

Institut für Visualisierung und Interaktive Systeme
Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

Diplomarbeit Nr. 3114

Hintergrundsubtraktion mit PTZ-Kameras

Davaadorj Tsendragchaa

Studiengang:	Informatik
Prüfer:	Prof. Dr. Gunther Heidemann
Betreuer:	Dipl.-Inf. Benjamin Höferlin Dipl.-Inf. Daniel Kauker
begonnen am:	29. November 2010
beendet am:	31. Mai 2011
CR-Klassifikation:	I.4.6, I.4.8, I.4.3, F.1.2

Zusammenfassung

Die Hintergrundsubtraktion ist eine bekannte Technik zur Erkennung von Vordergrundobjekten in Videodaten, die durch Subtraktion des Hintergrundmodells vom Kamerabild die Vordergrundobjekte extrahiert.

Ziel dieser Arbeit war die Entwicklung eines neuen Verfahrens zur Hintergrundsubtraktion in PTZ-Kameras (eng. *Pan-Tilt-Zoom cameras*).

Das entwickelte Verfahren verwendet ein sphärisches Hintergrundmodell, aus dem die Sicht eines beliebigen Winkels gesehen vom Standpunkt der Kamera aus hergeleitet werden kann. Das sphärische Hintergrundmodell ermöglicht es, den umgebenden Raum der Kamera vollständig zu modellieren. Daraus resultieren Vorteile hinsichtlich Speicherbedarf und Flexibilität im Vergleich zu anderen Lösungsansätzen. Ein weiterer Vorteil ist das einfache Erstellen von Panoramabildern. Andere Lösungsansätze werden im Kapitel 3 "Verwandte Arbeiten" näher beschrieben.

PTZ-Kameras liefern meist Bilder mit Bewegungsunschärfe. Im Rahmen dieser Arbeit wurden die Einsatzmöglichkeiten des Deblurring-Verfahrens in Hintergrundsubtraktion mit PTZ-Kameras untersucht. Es wird vorgeschlagen die Degradationsfunktion (eng. *Point Spread Function, PSF*) aus den optischen Flüssen der Bilder zu berechnen.

Für das aktuelle Kamerabild wird das entsprechende Hintergrundmodell mithilfe der Bildregistrierung unter Verwendung von "SURF"-Merkmalen ermittelt.

Um die Abbildung zwischen dem Hintergrundmodell und dem aktuellen Kamerabild zu ermöglichen, wird die Kartenprojektion "Rektangulärprojektion" benutzt, welche die Oberfläche einer Kugel auf eine 2D Fläche überträgt.

Abstract

Background subtraction is a well known technique for recognizing foreground objects in digital video, in which the background image is subtracted from the camera frame to reveal foreground objects.

The goal of the thesis is the development of a new background subtraction method suitable for PTZ (Pan-Tilt-Zoom) cameras. The proposed method uses a spherical background model from which the camera view at any angle can be generated. The ability to model the camera surroundings completely creates advantages in terms of memory usage and flexibility. In addition, the background model can be used to easily create panorama views with PTZ cameras. Related works and their shortcomings are discussed in detail in Chapter 3.

PTZ cameras often create videos with blurring effects. The thesis proposes the use of a deblurring technique as a solution, in which the PSF (Point Spread Function) is calculated from the optical flow of the frames.

To find the background model corresponding to the current camera frame, image registration technique based on SURF properties is used.

The relation between the current camera view and the background is created using the Equirectangular Projection, which projects the surface of a spherical body on a 2D surface.

Inhaltsverzeichnis

1. Einleitung	11
1.1. Motivation	11
1.2. Ziel der Arbeit	12
1.3. Aufbau	14
2. Grundlagen	17
2.1. Kameramodell	17
2.1.1. Grundbegriffe der Fotografie	17
2.1.2. Kameraparameter	20
2.1.3. Lochkameramodell	21
2.1.4. Optische Verzerrung und Kamerakalibrierung	23
2.2. Transformationen	26
2.2.1. Transformation zwischen kartesischen und sphärischen Koordinaten- systemen	26
2.2.2. Starre 3D Transformationen	27
2.2.3. Projektive Transformation	29
2.2.4. Hough Transformation	31
2.3. Kartenprojektion	32
2.3.1. Eigenschaften einer Karteprojektion	33
2.3.2. Projektionsflächen	33
2.3.3. Einige Kartenprojektionen	38
3. Verwandte Arbeiten	39
3.1. Hintergrundsubtraktion	39
3.2. Deblurring	41
4. Erstellung des Hintergrundmodells	43
4.1. Grundidee	43

4.2.	Messungen der Kameraeigenschaften	45
4.2.1.	Test der Parallaxenverschiebung	46
4.2.2.	Kamerakalibrierung	51
4.2.3.	Berechnung des initialen Bildwinkels	56
4.3.	Ermittlung der Kamerasichtrichtung	59
4.4.	Berechnung der Zoomstufe	70
4.5.	Sphärische Projektion	77
4.5.1.	Projektion eines Einzelbildes auf eine Sphäre	77
4.5.2.	Größe des Hintergrundmodells	80
4.5.3.	Abbildungen der Sphäre auf das Hintergrundmodell	80
4.6.	Kubische Projektion	85
5.	Deblurring	89
5.1.	Entfaltungsfiler	90
5.1.1.	Inverse-Filtering	90
5.1.2.	Wiener Filter	91
5.1.3.	Richardson-Lucy Algorithmus	91
5.2.	Eigenes Verfahren	92
6.	Zusammenfassung und Ausblick	95
A.	Hintergrundsubtraktion	99
A.1.	Einige Verfahren	103
A.2.	Das Verfahren "Gaussian Mixture Model"	106
	Literaturverzeichnis	113

Abbildungsverzeichnis

1.1. Hintergrundsubtraktion: (a) Hintergrundbild, (b) Einzelbild, (c) Vordergrundmaske. Quelle: [Br1]	12
1.2. (a) PTZ-Kamera "SONY SNC-RZ30", (b) Bild mit und ohne Bewegungsunschärfe. Quelle: (a)Benutzerhandbuch von SONY SNC-RZ30 (b)[CL09]	13
1.3. Gesamt Schema des Systems	15
2.1. Brennweite [Wik11b]	18
2.2. Abhängigkeit von Bildwinkel und Brennweite [Wik11a]	19
2.3. Optische Abbildung einer idealen Linse [Wik11f]	19
2.4. radiale Linsenverzerrung [Wik11i]	24
2.5. Beispiele für 3D, 2D Kalibriermuster	25
2.6. Sphärisches Koordinatensystem	27
2.7. Rechthändiges Koordinatensystem (Rechtssystem genannt). Im Rechtssystem wird gegen den Uhrzeigersinn rotiert. Eine elementare Rotation ist eine Rotation um eine einzelne Achse X, Y oder Z.	28
2.8. Drehbewegung der Kamera: Eine projektive Transformation ermöglicht die Zusammensetzung beider Bilder. Quelle: [Scho5]	30
2.9. Hough Transformation von zwei Linien in einem Raum aus Winkel und Abstand zur Geraden. Quelle: [Wik11d]	32
2.10. Projektionen: (a) gnomonische, (b) orthografische, (c) stereografische, (d) flächentreue und (e) mittabstandstreue Azimutalprojektion	34
2.11. Beispiele für die 5 Projektionsarten in gleicher Reihenfolge wie bei vorheriger Abbildung. Quelle: [Kö] und [Wik11g]	34
2.12. Zylindrische Abbildung	35
2.13. (a) Mercatorprojektion, (b) Rektangularprojektion Quelle: [Wik11j]	35
2.14. Kegelprojektion	36

2.15. Einige Kegelprojektionen: (a) Lambertsche, (b) Längentreue, (c) Albers. Quelle: [Wik11e]	37
2.16. Einige interessante Kartenprojektionen. Quelle: [Furo9]	38
4.1. Links: Projektion eines Einzelbildes auf einer Sphäre, Rechts: Abbildung von Oberfläche der Sphäre auf einer Ebene	45
4.2. Schwenkbereich und Neigewinkel der PTZ-Kamera "SONY SNC-RZ30" Quelle: Benutzerhandbuch von SONY SNC-RZ30	46
4.3. (a) Nodalpunkt, (b) Beispiel für Parallaxenverschiebung	47
4.4. Links: Vor der Kameradrehung sind mehrere Fenster sichtbar, Rechts: Fenster sind nach Kameradrehung durch Parallaxenverschiebung verdeckt.	48
4.5. Parallaxenverschiebung der Augen: Links - die Sicht mit dem linken Auge, Rechts - die Sicht mit dem rechten Auge. Quelle: [par]	49
4.6. Parallaxenverschiebung bei der verwendeten PTZ-Kamera für den kleinsten Zoomfaktor	49
4.7. Parallaxenverschiebung bei Neigebewegung	50
4.8. Parallaxenverschiebung für großen Zoomfaktor	50
4.9. Links: Schachbrettmuster für Kamerakalibrierung, Rechts: die innere Ecken sind erkannt	52
4.10. Bilder vom Schachbrettmuster mit verschiedenen Orientierungen für Kamerakalibrierung	53
4.11. Links: ein von der Kamera aufgenommene verzerrtes Bild, Rechts: entzerrtes Bild	56
4.12. Links: Erstellung einer Sphäre für einen zu kleinen Zoomfaktor. Die Bilder müssen vor der Projektion verkleinert werden, Rechts: Umgekehrter Fall: Die Bilder müssen im richtigen Abstand B vom Zentrum projiziert werden.	57
4.13. Links: Bild wird auf einen zu großen Bereich (roter Bereich im Kreis) projiziert, weil der Bildwinkel zu groß geschätzt war. Rechts: Im umgekehrten Fall wird das Bild auf einen zu kleinen Bereich projiziert (roter Bereich im Kreis).	58
4.14. die Winkel (θ, ϕ) beschreiben die Kamerasichtrichtung und entsprechen der Bildmitte	60
4.15. Schritte zur Berechnung der Kamerasichtrichtung	62
4.16. Merkmale und deren Zuordnung	63
4.17. Transparente Zusammensetzung von zwei Bildern	65

4.18. Darstellung der Geraden durch eine Hough-Transformation im Dualraum Winkel und Abstand zum Bildmittelpunkt	67
4.19. Winkel der Linien und deren Häufigkeiten. Meiste Linien haben einen Winkel von 178°	68
4.20. eliminierte Paare sind durch rote Linien verbunden.	68
4.21. Drehung der Kamera um ihre optische Achse (Kamera schaut nach unten und schwenkt)	69
4.22. Linien, die richtige Paare verbinden, haben verschiedene Winkel und schnei- den sich daher auch.	69
4.23. aktuelles und vorheriges Einzelbilder mit unterschiedlichen Zoomstufen. Bildweite und Bildwinkel ändern sich beim Zoomen mit. Gegenstand G hat in beiden Einzelbilder unterschiedliche Koordinaten (P_1, P_2)	71
4.24. Das rechte Bild (mit Bildwinkel $41,5^\circ$) wird im linken eingezoomt. Durch rote Linien verbundene Paare sind als falsche Paare erkannt und eliminiert.	72
4.25. Für gefundene 120 Paare jeweils ist der neue Bildwinkel berechnet. Durch falsche Paare verursachte Schwankungen sind sichtbar.	73
4.26. (a) Paare sind durch Linien in einem Bild verbunden, (b) Eliminierung der falschen Paare mit der Quotient $\frac{B_1}{B_2}$	74
4.27. Zusammensetzung mit projektiven Transformation. Das eingezoomte Bild ist nicht rechteckig wegen der falsch zugeordneten Paare.	74
4.28. Zusammensetzung der Bilder mit vielen erkannten Merkmalen	76
4.29. Projektion eines Einzelbildes auf eine Sphäre mit Radius R	77
4.30. Abbildung eines Einzelbildes auf Sphäre	78
4.31. Projektion eines Pixels auf die Sphäre	79
4.32. Hintergrundmodell bei Rektangularprojektion	81
4.33. Abbildungen eines Beispielmusters auf das Hintergrundmodell bei verschie- denen θ Winkeln: (a) $\theta = 0^\circ$, (b) $\theta = 70^\circ$, (c) $\theta = 90^\circ$	82
4.34. Abbildungen mehrerer Einzelbilder	83
4.35. Abbildungen mehrerer transparenter Einzelbilder	84
4.36. Hintergrundmodell bei Sinusoid	84
4.37. Abbildungen mit der Kartenprojektion Sinusoid bei $\theta = 90^\circ$	85
4.38. Hintergrundmodell bei der Segment-Kartenprojektion	85
4.39. Abbildungen mit der Kartenprojektion Segment bei $\theta = 90^\circ$	85
4.40. Hintergrundmodell bei der kubischen Projektion	86

4.41. Abbildungen der kubischen Projektion eines Musterbildes	87
4.42. Abbildungen bei der kubischen Projektion	88
5.1. Links: Faltung des idealen Bildes mit PSF. Rechts: Entfaltung des unscharfen Bildes mit PSF [Wik11c]	90
5.2. Bestimmung des optischen Flusses	93
A.1. Links: Hintergrundbild, Mitte: Einzelbild, Rechts: Differenzbild. Quelle: [hgS]	100
A.2. Links: Differenzbild, Mitte: Binärbild bei einem großen Schwellenwert, Rechts: Binärbild bei einem kleinen Schwellenwert. Quelle: [hgS]	101
A.3. Links: Binärbild mit Störungen, Rechts: Die Störungen mithilfe eines Median- filter beseitigt. Quelle: [hgS]	102

1. Einleitung

1.1. Motivation

Neue Entwicklungen im Bereich Computer Vision ermöglichen intelligente und automatisierte Überwachungssysteme, die rund um die Uhr beobachten und die bei Abweichungen von Normalverhalten automatisch reagieren können. Der Einsatz solcher Systeme wird immer beliebter und verbreitet sich in verschiedenen Bereichen, beginnend von der traditionellen Überwachung von Menschen in öffentlichen Bereichen bis hin zur Vermessung und Kontrolle in Industrie und Forschung.

Die Trennung der Vorder- und Hintergründe in Einzelbildern einer Videoaufnahme ist eine notwendige Aufgabe im Bildverstehen, um bewegte Objekte in Bildern erkennen und verfolgen zu können. Diesen Vorgang nennt man Segmentierung von Vordergrundobjekten. Er wird häufig in automatisierten Überwachungssystemen angewendet. Unter Segmentierung versteht man die Unterteilung eines Bildes in homogene Bereiche. Sie dient zur Untersuchung der Form des segmentierten Objekts und ist der erste Schritt bei den Verfahren zur Bildanalyse und Klassifizierung. Außerdem ermöglicht die Segmentierung die Verwendung spezieller Algorithmen auf bestimmten Bildteilen.

Die Hintergrundsubtraktion ist eine bekannte Technik, um Vordergrundobjekte aus Videoaufnahmen zu segmentieren. Die Technik ermöglicht eine Segmentierung der Vordergrundobjekte aus den Videodaten (d.h. nur mit Farb- bzw. Intensitätswerten der Pixel) und benötigt keine Tiefeninformationen. Zur Segmentierung der Vordergrundobjekte wird das aktuelle Kamerabild mit einem Hintergrundbild verglichen und diejenigen Bereiche des aktuellen Bildes, die eine große Abweichung vom Hintergrundbild zeigen, als Vordergrund ausgewählt. Die Hintergrundsubtraktion ist gut für Videodaten geeignet, die von einer unbeweglichen Kamera aufgenommen wurden. Bei beweglichen Kameras ist die Verwendung dieser Technik jedoch problematisch, weil bei der Kamerabewegung das Hintergrundmodell des

neuen Aufnahmebereichs bekannt sein muss, um Vordergrundobjekte des neuen Bereichs segmentieren zu können.

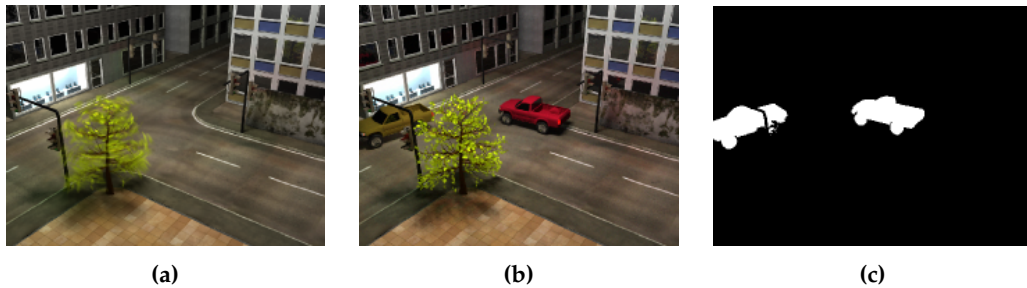


Abbildung 1.1.: Hintergrundsubtraktion: (a) Hintergrundbild, (b) Einzelbild, (c) Vordergrundmaske. Quelle: [Br1]

1.2. Ziel der Arbeit

Ziel dieser Diplomarbeit ist die Realisierung der Hintergrundsubtraktion bei den PTZ-Kameras (Abbildung 1.2a). PTZ ist eine englische Abkürzung und steht für **P**an-**T**ilt-**Z**oom, was auf deutsch Schwenken-Neigen-Zoomen bedeutet. PTZ-Kameras sind bewegliche Kameras, deren Sichtfeld über diese drei Funktionen fern gesteuert wird. In den vergangenen Jahren wurden solche Kameras bei Überwachungssystemen häufiger eingesetzt. Daher ist es wünschenswert, die Hintergrundsubtraktion bei diesen Kameras zu ermöglichen. Durch die PTZ-Funktionalität entstehen im allgemein zwei Probleme:

- Die Erstellung eines passenden Hintergrundmodells, das die gesamte Umgebung der PTZ-Kamera umfasst. Abbildungen müssen hierfür gefunden werden, mit denen das aktuelle Kamerabild auf den entsprechenden Teil des Hintergrundmodells abgebildet bzw. das entsprechende Hintergrundmodell zum aktuellen Kamerabild aus dem gesamten Hintergrundmodell gefunden werden können.
- Ermittlung der aktuellen Sichtrichtung und der Zoom-Stufe der PTZ-Kamera. Das heißt, man muss die PTZ-Parameter der Kamera (Schwenk- und Neigungswinkel und Zoomstufe) herausfinden. Mit diesen drei Größen und mithilfe der geeigneten Abbildung kann man den richtigen Ausschnitt aus dem gesamten Hintergrundmodell erhalten, der zu dem aktuellen Kamerabild passt. Zur Lösung dieses Problems kann

ein Bildregistrierungsverfahren verwendet werden, mit dem die aufgenommenen Kamerabilder zusammengesetzt (eng. stitching) werden können.

