of Equation 4.15 still depend on the variables and parameters according to Equation 4.13:

$$\begin{aligned} |\Delta x_{\text{img}}^{\text{omni}}| &= |\nabla \mathcal{F}_x| \left(x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}; p_x^{\text{omni}}, p_y^{\text{omni}}, p_x^{\text{per}}, p_y^{\text{per}} \right) \\ |\Delta y_{\text{img}}^{\text{omni}}| &= |\nabla \mathcal{F}_y| \left(x_{\text{img}}^{\text{per}}, y_{\text{img}}^{\text{per}}; p_x^{\text{omni}}, p_y^{\text{omni}}, p_x^{\text{per}}, p_y^{\text{per}} \right) \end{aligned} \quad (4.16)$$

The calculated values $|\Delta x_{\text{img}}^{\text{omni}}|$ can be assessed as follows:

- $|\Delta x_{\text{img}}^{\text{omni}}| < 1$: The source image is oversampled in x-direction. The target image with the configured pixel size p_x^{per} and p_y^{per} is not substantiated by an appropriate bandwidth of the source image.
- $|\Delta x_{\text{img}}^{\text{omni}}| > 1$: The source image is undersampled in x-direction. The target image with the configured pixel size p_x^{per} and p_y^{per} does not represent the full bandwidth of the source image.
- $|\Delta x_{\text{img}}^{\text{omni}}| = 1$: The sampling interval of the target image is the same as the source image in x-direction.

The statements hold true for the y-direction as well. Figure 4.16 demonstrates $|\Delta x_{\text{img}}^{\text{omni}}|$ (left) and $|\Delta y_{\text{img}}^{\text{omni}}|$ (right) for an *exemplary* pixel size p_x^{per} and p_y^{per} with $p_x^{\text{per}}/p_x^{\text{omni}} = p_y^{\text{per}}/p_y^{\text{omni}} = 0.5$.

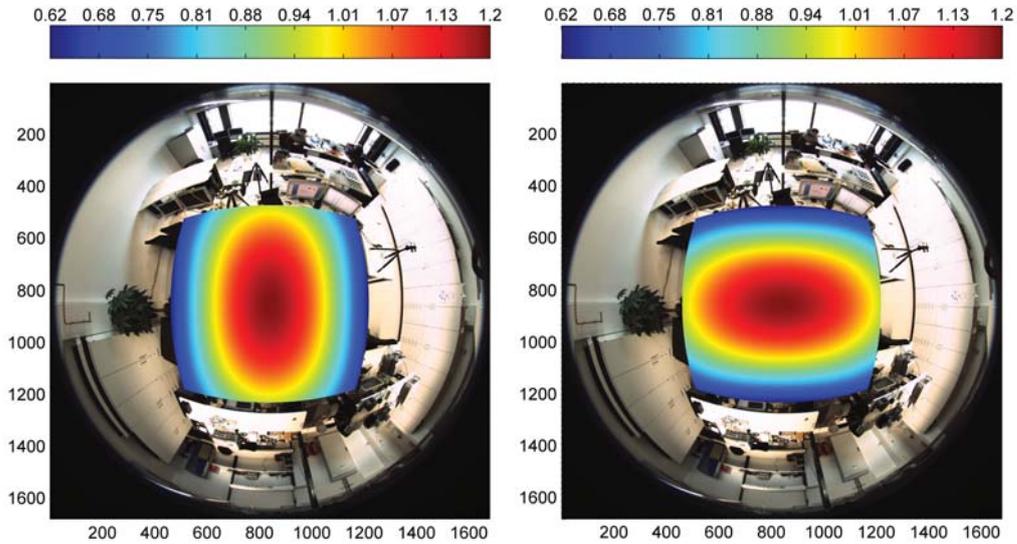


Figure 4.16: Sample interval of the target image in the source image (left: x-direction, right: y-direction)

It can be seen that this configuration causes both over-sampling as well as under-sampling. Due to the highly non-linear relationship between the two camera models, a homogeneous sampling cannot be applied. Hence for sampling the image there are two strategies conceivable: *under-sampling* and *over-sampling*.

- under-sampling: To obtain a virtual pixel size that is supported with equal underlying accuracy both directions must comply to $|\Delta x_{\text{img}}^{\text{omni}}| \geq 1 \wedge |\Delta y_{\text{img}}^{\text{omni}}| \geq 1$. In this configuration physical resolution is wasted.
- over-sampling: To obtain a virtual pixel size that does not waste any physical resolution both directions must comply to $|\Delta x_{\text{img}}^{\text{omni}}| \leq 1 \wedge |\Delta y_{\text{img}}^{\text{omni}}| \leq 1$.

The pixel size p_x^{per} and p_y^{per} where $|\Delta x_{\text{img}}^{\text{omni}}| = 1 \wedge |\Delta y_{\text{img}}^{\text{omni}}| = 1$ shall be called *marginal virtual pixel size*.

For under-sampling the marginal virtual pixel size is computed by evaluating the following test function:

$$\text{minimize} (|\nabla \mathcal{F}_x| - 1.0| + |\nabla \mathcal{F}_y| - 1.0|) \quad (4.17)$$

For over-sampling the marginal virtual pixel size is computed by evaluating the following test function:

$$\text{maximize} (|\nabla \mathcal{F}_x| - 1.0| + |\nabla \mathcal{F}_y| - 1.0|) \quad (4.18)$$

In order to solve this problem numerically a range of pixel size ratios $p_x^{\text{per}}/p_x^{\text{omni}}$ and $p_y^{\text{per}}/p_y^{\text{omni}}$ has been investigated in order to feed Equation 4.17 and Equation 4.18. The distinct value range originates from experimental trials and is outlined in Equation 4.19:

$$\begin{aligned} p_x^{\text{per}}/p_x^{\text{omni}} &\in [0.25, 2.50] \\ p_y^{\text{per}}/p_y^{\text{omni}} &\in [0.25, 2.50] \end{aligned} \quad (4.19)$$

The pixel coordinates $x_{\text{img}}^{\text{per}}$ and $y_{\text{img}}^{\text{per}}$, which are parameters for Equation 4.17 and Equation 4.18 as well are calculated by Equation 4.11:

$$\begin{aligned} x_{\text{img}}^{\text{per}} &\in [0, \text{width} - 1] \\ y_{\text{img}}^{\text{per}} &\in [0, \text{height} - 1] \end{aligned} \quad (4.20)$$

Figure 4.17 shows the overlapping contour plots of $|\Delta x_{\text{img}}^{\text{omni}}|$ and $|\Delta y_{\text{img}}^{\text{omni}}|$ for different target pixel sizes and the under-sampling (left chart) as well as over-sampling (right chart) strategy. The optimal parameter is marked with a red dot.

The ratios of pixel sizes have been calculated to $p_x^{\text{per}}/p_x^{\text{omni}} = p_y^{\text{per}}/p_y^{\text{omni}} = 0.810$ and $p_x^{\text{per}}/p_x^{\text{omni}} = p_y^{\text{per}}/p_y^{\text{omni}} = 0.417$ respectively. One may ask why a source camera following

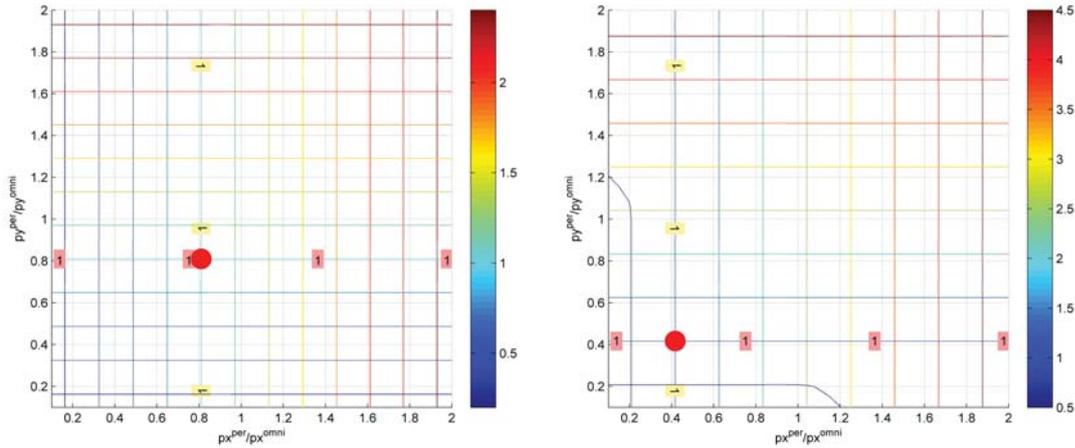


Figure 4.17: Overlapping contour plots of $|\Delta x_{\text{img}}^{\text{omni}}|$ and $|\Delta y_{\text{img}}^{\text{omni}}|$ for different ratios of pixel sizes $p_x^{\text{per}}/p_x^{\text{omni}}$ and $p_y^{\text{per}}/p_y^{\text{omni}}$, an generic projection model for the source camera, an under-sampling strategy (left) as well as over-sampling strategy (right) for the perspective target camera

the generic projection model causes pixel ratios that are below one. The reason for this is as follows: The radial symmetric projection model as presented in Section 2.2.4.1 Equation 2.29 together with the parameter set outlined in Table 4.2 (especially k_1) indicate a high radial distortion. Correcting this distortion as the target model does means "shrinking" the pixels of the source camera. The perspective characteristic of the target camera however gives rise to a subsequent slight "growth" of the pixels.

Figure 4.18 demonstrates $|\Delta x_{\text{img}}^{\text{omni}}|$ (left) and $|\Delta y_{\text{img}}^{\text{omni}}|$ (right) for the afore-calculated marginal pixel sizes. The upper images show the under-sampling and the lower ones the over-sampling strategy.

Table 4.5 and Table 4.6 outline the computed optimal virtual intrinsic parameters and pixel dimensions.

Table 4.5: Results of the intrinsic generic camera model calculation

Camera	Strategy	α_x	α_y	c_x	c_y
Cam ^{per}	under-sampling	247.23	247.62	247.50	248.00
Cam ^{per}	over-sampling	480.23	480.99	480.50	481.00

Table 4.6: Virtual pixel dimensions

Camera	Strategy	p_x	p_y
Cam ^{per}	under-sampling	1.09×10^{-5} [m]	1.09×10^{-5} [m]
Cam ^{per}	over-sampling	5.62×10^{-6} [m]	5.62×10^{-6} [m]

The values computed in the afore-presented example are only valid for this distinct configuration. They are crucially dependent on the underlying camera models (intrinsic) as well as their mutual extrinsic relationship, e.g. the rotation. The method can yet be applied for all configurations and camera models.

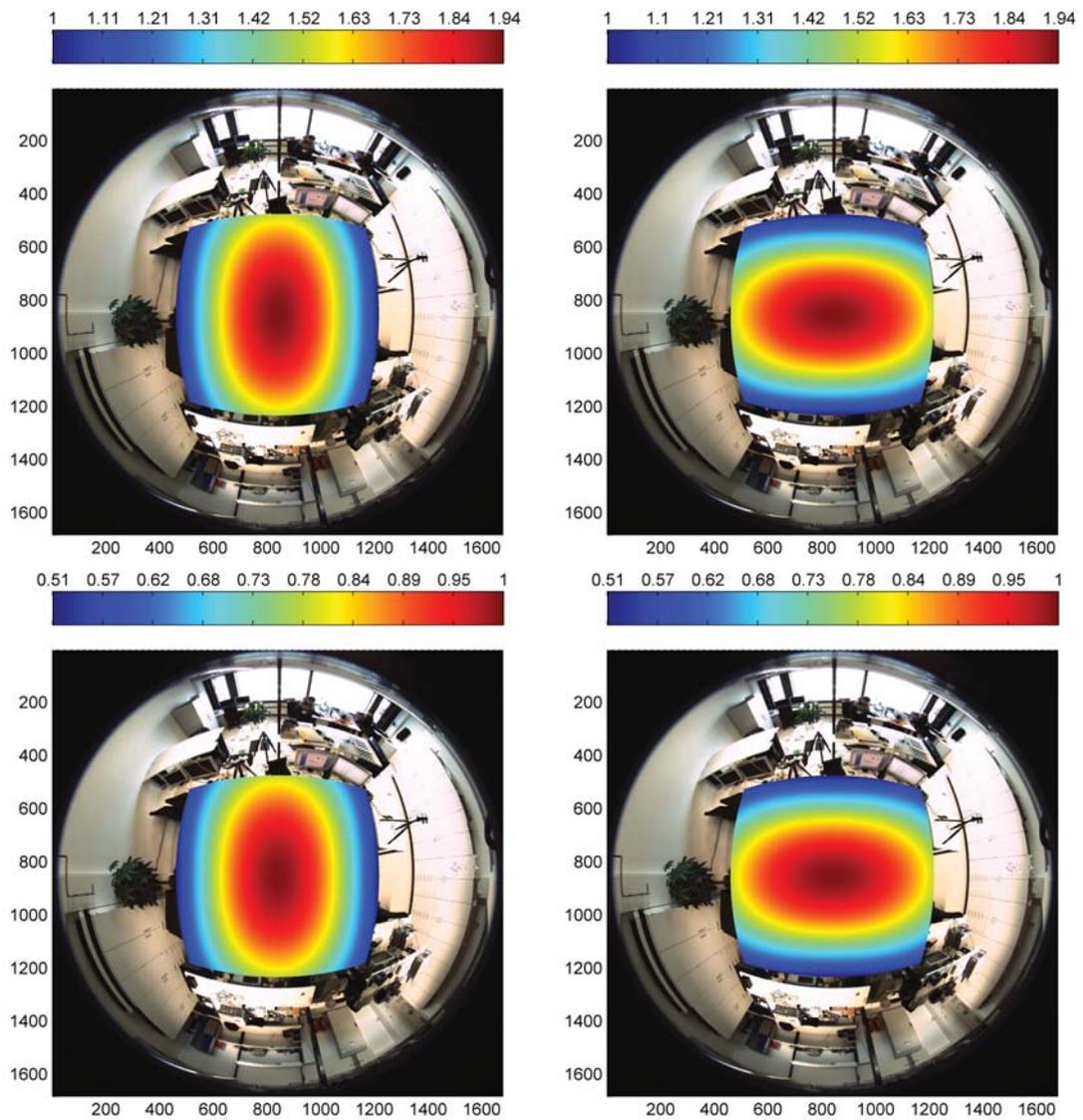


Figure 4.18: Sample interval of the target image in the source image (left: x-direction, right: y-direction, top: under-sampling, bottom: over-sampling)

4.5 Performance Measurements

The previously described algorithm for virtual camera generation by (fast-) backward mapping has been implemented in C++ as part of the professorships proprietary computer vision framework. By employing the graphical interface of this framework, the user can parameterize the underlying functionality. So it is possible to do a live parameterization of each individual perspective view. An impression is provided by Figure 4.19.

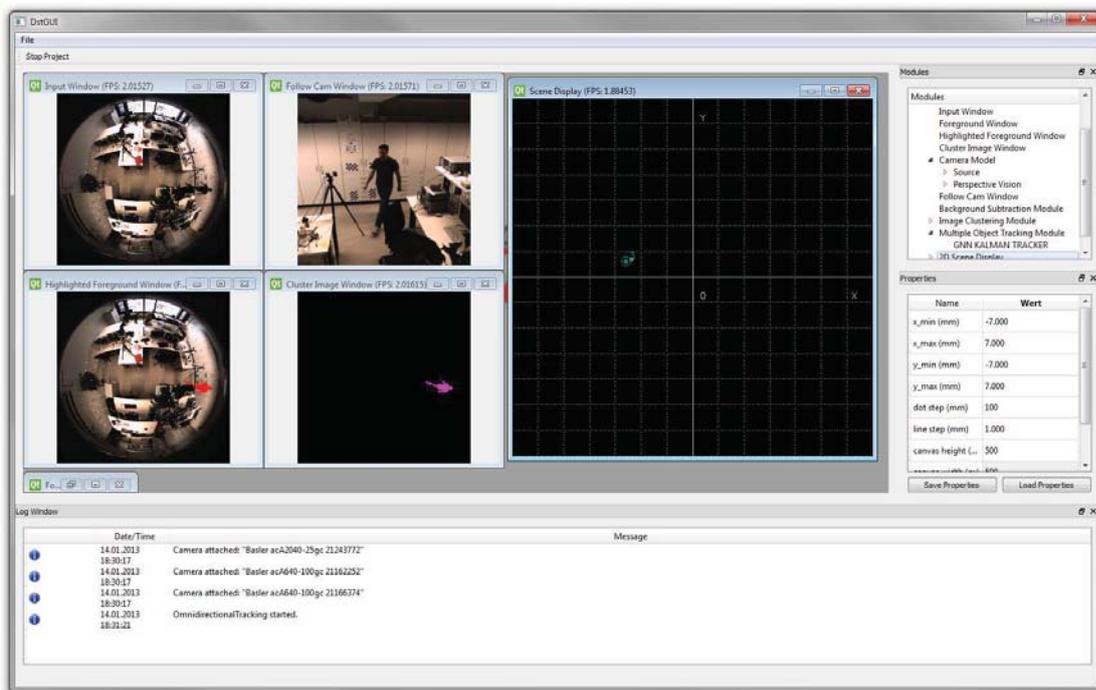


Figure 4.19: Live virtual camera generation [80]

The processing time for the conversion method was evaluated by means of eight LUTs of different sizes, described by the power of reduction n . The process, which was measured, consists of both the step for LUT generation together with the actual mapping procedure from the omnidirectional image to the perspective view. Two different resolutions of the target image were tested: 640×480 and 1280×960 . The results are shown in Figure 4.20.

As it can be seen, the processing time declines rapidly with growing n . So the aimed speed improvement in case of the optimized LUT calculation is confirmed. The fastest calculation of the 640×480 sized target image is reached with $n = 4$ (processing time: 8 ms). After that, the Gauss-filtering consumes more time than the calculation of the LUT values. For the generation of a 1280×960 perspective image, this minimum is located at $n \approx 5$ (processing time: 33 ms), because the initial LUT is 2^2 times larger than the initial one for a resolution of 640×480 . So the optimal power of reduction depends on the desired size of the target image.

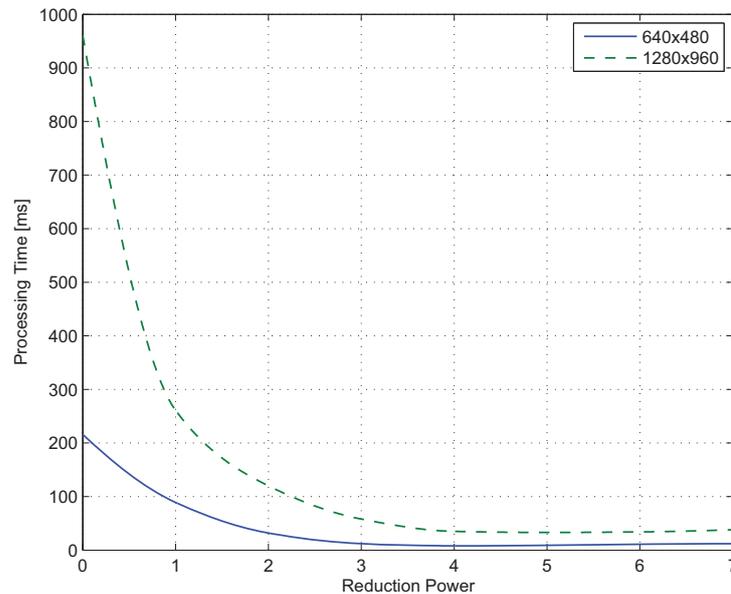


Figure 4.20: Measured processing times for the generation of perspective views with two different resolutions versus power of reduction

4.6 Virtual Perspective Views for Real-Time People Detection

In order to evaluate the practical relevance of virtual perspective views for indoor surveillance scenarios, the following test was performed: Omnidirectional images were compared with perspective ones for their eligibility of recognizing people. Therefore a system for automatic people detection and tracking was developed [82]. It comprises of the following successive stages:

1. A background subtraction algorithm is performed in order to generate hypotheses for walking persons on the omnidirectional input image.
2. A tracking algorithm is employed in order to smooth the trajectories of the moving hypotheses.
3. A vPTZ is generated and aligned in the direction of the assumed person.
4. A people detection method finally confirms the track as a human on the perspective image.

For the purpose of recognizing people the method presented by Dollar et al. [20] was employed. Figure 4.21 demonstrates a perspective view that *follows* a walking person automatically. For the announced evaluation an annotated data set of 256 frames of a moving person was created. The people detection algorithm was applied to both the omnidirectional input image and the generated virtual perspective view. Subsequently the results were compared. In the vPTZ case 228 people were positively detected compared

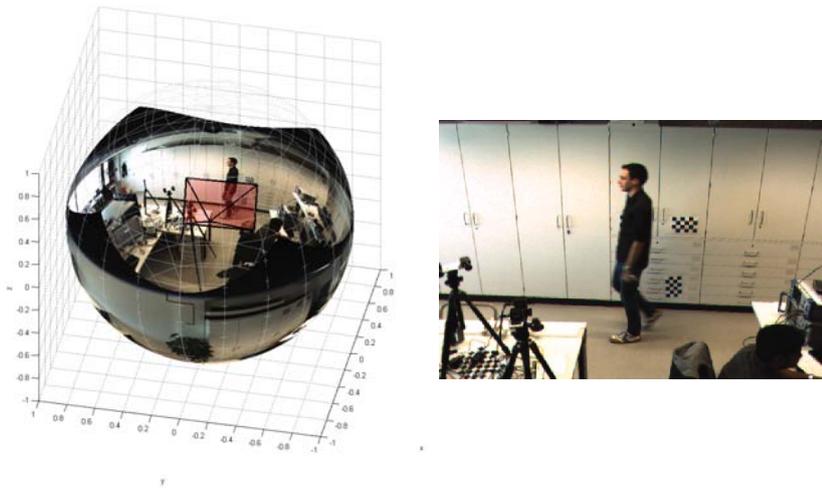


Figure 4.21: Automatic people tracking system employing an omnidirectional imaging device and a vPTZ [82]

to only 62 in the omnidirectional images. This corresponds to average detection rates of 0.89 and 0.24 respectively. More precisely Figure 4.22 shows the spatial distribution of the detection rates in polar coordinates of the ground plane. The polar coordinate system consists of 32 angular bins and 5 radial bins. The maximum radius is 8 m. Green areas show high detection rates whereas red areas indicate poor detection rates. White spaces represent areas where no measurements were recorded.

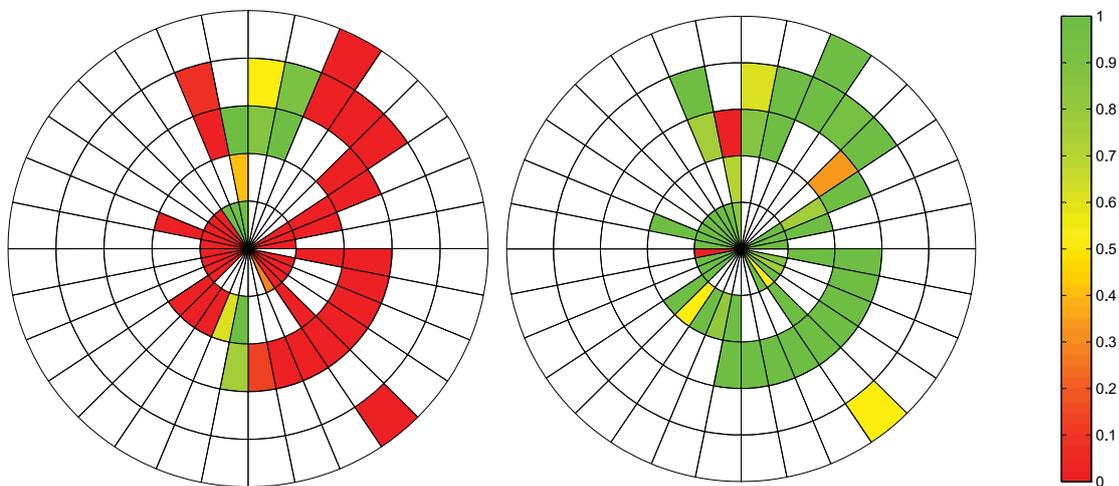


Figure 4.22: Distribution of detection rates using an omnidirectional image directly (left) compared to the use of virtual perspective views (right) [82]

Figure 4.22 visualizes that if a people detection algorithm is applied to the omnidirectional image directly, detection rates are only high when the person is visible in an upright (or upside down) position. By comparison the virtual perspective views produce consistently high detection rates for all positions. More details can be reviewed in the publication of Meinel et al. [82].

Chapter 5

Omnidirectional Stereo Vision

This chapter presents and examines the idea of *omnidirectional stereo vision* which has been motivated in Chapter 1. Therefore an extended overview about existing approaches is presented in Section 5.1. Especially the categorization of such systems with respect to the geometrical camera configuration will be taken into focus and major derivatives are compared. As stated in Chapter 3 parallel epipolar lines are a major precondition for stereo processing. Due to this, Section 5.2 outlines important rectification methods for the afore-presented stereo configurations.

Based on the findings from Section 5.1 and Section 5.2, an idea for a novel omnidirectional stereo vision setup is presented in Section 5.3.

5.1 Introduction and Related Works

The challenge of omnidirectional stereo vision has been discussed in several previous scientific publications. A comprehensive overview about the vast amount of approaches can be reviewed in the publication by Zhu [146].

A rough classification can be done by distinguishing strategies of *image generation*, different *geometrical configurations*, the achievable *field of view*, target *applications* and a few more as briefly explained in the following:

- **Image generation:** While special omnidirectional lens types can create a very wide-angle image, multiple ordinary perspective cameras have to be employed in order to achieve this. Besides this static approach, moving imaging devices can be employed to widen the FOV.
- **Geometrical configurations:** The number of used sensors as well as their mutual geometrical alignment differ. This has a high impact on properties like FOV, static or dynamic behaviour and type of sensor for instance.

- **Field of view:** The notion *omnidirectional stereo* is vague and does not explicitly describe a certain quantity of coverage. For instance it can refer to a panoramic projection, or to a spherical one respectively.
- **Application:** Heterogeneous fields of application drive the development of powerful optical sensors, like *robot navigation*, *visual surveillance* or *automotive*.
- **Miscellaneous:** A lot of alternative properties can be chosen to distinguish between different systems. While a stereo rig provides a *fixed* baseline between two cameras, two moving robots can form a stereo system as well, but with a *floating* baseline. So the *baseline* is one further property of classification amongst others.

It is obvious that those properties are merely mutual independent. In Figure 5.1 a mind-map presents an overview about this complex issue of classification.

Table 5.1 refers to important publications that try to organize and investigate the vast amount of possibilities in omnidirectional stereo imaging.

Table 5.1: Relevant publications that categorize omni-stereo vision approaches

Year	Author(s)	Remark	Publications
1996	Southwell, D. et. al	Catadioptrical are exploited for panoramic stereo.	[118]
1998	Sameer A. Nene and Shree K. Nayar	A single camera is used with multiple (planar, ellipsoidal, hyperboloidal, and paraboloidal) mirrors to create wide-angle stereo.	[88]
2001	Zhigang Zhu	Classifies multiple state-of-the-art omnidirectional stereo vision setups, like binocular, n-ocular, circular projection and dynamic omni-stereo.	[146]
2006	El Mustapha Mouaddib et. a.	Investigates catadioptric omnidirectional stereovision systems with horizontal and vertical shifts of the mirrors center positions.	[86]
2010	Tang Yi-ping et. al	Reviews the existing panoramic imaging technologies. Obtains $360^\circ \times 360^\circ$ full sphere panoramic image by integrating to images recorded by two symmetrical omnidirectional vision sensors.	[135]
2012	O. Schreer et. al	Geometrical concepts of omnidirectional and omni-stereoscopic multi-camera systems are presented.	[109]

In the following the geometrical configurations shall be exploited for categorization.

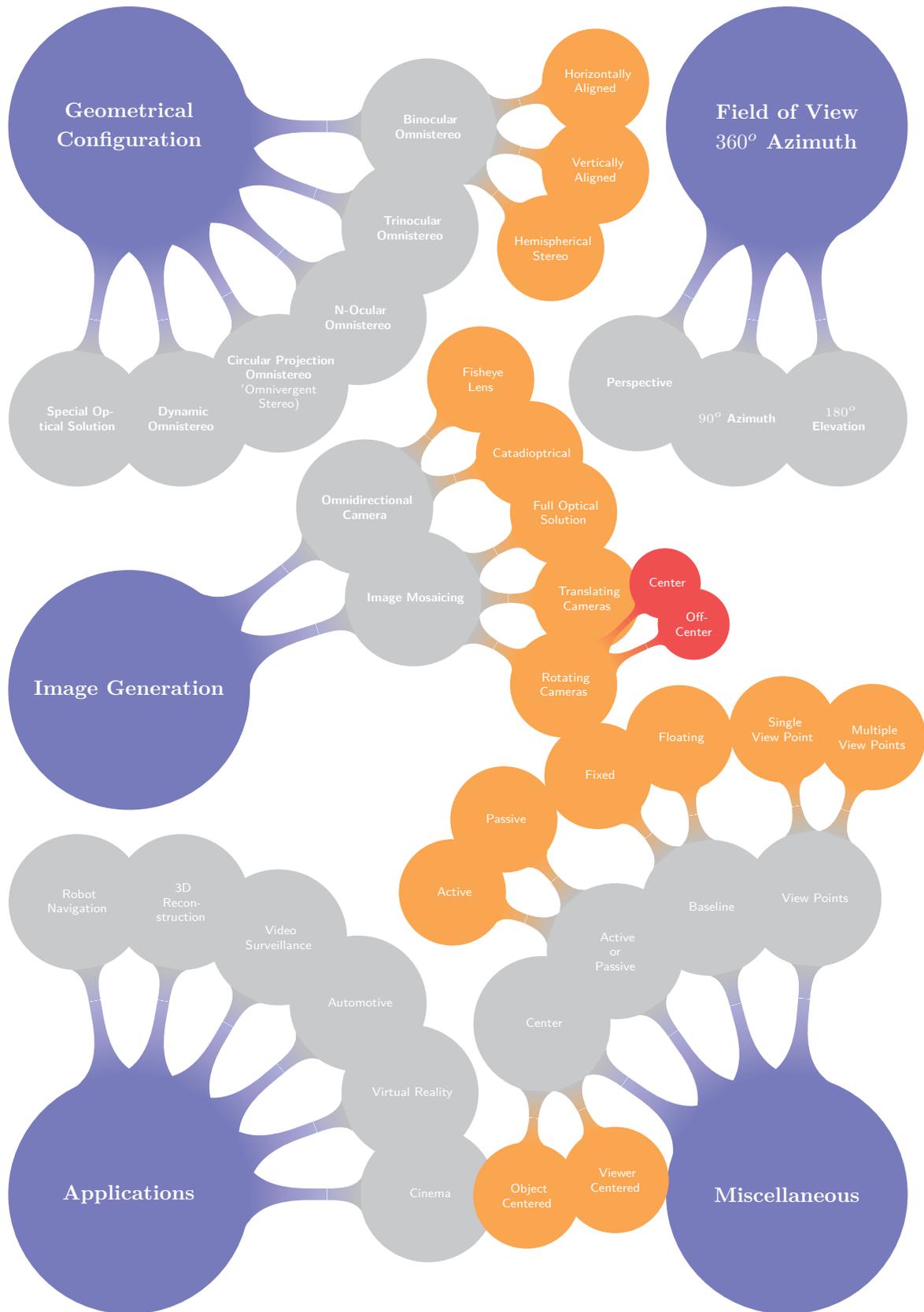


Figure 5.1: Overview of categorizations for omni-stereo vision approaches

5.1.1 Geometrical Configuration

The number of applied sensors is the crucial property for the geometrical configuration.

By means of one single sensor one can obtain 3D measurements by moving somehow the sensor itself. This can be achieved by an arbitrary motion, e.g. by a roboter, and is called *dynamic omni-stereo*. If the dynamic action constrains to a rotation, this technique is known as *circular projection omni-stereo* also denoted as *omnivergent stereo*.

The two view configuration is the most common variant not limited to applications that claim omnidirectional views. It is called *binocular omni-stereo*. Enclosing more views to the computation, one separates between *trinocular omni-stereo* (three views) and *n-ocular omni-stereo* (more than three views).

It is worth mentioning that dynamic stereo techniques are not constrained to single sensor systems. Rather, multiple moving sensors can form highly flexible 3D measurements configurations, e.g. in cooperative robotic swarm environments.

The publication by Zhu [146] compares the afore-mentioned geometrical configurations with respect to crucial parameters as can be seen in Table 5.2.

Table 5.2: Comparison of omni-stereo principles [146]

omni-stereo configuration	view-points	base-lines	vergence	epipolar geometry	depth error in directions	depth error in distance	mutual occlusion	stereo viewing
H-Binocular	2, fixed	fixed	Largely un-uniform	sine curve	non-isotropic	$\propto D^2$	yes	okay
V-Binocular	2, fixed	fixed	Similar to perspective	vertical lines	isotropic	$\propto D^2$	no	no
N-ocular	N (a few), fixed	fixed	Select from C_N^2 pairs	sine curve	roughly isotropic	$\propto D^2$	maybe no	okay
Dynamic	2, freely movable	optimal for given points	max for a few points	sine curve	optimal for given points	$\propto D^{1.5}$	mutual awareness	N/A
Viewer-centered Circular Projection	many on a small circle	fixed	max for all points	horizontal lines	isotropic	$\propto D^2$	no	good
Object-centered Circular projection	many on a small circle	fixed	max for all points	horizontal lines	isotropic	$\propto D^2$	no	good
Parallel-Perspective Projection	many on a camera path	optimal for all points	max for all points	horizontal lines	uniform everywhere	uniform or $\propto D$	no	good

The following subsections distinguish between different configurations and refer to relevant publications that have been released in that area so far.

5.1.1.1 H-Binocular Omni-Stereo with Panoramic Views

Most binocular setups employ panoramic images that are a 360-degree cylindrical projection from a single viewpoint where the projection in the vertical direction is perspective [146]. Figure 5.2 shows a horizontally-aligned omni-stereo vision configuration where two views are arranged in a horizontal setup with a fixed baseline.

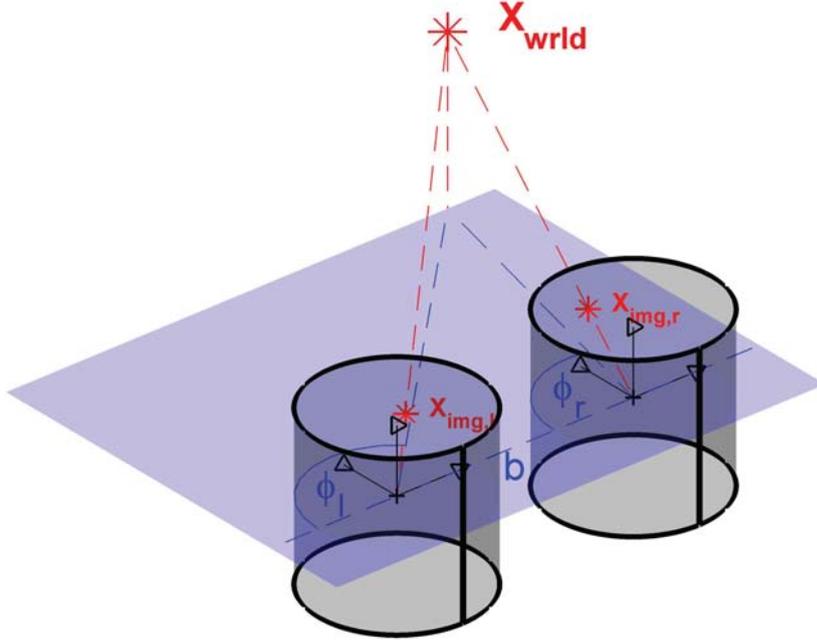


Figure 5.2: H-binocular omni-stereo with panoramic views

This approach guarantees a horizontal panoramic 3D measurement field while the observable FOV in vertical direction is constraint to common perspective projection characteristics.

According to a common stereo configuration as described in Section 3.5 a physical depth Z can be computed with respect to the left camera:

$$Z = b \cdot \frac{\sin \phi_r}{\sin(\phi_l - \phi_r)} = b \cdot \frac{\sin \phi_r}{\sin(\phi)} \quad (5.1)$$

where

- b is the physical baseline between both sensors (see Section 2.4.1). $[b] = mm$
- ϕ_l is the angle of the left incoming ray on the principle plane (blue plane in Figure 5.2) with respect to the baseline (blue line in Figure 5.2) connecting the centres of both imaging devices) . $[\phi_l] = rad$
- ϕ_r is the angle of the right incoming ray on the principle plane (blue plane in Figure 5.2) with respect to the baseline (blue line in Figure 5.2) connecting the centres of both imaging devices) . $[\phi_r] = rad$

- ϕ is the physical disparity $\phi_l - \phi_r$ of the world point projections between left and right camera. $[\phi] = rad$

According to Zhu [146] all epipolar curves are sinusoidal and a triangulation is not defined for the singularity points when $\phi_l = \phi_r = 0^\circ$. The obtained depth error is non-isotropic, proportional to the square of Z and inversely proportional to b . The latter fact is congruent with perspective stereo.

A mentionable representative of the principle of horizontally-aligned omni-stereo was presented by Kang et al. [52]. In their publication they compose cylindrical images from video sequences taken by vertically rotating cameras and recover 3-D point clouds by structure from motion. For the computation of full 3D maps, they overcome the blind-spot problem by means of three sensors. A schematic configuration as presented in [52] can be seen in Figure 5.3.

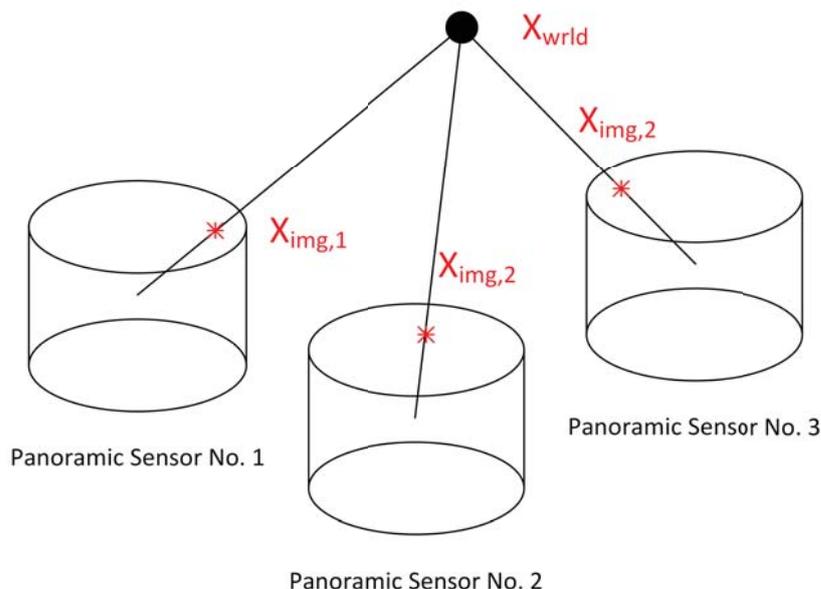


Figure 5.3: Example for horizontally-aligned omni-stereo [52]

Based on the results it can be concluded that the afore-presented method is not appropriate for fast real-time stereo matching since the epipolar curves are not parallel straight lines. Rather they are sinusoidal. Standard methods for dense stereo computation cannot be applied - it is more appropriate for approaches that triangulate sparse features. The limited vertical FOV makes it not very reasonable for indoor surveillance applications.

An overview about further related works with respect to this technology can be reviewed in Appendix B Table B.1.

5.1.1.2 V-Binocular Omnistereo with Panoramic Views

Figure 5.4 shows a vertically-aligned stereo configuration where two cylindrical imaging devices as published by Yamashita et al. [133] are arranged in a vertical setup with a fixed baseline.

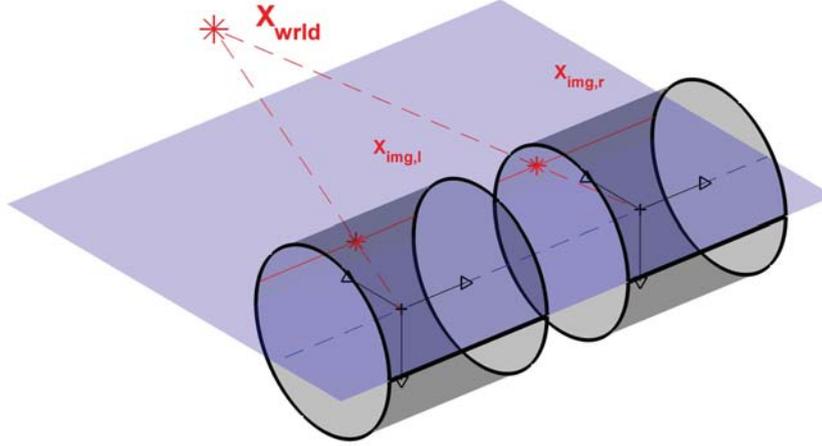


Figure 5.4: V-binocular omni-stereo with panoramic views

As stated by Zhu [146], the disparity computation for this configuration is equivalent to the method used for ordinary perspective stereo computation - cf. Section 3.4 and Section 3.5. Hence the depth Z computation in Equation 5.2 is equivalent to Equation 3.17.

$$Z = b \cdot \frac{\alpha_x}{x_{\text{img},l} - x_{\text{img},r}} \quad (5.2)$$

In contrast to the triangulation process for perspective stereo vision setups where the depth value Z expresses the z -component of the 3D point (given in camera coordinates), the current method represents the norm of the y - z -component for vertically aligned omnidirectional systems.