Eine mögliche Lösung des ersten Problems ist die Darstellung der Umgebung von PTZ-Kamera als Sphäre und die Verwendung einer Kartenabbildung (eng. map projection). Mithilfe einer Kartenabbildung kann die Oberfläche eines drei dimensional Objekts auf einer Ebene abgebildet werden. Die Auswahl einer geeigneten Kartenabbildung ist aber mit dem zweiten Problem verbunden.

Hat man das, dem aktuellen Kamerabild entsprechende, Hintergrundmodell gefunden, können die Vordergrundobjekte des aktuellen Bildes mittels Hintergrundsubtraktion segmentiert werden. Hierfür gibt es verschiedene Verfahren, die ihre Hintergrundmodelle unterschiedlich aufbauen und aktualisieren.



Abbildung 1.2.: (a) PTZ-Kamera "SONY SNC-RZ30", (b) Bild mit und ohne Bewegungsunschärfe. Quelle: (a)Benutzerhandbuch von SONY SNC-RZ30 (b)[CLog]

Das zweite Thema der Diplomarbeit ist das Reduzieren von Bewegungsunschärfe (eng. motion blur). Ein Beispiel ist in Abbildung 1.2b zu sehen. Der rechte Teil des Bildes enthält Bewegungsunschärfe. Die Bewegungsunschärfe bei den Bildern entstehen durch die relative Bewegung der Kamera zur Szene, während der Belichtung.

Die PTZ-Kameras liefern meist Bilder mit Bewegungsunschärfe. Der Grund dafür ist einerseits, dass die PTZ-Kameras im allgemein niedrige Bildwiederholfrequenzen haben. Die dadurch entstandene Bewegungsunschärfe sieht man bei Szenen, in denen sich Objekte bewegen. Andererseits verursacht die Bewegung der PTZ-Kamera die Bewegungsunschärfe. Bei der Kamerabewegung wird das gesamte Bild unscharf.

Der Wiederherstellungsprozess, der aus Bildern mit Bewegungsunschärfe ideale Bilder ohne bzw. mit verminderten Bewegungsunschärfe erstellt, nennt man "motion deblurring". Es gibt zahlreiche Deblurring Verfahren. Das Hauptproblem beim Deblurring ist die korrekte Bestimmung der unscharfen Regionen. Da ein Bildregistrierungsverfahren für Hintergrundsubtraktion mit PTZ-Kamera notwendig ist, besteht hier die Möglichkeit, die Bewegungen der PTZ-Kamera zusammen zu ermitteln und mit ihnen die entstandene Bewegungsunschärfe zu vermindern.

1.3. Aufbau

In diesem Abschnitt wird ein Überblick über den Aufbau der vorliegenden Arbeit gegeben. In **Kapitel 2** werden die Grundbegriffe erläutert, die zum besseren Verständnis der Arbeit beitragen. Diese umfassen Kameraeigenschaften, Transformationen in und zwischen Kartesischen und Sphärischen Koordinatensystemen. Zudem werden verschiedene Kartenabbildungen aufgeführt, die für die Erstellung des Hintergrundmodells der PTZ-Kamera eine Rolle spielen.

In **Kapitel 3** werden verwandte Arbeiten vorgestellt, in denen auf die Hintergrundsubtraktion mit und ohne PTZ-Kamera eingegangen wird. Zusätzlich werden einige Arbeiten vorgestellt, die sich mit Deblurring-Verfahren beschäftigen. Der Aufbau der weiteren Kapitel entspricht dem Schema der gesamten Architektur des Systems in Abbildung 1.3. Die Abbildung 1.3 zeigt den allgemeinen Ablauf der Einzelbild-Verarbeitung. Jedes Einzelbild der PTZ-Kamera wird in der gegebenen Reihenfolge bearbeitet. Die Verarbeitung besteht aus Deblurring und Hintergrundsubtraktion. In Rahmen des Deblurring-Prozesses wird das aktuelle Einzelbild mithilfe eines Registrierungsverfahrens mit dem vorherigen Einzelbild verglichen, um Bewegungen bzw. den optischen Fluss des aktuellen Einzelbildes zu bestimmen. Anhand des berechneten optischen Flusses wird ein Entfaltungsfiter erstellt, womit die erkannte Bewegungsunschärfe des Einzelbildes beseitigt werden kann.

Im zweiten Schritt beginnt die Hintergrundsubtraktion mit dem aktuellen von Bewegungsunschärfe befreiten Einzelbild. Die Werte der PTZ-Parameter werden durch den Vergleich des aktuellen Einzelbildes mit dem vorherigen Einzelbild berechnet. Der für das aktuelle Einzelbild entsprechende Hintergrund kann aus dem gesamten Hintergrund mit einer Abbildung bestimmt werden, wenn die PTZ-Parameter für das aktuelle Einzelbild bekannt sind. Dann wird die Hintergrundsubtraktion mit dem erhaltenen Hintergrundmodell und mit

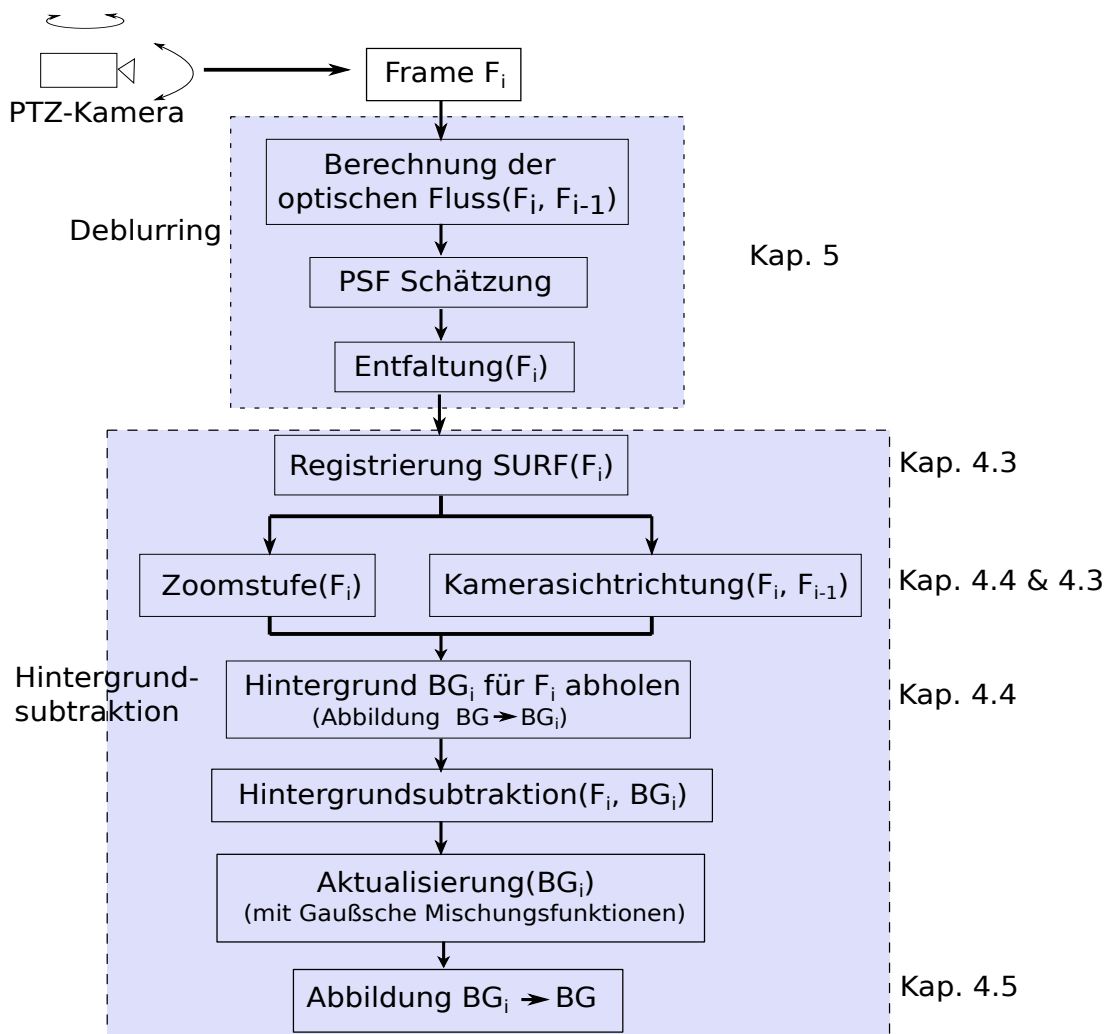


Abbildung 1.3.: Gesamt Schema des Systems

dem aktuellen Einzelbild durchgeführt. Die Aktualisierung des Hintergrundmodells wird entsprechend dem Verfahren von Stauffer[SG99] ("Gauß'sche Mischfunktionen") durchgeführt. Die Abbildung 1.3 gibt auch eine Übersicht darüber, welche Verarbeitungsschritte in welchem Kapitel erläutert werden.

Das **Kapitel 4** beschäftigt sich mit der Erstellung eines Hintergrundmodells für PTZ-Kamera mithilfe verschiedener Kartenabbildungen. Das Hintergrundmodell umfasst die ganze Umgebung der PTZ-Kamera. Die Kartenabbildungen ermöglichen für bekannte PTZ-Parameter

die Abbildungen zwischen dem gesamten Hintergrund und dem Hintergrund, der dem aktuellen Einzelbild entspricht.

Da die Bilder der PTZ-Kamera mit dem Hintergrund zusammenpassen müssen, wird vor der Erstellung des Hintergrundmodells überprüft, ob die Kamera verzerrungsfreie Bilder liefert und die Kamera sich um ihr optisches Zentrum dreht. Um dies zu gewährleisten muss die Kamera kalibriert und ein Test auf Parallaxenverschiebung gemacht werden. In dem Kapitel 4.2 wird Kamerakalibrierung und Parallaxenverschiebungstest beschrieben. Außerdem wird erklärt, wie die PTZ-Parameter für das aktuelle Einzelbild geschätzt werden.

In **Kapitel 5** wird das Deblurring-Verfahren im Detail beschrieben.

Anschließend fasst **Kapitel 6** die Ergebnisse der Arbeit zusammen und gibt einen Ausblick über mögliche Anknüpfungspunkte.

2. Grundlagen

Die vorliegende Arbeit stellt einen Lösungsansatz zur Hintergrundsubtraktion mit PTZ-Kameras mittels Projektionsverfahren dar. Deshalb sollte man die Grundbegriffe der Fotografie und die verschiedenen Projektionsarten kennen, um den Lösungsweg besser verstehen zu können.

2.1. Kameramodell

Kameras können durch eine bestimmte Projektion eine Szene auf einem Aufnahmemedium abbilden. Ein Kameramodell beschreibt diese Abbildung, wie 3D Weltpunkte auf 2D Bildpunkte projiziert werden.

Das einfachste Gerät mit dem optische Abbildungen erzeugt werden können, ist die Lochkamera. Eine Lochkamera besteht aus einem dunklen Raum mit einer kleinen Öffnung, durch die das Licht in den Raum eindringt.

Das Bild eines Gegenstandes, der vor dieser Öffnung der Lochkamera steht, wird auf die gegenüberliegende Wand umgedreht abgebildet. Die Öffnung nennt man Brennpunkt oder Projektionszentrum, die gegenüberliegende Wand wird als Bildebene bezeichnet. Die Lochkamera wird als mathematisches Modell für reale Kameras verwendet.

2.1.1. Grundbegriffe der Fotografie

Brennweite

Die Brennweite ist der Abstand zwischen der Linse und dem Punkt hinter der Linse, in dem die einfallenden Lichtstrahlen gebündelt werden. Die Brennweite f wird in Millimetern angegeben (siehe Abbildung 2.1).

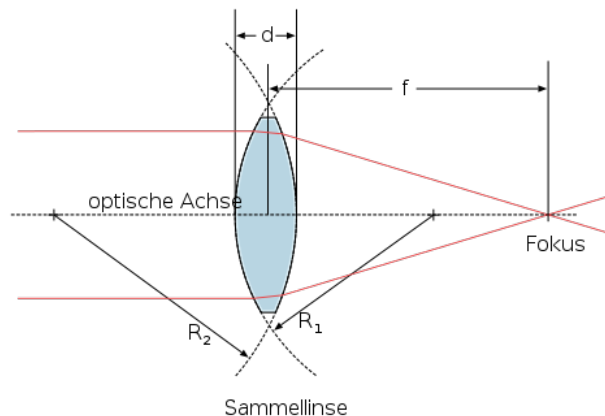


Abbildung 2.1.: Brennweite [Wik11b]

Bildwinkel

Unter dem Bildwinkel versteht man den Winkel zwischen den Bildrändern und dem optischen Mittelpunkt. Es wird unterschieden zwischen vertikalem, horizontalem und diagonalem Bildwinkel. Beim üblichen rechteckigen Bildformat ist mit dem Bildwinkel in der Regel der diagonale Bildwinkel gemeint.

Durch das Zusammenspiel von Bildwinkel und Brennweite wird die Größe des sichtbaren Bereichs definiert, der auf das Aufnahmemedium projiziert wird. Im Falle einer langen Brennweite ist der Bildwinkel klein, sodass nur ein kleiner Ausschnitt des Gegenstandsraums auf das Aufnahmemedium projiziert wird (siehe Abbildung 2.2).

Die Größe des Bildwinkels hängt von der Länge der Brennweite und der Sensorgröße ab.

- Ändert man die Brennweite und das Aufnahmeformat um den gleichen Faktor, bleibt der Bildwinkel gleich.
- Daraus folgt, dass verschiedene Aufnahmeformate unterschiedliche Brennweiten benötigen, um einen bestimmten Bildwinkel abzudecken.
- Das heißt: Je kleiner der Sensor wird, desto kürzer muss die Brennweite werden, wenn der Bildwinkel und damit der abgebildete Ausschnitt der Umwelt gleich bleiben sollen.

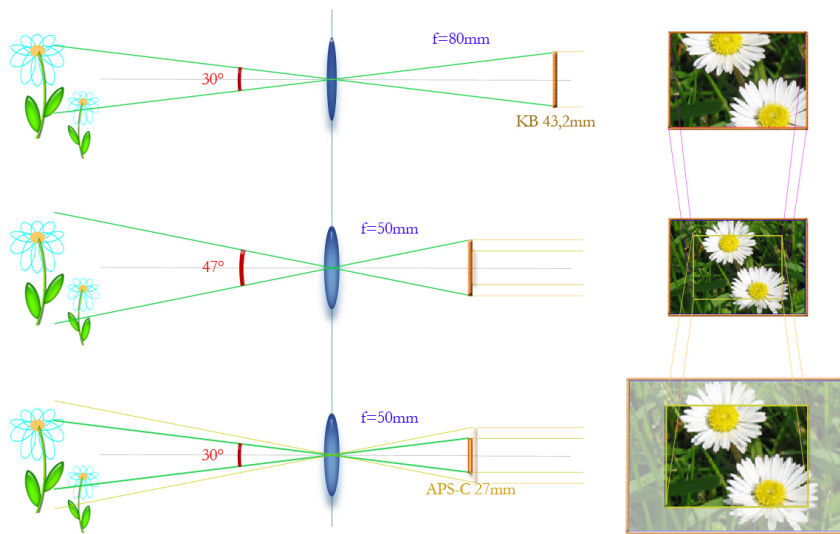


Abbildung 2.2.: Abhängigkeit von Bildwinkel und Brennweite [Wik11a]

Linsengleichung

Mit der Linsengleichung, auch bekannt als Abbildungsgleichung, berechnet man die optische Abbildung einer idealen Linse:

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{g}$$

wobei f die Brennweite, b die Bildweite und g die Gegenstandsweite ist. Die Größen kann man entsprechend Abbildung 2.3 ermitteln.

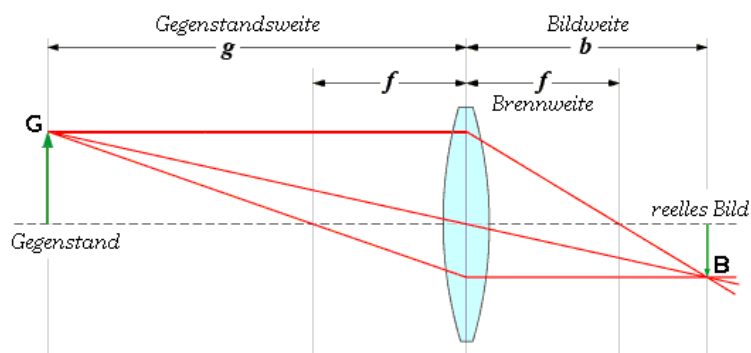


Abbildung 2.3.: Optische Abbildung einer idealen Linse [Wik11f]

Belichtungszeit

Die Belichtungszeit ist die Zeit, während der das Aufnahmemedium einer Kamera zur Aufzeichnung eines Bildes belichtet wird. Von der Länge dieser Zeit hängt die Helligkeit des Bildes ab. Bei einer kurzen Belichtungszeit wird das Bild dunkel. Während einer langen Belichtungszeit sammelt das Aufnahmemedium viel Licht, sodass das Bild sehr hell dargestellt wird. Falls jedoch die Dauer des Lichteinfalls zu lang wird, erhöht sich die Gefahr der Bewegungsunschärfe.

2.1.2. Kameraparameter

Ein Kameramodell wird durch mehrere Parameter beschrieben, die physikalische Eigenschaften und geometrische Lage der Kamera charakterisieren. Diese Parameter werden in extrinsische und intrinsische Parameter unterteilt.

Extrinsische Parameter

Die extrinsischen Parameter definieren die Position und die Orientierung der Kamera im Raum, also der Ort, wo die Kamera platziert ist. Das heißt, dass es in diesem Raum ein Koordinatensystem gibt, in dem die Kamera und die aufgenommene Objekte positioniert werden können. Dieses Koordinatensystem heißt Weltkoordinatensystem.

Unabhängig davon verfügt die Kamera auch über ein eigenes Koordinatensystem, dessen Ursprung das optische Zentrum der Kamera ist. Eine Achse dieses Systems ist die optische Achse der Kamera.

Die extrinsischen Parameter ermöglichen es, die Beziehung zwischen dem Kamera- und dem Weltkoordinatensystem zu beschreiben.

Intrinsische Parameter

Die intrinsischen Parameter beschreiben die physikalischen Eigenschaften der Kamera. Während sich die extrinsischen Parameter bei einer Bewegung der Kamera ändern, ändern sich die intrinsischen Parameter der Kamera nicht.

Folgende Parameter gehören zu den intrinsischen Parametern einer Kamera:

- **Brennweite:** siehe Abschnitt 2.1.1
- **Optisches Zentrum:** Definiert den Ursprung des Bildkoordinatensystems, durch den Schnittpunkt der optischen Achse mit dem Sensor.
- **Pixelgröße:** Die Größe eines Pixels hängt von dessen Länge und Breite auf dem Aufnahmemedium ab. Die Pixel auf dem Aufnahmemedium sind nicht notwendigerweise Quadrate, sondern Rechtecke. Aus der Pixelgröße werden die Pixelskalierungsfaktoren für die horizontale und vertikale Richtung berechnet.
- **Radiale Verzerrungskoeffizienten:** Die radiale Verzerrung der Kameralinse wird durch die Koeffizienten k_1 und k_2 beschrieben.
- **Scherung (eng. skew):** Das ist der Winkel zwischen dem Aufnahmemedium und der optischen Achse. Manche Kamerakoordinatensysteme sind aufgrund eines Herstellungsfehlers schief, sodass dieser Winkel sich von 90° unterscheidet.

Die intrinsischen Parameter beschreiben die interne Geometrie einer Kamera. Durch die intrinsischen Parameter ist definiert, wie Punkte aus dem 3D Kamerakoordinatensystem in das 2D Bildkoordinatensystem mittels perspektivischer Transformation abgebildet werden.

Die Kenntnis der intrinsischen Parameter ist die Voraussetzung für Entfernung der durch eine Linse verursachten Verzerrungen.

2.1.3. Lochkameramodell

In den vorherigen Abschnitten wurden verschiedene Parameter eines Kameramodells erklärt. Mithilfe dieser Parameter können in einem Kameramodell folgende Transformationen eines Objektpunktes durchgeführt werden:

- vom Weltkoordinatensystem in das Kamerakoordinatensystem (mithilfe extrinsischer Parameter)
- vom Kamerakoordinatensystem in das Koordinatensystem des Aufnahmemediums (mithilfe intrinsischer Parameter)
- vom Koordinatensystem des Aufnahmemediums in das Bildkoordinatensystem (mithilfe intrinsischer Parameter)

2. Grundlagen

Beim Lochkameramodell liegt ein Weltpunkt und der entsprechende Bildpunkt auf einer Geraden, die durch den Ursprung des Kamerakoordinatensystems verläuft. Deshalb können Weltpunkte mit Zentralprojektion auf Bildpunkte abgebildet werden. Die Bildebene hat vom Ursprung des Kamerakoordinatensystems den Abstand der Brennweite F .

Diese Eigenschaften ermöglichen es mithilfe des Lochkameramodells und dem Strahlensatz des Projektion des Weltpunktes P_W zum Bildpunkt P_B wie folgt zu bestimmen:

$$P_B = A[Rt]P_W$$

wobei R und t die extrinsischen Parameter sind und A die sogenannte intrinsische Kameramatrix ist.

Der extrinsische Parameter R definieren die Kamerarichtung durch eine Rotationsmatrix R während der extrinsische Parameter t die Position der Kamera durch eine Translationsmatrix t festlegt:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

R beschreibt eine beliebige Rotation, die sich aus drei elementaren Rotationen zusammensetzt (siehe Abschnitt 2.2.2).

Die Kameramatrix A beschreibt die Transformation vom Kamerakoordinatensystem direkt in das Bildkoordinatensystem. Die Kameramatrix ist wie folgt definiert:

$$\begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Die Matrix enthält fünf intrinsische Parameter der Kamera. Die Größen (c_x, c_y) definieren eine Translation des Bildursprungs relativ zum Schnittpunkt der optischen Achse mit dem Sensor. Dadurch wird der Ursprung in die Bildmitte verschoben.

(f_x, f_y) sind die Brennweiten der horizontalen und vertikalen Bildachsen. Fakt ist, dass es zwei unterschiedliche Brennweiten gibt. Der Grund hierfür ist, dass Pixel auf dem Aufnahmemedium kein Quadrat, sondern Rechtecke sind. Die beide Brennweiten entsprechen nicht der physikalischen Brennweite F der Kamera. Sie sind Produkte aus der physikalischen Brennweite und einem Faktor, der den durch die Rechtecksform entstandenen Unterschied

ausgleicht. Diese Faktoren werden auch Pixelskalierungsfaktoren genannt. Eine weitere Aufgabe dieser Faktoren ist es, die Längen-Einheit in die Pixel-Einheit umzuwandeln, damit alle Berechnungen in der gleichen Einheiten durchgeführt werden können.

$$\begin{cases} f_x = m_x \cdot F \\ f_y = m_y \cdot F \end{cases}$$

wobei F die physikalische Brennweite und (m_x, m_y) Skalierungsfaktoren mit der Einheit $[Pixel/mm]$ sind.

Die vier intrinsischen Parameter (c_x, c_y) und (f_x, f_y) sind in Pixelgrößen angegeben. Der fünfte intrinsische Parameter s ist der Faktor der Scherung. Wenn der Winkel zwischen der optischen Achse und dem Aufnahmemedium weit von 90° deutlich abweicht, ist dieser Faktor s entsprechend groß. Wenn der Winkel 90° beträgt, ist $s = 0$.

2.1.4. Optische Verzerrung und Kamerakalibrierung

Optische Verzerrung

Unter optischer Verzerrung versteht man geometrische Abbildungsfehler optischer Systeme. Liegt ein solcher Fehler vor, ist das projiziertes Bild an den Bildrändern stark verzogen und die Formen der Objekte auf dem Bild sind deformiert. Beispielsweise werden gerade Linien infolge einer optischen Verzerrung als gekrümmte Linien abgebildet. Optische Verzerrungen treten bei Lochkameras nicht auf, sondern lediglich bei Kameras mit Linsen.

Man unterscheidet zwischen radialen und tangentialen Verzerrungen. Durch eine radiale Verzerrung wird der Abstand zwischen dem Bildpunkt und der Bildmitte verändert. Je größer die radiale Verzerrung ist, desto größer ist der Abstand zwischen dem Bildpunkt und der Bildmitte. Bei der tangentialen Verzerrung verläuft der Bildpunkt jedoch senkrecht zum radial verzerrten Punkt. Das heißt, die tangentielle Verzerrung wächst wie die radiale Verzerrung mit dem Abstand zwischen dem Bildpunkt und der Bildmitte, aber nur in die tangentielle Richtung. Die tangentielle Verzerrung ist gering im Vergleich zu der radialen Verzerrung, sodass man sie vernachlässigen kann.

Radiale Verzerrungen werden in tonnenförmige und kissenförmige Verzerrungen unterteilt (siehe Abbildung 2.4). Bei einer tonnenförmigen Verzerrung sind die Linien an den Bildecken nach innen gebogen. Bei einer kissenförmigen Verzerrung sind die Linien an den Bildecken

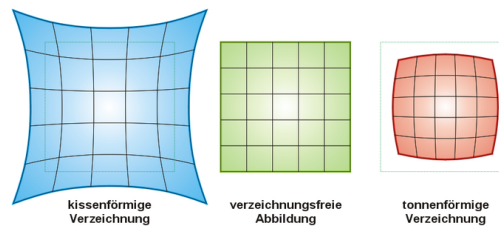


Abbildung 2.4.: radiale Linsenverzerrung [Wik11i]

nach außen gebogen. Um Verzerrungen aus einem Bild zu entfernen, benötigt man ein Modell, das radiale und tangentielle Verzerrungen charakterisiert. Das Lochkammermodell ist ein lineares Modell. Durch die Kameramatrix sind zwar die Eigenschaften der Kamera und der Linse beschrieben. Aber das Modell beschreibt eine Verzerrung linear und deshalb werden Geraden auf Geraden projiziert. Radiale und tangentielle Verzerrungen sind jedoch nicht-lineare Verzerrungen.

Die radiale Verzerrung kann mit einer Linsenfunktion $L(r)$ modelliert werden, die für einen Abstand r zwischen dem Bildpunkt $p_d(x, y)$ und der Bildmitte $p_0(x_0, y_0)$ einen Faktor für Abweichung der Verzerrung angibt: $p_d - p_0 = L(r) \cdot (p - p_0)$, wobei $p(x, y)$ der ideale Bildpunkt für den verzerrten Punkt $p_d(x_d, y_d)$ ist. Für den Abstand r gilt natürlich $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$. Die Linsenfunktion ist nur für die Bildmitte bekannt, nämlich $L(0) = 1$. Für die anderen Abstände ist die Linsenfunktion durch eine Taylorreihenentwicklung um $r = 0$ approximiert. Das Modell für die radiale Verzerrung sieht dann wie folgt aus:

$$p_d = p + (p - p_0)(k_1 \cdot r + k_2 \cdot r^2)$$

wobei k_1 und k_2 Verzerrungskoeffizienten¹ sind.

Kamerakalibrierung

Durch die Kamerakalibrierung können die extrinsischen und intrinsischen Kameraparameter einer Kamera ermittelt werden. Für die Kalibrierung einer Kamera wird ein Kalibrieremuster,

¹Weitere Verzerrungskoeffizienten können mithilfe der Taylorreihenentwicklung ermittelt werden. Aber die erste zwei Koeffizienten k_1 und k_2 beeinflussen auf die meisten Verzerrungen.

dessen Geometrie bekannt ist, verwendet. Es gibt verschiedene Kalibriermuster, die entweder aus einer 2D Ebene oder aus mehreren 2D Ebenen (3D) bestehen (siehe Abbildung 2.5). Die Kalibriermuster enthalten Markierungen, die bei der Aufnahme im Bild erkennbare Merkmale erzeugen, um geometrische Berechnungen zu ermöglichen. Ein Verfahren zur

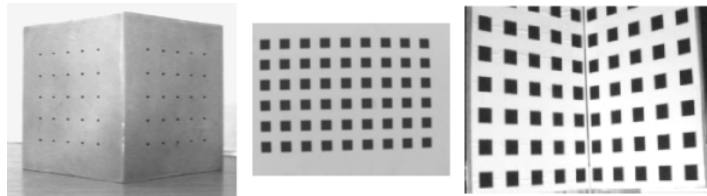


Abbildung 2.5.: Beispiele für 3D, 2D Kalibriermuster

Kamerakalibrierung ermittelt aus jedem durch die kalibrierende Kamera aufgenommenen Bild vom Kalibriermuster die Positionen der Markierungen des Kalibrierusters. Aus diesen ermittelten Positionen werden dann Gleichungssysteme erstellt, die dem verwendenden Kameramodell entsprechen. Durch die Lösung dieser Gleichungssysteme werden die Parameter der Kamera geschätzt und bei nächstem Bild werden sie verbessert.

Es gibt zahlreiche Verfahren zur Schätzung der Parameter einer Kamera durch Kalibrierung. Sie unterscheiden sich voneinander dadurch,

- ob sie einen linearen oder nicht-linearen Algorithmus zur Lösung der erstellten Gleichungssysteme verwenden,
- welche Kalibriermuster sie verwenden bzw. ob sie 3D oder 2D Kalibriermuster verwenden oder
- ob es sich um eine *Selbstkalibrierung* handelt.

Verfahren, die 3D Kalibriermuster verwenden, erstellen die Gleichungssysteme mithilfe einer Projektionsmatrix des Kameramodells. Verfahren, die 2D Kalibriermuster verwenden, erstellen die Gleichungssysteme mithilfe einer Homographiematrix (siehe Abschnitt 2.2.3).

Unter *Selbstkalibrierung* versteht man Verfahren, die keine Kalibriermuster verwenden. Bei diesen Verfahren werden die Kameraparameter ohne 3D Geometrie berechnet. Selbstkalibrierende Verfahren helfen dem Verständnis der Verfahren zur Kamerakalibrierung, aber sie sind wegen ihrer ungenauen Schätzungen nicht zu empfehlen.

Um die geschätzten Kameraparameter zu bewerten, können geometrische Fehler für die erkannten Positionen der Markierungen anhand des Kalibrierungsmusters berechnet werden.

Diesbezüglich sei angemerkt, dass man durch ein Kalibrierungsverfahren die physikalische Brennweite F nicht schätzen kann. Man kann die Kameramatrix bzw. die Brennweite (f_x, f_y) schätzen. Wie in Abschnitt 2.1.3 beschrieben, wird die Brennweite (f_x, f_y) in Pixeln angegeben und entspricht nicht der physikalischen Brennweite, die in Längeneinheit Millimetern angegeben wird.

2.2. Transformationen

2.2.1. Transformation zwischen kartesischen und sphärischen Koordinatensystemen

Ein Punkt im Raum kann durch verschiedene Koordinatensysteme dargestellt werden. Das kartesische Koordinatensystem ist ein orthogonales Koordinatensystem und besteht aus drei Koordinatenachsen (X, Y, Z) . Ein Punkt P kann in einem kartesischen Koordinatensystem entsprechend durch drei Koordinaten (x, y, z) eindeutig positioniert werden, wobei x , y , und z die Projektionen des Punktes auf die entsprechenden Achsen X , Y und Z sind. Das kartesische Koordinatensystem eignet sich nicht so gut, wenn man Drehbewegungen eines Punktes beschreiben möchte.

In einem sphärischen Koordinatensystem wird ein Punkt P dagegen durch zwei Winkel und eine Länge (θ, ϕ, R) dargestellt (siehe Abbildung 2.6). R ist der Abstand zwischen dem Punkt P und dem Ursprung des Koordinatensystems. θ gibt den Neigewinkel von der Z -Achse an und ϕ ist der Drehwinkel um die Z -Achse. Die Wertebereiche von (θ, ϕ, R) wurden in Abbildung 2.6 angegeben. Damit man die beiden Koordinatensysteme im Rahmen einer Aufgabe verwenden kann, müssen sphärische und kartesische Koordinaten umgerechnet werden. Möglich ist dies mit folgenden Formeln:

$$\begin{cases} x = R \sin \theta \cos \phi \\ y = R \sin \theta \sin \phi \\ z = R \cos \theta \end{cases} \quad \begin{cases} R = \sqrt{x^2 + y^2 + z^2} \\ \theta = \arccos\left(\frac{z}{R}\right) \\ \phi = \arccos\left(\frac{x}{R \cdot \sin \theta}\right) \end{cases}$$

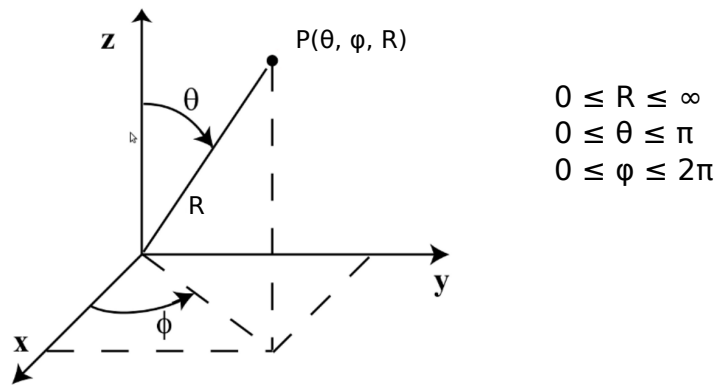


Abbildung 2.6.: Sphärisches Koordinatensystem

Bei einer Transformation vom kartesischen ins sphärische Koordinatensystem müssen die Sonderfälle $\theta = 0$ oder $R = 0$ beachtet werden, weil in manchen Fällen der Nenner Null werden kann.

2.2.2. Starre 3D Transformationen

Unter einer starren Transformation versteht man die Bewegung eines Starrkörpers. Starre Körper sind 3D Objekte, die nicht verformbar sind. Das heißt, die Form eines Starrkörpers verändert sich durch seine Bewegung nicht. Damit bleibt Abstand zwischen zwei beliebigen Punkten im Körper bei einer Bewegung erhalten. Die Bewegung eines starren Körpers setzt sich aus Translation und Rotation zusammen.