The depth error computation δZ for this configuration is equivalent to the methods presented in Section 3.5.

All epipolar curves are parallel (red solid lines in Figure 5.4) and a triangulation is constrained as for perspective cameras since the imaging is perspective in x -direction for cylindrical projections.

As an example Mokri et al. [84] apply vertically-aligned omni-stereo in order to establish a navigation system for an autonomous robot. They equip the device with two vertically stacked catadioptrical cameras and measure the depth in a 360-degree field of view by computing two panoramic images, as can be seen in Figure 5.5.

Finally it can be concluded that the v-binocular approach is relevant with reference to stereo correspondence computation. The epipolar constraint is formed by parallel

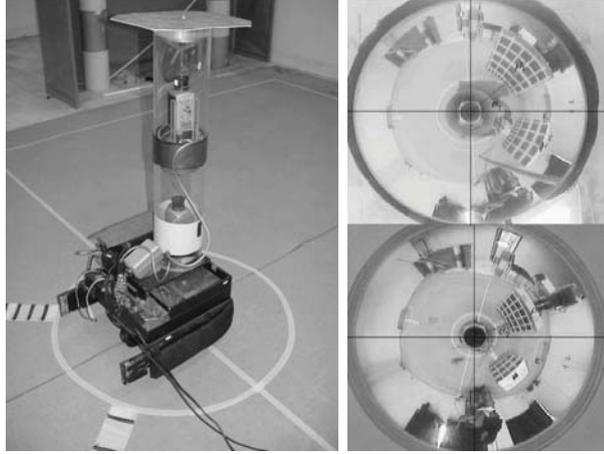


Figure 5.5: Example for vertically-aligned omni-stereo [84]

lines, which enables the straight application of standard correspondence methods. The images do not need to be rectified. Although the vertical field of view is omnidirectional, the horizontal one is perspective and hence too constrained for reducing the amount of employed sensors in visual indoor surveillance applications.

An overview about further related works with respect to this technology can be reviewed in Appendix B Table B.2.

5.1.1.3 Binocular Omnistereo with Hemispherical Views

Figure 5.6 shows a binocular omni-stereo setup with hemispherical views.

From Section 2.4.3 it is known that the epipolar constraint for two hemispherical views that naturally employ the equiangular projection model from Section 2.2.3 is not represented as parallel lines. Rather these curves are so-called *great circles* and hence sinusoidal as can be seen in Figure 5.7 (green solid lines).

A physical depth Z can be computed equivalently to Equation 5.1 with respect to the left camera as published by Li [67]:

$$Z = b \cdot \frac{\sin \phi_r}{\sin(\phi_l - \phi_r)} = b \cdot \frac{\sin \phi_r}{\sin(\phi)} \quad (5.3)$$

Although providing a full 180° FOV horizontally as well as vertically, a triangulation is not defined for the singularity points when $\phi_l = \phi_r = 0^\circ$ - cf. Section 5.1.1.1. This case is also expressed in the mathematical relationship that computes a distance error δZ :

$$\delta Z = \left[b \cdot \frac{\cos(\phi - \phi_l)}{\sin \phi} - \phi \cdot \frac{\sin(\phi - \phi_l) \cos \phi}{\sin^2 \phi} \right] \cdot \delta \phi \quad (5.4)$$

where

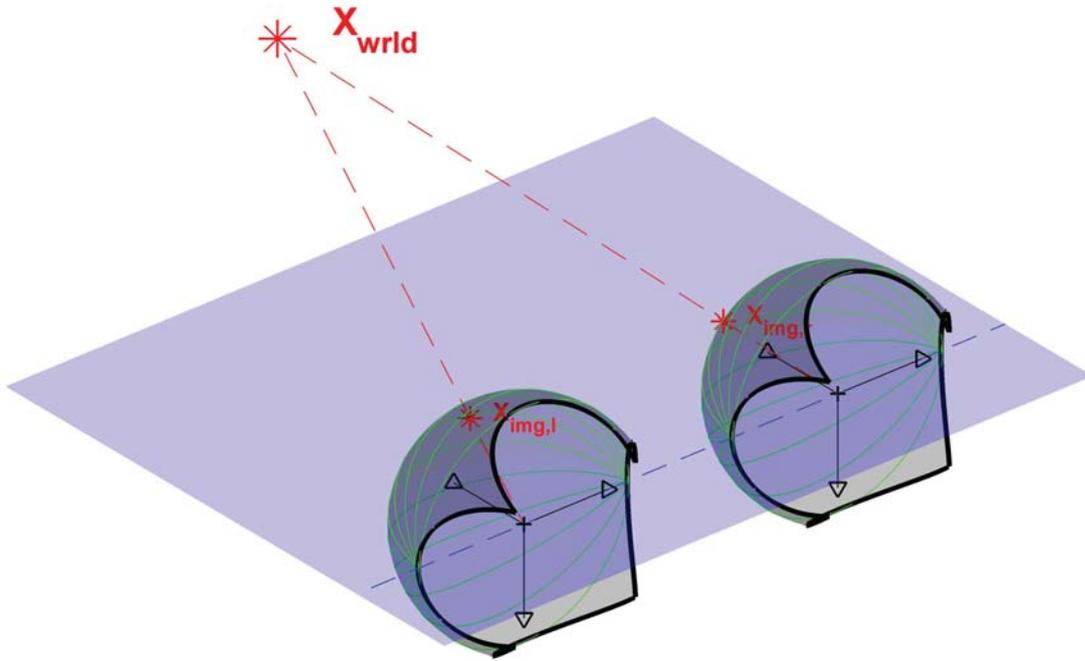


Figure 5.6: Binocular omni-stereo with hemispherical views

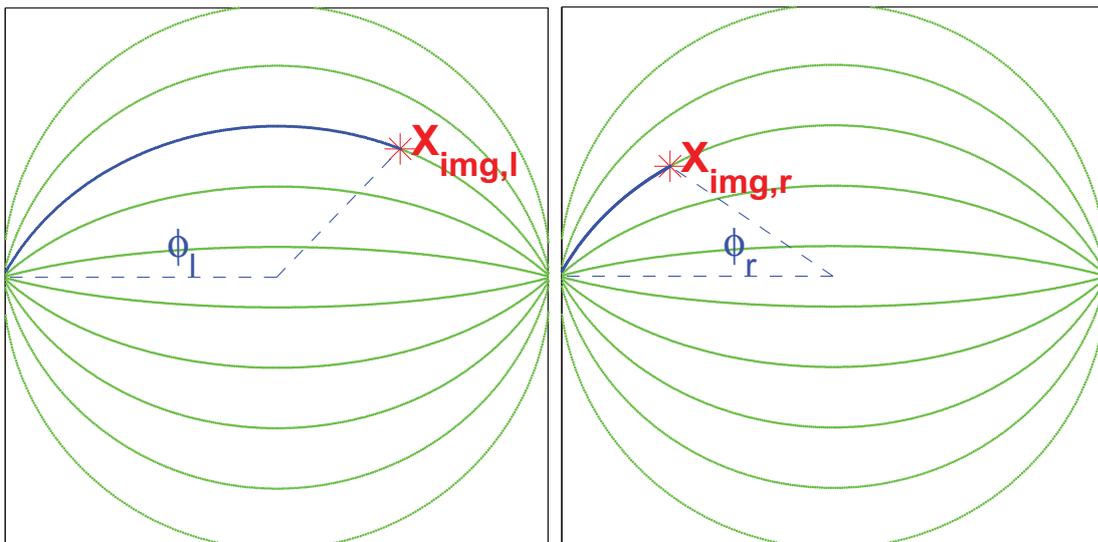


Figure 5.7: Epipolar lines of binocular omni-stereo with hemispherical views

- $\delta\phi$ is the accuracy a physical disparity can be determined with. $[\delta\phi] = rad$

The closer ϕ_l gets to zero or π , and consequently the closer the point to triangulate approaches the baseline (\equiv the singularity points) - the more the measurement error increases. The smallest error is reached, when $\phi_l = \pi/2$. Figure 5.8 plots qualitatively the circumstances of Equation 5.4. The graphic visualizes the ascending error beginning at the most accurate value δZ up to $5 \cdot \delta Z$. Beyond that value the error further heads for infinity.

For applications that require a full hemispherical depth map, a two-view approach is not sufficient. This notion will be picked up again in Chapter 8.

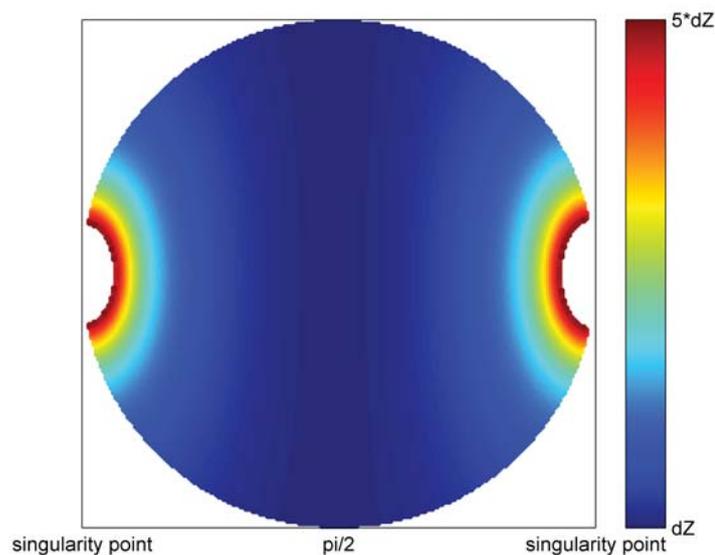


Figure 5.8: Relationship between azimuth angle ϕ_l and qualitative error measurement δZ

As an example Zhu [145] applied the principles of binocular omni-stereo with two fish-eye sensors as can be seen in Figure 5.9. The physical projection process is approximated by a Taylor model. To overcome the challenge of non-parallel epipolar lines in omnidirectional views he employed a virtual perspective projection for rectification.

An overview about further related works with respect to this technology can be reviewed in Appendix B Table B.3.

5.1.1.4 Trinocular Omnistereo

As stated in Section 5.1.1.1, Section 5.1.1.2 and Section 5.1.1.3, binocular solutions are subject to the drawback that a disparity is not defined for the singularity point. Li [67] overcomes this drawback by the application of a third camera. The configuration from

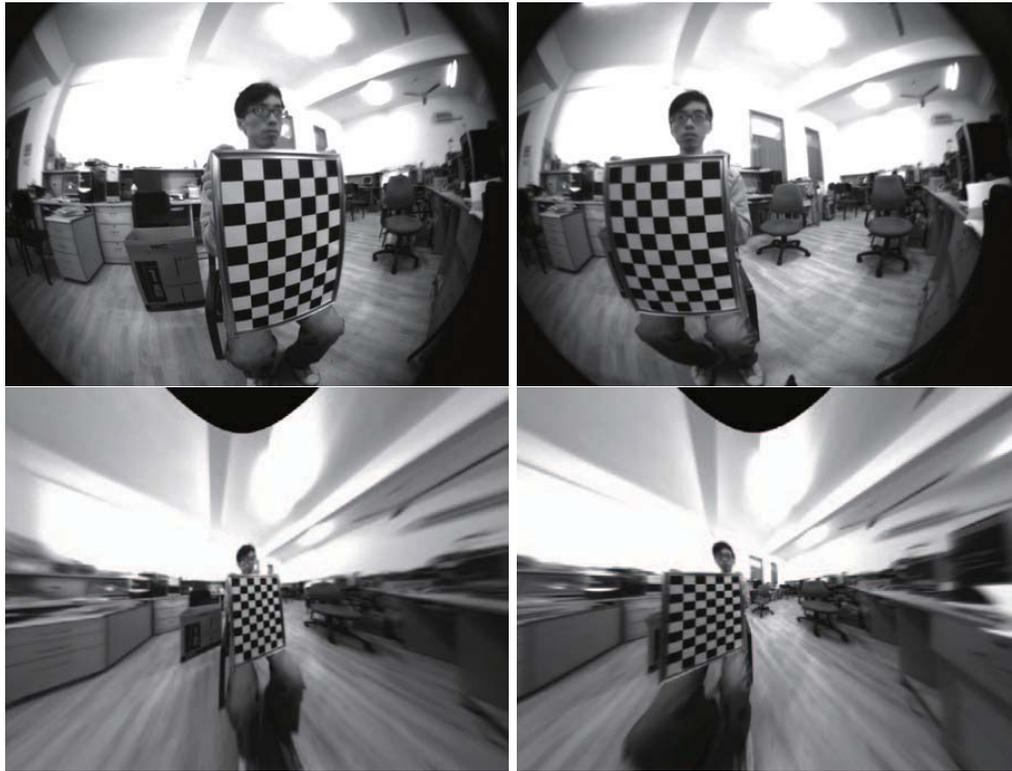


Figure 5.9: Example for binocular omni-stereo with hemispherical views (top left: left image, top right: right image, bottom left: left rectified image, bottom right: right rectified image) [145]

Section 5.1.1.3 can be enhanced to a trinocular omnidirectional stereo approach as can be seen in Figure 5.10.

The mathematical relationships for depth- Z computation are equivalent to Equation 5.3 and Equation 5.4. Moreover this configuration comprises two stereo cameras where each epipolar constraint is formed by *great circles* as can be seen in Figure 5.10 as green- and magenta-coloured epipolar curves.

The illustration of Figure 5.10 is inspired by Li [67]. The author generates a hemispherical depth map by means of three fish-eye cameras mounted in a right-angled configuration. Obviously the problem of singularity points can be solved thereby.

To overcome the problem of sinusoidal epipolar curves for real-time stereo matching as described by Zhu [146] a *gnomonic projection* i.e. a *latitude-longitude sampling* method is used. The method is treated in Section 5.2. In fact the author rectifies the dedicated images for both stereo cameras up to the epipols and correlates them separately. For the obtained disparity maps he reports from severe distortions for the calculated depth values close to the epipoles - cf. Figure 5.8. The generation of the final hemispherical depth map is subsequently done by the merge of both maps. Therefore he averaged the computed distances weighted by the calculated error to be expected. A quantitative error assessment for that configuration is not provided.

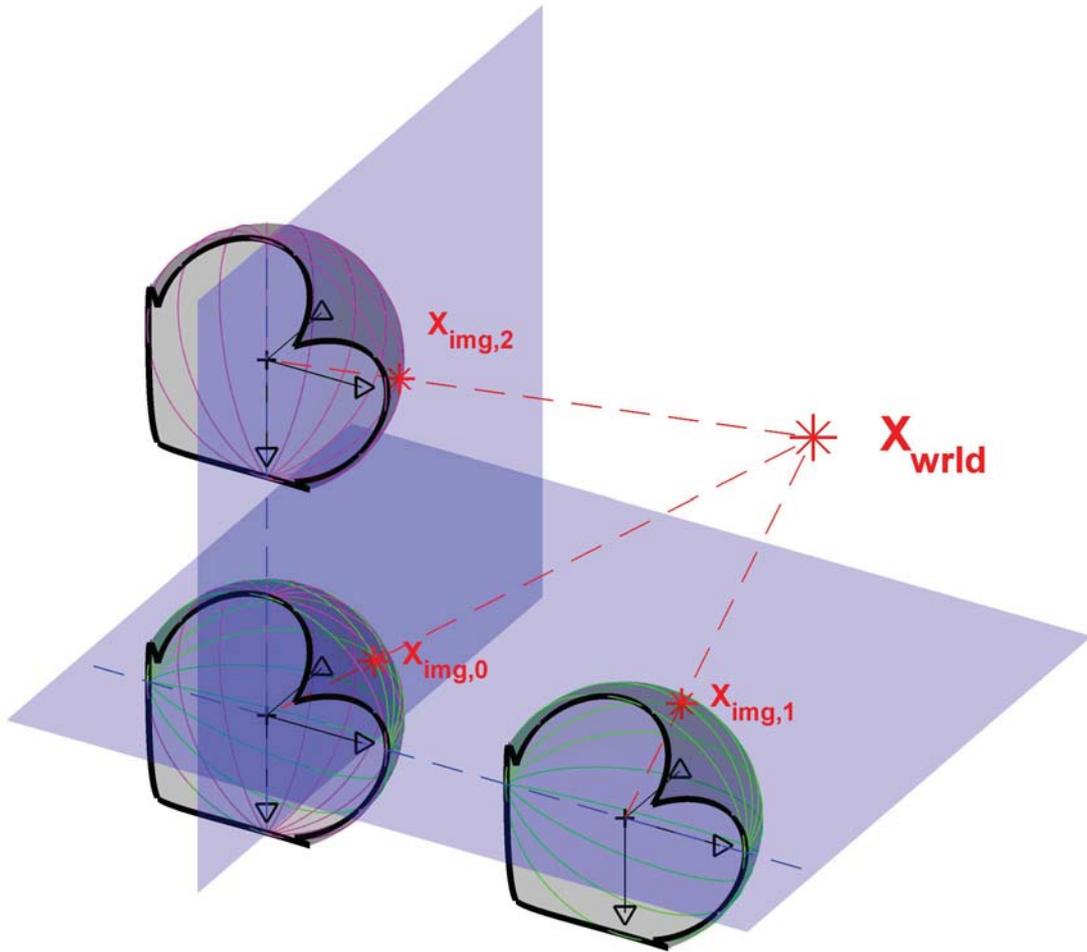


Figure 5.10: Trinocular omnidirectional stereo vision with hemispherical views

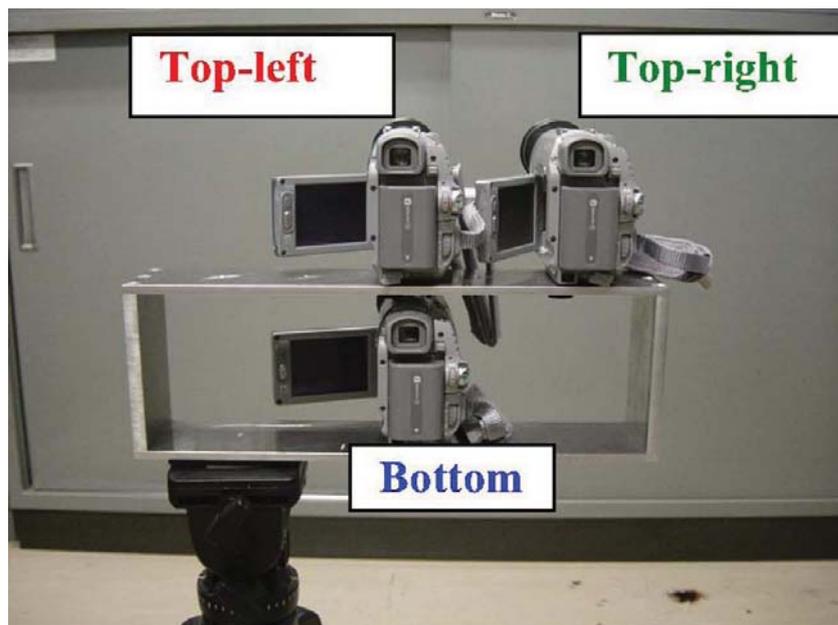


Figure 5.11: Example for trinocular omnidirectional stereo vision with hemispherical views [67]

An overview about further related works with respect to this technology can be reviewed in Appendix B Table B.4.

5.1.1.5 Miscellaneous Configurations

In addition to the afore-mentioned geometrical configurations, the principles of omnidirectional stereo have further been realized by means of alternative approaches. Although those approaches are of less importance for this work, yet they shall be mentioned for completeness:

- **N-ocular omnidirectional stereo vision** - More than two or three views (n-ocular) are used extensively to improve the results of 3D measurements. For instance Sogo et al. [116] use several catadioptrical sensors for real-time human tracking in order to tackle the correspondence problem amongst multiple targets and thereby increase the measuring accuracy of target locations. An overview about further related works with respect to this technology can be seen in Appendix B Table B.5.
- **Omnivergent stereo vision** - Rotating single or multiple cameras around a fixed axis realize so-called *omnivergent stereo*. Most approaches as published by Peleg [94–96] use ordinary perspective cameras in order to create panoramic projections and hence panoramic stereo vision. An overview about further related works with respect to this technology can be found in Appendix B Table B.6.
- **Dynamic omnidirectional stereo vision** - Creating 3D information by one or multiple moving camera(s) is an established method in robotics. For instance Zhu et al. [147–149] generate dynamic panoramic stereo vision in a robot navigation scenario in order to find and protect humans by a robot team in an emergency circumstance. They call this approach Panoramic Virtual Stereo Vision (PVS). An overview about further related works with respect to this technology can be viewed in Appendix B Table B.7.
- **Omnidirectional stereo vision with special optical solutions** - Besides the common imaging technologies based on perspective or omnidirectional (e.g. catadioptric and fisheye) optics, there exist multiple uncommon technologies. An overview about related works with respect to this technology can be reviewed in Appendix B Table B.8.

5.2 Epipolar Rectification

All stereo vision imaging devices share a common aspect: In order to reduce the processing load for stereo correspondence significantly an appropriate epipolar rectification approach has to be introduced. For perspective views the need for rectification results only from inaccurate stereo vision setups as described in Section 3.3. For divergent projection forms as spherical cameras (cf. Section 2.2.3 and Section 2.2.4) even perfectly aligned camera configurations give rise to non-parallel epipolar lines as can be observed in Figure 5.7.

According to Abraham et al. [3] an "epipolar image rectification can be defined as geometric transformation of an image pair to an image pair which has the special property, that every scene point" $\mathbf{X}_{\text{world}}$ "is projected in both images into the same row" $y_{\text{img},l} = y_{\text{img},r}$. The vertical parallax disappears.

From Figure 5.12 it can be observed that the baseline results in a pencil of planes in the three-dimensional space together with the epipolar lines of a binocular spherical stereo vision system. The angle between these planes and the optical axes is denoted by β .

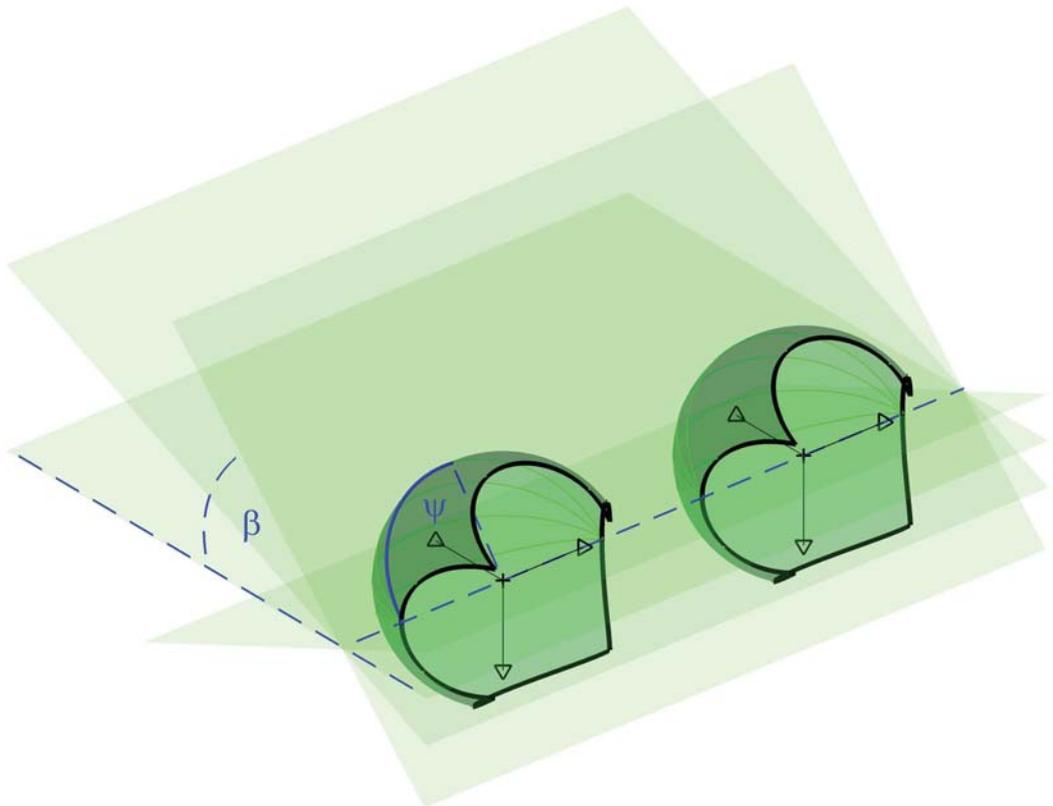


Figure 5.12: Epipolar planes in an binocular stereo vision system with hemispherical views

In order to project the same camera point in both images on the same row, the rows have to be composed of the identical epipolar planes, hence have to correspond to the

A depth value Z for an appropriate stereo configuration can be calculated in dependence of pixel disparity values according to Equation 3.17. Note that the depth values Z represent the norm of the y-z-component of a 3D point given in camera coordinates.

5.2.2 Epipolar Equi-Distance Rectification

Following Abraham et al. [3], a modified epipolar equidistant model with parallel epipolar lines can be described using the following projection function $\mathcal{F}_{\text{rect},2}$:

$$\tilde{\mathbf{X}}_{\text{norm}} = \mathcal{F}_{\text{rect},2}(\tilde{\mathbf{X}}_{\text{cam}}) = \left(\arctan \frac{x_{\text{cam}}}{d}, \arctan \frac{y_{\text{cam}}}{z_{\text{cam}}} \right)^T = (\psi, \beta)^T \quad (5.8)$$

It is obvious that the component x_{norm} corresponds with equidistant steps on the great circles from Figure 5.12 denoted as angle ψ . Hence this model can overcome the perspective projection limitations while keeping the rectified rows of constant β . Li [67] calls this projection method latitude-longitude sampling.

By the way, the angle ψ of Equation 5.8 is concurrent to the angle ϕ_l for the left image, and ϕ_r of the right image respectively, of Equation 5.3.

A depth value Z for an appropriate stereo configuration can be derived from Equation 2.15, Equation 5.3 and Equation 5.8 to the following relationship:

$$Z = b \cdot \frac{\sin \phi_r}{\sin(\phi_l - \phi_r)} = b \cdot \frac{\sin \left(\frac{\pi}{2} + \frac{x_{\text{img},r} - c_x}{\alpha_x} \right)}{\sin \frac{d}{\alpha_x}} \quad (5.9)$$

Note that the depth values Z represent the euclidean distance of a 3D point given in camera coordinates.

5.2.3 Epipolar Stereographic Rectification

Finally Abraham et al. [3] present the epipolar stereographic rectification method as an alternative for the epipolar equidistant model. It is appropriate for usage in real-time applications if no trigonometric functions (cf. Equation 5.8) shall be used and no lookup table operation is possible. The projection model can be described as follows:

$$\begin{aligned} \tilde{\mathbf{X}}_{\text{norm}} &= \mathcal{F}_{\text{rect},3}(\tilde{\mathbf{X}}_{\text{cam}}) \\ &= \left(\frac{x_{\text{cam}}}{\sqrt{x_{\text{cam}}^2 + y_{\text{cam}}^2 + z_{\text{cam}}^2} + \sqrt{y_{\text{cam}}^2 + z_{\text{cam}}^2}}, \frac{y_{\text{cam}}}{\sqrt{y_{\text{cam}}^2 + z_{\text{cam}}^2} + z_{\text{cam}}} \right)^T \quad (5.10) \end{aligned}$$

5.2.4 Comparison of Rectification Methods

In Figure 5.14 the horizontal projection behaviour of the different rectification approaches are outlined.

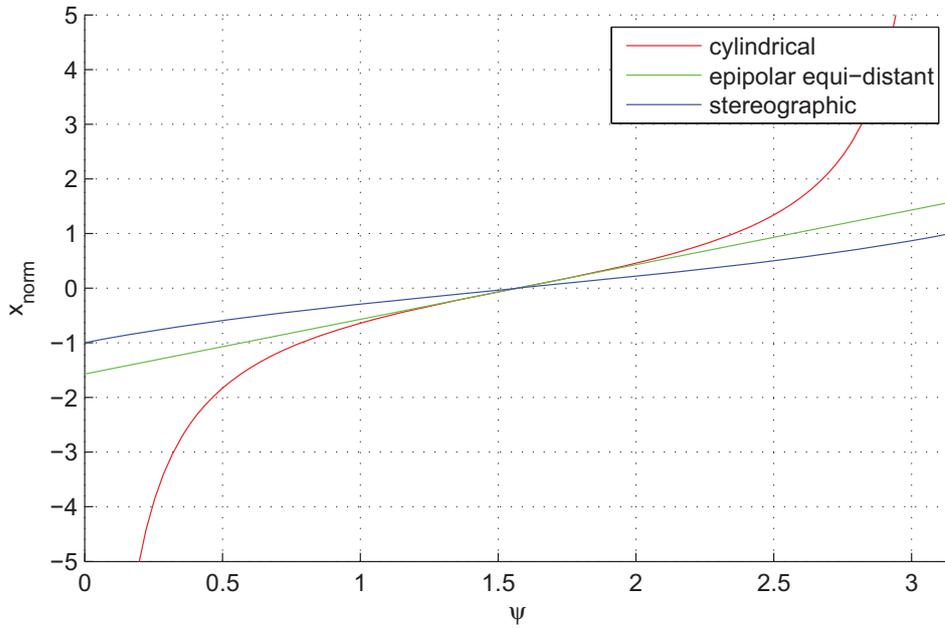


Figure 5.14: Horizontal projection behaviour for different rectification methods

It can be recognized that the epipolar equidistant method projects isogonal (equal of angle) as its mapping curve forms a straight line over the whole range of ψ . The stereographic approach approximates the latter one quite well without the use of trigonometric functions. Projecting onto a cylinder however only approximates equi-angular characteristics in the center where $\psi = \pi/2$ but the closer we approach the epipoles the more the perspective influence prevails.

5.3 A Novel Spherical Stereo Vision Setup

Based on the knowledge gained from Section 5.1 and Section 5.2 the trinocular camera configuration shall be chosen for further investigation since this setup is described to be the one with the smallest number of employed cameras for the generation of a full hemispherical depth map [67]. Hence it is the right approach for obtaining a vast observation field with one sensor (which comprises three cameras) in visual indoor surveillance applications.

As disadvantageous for the algorithm described in Section 5.1.1.4 can be considered that two full-hemispherical depth maps of approximately original image size are calculated with both showing high inaccurate depth data near the epipoles:

- The rectification and subsequently the stereo processing of image data near the epipols lead to multiple redundant operations. For example if the rectified image comprises 1000 rows, the singularity pixel will be mapped to each of these lines during epipolar equidistant rectification. After 1000 redundant stereo operations the obtained depth information is discarded in the merge stage.
- The weighted average method applied involves the computation of the actual depth error for each pixel under consideration. Since this procedure includes the calculation of multiple trigonometric functions (cf. Equation 5.4), it may not be appropriate for real-time processing.
- When a stereo sensor is designed, its error specification bases on the worst case error. Whatever the merge procedure of Li achieves, the worst case error value cannot be more accurate than the best value of the binocular stereo vision setup with hemispherical views - cf. Section 5.1.1.3 Figure 5.8. This is because in his setup four singularities are included in the merge process - each time with the best value from the complementary stereo camera. The question is now, whether another geometrical constellation and the employment of narrow-angled views without complicated merge procedure may lead to the same result - at least nearly.
- If nothing else the right-angled trinocular stereo vision configuration as outlined in Figure 5.11 is protected by at least U.S. patent claims¹ which motivates to investigate an alternative approach in addition.

As basis for the further investigations I propagate the following hypotheses:

1. It is possible to apply an alternative three-view camera configuration compared to Li [67] in order to calculate a full-hemispherical depth map.
2. Using only narrow-angled views can avoid subsequent merging procedures of high-computational effort.
3. The avoidance of calculating unreliable depth information near the epipoles (that is skipped during the merging process anyway) reduces the number of necessary stereo correspondence operations compared to Li [67].

The answers will be given in the following chapters. But let us start with a brief introduction of the suggested configuration:

¹ <http://www.google.de/patents/USD445818>

5.3.1 Physical Omnidirectional Camera Configuration

My proposed system comprises three omnidirectional fisheye cameras denoted as $\text{Cam}_0^{\text{omni}}$, $\text{Cam}_1^{\text{omni}}$ and $\text{Cam}_2^{\text{omni}}$. In the following, I refer to them as $\text{Cam}_i^{\text{omni}}$ with $i \in \{1, 2, 3\}$. The deployment of the cameras is organized as equilateral triangle as can be seen in Figure 5.15.

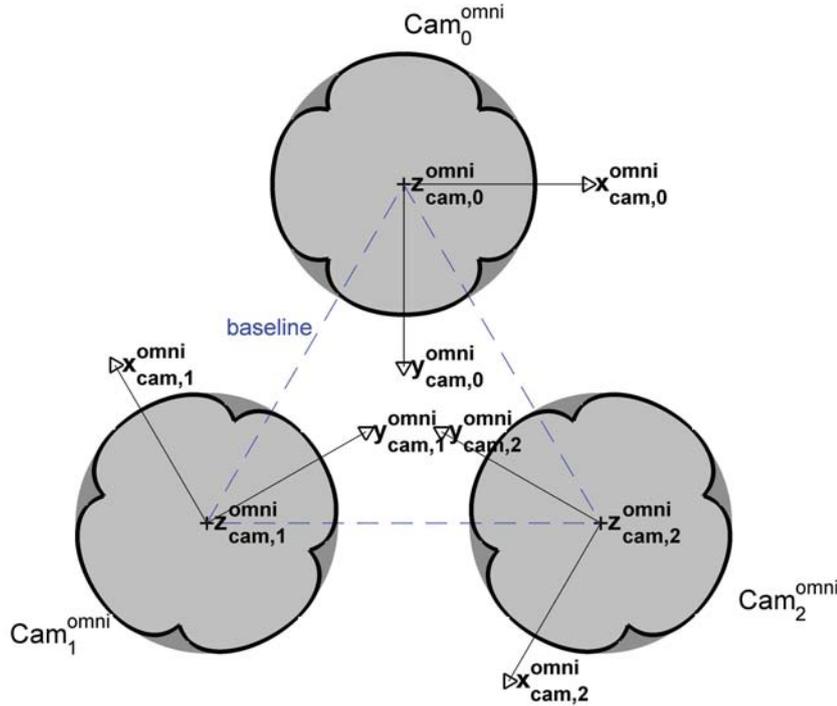


Figure 5.15: System configuration with physical omnidirectional cameras

In contrast to Li [67] the presented setup generates three (instead of two) virtual stereo cameras. This should lead to the benefit that each stereo camera has to cover merely an azimuth angle of approximately 60° . The omnidirectional images do not need to be sampled and correlated up to the singularity points as can be seen later.

5.3.2 Virtual Rectified Cameras

In the previous sections it has been shown that a rectification of the equiangular projections has to be carried out. This can be performed by introducing virtual rectified cameras.

Figure 5.16 demonstrates how the images of the cameras $\text{Cam}_0^{\text{omni}}$ and $\text{Cam}_1^{\text{omni}}$ can be rectified with respect to each other by means of two virtual cameras, denoted as $\text{Cam}_{0a}^{\text{rect}}$ and $\text{Cam}_{1a}^{\text{rect}}$. It is obvious that their rotation is different compared to the physical cameras. Indeed they have a roll of $\pm\pi/6$ around their z -axes z_{0a}^{rect} and z_{1a}^{rect} with respect to z_0^{omni} and z_1^{omni} respectively. The virtual cameras themselves can be fully described by their intrinsic parameters \mathbf{K}^{rect} and extrinsic parameters \mathbf{C}^{rect} and \mathbf{R}^{rect} [25].

$\text{Cam}_{0a}^{\text{rect}}$ and $\text{Cam}_{1a}^{\text{rect}}$ form together a virtual stereo camera denoted as $\text{SCam}_a^{\text{rect}} \in \{\text{Cam}_{0a}^{\text{rect}}, \text{Cam}_{1a}^{\text{rect}}\}$.

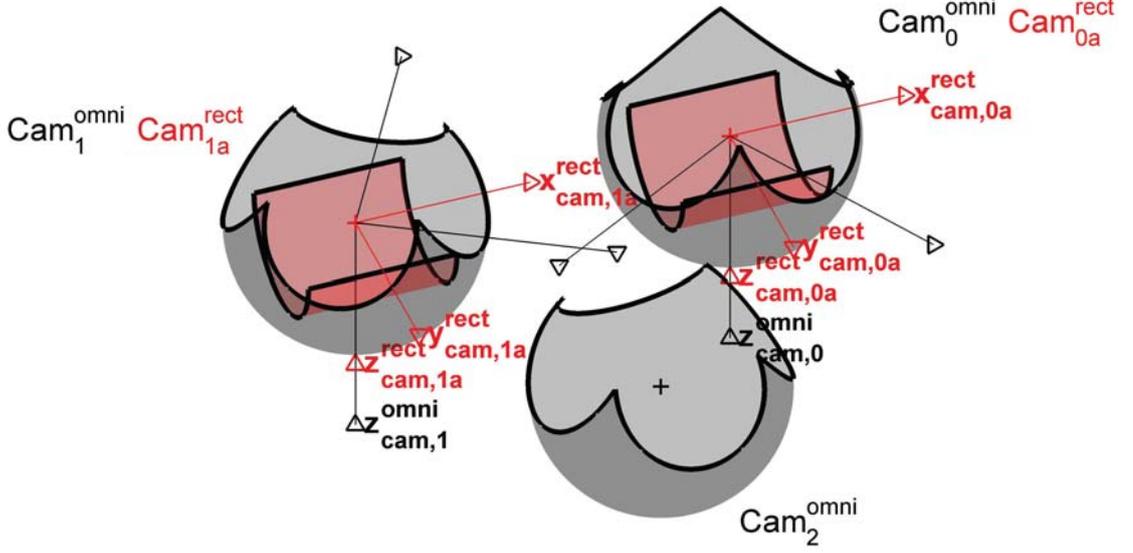


Figure 5.16: System configuration with two virtual rectified cameras

In order to exploit the geometrical setup to the whole extent, it is equipped with six virtual rectified cameras denoted as $\text{Cam}_{0a}^{\text{rect}}$, $\text{Cam}_{0b}^{\text{rect}}$, $\text{Cam}_{1a}^{\text{rect}}$, $\text{Cam}_{1c}^{\text{rect}}$, $\text{Cam}_{2b}^{\text{rect}}$ and $\text{Cam}_{2c}^{\text{rect}}$, or $\text{Cam}_j^{\text{rect}}$ with $j \in \{0a, 0b, 1a, 1c, 2b, 2c\}$ respectively. Together these virtual views generate three stereo cameras $\text{SCcam}_a^{\text{rect}}$, $\text{SCcam}_b^{\text{rect}}$, $\text{SCcam}_c^{\text{rect}}$, or $\text{SCcam}_k^{\text{rect}}$ with $k \in \{a, b, c\}$ respectively. This is summarized in Equation 5.11.

$$\begin{aligned}
 \text{SCam}_a^{\text{rect}} &= \{\text{Cam}_{0a}^{\text{rect}}, \text{Cam}_{1a}^{\text{rect}}\} \\
 \text{SCam}_b^{\text{rect}} &= \{\text{Cam}_{2b}^{\text{rect}}, \text{Cam}_{0b}^{\text{rect}}\} \\
 \text{SCam}_c^{\text{rect}} &= \{\text{Cam}_{1c}^{\text{rect}}, \text{Cam}_{2c}^{\text{rect}}\}
 \end{aligned} \tag{5.11}$$

Figure 5.17 shows the complete configuration of three physical cameras (black) together with six virtual rectified views organized in three pairs (red, green and blue).

In principle each method for rectification that has been presented in Section 5.2 could possibly be used for this configuration, because a full rectification of the spherical images is not necessary. Since the epipolar equidistant method has been successfully applied by Li [67] and the cylindrical method is predestined for stereo correspondence processing as stated by Abraham et al. [3], I will continue the feasibility study by comparing both. A real-time implementation as can be done by means of epipolar stereographic rectification is not yet the major concern of this investigation. Since the latter method shows almost identical projection behaviour compared to the epipolar equidistant principle, its explicit treatment is skipped for the time being.

In the following chapter the hereby presented configuration will be exploited in order to generate a full hemispherical depth map.

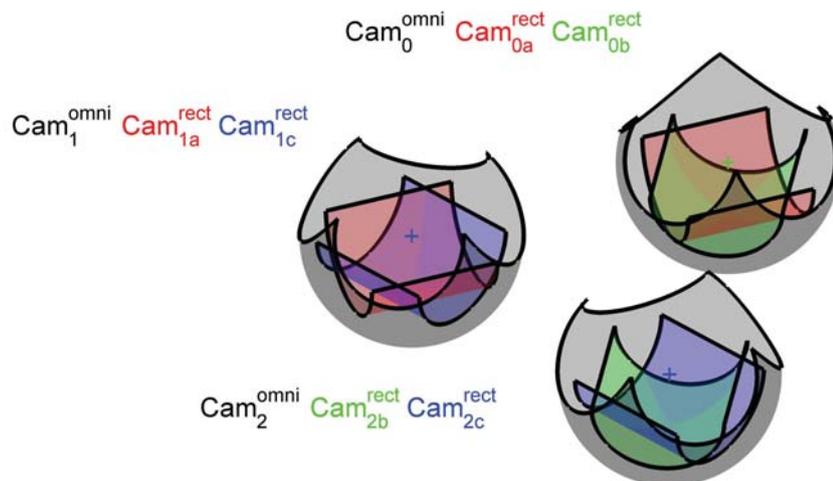


Figure 5.17: System configuration with six virtual rectified cameras

Chapter 6

A Novel Spherical Stereo Vision Algorithm

The concept of a novel three-view camera setup as presented in Section 5.3 shall be used in this chapter for the generation of a full hemispherical depth map.

It starts with introducing a MATLAB[®] simulation environment in Section 6.1. The camera deployment as well as its intrinsic parametrization is explained in Section 6.2, Section 6.3 and Section 6.4. Successively based on the simulation setup the depth map computation is performed in Section 6.5. Finally an error estimation is presented in Section 6.6.

The basic idea as presented in Chapter 6 has been published by the author in scientific publications [23,24].

6.1 Matlab Simulation Environment

For the purpose of simulation and evaluation of the novel spherical stereo vision algorithm, a reproducible data set shall be utilized. Therefore a MATLAB[®] test environment provides virtual 3D data of an exemplary domestic environment.

A single artificial room of the size of $5m \times 5m \times 2.5m$ is presented in Figure 6.1. Mathematically it forms a quantity of scattered world points $\mathbf{X}_{\text{world}}$.

The physical omnidirectional imaging devices $\text{Cam}_0^{\text{omni}}$, $\text{Cam}_1^{\text{omni}}$ and $\text{Cam}_2^{\text{omni}}$ are indicated by their coordinate systems (red, green and blue). Since $\text{Cam}_0^{\text{omni}}$ represents the reference camera in the stereo vision system with respect to the other cameras, it is located at the ceiling in the middle of the virtual room.

To form the scenario realistic and comparable, different configurations of the omnidirectional stereo setup will be simulated.

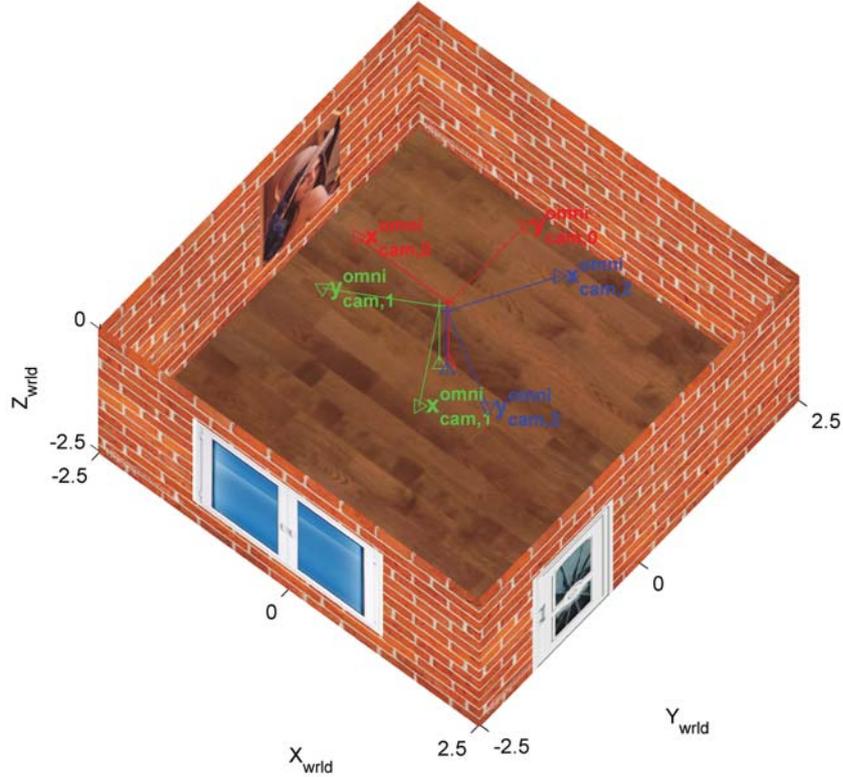


Figure 6.1: Virtual 3D test environment equipped with a spherical stereo vision sensor

6.2 Extrinsic Configuration

The general deployment of $\text{Cam}_i^{\text{omni}}$ and $\text{Cam}_j^{\text{rect}}$ has been introductorily outlined in Section 5.3. Now the cameras shall be parametrized realistically. Therefore two configurations with different baselengths will be investigated: $b = 10\text{cm}$ and $b = 15\text{cm}$.

Initially $\mathbf{C}_0^{\text{omni}}$ is determined to be the origin of the world coordinate system, i.e. we define $\mathbf{R}_0^{\text{omni}} = \mathbf{I}$ and $\mathbf{C}_0^{\text{omni}} = \mathbf{0}$. Since all physical cameras are to form an equilateral triangle and shall have a mutual differential rotation of $\pi/2$, one can calculate the camera origins $\mathbf{C}_i^{\text{omni}}$, the rotation matrices $\mathbf{R}_i^{\text{omni}}$ and consequently the translation vectors $\mathbf{T}_i^{\text{omni}}$ by following the mathematics of Section 2.2.1.3.

The virtual cameras $\text{Cam}_j^{\text{rect}}$ have the identical origins $\mathbf{C}_j^{\text{rect}}$ as their corresponding physical equivalent. However, in order to achieve the configuration of Section 5.3.2 the cameras $\text{Cam}_j^{\text{rect}}$ are rotated by $\pm\pi/6$ around their z-axis. Subsequently, they obtain new rotation matrices and $\mathbf{R}_j^{\text{rect}}$ and translation vectors $\mathbf{T}_j^{\text{rect}}$.

6.3 Physical Camera Configuration

While an ideal omnidirectional camera obeys the equiangular projection principle as presented in Section 2.2.3, real devices can better be described using the generic model of Section 2.2.4. In the following both variants shall be considered.

Further parameters will orientate towards the industry camera of Section 4.4.1 having a resolution (height \times width) of 1680×1680 with a focal length of $f = 2.7$ mm. The intrinsic parameters already presented in Table 4.2 and Table 4.3 shall be used for example.

The parameters of the equiangular model can be calculated as follows: It is intended to observe a FOV of 180° in horizontal and vertical direction exploiting the full pixel range. Hence the pixel densities can be computed as $m_x = \text{width}/(f \cdot \pi) = (2 \cdot c_x)/(f \cdot \pi)$ and $m_y = \text{height}/(f \cdot \pi) = (2 \cdot c_y)/(f \cdot \pi)$ where c_x and c_y are the horizontal and vertical offset values of the principal point from the pixel frame - cf. Section 2.2.1.2. The obtained intrinsic parameters can be seen in Table 6.1 and Table 6.2.

Table 6.1: Intrinsic parameters of the equiangular cameras $\text{Cam}_i^{\text{omni}}$

Camera	Projection Model	α_x	α_y	c_x	c_y
$\text{Cam}_i^{\text{omni}}$	Equiangular	534.76	534.76	840.50	840.50

Table 6.2: Physical pixel dimensions of the equiangular cameras $\text{Cam}_i^{\text{omni}}$

Camera	Projection Model	p_x	p_y
$\text{Cam}_i^{\text{omni}}$	Equiangular	5.05×10^{-6} [m]	5.05×10^{-6} [m]

Using the mathematical relation of Equation 2.16 described in Section 2.2.1.2 in combination with the afore-computed intrinsic values, an intrinsic matrix $\mathbf{K}_i^{\text{omni}}$ can be established for each physical camera and each configuration.

6.4 Virtual Camera Configuration

In Section 5.3 the general deployment of the virtual cameras (extrinsic parameters) has been discussed. The crucial issue is now to investigate what intrinsic parameters are appropriate for covering the whole environment. As a result of Chapter 4 a sensible resolution of a virtual target camera highly depends on the underlying physical resolution of the source camera. Furthermore in Chapter 3 the impact of the (virtual) stereo cameras resolution on the 3D measurement accuracy has been stated. A suggestion for deriving virtual camera parameters is given in the following subsections.

6.4.1 The Focal Length

The focal length f can be chosen freely and is determined to $f = 2.7$ mm in analogy to Section 6.3. Setting an identical focal length of source and target camera makes the underlying pixel sizes directly comparable since they scale proportional to f .

6.4.2 Prediscussion of the Field of View

According to Section 5.3 each virtual stereo camera $\text{SCam}_k^{\text{rect}}$ has to cover a minimum horizontal field of view FOV_h of 60° and a vertical field of view FOV_v of 180° for a hemispherical scene coverage.

All rectification methods presented in Section 5.2 are characterized by an equiangular behaviour in vertical direction. Equation 6.1 shows that FOV_v for the equiangular model can be calculated in dependence of the vertical resolution of the image and a given virtual pixel height p_y . In other words: With a given field of view a predefined pixel height is required in order to determine the vertical number of pixels (height).

$$\text{FOV}_v = \frac{2 \cdot c_y - 1}{\alpha_y} = \frac{\text{height} \cdot p_y}{f} \quad (6.1)$$

The afore-presented rectification methods of Section 5.2 show different horizontal projection behaviour: While the cylindrical method rectifies perspectively, the epipolar equidistance method performs this operation equiangular. The following equations demonstrate the calculation of FOV_h for both approaches - cylindrical (Equation 6.2) and epipolar equidistance (Equation 6.3) rectification:

$$\text{FOV}_h = 2 \cdot \arctan\left(\frac{2 \cdot c_y - 1}{2 \cdot \alpha_x}\right) = 2 \cdot \arctan\left(\frac{\text{width} \cdot p_x}{2 \cdot f}\right) \quad (6.2)$$

$$\text{FOV}_h = \frac{2 \cdot c_x - 1}{\alpha_x} = \frac{\text{width} \cdot p_x}{f} \quad (6.3)$$

Both can be computed in dependence of the horizontal resolution of the image and a given virtual pixel width p_x (cf. Equation 4.11). I.e. with a given FOV a predefined pixel width is required to determine the horizontal number of pixels.

In addition Section 3.5.2 discussed that the horizontal field of view FOV_h of the virtual stereo cameras $\text{SCam}_k^{\text{rect}}$ cannot be identical to the FOV_h of $\text{Cam}_j^{\text{rect}}$. A maximal measurable disparity d_{max} (depending on a configurable minimal measurable depth Z_{min} and the baseline b) together with the pixel length p_x reduce the actual observable horizontal area.

For the perspective view in accordance with Equation 3.16, d_{\max} can be calculated as follows:

$$d_{\max} = b \cdot \frac{f}{p_x \cdot Z_{\min}} \quad (6.4)$$

For the equiangular sampling method a maximum disparity d_{\max} can be derived from Equation 5.9:

$$d_{\max} = \arcsin \left[\frac{b}{Z_{\min}} \cdot \sin \left(\frac{\pi}{2} + \frac{x_{\text{img},1} - c_x}{\alpha_x} \right) \right] \cdot \alpha_x \quad (6.5)$$

In order to make the parametrizations for both rectification methods comparable, we claim Z_{\min} to hold true for the mid of the camera where $x_{\text{img},1} = c_x$. Here the euclidean distance equals to the y-z-component of a 3D camera point under consideration. Equation 6.5 simplifies to:

$$d_{\max} = \arcsin \left(\frac{b}{Z_{\min}} \right) \cdot \alpha_x = \arcsin \left(\frac{b}{Z_{\min}} \right) \cdot \frac{f}{p_x} \quad (6.6)$$

Considering that d_{\max} reduces the visible field of view, a reduced (symmetrical) field of view, FOV_{hr} , for $\text{SCam}_k^{\text{rect}}$ can be calculated for both approaches - cylindrical (Equation 6.7) and epipolar equidistance (Equation 6.8) rectification:

$$\text{FOV}_{\text{hr}} = 2 \cdot \arctan \left[\frac{(\text{width} - 2 \cdot d_{\max}) \cdot p_x}{2 \cdot f} \right] \quad (6.7)$$

$$\text{FOV}_{\text{hr}} = \frac{(\text{width} - 2 \cdot d_{\max}) \cdot p_x}{f} \quad (6.8)$$

Having predefined FOV_{hr} to be 60° , the remaining parameters shall be fixed by starting to observe appropriate virtual pixel sizes for different FOV_h .

6.4.3 Marginal Virtual Pixel Sizes

In Section 4.4.3 the computation of marginal virtual pixel sizes has been discussed for virtual perspective cameras. This principle is now adapted for virtual rectified views. The relationship of Equation 4.13 changes to:

$$\begin{aligned} x_{\text{img}}^{\text{omni}} &= \mathcal{F}_x \left(x_{\text{img}}^{\text{rect}}, y_{\text{img}}^{\text{rect}}, p_x^{\text{omni}}, p_y^{\text{omni}}, p_x^{\text{rect}}, p_y^{\text{rect}} \right) \\ y_{\text{img}}^{\text{omni}} &= \mathcal{F}_y \left(x_{\text{img}}^{\text{rect}}, y_{\text{img}}^{\text{rect}}, p_x^{\text{omni}}, p_y^{\text{omni}}, p_x^{\text{rect}}, p_y^{\text{rect}} \right) \end{aligned} \quad (6.9)$$

The underlying projection models of the functions \mathcal{F}_x and \mathcal{F}_y are in compliance with Section 6.3 either \mathcal{F}_e (Section 2.2.3) or \mathcal{F}_g (Section 2.2.4) for the source camera and $\mathcal{F}_{\text{rect},1}$ (Section 5.2.1) and $\mathcal{F}_{\text{rect},2}$ (Section 5.2.2) for the target camera.

Based on Section 4.4.3 we can determine the virtual pixel size according to the under-sampling as well as over-sampling strategy. Since reliable depth measurements by stereo computation base on virtual pixel information that is substantiated by underlying physical pixels appropriately, only the under-sampling method shall be further considered. The actual computation can now be performed by evaluating the following mathematical expression:

$$\text{minimize} (|\nabla \mathcal{F}_x| - 1.0| + |\nabla \mathcal{F}_y| - 1.0|) \quad (6.10)$$

This evaluation has been performed for all projection models under consideration interchangeably. A virtual camera which is fed by the physical one is simulated with its rotated orientation (more precisely a roll of $\pm\pi/6$ around its z-axis as shown in Section 5.3.2), a FOV_v of 180° and a varying FOV_h between 60° and 90° :

- $\text{FOV}_h = 60^\circ$: This is the theoretical minimum in the presented approach which has to be covered by each virtual stereo camera $\text{SCam}_k^{\text{rect}}$ and hence by each virtual camera $\text{Cam}_j^{\text{rect}}$.
- $\text{FOV}_h = 90^\circ$: This is the theoretical minimum in the approach presented by Li [67] which has to be covered by each of both stereo cameras.

In order to solve this problem numerically for each FOV_h between 60° and 90° , a range of pixel size ratios $p_x^{\text{rect}}/p_x^{\text{omni}}$ and $p_y^{\text{rect}}/p_y^{\text{omni}}$ has been investigated in order to feed Equation 6.10. The distinct value range originates from results of Chapter 4 and is outlined in Equation 6.11:

$$\begin{aligned} p_x^{\text{rect}}/p_x^{\text{omni}} &\in [0.25, 2.50] \\ p_y^{\text{rect}}/p_y^{\text{omni}} &\in [0.25, 2.50] \end{aligned} \quad (6.11)$$

The pixel coordinates $x_{\text{img}}^{\text{rect}}$ and $y_{\text{img}}^{\text{rect}}$, which are further parameters for Equation 6.10 are calculated by Equation 6.1, Equation 6.2 and Equation 6.3:

$$\begin{aligned} x_{\text{img}}^{\text{rect}} &\in [0, \text{width} - 1] \\ y_{\text{img}}^{\text{rect}} &\in [0, \text{height} - 1] \end{aligned} \quad (6.12)$$

In accordance with Figure 4.17 of Section 4.4.3 one can observe overlapping contour plots of $|\Delta x_{\text{img}}^{\text{omni}}|$ and $|\Delta y_{\text{img}}^{\text{omni}}|$ for two *exemplary* configurations in Figure 6.2. It can be observed that the optimal pixel ratios indicated by red dots differ for each configuration. One can state that the more the horizontal field of view increases the more the obtained resolution for the virtual (cylindrical) camera decreases in horizontal as well as vertical direction. In other words: The size of the virtual pixels p_x^{rect} and p_y^{rect} grows with an increasing horizontal field of view.

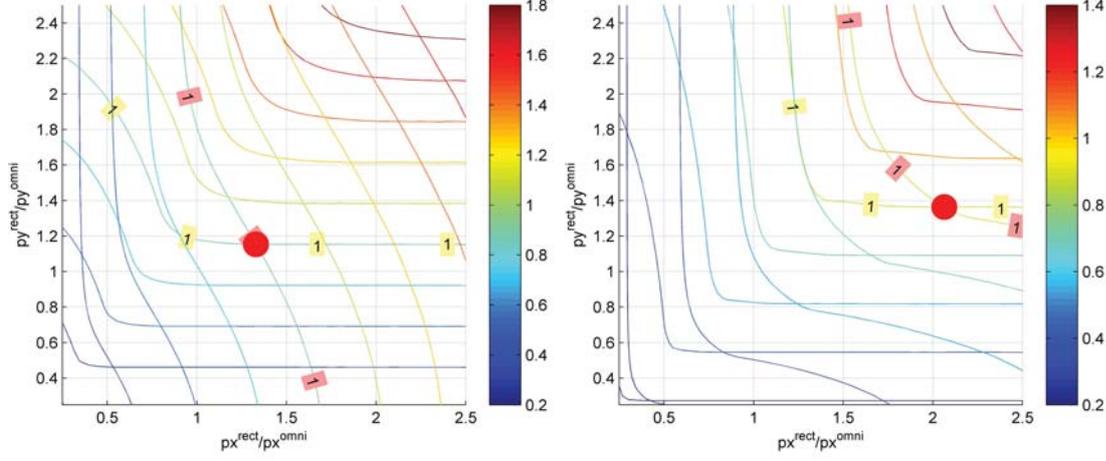


Figure 6.2: Overlapping contour plots of $|\Delta x_{\text{img}}^{\text{omni}}|$ and $|\Delta y_{\text{img}}^{\text{omni}}|$ for different ratios of pixel sizes $p_x^{\text{rect}}/p_x^{\text{omni}}$ and $p_y^{\text{rect}}/p_y^{\text{omni}}$, an equiangular projection model for the source camera, a cylindrical projection model for the target camera, an under-sampling strategy and $\text{FOV}_h = 60^\circ$ (left) and $\text{FOV}_h = 90^\circ$ (right)

Figure 6.3 demonstrates how the sampling interval of the source images x-direction $|\Delta x_{\text{img}}^{\text{omni}}|$ (left plot) and the source images y-direction $|\Delta y_{\text{img}}^{\text{omni}}|$ (right plot) deploys over the target image for the exemplary configuration of $\text{FOV}_h = 60^\circ$. The outer dark blue coded areas specify the actual virtual pixel size since its sampling interval coincides with the physical one. The red areas indicate high under-sampling.

Remark: Figure 6.3 visualizes the roll of $+\pi/6$ of the target camera in comparison to the source camera. For the alternative roll of $-\pi/6$, Equation 6.10 shows symmetrical behaviour.

In Figure 6.4 the optimal pixel ratios for each configuration can be observed depending on the different FOV_h . The optimal virtual pixel size becomes a function f of the physical pixel size and the field of view:

$$\begin{aligned} p_x^{\text{rect}} &= f(p_x^{\text{omni}}, \text{FOV}_h) \\ p_y^{\text{rect}} &= f(p_y^{\text{omni}}, \text{FOV}_h) \end{aligned} \quad (6.13)$$

Equally to Section 4.4 one can see that a source camera following the generic projection model (blue and cyan graphs) causes pixel ratios that are below one - in comparison to the equiangular model (red and green graphs). It is interesting to discover that the optimal pixel ratios $p_x^{\text{rect}}/p_x^{\text{omni}}$ in horizontal direction strongly grow for the cylindrical model (red and blue graphs, left chart) with the increase of FOV_h while $p_x^{\text{rect}}/p_x^{\text{omni}}$ only slightly change for the epipolar equidistant projection principle. However, for the vertical direction $p_y^{\text{rect}}/p_y^{\text{omni}}$ both rectification models behave identically.

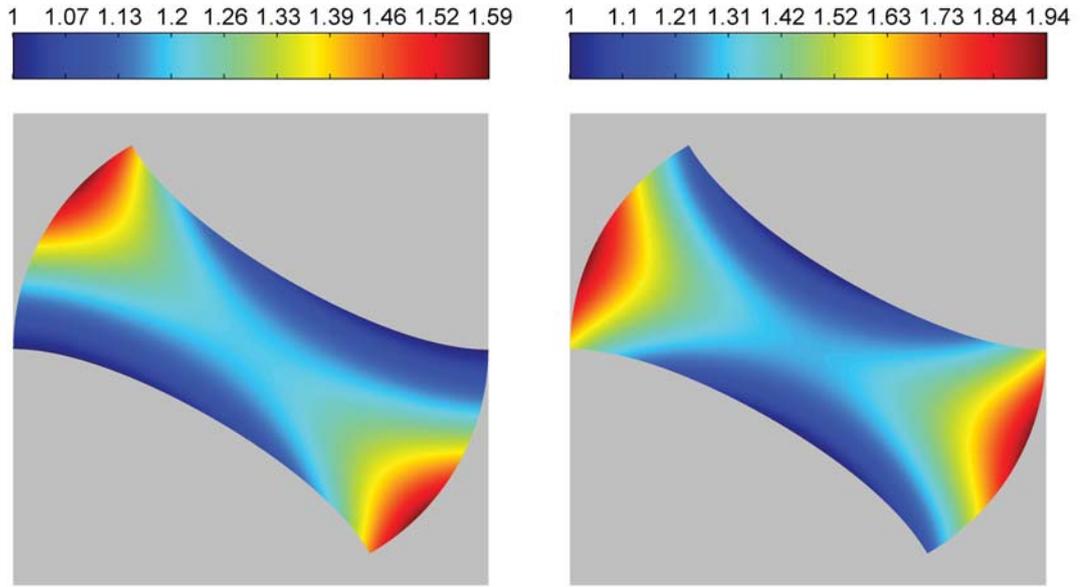


Figure 6.3: Colour coded sampling interval $|\Delta x_{\text{img}}^{\text{omni}}|$ (left) and $|\Delta y_{\text{img}}^{\text{omni}}|$ (right) for the optimal pixel ratios $p_x^{\text{cyl}}/p_x^{\text{omni}}$ and $p_y^{\text{cyl}}/p_y^{\text{omni}}$, an equiangular projection model for the source camera, a cylindrical projection model for the target camera, an under-sampling strategy and $\text{FOV}_h = 60^\circ$

6.4.4 Calculation of the Field of View

Combining Equation 6.3 and Equation 6.4 with Equation 6.13 the disparity range d_{max} can be calculated for both target models depending on different FOV_h and the underlying physical pixel size p_x^{omni} - parametrized by a predefined minimal measurable depth Z_{min} . Assuming the sensor to be mounted $2.50m$ over ground and a person to be of $1.75m$ height, a suitable *exemplary* Z_{min} shall be defined to be $0.75m$. Equation 6.14 and Equation 6.15 outline the calculation of d_{max} for cylindrical (Equation 6.14) and epipolar equidistance (Equation 6.15) projection:

$$d_{\text{max}} = b \cdot \frac{f}{p_x^{\text{rect}} \cdot Z_{\text{min}}} = b \cdot \frac{f}{f(p_x^{\text{omni}}, \text{FOV}_h) \cdot Z_{\text{min}}} \quad (6.14)$$

$$d_{\text{max}} = \arcsin\left(\frac{b}{Z_{\text{min}}}\right) \cdot \frac{f}{p_x^{\text{rect}}} = \arcsin\left(\frac{b}{Z_{\text{min}}}\right) \cdot \frac{f}{f(p_x^{\text{omni}}, \text{FOV}_h)} \quad (6.15)$$

Figure 6.5 presents the simulation results for the afore-discussed configurations under consideration.

Figure 6.5 demonstrates the impact of FOV_h on d_{max} . For the cylindrical rectification method the disparity range rapidly diminishes because the underlying pixel size increases significantly - as shown in Figure 6.4. For the epipolar equidistant approach, d_{max} changes only slightly.

With a given maximum disparity range d_{\max} and a given virtual pixel size p_x^{rect} , both depending on FOV_h , the reduced field of view FOV_{hr} can be computed depending on FOV_h as well, according to Equation 6.5 and Equation 6.7. The results are presented in Figure 6.6.

It has turned out that the computation of the reduced field of field FOV_{hr} that already includes the stereo disparity range pixel trade-off based on FOV_h only depends on the baselength b and the target rectification model. For a given minimum $\text{FOV}_{\text{hr}} = 60^\circ$ the minimum FOV_h can be easily extracted from Figure 6.6 and is summarized in Table 6.3.

Table 6.3: Calculated fields of view FOV_h for the virtual cameras $\text{Cam}_j^{\text{rect}}$ according to the variety of configurations

No.	$\text{Cam}_i^{\text{omni}}$	$\text{Cam}_j^{\text{rect}}$	b [m]	FOV_{hr} [deg]	Z_{\min} [m]	FOV_h [deg]
1	equiangular	cylindrical	0.10	60.0	0.75	70.90
2	equiangular	cylindrical	0.15	60.0	0.75	75.80
3	kannala9	cylindrical	0.10	60.0	0.75	70.90
4	kannala9	cylindrical	0.15	60.0	0.75	75.80
5	equiangular	epipolar equidistance	0.10	60.0	0.75	75.40
6	equiangular	epipolar equidistance	0.15	60.0	0.75	83.20
7	kannala9	epipolar equidistance	0.10	60.0	0.75	75.40
8	kannala9	epipolar equidistance	0.15	60.0	0.75	83.20

6.4.5 Calculation of the Virtual Pixel Size Ratios

Having determined FOV_h for each configuration, the appropriate virtual pixel size ratios can be extracted from Figure 6.4. The virtual image sizes are calculated using Equation 6.1, Equation 6.2 and Equation 6.3. Table 6.4 summarizes the results.

Table 6.4: Calculated virtual pixel size ratios and image sizes

No.	FOV_h [deg]	$p_x^{\text{rect}}/p_x^{\text{omni}}$	$p_y^{\text{rect}}/p_y^{\text{omni}}$	height	width	d_{\max}
1	70.90	1.5060	1.2195	1377	505	48
2	75.80	1.6120	1.2526	1341	516	67
3	70.90	0.6223	0.4908	1283	458	43
4	75.80	0.6632	0.5024	1254	470	61
5	75.40	1.0064	1.2498	1344	699	72
6	83.20	1.0176	1.3065	1285	763	106
7	75.40	0.4140	0.5022	1254	636	65
8	83.20	0.4140	0.5246	1201	702	98

6.4.6 Results of the Virtual Camera Parameters

As shown in Section 2.2.1.2 the given information can now be employed to obtain the intrinsic parameters by applying the fundamentals presented in Section 2.2.1. The results are outlined in Table 6.5 and Table 6.6. Using the relations described in Section 2.2.1.2

Table 6.5: Intrinsic parameters of the virtual cameras $\text{Cam}_j^{\text{rect}}$

No.	Camera	α_x	α_y	c_x	c_y
1	$\text{Cam}_j^{\text{rect}}$	355.09	438.51	253.00	689.00
2	$\text{Cam}_j^{\text{rect}}$	331.74	426.92	258.50	671.00
3	$\text{Cam}_j^{\text{rect}}$	321.80	408.67	229.50	642.00
4	$\text{Cam}_j^{\text{rect}}$	301.95	399.23	235.50	627.50
5	$\text{Cam}_j^{\text{rect}}$	531.36	427.88	350.00	672.50
6	$\text{Cam}_j^{\text{rect}}$	525.51	409.31	382.00	643.00
7	$\text{Cam}_j^{\text{rect}}$	483.71	399.39	318.50	627.50
8	$\text{Cam}_j^{\text{rect}}$	483.71	382.34	351.50	601.00

and the afore-computed intrinsic values, an intrinsic matrix $\mathbf{K}_j^{\text{rect}}$ can be established for each virtual camera and each configuration.

Table 6.6: Pixel dimensions of the virtual cameras $\text{Cam}_j^{\text{rect}}$

No.	Camera	$p_x[m]$	$p_y[m]$
1	$\text{Cam}_j^{\text{rect}}$	7.60×10^{-6}	6.16×10^{-6}
2	$\text{Cam}_j^{\text{rect}}$	8.14×10^{-6}	6.32×10^{-6}
3	$\text{Cam}_j^{\text{rect}}$	8.39×10^{-6}	6.61×10^{-6}
4	$\text{Cam}_j^{\text{rect}}$	8.94×10^{-6}	6.76×10^{-6}
5	$\text{Cam}_j^{\text{rect}}$	5.08×10^{-6}	6.31×10^{-6}
6	$\text{Cam}_j^{\text{rect}}$	5.14×10^{-6}	6.60×10^{-6}
7	$\text{Cam}_j^{\text{rect}}$	5.58×10^{-6}	6.76×10^{-6}
8	$\text{Cam}_j^{\text{rect}}$	5.58×10^{-6}	7.06×10^{-6}

From Table 6.6 it becomes visible that the epipolar equidistant method produces higher resolute target images compared to a cylindrical projection, which allows more accurate depth measurements. Since this approach clearly outperforms the cylindrical rectification it will be the rectification method of choice for further investigations.

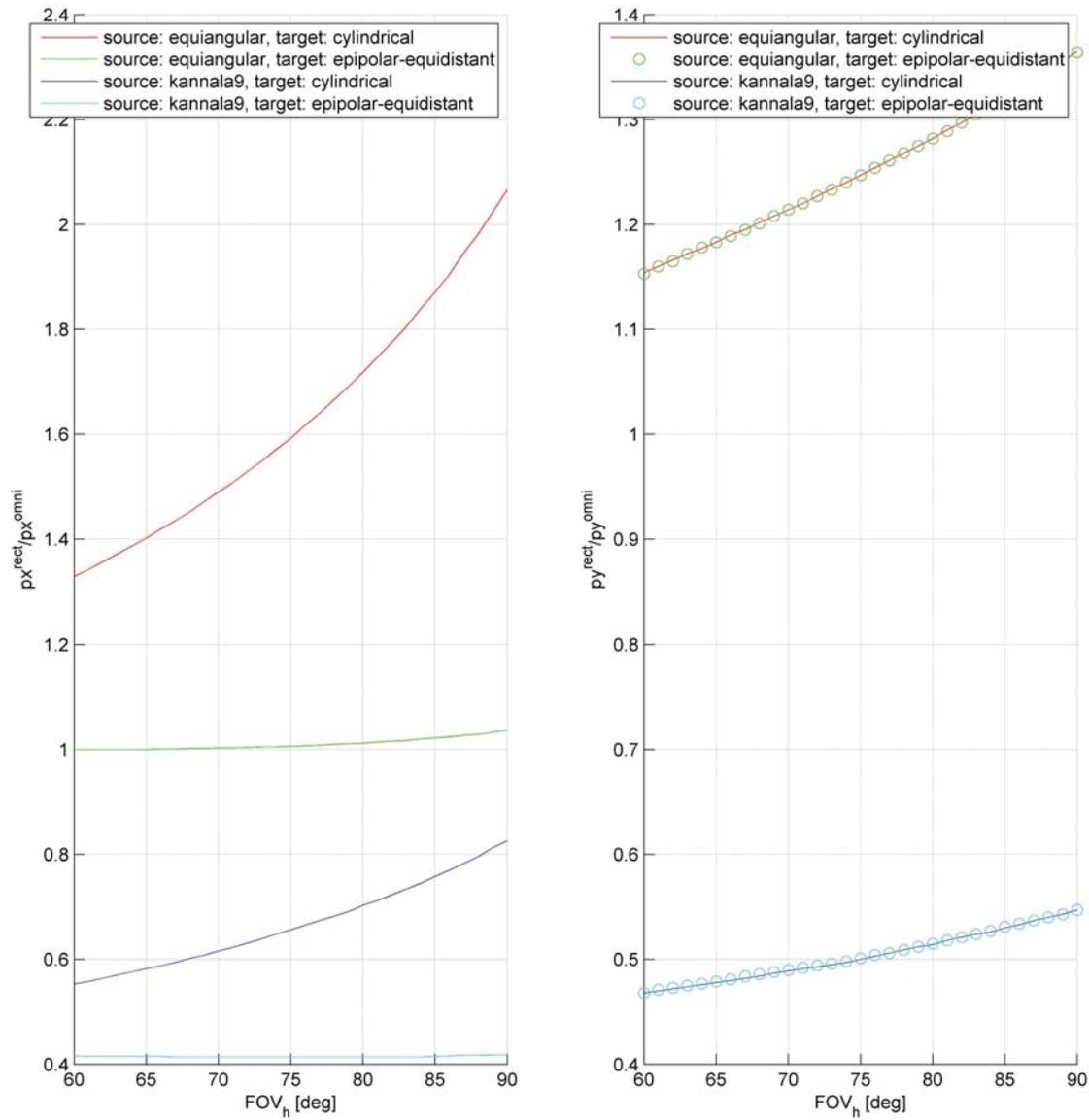


Figure 6.4: Calculated optimal pixel ratios $p_x^{\text{rect}}/p_x^{\text{omni}}$ and $p_y^{\text{rect}}/p_y^{\text{omni}}$ of physical and virtual pixel dimensions based on different horizontal fields of view FOV_h

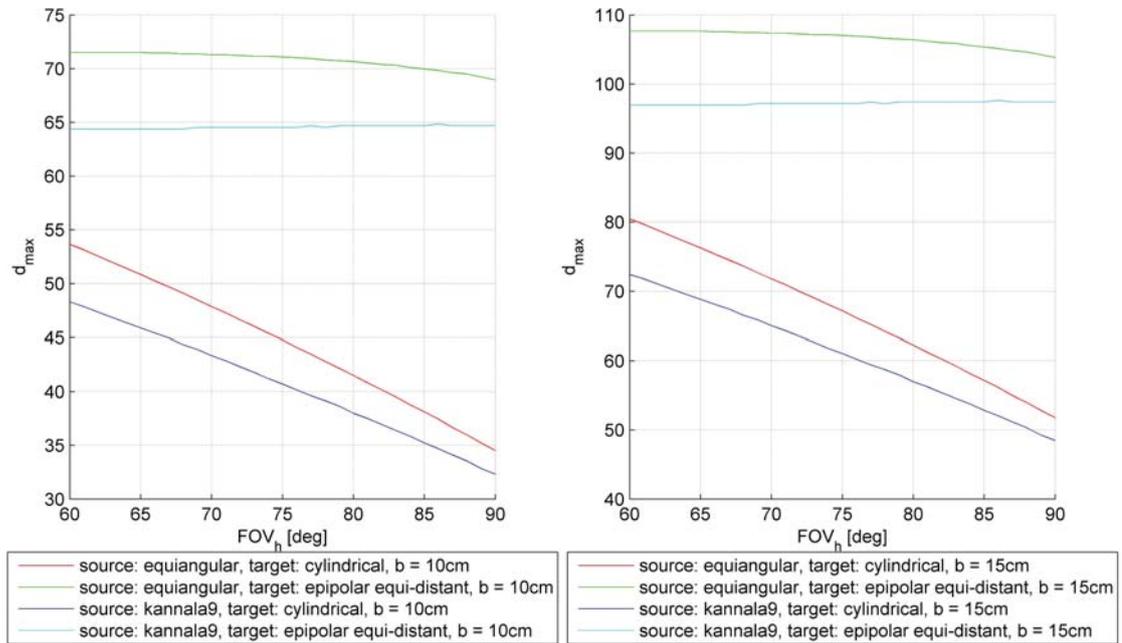


Figure 6.5: Calculated disparity ranges based on different horizontal fields of view

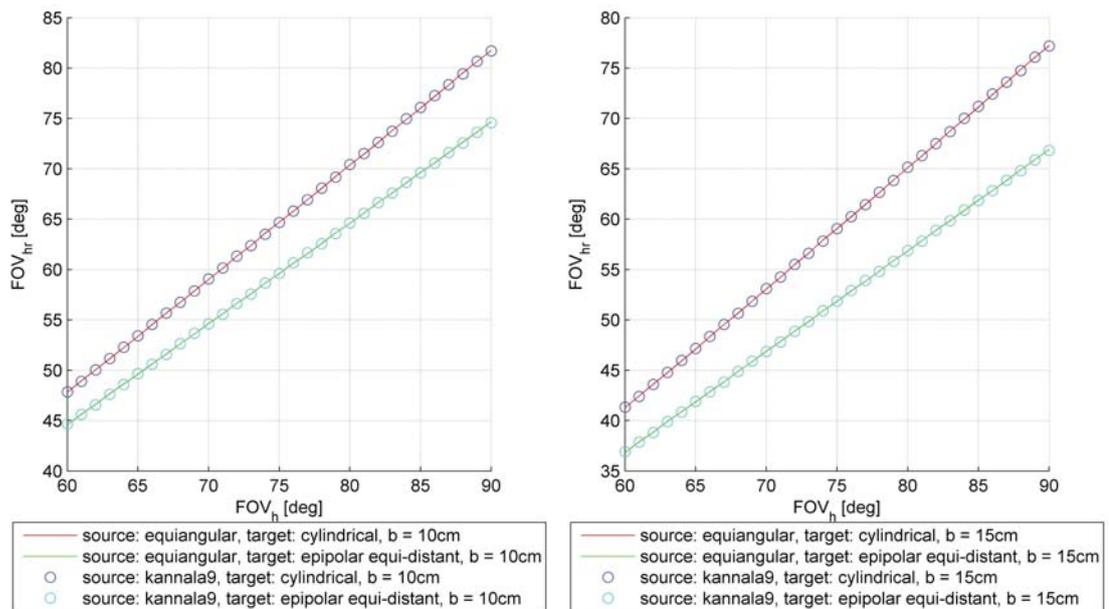


Figure 6.6: Calculated reduced horizontal field of view based on different horizontal fields of view

6.5 Spherical Depth Map Generation

The experimental setup from Section 6.1 together with the parametrizations of Section 6.5 and Section 6.6 are now employed to generate a spherical depth map Z_0^{omni} that refers to $\text{Cam}_0^{\text{omni}}$. The precondition for this simulation is the consideration of the scene itself as a quantity of scattered 3D world points X_{world} . The complete algorithm is summarized in Figure 6.7.