Rotation

Eine starre Rotation um eine beliebige Gerade, die durch den Ursprung verläuft, kann aus drei elementaren Drehungen um drei kartesische Koordinatenachsen zusammengesetzt werden:

- Rotation um die X-Achse mit dem Winkel α
- Rotation um die Y-Achse mit dem Winkel β
- Rotation um die Z-Achse mit dem Winkel γ

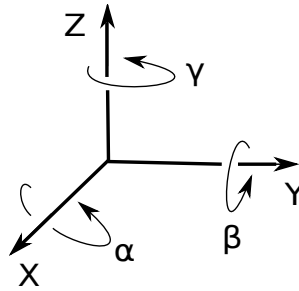


Abbildung 2.7.: Rechtshändiges Koordinatensystem (Rechtssystem genannt). Im Rechtssystem wird gegen den Uhrzeigersinn rotiert. Eine elementare Rotation ist eine Rotation um eine einzelne Achse X, Y oder Z.

Sowohl elementare als auch zusammengesetzte Rotation können mit 3×3 Matrizen dargestellt werden: $P' = R_Z(\gamma) \cdot R_Y(\beta) \cdot R_X(\alpha) \cdot P$ bzw.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

In der obigen Gleichung wird ein Punkt $P(x, y, z)$ um die Achsen X, Y und Z (erst um X, dann um die Y Achse und anschließend um die Z Achse) mit (α, β, γ) Winkeln rotiert. Der Punkt hat nach der Rotation die Koordinaten $P(x', y', z')$. Man kann aber auch sagen, dass das Koordinatensystem rotiert wurde und die Position des Punktes im neuen Koordinatensystem (x', y', z') ist.

Im Rechtssystem (siehe Abbildung 2.7) wird der Punkt entgegen dem Uhrzeigersinn um die Drehachse rotiert, wenn der Drehwinkel positiv ist. In diesem Fall spricht man von einer aktiven Drehung. Bei einer passiven Drehung wird der Punkt bzw. das Koordinatensystem mit einem negativen Winkel rotiert.

Wenn man die Multiplikation der Matrizen $R_X(\alpha), R_Y(\beta)$ und $R_Z(\gamma)$ berechnet, erhält man die zusammengesetzte 3×3 Rotationsmatrix R , die den drei elementaren Rotationen entspricht: $P' = R \cdot P$ bzw.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\gamma \cdot \cos\beta & (\cos\gamma \cdot \sin\beta \cdot \sin\alpha - \sin\gamma \cdot \cos\alpha) & (\sin\gamma \cdot \sin\alpha + \cos\gamma \cdot \sin\beta \cdot \cos\alpha) \\ \sin\gamma \cdot \cos\beta & (\cos\gamma \cdot \cos\alpha + \sin\gamma \cdot \sin\beta \cdot \sin\alpha) & (\sin\gamma \cdot \sin\beta \cdot \cos\alpha - \cos\gamma \cdot \sin\alpha) \\ -\sin\beta & \cos\beta \cdot \sin\alpha & \cos\beta \cdot \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Eine Eigenschaft einer Rotationsmatrix ist es, dass sie orthogonal und ihre Determinante 1 ist. Das heißt, es gilt für eine Rotationsmatrix R :

$$R \cdot R^T = E \rightarrow R^T = R^{-1}$$

wobei R^T die transponierte Matrix von R ist und E die Einheitsmatrix ist. Daraus folgt, dass die transponierte Matrix gleichzeitig auch die inverse Matrix von R ist.

Translation

Die andere Bewegung des starren Körpers ist die Translation. Im Rahmen einer Translation wird der starre Körper bzw. sein Koordinatensystem verschoben. Um eine beliebige Translation eines Punktes $P(x, y, z)$ zu berechnen, benötigt man 3 Verschiebungen (T_X, T_Y, T_Z) des Punktes in X, Y und Z Richtungen. Diese Berechnung lässt sich auch durch Matrizen beschreiben: $P' = P + T$ bzw.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

Bei mehreren Transformationen muss die Reihenfolge der Berechnungen beachten werden, da die Matrixmultiplikation nicht kommutativ ist und es so bei unterschiedlichen Reihenfolgen zu verschiedenen Ergebnissen kommen kann.

2.2.3. Projektive Transformation

Eine projektive Transformation wird in der Bildverarbeitung meist zur Zusammensetzung mehrerer Bilder, die eine aus unterschiedlichen Positionen und Orientierungen der Kamera aufgenommene Szene darstellen, verwendet. Abbildung 2.8 zeigt zwei Bilder einer Kamera als Beispiel. Die Position der Kamera wurde hier nicht verändert, sondern lediglich ihre Sichtrichtung. Wenn man für den Fall (Kamera macht keine Verschiebungen, nur Drehbewegungen) das Lochkameramodell verwendet, kann die projektive Transformation durch

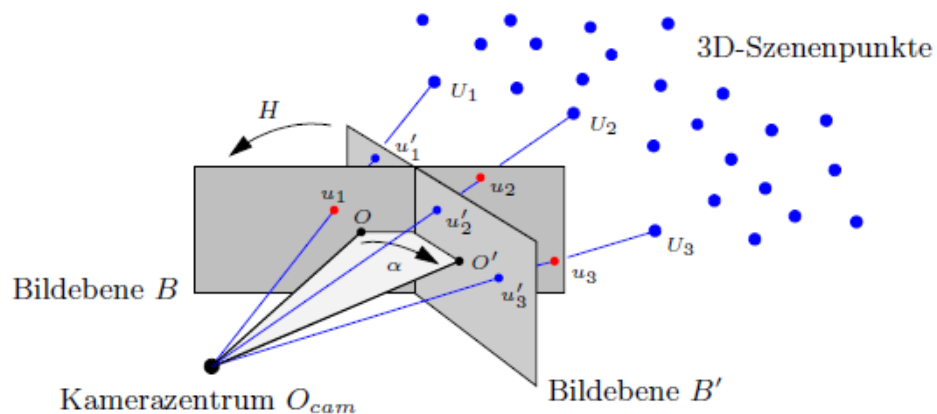


Abbildung 2.8.: Drehbewegung der Kamera: Eine projektive Transformation ermöglicht die Zusammensetzung beider Bilder. Quelle: [Scho5]

eine 3×3 Matrix H beschrieben werden. Die Matrix H wird auch Homographie-Matrix genannt.

$$H = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix}$$

Die Matrix H transformiert Punkte auf Punkte und Linien auf Linien. In dieser Transformation ist die affine Transformation eingeschlossen. Alle affinen Transformationen sind projektive Transformationen. Umgekehrt gilt dies allerdings nicht.

Die projektive Transformation in Ebene hat 8 Freiheitsgrade und kann deshalb durch 8 Parameter bestimmt werden. Das heißt, man benötigt für die Bestimmung von zwei Ebenen 8 Gleichungen bzw. man muss in jeder Ebene 4 Punkte, die zueinander zugeordnet sind, kennen. Verschiedene Algorithmen zur Bestimmung der Homographie können mit mehreren (mehr als 8) Punkten arbeiten, weil in der Praxis 4 nicht ideale Punktpaare durch Messungen gefunden werden. Oft wird das beste Ergebnis durch Minimierung der quadratischen Fehler berechnet.

Wenn man die Homographie von zwei Ebenen berechnet hat, kann man einen Punkt $p(x, y)$ einer Ebene zum Punkt $p'(x', y')$ in der anderen Ebene wie folgt transformieren:

$$x' = \frac{h_{00} \cdot x + h_{01} \cdot y + h_{02}}{h_{20} \cdot x + h_{21} \cdot y + h_{22}} \qquad y' = \frac{h_{10} \cdot x + h_{11} \cdot y + h_{12}}{h_{20} \cdot x + h_{21} \cdot y + h_{22}}$$

2.2.4. Hough Transformation

Durch die sogenannte Hough Transformation wird ein geometrisches Objekt (z.B. eine Gerade, ein Kreis, eine Ellipse usw.) zu einem Punkt in einem Dualraum transformiert, der durch die Parameter des Objekts aufgespannt wird. Diese Transformation ermöglicht es, geometrische Objekte in einem Bild zu erkennen. Die Hough Transformation wird daher oft in der Bildverarbeitung zur Detektion kollinearere Punkte, zur Konturbeschreibung und zur Bildsegmentierung verwendet.

Als Beispiel betrachten wir eine Gerade, die durch die Gleichung $y = mx + b$ definiert ist. Man kann z.B. als Parameter die Steigung m und der Y -Achsenabschnitt b für die Transformation auswählen. Dann fallen alle zu einander kollinearen Geraden im Raum $m \times b$ auf einen Punkt. In der Praxis inkrementieren alle gefundenen Kantenpunkte im Bildraum den Dualraum (auch Akkumulator genannt) an den Punkten der Geraden, zu denen der untersuchte Kantenpunkt gehören kann. Durch die Suche nach Extrempunkte im Dualraum kann man im Bildraum auftretende Geraden detektieren und die gesuchten Parameter bestimmen.

Bei Geraden sind allerdings die Parameter m und b für die Hough Transformation nicht geeignet, weil die Steigung m einer zur Y -Achse parallele Gerade unendlich ist und solche Geraden nicht transformiert werden können. Andere zwei Parameter, die eine Gerade charakterisiert, sind der Abstand d und der Winkel α , wobei d der Abstand zwischen dem Ursprung und der Gerade ist und α der Winkel von der Normale der Gerade zur X -Achse ist. Die beide Parameter sind besser für Berechnungen geeignet und daher werden sie meist bei Hough Transformation für Geraden verwendet. Mit Hilfe der Hesseschen Normalform sieht die Gleichung, die eine Gerade durch den Abstand d und den Winkel α beschreibt, wie folgt aus: $x \cdot \cos(\alpha) + y \cdot \sin(\theta) = d$

In Abbildung 2.9 ist die Darstellung der Hough-Transformation von zwei Geraden zu sehen.

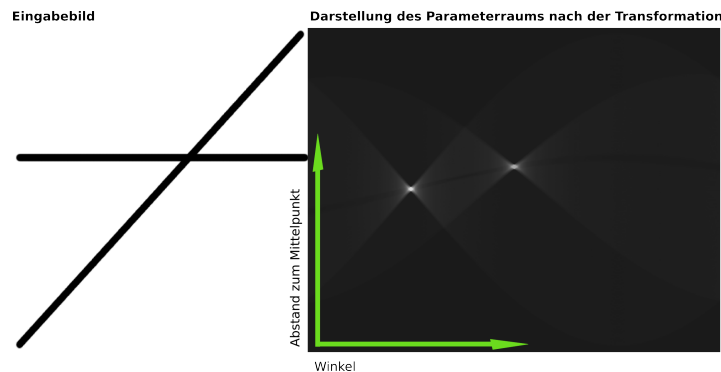


Abbildung 2.9.: Hough Transformation von zwei Linien in einem Raum aus Winkel und Abstand zur Geraden. Quelle: [Wik11d]

2.3. Kartenprojektion

Die Kartenprojektion (auch Kartennetzentwurf genannt) ist eine Methode, die durch eine Abbildung die Oberfläche der Erde auf einer ebenen Fläche darstellt. In diesem Abschnitt werden verschiedene Kartenprojektionen und deren Eigenschaften sowie Klassifikationen erklärt und grafisch veranschaulicht.

Bei der Kartenprojektion entstehen immer Verzerrungen, sodass nach der Abbildung der Erdoberfläche z.B. Strecken, Winkel und/oder Flächen nicht mit den echten Größen übereinstimmen. Je größer die projizierte Fläche ist, desto stärker wirken sich die Verzerrungen aus. Daher wurden zahlreiche Kartenprojektionen entwickelt, die bestimmte Eigenschaften ohne Verzerrung auf einer Ebene darstellen zu können.

In der Regel werden Kugel oder Ellipsoid dabei als die Form der Erde angenommen. Man kann aber auch andere 3D-Objekte (z.B. planetarische Körper) als abzubildende Objekt verwenden. Andererseits können Kegel, Zylinder, Kubus oder ein anderer Körper, der ohne Verzerrung als ebenen Fläche aufgefaltet werden kann, für die Kartenprojektion als Hilfsfläche verwendet werden. Solche Kartenprojektionen bilden eine Kugel zu erst auf eine Hilfsfläche. Anschließend kann die Hilfsfläche verzerrungsfrei zur Ebene aufgefaltet werden.

Man kann eine Kartenprojektion grob nach ihren Eigenschaften und ihrer Projektionsfläche klassifizieren.

2.3.1. Eigenschaften einer Karteprojektion

Die Eigenschaften einer Kartenprojektion werden durch ihre Verzerrungen definiert. Zu den wichtigsten Eigenschaften gehören:

- **Längentreue:** Bei der Abbildung bleibt der Abstand zwischen zwei Punkten bis auf einen Maßstabsfaktor m erhalten. Es gibt jedoch keine Kartenprojektion, die absolut längentreu ist, d.h. $m = 1$ ist. Einige Kartenprojektionen bilden nur bestimmte Linien längentreu ab. Bei manchen Kartenprojektionen kann man bestimmen, welche Linien längentreu abgebildet werden sollen.
- **Flächentreue:** Die Größe aller Flächenelemente bleibt bei der Abbildung erhalten.
- **Winkeltreue:** Bei der Abbildung bleibt der Winkel zwischen zwei Linien erhalten. Der Winkel zwischen zwei Linien in einer Sphäre wird als Winkel zwischen den Tangenten der Sphäre am Schnittpunkt der Linien mit der Sphäre definiert.

Wenn eine dieser Eigenschaften bei einer Kartenprojektion fehlt, treten entsprechende Verzerrungen auf. Jedoch gibt es keine Kartenprojektion, die all diese Eigenschaften besitzt.

2.3.2. Projektionsflächen

Wie schon zu Beginn dieses Abschnittes erwähnt, gibt es verschiedene Hilfsflächen, über die eine Kartenprojektion eine Kugeloberfläche abbilden kann. Entsprechend dieser Hilfsflächen werden die Kartenprojektionen im folgenden klassifiziert. Dabei können die Hilfsflächen unterschiedliche Größen und Positionen haben. In diesem Fall muss man je nach Zweck entscheiden, welche Region von besonderer Bedeutung ist.

Azimutale Abbildungen Kartenprojektionen, die die Kugeloberfläche direkt auf einer ebenen Fläche abbilden, heißen azimutale Abbildungen. Bei diesen Abbildungen berührt die Kugeloberfläche an einem Punkt die Ebene der Abbildung. Verschiedene Projektionsarten der azimutalen Abbildung sind in Abbildung 2.10 dargestellt.

Bei der gnomonischen Projektion liegt das Projektionszentrum im Mittelpunkt, während es bei der stereografischen Projektion auf der dem Berührungspunkt entgegengesetzten Seite auf der Kugeloberfläche liegt. Bei der orthografischen Projektion sitzt das Projektionszentrum im Unendlichen, sodass die Kugeloberfläche mittels parallelen Strahlen projiziert wird.

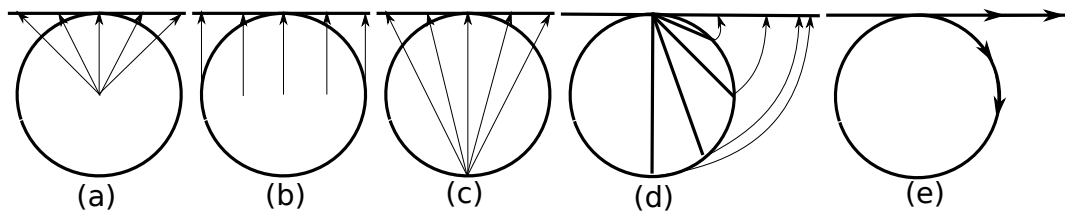


Abbildung 2.10.: Projektionen: (a) gnomonische, (b) orthografische, (c) stereografische, (d) flächentreue und (e) mittabstandstreue Azimutalprojektion

Mittabstandstreue und flächentreue Azimutalprojektionen sind mathematisch definiert und sind keine physikalischen Projektionen. Mit diesen beiden Projektionen kann man die gesamte Kugeloberfläche abbilden, wobei die Verzerrung zum Rand hin jedoch stark zunimmt. Bei der mittabstandstreuen Azimutalprojektion werden Entfernung und Richtung der projizierten Punkte vom Berührungspunkt nicht verzerrt. Abbildung 2.11 zeigt Beispiele für die fünf Projektionsarten.

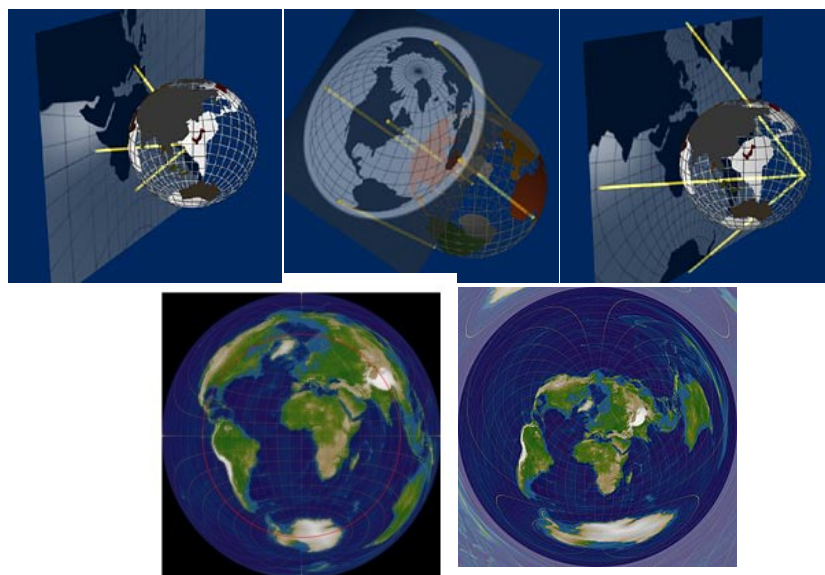


Abbildung 2.11.: Beispiele für die 5 Projektionsarten in gleicher Reihenfolge wie bei vorheriger Abbildung. Quelle: [Kö] und [Wik11g]

Zylindrische Abbildungen Bei diesen Abbildungen wird ein Zylinder um die Kugel herum konstruiert und die Kugelfläche auf den Zylinder projiziert. Es gibt unterschiedliche zylindrische Projektionen (siehe Abbildung 2.12).