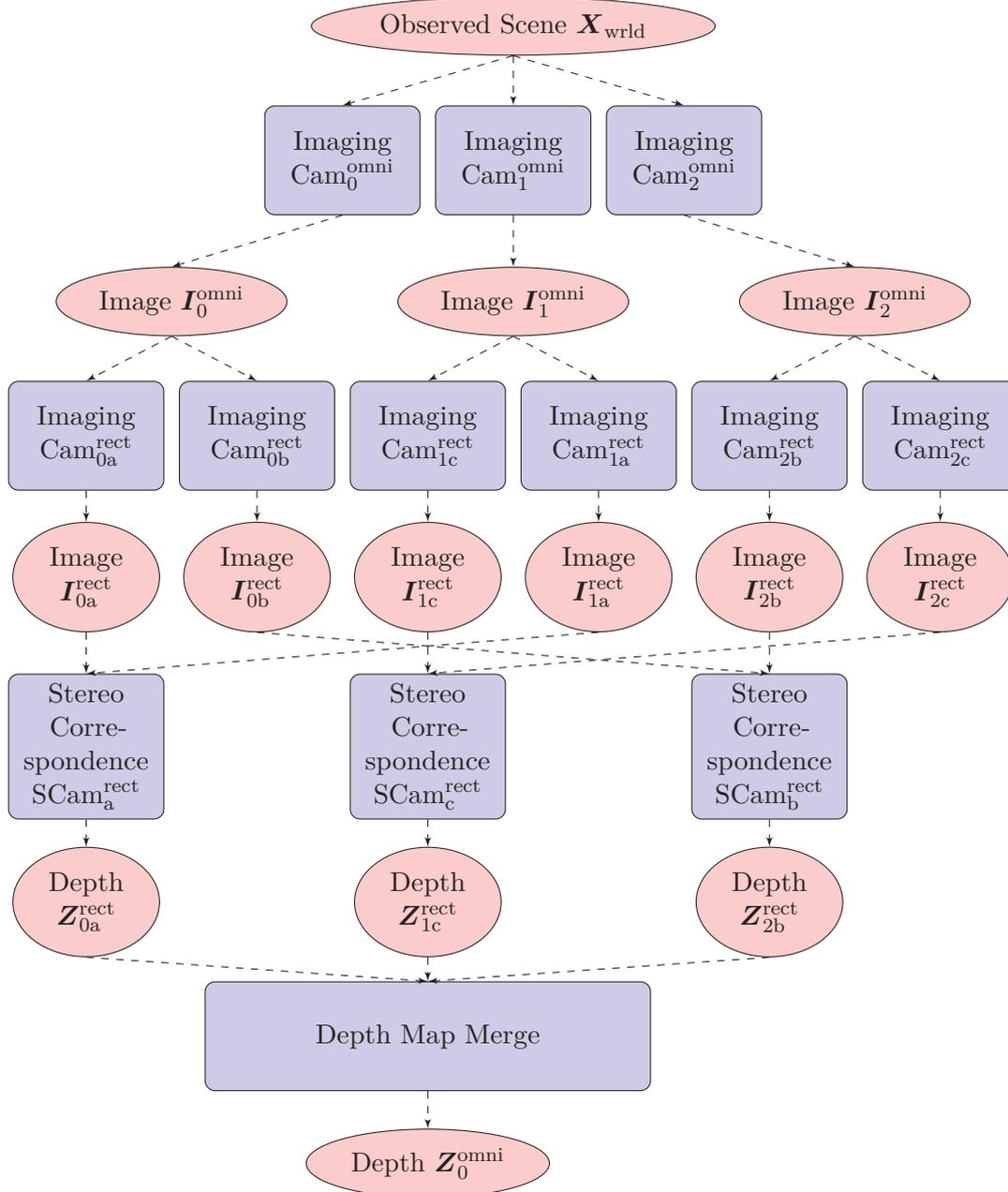


Figure 6.7: Summary of the spherical depth map generation algorithm

The steps to follow investigate each stage of the procedure illustrated in Figure 6.7. Since the approach of equiangular projection together with epipolar equidistant rectification (configuration no. 5 and no. 6, cf. Table 6.3, Table 6.4, Table 6.5 and Table 6.6) showed the best performance in terms of resolution, it will be the primary algorithm of choice for this explanation. For the accompanying plots, configuration no. 6 has been elected. However, the algorithm works with all the afore-mentioned configurations.

6.5.1 Omnidirectional Imaging Process

In this step the physical imaging process of the omnidirectional camera devices is simulated. Since the deployment of the scene points is not analytically described, the projection is done by forward mapping - cf. Section 4.2.1. The observed environment is projected to $\text{Cam}_0^{\text{omni}}$, $\text{Cam}_1^{\text{omni}}$ and $\text{Cam}_2^{\text{omni}}$ using \mathcal{F}_e or alternatively \mathcal{F}_g .

Applying the equiangular camera model we can utilize the projection function \mathcal{F}_e from Section 2.2.3 for processing:

$$\hat{\mathbf{X}}_{\text{img},i}^{\text{omni}} = \mathbf{K}_i^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_i^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{world}} + \mathbf{T}_i^{\text{omni}} \right) \quad (6.16)$$

The result is real number image points $\hat{\mathbf{X}}_{\text{img},0}^{\text{omni}}$, $\hat{\mathbf{X}}_{\text{img},1}^{\text{omni}}$ and $\hat{\mathbf{X}}_{\text{img},2}^{\text{omni}}$. In order to obtain pixel intensities $\mathbf{I}_0^{\text{omni}}$, $\mathbf{I}_1^{\text{omni}}$ and $\mathbf{I}_2^{\text{omni}}$ of the integer coordinates $\mathbf{X}_{\text{img},0}^{\text{omni}}$, $\mathbf{X}_{\text{img},1}^{\text{omni}}$ and $\mathbf{X}_{\text{img},2}^{\text{omni}}$, a standard image interpolation procedure (e.g. bilinear or bi-cubic) is employed. The final omnidirectional images can be viewed in Figure 6.8.

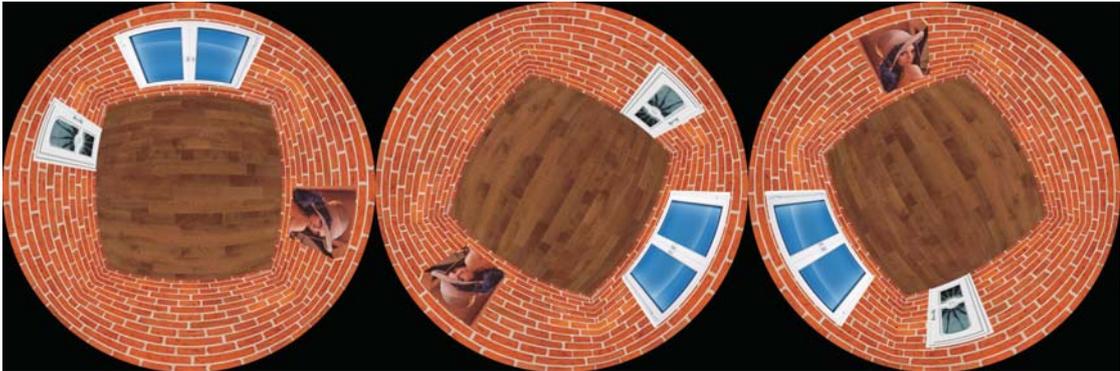


Figure 6.8: Result of the equiangular imaging process ($\text{Cam}_0^{\text{omni}}$, $\text{Cam}_1^{\text{omni}}$ and $\text{Cam}_2^{\text{omni}}$)

6.5.2 Rectification Process

At this stage the virtual rectified images are generated. This is done by backward mapping - cf. Section 4.2.2 and Section 4.2.3.

Equivalently to Section 4.2.2, the target image mesh of pixel positions $\mathbf{X}_{\text{img},j}^{\text{rect}}$ is set up and back-projected to its appropriate world points $\mathbf{X}_{\text{wld},j}^{\text{rect}}$ with respect to its dedicated omnidirectional camera using $\mathcal{F}_{\text{rect},1}$ (or alternatively $\mathcal{F}_{\text{rect},2}$) inversely:

$$\tilde{\mathbf{X}}_{\text{cam},j}^{\text{rect}}(\lambda) = \lambda \cdot \mathbf{R}_j^{\text{rect}} \cdot \mathcal{F}_{\text{rect},1}^{-1} \left(\mathbf{K}_j^{\text{rect}-1} \cdot \mathbf{X}_{\text{img},j}^{\text{rect}} \right) \quad (6.17)$$

In accordance to Section 4.2.2, the arbitrary scale factor λ expresses the one dimensional space of points on the ray. To overcome the uncertainty of distance from the camera center, we define $\|\tilde{\mathbf{X}}_{\text{wld},j}^{\text{rect}}\|_2 = 1$.

In the following, those points are projected to their attached omnidirectional camera. Applying the equiangular camera model we can utilize the projection function \mathcal{F}_e :

$$\begin{aligned} \hat{\mathbf{X}}_{\text{img},0a}^{\text{omni}} &= \mathbf{K}_0^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_0^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{cam},0a}^{\text{rect}} \right) \\ \hat{\mathbf{X}}_{\text{img},0b}^{\text{omni}} &= \mathbf{K}_0^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_0^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{cam},0b}^{\text{rect}} \right) \\ \hat{\mathbf{X}}_{\text{img},1a}^{\text{omni}} &= \mathbf{K}_1^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_1^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{cam},1a}^{\text{rect}} \right) \\ \hat{\mathbf{X}}_{\text{img},1c}^{\text{omni}} &= \mathbf{K}_1^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_1^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{cam},1c}^{\text{rect}} \right) \\ \hat{\mathbf{X}}_{\text{img},2b}^{\text{omni}} &= \mathbf{K}_2^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_2^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{cam},2b}^{\text{rect}} \right) \\ \hat{\mathbf{X}}_{\text{img},2c}^{\text{omni}} &= \mathbf{K}_2^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{R}_2^{\text{omni}} \cdot \tilde{\mathbf{X}}_{\text{cam},2c}^{\text{rect}} \right) \end{aligned} \quad (6.18)$$

The result is real-numbered image points $\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}}$. In order to obtain pixel intensities $\mathbf{I}_j^{\text{rect}}$, a standard image interpolation procedure (e.g. bilinear or bicubic) is used. The final rectified images can be seen in Figure 6.9.



Figure 6.9: Result of the epipolar equidistant imaging process ($\text{Cam}_{0a}^{\text{rect}}$, $\text{Cam}_{1a}^{\text{rect}}$, $\text{Cam}_{2b}^{\text{rect}}$, $\text{Cam}_{0b}^{\text{rect}}$, $\text{Cam}_{1c}^{\text{rect}}$, $\text{Cam}_{2c}^{\text{rect}}$)

It can be seen that either of the two images are rectified with respect to each other and therefore, since they have different positions $\mathbf{C}_j^{\text{rect}}$, form rectified stereo vision pairs $\text{SCam}_k^{\text{rect}}$. This shows the effect described in Section 5.3.

6.5.3 Rectified Depth Map Generation

The computation of pixel disparities d as described in Section 3.4.1 can be done for each $\text{SCam}_k^{\text{rect}}$ separately. Three disparity maps, denoted by D_{0a}^{rect} , D_{2b}^{rect} and D_{1c}^{rect} , can be calculated.

For processing the correspondence operation \mathcal{D} (cf. Section 3.4.2), a *Local Algorithm*¹ based on correlation windows is employed:

$$\begin{aligned} D_{0a}^{\text{rect}} &= \mathcal{D}\left(\mathbf{X}_{\text{img},0a}^{\text{rect}}, \mathbf{X}_{\text{img},1a}^{\text{rect}}\right) \\ D_{2b}^{\text{rect}} &= \mathcal{D}\left(\mathbf{X}_{\text{img},2b}^{\text{rect}}, \mathbf{X}_{\text{img},0b}^{\text{rect}}\right) \\ D_{1c}^{\text{rect}} &= \mathcal{D}\left(\mathbf{X}_{\text{img},1c}^{\text{rect}}, \mathbf{X}_{\text{img},2c}^{\text{rect}}\right) \end{aligned} \quad (6.19)$$

The result of the simulation can be seen in Figure 6.10. The presented maps show the horizontal pixel offset.

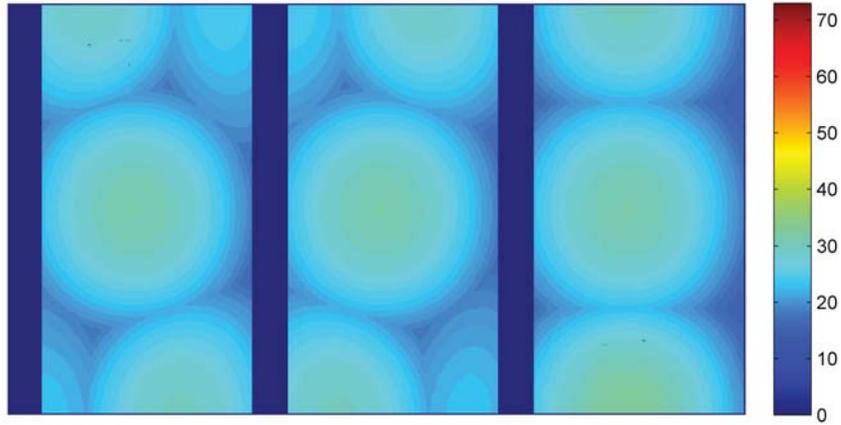


Figure 6.10: Result of the stereo correspondence operation (D_{0a}^{rect} , D_{2b}^{rect} and D_{1c}^{rect}) for the epipolar-equidistant rectified images

The actual depth maps called Z_{0a}^{rect} , Z_{2b}^{rect} and Z_{1c}^{rect} can be computed by triangulation as presented in Section 5.2.2 (epipolar equidistant rectification) and Section 3.5 (cylindrical rectification).

For the epipolar equidistant maps, the mathematical relationship of Equation 5.9 changes to:

$$Z_{0a}^{\text{rect}} = \frac{b \cdot \sin \phi_l}{\sin \frac{D_{0a}^{\text{rect}} \cdot p_x^{\text{rect}}}{f}} \quad Z_{1b}^{\text{rect}} = \frac{b \cdot \sin \phi_l}{\sin \frac{D_{1b}^{\text{rect}} \cdot p_x^{\text{rect}}}{f}} \quad Z_{2c}^{\text{rect}} = \frac{b \cdot \sin \phi_l}{\sin \frac{D_{2c}^{\text{rect}} \cdot p_x^{\text{rect}}}{f}} \quad (6.20)$$

¹ <http://de.mathworks.com/help/vision/ref/disparity.html>

with ϕ_l being the two-dimensional matrix (according to the depth map size) of azimuth angles ψ on the great circle with respect to the dedicated rectified image column $x_{\text{img},l}$ - cf. Section 5.2. Note that the depth values Z_{0a}^{rect} , Z_{2b}^{rect} and Z_{1c}^{rect} represent the euclidean norm of a 3D point given in camera coordinates. Figure 6.11 presents the obtained distance maps information for the current configuration.

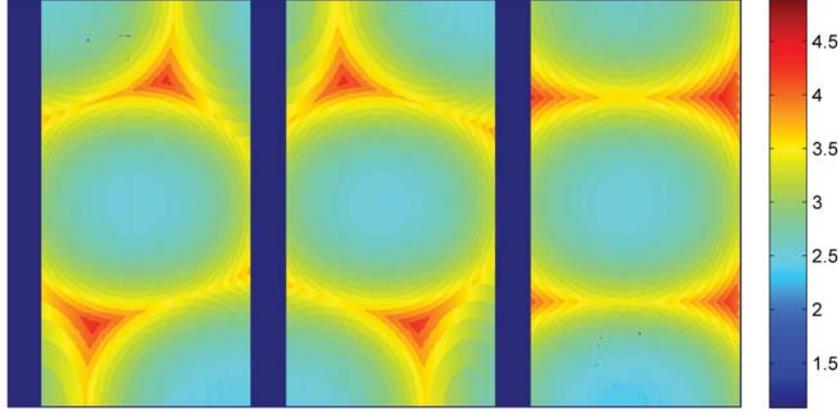


Figure 6.11: Result of the triangulation operation (Z_{0a}^{rect} , Z_{2b}^{rect} and Z_{1c}^{rect}) for the epipolar equidistant disparity maps

Alternatively the re-projection can be done for cylindrical disparity maps according to Section 3.5, Equation 3.16:

$$Z_{0a}^{\text{rect}} = b \cdot \frac{f}{D_{0a}^{\text{rect}} \cdot p_x^{\text{rect}}} \quad Z_{2b}^{\text{rect}} = b \cdot \frac{f}{D_{2b}^{\text{rect}} \cdot p_x^{\text{rect}}} \quad Z_{1c}^{\text{rect}} = b \cdot \frac{f}{D_{1c}^{\text{rect}} \cdot p_x^{\text{rect}}} \quad (6.21)$$

Note that the depth values Z_{0a}^{rect} , Z_{2b}^{rect} and Z_{1c}^{rect} for the latter case represent the norm of the y-z-component of a 3D point given in camera coordinates.

6.5.4 Spherical Depth Map Generation

The overall goal is to obtain a hemispherical depth map Z_0^{omni} with respect to the image I_0^{omni} of $\text{Cam}_0^{\text{omni}}$.

Equivalently to Section 6.5.2, the rectified image mesh of pixel positions $\mathbf{X}_{\text{img},j}^{\text{rect}}$ is set up and back-projected to its appropriate camera points $\mathbf{X}_{\text{cam},j}^{\text{rect}}$ using $\mathcal{F}_{\text{rect},1}$ (or alternatively $\mathcal{F}_{\text{rect},2}$) inversely as described by Equation 6.22.

$$\mathbf{X}_{\text{cam},j}^{\text{rect}}(\lambda) = \lambda \cdot \mathcal{F}_{\text{rect},1}^{-1} \left(\mathbf{K}_j^{\text{rect}-1} \cdot \mathbf{X}_{\text{img},j}^{\text{rect}} \right) \quad (6.22)$$

Subsequently the computed camera rays are normalized according to its depth values. For the epipolar equidistant depth maps this can be done as follows:

$$\mathbf{X}_{\text{cam},j}^{\text{rect}} = \left[\frac{x_{\text{cam},j}^{\text{rect}}}{\lambda} \cdot Z_j^{\text{rect}} \quad \frac{y_{\text{cam},j}^{\text{rect}}}{\lambda} \cdot Z_j^{\text{rect}} \quad \frac{z_{\text{cam},j}^{\text{rect}}}{\lambda} \cdot Z_j^{\text{rect}} \quad 1 \right]^T \quad (6.23)$$

Now the points are expressed with respect to the camera coordinate system of $\text{Cam}_0^{\text{omni}}$:

$$\mathbf{X}_{\text{cam},j}^{\text{omni}} = \mathbf{H}_j^{\text{rect}^{-1}} \cdot \mathbf{X}_{\text{cam},j}^{\text{rect}} \quad (6.24)$$

In the following those points are projected to $\text{Cam}_0^{\text{omni}}$ using \mathcal{F}_e (or alternatively \mathcal{F}_g):

$$\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}} = \mathbf{K}_0^{\text{omni}} \cdot \mathcal{F}_e \left(\mathbf{X}_{\text{cam},j}^{\text{omni}} \right) \quad (6.25)$$

The result is real number image points $\hat{\mathbf{X}}_{\text{img},0a}^{\text{omni}}$, $\hat{\mathbf{X}}_{\text{img},2b}^{\text{omni}}$ and $\hat{\mathbf{X}}_{\text{img},1c}^{\text{omni}}$ with pixel intensities $\hat{\mathbf{I}}_{0a}^{\text{omni}}$, $\hat{\mathbf{I}}_{2b}^{\text{omni}}$ and $\hat{\mathbf{I}}_{1c}^{\text{omni}}$.

The appropriate depth maps $\hat{\mathbf{Z}}_{0a}^{\text{omni}}$, $\hat{\mathbf{Z}}_{2b}^{\text{omni}}$ and $\hat{\mathbf{Z}}_{1c}^{\text{omni}}$ are determined by computing the euclidean norm of $\tilde{\mathbf{X}}_{\text{cam},0a}^{\text{omni}}$, $\tilde{\mathbf{X}}_{\text{cam},1c}^{\text{omni}}$ and $\tilde{\mathbf{X}}_{\text{cam},0a}^{\text{omni}}$ respectively:

$$\hat{\mathbf{Z}}_j^{\text{omni}} = \left\| \tilde{\mathbf{X}}_{\text{cam},j}^{\text{omni}} \right\| \quad (6.26)$$

See Figure 6.12. Note that the depth values z represent the norm of the 3D points given in camera coordinates.

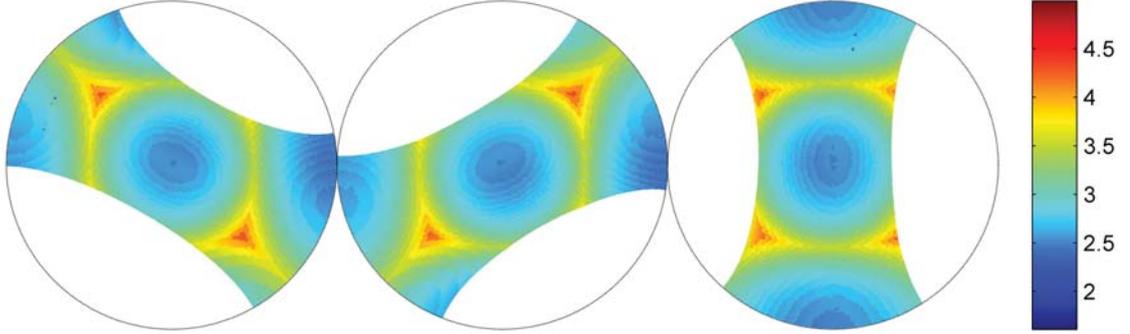


Figure 6.12: Result of the projection process of rectified depth maps ($\mathbf{Z}_{0a}^{\text{omni}}$, $\mathbf{Z}_{2b}^{\text{omni}}$ and $\mathbf{Z}_{1c}^{\text{omni}}$).

Together they can be aggregated to form a dense depth map $\mathbf{Z}_0^{\text{omni}}$ with respect to the integer pixel grid $\mathbf{X}_{\text{img},0}^{\text{omni}}$ and pixel values $\mathbf{I}_0^{\text{omni}}$ respectively. This is achieved by introducing a nearest neighbour or bilinear interpolation method \mathcal{I} .

$$\mathbf{Z}_0^{\text{omni}} = \mathcal{I} \left(\hat{\mathbf{Z}}_{0a}^{\text{omni}} \cup \hat{\mathbf{Z}}_{2b}^{\text{omni}} \cup \hat{\mathbf{Z}}_{1c}^{\text{omni}} \right) \quad (6.27)$$

An investigation of certain merge methods has not been done at this point of time. The finally obtained depth map $\mathbf{Z}_0^{\text{omni}}$ is visualized in Figure 6.13.

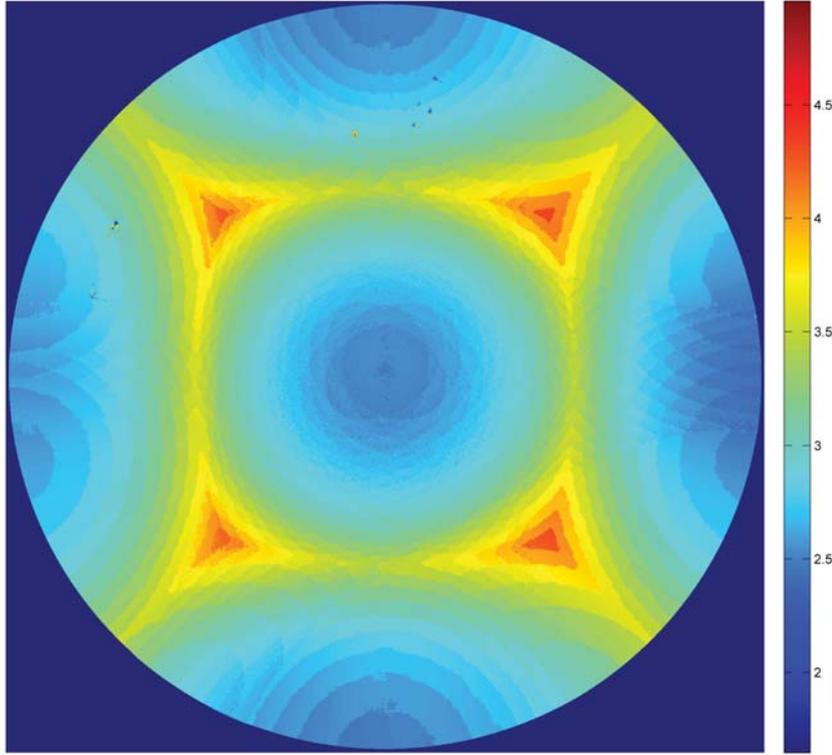


Figure 6.13: Merged depth map Z_0^{omni}

6.5.5 3D Reprojection

In order to check the calculated spherical depth map, the result shall be validated by a 3D reprojection. The integer spherical image grid $\mathbf{X}_{\text{img},0}^{\text{omni}}$ is reprojected employing Equation 2.15 and Equation 2.28 inversely:

$$\mathbf{X}_{\text{cam},0}^{\text{omni}}(\lambda) = \lambda \cdot \mathcal{F}_e^{-1} \left(\mathbf{K}_0^{\text{omni}^{-1}} \cdot \mathbf{X}_{\text{img},0}^{\text{omni}} \right) \quad (6.28)$$

The computed camera rays are subsequently scaled according to its depth values:

$$\mathbf{X}_{\text{cam},0}^{\text{omni}} = \left[\frac{x_{\text{cam},0}^{\text{omni}}}{\lambda} \cdot Z_0^{\text{omni}} \quad \frac{y_{\text{cam},0}^{\text{omni}}}{\lambda} \cdot Z_0^{\text{omni}} \quad \frac{z_{\text{cam},0}^{\text{omni}}}{\lambda} \cdot Z_0^{\text{omni}} \quad 1 \right]^T \quad (6.29)$$

Since $\text{Cam}_0^{\text{omni}}$ represents the reference camera, its coordinate system coincides with the world frame. That means the computed camera points $\mathbf{X}_{\text{cam},0}^{\text{omni}}$ form the estimate $\hat{\mathbf{X}}_{\text{world}}$ of the original world points $\mathbf{X}_{\text{world}}$:

$$\mathbf{X}_{\text{cam},0}^{\text{omni}} \hat{=} \hat{\mathbf{X}}_{\text{world}} \quad (6.30)$$

The illustration in Figure 6.14 depicts the pixel values $\mathbf{I}_0^{\text{omni}}$ of the reference camera at its estimated world positions $\hat{\mathbf{X}}_{\text{world}}$.

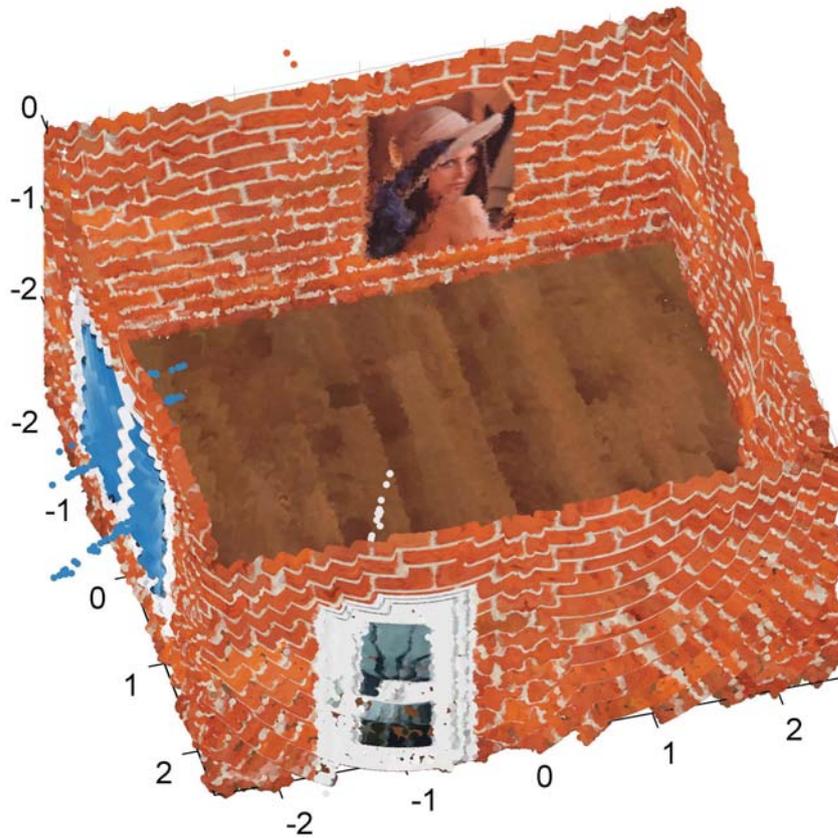


Figure 6.14: 3D re-projection of the omnidirectional depth map Z_0^{omni}

6.6 Error Analysis

The accuracy and error effects that appear when stereo vision is processed have been discussed in Section 3.5 for an ordinary perspective two-view configuration. Accuracy preservation for virtual camera generation has been discussed in Section 4.4. At this point the expectable measurement errors for the presented stereo vision configuration shall be numerically evaluated and discussed.

In Chapter 3 it has been discussed that 3D measurement devices which rely on stereo vision face a considerable amount of error sources. Primarily these are inaccurate intrinsically and extrinsically calibrated setups or defective stereo correspondence results as a matter of principle. However in this section the focus is on imperfect depth measurements due to a limited accuracy of the underlying pixel information and errors that are introduced due to the special characteristics of the present system.

Although the configuration works in principle with each and every afore-discussed configuration (cf. Table 6.3, Table 6.4, Table 6.5 and Table 6.6) I will restrict this investigation to epipolar equidistant rectification. It has shown the best achievable resolution.

As the plot in Figure 6.15 shows, a depth measurement error analysis is derived for a distinct hemispherical distance around the sensor with an exemplary value of $Z = 2.5m$.

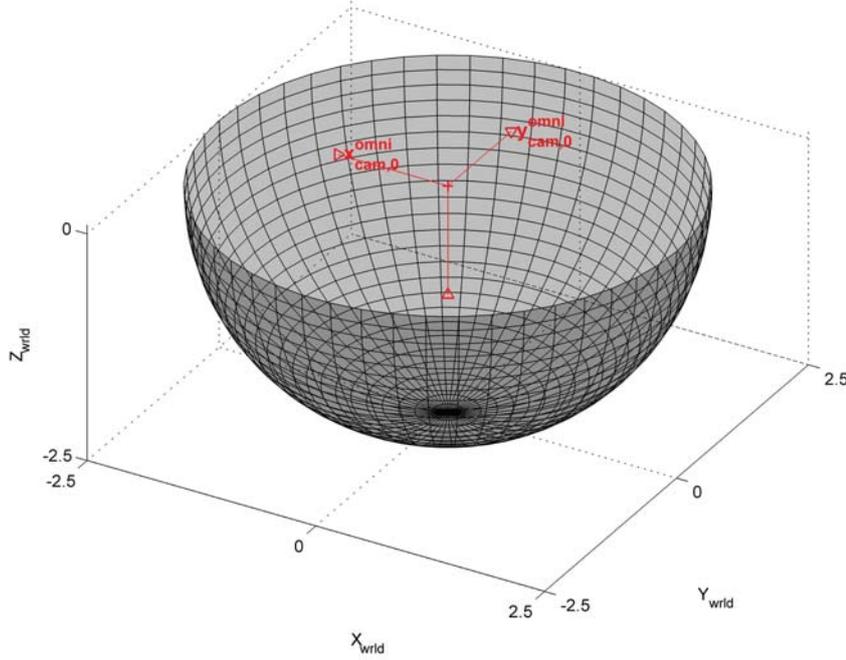


Figure 6.15: Hemispherical measurement space around the sensor with an exemplary distance of $Z = 2.5m$

For the given environment (Figure 6.15) the rectified distance maps $\mathbf{Z}_j^{\text{rect}}$ (representing $\mathbf{Z}_{0a}^{\text{rect}}$, $\mathbf{Z}_{2b}^{\text{rect}}$ and $\mathbf{Z}_{1c}^{\text{rect}}$) are computed by following the algorithm of Section 6.5.3. For the underlying image point meshes $\mathbf{X}_{\text{img},j}^{\text{rect}}$ (representing $\mathbf{X}_{\text{img},0a}^{\text{rect}}$, $\mathbf{X}_{\text{img},2b}^{\text{rect}}$ and $\mathbf{X}_{\text{img},1c}^{\text{rect}}$) the appropriate camera points $\mathbf{X}_{\text{cam},j}^{\text{rect}}$ are obtained by back-projection and normalization to the afore-calculated distance maps as presented in Section 6.5.4 - cf. Equation 6.22 and Equation 6.23. For the epipolar equidistant method it is achieved as follows:

$$\mathbf{X}_{\text{cam},j}^{\text{rect}} = \left[x_{\text{cam},j}^{\text{rect}} \cdot Z_j^{\text{rect}} \quad y_{\text{cam},j}^{\text{rect}} \cdot Z_j^{\text{rect}} \quad z_{\text{cam},j}^{\text{rect}} \cdot Z_j^{\text{rect}} \quad 1 \right]^T \quad (6.31)$$

Generally, the triangulation process produces erroneous distance values. For the epipolar equidistance method this has been initially discussed in Chapter 5. An expectable magnitude for this value can be derived from Equation 5.4:

$$|\delta Z| = \left[\left| b \cdot \frac{\cos(\phi - \phi_l)}{\sin \phi} \right| + \left| b \cdot \frac{\sin(\phi - \phi_l) \cos \phi}{\sin \phi^2} \right| \right] \cdot |\delta \phi| \quad (6.32)$$

The minimum measurable disparity angle $\delta \phi$ for this configuration is in accordance with Section 3.5.4:

$$\delta \phi = \pm \frac{1}{2 \cdot \alpha_x} \quad (6.33)$$

As stated in Section A.3 the error value δZ can be considered as random variable complying with a conditional probability function $p(\delta Z|Z)$, which itself implies a certain depth value Z . An approximation of $p(\delta Z|Z)$ has been derived in Section A.4. According to this calculation the first order approximation of $p(\delta Z|Z)$ can be fully described by $\pm\delta Z(\delta\phi)$.

Hence in order to simplify the overall error computation, two measurement points parametrized by $\pm\delta Z$ are calculated instead of applying the whole probability density function:

$$\mathbf{X}_{\text{cam},j+}^{\text{rect}} = \begin{bmatrix} x_{\text{cam},j}^{\text{rect}} \cdot (Z_j^{\text{rect}} + |\delta Z|) \\ y_{\text{cam},j}^{\text{rect}} \cdot (Z_j^{\text{rect}} + |\delta Z|) \\ z_{\text{cam},j}^{\text{rect}} \cdot (Z_j^{\text{rect}} + |\delta Z|) \\ 1 \end{bmatrix}; \mathbf{X}_{\text{cam},j-}^{\text{rect}} = \begin{bmatrix} x_{\text{cam},j}^{\text{rect}} \cdot (Z_j^{\text{rect}} - |\delta Z|) \\ y_{\text{cam},j}^{\text{rect}} \cdot (Z_j^{\text{rect}} - |\delta Z|) \\ z_{\text{cam},j}^{\text{rect}} \cdot (Z_j^{\text{rect}} - |\delta Z|) \\ 1 \end{bmatrix} \quad (6.34)$$

Following Equation 6.24, $\mathbf{X}_{\text{cam},j}^{\text{rect}}$, $\mathbf{X}_{\text{cam},j+}^{\text{rect}}$ and $\mathbf{X}_{\text{cam},j-}^{\text{rect}}$ are subsequently expressed with respect to the reference camera $\text{Cam}_0^{\text{omni}}$ as $\mathbf{X}_{\text{cam},j}^{\text{omni}}$, $\mathbf{X}_{\text{cam},j+}^{\text{omni}}$ and $\mathbf{X}_{\text{cam},j-}^{\text{omni}}$. After re-projection following Equation 6.25 we obtain $\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}}$, $\hat{\mathbf{X}}_{\text{img},j+}^{\text{omni}}$ and $\hat{\mathbf{X}}_{\text{img},j-}^{\text{omni}}$. It turns out that we will observe a *mapping error* denoted as $\Delta\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}}$ for the image point $\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}}$:

$$\Delta\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}} \approx \frac{1}{2} \left\| \hat{\mathbf{X}}_{\text{img},j+}^{\text{omni}} - \hat{\mathbf{X}}_{\text{img},j-}^{\text{omni}} \right\| \quad (6.35)$$

The largest displacement to be expected on the target image can be calculated subsequently by the maximum value of the projected points from each of the rectified cameras:

$$\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}} = \max \left(\Delta\hat{\mathbf{X}}_{\text{img},0a}^{\text{omni}}, \Delta\hat{\mathbf{X}}_{\text{img},2b}^{\text{omni}}, \Delta\hat{\mathbf{X}}_{\text{img},1c}^{\text{omni}} \right) \quad (6.36)$$

Figure 6.16 visualizes the deployment and the magnitude $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ on the reference camera for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$ and configuration no. 6.

It becomes obvious that the mapping error is deployed quite inhomogeneously. This is due to the different origins of the stereo measuring cameras and the subsequent re-mapping to one of them. Naturally, the reference camera $\text{Cam}_0^{\text{omni}}$ does not introduce any displacement as can be seen as dark blue area. The remaining areas show errors $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ up to little more than one pixel. However, common stereo vision correspondence operations like block matching diminish the achievable spacial resolution for depth estimation anyway. This effect actually depends on the applied block size for image correlation. Commonly one utilizes block sizes of not less than 11×11 pixels which smooth surfaces and therefore should absorb the impact of the mapping error. More detailed investigations for this error have not yet been done.

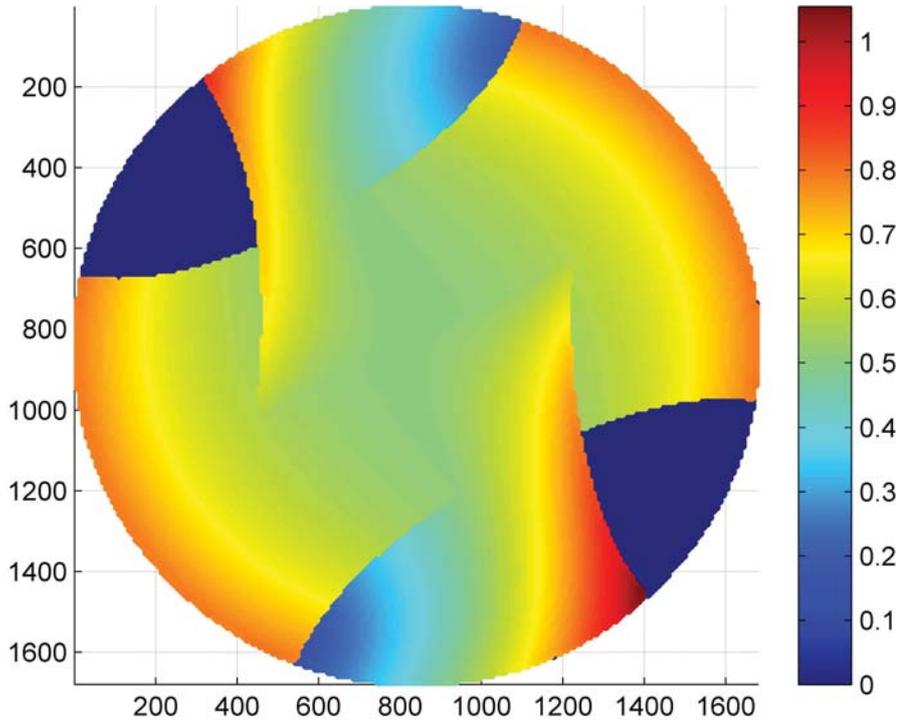


Figure 6.16: Maximum mapping error $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$, an equian-gular projection model for the source camera, an epipolar equidistant projection model for the target camera, a $\text{FOV}_H = 83.20^\circ$ and a baselength of $b = 15cm$.

Actually $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}} = 1$ represents a certain observation angle which is numerically small. For configuration no. 6 it can be calculated as follows:

$$\frac{\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}}{\alpha_x} = \frac{1}{526} = 0.0019 \cong 0.1^\circ \quad (6.37)$$

Furthermore $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ varies according to multiple parameters:

- Stereo vision configuration: The larger the baselength b is, the higher the expected maximum mapping error $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ becomes for a certain distance under observation and for a distinct parametrized minimum measurable distance.
- Measurement distance: The mapping error decreases with increasing measurement distance Z . At infinity the different locations of the three cameras will become meaningless and hence $\Delta\hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ approaches zero.
- Virtual camera configuration: Currently the virtual views are sampled according to the under-sampling strategy - cf. Chapter 4. In the center of each virtual image its resolution coincides with the underlying omnidirectional image. In contrast, the outer areas highly under-sample the source image. This leads to the effect that

the mapping error increases especially in the outer areas, when erroneous distance information from the virtual views is mapped to the omnidirectional reference camera. Hence, a change to over-sampling would reduce $\Delta \hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ considerably.

Figure 6.17 outlines the calculated maximum mapping error for different configurations and distances Z .

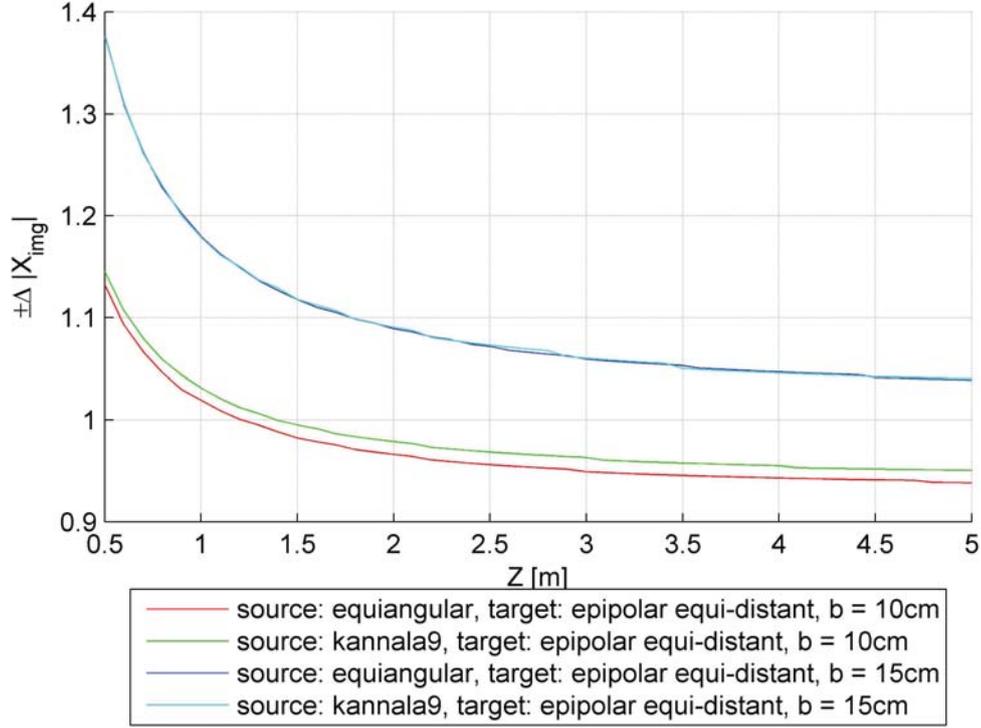


Figure 6.17: Maximum mapping error $\Delta \hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ for multiple measurement distances and camera configurations

The actual depth error $|\delta Z_j^{\text{omni}}|$ for $\hat{\mathbf{X}}_{\text{img},j}^{\text{omni}}$ can hence be estimated by orthogonal projection of $\mathbf{X}_{\text{cam},j+}^{\text{omni}}$ and $\mathbf{X}_{\text{cam},j-}^{\text{omni}}$ to $\mathbf{X}_{\text{cam},j}^{\text{omni}}$ and subsequently subtracting the magnitudes as follows:

$$|\delta Z_j^{\text{omni}}| = \frac{1}{2 \|\mathbf{X}_{\text{cam},j}^{\text{omni}}\|} \left(\mathbf{X}_{\text{cam},j}^{\text{omni}T} \cdot \mathbf{X}_{\text{cam},j+}^{\text{omni}} - \mathbf{X}_{\text{cam},j}^{\text{omni}T} \cdot \mathbf{X}_{\text{cam},j-}^{\text{omni}} \right) \quad (6.38)$$

A maximum distance error $|\delta Z_{\text{max}}^{\text{omni}}|$ on the target image can be calculated subsequently:

$$|\delta Z_{\text{max}}^{\text{omni}}| = \max \left(|\delta Z_{0a}^{\text{omni}}|, |\delta Z_{2b}^{\text{omni}}|, |\delta Z_{1c}^{\text{omni}}| \right) \quad (6.39)$$

Figure 6.18 visualizes the deployment and the magnitude $|\delta Z_{\text{max}}^{\text{omni}}|$ on the reference camera for an exemplary distance of $Z = 2.5m$ and configuration no. 6. While the left plot shows

the maximum error of my approach, the right one outlines the same for the system setup of [67].

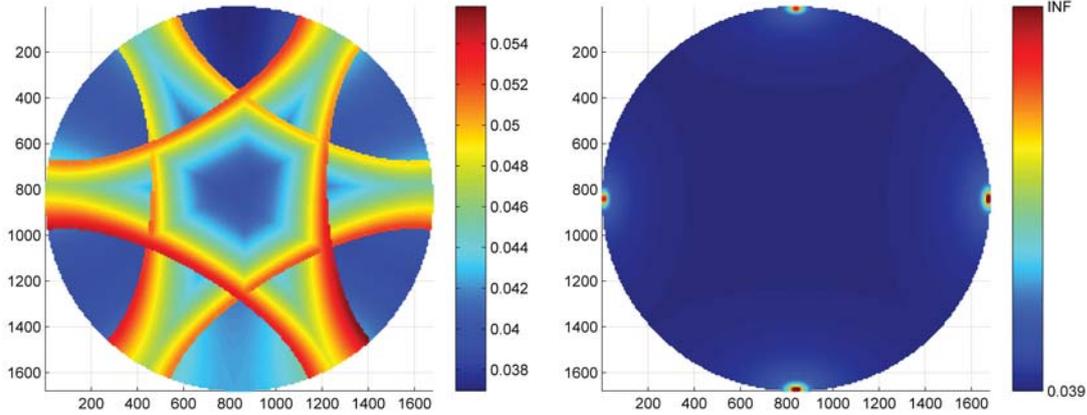


Figure 6.18: Maximum distance error $|\delta Z_{\max}^{\text{omni}}|$ for an exemplary distance of $Z_0^{\text{omni}} = 2.5\text{m}$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 15\text{cm}$; Left: our approach ($\text{FOV}_H = 83.20^\circ$), Right: Approach of Li Shigang [67]

Reviewing the left plot of Figure 6.18 it is apparent that the maximum distance error is deployed heterogeneously. Firstly the non-isotrope error distribution of the virtual cameras, which has already been discussed in Chapter 5, causes each partial distance map to be most accurate at the center. Secondly, the different location of each measurement camera, with respect to the hemispherical measurement area (Z) around the sensors center, causes each camera to measure varying distances at different directions. Subsequently, the sensing is carried out with versatile accuracy.

Comparing both plots of Figure 6.18 it can be concluded: The expected minimum error is identical for both approaches - given the same underlying virtual pixel configuration and baselength. In the left plot the ratio between the smallest and the largest error is determined to ≈ 0.70 . This leads to the consequence that in the worst case a measurement can be carried out with $\approx 70\%$ of the best achievable accuracy without any depth averaging method. Since Li [67] includes the epipoles in this process the worst case are a very coarse accuracy in the near of the singularities up to zero accuracy at the singularities.

The final accuracy that will be achieved depends on different aspects:

- Merge strategy: Overlapping measurements can be used to average the error by integration over multiple calculated depth values and obtain $|\delta Z_{\text{avg}}^{\text{omni}}|$. Since the epipoles introduce very inaccurate depth data or "*great shape distortion*" [67] for reconstruction Li [67] employs a weighted-averaging process in order to handle this circumstance. By means of that method "*the shape distortion [...] is apparently improved.*" [67]. A quantitative assessment has not been presented.

- **Reliable disparity information:** An adequate depth information (with underlying accuracy) can only be obtained when the calculation of the disparity maps works properly. Since common correspondence algorithms fail near the epipoles, partially sparse depth maps and unreliable depth information are the consequence. Especially the latter issue may cause a weighted merge strategy to be annulled. However, a more detailed treatment of this problem is beyond the scope of this work.

Figure 6.18 showed the deployment of the maximum error $|\delta Z_{\max}^{\text{omni}}|$. Applying an arbitrary merge strategy will result in an averaged error $|\delta Z_{\text{avg}}^{\text{omni}}|$ with

$$|\delta Z_{\max}^{\text{omni}}| > |\delta Z_{\text{avg}}^{\text{omni}}| > |\delta Z_{\min}^{\text{omni}}| \quad (6.40)$$

where $|\delta Z_{\min}^{\text{omni}}|$ is a lower theoretical error boundary. For roughly estimating $|\delta Z_{\min}^{\text{omni}}|$ (at least for a known distance Z) the joint probability density function $p^{\text{omni}}(\delta Z|Z)$ shall be exploited which is calculated as follows:

$$p^{\text{omni}}(\delta Z|Z) = p_{0a}^{\text{omni}}(\delta Z|Z) \cdot p_{2b}^{\text{omni}}(\delta Z|Z) \cdot p_{1c}^{\text{omni}}(\delta Z|Z) \quad (6.41)$$

where $p_j^{\text{omni}}(\delta Z|Z)$ represents the conditional probability of δZ implied a certain depth value Z . The first order approximation of this function can be constituted according to Appendix A, Section A.4 by employing the afore-calculated error $|\delta Z_j^{\text{omni}}|$. The error δZ of $p^{\text{omni}}(\delta Z|Z)$ with an probability density of 0.5 is considered as lower boundary $|\delta Z_{\min}^{\text{omni}}|$.

Figure 6.19 outlines the obtained colour coded maps for $|\delta Z_{\min}^{\text{omni}}|$ and the configuration under current investigation.

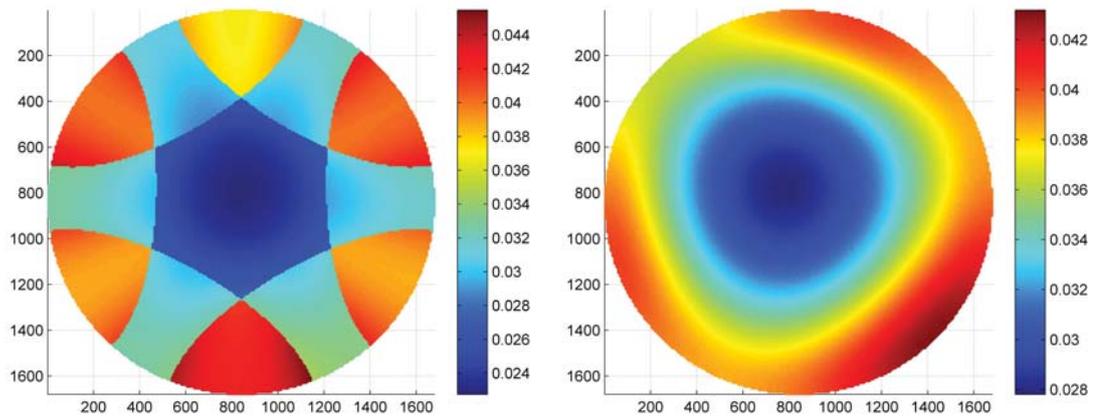


Figure 6.19: Minimum distance error $|\delta Z_{\min}^{\text{omni}}|$ for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 15cm$; Left: our approach ($\text{FOV}_H = 83.20^\circ$), Right: Approach of Li Shigang [67]

Figure 6.19 demonstrates that both approaches perform quite similarly although the largest $|\delta Z_{\min}^{\text{omni}}|$ of Li is slightly smaller than ours.

Equivalently to the mapping error, the distance error depends on the distinct stereo vision configuration, the actual measurement distance and the virtual camera configuration. The following illustrations in Figure 6.20 and Figure 6.21 show the error behaviour for different configurations and distances.

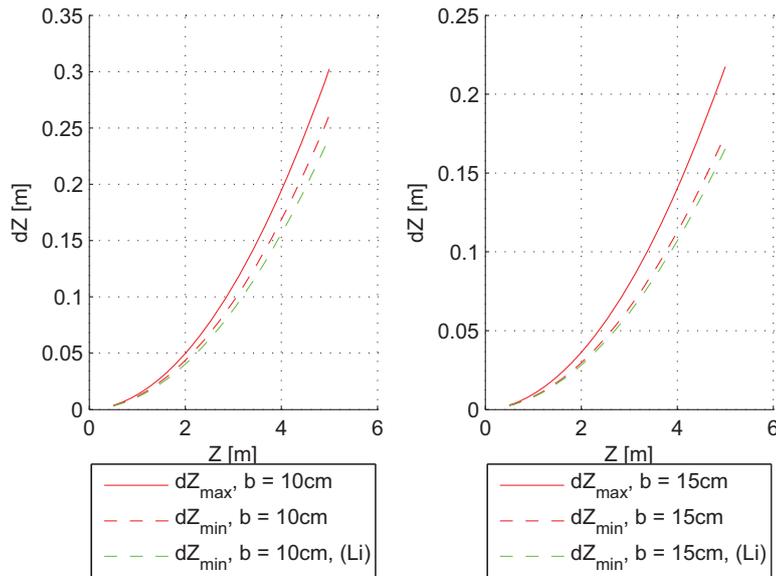


Figure 6.20: Largest minimal Error $|\delta Z_{\min}^{\text{omni}}|$ and maximal Error $|\delta Z_{\max}^{\text{omni}}|$ for multiple measurement distances, an equiangular projection model for the source camera and two baselengths

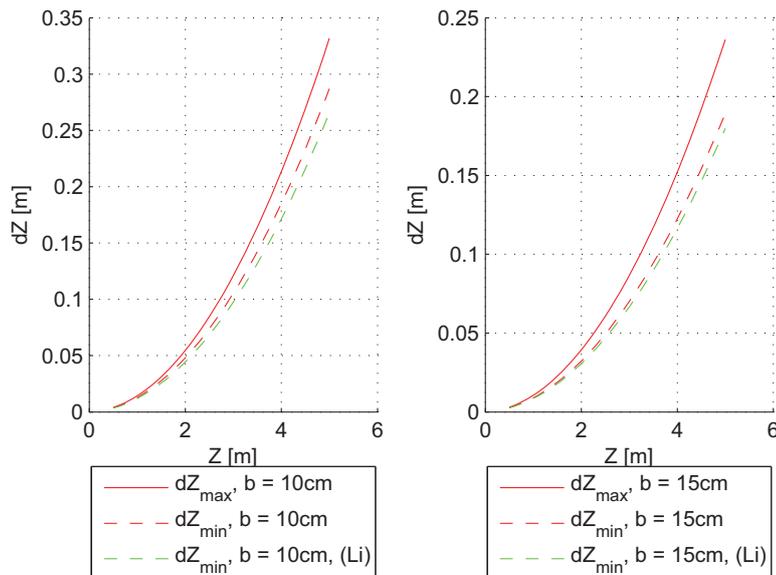


Figure 6.21: Largest minimal Error $|\delta Z_{\min}^{\text{omni}}|$ and maximal Error $|\delta Z_{\max}^{\text{omni}}|$ for multiple measurement distances, a generic projection model for the source camera and two baselengths

The presented charts clarify that the author’s approach shows a similar behaviour on the theoretical achievable accuracy $|\delta Z_{\min}^{\text{omni}}|$ in comparison to Li’s method. The green and red dashed lines hardly differ.

In contrast to Li [67] where the maximum expectable error $|\delta Z_{\max}^{\text{omni}}|$ is at infinity (a green solid line cannot be plotted), our maximum error outlined as a red solid line is *not far away* from $|\delta Z_{\min}^{\text{omni}}|$. Assuming a baselength $b = 15\text{cm}$, a distance $Z = 2.5\text{m}$ and an equiangular projection model for the source camera (configuration no. 6, Figure 6.20 on the right) for instance, we obtain a minimum percentage depth error of:

$$\frac{|\delta Z_{\min}^{\text{omni}}|}{Z} = \frac{0.17}{5.0} = 0.034 \hat{=} 3.4\% \quad (6.42)$$

and a maximum percentage depth error of:

$$\frac{|\delta Z_{\max}^{\text{omni}}|}{Z} = \frac{0.22}{5.0} = 0.084 \hat{=} 4.4\% \quad (6.43)$$

The actual measurement error $|\delta Z_{\text{avg}}^{\text{omni}}|$ will appear somewhere in between depending on the merge strategy. This aspect proofs that using narrow-angled virtual cameras for composing the hemispherical depth map makes sophisticated merging methods obsolete.

Finally Figure 6.22 demonstrates how $|\delta Z_{\max}^{\text{omni}}|$ changes over distance Z and the parameter FOV_H of the virtual cameras. One can conclude: When there is a possibility to further reduce FOV_H , an enhanced accuracy can be obtained.

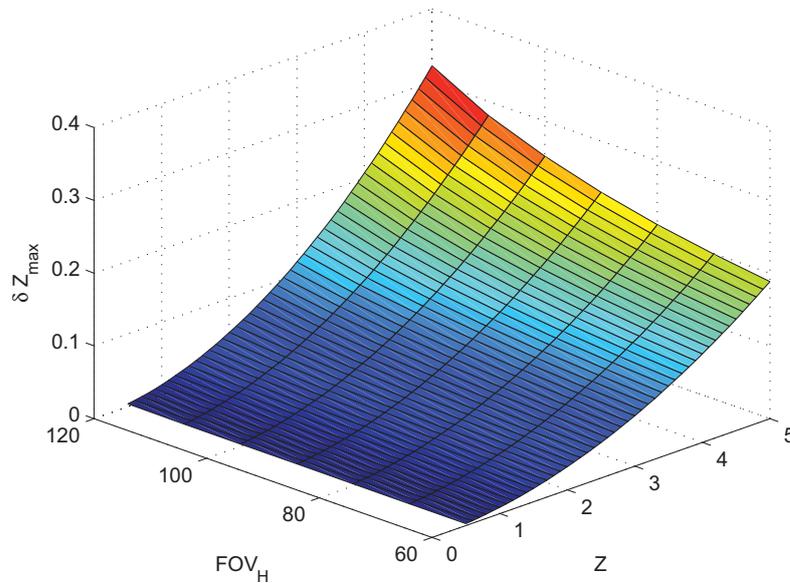


Figure 6.22: Maximum distance error $|\delta Z_{\max}^{\text{omni}}|$ for an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 15\text{cm}$ depending on different measurement distances Z and the parameter FOV_H of the virtual cameras

Chapter 7

Stereo Vision Demonstrator

This chapter treats a real system that shall demonstrate the practical feasibility of the afore-presented omnidirectional stereo method. In Section 7.1 the actual compiled camera configuration is presented. Successively, the currently employed calibration strategy for the system is explained in Section 7.2. In order to calculate a depth map, appropriate rectified views have to be configured. This is done in Section 7.3.

Finally, the C++ implementation that currently runs the system as well as first experimental results are outlined in Section 7.4 and Section 7.5.

7.1 Physical System Setup

The applied imaging system comprises three Allied GC940GC industry cameras (1-inch CMOSIS CMV4000 sensor with a full resolution of 2048×2048 pixels), each equipped with a FE185C086HA-1 fisheye lens that is characterized by a focal length of $f = 2.7$ mm and a field of view of 185° . The captured images are cropped to an effective resolution of 1680×1680 pixels.

The stereo vision setup, as can be seen in Figure 7.1, is specified in order to provide an approximate baselength of $b \approx 10$ cm for each rectified stereo camera.

For capturing images the cameras have been triggered in order to record isochronously. The illustrations in Figure 7.2 show images that have been generated in the professorship's lab apartment.



Figure 7.1: A spherical stereo vision demonstrator



Figure 7.2: Sample images from the professorships lab apartment

7.2 System Calibration Strategy

The stereo vision system has been initially calibrated, intrinsically as well as extrinsically, in order to enable the systems 3D measurement capability.

7.2.1 Intrinsic Calibration of the Physical Cameras

Although the employed lenses project very similarly to the equiangular projection method (cf. Section 2.2.3), they do not comply with this principle exactly. That is why the projection behaviour of each imaging device $\text{Cam}_0^{\text{omni}}$, $\text{Cam}_1^{\text{omni}}$ and $\text{Cam}_2^{\text{omni}}$ is approximated by means of the generic camera model by Juho Kannala - cf. Section 2.2.4. For the generation of reference points the author employs a special three-dimensional calibration target that comprises several planar calibration patterns at once, as pictured in Figure 7.3. The results of this calibration step can be reviewed in Table 7.1.

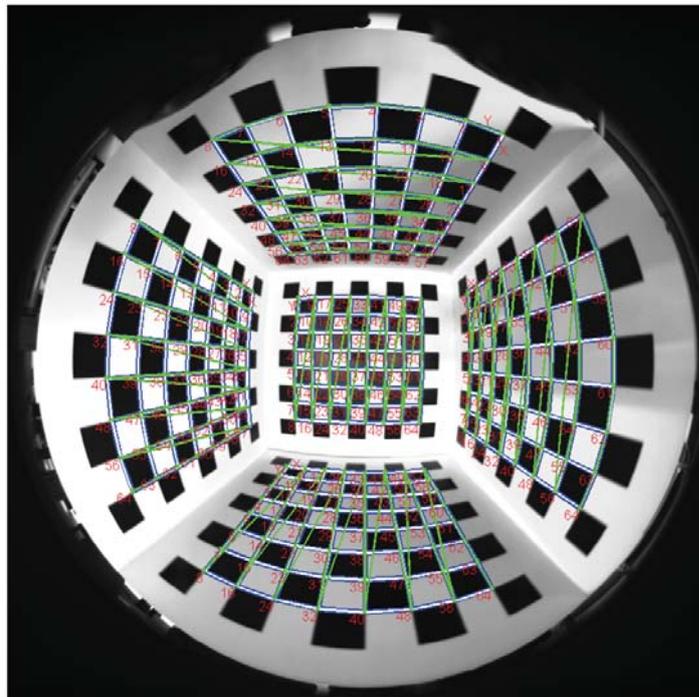


Figure 7.3: Special three-dimensional calibration target

One can observe that the applied lenses represent a quite similar projection behaviour. Furthermore, the distortion effects compared to the ideal equiangular projection model is basically determined by the first coefficient k_1 - cf. Section 2.2.4 Equation 2.29.

Table 7.1: Results of the intrinsic generic camera models estimation

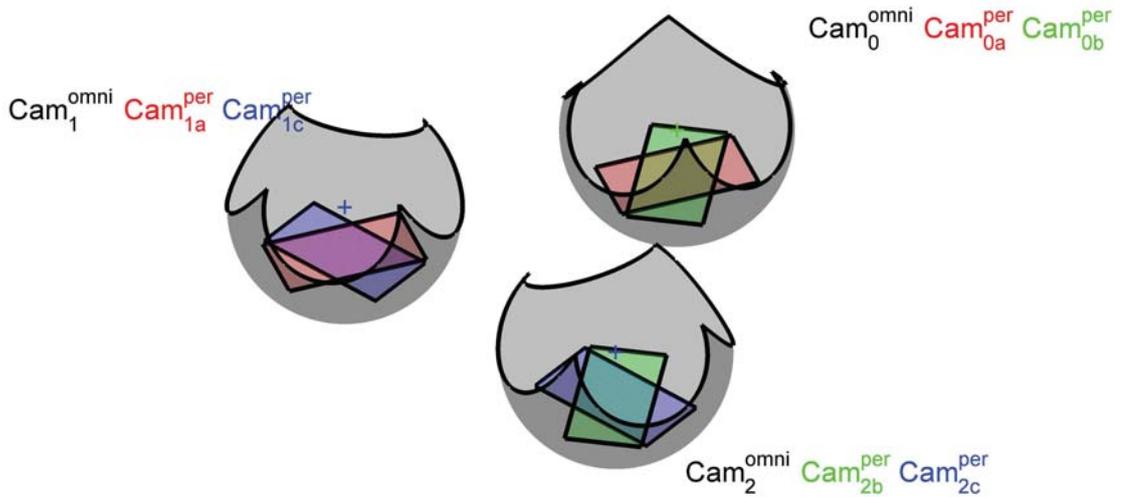
Camera	k_1	k_2	α_x	α_y	c_x	c_y	k_3	k_4	k_5
Cam ₀ ^{omni}	2.43	0.08	198.07	198.04	830.37	844.53	-0.04	0.03	-0.01
Cam ₁ ^{omni}	2.43	0.11	198.14	197.97	843.72	849.74	-0.11	0.07	-0.01
Cam ₂ ^{omni}	2.44	0.09	198.00	198.12	857.28	846.02	-0.07	0.04	-0.01

7.2.2 Extrinsic Calibration of the Physical and the Virtual Cameras

In Chapter 2 and Chapter 4 it is explained that the extrinsic parameters of each physical and virtual camera are fully described by the homography \mathbf{H} (cf. Chapter 2 Equation 2.20), comprising the cameras origin \mathbf{C} and its rotation matrix \mathbf{R} (composed of Euler Angles φ_x , φ_y and φ_z).

The applied calibration method utilizes *two view geometry*, as discussed in Section 2.4 - without optimization amongst all three views. Therefore, six virtual perspective cameras, denoted as Cam_j^{per}, are utilized. In combination, they compose three virtual perspective stereo cameras, denoted as SCam_k^{per}, for calibration. Their extrinsic parameters coincide with those of Cam_j^{rect}, and SCam_k^{rect} respectively. This relationship is presented in Equation 7.1. An adequate illustration, comparable with Figure 5.17 of Section 5.3, can be reviewed in Figure 7.4.

$$\begin{aligned}
 \text{SCam}_a^{\text{per}} &= \{\text{Cam}_{0a}^{\text{per}}, \text{Cam}_{1a}^{\text{per}}\} \\
 \text{SCam}_b^{\text{per}} &= \{\text{Cam}_{2b}^{\text{per}}, \text{Cam}_{0b}^{\text{per}}\} \\
 \text{SCam}_c^{\text{per}} &= \{\text{Cam}_{1c}^{\text{per}}, \text{Cam}_{2c}^{\text{per}}\}
 \end{aligned} \tag{7.1}$$

**Figure 7.4:** Geometrical configuration for extrinsic calibration of the physical and virtual cameras

7.2.2.1 Extrinsic Initialization of the Physical Cameras

Initially, it is assumed that each omnidirectional camera $\text{Cam}_i^{\text{omni}}$ is located at the world's origin $\mathbf{C}_i^{\text{omni}} = \mathbf{0}$. Furthermore, two cameras are characterized by an ideal mutual rotation of $\frac{2}{3}\pi$ around their own z-axes: $\varphi_{z0}^{\text{omni}} = 0$, $\varphi_{z1}^{\text{omni}} = +\frac{2}{3}\pi$ and $\varphi_{z2}^{\text{omni}} = -\frac{2}{3}\pi$. This can be justified by reviewing Figure 5.15 in Chapter 5. Thereby, initial extrinsic parameters are determined, denoted as $\mathbf{Hi}_0^{\text{omni}}$, $\mathbf{Hi}_1^{\text{omni}}$ and $\mathbf{Hi}_2^{\text{omni}}$.

7.2.2.2 Extrinsic Initialization of the Virtual Cameras

At the beginning the virtual cameras inherit the extrinsic configuration from the omnidirectional cameras. To roughly pre-rectify each virtual stereo camera pair $\text{SCam}_k^{\text{per}}$, each left camera ($\text{Cam}_{0a}^{\text{per}}$, $\text{Cam}_{2b}^{\text{per}}$, $\text{Cam}_{1c}^{\text{per}}$) of each pair is rotated around its z-axis towards the associate right camera ($\text{Cam}_{1a}^{\text{per}}$, $\text{Cam}_{0b}^{\text{per}}$, $\text{Cam}_{2c}^{\text{per}}$) by an angle of $\frac{1}{3}\pi$, and vice versa. In combination with the initial extrinsic parameters, determined in Section 7.2.2.1, the initial homographies $\mathbf{Hi}_j^{\text{per}}$ for the virtual perspective cameras can be calculated as follows:

$$\begin{aligned}
\mathbf{Hi}_{0a}^{\text{per}} &= d\mathbf{Hi}_+^{\text{per}} \cdot \mathbf{Hi}_0^{\text{omni}} \\
\mathbf{Hi}_{0b}^{\text{per}} &= d\mathbf{Hi}_-^{\text{per}} \cdot \mathbf{Hi}_0^{\text{omni}} \\
\mathbf{Hi}_{1a}^{\text{per}} &= d\mathbf{Hi}_-^{\text{per}} \cdot \mathbf{Hi}_1^{\text{omni}} \\
\mathbf{Hi}_{1c}^{\text{per}} &= d\mathbf{Hi}_+^{\text{per}} \cdot \mathbf{Hi}_1^{\text{omni}} \\
\mathbf{Hi}_{2b}^{\text{per}} &= d\mathbf{Hi}_+^{\text{per}} \cdot \mathbf{Hi}_2^{\text{omni}} \\
\mathbf{Hi}_{2c}^{\text{per}} &= d\mathbf{Hi}_-^{\text{per}} \cdot \mathbf{Hi}_2^{\text{omni}}
\end{aligned} \tag{7.2}$$

where $d\mathbf{Hi}_+^{\text{per}}$ and $d\mathbf{Hi}_-^{\text{per}}$ determine the homographies describing a differential rotation around the z-axis in positive direction, and negative direction respectively. Hence $\mathbf{Hi}_j^{\text{per}}$, comprising $\mathbf{Ri}_j^{\text{per}}$ and $\mathbf{Ci}_j^{\text{per}}$, represent the initial extrinsic estimates of the virtual perspective cameras which roughly complies with the mechanical system configuration.

7.2.2.3 Two-View Stereo Calibration and Rectification

The subsequent calibration procedure is accomplished for each stereo camera $\text{SCam}_k^{\text{per}}$ independently. For instance, Figure 7.5 shows sample images that have been generated for calibrating stereo camera $\text{SCam}_a^{\text{per}}$.

This process is performed in accordance with the description in Section 3.2. Therefore, the calibration toolbox by Bouguet [11] is applied for each stereo pair in order to obtain calibration extrinsic estimates, denoted as $\hat{\mathbf{H}}\mathbf{c}_{1a}^{\text{per}}$, $\hat{\mathbf{H}}\mathbf{c}_{0b}^{\text{per}}$ and $\hat{\mathbf{H}}\mathbf{c}_{2c}^{\text{per}}$, for each of the right

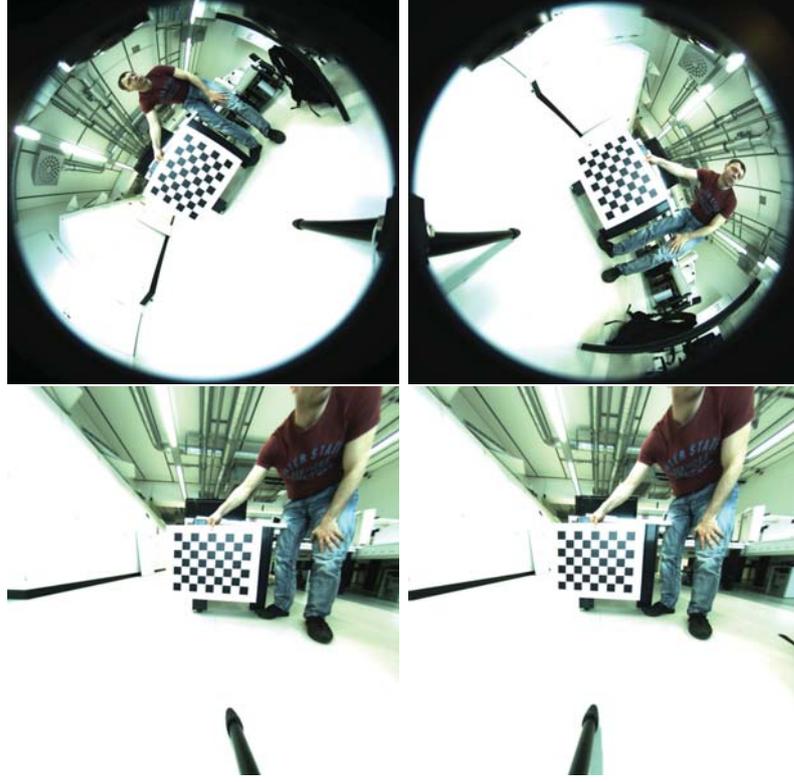


Figure 7.5: Example calibration images for $S\text{Cam}_a^{\text{per}}$ with original images from $\text{Cam}_0^{\text{omni}}$ ($H i_0^{\text{omni}}$) and $\text{Cam}_1^{\text{omni}}$ ($H i_1^{\text{omni}}$) as well as pre-rectified images from $\text{Cam}_{0a}^{\text{per}}$ ($H i_{0a}^{\text{per}}$) and $\text{Cam}_{1a}^{\text{per}}$ ($H i_{1a}^{\text{per}}$)

cameras ($\text{Cam}_{1a}^{\text{per}}$, $\text{Cam}_{0b}^{\text{per}}$ and $\text{Cam}_{2c}^{\text{per}}$) with respect to their associate left ones ($\text{Cam}_{0a}^{\text{per}}$, $\text{Cam}_{2b}^{\text{per}}$ and $\text{Cam}_{1c}^{\text{per}}$).

The utilization of a rectification procedure for the calibrated images, as explained in Section 3.2 and described by Fusiello [28], results in new corrected homographies, denoted as $\hat{H}r_a^{\text{per}}$, $\hat{H}r_b^{\text{per}}$ and $\hat{H}r_c^{\text{per}}$, for each left and right perspective camera:

$$\begin{aligned}
 \hat{H}c_{1a}^{\text{per}} &\longrightarrow \hat{H}r_a^{\text{per}} \\
 \hat{H}c_{0b}^{\text{per}} &\longrightarrow \hat{H}r_b^{\text{per}} \\
 \hat{H}c_{2c}^{\text{per}} &\longrightarrow \hat{H}r_c^{\text{per}}
 \end{aligned} \tag{7.3}$$

7.2.2.4 Three-View Stereo Rectification

The results of Section 7.2.2.3 are now used to locate the omnidirectional cameras $\text{Cam}_i^{\text{omni}}$ extrinsically, as described by Equation 7.4. Three two-view calibrations for three concatenated stereo cameras form an overdetermined configuration. By now, no optimization method has been performed. Hence, the determination of the extrinsic parameters for the fisheye cameras, denoted as $\hat{H}c_0^{\text{omni}}$, $\hat{H}c_1^{\text{omni}}$ and $\hat{H}c_2^{\text{omni}}$, is performed on the basis

of stereo calibration results $\hat{\mathbf{H}}\mathbf{c}_{1a}^{\text{per}}$ and $\hat{\mathbf{H}}\mathbf{c}_{0b}^{\text{per}}$:

$$\begin{aligned}\hat{\mathbf{H}}\mathbf{c}_0^{\text{omni}} &= \mathbf{I} \\ \hat{\mathbf{H}}\mathbf{c}_1^{\text{omni}} &= (\mathbf{d}\mathbf{H}\mathbf{i}_-^{\text{per}})^{-1} \cdot \hat{\mathbf{H}}\mathbf{c}_{1a}^{\text{per}} \cdot \mathbf{d}\mathbf{H}\mathbf{i}_+^{\text{per}} \\ \hat{\mathbf{H}}\mathbf{c}_2^{\text{omni}} &= (\mathbf{d}\mathbf{H}\mathbf{i}_+^{\text{per}})^{-1} \cdot (\hat{\mathbf{H}}\mathbf{c}_{0b}^{\text{per}})^{-1} \cdot \mathbf{d}\mathbf{H}\mathbf{i}_-^{\text{per}}\end{aligned}\tag{7.4}$$

Furthermore, final extrinsics of the virtual perspective cameras $\text{Cam}_j^{\text{per}}$ are computed with respect to $\text{Cam}_0^{\text{omni}}$. See Equation 7.5 and Equation 7.6 for $\text{SCam}_a^{\text{per}}$ as an example:

$$\begin{aligned}\mathbf{d}\hat{\mathbf{H}}\mathbf{r}_{0a}^{\text{per}} &= \hat{\mathbf{H}}\mathbf{r}_a^{\text{per}} \\ \mathbf{d}\hat{\mathbf{H}}\mathbf{r}_{1a}^{\text{per}} &= \hat{\mathbf{H}}\mathbf{r}_a^{\text{per}} \cdot (\hat{\mathbf{H}}\mathbf{c}_{1a}^{\text{per}})^{-1}\end{aligned}\tag{7.5}$$

$$\begin{aligned}\hat{\mathbf{H}}\mathbf{r}_{0a}^{\text{per}} &= \mathbf{d}\hat{\mathbf{H}}\mathbf{r}_{0a}^{\text{per}} \cdot \mathbf{d}\mathbf{H}\mathbf{i}_+^{\text{per}} \cdot \hat{\mathbf{H}}\mathbf{c}_0^{\text{omni}} \\ \hat{\mathbf{H}}\mathbf{r}_{1a}^{\text{per}} &= \mathbf{d}\hat{\mathbf{H}}\mathbf{r}_{1a}^{\text{per}} \cdot \mathbf{d}\mathbf{H}\mathbf{i}_-^{\text{per}} \cdot \hat{\mathbf{H}}\mathbf{c}_1^{\text{omni}}\end{aligned}\tag{7.6}$$

The variables $\mathbf{d}\hat{\mathbf{H}}\mathbf{r}_j^{\text{per}}$ are the correction homographies that have to be superimposed on the initial pre-rotations $\mathbf{d}\mathbf{H}\mathbf{i}_+^{\text{per}}$ and $\mathbf{d}\mathbf{H}\mathbf{i}_-^{\text{per}}$ with respect to the calibrated extrinsics of the omnidirectional cameras $\text{Cam}_i^{\text{omni}}$. The calculation of the parameters for the remaining cameras is identical.

The extrinsic results can directly be used for the rectified cameras:

$$\hat{\mathbf{H}}\mathbf{r}_j^{\text{rect}} \equiv \hat{\mathbf{H}}\mathbf{r}_j^{\text{per}}\tag{7.7}$$

7.2.2.5 Extrinsic Calibration Results

For the purpose of calibration, perspective images of 1000×1000 pixels in size are employed. Every stereo camera is calibrated individually. The results are outlined in Table 7.2.

Table 7.2: Image calibration error

Stereo Camera	Total Average Calibration Error
$\text{SCam}_a^{\text{per}}$	0.1700 px
$\text{SCam}_b^{\text{per}}$	0.2323 px
$\text{SCam}_c^{\text{per}}$	0.1623 px

The rectification of each camera is successively computed as described in Section 7.2.2.4. The extrinsic results are summarized in Table 7.3. The extrinsic parameters reveal that two virtual cameras always share the same position \mathbf{C} (c_x, c_y, c_z) and, furthermore, that

Table 7.3: Rectification results

Camera	$c_x[mm]$	$c_y[mm]$	$c_z[mm]$	$\varphi_x[rad]$	$\varphi_y[rad]$	$\varphi_z[rad]$
Cam _{0a} ^{per}	0.00	0.00	0.00	0.0100	-0.0060	1.0573
Cam _{1a} ^{per}	48.00	-88.82	-0.61	0.0100	-0.0060	1.0573
Cam _{2b} ^{per}	-52.90	-84.87	-0.20	-0.0023	0.0020	-1.0135
Cam _{0b} ^{per}	0.00	0.00	0.00	-0.0023	0.0020	-1.0135
Cam _{1c} ^{per}	48.00	-88.82	-0.61	0.0042	0.0020	-3.1058
Cam _{2c} ^{per}	-52.90	-84.87	-0.20	0.0046	0.0015	-3.1038

three pairs of cameras coincide in terms of orientation $\mathbf{R}(\varphi_x, \varphi_y, \varphi_z)$, as a consequence of rectification. However, this does not completely hold true for the stereo camera SCam_c^{per}, since the rotation parameters of the involved views slightly differ. This issue is indicated by red color in Table 7.3.

However, Table 7.3 provides no information about the error in localizing Cam₁^{omni} and Cam₂^{omni} with respect to Cam₀^{omni}. In order to get an indication about that issue, a cycle homography, denoted as \mathbf{H}_{err} , is computed by incorporating all the calibration information calculated. Initially, the actual cycle starts at Cam₀^{omni} computing the homography to Cam₁^{omni}, based on this it computes the homography to Cam₂^{omni} and finally it closes the circle by computing the homography back to Cam₀^{omni}. If the calibration process had provided perfect results, the calculation as seen in Equation 7.8 should have led to the identity matrix.

$$\begin{aligned}
\mathbf{H}_{\text{err}} &= \underbrace{(d\mathbf{H}i_-^{\text{per}})^{-1} \cdot \hat{\mathbf{H}}\mathbf{c}_{0b}^{\text{per}} \cdot d\mathbf{H}i_+^{\text{per}}}_{\text{Homography from Cam}_2^{\text{omni}} \text{ to Cam}_0^{\text{omni}}} \cdot \underbrace{(d\mathbf{H}i_-^{\text{per}})^{-1} \cdot \hat{\mathbf{H}}\mathbf{c}_{2c}^{\text{per}} \cdot d\mathbf{H}i_+^{\text{per}}}_{\text{Homography from Cam}_1^{\text{omni}} \text{ to Cam}_2^{\text{omni}}} \\
&\quad \cdot \underbrace{(d\mathbf{H}i_-^{\text{per}})^{-1} \cdot \hat{\mathbf{H}}\mathbf{c}_{1a}^{\text{per}} \cdot d\mathbf{H}i_+^{\text{per}}}_{\text{Homography from Cam}_0^{\text{omni}} \text{ to Cam}_1^{\text{omni}}} \tag{7.8} \\
\mathbf{H}_{\text{err}} &= \begin{bmatrix} \mathbf{R}_{\text{err}} & \mathbf{T}_{\text{err}} \\ \mathbf{0} & 1 \end{bmatrix}
\end{aligned}$$

The evaluation of this equation by employing the afore-computed numerical values leads to $\mathbf{R}_{\text{err}} \approx \mathbf{I}$ and $|\mathbf{T}_{\text{err}}| > 5mm$. This reveals that the utilized method, for calibration of three cameras, is quite accurate in terms of rotation but not very satisfactory with respect to translation. While one can now assume good results from the stereo matching stage, the triangulation and the merge process of the depth maps leads us to expect perceptible errors.

7.3 Virtual Camera Setup

Since the results of the intrinsic generic camera model estimation stage (Table 7.1) are quite similar to the generic parameters used for simulation (Table 4.2) and distinct requirements are not claimed, configuration no. 7 is chosen for initial practical tests. The calculated parameters can be reviewed in Table 7.4 to Table 7.7. In contrast to configuration no. 7, the parameters for the vertical resolution are determined according to the over-sampling strategy. This does not have any influence on the accuracy but causes the final merged map to be *more dense* as can be seen in Section 7.5. The variations applied have been marked with blue color in the tables.