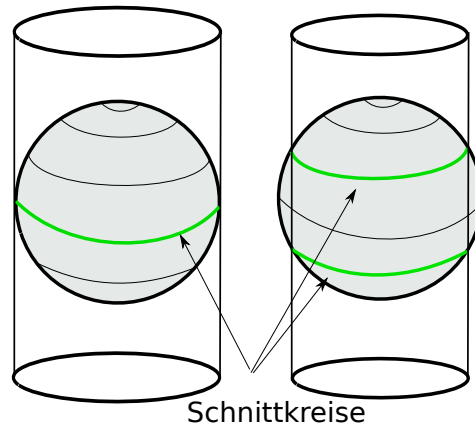


Abbildung 2.12.: Zylindrische Abbildung

In Abbildung 2.13 sind die zylindrischen Mercator- und Rektangularprojektionen der Erdoberfläche dargestellt. Bei beiden Projektionen sind die Linien der Breiten- und Längengrade

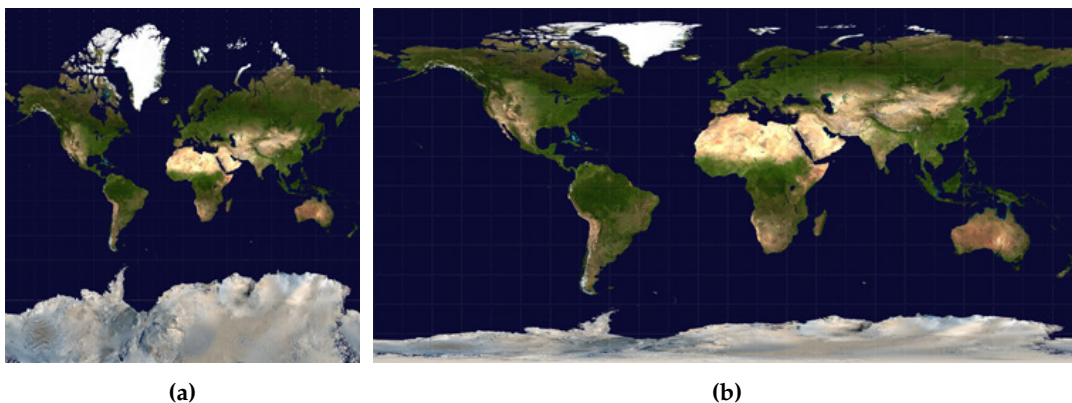


Abbildung 2.13.: (a) Mercatorprojektion, (b) Rektangularprojektion Quelle: [Wik11]

parallele gerade Linien. Der Längengrad ist eine der beiden geographischen Koordinaten und beschreibt die Position eines Punktes auf der Erde in Grad, z.B. wie weit östlich oder westlicher vom Nullmeridian entfernt ist. Die andere geographische Koordinate ist der

2. Grundlagen

Breitengrad. Er gibt die nördliche oder südliche Entfernung eines Punktes vom Äquator an.

Die Mercatorprojektion ist eine winkeltreue Abbildung. Die Breitengrade bekommen aber dadurch ungleiche Abstände, sodass die Flächenelemente vom Äquator in Richtung der Zylinderachse immer stärker verzerrt werden. Die Längengrade haben dagegen regelmäßige Abstände. Wegen der starken Verzerrung wird ein Kugel mit dieser Projektion nicht vollständig dargestellt. Die Darstellung endet bei ca. 80 Grad Breite. Die Polgebiete werden deshalb meist mit einer anderen Kartenprojektion (z.B. mit einer stereografischen Projektion) dargestellt. Bei der Rektangularprojektion wird die Kugeloberfläche abgebildet, indem die Längen- und Breitengrade direkt als kartesische Koordinaten verwendet werden. Dadurch haben die Pole und der Äquator gleiche Länge (normalerweise haben die Pole eine Länge von Null). Die Flächenverzerrung zum Pol hin wird auch hier stärker, ebenso wie bei der Mercatorprojektion. Der Unterschied ist hier jedoch, dass starke Verzerrungen an den Polen entlang der Ost-West Achse existieren. In Nord-Süd-Richtung ist die Rektangularprojektion jedoch längentreu.

Es gibt auch die Möglichkeit eine Rektangularprojektion mithilfe eines kleinen Zylinders abzubilden. Das heißt, die Erdkugel und der Zylinder schneiden sich und bilden so zwei Kreise. Falls sie einen Kreis am Äquator bilden, hat die Rektangularprojektion eine projizierte Fläche mit gleicher Breite wie die Mercatorprojektion.

Kegelprojektion Bei der Kegelprojektion wird die Kugeloberfläche auf einem Kegel abgebildet. Die Kegelachse verläuft durch den Erdmittelpunkt. Der Abstand zwischen Kugel und Kegel kann bei Kegelprojektionen unterschiedlich sein, sodass sie sich entweder in einem Kreis berühren oder sich in zwei Kreisen schneiden (siehe Abbildung 2.14).

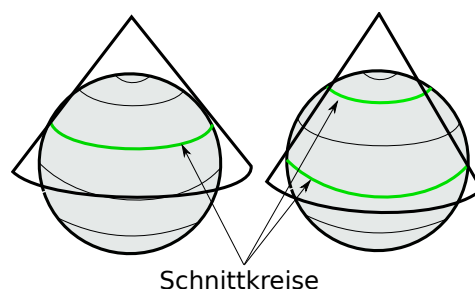


Abbildung 2.14.: Kegelprojektion

Abbildung 2.16 zeigt einige Kegelprojektionen.



Abbildung 2.15.: Einige Kegelprojektionen: (a) Lambertsche, (b) Längentreue, (c) Albers.
Quelle: [Wik11e]

2.3.3. Einige Kartenprojektionen

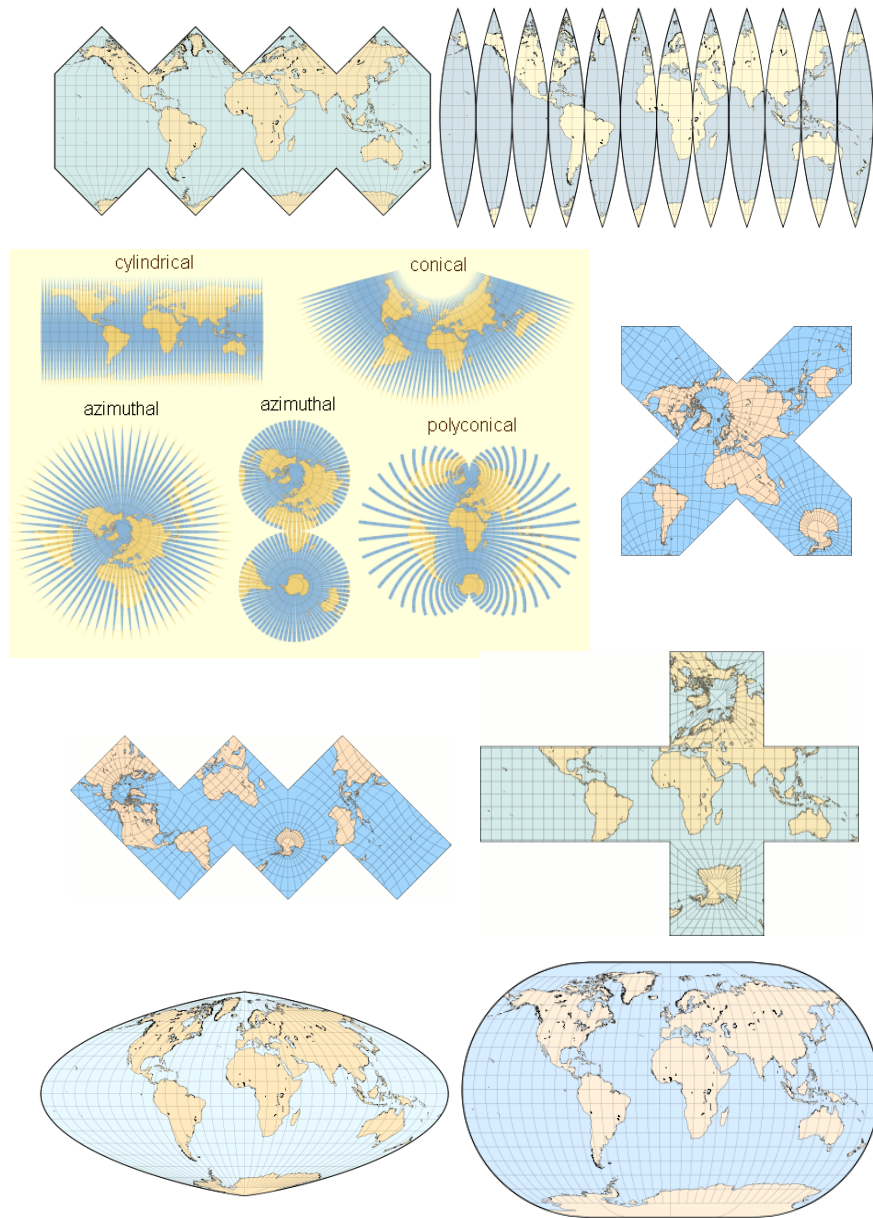


Abbildung 2.16.: Einige interessante Kartenprojektionen. Quelle: [Furo9]

3. Verwandte Arbeiten

3.1. Hintergrundsubtraktion

In der Literatur wird über viele Verfahren zur Hintergrundsubtraktion berichtet. Eine Zusammenfassung wichtiger Verfahren wurde von Piccardi [Pico4] veröffentlicht. Die meisten Verfahren wurden für stationäre Kameras entwickelt. Ein sehr bekanntes Verfahren für diese Kameras ist das von Stauffer und Grimson [SG99], das aus den vorangegangenen Einzelbildern Hintergrundwerte durch mehrere gewichtete Gauß-Verteilungen pro Pixel als Hintergrundmodell darstellt. Basierend auf diesem Verfahren wurden weitere Verfahren veröffentlicht, die das Original-Verfahren auf unterschiedliche Weisen verbessern. Ein Überblick über die weiter entwickelten Verfahren findet man in [BEBVo8].

Neben den Verfahren für stationäre Kameras wurden auch einige Verfahren für PTZ-Kameras entwickelt. Sie unterscheiden sich hauptsächlich durch die Art, wie sie die projektive Transformation (Homographie-Matrix) für das aktuelle Einzelbild erstellen. Manche Verfahren erstellen diese anhand bekannter Kamerabewegung. Kang et al. [KPK⁺03] haben in ihrem Verfahren angenommen, dass die PTZ-Kamera sich mit einheitlicher Geschwindigkeit dreht. Das heißt, die Schwenk- und Neigungswinkel der PTZ-Kamera sind konstante Größen. Sie haben nur einmal die projektive Transformation für die bekannten Schwenk- und Neigungswinkel berechnet und verwenden diese für jedes Einzelbild. Bhat et al. [BSK00] verfolgten einen ähnlichen Ansatz. Sie haben bei der Erstellung des Hintergrundmodells die berechneten projektiven Transformationen des aktuellen Einzelbildes zusammen mit den Winkeln gespeichert und verwendeten diese wieder bei einer ähnlichen Positionierung der Kamera. Ein Problem für solche Verfahren ist, dass sich die Kamera nicht genau mit den vorgegebenen Winkeln dreht. Dreht sich die Kamera ein halbes Grad zu kurz oder zu weit, so wird das aktuelle Einzelbild durch die gegebene Matrix nicht richtig transformiert und stimmt mit dem Hintergrundbild nicht überein. Um eine solche Versetzung zu korrigieren, ist eine Bildregistrierung nötig.

Manche Verfahren ermitteln die notwendige projektive Transformation für das aktuelle Einzelbild aus einem Translationsmodell, obwohl PTZ-Kameras sich nur drehen können. Die Verfahren von Cucchiara et al. [CPVo6] und von Bhat et al. [BSKoo] verwenden ein Translationsmodell zur Berechnung der Kamerabewegung. Solche Verfahren sind rechenaufwendig, weil Translationen bestimmter Pixel anhand ihrer Intensitätswerte bestimmt werden und aus den Translationen dann die projektive Transformation berechnet wird. Dabei wird angenommen, dass sich die Kamera zwischen zwei Einzelbildern lediglich um einen kleinen Winkel dreht.

Es gibt aber auch Verfahren, die mit Hilfe von Bildregistrierung die projektive Transformation für das aktuelle Einzelbild berechnet. Bei solchen Verfahren sind Bewegungsrichtung und Bewegungsgröße der Kamera unbekannt und müssen durch einen Vergleich der extrahierten Merkmale geschätzt werden. Das heißt, die Kamera kann sich in eine beliebige Richtung drehen und die Bewegung muss nicht vorher bekannt sein. Bei solchen Verfahren ist es wichtig, dass die Bilder fehlerfrei und möglichst in Echtzeit registriert werden. In [ADSB05] und in [BDBHo2] wurden solche Verfahren verwendet. Dabei werden die Merkmale des aktuellen Einzelbildes mit den Merkmalen des vorherigen Einzelbildes und desweiteren -zur Korrektur eines möglichen Fehlers- mit Merkmalen des Hintergrundmodells verglichen.

Die meisten der veröffentlichten Verfahren (alle oben zitierten Verfahren) erstellen ein sogenanntes Mosaik-Hintergrundmodell, welches durch Zusammensetzung (eng. stitching) der PTZ-Kamerabilder erstellt wird. Das heißt, die aufgenommenen Einzelbilder werden auf einer Ebene zusammengesetzt. Dies hat den Vorteil, dass keine Informationen der Kameraeigenschaften wie Brennweite, Bildwinkel, Koeffizienten der Verzerrung usw. benötigt werden, weil die Bilder bei der Zusammensetzung passend übereinander gelegt werden. Der Nachteil aber ist, dass ein Hintergrundmodell für einen großen Bereich der Schwenk- und Neigungswinkel bzw. für die gesamte Umgebung der PTZ-Kamera nicht erstellen werden kann. Das heißt, das gesamte Hintergrundmodell stellt nur ein kleinen Bereich der Kameraumgebung dar. Zhu et al. [ZXRH99] konnten mithilfe einer Transformation ein zylindrisches Panorama durch die Zusammensetzung dieser Bilder erstellen. Mithilfe bekannter Brennweite haben Chen [Che95] und McMillan et al. [MB95] auch ein zylindrisches Panorama erstellt. Sankaranarayanan et al. [SD10] veröffentlichten ein Verfahren zur Erstellung eines Halbkugel-Panoramas ($360^\circ \times 90^\circ$) für PTZ-Kameras. Sie verwendeten eine Methode zur Berechnung der Brennweite unter Verwendung von "SURF"-Merkmalen. Coorg et al. haben [CToo] auch ein Verfahren zur Erstellung von Halbkugel-Panoramas entwickelt.

Eine ganz andere Methode zur Hintergrundsubtraktion für PTZ-Kamera beschreiben Guillot et al. [GTS⁺10]. Ihre Hintergrundsubtraktion basiert bei dieser Methode auf einer Wahrscheinlichkeitsdichtefunktion. Die Regionen der Vordergrundobjekte werden anhand erkannter Merkmale des aktuellen Einzelbildes und des Hintergrundmodells segmentiert. Das heißt, wenn ein Merkmal eines Vordergrundobjektes zu keinem Merkmal des Hintergrundmodells zugeordnet werden kann, wird das entsprechende Pixel als Vordergrund angesehen. Solche Pixel bilden dann ein Vordergrundobjekt. Diese Hintergrundsubtraktion wird auch mit PTZ-Kamera verwendet. Dabei muss die Bewegung der Kamera vorgegeben sein und als Hintergrundmodell wird eine Gruppe von registrierten Einzelbildern mit bekannten Positionen verwendet. Nachteil dieses Verfahrens ist es, dass man bei einer Aktualisierung des Hintergrundmodells mehrere Einzelbilder, die als Hintergrundmodell definiert sind, aktualisieren muss. Zudem muss man mit einem geringen Grenzwert die Merkmale extrahieren, damit viele Merkmale überall im Bild gefunden werden können und die Dichte der Merkmale für Erkennung eines Vordergrundobjekts ausreicht. Die Merkmale, die mit einem geringen Grenzwert gefunden werden, haben den Nachteil, dass sie anhand ihrer Beschreibung nicht eindeutig identifiziert werden können und so falsche Zuordnungen der Merkmale entstehen. Ein ähnlicher Ansatz wurde auch von Trichet et al. [TMO8] veröffentlicht. In diesem Verfahren wurden erkannte Merkmale mit vier Eigenschaften (Merkmal, Farbe, Bewegung und Position der Bounding Box) als Vorder- oder Hintergrund beschrieben.