Table 7.4: Calculated fields of view FOV_h for the virtual cameras Cam_j^{rect} according to the variety of configurations

No.	Cam_i^{omni}	Cam_j^{rect}	$b [m]$	$FOV_{hr} [deg]$	$Z_{min} [m]$	$FOV_h [deg]$
7	kannala9	epipolar equidistance	0.10	60.0	0.75	75.40

As can be seen in Table 7.5, indicated by red color, the parameter d_{max} for the maximum disparity has been slightly adapted in order to comply with available procedures for stereo correspondence computation. Since for the stereo matching process the standard algorithm from OpenCV *StereoBM*¹ is employed, which claims d_{max} to be a power of two, it has been changed from 65 to 64.

Table 7.5: Calculated virtual pixel size ratios and image sizes

No.	$FOV_h [deg]$	p_x^{rect}/p_x^{omni}	p_y^{rect}/p_y^{omni}	height	width	d_{max}
7	75.40	0.4140	0.4577	1680	636	64

Table 7.6: Intrinsic parameters of the virtual cameras Cam_j^{rect}

No.	Camera	α_x	α_y	c_x	c_y
7	Cam_j^{rect}	483.71	534.76	318.50	839.5

Table 7.7: Pixel dimensions of the virtual cameras Cam_j^{rect}

No.	Camera	$p_x [m]$	$p_y [m]$
7	Cam_j^{rect}	5.58×10^{-6}	5.05×10^{-6}

¹ http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

7.4 Software Realization

The previously described algorithm has been implemented in C++ as part of the professorship's proprietary computer vision framework. A screenshot of the running application can be reviewed in Figure 7.6.

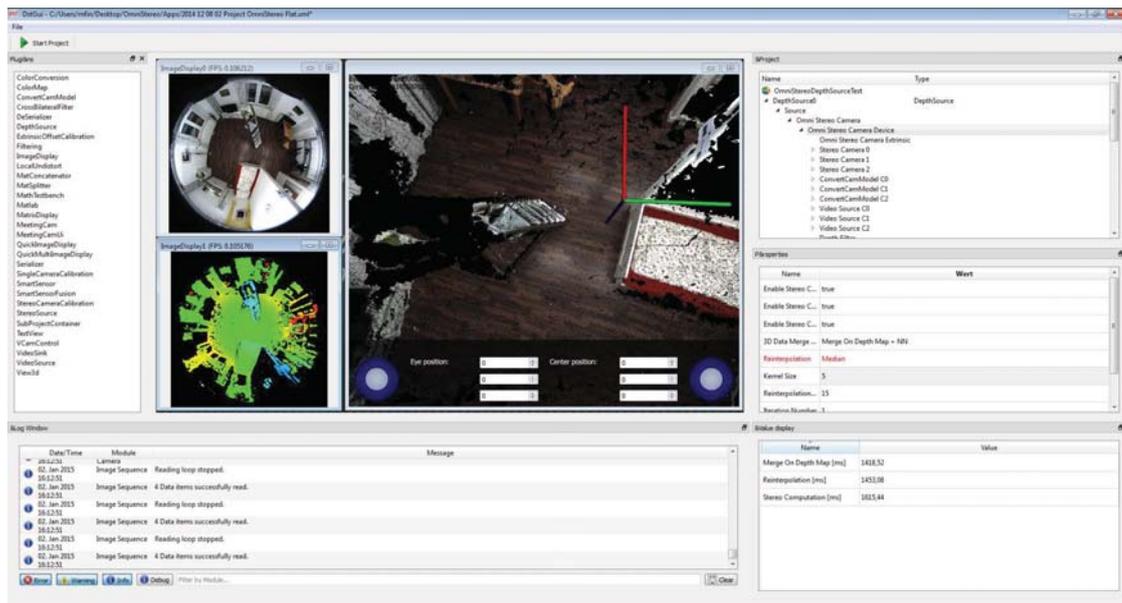


Figure 7.6: Computer vision framework with running spherical stereo project

The whole processing chain has been realized self-dependent.

7.5 Experimental Results

For initial test measurements the stereo vision demonstrator, as outlined in Figure 7.1, was installed at the ceiling of the professorships lab apartment. The objective was to capture RGB-D data for nearly the whole living space.

7.5.1 Qualitative Assessment

Figure 7.7 outlines a calculated hemispherical depth map with respect to a reference image. The underlying partial depth maps have been determined by means of the afore-mentioned block matching procedure, using a block size of 15×15 pixels. The subsequent merge has been done with an ordinary nearest neighbour method.

By assessing the depth map of Figure 7.7 closer, one can recognize small gaps in the map where not at least two views overlap. That issue results from the under-sampling strategy and the forward mapping process. For successive processing stages, e.g. for

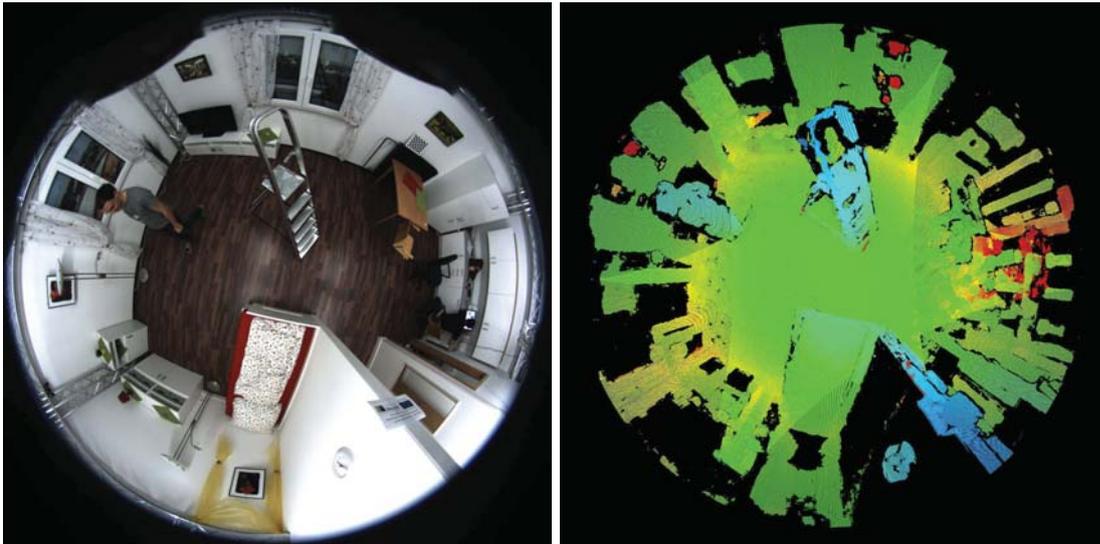


Figure 7.7: Reference image with calculated hemispherical depth map

generating a dense point cloud, it might be sensible to *close* these gaps. A first endeavour was made by Kaden [51], in order to investigate this issue. Up to now the application of a post processing filter, more precisely a median filter, has shown the best improvement. Figure 7.8 demonstrates the impact of a median filter stage, which is parametrized to a kernel size of 5×5 .

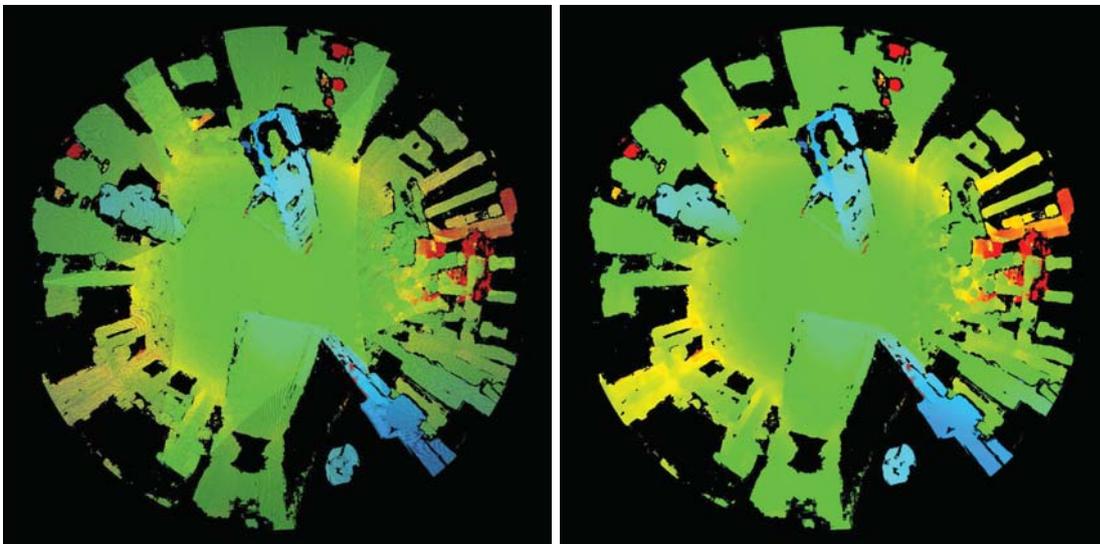


Figure 7.8: Hemispherical depth map without (left) and with (right) post processing median filter of kernel size 5×5

Small gaps in the depth map can be considered as *impulsive noise*, also denoted as *salt and pepper noise*. Thus, the median filter is appropriate for cancelling out these minor areas with no depth information by replacing it with the *median depth value* of its vicinity. However, pixels representing distance information are not impacted.

Subsequently, the calculated hemispherical depth map can be utilized in order to re-project the reference image of Figure 7.7 to a point cloud, as done in the simulation of Section 6.5. An impression of how the whole hemispherical vicinity is observed can be obtained by reviewing Figure 7.9.



Figure 7.9: Re-projection of reference image

A closer assessment with respect to certain details of the point cloud reveals remaining challenges that have to be engaged: Figure 7.10 shows an example where a table-leg is partially observed by two stereo cameras. In the merged point cloud both parts do not appear at the same position. Their associate distance information, calculated from different stereo cameras, differs with respect to the reference camera. As stated in Section 7.2.2.5, an accumulative error of $|\mathbf{T}_{\text{err}}| > 5\text{mm}$ has been calculated for the afore-determined camera origins $\mathbf{C}_0^{\text{omni}}$, $\mathbf{C}_1^{\text{omni}}$ and $\mathbf{C}_2^{\text{omni}}$. Thus, this uncertainty results in deficient baselengths for the underlying stereo cameras and, consequently, in erroneous distance measurements. For this reason, a calibration procedure, that performs an optimization amongst all three views, has to necessarily replace the currently applied method.

7.5.2 Performance Measurements

The implementation of the concept for hemispherical depth map generation has not yet been optimized for real-time processing. It can rather be considered as straight forward implementation: Firstly, the re-projection of the partial disparity maps by means of the epipolar equidistant method requires the calculation of trigonometric functions and is hence time consuming. Secondly, the fusion process of *each* partial depth map is

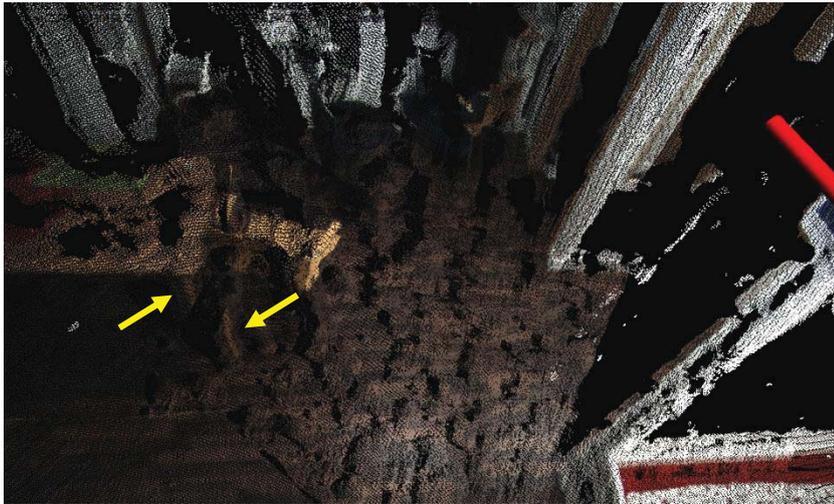


Figure 7.10: Example case for an erroneous fusion result

done by a forward mapping process. Further non-optimized parts include the nearest neighbour interpolation stage as well as the post processing filter. Besides the need for improvement with respect to practical implementations, Chapter 8 will introduce optimization strategies that will accelerate the current algorithm significantly. Initial performance measurements have been realized on an Intel(R) Core (TM) i7-3520M CPU @ 2.90 GHz and are presented in Table 7.8. The actual timing values have been rounded sensibly. The obtained speed in terms of Frames Per Second (FPS) can be considered as *near real-time*.

Table 7.8: Performance measurements

Stereo Blocksize	Depth map merge	Post Processing	Stereo computation [ms]	Re-projection [ms]	Depth map merge [ms]	Post filter [ms]	Total fps
11 × 11	nearest neighbour	no post processing	750	100	480	0	0.75
15 × 15	nearest neighbour	no post processing	780	100	470	0	0.74
21 × 21	nearest neighbour	no post processing	800	100	450	0	0.74
11 × 11	nearest neighbour	median filter 5 × 5	750	100	480	150	0.67
15 × 15	nearest neighbour	median filter 5 × 5	780	100	470	130	0.67
21 × 21	nearest neighbour	median filter 5 × 5	800	100	450	120	0.68

Chapter 8

Discussion and Outlook

The so far achieved results with respect to the novel concept for spherical stereo vision are revisited in this final chapter. A discussion of the achievements is presented in Section 8.1. Based on this assessment, further need for research is derived.

Furthermore, the idea of omnidirectional stereo vision, as investigated in this thesis, is compared to the approach by Li [67], and moreover, to a solution employing two hemispherical views in Section 8.2.

The end of this thesis is constituted by the presentation of a demo application that utilizes spherical depth data in Section 8.3, and last but not least, by summarizing remarks in Section 8.4.

8.1 Discussion of the Current Results and Further Need for Research

The major objective of this thesis was to investigate a new principle in order generate a full hemispherical depth map by means of stereo vision. Based on known methods, as outlined in Chapter 5, an alternative approach to Li [67] has been presented that shall fulfil the following requirements:

1. A full hemispherical depth map shall be generated without the need of applying a merging procedure of high computational effort.
2. The proposed approach shall generate a full hemispherical depth map by employing fewer correspondence operations.

In the following, the investigations that have been performed in order to achieve the afore-mentioned goals shall be reflected and assessed. In addition, proposals for further research activities and for algorithm optimization are presented.

8.1.1 Assessment of the Geometrical Camera Configuration

The proposed geometrical camera configuration comprises a three-view stereo vision setup of equilateral triangle configuration. This special characteristic leads to certain effects that shall be discussed briefly.

Availability

According to a performed search, this special camera constellation is not protected by any patent claims. Hence it can be employed without restriction.

Forward Mapping

The hemispherical depth information, calculated by means of this setup, is obtained by mapping each of the partial depth maps to the reference camera. For $\text{Cam}_{0a}^{\text{rect}}$ this can be achieved by a LUT operation, since its origin coincides with $\text{Cam}_0^{\text{omni}}$. For $\text{Cam}_{1b}^{\text{rect}}$ and $\text{Cam}_{2c}^{\text{rect}}$ this operation is performed via *forward mapping* (cf. Chapter 4), which causes the effect of gaps within the target map - cf. Section 7.5. Thus, a post processing filter becomes reasonable.

Since each stereo camera is currently characterized by another origin, the author suggests referring *two* stereo cameras ($\text{SCam}_a^{\text{rect}}$ and $\text{SCam}_b^{\text{rect}}$) to $\text{Cam}_0^{\text{omni}}$ in contrary to the presented approach. Perspectively, just one partial depth map will have to be processed using forward mapping. Furthermore, this improvement will reduce the employment of the post processing filter stage to a minimum, i.e. to the depth information of $\text{SCam}_c^{\text{rect}}$.

Mapping Error

As discussed in Chapter 6, the applied geometrical configuration and the method for depth map fusion give rise to a mapping error $\Delta \hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ of maximal 1.1 pixels down to 0.9 pixels for $b = 10\text{cm}$ and 1.4 pixels down to 1.0 pixels for $b = 15\text{cm}$, assuming a measurement range of 0.5m to 5.0m . This aspect indicates that the presented approach with the underlying virtual camera configuration is most appropriate for small baselengths. The more the baselength is increased, and thereby the cameras FOV, the more this error will grow. As stated in Chapter 6, yet this uncertainty can be reduced significantly in case the over-sampling strategy is applied.

However, a reduced spacial resolution of depth maps, due to image block based correlation methods for stereo computation, diminishes or completely cancels out this effect. Dedicated investigations are necessary in order to confirm this statement.

Occlusion Effects

Due to the different locations of the 3D data generating devices, occlusions may occur. This case is sketched in Figure 8.1. Imagine a measurement camera calculates a distance with respect to a background object (red arrow on the left) that appears occluded by another object, with respect to the viewpoint of the reference camera. In that case, the forward mapping procedure might result in a wrong distance information with respect to the reference camera (red arrow on the right). The correct information would be the distance to the foreground object (green arrows).

Since the utilized baselength is small compared to the camera-object-distance, this occurrence should be on the fringes. Anyway, stereo measurement devices correlate images captured from different positions principally, which means that this issue belongs to the basic characteristics of a stereo vision system. However, during the practical tests performed so far, this effect has not been observed.

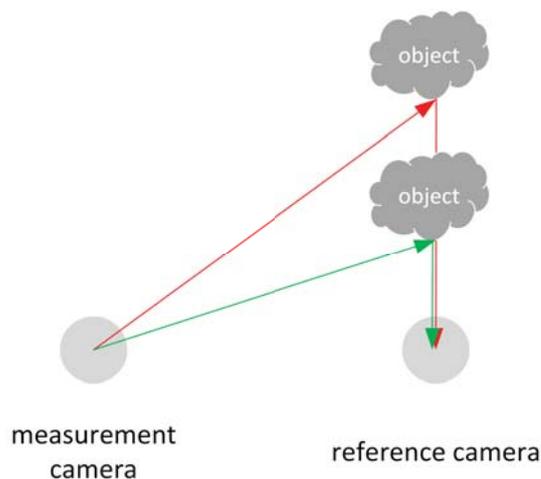


Figure 8.1: Occlusion case that might appear if a depth measurement information obtained by one imaging device (measurement camera) is mapped to another device (reference camera)

8.1.2 Assessment of the Depth Map Computation

For generating a full hemispherical depth map three narrow-angled stereo cameras have been parametrized as part of the chosen geometrical setup. The achievements of this configuration as well as further suggestions for improvement shall be assessed now.

Parametrization of the Narrow-Angled Views

The presumption was that each view ought to cover a horizontal field of view of 60° approximately - cf. Chapter 5 Section 5.3.1. Appropriate investigations in Chapter 6 identified the necessity of a larger FOV for each rectified image, since the maximum disparity value for guaranteeing a minimum measurable distance had to be included into the calculations. The actual values for the current configuration have been determined to 75.80° ($b = 10\text{cm}$) and 83.20° ($b = 15\text{cm}$) provided an epipolar equidistance rectification model. The cylindrical projection method has been declined since it turned out to generate target images of lower resolution and hence less accuracy compared to the alternative method. However, Abraham et al. [3] describe this projection principle to be beneficial for stereo matching performance. Thus it could be beneficial to perform a competitive evaluation for rectification methods with respect to their stereo matching performance in future.

Considering a real-time implementation, the usability of a rectification method that does not employ any trigonometric functions should be investigated. Thereby, the performance for the triangulation process could be accelerated. As an example the stereographic rectification method has been briefly introduced in Section 5.2.

Stereo Correspondence Reduction

For an estimation of the currently achieved savings in terms of stereo correspondence operations the author assumes that Li [67] chooses the same sampling rate as presented in Chapter 6. As mentioned in Chapter 5 he correlates two full images, while the author performs this operation on three partial images. Table 8.1 outlines the current achievements.

Table 8.1: Savings in terms of stereo correspondence operations

No.	Cam _i ^{omni}	b [m]	Z _{min} [m]	FOV _h [deg]	height	width	savings [%]
5	equiangular	0.10	0.75	75.40	1344	699	30.0
6	equiangular	0.15	0.75	83.20	1285	763	23.0
7	kannala9	0.10	0.75	75.40	1254	636	37.2
8	kannala9	0.15	0.75	83.20	1201	702	30.6

It is important to realize that the achievements of Table 8.1 highly depend on the parametrization, above all, of the baselength and of the minimum allowable measurement distance.

It can be concluded that the strategy of employing narrow-angled stereo cameras diminishes the number of necessary correspondence operations considerably. However, the high redundancy in the center of the observed scene, where each distance is measured three times, can be considered as disadvantageous. A perspective for further improvement is developed in the following:

Further Optimization of the Rectified Views

By reviewing the strategy for hemispherical depth map computation in Section 6.5 the following fact becomes obvious: While the rectified views seem to hardly cover any scene point twice at the edges of the omnidirectional view, they see the observed vicinity three times at the center. This issue is visualized in Figure 8.2. At this point the presented method yet bears a high redundancy.

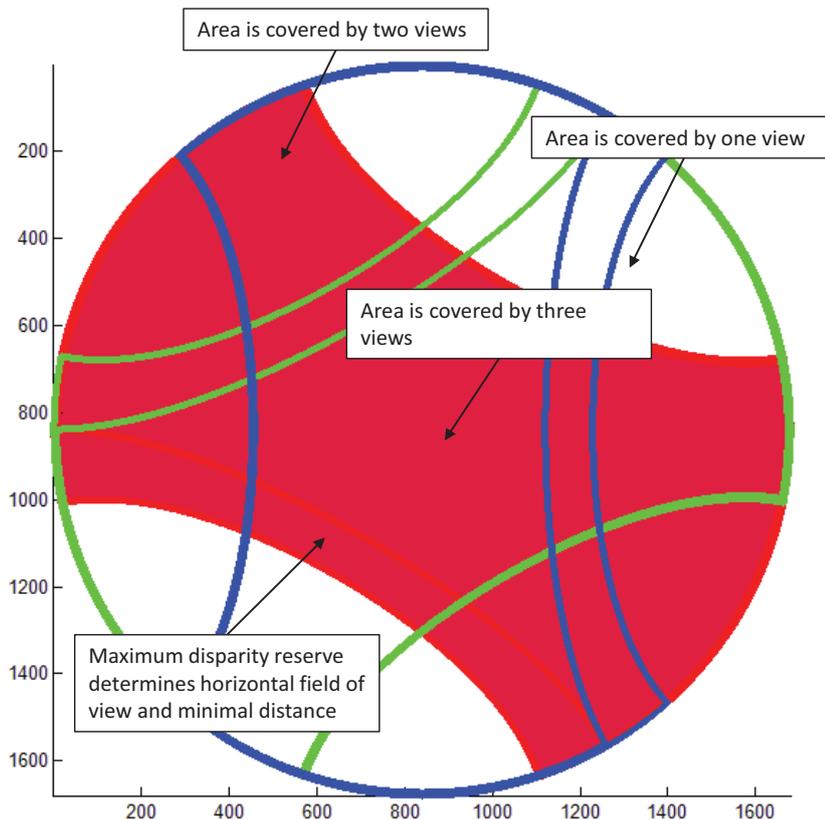


Figure 8.2: Overlapping depth maps (red, green and blue) with highlighted area of maximal disparity

For a further optimization, two major strategies can be taken into consideration:

1. If we assume the sensor to be mounted at the center of a rooms ceiling, it is presumable that the measuring device could be adjusted to guarantee a smaller minimal measurement distance in the middle ($\phi_l = \pi/2$) than at the edges ($\phi_l = 0$ or $\phi_l = \pi$). For example when the sensor is located $2.5m$ above the floor it should

allow a minimal distance of $0.5m$ in order to recognize a person that stands beneath and is $2m$ in height. At the edges the minimal distance could be parametrized to $0.75m$ for example. The basic idea is to enlarge the disparity reserve in the middle and decrease it at the edges. By that the presented concept could also work properly in environments with reduced ceiling heights.

2. Since realistic requirements naturally demand the algorithm to process as fast as possible, it makes sense to optimize the stereo processing itself with respect to the partial rectified views: At the top of the rectified images (first image row) as well as the bottom (last image row), the originally parametrized FOV for 3D calculation should be preserved which means that the whole image line is considered by the stereo correspondence algorithm. When approaching the center row of the image from either direction, the FOV for stereo processing could be reduced. By means of that, an almost non-recurring coverage of each scene point could be achieved. Hence, the computational load for stereo processing as well as the forward mapping procedure would further decrease considerably.

8.1.3 Assessment of the Depth Measurement Error

By employing narrow-angled views for stereo processing the variance of determined depth measurement uncertainties, within the same partial depth map, is constrained at close range. The example of Chapter 6, Figure 6.18 plotted the error values $|\delta Z_{\max}^{\text{omni}}|$ for the overlapping partial depth maps of configuration no. 6 measuring a distance of $2.5m$. The minimum error to be observed has been determined to $\approx 3.8cm$, and the maximum error to $\approx 5.5cm$ respectively. The observed differential of $\approx 30\%$ leads the author to the conclusion that a merge method of high computational effort as used by Li [67] has not to be implemented. Because of the small range of the error values, a simple averaging method is adequate. However, until now methods for merging depth maps have not been investigated sufficiently.

Comparing the largest error value of $|\delta Z_{\min}^{\text{omni}}| \approx 4.3cm$ in the right plot of Chapter 6, Figure 6.19 (approach Li [67]) with the largest error value $|\delta Z_{\max}^{\text{omni}}| \approx 5.5cm$ in the left plot of Chapter 6, Figure 6.18 (authors approach), one can conclude the following: Without applying any depth averaging method, the presented approach obtains for the configuration in discussion at least approximately 80% of the accuracy achievable by competitors. The measurement errors can hence be considered as *comparable*. This statement holds true if the following assumption is made: The significant accuracy of a 3D sensor is specified by the worst case error value that occurs in its observation area (for a certain distance).

8.1.4 Assessment of the Spherical Stereo Vision Demonstrator

As part of the thesis a demo setup has been assembled and presented in Chapter 7. The current status of this development and open points are briefly summarized in the following:

Calibration Results

For the intrinsic calibration stage a state-of-the-art generic camera model has been applied. The utilized extrinsic calibration method (based on standard two-view geometry) turned out not to provide the required precision. While the mutual orientation of the cameras could be determined quite well, the translational offset between them led to a cumulative error of about $5mm$. This gives rise to erroneous measurements due to deficiently determined baselengths, as could be seen in Chapter 7. Any MATLAB[®] simulation the author performed, did not reveal inaccuracies of that kind. Hence, for an accurate alignment of the partial point clouds, a more sophisticated calibration method will have to be applied in future necessarily. Advanced methods compute a trifocal tensor that, equivalently to the fundamental matrix for two views, directly describe the inherent relationship between three views. This issue represents a further field for research.

Results of the Depth Map Computation

As a result of Chapter 7, a full hemispherical depth map was presented, obtained by realizing the author's concept. The reprojection of this 3D data set generated a considerable point cloud. However, since the calibration process has yet to be optimized for accurate point cloud alignment, the validation of achieved measurements versus computed error values could not be carried out so far.

Computational Performance

As stated in Chapter 7 the current performance, which has been determined to slightly less than $1FPS$ on a standard office PC hardware, can be considered as *near real-time*. Nevertheless, the current implementation has not yet been optimized, because the thesis focuses on the principle investigation of the new method.

In comparison to Li [67], a real-time implementation of the author's concept should be fast due to the following facts:

- Currently the necessary stereo correspondence operations for obtaining a full hemispherical depth map could be reduced to approximately 70%. A further reduction is possible as already stated.

- A computational costly merge process based on calculating multiple trigonometric functions is not necessary.

However, the finally achievable performance will primarily depend on the underlying hardware architecture.

8.2 Review of the Different Approaches for Hemispherical Depth Map Generation

In the following, the most significant facts that distinguish both three-view configurations in discussion are reflected. A subsequent comparison with a two-view approach highlights their practical relevance.

8.2.1 Comparison of the Equilateral and the Right-Angled Three-View Approach

As a matter of fact, both approaches are appropriate in order to generate a full hemispherical depth map. Since it averages two full depth maps, the right-angled approach generates more accurate results, with the drawback of a more complex data fusion process. On the other hand, the equilateral approach produces error values that are slightly more inaccurate. However, the latter concept has high potential for showing more performance, provided that the afore-proposed optimizations have been realized. Considering a realization on an embedded system of either concept, one can state additional notions:

- The right-angled camera configuration is protected by at least U.S. patent claims, while the equilateral configuration is not - according to the current knowledge of the author.
- An embedded sensor would be slightly more compact with a surface area of $\approx \frac{1}{2}b^2$ using the equilateral approach compared to an area of $\approx \frac{\sqrt{3}}{4}b^2$ by means of the right-angled concept.
- In contrast to the right-angled approach, the equilateral method is suitable for small baselengths, since the mapping error increases along with it.
- The practical demonstrator as shown in Chapter 7 has been assembled by means of high-performance fish-eye lenses, each providing support for up to 5 megapixel camera resolution. It is yet ambiguous how both concepts perform, if fish-eye lenses complying to the S-mount standard are employed. Related characteristics of comparatively lower performance like *maximal aperture*, *vignette*, *distortions*, *chromatic aberration*, *digital noise* (due to smaller dimensions of the lens and the sensor), *resolution* etcetera may influence the result negatively.

8.2.2 Review of the Three-View Approach in Comparison with the Two-View Method

In Chapter 5 the two-view method employing hemispherical views turned out to cover a huge FOV as well. In theory, this approach could provide a full hemispherical depth map, if measurement uncertainties were not a concern. According to Chapter 5, Figure 5.8 such a configuration is characterized by a rising depth error along the baseline, but a constantly low one orthogonal to it.

Figure 8.3 depicts two comparative plots for error estimation $|\delta Z_{\max}^{\text{omni}}|$ of the authors three-view approach, and a two-view method respectively. The value coding colors have been scaled appropriately in order to cover a range starting from the minimum error ($|\delta Z_{\max}^{\text{omni}}| = 0.0562m$) up to twice the maximum error ($|\delta Z_{\max}^{\text{omni}}| = 2 \cdot 0.0788m$), as observable in the left plot.

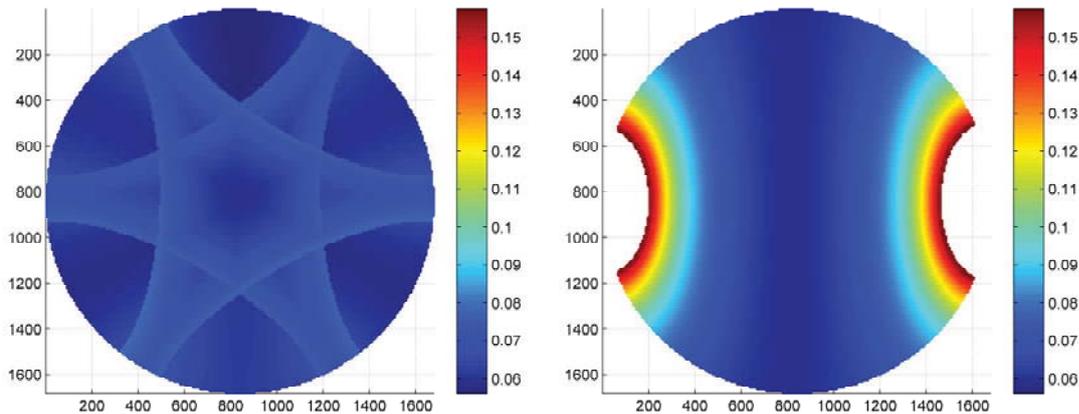


Figure 8.3: Maximum distance error $|\delta Z_{\max}^{\text{omni}}|$ for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 10cm$; Left: the author's three-view approach ($\text{FOV}_H = 75.40^\circ$), Right: two-view approach with restricted FOV that complies with an error smaller than two times of the maximum error observable in the left plot

Subsequently, one can interpret the right plot as follows: If an omnidirectional, two-view 3D sensor is designed and specified to a maximum measurement uncertainty twice as high compared to the author's approach, its circular FOV is restricted to 132° for the current configuration. The more coarse the sensor is admitted to measure in the worst case, the more its FOV converges to a full hemispherical depth map. Furthermore, as soon as a depth value averaging method for the three-view method is introduced in addition, the divergence between both configurations will further increase - cf. Chapter 6.

However, the distinct variant of choice finally depends on certain requirements for a distinct scenario of application. Nevertheless, for applications that require a full hemispherical depth map, e.g. in order to observe a single room from floor to ceiling by employing one single sensor, a two-view approach is not sufficient.

Last but not least, a descriptive example shall be assessed: Figure 8.4 outlines the reprojection of artificial 3D data, as presented in Chapter 6, that is captured by an omnidirectional two-view camera configuration. Since the sensor is located at a height of $2.5m$ above the floor, it is parametrized to allow a minimum measurable distance of $0.5m$. Furthermore, the sensor is *mounted* in a diagonal orientation. Examining this plot, the following characteristics can be extracted:

- The lower left corner as well as the upper right one of the artificial 3D environment demonstrate the best accuracy since they are located orthogonal to the sensors baseline.
- The lower right corner reveals distortions that become larger the more the epipole is reached.
- In addition to distortions, the upper left corner is characterized by a blind spot. This issue originates from the sensors parametrized minimal measurable distance, and hence the adjusted maximal disparity value, which diminishes the observable area. For this concrete scenario the omnidirectional region, that can be perceived by the sensor, is restricted vertically to a range from $0m$ up to $\approx 1.75m$ above the floor. For example, the head of a person with $2m$ in height would no longer be visible in that blind spot area.

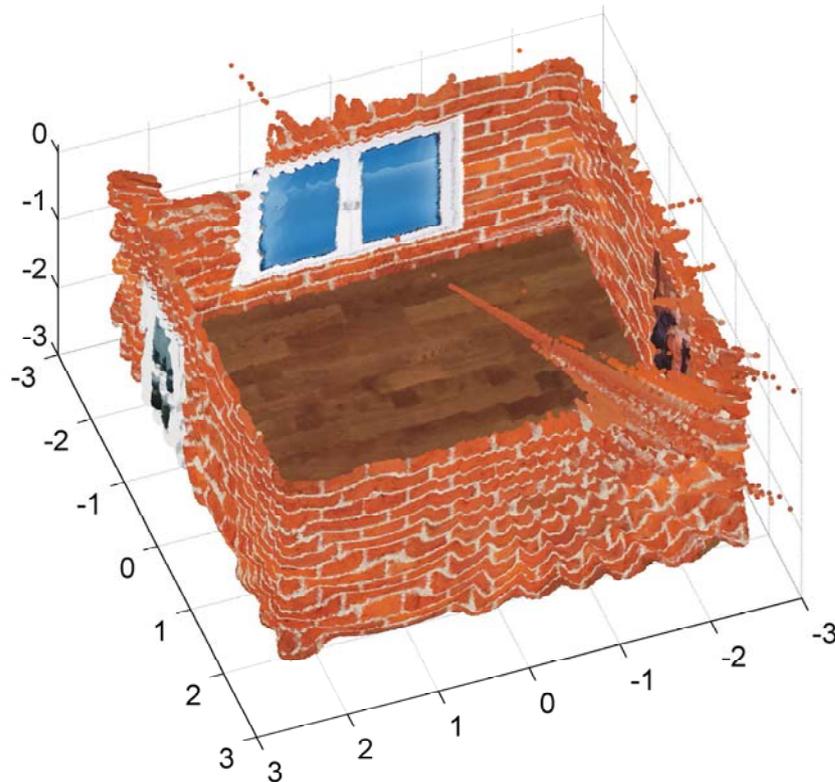


Figure 8.4: 3D reprojection of an artificial home environment observed by a two-view stereo vision setup with hemispherical views ($b = 10cm$)

8.3 A Sample Algorithm for Human Behaviour Analysis

For a convenient closure of this thesis, a detour to a practical application is made. In Figure 8.5 a system is depicted that applies an omnidirectional 3D sensor for human behaviour analysis. The algorithm as outlined has been published by the author in a scientific paper [26].

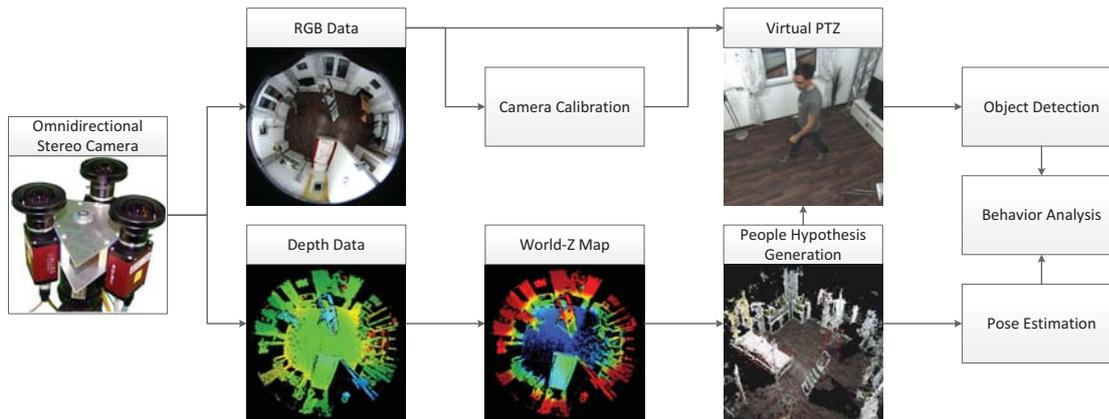


Figure 8.5: Structure of the proposed algorithm

In the following each single step is briefly explained.

Omnidirectional RGB-D Data Generation

The omnidirectional stereo camera provides hemispherical RGB-D data. Since the depth measuring device is extrinsically calibrated, with respect to a world coordinate system on the floor of the observed scene, a world-z map can be calculated. Actually, this specifies a map of distance values representing the height of each scene detail over ground.

People Hypothesis Generation

Position information about present persons in a room constitutes the precondition for a successive analysis of their behaviour. Robust person detection algorithms that employ 3D data have recently been proposed [99]. In this study hypotheses about moving foreground objects, e.g. persons, are generated by applying a Gaussian Mixture Model (GMM) to the world-z map. In order to localize the potential persons with respect to world coordinates, points belonging to these hypotheses are projected into a virtual overhead view on which separate clusters are extracted. The detected persons are defined by the center, the orientation and the expansions of each cluster (Figure 8.6). Details can be reviewed in [99].

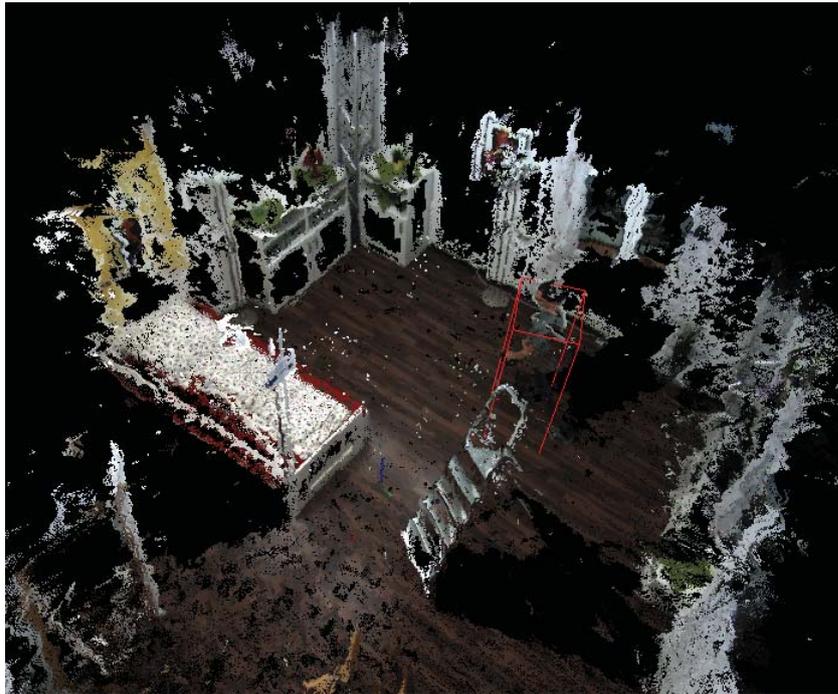


Figure 8.6: People hypothesis generation for 3D localization

Virtual Perspective View

Subsequently, a vPTZ camera is employed in order to identify hypotheses as persons by means of a people recognition algorithm, such as presented by Dollar et al. [20]. This stage complies with the approach that has been presented in Chapter 4, Section 4.6.

Behaviour Analysis

In order to analyse a person's behaviour, further processing steps are performed. The person's general pose, i. e. *standing*, *sitting* and *lying*, is determined on the basis of the world points belonging to the person. Moreover, objects the person is interacting with are detected on the perspective view of the vPTZ camera. Based on this generated data, conclusions about a person's behaviour, e. g. sleeping behaviour, can be drawn.

8.4 Closing Remarks

During the elaboration of this thesis, a couple of aspects related to spherical stereo vision have been investigated. First of all, the primary objective of verifying the feasibility of the proposed concept could be achieved. A measurement system has been developed, which has great potential for future application fields in industry, security in public spaces as well as home environments. Furthermore, the amount of possible research topics, which has been generated in the course of this work, is considerable and thus provides a scientific prospect as well. Anyway, I hope the perusal of this thesis has been both informative and interesting to the reader.

Appendix A

Relevant Mathematics

This part outlines important mathematics and derivations.

A.1 Cross Product by Skew Symmetric Matrix

The cross product of two vectors $\mathbf{a} = (a_1, a_2, a_3)^T$ and $\mathbf{b} = (b_1, b_2, b_3)^T$ can be represented as dot product:

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_{\times} \mathbf{b} = \left(\mathbf{a}^T [\mathbf{b}]_{\times} \right)^T \quad (\text{A.1})$$

with

$$[\mathbf{a}]_{\times} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \quad (\text{A.2})$$

A.2 Derivation of the Quantization Error

A depth Z for stereo vision devices can be computed as follows:

$$Z = b \cdot \frac{f}{ad} \quad (\text{A.3})$$

An depth measurement error can be specified as *absolute depth error* δZ :

$$\delta Z = \delta Z_b + \delta Z_f + \delta Z_d = \frac{f}{d} \delta b + \frac{b}{d} \delta f - \frac{bf}{d^2} \delta d \quad (\text{A.4})$$

According to [17] this expression can be simplified to:

$$\delta Z \approx \delta Z_d = -b \cdot \frac{f}{ad^2} \delta d = \left(b \cdot \frac{f}{ad} \right)^2 \cdot -\frac{1}{bf} \delta d \quad (\text{A.5})$$

Finally the expression can be reduced to:

$$\delta Z \approx -\frac{Z^2}{bf} \delta d \quad (\text{A.6})$$

A.3 Derivation of the Statistical Distribution of Quantization Errors

From Equation 3.14 it is known that a disparity d can be computed as:

$$d = x_{\text{img},l} - x_{\text{img},r} \quad (\text{A.7})$$

According to [9, 17, 100] the variables $x_{\text{img},l}$ and $x_{\text{img},r}$ can be considered as mutual independent random variables of uniform distribution in the interval $[-\frac{a}{2}, +\frac{a}{2}]$.

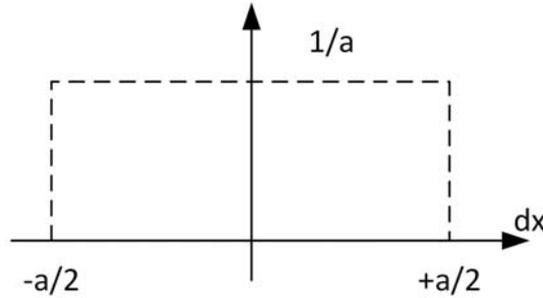


Figure A.1: Random variables $x_{\text{img},l}$ and $x_{\text{img},r}$ of uniform distribution

The disparity error δd can hence be obtained from the quantization noise $\delta x_{\text{img},l}$ and $\delta x_{\text{img},r}$ for each pixel in the left and right image:

$$\delta d = \delta x_{\text{img},l} - \delta x_{\text{img},r} \quad (\text{A.8})$$

The probability $p(\delta d)$ is the convolution of the probabilities $p(\delta x_{\text{img},l})$ and $p(\delta x_{\text{img},r})$:

$$p(\delta d) = p(\delta x_{\text{img},l}) * p(\delta x_{\text{img},r}) \quad (\text{A.9})$$

$$p(\delta d) = \begin{cases} \frac{a+\delta d}{a^2} & \text{if } -a \leq \delta d \leq 0 \\ \frac{a-\delta d}{a^2} & \text{if } 0 \leq \delta d \leq +a \end{cases} = \frac{a - |\delta d|}{a^2} \quad (\text{A.10})$$

The probability $p(\delta Z)$ can be computed as follows:

$$p(\delta Z) = \int_{-\infty}^{+\infty} p(\delta Z|Z)p(Z)dZ \quad (\text{A.11})$$

where $p(\delta Z|Z)$ is the conditional probability of δZ implied a certain depth value Z that can be measured with the probability $p(Z)$.

It is known that the depth error δZ can be computed as:

$$\delta Z = -\frac{Z^2}{bf}(\delta x_{\text{img,l}} - \delta x_{\text{img,r}}) = -\frac{Z^2}{bf}\delta d \quad (\text{A.12})$$

which means that the disparity value δd is a monotone function of δZ :

$$\delta d = -\frac{bf}{Z^2}\delta Z = -\frac{a^*Z^*}{Z^2}\delta Z = -\frac{aZ^*}{2Z^2}\delta Z \quad (\text{A.13})$$

Since Equation A.13 is a monotone function, the probability $p(\delta Z|Z)$ can be modelled as follows:

$$\begin{aligned} p(\delta Z|Z) &= p(\delta d) \cdot \left| \frac{d(\delta d)}{d(\delta Z)} \right| = \frac{a - |\delta d|}{a^2} \cdot \frac{aZ^*}{2Z^2} \\ &= \left[1 - \frac{|\delta d|}{a} \right] \cdot \frac{Z^*}{2Z^2} \end{aligned} \quad (\text{A.14})$$

Using Equation A.13 this expression can be simplified further:

$$p(\delta Z|Z) = p(\delta d) \cdot \left| \frac{d(\delta d)}{d(\delta Z)} \right| = \frac{Z^*}{2Z^2} - \left(\frac{Z^*}{2Z^2} \right)^2 |\delta Z| \text{ with } 0 \leq \delta Z \leq \frac{Z^*}{2Z^2} \quad (\text{A.15})$$

Now Equation A.11 can be rewritten as follows:

$$p(\delta Z) = \int_{-\infty}^{+\infty} \left[\frac{Z^*}{2Z^2} - \left(\frac{Z^*}{2Z^2} \right)^2 |\delta Z| \right] p(Z)dZ \quad (\text{A.16})$$

According to [17], the measured depth Z can only be randomly distributed over a limited range between Z_{\min} and Z_{\max} :

$$Z_{\min} \leq Z \leq Z_{\max} \quad (\text{A.17})$$

Furthermore from Equation A.15 it holds true that:

$$Z >= \sqrt{\frac{1}{2}Z^* \cdot |\delta Z|} \quad (\text{A.18})$$

For Equation A.16 it follows that:

$$p(\delta Z) = \int_{\max\left(Z_{\min}, \sqrt{\frac{1}{2}Z^* \cdot |\delta Z|}\right)}^{Z_{\max}} \left[\frac{Z^*}{2Z^2} - \left(\frac{Z^*}{2Z^2} \right)^2 |\delta Z| \right] p(Z)dZ \quad (\text{A.19})$$

From Equation A.19 it can be concluded that the distribution of the depth error can be outlined if a certain depth measurement distribution is known or assumed, e.g. uniform distribution.

Since the disparity d can be assumed to be distributed uniformly in the interval $[-a, +a]$ as mentioned above, the distribution of Z can be assumed to be distributed uniformly in the interval $[Z_{\min}, Z_{\max}]$:

$$\begin{aligned} Z_{\min} &= Z - \frac{1}{2} \frac{Z^2}{bf} \delta Z \\ Z_{\max} &= Z + \frac{1}{2} \frac{Z^2}{bf} \delta Z \end{aligned} \quad (\text{A.20})$$

Finally Equation A.19 can be solved for the intervals of Z separately.

When $0 \leq |\delta Z| \leq \frac{2Z_{\min}^2}{Z^*}$ it can be written:

$$p(\delta Z) = \int_{Z_{\min}}^{Z_{\max}} \left[\frac{Z^*}{2Z^2} - \left(\frac{Z^*}{2Z^2} \right)^2 |\delta Z| \right] p(Z) dZ \quad (\text{A.21})$$

When $\frac{2Z_{\min}^2}{Z^*} \leq |\delta Z| \leq \frac{2Z_{\max}^2}{Z^*}$ it can be written:

$$p(\delta Z) = \int_{\sqrt{\frac{1}{2}Z^* \cdot |\delta Z|}}^{Z_{\max}} \left[\frac{Z^*}{2Z^2} - \left(\frac{Z^*}{2Z^2} \right)^2 |\delta Z| \right] p(Z) dZ \quad (\text{A.22})$$

Solving both integrals, the final probability distribution $p(\delta Z)$ can be stated as follows:

$$p(\delta Z) = \begin{cases} \frac{Z^*}{2Z_{\max}Z_{\min}} - \frac{(Z^*)^2(Z_{\min}^2 + Z_{\min}Z_{\max} + Z_{\max}^2)}{12(Z_{\min}Z_{\max})^3} \cdot |\delta Z| & \text{if } 0 \leq |\delta Z| \leq \frac{2Z_{\min}^2}{Z^*} \\ \frac{1}{Z_{\max} - Z_{\min}} \left[\frac{2}{3} \sqrt{\frac{Z^*}{2|\delta Z|}} + \frac{(Z^*)^2}{12Z_{\max}^3} \cdot |\delta Z| - \frac{Z^*}{2Z_{\max}} \right] & \text{if } \frac{2Z_{\min}^2}{Z^*} \leq |\delta Z| \leq \frac{2Z_{\max}^2}{Z^*} \end{cases} \quad (\text{A.23})$$

A.4 Approximation of the Quantization Error for Equiangular Geometry

A distance Z for fish-eye stereo devices can be calculated with respect to the left camera as follows:

$$Z = b \cdot \frac{\sin \phi_r}{\sin(\phi_l - \phi_r)} = b \cdot \frac{\sin \phi_r}{\sin(\phi)} \quad (\text{A.24})$$

A first order error δZ for a certain given Z (and hence ϕ_l as well as ϕ) can be estimated by deviation:

$$\delta Z = \left[b \cdot \frac{\cos(\phi - \phi_l)}{\sin \phi} - \phi \cdot \frac{\sin(\phi - \phi_l) \cos \phi}{\sin \phi^2} \right] \cdot \delta \phi \quad (\text{A.25})$$

The disparity error $\delta \phi$ results for the statistical treatment from an error of the measured angles ϕ_l in the left image and ϕ_r in the right image respectively:

$$\delta Z = \left[b \cdot \frac{\cos(\phi - \phi_l)}{\sin \phi} - \phi \cdot \frac{\sin(\phi - \phi_l) \cos \phi}{\sin \phi^2} \right] \cdot \frac{1}{\alpha_x} (\delta x_{\text{img},l} - \delta x_{\text{img},r}) \quad (\text{A.26})$$

As stated in Section A.3 the variables $x_{\text{img},l}$ and $x_{\text{img},r}$ can be considered as mutual independent random variables of uniform distribution in the interval $[-\frac{a}{2}, +\frac{a}{2}]$ [17].

A probability density function $p(\delta Z|Z)$ for a given Z can hence be approximated by:

$$p(\delta Z|Z) = \underbrace{\left[b \cdot \frac{\cos(\phi - \phi_l)}{\sin \phi} - \phi \cdot \frac{\sin(\phi - \phi_l) \cos \phi}{\sin \phi^2} \right] \cdot \frac{1}{\alpha_x}}_{\left| \frac{d(\delta \phi)}{d(\delta Z)} \right|} \cdot p(\delta \phi) \quad (\text{A.27})$$

with

$$p(\delta \phi) = p(\delta x_{\text{img},l}) * p(\delta x_{\text{img},r}) \quad (\text{A.28})$$

Note that Equation A.27 represents a first order approximation of $p(\delta Z|Z)$. If required, higher order approximations can be generated corresponding to the *Taylor series approximation*. As an example the parametrization $b = 1.0\text{e} - 1 \text{ m}$, $f = 2.7\text{e} - 3 \text{ m}$, $p_x = 5.0\text{e} - 6 \text{ m}$, $Z = 2.5 \text{ m}$, $\phi_l = \pi/2$ shall be employed. Figure A.2 outlines approximated normalized probability density functions p_n of first and second order as well as a numerically evaluated density function for comparison. The latter has been determined using a MATLAB[®] kernel smoothing density function¹ estimation command following Bowman et al. [14].

Furthermore a first order error value $\pm \|\delta Z\|$ is outlined in Figure A.2 as black dashed lines based on the minimum measurable physical disparity angle $\delta \phi$ [17]:

$$\delta \phi = \pm \frac{1}{2 \cdot \alpha_x} \quad (\text{A.29})$$

It is recognizable that it coincides with a probability density value of exactly 0.5 for the first order approximation of the probability density function $p(\delta Z|Z)$.

¹ <http://de.mathworks.com/help/stats/ksdensity.html>

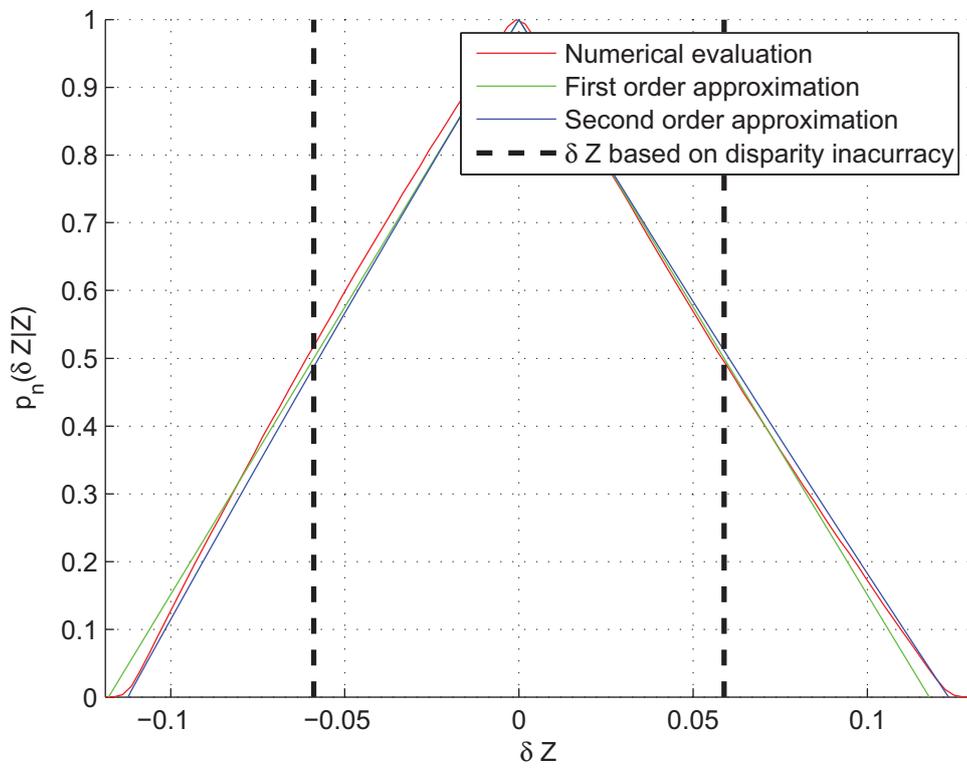


Figure A.2: Normalized probability density function

Appendix B

Further Relevant Publications

The following tables present further results of the search for literature related to omnistereo approaches.

B.1 H-Binocular Omnidirectional Stereo Vision with Panoramic Views

Table B.1: Related works in h-binocular omnidirectional stereo vision with panoramic views

Year	Author(s)	Remark	Publications
1990	Hiroshi Ishiguro et. al	Investigation of omnidirectional views to reconstruct a world model by an autonomous mobile robot.	[42–44]
1996	Sing Bing Kang and Richard Szeliski	Multiple rotating cameras create omnidirectional views and are applied for stereo processing.	[52]
2004	Yin Li et. al	A large collection of images taken by cameras that move in concentric circles is assembled to multiple panoramic images. Successively they compute multiperspective stereo vision.	[70]

Year	Author(s)	Remark	Publications
2004	Wei Jiang et. al	A large collection of images taken by stereo cameras that move in concentric circles is assembled to multiple panoramic images. The two collections of regular perspective images are resampled into four multi-perspective panoramas. Afterwards depth information is computed using three types of epipolar constraints: horizontal, vertical and combination of them.	[50]

B.2 V-Binocular Omnidirectional Stereo Vision with Panoramic Views

Table B.2: Related works in v-binocular omnidirectional stereo vision with panoramic views

Year	Author(s)	Remark	Publications
1995	Leonard McMillan and Gary Bishop	A plenoptic function is used for an image-based rendering system. This relies on sampling, reconstructing and re-sampling the plenoptic function.	[78]
1996 1998	R.Benosman et. al	A special sensor formed by a pair of cameras is rotating about the vertical axis and creates two panoramic images for 3D scene reconstruction.	[7, 8]
1998	Joshua Gluckman et. al	Real-time panoramic stereo is created by means of two vertically aligned orthographic projecting sensors that use parabolic reflecting surface.	[31]
2003	Shih-Schon Lin and Ruzena Bajcsy	A novel catadioptric sensor for omnidirectional stereo vision is presented. This sensor bases on two regular perspective cameras, a beam splitter and one reflective surface.	[71]

Year	Author(s)	Remark	Publications
2004	Wei Jiang et. al	A large collection of images taken by stereo cameras that move in concentric circles is assembled to multiple panoramic images. The two collections of regular perspective images are resampled into four multi-perspective panoramas. Afterwards depth information is computed using three types of epipolar constraints: horizontal, vertical and combination of them.	[50]
2005	Yasamin Mokri and Mansour Jamzad	Two vertically aligned catadioptrical sensors are applied on a mobile robot. Omnidirectional stereo is computed by means of neural networks.	[84]
2006	Sooyeong Yi and Narendra Ahuja	A new omnidirectional stereo imaging system is described that uses a concave lens and a convex mirror to produce a panoramic stereo pair of images on the sensor of a conventional camera.	[134]
2010	Tang Yi-ping et. al	Review of the existing panoramic imaging technologies. Obtains $360^\circ \times 360^\circ$ full sphere panoramic image by integrating to images that have been recorded by two symmetrical omnidirectional vision sensors.	[135]
2011	Zhang Bo et. al	Two vertically aligned catadioptric sensors are employed for computing panoramic stereo for surveying traffic scenes.	[10]

B.3 Binocular Omnidirectional Stereo Vision with Hemispherical Views

Table B.3: Related works in binocular omnidirectional stereo vision with hemispherical views

Year	Author(s)	Remark	Publications
2000	Toshiyuki Yamashita et. al	Two <i>Hyper-Omni-Vision-Sensors</i> are horizontally placed for reconstructing a 3-D model of a real environment by stereo vision. In order to make epipolar lines straight, an omnidirectional input image is projected onto a cylindrical surface of which axis is on the base line for stereo vision.	[133]
2005	Steffen Abraham and Wolfgang Forstner	The paper discusses binocular stereo approaches with fish-eye cameras. It surveys mathematical projection methods and discusses epipolar rectification as pre-requisite for stereo processing with fish-eyes.	[3]
2005	Shigang Li and Kiyotaka Fukumori	Binocular spherical stereo is applied using catadioptric devices. Instead of rectification, they consider the spherical image pair as sum of multiple image pairs and process them separately.	[69]
2006 2008	Shigang Li	A real-time binocular spherical stereo setup is presented using two fish-eye cameras. The rectification is carried out using <i>latitude-longitude representation</i> . See Section 5.2.	[66, 68]
2007	Takeshi Nishimoto and Junichi Yamaguchi	Binocular spherical stereo with fish-eye optics is employed for object detection. The epipolar rectification is carried using virtual perspective views.	[92]

Year	Author(s)	Remark	Publications
2009	Paul Bourke	Optical requirements for stereoscopic projection into hemispherical domes are discussed. The creation of optimal omni-directional stereoscopic fish-eye pairs for view independent projections is emphasised.	[13]
2010	Yun-fang Zhu	Binocular spherical stereo with fish-eye optics and intrinsic calibration based on the taylor model is investigated. The epipolar rectification is carried using virtual perspective views.	[145]
2011	Nobuyuki Kita	Dense 3D measurements are investigated by using parallel stereo camera system with fish-eye lenses. The measurement space is almost spherical at the distance of two or three times of baseline length from the stereo camera system.	[61]

B.4 Trinocular Omnidirectional Stereo Vision

Table B.4: Related works in trinocular omnidirectional stereo vision

Year	Author(s)	Remark	Publications
1996	Sing Bing Kang and Richard Szeliski	Multiple rotating cameras create omnidirectional views and are applied for stereo processing.	[52]
2001	Jane Mulligan et. al	To achieve high speed and accuracy in a stereo vision application, a trinocular stereo vision approach is applied. The employed optics base on perspective vision and do not target largeFOVsolutions.	[87]

Year	Author(s)	Remark	Publications
2006	Shigang Li	A trinocular spherical stereo vision approach based on fisheye optics and latitude-longitude sampling is presented. Theoretically, a hemispherical depth map can be generated.	[67]

B.5 Miscellaneous Configurations

Table B.5: N-ocular omnidirectional stereo vision

Year	Author(s)	Remark	Publications
1999 2001 2002	Kim C. Nga et. al	Multiple deployed omnidirectional vision sensors are employed for 3D ranging and generation of virtual scene views.	[89–91]
2000 2001 2004	Hideki Tanahashi et. al	A stereo omnidirectional imaging system named SOS is presented. It consists of multiple (e.g. 60) perspective cameras arranged as stereo units each on a face of a regular icosahedron. A full spherical depth map can be obtained.	[112, 122, 123]
2000 2001	Takushi Sogo et. al	Multiple deployed catadioptric vision sensors form a n-ocular stereo system for real-time human tracking.	[116, 117]
2009 2013	Hansung Kim and Adrian Hilton	For the purpose of 3D reconstruction multiple spherical images are generated by a bunch of rotating line cameras. A 3D mesh model for each pair of spherical images is reconstructed by stereo matching.	[59, 60]
2010	Tomokazu Sato and Naokazu Yokoya	A method is presented for estimating a depth map from a calibrated moving omnidirectional multi-camera system.	[101]

Table B.6: Omnivergent stereo vision

Year	Author(s)	Remark	Publications
1990 1992	Hiroshi Ishiguro et. al	Panoramic stereo is constructed by a swivelling camera that is mounted on a mobile robot. Three approaches for 3D computation are discussed: (1) Cylindrical projection, (2) Epipolar curves for stereo matching, (3) optical flow.	[42–44]
1996	Sing Bing Kang and Richard Szeliski	Multiple rotating cameras create omnidirectional views and are applied for stereo processing.	[52]
1996 1998	R.Benosman et. al	A special sensor formed by a pair of cameras is rotating about the vertical axis and creates two panoramic images for 3D scene reconstruction.	[7, 8]
1998	Ho-Chao Huang and Yi-Ping Hung	A panoramic stereo imaging (PSI) system for virtual reality is presented. A fixed baseline stereo rig is rotating around the center of the left camera and allows panoramic stereo recordings. Misalignments in rectification are corrected automatically.	[41]
1998 1999	Heung-Yeung Shum et. al	Multiple approaches for omnivergent stereo are investigated. Multiperspective panoramic, panoramic mosaics as well as spherical technics are presented.	[113–115]
1999 - 2001	Shmuel Peleg et. al	Panoramic stereo imaging approaches are presented by use of a single camera that rotates with an offset around a center point. Furthermore a special lens is presented that enables the easy generation of omni-stereo panoramas.	[94–96]

Year	Author(s)	Remark	Publications
2004	Yin Li et. al	A large collection of images taken by cameras that move in concentric circles is assembled to multiple panoramic images. Successively they compute multi-perspective stereo vision.	[70]
2004	Wei Jiang et. al	A large collection of images taken by stereo cameras that move in concentric circles is assembled to multiple panoramic images. The two collections of regular perspective images are resampled into four multi-perspective panoramas. Afterwards depth information is computed using three types of epipolar constraints: horizontal, vertical and combination of them.	[50]
2009 2013	Hansung Kim and Adrian Hilton	For the purpose of 3D reconstruction multiple spherical images are generated by a bunch of rotating line cameras. A 3D mesh model for each pair of spherical images is reconstructed by stereo matching.	[59, 60]
2010	Paul Bourke	By means of a single camera and a fish-eye lens, the capture of omni-directional stereoscopic spherical projections is presented. Furthermore the use of a single camera is compared to the employment of two rotating cameras.	[12]

Table B.7: Dynamic stereo vision

Year	Author(s)	Remark	Publications
1990 1992	Hiroshi Ishiguro et. al	Panoramic stereo is constructed by a swivelling camera that is mounted on a mobile robot. Three approaches for 3D computation are discussed: (1) Cylindrical projection, (2) Epipolar curves for stereo matching, (3) optical flow.	[42–44]
1998	Ho-Chao Huang and Yi-Ping Hung	A panoramic stereo imaging (PSI) system for virtual reality is presented. A fixed baseline stereo rig is rotating around the center of the left camera and allows panoramic stereo recordings. Misalignments in rectification are corrected automatically.	[41]
1999 - 2001	Zhigang Zhu et. al	A team of robots, each equipped with a panoramic imaging device, forms a multi-baseline cooperative panoramic stereo sensing approach. The application finds and protects humans by a robot team in such emergent circumstances, for example, during a fire in an office building.	[147–151]
2002	Tatsuya Yagishita et. al	A new technology related to a single dynamic 360-degree vision stereo camera by using one hyperboloid mirror has been developed. Specifically, the single camera is continuously moved in the vertical direction, the time (the number of frames) to travel a fixed distance on an image is measured, and the depth of the space is extracted.	[132]
2005	Yasamin Mokri and Mansour Jamzad	Two vertically aligned catadioptrical sensors are applied on a mobile robot. Omnidirectional stereo is computed by means of neural networks.	[84]

Year	Author(s)	Remark	Publications
2008	Ryosuke Kawanishi et. al	A mobile robot equipped with a single omnidirectional camera acquires sequences of images and computes the 3D surrounding with a structure from motion approach.	[58]

Table B.8: Omnidirectional stereo vision with special optical solutions

Year	Author(s)	Remark	Publications
1996 2000	David Southwell et. al	Catadioptric cameras are equipped with special double lobbed mirrors in order to perform stereoscopic measurements with one single camera.	[6, 118]
1999 - 2001	Shmuel Peleg et. al	Panoramic stereo imaging approaches are presented by use of a single camera that rotates with an offset around a center point. Furthermore a special lens is presented that enables the easy generation of omni-stereo panoramas.	[94-96]
2000 2001 2004	Hideki Tanahashi et. al	A stereo omnidirectional imaging system named SOS is presented. It consists of multiple (e.g. 60) perspective cameras arranged as stereo units each on a face of a regular icosahedron. A full spherical depth map can be obtained.	[112, 122, 123]
2003	Shih-Schon Lin and Ruzena Bajcsy	A novel catadioptric sensor for omnidirectional stereo vision is presented. This sensor bases on two regular perspective cameras, a beam splitter and one reflective surface.	[71]
2005	Yuichiro Kojima et. al	Catadioptric sensors in combination with special compound spherical mirrors are applied in order to generate an omnidirectional 3D measurement system. An application is seen for instance in wearable security systems.	[62]

Year	Author(s)	Remark	Publications
2006	Sooyeong Yi and Narendra Ahuja	A new omnidirectional stereo imaging system is described that uses a concave lens and a convex mirror to produce a panoramic stereo pair of images on the sensor of a conventional camera.	[134]
2006 2007	Liancheng Su et. al	A single catadioptric stereo vision system equipped with a pair of mirrors is used to establish a 3D obstacle information system for autonomous robot navigation purposes.	[74, 121, 144]
2010	Tang Yi-ping et. al	Review of the existing panoramic imaging technologies. Obtains $360^\circ \times 360^\circ$ full sphere panoramic image by integrating to images that have been recorded by two symmetrical omnidirectional vision sensors.	[135]
2012	Christian Weissig et. al	A new approach on an omni-directional omni-stereo multi-camera system is presented. It allows the recording of panoramic 3D video with high resolution and quality and display in stereo 3D on a cylindrical screen.	[126]
2012	Z.-H. Xiong et. al	A depth-space partitioning algorithm for performing object tracking using single-camera omni-stereo imaging system is presented. The proposed method uses a catadioptric omnidirectional stereo-imaging system to capture omni-stereo image pairs.	[130]
2012	Chi Zhang et. al	A novel structured light based omnidirectional 3D camera is presented, which consists of a projector, a camera, and two hyperbolic mirrors.	[137]

Bibliography

- [1] *Vision3D*. <http://www.vision3d.com/stereo.html>, . – Accessed: 2015-01-27
- [2] ABDEL-AZIZ, Y.I. ; KARARA, H.M.: Direct linear transformation into object space coordinates in closerange photogrametry. In: *Close-Range Photogrametry* (1971), S. 1–18
- [3] ABRAHAM, Steffen ; FOERSTNER, Wolfgang: Fish-eye-stereo calibration and epipolar rectification. In: *ISPRS Journal of Photogrametry and Remote Sensing* 59 (2005), S. 278–288
- [4] AMAT, J. ; FRIGOLA, M. ; CASALS, A.: Selection of the best stereo pair in a multi-camera configuration. In: *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on* Bd. 4, 2002, S. 3342–3346 vol.4
- [5] ARDOUIN, Jérôme ; LÉCUYER, Anatole ; MARCHAL, Maud ; RIAN, Clément ; MARCHAND, Eric: FlyVIZ: a novel display device to provide humans with 360 degree vision by coupling catadioptric camera with hmd. In: *Proceedings of the 18th ACM symposium on Virtual reality software and technology*. New York, NY, USA : ACM, 2012 (VRST '12). – ISBN 978-1-4503-1469-5, 41–44
- [6] BALDWIN, Jonathan ; BASU, Anup: 3D estimation using panoramic stereo. In: *Pattern Recognition, 2000. Proceedings. 15th International Conference on* Bd. 1, 2000. – ISSN 1051-4651, S. 97–100 vol.1
- [7] BENOSMAN, R. ; DEVARS, J.: Panoramic stereo vision sensor. In: *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on* Bd. 1, 1998. – ISSN 1051-4651, S. 767–769 vol.1
- [8] BENOSMAN, R. ; MANIERE, T. ; DEVARS, J.: Real time omni-directional stereovision and planes detection. In: *Electrotechnical Conference, 1996. MELECON '96., 8th Mediterranean* Bd. 2, 1996, S. 1151–1155 vol.2
- [9] BLOSTEIN, S.D. ; HUANG, T.S.: Error Analysis in Stereo Determination of 3-D Point Positions. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-9* (1987), Nr. 6, S. 752–765. – ISSN 0162-8828

- [10] BO, Zhang ; HAITAO, Xu ; LULU, Lin ; JUN, Jiang ; JIANYUE, Zhang ; YIPING, Tang: Development of Coordinate-machine of the Scene of Road Traffic Accident based on Binocular Stereo Omni-Directional Vision Sensors. In: *Procedia Engineering* 24 (2011), Nr. 0, S. 262 – 266. – ISSN 1877–7058. – International Conference on Advances in Engineering 2011
- [11] BOUGUET, Jean-Yves: Camera calibration toolbox for matlab. (2004). http://www.vision.caltech.edu/bouguetj/calib_doc/index.html
- [12] BOURKE, P.: Capturing omni-directional stereoscopic spherical projections with a single camera. In: *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on*, 2010, S. 179–183
- [13] BOURKE, Paul: Omni-Directional stereoscopic fisheye images for immersive hemispherical dome environments. In: *Comput. Games Allied Technol* (2009), S. 136–143
- [14] BOWMAN, Adrian W. ; AZZALINI, Adelchi: *Applied smoothing techniques for data analysis*. Clarendon Press, 2004
- [15] BRAEUER-BURCHARDT, Christian ; VOSS, Klaus: A new algorithm to correct fish-eye- and strong wide-angle-lens-distortion from single images. In: *Proceedings of the 2001 International Conference on Image Processing 1* (2001), S. 225–228
- [16] BROWN, D.C.: Decentering distortion of lenses. In: *Photogrammetric Eng. Remote Sensing* (1966), S. 444–462
- [17] CHANG, C. ; CHATTERJEE, S.: Quantization error analysis in stereo vision. In: *Signals, Systems and Computers, 1992. 1992 Conference Record of The Twenty-Sixth Asilomar Conference on*, 1992. – ISSN 1058–6393, S. 1037–1041 vol.2
- [18] CSURKA, Gabriella ; ZELLER, Cyril ; ZHANG, Zhengyou ; FAUGERAS, Olivier D.: Characterizing the Uncertainty of the Fundamental Matrix. In: *Computer Vision and Image Understanding* 68 (1997), Nr. 1, S. 18 – 36. – ISSN 1077–3142
- [19] DEVERNAY ; FAUGERAS: Straight Lines Have to Be Straight: Automatic Calibration and Removal of Distortion from Scenes of Structured Environments. In: *Machine Vision and Applications* 1 (2001), S. 14–24
- [20] DOLLÁR, P. ; TU, Z. ; PERONA, P. ; BELONGIE, S.: Integral Channel Features. In: *BMVC*, 2009
- [21] FAUGERAS, Oliver: *Three dimensional computer vision: A geometric viewpoint*. the MIT Press, 1993
- [22] FAUGERAS, Olivier ; TOSCANI, Giorgio: Camera calibration for 3D computer vision. In: *International Workshop on Machine Vision and Machine Intelligence* (1987), S. 240–247

- [23] FINDEISEN, M. ; MEINEL, L. ; HIRTZ, G.: A trinocular omnidirectional stereo vision system for high-precision RGB-D acquisition. In: *ELMAR (ELMAR), 2014 56th International Symposium*, 2014, S. 1–4
- [24] FINDEISEN, Michel ; HIRTZ, Gangolf: Trinocular Spherical Stereo Vision for Indoor Surveillance. In: *11th Conference on Computer and Robot Vision (CRV)*. Montreal, Quebec, Canada : IEEE Xplore Digital Library, May 2014
- [25] FINDEISEN, Michel ; MEINEL, Lars ; HESS, Markus ; APITZSCH, André ; HIRTZ, Gangolf: A Fast Approach For Omnidirectional Surveillance With Multiple Virtual Perspective Views. In: *2013 IEEE Eurocon International conference on computer as a tool*. Zagreb, Croatia, 2013, S. 1578–1585
- [26] FINDEISEN, Michel ; MEINEL, Lars ; RICHTER, Julia ; HIRTZ, Gangolf: An Omnidirectional Stereo Sensor for Human Behavior Analysis in Complex Indoor Environments IEEE International Conference on Consumer Electronics, Las Vegas, 2015
- [27] FOOLADGAR, F. ; SAMAVI, S. ; SOROUSHMEHR, S.M.R. ; SHIRANI, S.: Geometrical Analysis of Localization Error in Stereo Vision Systems. In: *Sensors Journal, IEEE* 13 (2013), Nr. 11, S. 4236–4246. – ISSN 1530–437X
- [28] FUSIELLO, Andrea ; TRUCCO, Emanuele ; VERRI, Alessandro: A compact algorithm for rectification of stereo pairs. In: *Machine Vision and Applications* 12 (2000), Nr. 1, S. 16–22. <http://dx.doi.org/10.1007/s001380050120>. – DOI 10.1007/s001380050120. – ISSN 0932–8092
- [29] GANAPATHY, S.: Decomposition of transformation matrices for robot vision. In: *Proceedings of the 1984 IEEE International Conferece on Robotics and Automation* (1984), S. 130–139
- [30] GANDHI, Tarak: Motion analysis of omni-directional video streams for a mobile sentry. In: *ACM International Workshop on Video Surveillance*, ACM Press, 2003, S. 49–58
- [31] GLUCKMAN, Joshua ; NAYAR, Shree K. ; THORESZ, Keith J.: Real-time omnidirectional and panoramic stereo. In: *Proc. of DARPA Image Understanding Workshop* Bd. 1, 1998, S. 299–303
- [32] HANNING, Tobias: Re-projective calibration of a stereo camera system. In: *International Conference on Visualization, Imaging and Image Processing (VIIP)*, 2006, S. 7–12
- [33] HANNING, Tobias: *High precision camera calibration*. Springer, 2011

- [34] HANNING, Tobias ; GRAF, Simone ; PISINGER, Georg: Extrinsic calibration of a stereo camera system fulfilling generalized epipolar constraints. In: *International Conference on Visualization, Imaging and Image Processing Conference (VIIP)*, 2004, S. 1–5
- [35] HARTLEY, Richard: Theory and practice of projective rectification. In: *International Journal of Computer Vision* 35 (1999), Nr. 2, S. 115–127
- [36] HARTLEY, Richard ; ZISSERMAN, Andrew: *Multiple view geometry in computer vision*. Bd. 2. Cambridge Univ Press, 2000
- [37] HARTLEY, Richard ; ZISSERMAN, Andrew: *Multiple view geometry in computer vision*. Cambridge university press, 2003
- [38] HEIKKILÄ, Janne: Geometric Camera Calibration Using Circular Control Points. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), S. 1066–1077
- [39] HEIKKILÄ, Janne ; SILVÉN, Olli: A Four-step Camera Calibration Procedure with Implicit Image Correction. In: *CVPR '97 Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition* (1997)
- [40] HIRSCHMULLER, H. ; SCHARSTEIN, D.: Evaluation of Cost Functions for Stereo Matching. In: *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007. – ISSN 1063–6919, S. 1–8
- [41] HUANG, Ho-Chao ; HUNG, Yi-Ping: Panoramic stereo imaging system with automatic disparity warping and seaming. In: *Graphical Models and Image Processing* 60 (1998), Nr. 3, S. 196–208
- [42] ISHIGURO, Hiroshi ; YAMAMOTO, Masashi ; TSUJI, Saburo: Analysis of omnidirectional views at different location. In: *Intelligent Robots and Systems '90. 'Towards a New Frontier of Applications', Proceedings. IROS '90. IEEE International Workshop on*, 1990, S. 659–664 vol.2
- [43] ISHIGURO, Hiroshi ; YAMAMOTO, Masashi ; TSUJI, Saburo: Omnidirectional stereo for making global map. In: *Computer Vision, 1990. Proceedings, Third International Conference on*, 1990, S. 540–547
- [44] ISHIGURO, Hiroshi ; YAMAMOTO, Masashi ; TSUJI, Saburo: Omnidirectional stereo. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 14 (1992), Nr. 2, S. 257–262. <http://dx.doi.org/10.1109/34.121792>. – DOI 10.1109/34.121792. – ISSN 0162–8828

- [45] ISHII, C. ; SUDO, Y. ; HASHIMOTO, H.: An image conversion algorithm from fish eye image to perspective image for human eyes. In: *Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings. 2003 IEEE/ASME International Conference on* Bd. 2, 2003, S. 1009–1014 vol.2
- [46] JAEHNE, Bernd: *Digitale Bildverarbeitung*. Springer DE, 2012
- [47] JENG, S. W. ; TSAI, W.H.: Using pano-mapping tables for unwarping of omni-images into panoramic and perspective-view images. In: *Image Processing, IET 1* (2007), June, Nr. 2, S. 149–155. <http://dx.doi.org/10.1049/iet-ipr:20060201>. – DOI 10.1049/iet-ipr:20060201. – ISSN 1751–9659
- [48] JENG, Sheng-Wen ; TSAI, Wen-Hsiang: Construction of perspective and panoramic images from omni-images taken from hypercatadioptric cameras for visual surveillance. In: *Networking, Sensing and Control, 2004 IEEE International Conference on* Bd. 1, 2004. – ISSN 1810–7869, S. 204–209 Vol.1
- [49] JENG, Sheng-Wen ; TSAI, Wen-Hsiang: Analytic image unwarping by a systematic calibration method for omni-directional cameras with hyperbolic-shaped mirrors. In: *Image and Vision Computing 26* (2008), Nr. 5, S. 690 – 701. – ISSN 0262–8856
- [50] JIANG, Wei ; SUGIMOTO, S. ; OKUTOMI, M.: Omnidirectional 3D reconstruction using stereo multi-perspective panoramas. In: *SICE 2004 Annual Conference* Bd. 2, 2004, S. 1128–1133 vol. 2
- [51] KADEN, Sascha: Merging Depth Maps from the Omnidirectional Stereo Vision / Chemnitz University of Technology, Professorship Digital Signal Processing and Circuit Technology. 2014. – Forschungsbericht
- [52] KANG, Sing B. ; SZELISKI, R.: 3-D scene data recovery using omnidirectional multi-baseline stereo. In: *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, 1996. – ISSN 1063–6919, S. 364–370
- [53] KANNALA, Juho: *Camera Calibration Toolbox for Generic Lenses*. <http://www.ee.oulu.fi/~jkannala/calibration/calibration.html>. Version:2008
- [54] KANNALA, Juho: Models and methods for geometric computer vision. In: *Academic dissertation* (2010)
- [55] KANNALA, Juho ; BRANDT, Sami S.: A Generic Camera Calibration Method for Fish-Eye Lenses. In: *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)* (2004)