3.2. Deblurring

Es gibt zahlreiche Verfahren zur Beseitigung von Bewegungsunschärfe (Deblurring-Verfahren). Sie unterscheiden sich von einander hauptsächlich dadurch, wie sie die PSF schätzen. Verfahren, die die PSF nur aus den unscharfen Bildern selbst schätzen, gehören zu den Verfahren mit dem Oberbegriff *Blind Deconvolution*. Solche Verfahren verwenden meist ein einfaches Bewegungsmodell und geben in der Regel nur in bestimmten Situationen gute Ergebnisse. Deshalb sind solche Verfahren für PTZ-Kameras ungeeignet.

Der Richardson-Lucy Algorithmus [Ric72], [Luc74] und der Wiener Filter (siehe Abschnitt 5.1) werden bei Deblurring-Verfahren häufig verwendet.

Ein häufig zitiertes Verfahren ist das von Ben-Ezra et al. [BEN04] entwickelte Verfahren zum Deblurring. Sie haben angenommen, dass die Belichtungszeit proportional abhängig ist zur Bildauflösung. Dementsprechend haben kleine Bilder weniger Bewegungsunschärfe

3. Verwandte Arbeiten

als großer Bilder. Deshalb verwendeten sie zwei Kameras zum Deblurring. Während die erste Kamera mit hoher Auflösung wenige Bilder pro Sekunde aufnimmt, kann die zweite Kamera mehrere Bilder mit niedriger Auflösung aufnehmen. Aus den Bildern der zweiten Kamera wird die PSF bestimmt und diese wird zum Deblurring der Bilder der ersten Kamera verwendet. Als Entfaltungsfiler verwendeten sie den Richardson-Lucy Algorithmus. Das Verfahren liefert gute Resultate bei Bewegung der Kamera, erfordert jedoch eine zweite Kamera, die die Kamerabewegung misst.

Ein anderes bekanntes Verfahren veröffentlichten Bascle et al. [BBZ96]. Das Verfahren berücksichtigt neben der Bewegungsunschärfe auch die optische Unschärfe und das Sensorrauschen. Durch eine rekursive Minimierung des Unterschiedes zwischen dem aktuellen unscharfen und einem aus den letzten Einzelbildern simulierten Bild wird das Bild geschärft. Ein Vorteil dieses Verfahrens ist es, dass ein Entfaltungsfiler (*Inverse-Filter*) nicht explizit berechnet werden muss.

4. Erstellung des Hintergrundmodells

In diesem Kapitel wird die Erstellung des Hintergrundmodells der PTZ-Kamera erläutert. Zunächst wird die Grundidee vorgestellt, die zur Erstellung des Hintergrundmodells der PTZ-Kamera geführt hat.

Die Kameraeigenschaften spielen bei der Erstellung des Hintergrundmodells eine wichtige Rolle. Der Abschnitt 4.2 gibt einen Überblick über die Eigenschaften einer PTZ Kamera.

Im darauf folgenden Abschnitt wird erläutert, wie die PTZ-Parameter anhand der aufgenommenen Bilder geschätzt werden. Abschließend werden einige Kartenprojektionen vorgestellt, die zur Erstellung des Hintergrundmodells erprobt wurden.

4.1. Grundidee

PTZ-Kameras haben drei Freiheitsgrade. Daher erfordert die Hintergrundsubtraktion für eine PTZ-Kameras, dass

- das Hintergrundmodell die gesamte Umgebung der Kamera umfasst, damit es ein entsprechendes Hintergrundmodell für eine beliebige Kombination der PTZ-Parameter gibt und
- das Hintergrundmodell des aktuellen Einzelbildes aus dem gesamten Hintergrundmodell abhängig von den PTZ-Parametern dynamisch ermittelt werden kann.

Um diese Anforderungen zu erfüllen, mussten folgende Ziele erreicht werden.

1. Als Hintergrundmodell einer PTZ-Kamera muss eine 2D Ebene verwendet werden, sodass das aktuelle Einzelbild auf der Ebene abgebildet werden kann.

4. Erstellung des Hintergrundmodells

2. Die vorhandenen Verfahren zur Projektion von 3D auf 2D müssen untersucht werden, mit dem Ziel, ein geeignetes Verfahren zur Abbildung des Kamerabildes auf das Hintergrundmodell und umgekehrt auszuwählen.
3. Falls so ein Verfahren nicht vorhanden ist, muss ein geeignetes Verfahren entwickelt werden.
4. Es muss eine Methode zur Schätzung von PTZ-Parametern eines Kamerabildes entwickelt werden, die das Bildregistrierungsverfahren anwendet.

Die als Hintergrundmodell zu verwendende 2D Ebene sollte eine Kartenprojektion eines 3D Objekts sein. Bei der Erstellung des Hintergrundmodells wird das aktuelle Einzelbild auf die Oberfläche eines geschlossenen 3D Objekts projiziert und dann die Oberfläche des 3D Objekts mittels einer Kartenprojektion auf die 2D Ebene abgebildet.

Die Kartenprojektion sollte es ermöglichen, mit einem geringen Aufwand und möglichst genau ein Kamerabild zu einer 2D Projektion des den PTZ-Parametern (Zoom, hor. und vert. Winkel) entsprechenden Raumbereichs in dem Hintergrundmodell zuzuordnen.

Das Projektionsverfahren sollte sowohl zur Erstellung des Hintergrundmodells als auch zur Ermittlung eines Hintergrundmodells entsprechend einem Kamerabild verwendet werden können.

Es standen verschiedene 3D Objekte wie Kugel, Würfel oder Zylinder zur Auswahl. Für die 3D Objekte gibt es wiederum auch verschiedene Kartenprojektionen, die das 3D Objekt auf eine Ebene abbilden. Durch die Kartenprojektionen entstehen Verzerrungen im abgebildeten Bild auf der 2D Ebene für bestimmte Kamerasichtrichtungen. Die Stärke dieser Verzerrungen und der Berechnungsaufwand der Abbildungen für ein Einzelbild hängen von der Auswahl des 3D Objekts und der Kartenprojektion ab. Die Kugel wurde ausgewählt, weil ihre Oberfläche mit der Bewegung der PTZ-Kamera voll abgedeckt werden kann und sie es ermöglichte, die Umgebung der Kamera vollständig abzubilden.

Der Zylinder hatte den Nachteil, dass der Bereich direkt über der Kamera auf die 2D Ebene nicht projiziert werden konnte. Die Rektangularprojektion wurde für das Projizieren der Kugel auf die 2D Ebene ausgewählt.

Die Abbildung 4.1 zeigt die Erstellung des Hintergrundmodells für die PTZ-Kamera. Das Hintergrundmodell umfasst die komplette Umgebung der PTZ-Kamera. Nachdem die PTZ-Parameter ermittelt wurden, wird das aktuelle Einzelbild auf die Oberfläche einer Kugel

projiziert. Die Position auf die das Einzelbild projiziert wird, wird durch die PTZ-Parameter bestimmt. Anschließend wird die Oberfläche der Sphäre mit der Rektangularprojektion auf die 2D Hintergrunde ebene abgebildet. Man sieht, dass das Einzelbild verzerrt auf die 2D Hintergrunde ebene abgebildet wird.

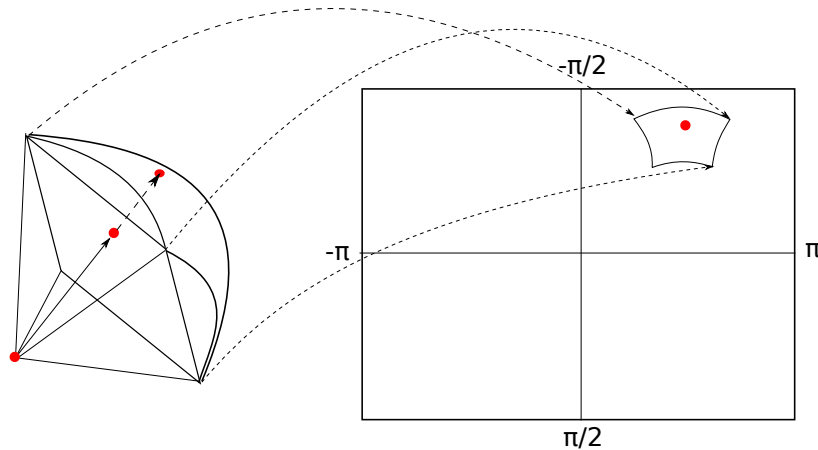


Abbildung 4.1.: Links: Projektion eines Einzelbildes auf einer Sphäre, Rechts: Abbildung von Oberfläche der Sphäre auf einer Ebene

Die PTZ-Parameter wurden ermittelt durch das Wiederfinden von Merkmalen in zwei aufeinander folgenden Kamerabildern. Die Merkmale wurden durch Bildregistrierungsverfahren gefunden. Das Verfahren wird in Abschnitt 4.3 im Detail beschrieben.

Ursprünglich war geplant, die PTZ-Parameter des aktuellen Einzelbildes durch Wiederfinden des Einzelbildes im Hintergrundmodell zu schätzen, die Sphäre als 3D Objekt und die Rektangularprojektion als Kartenprojektion zu verwenden.

Bei der Rektangularprojektion entstanden bei Einzelbildern, die in der Nähe von $-\frac{\pi}{2}\theta$ bzw. $\frac{\pi}{2}\theta$ (in der Nähe von Polen) aufgenommen wurden, große Verzerrungen. Aus diesem Grund mussten weitere Kartenprojektionen erprobt werden.

4.2. Messungen der Kameraeigenschaften

Wie in der Einleitung beschrieben, ist die PTZ-Kamera eine bewegliche Kamera, deren Sicht- richtung und Zoomfaktor über die Funktionen Schwenken-Neigen-Zoomen ferngesteuert

4. Erstellung des Hintergrundmodells

sind (Abbildung 1.2a). Eine Fahrbewegung kann sie nicht ausführen. In der Regel werden PTZ-Kameras an einem bestimmten Ort befestigt und haben als Haupteinsatzgebiet Überwachungsaufgaben. Bei intelligenten Überwachungssystemen werden sie oft in Kombination mit anderen unbeweglichen Kameras zusammen verwendet. Die meisten PTZ-Kameras können sich mithilfe ihrer Schwenkfunktion um ihre vertikale Achse so weit drehen, dass die aufgenommenen Bilder einen geschlossenen Kreis bilden können. Der Neigungswinkel ist dagegen meist weniger als 180° , weil eine PTZ-Kamera entweder oben oder unten befestigt wird. Auf der Abbildung 4.2 ist die Drehbereiche der PTZ-Kamera "SONY SNC-RZ30" (Abbildung 1.2a), die für diese Arbeit verwendet wurde, dargestellt. Man sieht, dass sich diese Kamera um 340° schwenken und um 115° neigen kann. Der optische Zoomfaktor der Kamera kann zwischen x1 und x25 liegen. Bevor das Hintergrundmodell einer PTZ-

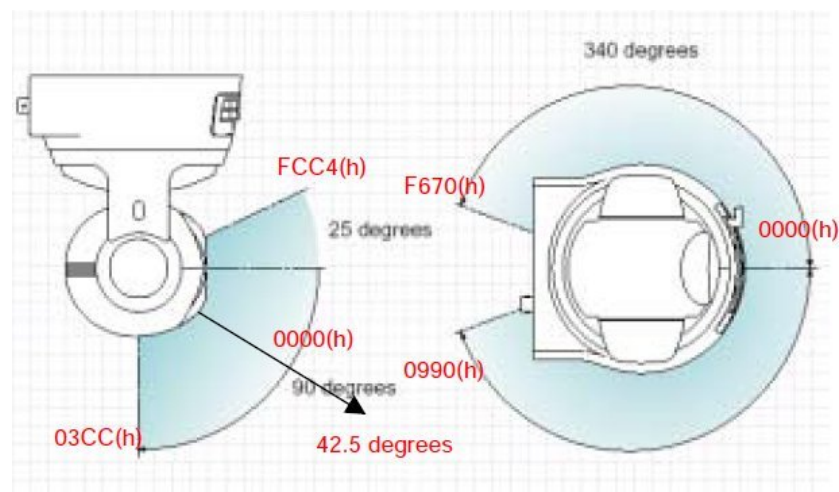


Abbildung 4.2.: Schwenkbereich und Neigungswinkel der PTZ-Kamera "SONY SNC-RZ30"
Quelle: Benutzerhandbuch von SONY SNC-RZ30

Kamera erstellt wird, müssen einige ihrer Eigenschaften überprüft werden und die dabei festgestellten Fehler korrigiert werden.

4.2.1. Test der Parallaxenverschiebung

PTZ-Kameras haben zwei Drehachsen. Beim Schwenken dreht sich die Kamera um ihre vertikale Achse und beim Neigen dreht sie sich um ihre horizontale Achse. Für die Erstellung des Hintergrundmodells ist es von großer Bedeutung, dass sich eine PTZ-Kamera in beiden

Drehungen um ihr optisches Zentrum dreht. Das optische Zentrum einer Kamera nennt man Nodalpunkt (Knotenpunkt). In Abbildung 4.3a wird der Nodalpunkt der gezeichneten Kamera durch einen roten Punkt dargestellt. Der Nodalpunkt liegt auf der optischen Achse einer Kamera. Die exakte Position des Punktes hängt von der Brennweite der Kamera ab. Das heißt, dass sich bei einer PTZ-Kamera die Position des Nodalpunktes ändert, wenn man den Zoomfaktor der Kamera ändert. Der Nodalpunkt hat die Eigenschaft, dass die Kamera Bilder ohne Parallaxenverschiebung liefert, wenn sie sich bei der Aufnahme der Bilder um ihren Nodalpunkt dreht. Unter einer Parallaxenverschiebung versteht man eine Änderung des Abstandes zwischen zwei Objekten aus der Perspektive eines Beobachters in verschiedenen Positionen. Als Beispiel zeigt Abbildung 4.3b, dass der Abstand zwischen der Laterne und der Sonne wesentlich kleiner ist, als wenn man ihn von der Position sieht, bei der die Spiegelung im Wasser sich befindet.

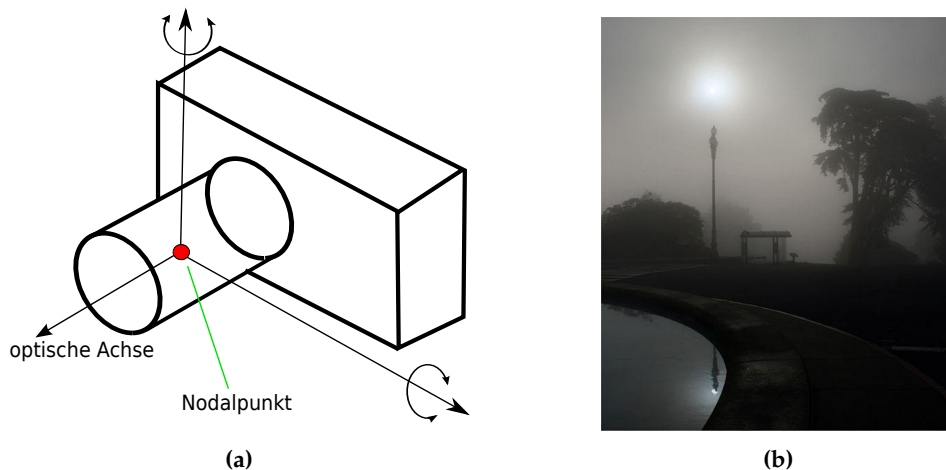


Abbildung 4.3.: (a) Nodalpunkt, (b) Beispiel für Parallaxenverschiebung

Dreht sich eine PTZ-Kamera nicht um ihren Nodalpunkt, ändert sich die Position des Nodalpunktes bei der Drehung der Kamera und die durch die Kamera aufgenommenen Objekte werden aus unterschiedlichen Positionen und von verschiedenen Seiten betrachtet. Dadurch entsteht die Parallaxenverschiebung bei den aufgenommenen Bildern. Man muss bei Objekten, die kleinere Abstände zur Kamera haben, mit einer stärkeren Parallaxenverschiebung rechnen. Nachfolgend wird erklärt, weshalb die Parallaxenverschiebung die Erstellung des Hintergrundmodells erschwert:

4. Erstellung des Hintergrundmodells

- Durch die Parallaxenverschiebung werden manche Objekte in einigen Einzelbildern verdeckt, sodass deren Merkmalen nicht gefunden und zugeordnet werden können. Ein Beispiel hierfür zeigt Abbildung 4.4. Die Fenster im linken Bild sind im rechten Bild verdeckt. Infolgedessen sind die aus den Fenstern erkannten Merkmalen unbrauchbar bei der Berechnung der projektiven Transformation zwischen den beiden Bildern. Die Parallaxenverschiebung vermindert also die Anzahl der zuzuordnenden Merkmale. Dies kann sich negativ auf die Schätzung der PTZ-Parameter für das aktuelle Einzelbild auswirken.



Abbildung 4.4.: Links: Vor der Kameradrehung sind mehrere Fenster sichtbar, Rechts: Fenster sind nach Kameradrehung durch Parallaxenverschiebung verdeckt.

- Des Weiteren gibt es verschiedene Möglichkeiten für die Zusammensetzung von zwei Einzelbildern durch die Parallaxenverschiebung. Zum Beispiel können zwei verschiedene projektive Transformationen für die in Abbildung 4.5 dargestellten Bilder berechnet werden. Eine ist die Berechnung, dass die Zeigefinger der Bilder zusammenpassen und andere ermöglicht es die Hintergrundmodelle der Bilder zu vergleichen. Welche Berechnung durchgeführt werden soll, hängt von der Anzahl der erkannten Merkmale für die jeweilige Berechnung ab. Für jede Berechnung erhält man unterschiedliche PTZ-Parameter und unterschiedliche Projektionen des aktuellen Einzelbildes auf die Hintergrundebene. Im Falle einer falschen Projektion kann ein Objekt jedoch mehrmals im Hintergrund vorkommen.