- [56] KANNALA, Juho ; BRANDT, Sami S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28 (2006), Aug, Nr. 8, S. 1335–1340. <http://dx.doi.org/10.1109/TPAMI.2006.153>. – DOI 10.1109/TPAMI.2006.153. – ISSN 0162–8828
- [57] KANNALA, Juho ; HEIKKILA, Janne ; BRANDT, Sami S.: Geometric Camera Calibration. In: *Wiley Encyclopedia of Computer Science and Engineering* (2008)
- [58] KAWANISHI, Ryosuke ; YAMASHITA, Atsushi ; KANEKO, Toru: Construction of 3D Environment Model from an Omni-Directional Image Sequence. In: *Asia International Symposium on Mechatronics, 2008*
- [59] KIM, Hansung ; HILTON, A.: Environment modelling using spherical stereo imaging. In: *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009*, S. 1534–1541
- [60] KIM, Hansung ; HILTON, Adrian: 3D Scene Reconstruction from Multiple Spherical Stereo Pairs. In: *International Journal of Computer Vision* 104 (2013), Nr. 1, S. 94–116. – ISSN 0920–5691
- [61] KITA, Nobuyuki: Dense 3D Measurement of the Near Surroundings by Fisheye Stereo. In: *Proc. The Twelfth IAPR Conference on Machine Vision Applications, 2011*
- [62] KOJIMA, Yuichiro ; SAGAWA, Ryusuke ; ECHIGO, Tomio ; YAGI, Yasushi: Calibration and Performance Evaluation of Omnidirectional Sensor with Compound Spherical Mirrors. In: *IPSJ SIG Technical Reports* Bd. 38, 2005, S. 31–38
- [63] KWON, Hyukseong ; PARK, Johnny ; KAK, Avinash C.: A New Approach for Active Stereo Camera Calibration. In: *Robotics and Automation, 2007 IEEE International Conference on, 2007*. – ISSN 1050–4729, S. 3180–3185
- [64] LENZ, R.K. ; TSAI, R.Y.: Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. In: *Proceedings of the IEEE International Conference on Robotics Automation* (1987), S. 68–75
- [65] LI, Fuxing ; BRADY, Michael ; WILES, Charles: Fast computation of the fundamental matrix for an active stereo vision system. Version: 1996. <http://dx.doi.org/10.1007/BFb0015532>. In: BUXTON, Bernard (Hrsg.) ; CIPOLLA, Roberto (Hrsg.): *Computer Vision - ECCV '96* Bd. 1064. Springer Berlin Heidelberg, 1996. – DOI 10.1007/BFb0015532. – ISBN 978–3–540–61122–6, 157–166
- [66] LI, Shigang: Real-Time Spherical Stereo. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* Bd. 3, 2006. – ISSN 1051–4651, S. 1046–1049

-
- [67] LI, Shigang: Trinocular Spherical Stereo. In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, S. 4786–4791
- [68] LI, Shigang: Binocular Spherical Stereo. In: *Intelligent Transportation Systems, IEEE Transactions on* 9 (2008), Nr. 4, S. 589–600. <http://dx.doi.org/10.1109/TITS.2008.2006736>. – DOI 10.1109/TITS.2008.2006736. – ISSN 1524–9050
- [69] LI, Shigang ; FUKUMORI, Kiyotaka: Spherical stereo for the construction of immersive VR environment. In: *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, 2005, S. 217–222
- [70] LI, Yin ; SHUM, Heung-Yeung ; TANG, Chi-Keung ; SZELISKI, R.: Stereo reconstruction from multiperspective panoramas. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26 (2004), Nr. 1, S. 45–62. <http://dx.doi.org/10.1109/TPAMI.2004.1261078>. – DOI 10.1109/TPAMI.2004.1261078. – ISSN 0162–8828
- [71] LIN, Shih-Schon ; BAJCSY, Ruzena: High resolution catadioptric omni-directional stereo sensor for robot vision. In: *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on* Bd. 2, 2003. – ISSN 1050–4729, S. 1694–1699 vol.2
- [72] LONGUET-HIGGINS, H. C.: A computer algorithm for reconstructing a scene from two projections. In: *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, MA Fischler and O. Firschein, eds (1987), S. 61–62
- [73] LOOP, Charles ; ZHANG, Zhengyou: Computing rectifying homographies for stereo vision. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. Bd. 1, 1999. – ISSN 1063–6919, S. –131 Vol. 1
- [74] LUO, Chuanjiang ; SU, Liancheng ; ZHU, Feng: A novel omnidirectional stereo vision system via a single camera. In: *Scene Reconstruction Pose Estimation and Tracking, InTech* (2007), S. 19–38
- [75] LUONG, Quan-Tuan ; FAUGERAS, Olivier D.: The fundamental matrix: Theory, algorithms, and stability analysis. In: *International Journal of Computer Vision* 17 (1996), Nr. 1, S. 43–75
- [76] MATOUSEK, Martin ; SARA, Radim ; HLAVAC, Vaclav: Data-optimal rectification for fast and accurate stereovision. In: *Multi-Agent Security and Survivability, 2004 IEEE First Symposium on*, 2004, S. 212–215
- [77] MATTOCCIA, Stefano: *Stereo Vision: Algorithms and Applications*. October 2013

- [78] MCMILLAN, Leonard ; BISHOP, Gary: Plenoptic modeling: an image-based rendering system. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. New York, NY, USA : ACM, 1995 (SIGGRAPH '95). – ISBN 0-89791-701-4, S. 39–46
- [79] MEINEL, Lars: *Verfahren zur Kalibrierung von perspektivischen und omnidirektionalen Kamerasystemen*, Chemnitz University of Technology, Diplomarbeit, 2012
- [80] MEINEL, Lars: *Multi-Objektverfolgung zur Innenraumüberwachung unter Nutzung eines omnidirektionalen Kamerasystems*, Chemnitz University of Technology, Diplomarbeit, 2013
- [81] MEINEL, Lars ; FINDEISEN, Michel ; HESS, Markus ; APITZSCH, André ; HIRTZ, Gangolf: Automated Real-Time Surveillance for Ambient Assisted Living Using an Omnidirectional Camera. In: *2014 IEEE International Conference on Consumer Electronics (ICCE) (2014 IEEE ICCE)*. Las Vegas, USA, Januar 2014, S. 402–405
- [82] MEINEL, Lars ; WIEDE, Christian ; FINDEISEN, Michel ; APITZSCH, Andre ; HIRTZ, Gangolf: Virtual perspective views for real-time people detection using an omnidirectional camera. In: *Imaging Systems and Techniques (IST), 2014 IEEE International Conference on*, 2014, S. 312–315
- [83] MELEN, Trond: Geometrical Modelling and Calibration of Video Cameras for Underwater Navigation. In: *Dissertation zum Dr.-Ing., Univ. of Science and Technology, Trondheim, Norwegen* (1994)
- [84] MOKRI, Yasamin ; JAMZAD, Mansour: Omni-stereo vision system for an autonomous robot using neural networks. In: *Electrical and Computer Engineering, 2005. Canadian Conference on*, 2005. – ISSN 0840-7789, S. 1590–1593
- [85] MONASSE, Pascal ; MOREL, Jean-Michel ; TANG, Zhongwei: Three-step image rectification. In: *Proceedings of the British Machine Vision Conference*, 2010
- [86] MOUADDIB, El M. ; SAGAWA, Ryusuke ; ECHIGO, Tomio ; YAGI, Yasushi: Two or more mirrors for the omnidirectional stereovision? In: *eurasip* 52 (2006)
- [87] MULLIGAN, Jane ; ISLER, Volkan ; DANILIDIS, Kostas: Trinocular stereo: a real-time algorithm and its evaluation. In: *Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on*, 2001, S. 10–17
- [88] NENE, S.A. ; NAYAR, S.K.: Stereo with mirrors. In: *Computer Vision, 1998. Sixth International Conference on*, 1998, S. 1087–1094
- [89] NG, Kim C. ; TRIVEDI, Mohan M. ; ISHIGUROB, Hiroshi: 3 D ranging and virtual view generation using omni-view cameras. In: *Proceedings of SPIE- The International Society for Optical Engineering* Bd. 3528, 1999, S. 11–14

-
- [90] NG, Kim C. ; TRIVEDI, Mohan M. ; ISHIGUROB, Hiroshi: Range-space approach for generalized multiple baseline stereo and direct virtual view synthesis. In: *Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on, 2001*, S. 62–71
- [91] NG, Kim C. ; TRIVEDI, Mohan M. ; ISHIGUROB, Hiroshi: Generalized multiple baseline stereo and direct virtual view synthesis using range-space search, match and render. In: *International Journal of Computer Vision* 47 (2002), S. 131–147
- [92] NISHIMOTO, Takeshi ; YAMAGUCHI, Junichi: Three Dimensional Measurement Using Fisheye Stereo Vision. In: *SICE, 2007 Annual Conference, 2007*, S. 2008–2012
- [93] ONOE, Y. ; YOKOYA, N. ; YAMAZAWA, K. ; TAKEMURA, H.: Visual surveillance and monitoring system using an omnidirectional video camera. In: *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on* Bd. 1, 1998, S. 588 –592 vol.1
- [94] PELEG, Shmuel ; BEN-EZRA, Moshe: Stereo panorama with a single camera. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. Bd. 1, 1999. – ISSN 1063–6919, S. –401 Vol. 1
- [95] PELEG, Shmuel ; BEN-EZRA, Moshe ; PRITCH, Yael: Cameras for stereo panoramic imaging. In: *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on* Bd. 1, 2000. – ISSN 1063–6919, S. 208–214 vol.1
- [96] PELEG, Shmuel ; BEN-EZRA, Moshe ; PRITCH, Yael: Omnistere: panoramic stereo imaging. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23 (2001), Nr. 3, S. 279–290. <http://dx.doi.org/10.1109/34.910880>. – DOI 10.1109/34.910880. – ISSN 0162–8828
- [97] PERI, Venkata N. ; NAYAR, Shree K.: Generation of Perspective and Panoramic Video from Omnidirectional Video. In: *Proc. of DARPA Image Understanding Workshop* (1997)
- [98] PETERSSON, Niklas ; PETERSSON, Lars: Online stereo calibration using FPGAs. In: *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE, 2005*, S. 55–60
- [99] RICHTER, Julia ; FINDEISEN, Michel ; HIRTZ, Gangolf: Assessment and Care System Based on People Detection for Elderly Suffering From Dementia International Conference on Consumer Electronics (ICCE-Berlin), 2014, S. 59 – 63
- [100] RODRIGUEZ, J.J. ; AGGARWAL, J.K.: Stochastic analysis of stereo quantization error. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12 (1990), May, Nr. 5, S. 467–470. – ISSN 0162–8828

- [101] SATO, Tomokazu ; YOKOYA, Naokazu: Efficient hundreds-baseline stereo by counting interest points for moving omni-directional multi-camera system. In: *Journal of Visual Communication and Image Representation* 21 (2010), Nr. 56, S. 416 – 426. – ISSN 1047–3203. – Special issue on Multi-camera Imaging, Coding and Innovative Display
- [102] SCARAMUZZA, Davide: *Omnidirectional Vision: from Calibration to Robot Motion Estimation*, ETH Zuerich, Diss., February 22 2008
- [103] SCARAMUZZA, Davide ; MARTINELLI, Agostino ; SIEGWART, Roland: A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. In: *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems* (2006)
- [104] SCARAMUZZA, Davide ; SIEGWART, Roland: A New Method and Toolbox for Easily Calibrating Omnidirectional Cameras. In: *Proceedings of the 5th International Conference on Computer Vision Systems* (2007)
- [105] SCHARSTEIN, D. ; PAL, Chris: Learning Conditional Random Fields for Stereo. In: *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007.* – ISSN 1063–6919, S. 1–8
- [106] SCHARSTEIN, D. ; SZELISKI, R.: High-accuracy stereo depth maps using structured light. In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* Bd. 1, 2003. – ISSN 1063–6919, S. I–195–I–202 vol.1
- [107] SCHARSTEIN, D. ; SZELISKI, R. ; ZABIH, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: *Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on, 2001*, S. 131–140
- [108] SCHARSTEIN, Daniel ; SZELISKI, Richard: Middlebury stereo vision page. In: *Online at [http://www. middlebury. edu/stereo](http://www.middlebury.edu/stereo)* (2002)
- [109] SCHREER, O. ; KAUFF, P. ; EISERT, P. ; WEISSIG, C. ; ROSENTHAL, J.: Geometrical design concept for panoramic 3D video acquisition. In: *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, 2012.* – ISSN 2219–5491, S. 2757–2761
- [110] SCHWALBE, Ellen: Geometric Modelling and Calibration of Fisheye Lens Camera Systems. In: *International archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVI (2005)
- [111] SCHWALBE, Ellen ; MAAS, Hans-Gerd: Ein Ansatz zur Elimination der chromatischen Aberration bei der Modellierung und Kalibrierung von Fisheye-Aufnahmesystemen. In: *Beitraege Oldenburger 3D-Tage* (2006)

-
- [112] SHIMIZU, Sanae ; YAMAMOTO, Kazuhiko ; WANG, Caihua ; SATOH, Yutaka ; TANAHASHI, Hideki ; NIWA, Yoshinori: Moving object detection by mobile Stereo Omni-directional System (SOS) using spherical depth image. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* Bd. 3, 2004. – ISSN 1051–4651, S. 248–251 Vol.3
- [113] SHUM, Heung-Yeung ; HAN, Mei ; SZELISKI, Richard: Interactive construction of 3D models from panoramic mosaics. In: *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, 1998. – ISSN 1063–6919, S. 427–433
- [114] SHUM, Heung-Yeung ; KALAI, Adam ; SEITZ, Steven M.: Omnivergent stereo. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* Bd. 1, 1999, S. 22–29 vol.1
- [115] SHUM, Heung-Yeung ; SZELISKI, Richard: Stereo reconstruction from multiperspective panoramas. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* Bd. 1, 1999, S. 14–21 vol.1
- [116] SOGO, T. ; ISHIGURO, H. ; TRIVEDI, M.M.: N-Ocular Stereo for Real-Time Human Tracking. In: BENOSMAN, Ryad (Hrsg.) ; KANG, SingBing (Hrsg.): *Panoramic Vision*. Springer New York, 2001 (Monographs in Computer Science). – ISBN 978–1–4419–2880–1, S. 359–375
- [117] SOGO, Takushi ; ISHIGURO, Hiroshi ; TRIVEDI, Mohan M.: Real-Time Target Localization and Tracking by N-Ocular Stereo. In: *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, 2000, S. 153–160
- [118] SOUTHWELL, David ; BASU, Anup ; FIALA, Mark ; REYDA, Jerome: Panoramic stereo. In: *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* Bd. 1, 1996. – ISSN 1051–4651, S. 378–382 vol.1
- [119] STURM, Peter: Mixing catadioptric and perspective cameras. In: *Omnidirectional Vision, 2002. Proceedings. Third Workshop on IEEE*, 2002, S. 37–44
- [120] SU, Huihuang ; HE, Bingwei: Stereo rectification of calibrated image pairs based on geometric transformation. In: *International Journal of Modern Education and Computer Science (IJMECS)* 3 (2011), Nr. 4, S. 17
- [121] SU, Liancheng ; LUO, Chuanjiang ; ZHU, Feng: Obtaining Obstacle Information by an Omnidirectional Stereo Vision System. In: *Information Acquisition, 2006 IEEE International Conference on*, 2006, S. 48–52

- [122] TANAHASHI, Hideki ; NIWA, Yoshinori ; WANG, Caihua ; NIWA, Yoshinori: Development of a stereo omnidirectional imaging system (SOS). In: *Industrial Electronics Society, 2000. IECON 2000. 26th Annual Conference of the IEEE* Bd. 1, 2000, S. 289–294 vol.1
- [123] TANAHASHI, Hideki ; NIWA, Yoshinori ; WANG, Caihua ; NIWA, Yoshinori: Acquisition of three-dimensional information in a real environment by using the Stereo Omni-directional System (SOS). In: *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, 2001, S. 365–371
- [124] THIBAUT, Simon: Enhanced surveillance system based on panomorph panoramic lenses. In: *Defense and Security Symposium International Society for Optics and Photonics*, 2007, S. 65400E–65400E
- [125] TSAI, Roger Y.: A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses. In: *IEEE JOURNAL OF ROBOTICS AND AUTOMATION* RS-3 (1987), S. 323–344
- [126] WEISSIG, Christian ; SCHREER, Oliver ; EISERT, Peter ; KAUFF, Peter: The Ultimate Immersive Experience: Panoramic 3D Video Acquisition. In: SCHOEFFMANN, Klaus (Hrsg.) ; Merialdo, Bernard (Hrsg.) ; HAUPTMANN, AlexanderG. (Hrsg.) ; NGO, Chong-Wah (Hrsg.) ; ANDREOPOULOS, Yiannis (Hrsg.) ; BREITENEDER, Christian (Hrsg.): *Advances in Multimedia Modeling* Bd. 7131. Springer Berlin Heidelberg, 2012. – ISBN 978–3–642–27354–4, S. 671–681
- [127] WENG, Juyang ; COHEN, Paul ; HERNIOU, Marc: Calibration of stereo cameras using a non-linear distortion model. In: *Pattern Recognition, 1990. Proceedings., 10th International Conference on* Bd. i, 1990, S. 246–253 vol.1
- [128] WENG, Juyang ; COHEN, Paul ; HERNIOU, Marc: Camera Calibration with Distortion Models and Accuracy Evaluation. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1992), S. 965–980
- [129] WONG, K.M.: Mathematical formulation and digital analysis in closerange photogrammetry. In: *Photogrammetric Eng. Remote Sensing* 41 (1975), S. 1355–1373
- [130] XIONG, Z.-H. ; CHENG, I. ; CHEN, W. ; BASU, A. ; ZHANG, M.-J.: Depth space partitioning for omni-stereo object tracking. In: *Computer Vision, IET* 6 (2012), Nr. 2, S. 153–163. <http://dx.doi.org/10.1049/iet-cvi.2010.0115>. – DOI 10.1049/iet-cvi.2010.0115. – ISSN 1751–9632
- [131] XU, Gang ; ZHANG, Zhengyou: Epipolar Geometry in Stereo, Motion and Object Recognition. In: *Journal of Electronic Imaging* 8 (1999), Nr. 3, 322–322. <http://dx.doi.org/10.1117/1.482719>. – DOI 10.1117/1.482719. ISBN 1017–9909

-
- [132] YAGISHITA, Tatsuya ; SHIMADA, Ryuya ; SHIMAMURA, Yohei ; MIZUTANI, Kota ; AMANO, Hirotaka ; SHIRASU, Hironori ; YAMADA, Hiromitsu ; MORI, Terunori: Development of 360-degree Vision Stereo Camera System. In: *IEIC Technical Report (Institute of Electronics, Information and Communication Engineers)* 102 (2002), Nr. 532, S. 31–36
- [133] YAMASHITA, Toshiyuki ; YAMAZAWA, Kazumasa ; TAKEMURA, Haruo ; YOKOYA, Naokazu: A 3D-model reconstruction of a real environment by using a stereo vision with omnidirectional image sensors. In: *IEIC Technical Report (Institute of Electronics, Information and Communication Engineers)* 99 (2000), Nr. 710, S. 43–48
- [134] YI, Sooyeong ; AHUJA, Narendra: An Omnidirectional Stereo Vision System Using a Single Camera. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* Bd. 4, 2006. – ISSN 1051–4651, S. 861–865
- [135] YI-PING, Tang ; BEI, Lin ; MIN-ZHI, Chen ; JUN, Sun: Design of Stereo Omnidirectional Vision Sensors with Full Sphere View and without Dead Angle. (2010)
- [136] ZHANG, Baofeng ; QI, Zhiqiang ; ZHU, Junchao ; CAO, Zuoliang: Omnidirection image restoration based on spherical perspective projection. In: *Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on*, 2008, S. 922–925
- [137] ZHANG, Chi ; XU, Jing ; XI, Ning ; JIA, Yunyi ; LI, Weixian: Development of an omni-directional 3D camera for robot navigation. In: *Advanced Intelligent Mechatronics (AIM), 2012 IEEE/ASME International Conference on*, 2012. – ISSN 2159–6247, S. 262–267
- [138] ZHANG, Zhengyou: Motion of a stereo rig: Strong, weak and self calibration. In: LI, StanZ. (Hrsg.) ; MITAL, DineshP. (Hrsg.) ; TEOH, EamKhwang (Hrsg.) ; WANG, Han (Hrsg.): *Recent Developments in Computer Vision* Bd. 1035. Springer Berlin Heidelberg, 1996. – ISBN 978–3–540–60793–9, S. 241–253
- [139] ZHANG, Zhengyou: On the epipolar geometry between two images with lens distortion. In: *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* Bd. 1, 1996. – ISSN 1051–4651, S. 407–411 vol.1
- [140] ZHANG, Zhengyou: *A Stereovision System for a Planetary Rover: Calibration, Correlation, Registration, and Fusion*. 1996
- [141] ZHANG, Zhengyou: Determining the epipolar geometry and its uncertainty: A review. In: *International Journal of Computer Vision* 27 (1998), Nr. 2, S. 161–195
- [142] ZHANG, Zhengyou: A Flexible New Technique for Camera Calibration. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1998), S. 1330–1334

- [143] ZHOU, Chuan ; TAN, Dalong ; GAO, Hongwei: A High-Precision Calibration and Optimization Method for Stereo Vision System. In: *Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on*, 2006, S. 1–5
- [144] ZHU, Feng ; SU, Liancheng: Omnidirectional Depth Estimation by a Single Perspective Camera. In: *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on Bd. 2*, 2006, S. 9854–9857
- [145] ZHU, Yun-fang: Fish-eye Stereo Based on the Taylor Model. In: *Information Science and Engineering (ICISE), 2010 2nd International Conference on*, 2010, S. 3709–3713
- [146] ZHU, Zhigang: Omnidirectional Stereo Vision. In: *Workshop on Omnidirectional Vision, in the 10th IEEE ICAR* (2001)
- [147] ZHU, Zhigang ; RAJASEKAR, K D. ; RISEMAN, Edward M. ; HANSON, Allen R.: 3D localization of multiple moving people by a omnidirectional stereo system of cooperative mobile robots. In: *U Mass CS TR 14* (2000), S. 2000
- [148] ZHU, Zhigang ; RAJASEKAR, K.D. ; RISEMAN, E.M. ; HANSON, A.R.: Panoramic virtual stereo vision of cooperative mobile robots for localizing 3D moving objects. In: *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, 2000, S. 29–36
- [149] ZHU, Zhigang ; RISEMAN, Edward M. ; HANSON, Allen R.: Geometrical modeling and real-time vision applications of a panoramic annular lens (PAL) camera system. In: *University of Massachusetts, Amherst, MA* (1999)
- [150] ZHU, Zhigang ; RISEMAN, E.M. ; HANSON, A.R.: Parallel-perspective stereo mosaics. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on Bd. 1*, 2001, S. 345–352 vol.1
- [151] ZHU, Zhigang ; RISEMAN, EM ; HANSON, AR: Theory and practice in making seamless stereo mosaics from airborne video / Citeseer. 2001. – Forschungsbericht

List of Figures

1.1	Optical 3D measurement techniques	36
1.2	Optical smart sensor based on passive stereo vision (source: Intenta GmbH, 2014)	36
1.3	Apartment equipped with a smart sensor network	37
1.4	Emergency detection in a domestic environment (source: Vitracom AG, 2015)	37
1.5	Standard perspective camera versus omnidirectional camera in a home environment	39
1.6	Floor plan of a single apartment equipped with optical stereo vision sensors (indicated by their camera frames)	40
1.7	Exemplary application fields (upper left: home automation, upper right: industry, lower left: security in public transportation, lower right: security in public buildings)	40
2.1	Coordinate transformations in the geometrical imaging process	46
2.2	Projection process	47
2.3	Radially symmetric projections	49
2.4	World coordinate system	50
2.5	Perspective camera model	52
2.6	Decomposition of P (row vectors)	54
2.7	Equiangular camera model	55
2.8	Omnidirectional image	55
2.9	Epipolar geometry	62
2.10	Parallel camera configuration	63
2.11	Parallel camera configuration	64
3.1	Human and machine based stereo vision [1]	68
3.2	The principle of stereo calibration	70
3.3	The principle of stereo rectification	72
3.4	Rectified images with exemplary epipolar lines	75
3.5	The principle of disparity computation	76

3.6	Colour coded disparity map	77
3.7	Triangulation	79
3.8	Horopter	81
3.9	Example quantization error	83
3.10	Sample probability distribution of δZ at $Z = 2.5m$ (2D)	85
3.11	Sample probability distribution of δZ at $Z = 2.5m$ (3D)	85
4.1	A virtual perspective camera	91
4.2	Summary of the forward transformation algorithm	91
4.3	Result forward projection	93
4.4	Summary of the backward transformation algorithm	94
4.5	Reprojection of 2D target image coordinates to 3D camera coordinates onto unit sphere	95
4.6	Projection of 3D camera coordinates to 2D source image coordinates . . .	95
4.7	Result backward projection	96
4.8	Exemplary lookup-table for the backward mapping of target pixel x- coordinates	96
4.9	Summary of the fast backward transformation algorithm	97
4.10	Reprojection of sub-sampled 2D target image coordinates to 3D camera coordinates onto unit sphere	98
4.11	Projection of sub-sampled 3D camera coordinates to 2D source image coordinates	98
4.12	Exemplary reduced lookup-table for the fast backward mapping of target pixel x-coordinates	99
4.13	Residuum error for the fast mapping of pixel x-coordinates	100
4.14	The total pixel error depending on elevation from equator to south pole for each power of reduction	101
4.15	Sample area of the target image in the source image	103
4.16	Sample interval of the target image in the source image (left: x-direction, right: y-direction)	105
4.17	Overlapping contour plots of $ \Delta x_{img}^{omni} $ and $ \Delta y_{img}^{omni} $ for different ratios of pixel sizes p_x^{per}/p_x^{omni} and p_y^{per}/p_y^{omni} , an generic projection model for the source camera, an under-sampling strategy (left) as well as over-sampling strategy (right) for the perspective target camera	107
4.18	Sample interval of the target image in the source image (left: x-direction, right: y-direction, top: under-sampling, bottom: over-sampling)	108
4.19	Live virtual camera generation [80]	109
4.20	Measured processing times for the generation of perspective views with two different resolutions versus power of reduction	110
4.21	Automatic people tracking system employing an omnidirectional imaging device and a vPTZ [82]	111

4.22	Distribution of detection rates using an omnidirectional image directly (left) compared to the use of virtual perspective views (right) [82]	111
5.1	Overview of categorizations for omni-stereo vision approaches	115
5.2	H-binocular omni-stereo with panoramic views	117
5.3	Example for horizontally-aligned omni-stereo [52]	118
5.4	V-binocular omni-stereo with panoramic views	119
5.5	Example for vertically-aligned omni-stereo [84]	120
5.6	Binocular omni-stereo with hemispherical views	121
5.7	Epipolar lines of binocular omni-stereo with hemispherical views	121
5.8	Relationship between azimuth angle ϕ_l and qualitative error measurement δZ	122
5.9	Example for binocular omni-stereo with hemispherical views (top left: left image, top right: right image, bottom left: left rectified image, bottom right: right rectified image) [145]	123
5.10	Trinocular omnidirectional stereo vision with hemispherical views	124
5.11	Example for trinocular omnidirectional stereo vision with hemispherical views [67]	124
5.12	Epipolar planes in an binocular stereo vision system with hemispherical views	126
5.13	Cylindrical rectification	127
5.14	Horizontal projection behaviour for different rectification methods	129
5.15	System configuration with physical omnidirectional cameras	131
5.16	System configuration with two virtual rectified cameras	132
5.17	System configuration with six virtual rectified cameras	133
6.1	Virtual 3D test environment equipped with a spherical stereo vision sensor	136
6.2	Overlapping contour plots of $ \Delta x_{img}^{omni} $ and $ \Delta y_{img}^{omni} $ for different ratios of pixel sizes p_x^{rect}/p_x^{omni} and p_y^{rect}/p_y^{omni} , an equiangular projection model for the source camera, a cylindrical projection model for the target camera, an under-sampling strategy and $FOV_h = 60^\circ$ (left) and $FOV_h = 90^\circ$ (right)	141
6.3	Colour coded sampling interval $ \Delta x_{img}^{omni} $ (left) and $ \Delta y_{img}^{omni} $ (right) for the optimal pixel ratios p_x^{cyl}/p_x^{omni} and p_y^{cyl}/p_y^{omni} , an equiangular projection model for the source camera, a cylindrical projection model for the target camera, an under-sampling strategy and $FOV_h = 60^\circ$	142
6.4	Calculated optimal pixel ratios p_x^{rect}/p_x^{omni} and p_y^{rect}/p_y^{omni} of physical and virtual pixel dimensions based on different horizontal fields of view FOV_h	145
6.5	Calculated disparity ranges based on different horizontal fields of view	146
6.6	Calculated reduced horizontal field of view based on different horizontal fields of view	146
6.7	Summary of the spherical depth map generation algorithm	147

6.8	Result of the equiangular imaging process ($\text{Cam}_0^{\text{omni}}$, $\text{Cam}_1^{\text{omni}}$ and $\text{Cam}_2^{\text{omni}}$)	148
6.9	Result of the epipolar equidistant imaging process ($\text{Cam}_{0a}^{\text{rect}}$, $\text{Cam}_{1a}^{\text{rect}}$, $\text{Cam}_{2b}^{\text{rect}}$, $\text{Cam}_{0b}^{\text{rect}}$, $\text{Cam}_{1c}^{\text{rect}}$, $\text{Cam}_{2c}^{\text{rect}}$)	149
6.10	Result of the stereo correspondence operation ($\mathbf{D}_{0a}^{\text{rect}}$, $\mathbf{D}_{2b}^{\text{rect}}$ and $\mathbf{D}_{1c}^{\text{rect}}$) for the epipolar-equidistant rectified images	150
6.11	Result of the triangulation operation ($\mathbf{Z}_{0a}^{\text{rect}}$, $\mathbf{Z}_{2b}^{\text{rect}}$ and $\mathbf{Z}_{1c}^{\text{rect}}$) for the epipolar equidistant disparity maps	151
6.12	Result of the projection process of rectified depth maps ($\mathbf{Z}_{0a}^{\text{omni}}$, $\mathbf{Z}_{2b}^{\text{omni}}$ and $\mathbf{Z}_{1c}^{\text{omni}}$)	152
6.13	Merged depth map $\mathbf{Z}_0^{\text{omni}}$	153
6.14	3D re-projection of the omnidirectional depth map $\mathbf{Z}_0^{\text{omni}}$	154
6.15	Hemispherical measurement space around the sensor with an exemplary distance of $Z = 2.5m$	155
6.16	Maximum mapping error $\Delta \hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera, a $\text{FOV}_H = 83.20^\circ$ and a baselength of $b = 15cm$	157
6.17	Maximum mapping error $\Delta \hat{\mathbf{X}}_{\text{img}}^{\text{omni}}$ for multiple measurement distances and camera configurations	158
6.18	Maximum distance error $ \delta Z_{\text{max}}^{\text{omni}} $ for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 15cm$; Left: our approach ($\text{FOV}_H = 83.20^\circ$), Right: Approach of Li Shigang [67]	159
6.19	Minimum distance error $ \delta Z_{\text{min}}^{\text{omni}} $ for an exemplary distance of $Z_0^{\text{omni}} = 2.5m$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 15cm$; Left: our approach ($\text{FOV}_H = 83.20^\circ$), Right: Approach of Li Shigang [67]	160
6.20	Largest minimal Error $ \delta Z_{\text{min}}^{\text{omni}} $ and maximal Error $ \delta Z_{\text{max}}^{\text{omni}} $ for multiple measurement distances, an equiangular projection model for the source camera and two baselengths	161
6.21	Largest minimal Error $ \delta Z_{\text{min}}^{\text{omni}} $ and maximal Error $ \delta Z_{\text{max}}^{\text{omni}} $ for multiple measurement distances, an generic projection model for the source camera and two baselengths	161
6.22	Maximum distance error $ \delta Z_{\text{max}}^{\text{omni}} $ for an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 15cm$ depending on different measurement distances Z and the parameter FOV_H of the virtual cameras	162
7.1	A spherical stereo vision demonstrator	164

7.2	Sample images from the professorships lab apartment	164
7.3	Special three-dimensional calibration target	165
7.4	Geometrical configuration for extrinsic calibration of the physical and virtual cameras	166
7.5	Example calibration images for $\text{SCam}_a^{\text{per}}$ with original images from $\text{Cam}_0^{\text{omni}}$ ($\mathbf{Hi}_0^{\text{omni}}$) and $\text{Cam}_1^{\text{omni}}$ ($\mathbf{Hi}_1^{\text{omni}}$) as well as pre-rectified images from $\text{Cam}_{0a}^{\text{per}}$ ($\mathbf{Hi}_{0a}^{\text{per}}$) and $\text{Cam}_{1a}^{\text{per}}$ ($\mathbf{Hi}_{1a}^{\text{per}}$)	168
7.6	Computer vision framework with running spherical stereo project	172
7.7	Reference image with calculated hemispherical depth map	173
7.8	Hemispherical depth map without (left) and with (right) post processing median filter of kernel size 5×5	173
7.9	Re-projection of reference image	174
7.10	Example case for an erroneous fusion result	175
8.1	Occlusion case that might appear if a depth measurement information obtained by one imaging device (measurement camera) is mapped to another device (reference camera)	179
8.2	Overlapping depth maps (red, green and blue) with highlighted area of maximal disparity	181
8.3	Maximum distance error $ \delta Z_{\text{max}}^{\text{omni}} $ for an exemplary distance of $Z_0^{\text{omni}} =$ $2.5m$, an equiangular projection model for the source camera, an epipolar equidistant projection model for the target camera and a baselength of $b = 10cm$; Left: the author's three-view approach ($\text{FOV}_H = 75.40^\circ$), Right: two-view approach with restricted FOV that complies with an error smaller than two times of the maximum error observable in the left plot	185
8.4	3D reprojection of an artificial home environment observed by a two-view stereo vision setup with hemispherical views ($b = 10cm$)	186
8.5	Structure of the proposed algorithm	187
8.6	People hypothesis generation for 3D localization	188
A.1	Random variables $x_{\text{img},l}$ and $x_{\text{img},r}$ of uniform distribution	192
A.2	Normalized probability density function	196

List of Tables

1.1	Markets and application fields	38
2.1	Radially symmetric projection models	48
2.2	Generic projection models	56
2.3	Summary of common methods for perspective camera calibration	59
2.4	Summary of common methods for omnidirectional camera calibration	60
2.5	Summary of publications that treat common two-view geometry	61
3.1	Overview of a stereo vision system	69
3.2	Summary of publications that consider stereo calibration	71
3.3	Summary of publications that consider stereo rectification	73
3.4	Evaluation dataset of Middlebury website	78
4.1	Summary of scientific contributions to methods of camera model conversion	89
4.2	Results of the intrinsic generic camera models estimation	102
4.3	Physical pixel dimensions	102
4.4	Exemplary FOV for the target camera	103
4.5	Results of the intrinsic generic camera model calculation	107
4.6	Virtual pixel dimensions	107
5.1	Relevant publications that categorize omni-stereo vision approaches	114
5.2	Comparison of omni-stereo principles [146]	116
6.1	Intrinsic parameters of the equiangular cameras $\text{Cam}_i^{\text{omni}}$	137
6.2	Physical pixel dimensions of the equiangular cameras $\text{Cam}_i^{\text{omni}}$	137
6.3	Calculated fields of view FOV_h for the virtual cameras $\text{Cam}_j^{\text{rect}}$ according to the variety of configurations	143
6.4	Calculated virtual pixel size ratios and image sizes	143
6.5	Intrinsic parameters of the virtual cameras $\text{Cam}_j^{\text{rect}}$	144
6.6	Pixel dimensions of the virtual cameras $\text{Cam}_j^{\text{rect}}$	144
7.1	Results of the intrinsic generic camera models estimation	166
7.2	Image calibration error	169

7.3	Rectification results	170
7.4	Calculated fields of view FOV_h for the virtual cameras Cam_j^{rect} according to the variety of configurations	171
7.5	Calculated virtual pixel size ratios and image sizes	171
7.6	Intrinsic parameters of the virtual cameras Cam_j^{rect}	171
7.7	Pixel dimensions of the virtual cameras Cam_j^{rect}	171
7.8	Performance measurements	175
8.1	Savings in terms of stereo correspondence operations	180
B.1	Related works in h-binocular omnidirectional stereo vision with panoramic views	197
B.2	Related works in v-binocular omnidirectional stereo vision with panoramic views	198
B.3	Related works in binocular omnidirectional stereo vision with hemispherical views	200
B.4	Related works in trinocular omnidirectional stereo vision	201
B.5	N-ocular omnidirectional stereo vision	202
B.6	Omnivergent stereo vision	203
B.7	Dynamic stereo vision	205
B.8	Omnidirectional stereo vision with special optical solutions	206

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten:

.....

.....

.....

.....

Weitere Personen waren an der Abfassung der vorliegenden Arbeit nicht beteiligt. Die Hilfe eines Promotionsberaters habe ich nicht in Anspruch genommen. Weitere Personen haben von mir keine geldwerten Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

.....
Ort, Datum

.....
Unterschrift

Theses

1. The application of a trinocular camera configuration (three-view camera setup) makes it possible to realize a three-dimensional space monitoring by means of one sensor.
2. A full hemispherical depth map can be generated by using a trinocular camera system in geometry of an equilateral triangle.
3. The creation of virtual images by converting the underlying projection model enables the direct application of standard stereo processing methods for omnidirectional sensors.
4. In contrast to the cylindrical projection principle, the utilization of an epipolar equidistant imaging model for image rectification preserves considerably more resolution of the omnidirectional source image.
5. The generation of virtual views can be significantly accelerated, as soon as the number of transformed coordinates is reduced. The resulting error highly depends on the actual level of reduction as well as the projection models utilized.
6. The composition of a hemispherical depth map by means of three partial depth maps, rather than two full depth maps, reduces the number of required numerical operations for the stereo correspondence process significantly. Furthermore, this approach avoids a computationally intensive step for merging the distance information.
7. The merge of multiple depth maps, generated by distributed stereo cameras, to a single depth map is possible in principle. This implies that the mapping of each distance information, characterized by a statistical measurement error, to the target map results in a negligible uncertainty of less than 1 pixel.

Thesen

1. Die Verwendung einer trinokularen Kamerakonfiguration (Dreikamerasystem) ermöglicht es, eine dreidimensionale Raumüberwachung mit lediglich einem Sensor zu realisieren.
2. Die Erzeugung einer vollständigen hemisphärischen Tiefenkarte unter Verwendung eines trinokularen Kamerasystems in Geometrie eines gleichseitigen Dreiecks ist möglich.
3. Die Generierung virtueller Abbildungen durch Konvertierung des zugrundeliegenden Projektionsmodells ermöglicht die direkte Anwendung von Standardalgorithmen der Stereoberechnung in der omnidirektionalen optischen Überwachung.
4. Im Gegensatz zum zylindrischen Abbildungsmodell erhöht die Nutzung einer epipolar-equidistanten Projektionsmethode die nutzbare Auflösung des omnidirektionalen Quellbildes für das Stereoverfahren signifikant.
5. Die Erzeugung virtueller Abbildungen kann erheblich beschleunigt werden, wenn die Anzahl der transformierten Koordinaten reduziert wird. Der dabei entstehende Fehler ist stark abhängig vom tatsächlichen Grad der Reduktion sowie von den eingesetzten Projektionsmodellen.
6. Die Berechnung einer hemisphärischen Tiefenkarte unter Verwendung dreier partieller Tiefenkarten, anstatt zweier vollständiger Tiefenkarten, reduziert den Rechenaufwand für die Stereokorrespondenzberechnung signifikant. Des Weiteren vermeidet dieses Vorgehen einen rechenintensiven Fusionsprozess von Entfernungsinformationen.
7. Das Zusammenführen mehrerer Tiefenkarten verteilter Stereokameras zu einer einzigen Tiefenkarte ist prinzipiell möglich. Voraussetzung dafür ist, dass die Zuordnung jeder, durch einen statistischen Messfehler charakterisierten, Entfernungsinformation zur fusionierten Tiefenkarte lediglich eine kleine Unsicherheit (≤ 1 Pixel) aufweist.

Curriculum Vitae

Personal Profile

Name	Michel Findeisen
Date and Place of Birth	November 17, 1979 in Rochlitz
Family Status	Single
Nationality	German
Address	Crossener Strasse 22; 09306 Erlau OT Milkau
Mobile	0176 20 16 70 25
Email	michel.findeisen@googlemail.com

Education and Career

May 2009 - Sept 2014	Scientific Assistant at Chemnitz University of Technology, Professorship of Digital Signal Processing and Circuit Technology
Oct 2008 - Mar 2009	Internship at Siemens Numerical Control Ltd (SNC), RD IT, Nanjing, China
Oct 2003 - Sept 2008	University Studies of the diploma degree program <i>Informations- und Kommunikationstechnik</i> at Chemnitz University of Technology
Oct 2001 - Sept 2003	Employment at Siemens AG, Department A& D SE WKC B in Chemnitz as Associate Engineer
Oct 1999 - Sept 2001	Education at the <i>Siemens-Technik-Akademie</i> in Erlangen to obtain the degree <i>Associate Engineer of Computer Systems and Information</i>
Nov 1998 - Aug 1999	Military Service
Sept 1992 - July 1998	Education at the <i>Johann-Mathesius-Gymnasium</i> in Rochlitz to obtain the general qualification for university entrance
Sept 1986 - July 1992	Education at the <i>Grund- und Mittelschule</i> in Milkau