Ob die verwendete PTZ-Kamera sich um ihren Nodalpunkt dreht bzw. wie groß die Parallaxenverschiebung bei der Kamera ist, kann mit einem Test geprüft werden. Hierfür werden zwei Objekte, die auf einem Bild als erkennbare Punkte bzw. Linien erscheinen können, werden dafür verwendet. Sie müssen von der PTZ-Kamera unterschiedlich weit entfernt sein (ein Objekt nur geringfügig entfernt und das andere Objekt möglichst weit entfernt) und so befestigt sein, dass sie mit der Kamera ungefähr auf einer Gerade liegen. Dann werden zwei

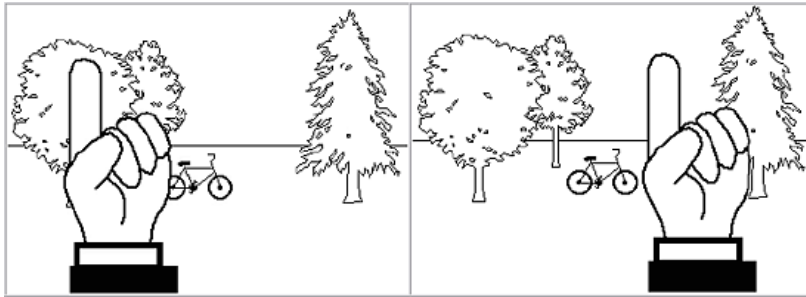


Abbildung 4.5.: Parallaxenverschiebung der Augen: **Links** - die Sicht mit dem linken Auge, **Rechts** - die Sicht mit dem rechten Auge. Quelle: [par]

Bilder aus verschiedenen Sichtrichtungen der Kamera aufgenommen. Die Bilder müssen die Messobjekte enthalten. Die Abbildung 4.6 zeigt die zwei beispielhafte Bilder. Als Messobjekte wurden ein senkrecht von der Decke hängender Bleistift und eine dahinter stehende Befestigungsstange eines Projektors verwendet. Damit eine mögliche Parallaxenverschiebung leichter erkannt wird, wurde die Kamera bei der Aufnahme der zwei Bilder soweit gedreht, dass die Messobjekte auf einem Bild im linken Bereich und auf dem anderen Bild im rechten Bereich aufgenommen wurden.



Abbildung 4.6.: Parallaxenverschiebung bei der verwendeten PTZ-Kamera für den kleinsten Zoomfaktor

Durch den Vergleich der Abstände zwischen den beiden Messobjekten in den aufgenommenen Bildern kann die Parallaxenverschiebung geprüft werden. Die Kamera dreht sich nicht um ihren Nodalpunkt, wenn der Abstand in den Bildern unterschiedlich groß ist. Auf dem linken Bild sieht man eine Lücke zwischen den Messobjekten. Auf dem rechten Bild ist eine solche Lücke nicht zu erkennen. Um die Parallaxenverschiebung bei der Neigebe-

4. Erstellung des Hintergrundmodells

wegung der Kamera zu zeigen, wird der Bleistift horizontal von der Decke gehängt und mit der oberen Kante der dahinter liegenden Wand verglichen (siehe Abbildung 4.7). Die Parallaxenverschiebung ist auch hier zu sehen. Der Abstand zwischen den Messobjekten ist auf dem rechten Bild nicht mehr erkennbar. Somit dreht sich die PTZ-Kamera bei der eingestellten Zoomstufe nicht um ihren Nodalpunkt.



Abbildung 4.7.: Parallaxenverschiebung bei Neigebewegung

Der gleiche Test wurde auch mit einem höheren Zoomfaktor durchgeführt. Das Ergebnis dieses Tests wird in Abbildung 4.8 dargestellt. Man sieht, dass die PTZ-Kamera mit dem eingestellten Zoomfaktor fast keine Parallaxenverschiebung erzeugt.



Abbildung 4.8.: Parallaxenverschiebung für großen Zoomfaktor

4.2.2. Kamerakalibrierung

Für die Erstellung des Hintergrundmodells einer PTZ-Kamera ist es wichtig, dass die Bilder keine optischen Verzerrungen enthalten. Die Ursachen einer optischen Verzerrung und die sogenannte Kamerakalibrierung wurden im Abschnitt 2.1.4 erklärt. Sowohl radiale als auch tangentielle Verzerrungen eines Bildes können eine falsche Berechnung für die Kamerasicht-richtung verursachen. Dies wirkt sich dann negativ auf die Abbildungen des aktuellen Einzelbildes aus, sodass die Bilder in dem Hintergrundmodell nicht zusammenpassen und schließlich bei der Hintergrundsubtraktion falsche Vordergründe segmentiert werden. Darüber hinaus verwendet die Berechnung vom initialen Bildwinkel der PTZ-Kamera eine Kameramatrix, die durch die Kamerakalibrierung erstellt wird (siehe Abschnitt 4.2.3).

Für die Kalibrierung einer Kamera kann die OpenCV-Bibliothek verwendet werden. Das Kamerakalibrierungsverfahren von Zhang [Zhaoo] wurde in OpenCV umgesetzt. Das Verfahren basiert auf einem Schachbrettmuster. Im Vergleich zu manchen Verfahren ist dieses Verfahren ziemlich einfach, weil sich die Geometrie eines Schachbrettmusters einfacher berechnen lässt als die Geometrie von 3D Objekte und die Erstellung des Musters auch nicht kompliziert ist. In OpenCV sind die notwendigen Funktionen für die Kamerakalibrierung bereits implementiert. Mithilfe dieser Funktionen lassen sich die intrinsischen und extrinsischen Parameter einer Kamera berechnen und optisch verzerrte Bilder entzerren.

Man nimmt dafür mehrere Bilder des Schachbrettmusters aus unterschiedlichen Sichten auf und führt anschließend anhand der bekannten Geometrie des Musters mit den in OpenCV bereitgestellten Funktionen die Kamerakalibrierung durch. Ein Programm zur Kamerakalibrierung *calibration.cpp* ist in den Beispielen der OpenCV-Dokumentation vorhanden. Dieses Programm wurde in der vorliegenden Arbeit verwendet.

Die Funktionen der OpenCV-Bibliothek für Kamerakalibrierung basieren auf dem Lochkammermodell (siehe Abschnitt 2.1.3). Die Kameramatrix enthält aber bei OpenCV vier intrinsische Parameter. Es wird angenommen, dass der Wert des fünften Parameters *Scherung* s den Wert Null hat, sodass die optische Achse und die Bildebene der kalibrierenden Kamera einen 90° -Winkel bilden. Die intrinsischen Parameter wurden im Abschnitt 2.1.3 vorgestellt und erklärt.

Die Kameramatrix ist unabhängig von der Bewegung der Kamera, aber sie ist abhängig von der Brennweite der Kamera. Das heißt, eine durch die Kamerakalibrierung erstellte Kameramatrix kann solange verwendet werden, wie der Zoomfaktor der Kamera unverändert

4. Erstellung des Hintergrundmodells

bleibt. Dies stellt ein Problem bei PTZ-Kameras dar, weil während der Bearbeitung die Zoomfunktion dieser Kameras genutzt werden kann. Eine Möglichkeit zur Lösung dieses Problems wäre, dass man schon vor der Bearbeitung durch die Kamerakalibrierung mehrere Kameramatrizen für alle Zoomfaktoren erstellt und beim Zoomen für den ermittelten Zoomfaktor die entsprechende Kameramatrix verwendet.

Vor dem Beginn der Kalibrierung muss man ein Schachbrettmuster beschaffen, weil OpenCV ein solches Muster verwendet. Die Felder des Musters müssen quadratisch sein und die Farben Schwarz und Weiß haben, damit der Unterschied zwischen den Farbwerten groß ist und die inneren Ecken der Schachbrettfelder gut erkennbare Merkmale im Bild erzeugen. Ein solches Muster ist auch in der OpenCV-Dokumentation zu finden. Für genauere Berechnungen ist es besser, wenn das Schachbrett über mehrere Felder verfügt, sodass mehrere Markierungen gefunden und zur Berechnung verwendet werden können. Das Muster kann auf einem Stück Papier ausgedruckt werden. Eine Voraussetzung für korrekte Messungen ist, dass das Muster mit hoher Auflösung gedruckt wird und das bedruckte Papier keine Knicke hat und nicht gewellt ist. Optimal ist es, wenn das Papier auf einer festen und glatten Ebene angebracht wird oder in einen Bilderrahmen mit Glasscheibe eingesetzt wird, sodass es während der Messungen immer auf einer Ebene bleibt (siehe linkes Bild in Abbildung 4.9).

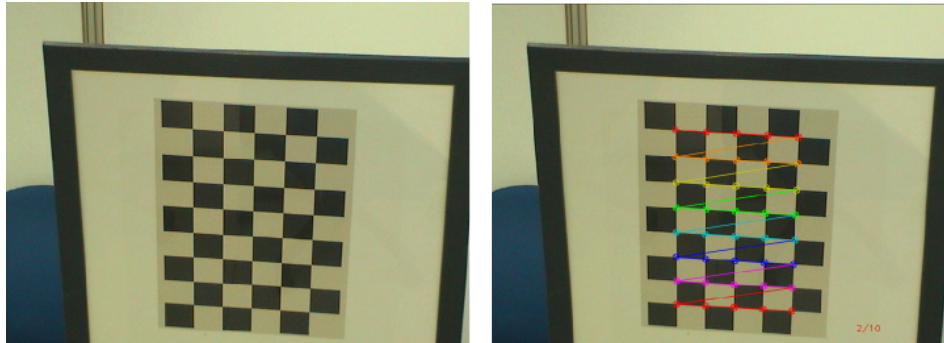


Abbildung 4.9.: Links: Schachbrettmuster für Kamerakalibrierung, Rechts: die innere Ecken sind erkannt

Anschließend kann man Bilder des Schachbrettmusters mit unterschiedlichen Orientierungen durch die Kamera aufnehmen. In Abbildung 4.10 wurden einige der aufgenommenen Bilder dargestellt. Vorteilhaft ist, wenn das Muster den Großteil des Bildes abdeckt, weil dann die Verzerrungen am Rand des Bildes stärker sind. Aus diesen Bildern ermittelt die Funktion `findChessboardCorners()` die Positionen der inneren Ecken der Schachbrettfelder. Mit den

gegebenen Größen des Musters können dann die Lage und die intrinsischen Parameter der Kamera berechnet werden.

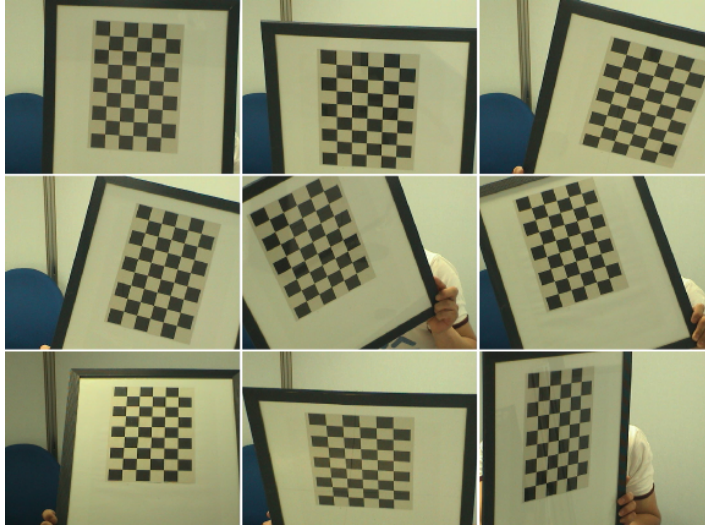


Abbildung 4.10.: Bilder vom Schachbrettmuster mit verschiedenen Orientierungen für Kamerakalibrierung

Im nächsten Schritt werden die aufgenommenen Bilder in einer XML-Datei gespeichert und an das Kamerakalibrierungs-Programm übermittelt. Die XML-Datei sieht wie folgt aus:

Listing 4.1 Eingabe XML-Datei, die aufgenommene Bilder listet

```
<?xml version="1.0"?>
<opencv_storage>
<images>
  bilder/s1.jpg
  bilder/s2.jpg
  bilder/s3.jpg
  bilder/s4.jpg
  bilder/s5.jpg
  ...
</images>
</opencv_storage>
```

Nun kann das Programm aufgerufen werden. Es müssen die Anzahl der inneren Ecken für jede Reihe und Spalte des Musters und die Seitenlänge der Felder des Schachbrettmusters in irgendeiner Einheit (z.B. cm oder mm) eingegeben werden. Das verwendete Schachbrettmus-

4. Erstellung des Hintergrundmodells

Algorithmus 4.1 Beispielalgorithmus zum Entzerren eines Bildes

```
void undistSample(){
    Mat imgSrc = imread("verzerrtesBild.jpg"); // optisch verzerrtes Bild
    Mat imgDst;                               // entzerrtes Bild
    Mat cameraMatrix;                          // Kameramatrix
    Mat distCoeffs;                            // Verzerrungskoeffizienten

    FileStorage fs("nexusCalib.xml",FileStorage::READ); // XML-Datei laden

    fs["camera_matrix"] >> cameraMatrix;
    fs["distortion_coefficients"] >> distCoeffs;

    undistort(imgSrc, imgDst, cameraMatrix, distCoeffs); // entzerren
}
```

ter hat 6×9 Felder bzw. 5×8 innere Ecken und die Seitenlänge der Felder ist 3,25cm. Der Aufruf des Programms kann z.B. so aussehen:

```
./calibration -w 5 -h 8 -s 3.25 -o nexusCalib.xml imageList.xml
```

wobei die XML-Datei imageList.xml die Eingabe-Bilder listet und in der Datei nexusCalib.xml die aus Kamerakalibrierung berechnete Kameraparameter als Ausgabe gespeichert sein werden. Mit dem Programm können auch Videodaten anstelle von Bildern als Eingabe gegeben werden. Nach dem Aufruf lädt das Programm die Bilder und zeigt sie nacheinander an. Wenn bei einem Bild alle inneren Ecken erkannt werden konnten, sollten verschiedenfarbige Linien auf dem angezeigten Bild zu sehen sein (siehe rechte Seite in Abbildung 4.9). Am Schluss speichert das Programm die Verzerrungskoeffizienten, die intrinsischen Parameter und andere Parameter der Kamera in der XML-Datei ab. Die Kameramatrix und die Verzerrungskoeffizienten der Kamera werden in Listing 4.2 dargestellt.

Somit ist die Kamera kalibriert und man kann die in der XML-Datei gespeicherten Parameter in anderen Programmen zum Entzerren der Kamerabilder verwenden. Der Algorithmus 4.1 zeigt ein Beispiel, wie die radiale oder die tangentielle Verzerrung eines Bildes mithilfe der Funktion `undistort()` von OpenCV entfernt werden kann.

Als Ergebnis sind die in Abbildung 4.11 dargestellte zwei Bilder zu sehen. Das linke Bild ist von der PTZ-Kamera aufgenommen und zeigt eine Decke mit Punkten, die eigentlich

Listing 4.2 Ausgabe XML-Datei: die berechnete Kameramatrix und Verzerrungskoeffizienten sind gezeigt.

```
<?xml version="1.0"?>
<opencv_storage>
...
<camera_matrix type_id="opencv-matrix">
  <rows>3</rows>
  <cols>3</cols>
  <dt>d</dt>
  <data>
    9.7223566126106198e+02    0.                3.7986276467636276e+02
    0.                    9.7978047405608356e+02  2.4456526251388686e+02
    0.                    0.                1.
  </data>
</camera_matrix>
<distortion_coefficients type_id="opencv-matrix">
  <rows>5</rows>
  <cols>1</cols>
  <dt>d</dt>
  <data>
    -2.5937604981904311e-01  2.0938212810302501e-01  -3.3306064307160595e-03
    1.0037037229486826e-04  1.7075931413032088e+00
  </data>
</distortion_coefficients>
...
</opencv_storage>
```

in einer Gerade stehen. Das Bild enthält kleine tonnenförmige Verzerrung und diese sieht man am oberen Rand des Bildes deutlich, wo die Verzerrung maximal ist. Die dort liegende Punkte weichen von einer Gerade leicht ab und stellen eine parabelförmige Kurve dar. Um dies deutlicher zu veranschaulichen ist eine Vergrößerung des oberen Bereichs in der Mitte des Bildes gezeigt und die Punkte sind durch eine roten Linie verbunden. Zum Entzerren des Bildes wird der Algorithmus 4.1 auf der vorherigen Seite mit den berechneten Parametern ausgeführt. Das rechte Bild ist das Ergebnis dieser Entzerrung. Im rechten Bild ist auch eine Vergrößerung vom gleichen Bereich zu sehen. Man sieht, dass die krumme rote Linie durch die Entzerrung ziemlich gerade geworden ist (die rote Linie auf dem rechten Bild ist nicht nachgezeichnet. Sie ist das Ergebnis der roten Linie vom Links).

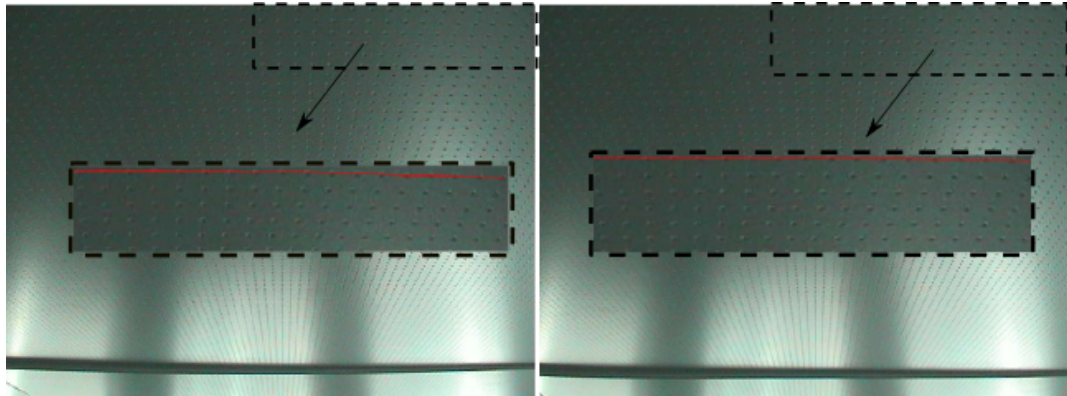


Abbildung 4.11.: Links: ein von der Kamera aufgenommene verzerrtes Bild,
Rechts: entzerrtes Bild

4.2.3. Berechnung des initialen Bildwinkels

In diesem Abschnitt wird die Berechnung eines initialen Bildwinkels, der für die Erstellung des Hintergrundmodells verwendet wird, erläutert. Wie der Bildwinkel einer PTZ-Kamera während der Verarbeitung geschätzt werden kann, wurde im Abschnitt 4.4 behandelt.

Der Bildwinkel der Kamera wird für die Erstellung des Hintergrundmodells benötigt, das die Umgebung der PTZ-Kamera darstellt. Mithilfe des Bildwinkels wird ermittelt, wie groß das Hintergrundmodell sein soll.

Bevor der Bildwinkel einer PTZ-Kamera ermittelt wird, muss für das zu erstellende Hintergrundmodell eine passenden Zoomstufe ausgewählt werden, weil der Bildwinkel einer Kamera hängt von ihrem Zoomfaktor ab. Das heißt, wenn der Benutzer während der Verarbeitung den Zoomfaktor der PTZ-Kamera ändert, ändert sich auch der Bildwinkel der Kamera. Ist der Zoomfaktor für die Berechnung des Bildwinkels bzw. für die Erstellung des Hintergrundmodells zu klein gewählt, wird der Bildwinkel, den wir für die Erstellung des Hintergrundmodells ermitteln, groß sein und es wird infolgedessen ein zu kleines Hintergrundmodell erstellt. In diesem Fall passen Bilder, die mit großen Zoomfaktoren aufgenommen werden, nicht auf das Hintergrundmodell und müssen deshalb klein skaliert werden, bevor sie auf dem Hintergrundmodell abgebildet werden. Dieses Szenario wurde links auf der Abbildung 4.12 dargestellt. Die Bilder werden auf einer Sphäre projiziert. Die Größe der Sphäre wurde entsprechend für den Bildwinkel β , der einem kleinen Zoomfaktor der Kamera entspricht, berechnet. Für die Projektion eines Bildes mit dem Bildwinkel α

(wobei $\alpha < \beta$ gilt) wird das Bild verkleinert, sodass es in die Sphäre passt. Der Nachteil dieser Vorgehensweise ist, dass durch die Skalierungen Informationen verloren gehen können.

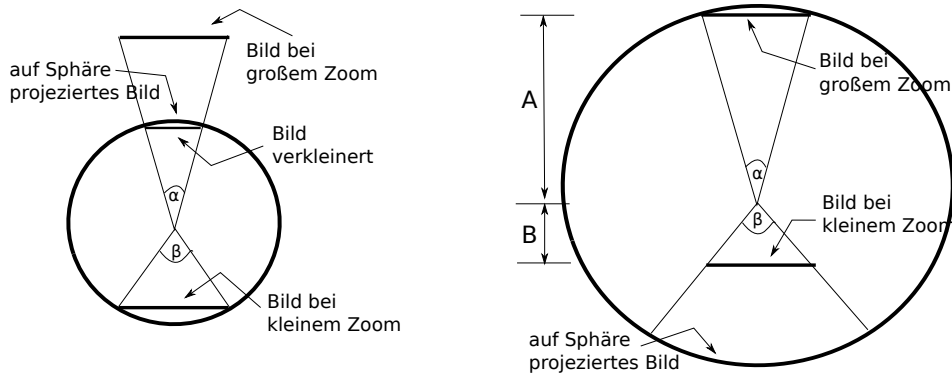


Abbildung 4.12.: **Links:** Erstellung einer Sphäre für einen zu kleinen Zoomfaktor. Die Bilder müssen vor der Projektion verkleinert werden, **Rechts:** Umgekehrter Fall: Die Bilder müssen im richtigen Abstand B vom Zentrum projiziert werden.

Wenn der Zoomfaktor zu groß ausgewählt wurde, wird das Hintergrundmodell sehr groß sein und die Bilder können ohne Skalierung auf dem 3D Objekt projiziert werden. Man muss jedoch bei der Projektion eines Bildes den richtigen Abstand zwischen dem Bild und dem Zentrum der Projektion berechnen. Dieser Fall wurde rechts in Abbildung 4.12 dargestellt. Hier wird der Radius der Sphäre mit dem Bildwinkel α berechnet und die Bilder mit dem Bildwinkel β (wobei $\beta > \alpha$ gilt) müssen bei einem kleinerem Abstand vom Zentrum auf die Sphäre projiziert werden. Der Nachteil ist aber, dass in diesem Fall die Bearbeitung der Bilder wegen des großen Hintergrundmodells rechenaufwendig sein kann. Die Größe des Hintergrundmodells ist z.B. 6750×3375 Pixel für den mit dem kleinsten Zoomfaktor der verwendeten PTZ-Kamera berechneten Bildwinkel. Deshalb muss man bei der Auswahl einer Zoomstufe für die Erstellung des Hintergrundmodells die daraus resultierenden Folgen beachten.

Wie im Abschnitt 2.1.1 beschrieben, gibt es für ein Bild drei Bildwinkel. Man kann sie mit der Formel $\alpha = 2 \cdot \arctan\left(\frac{d}{2 \cdot f}\right)$ berechnen, wobei f die Brennweite und d eine Seitenlänge bzw. die Diagonale des Bildsensors der Kamera sind.

In dieser Arbeit wird der horizontale Bildwinkel berechnet und für die Erstellung des Hintergrundmodells verwendet. Weil die Länge der Brennweite der Kamera für die ausgewählte Zoomstufe nicht bekannt war, wurde die Brennweite f_x , die bei der Kamerakalibrierung

4. Erstellung des Hintergrundmodells

berechnet und in der Kameramatrix gespeichert wurde, für die Berechnung verwendet. Der horizontale Bildwinkel wird wie folgt geschätzt:

$$\alpha = 2 \cdot \operatorname{atan} \left(\frac{\text{Bildbreite}}{2 \cdot f_x} \right)$$

Der so berechnete Bildwinkel beträgt $41,46^\circ$ für den kleinsten Zoomfaktor der verwendeten PTZ-Kamera. Für die Berechnung wurde die Bildbreite verwendet. Wenn man die Bildhöhe oder die Diagonale verwendet, erhält man kleinere bzw. größere Winkel. Des Weiteren kann mit diesem geschätzten Bildwinkel ein zu großes oder ein zu kleines Hintergrundmodell berechnet werden, weil die Brennweite f_x nicht der physikalischen bzw. der echten Brennweite entspricht (siehe Abschnitt 2.1.3). Wenn der Bildwinkel z.B. größer als der korrekte Winkel geschätzt wurde, wird ein zu kleines Hintergrundmodell erstellt. In diesem Fall wird das aktuelle Einzelbild auf den zu großem Bereich im 3D Objekt projiziert und dies führt dazu, dass das nächste Einzelbild unpassend überlappend auf vorherigen projiziert wird. Ein Beispiel hierfür wird in Abbildung 4.13 (linkes Bild) dargestellt. Das Bild wurde mit einem zu großen Winkel projiziert und den großen Projektionsbereich sieht man durch die rote Kurve auf der Sphäre. Der richtige Bildwinkel und der richtige Projektionsbereich wurden mit grüner Farbe gezeichnet. Bei der Betrachtung des Bildes erkennt man, dass die Sphäre größer erstellt bzw. der Bildwinkel kleiner geschätzt werden sollte. Im schlimmsten Fall kann ein Bild auf dem 3D Objekt nicht projiziert werden, wenn das Hintergrundmodell so klein erstellt wurde, dass ein Bild im 3D Modell nicht hineinpasst.

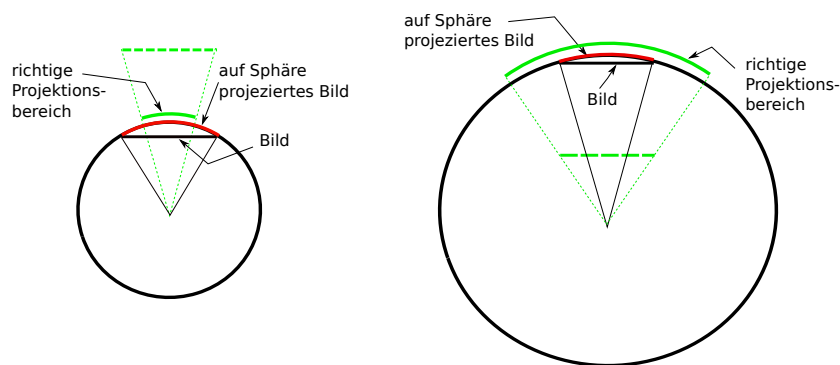


Abbildung 4.13.: Links: Bild wird auf einen zu großen Bereich (roter Bereich im Kreis) projiziert, weil der Bildwinkel zu groß geschätzt war. Rechts: Im umgekehrten Fall wird das Bild auf einen zu kleinen Bereich projiziert (roter Bereich im Kreis).

Wurde der Bildwinkel kleiner als der korrekte Winkel geschätzt, wird ein zu großes Hintergrundmodell erstellt und die PTZ-Kamera umfassende Bilder können das Hintergrundmodell nicht komplett füllen. In diesem Fall wird das Hintergrundmodell Lücken und doppelte Bereiche enthalten. Das heißt, bei der Erstellung des Hintergrundmodells, wenn sich die PTZ-Kamera vollständig um ihre vertikale Achse dreht (z.B. 360°), wird ein bestimmter Bereich doppelt als Hintergrund initialisiert und wenn sich die PTZ-Kamera vollständig um ihre horizontale Achse dreht (z.B. 180°), bleiben bestimmte Bereiche des Hintergrundmodells leer. Daraus resultieren weitere Probleme bei der Hintergrundsubtraktion, weil aus dem gesamten Hintergrundmodell für alle Sichtrichtungen der Kamera ein eindeutiges Hintergrundmodell zu finden sein sollte und die eine Seite des Hintergrundmodells die Fortsetzung der anderen Seite sein kann.

Den Bildwinkel, der mit der angegebenen Formel geschätzt wurde, kann man grob überprüfen, indem man die PTZ-Kamera in einer waagerechten Position (die optische Achse der Kamera ist horizontal) um ihre vertikale Achse dreht und dabei zählt, wieviele Drehungen sie für volle 360° benötigt. Zum Beispiel sollte die Kamera etwa 10 Drehungen benötigen, wenn der Bildwinkel 36° beträgt. Eine andere Möglichkeit, den Bildwinkel zu bestimmen, ist die rechnerische Ermittlung der korrekten Brennweite der Kamera über die Linsengleichung oder über die Linsenschleiferformel. Um solche Berechnungen durchführen zu können, müssen allerdings einige Größen der Kamera wie die Bildweite (siehe Abbildung 2.3) oder die Kugelradien der Kameralinsen bekannt sein.

Wurde der Bildwinkel für einen initial ausgewählten Zoomfaktor berechnet bzw. geschätzt, kann dann das Hintergrundmodell erstellt werden. Nach dem Erstellen bleibt die Größe des Hintergrundmodells unverändert – auch dann, wenn der Zoomfaktor im Laufe der Bearbeitung verändert wird.

4.3. Ermittlung der Kamerasichtrichtung

Die Kamerasichtrichtung ist die Richtung entlang der optischen Achse der Kamera, wo sie gerade schaut. Für die Ermittlung der Kamerasichtrichtung werden die von der Kamera aufgenommenen Einzelbilder verwendet. Die Ermittlung basiert daher auf Bildregistrierungsverfahren. Mit anderen Worten, für eine Position der Kamera wird ihre Sichtrichtung mit dem von der Position aufgenommenen Einzelbild ermittelt.

4. Erstellung des Hintergrundmodells

Um die Kamerasichtrichtung zu beschreiben, wird ein Koordinatensystem aufgebaut, in dessen Ursprung sich das optische Zentrum der Kamera befindet. Die Kamerasichtrichtung kann dann durch zwei Kugelkoordinaten (θ, ϕ) , die die Richtung vom Ursprung in die Bildmitte zeigen, definiert werden (siehe Abbildung 4.14).

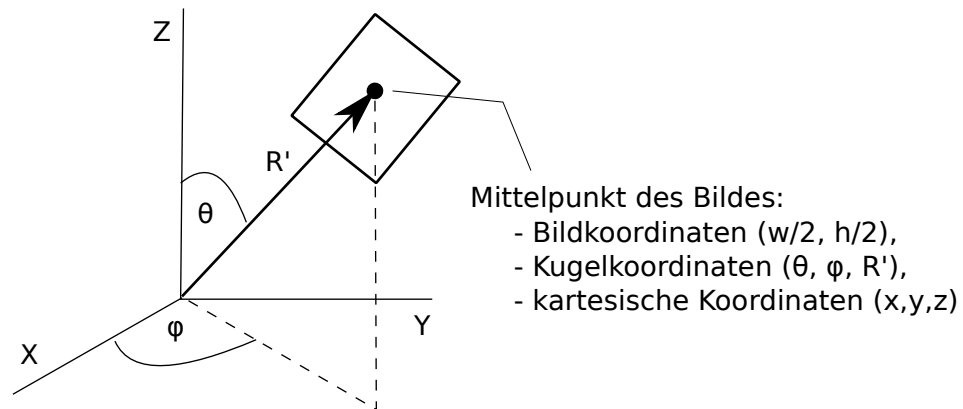


Abbildung 4.14.: die Winkel (θ, ϕ) beschreiben die Kamerasichtrichtung und entsprechen der Bildmitte

Für die Bestimmung der Kamerasichtrichtung müssen die beiden Winkel θ und ϕ des aktuellen Einzelbildes berechnet werden. Für die Berechnung wird das aktuelle Einzelbild erst registriert und die gefundenen Merkmale werden dann mit den Merkmalen des vorherigen Einzelbildes verglichen. Es gibt zwei Möglichkeiten, die Sichtrichtung zu schätzen. Eine davon ist die **lokale Methode**, die in dieser Arbeit verwendet wurde. Bei der lokalen Methode wird das aktuelle Einzelbild mit dem vorherigen Einzelbild verglichen und mithilfe der Winkel der Kamerasichtrichtung des vorherigen Einzelbildes wird die neue Kamerasichtrichtung berechnet. Die andere Methode heißt **globale Methode**. Bei globaler Methode wird die Kamerasichtrichtung aus dem Vergleich zwischen dem aktuellen Einzelbild und dem gesamten Hintergrundmodell ermittelt. Beide Methoden haben ihre Vor- und Nachteile.

Es gibt einige Probleme bei der Ermittlung der Kamerasichtrichtung.

- Zum Beispiel kann die Kamerasichtrichtung nicht berechnet werden, wenn das aktuelle Einzelbild keine oder zu wenige Merkmale enthält (z.B. weiße Wand) oder wenn aus den Vergleichen der Einzelbildes keine oder zu wenige gemeinsame Merkmale gefunden werden. In solchen Fällen kann das aktuelle Einzelbild bei der globalen Methode einfach übersprungen werden und die Verarbeitung mit dem nächsten Einzelbild fortgesetzt werden. Bei der lokalen Methode braucht man jedoch in diesem Fall

eine spezielle Lösung, weil keine weitere Sichtrichtungen der Kamera ermittelt und keine weitere Verarbeitungen durchgeführt werden können, wenn die Richtung für das aktuelle Einzelbild nicht gefunden werden konnte. Um dieses Problem zu lösen, kann die lokale Methode mit der globalen Methode kombiniert werden. Die globale Methode kann in solchen Fällen angewendet werden, bis die Kamerasichtrichtung für ein Einzelbild ermittelt werden konnte. Danach kann die lokale Methode angewendet werden. Ansonsten muss man die Kamera in eine bestimmte Position drehen und die Sichtrichtung manuell eingeben.

- Ein weiteres Problem ist, dass für das aktuelle Einzelbild eine falsche Kamerasichtrichtung ermittelt wurde. Um dies zu vermeiden, kann man festlegen, dass die Anzahl der gefundenen gemeinsamen Merkmale einen bestimmten, vorab definierten Mindestwert erreichen muss.

Weil bei der lokalen Methode die Kamerasichtrichtung des aktuellen Einzelbildes für die Berechnung der Kamerasichtrichtung des nächsten Einzelbildes verwendet wird, ist es bei der lokalen Methode besonders wichtig, die richtige Kamerasichtrichtung zu ermitteln. Ansonsten besteht die Gefahr, dass alle Kamerasichtrichtungen falsch berechnet werden können, wenn eine Kamerasichtrichtung falsch geschätzt ist.

- Ein weiterer Nachteil der lokalen Methode ist, dass die Kamerasichtrichtung nicht berechnet werden kann, wenn das aktuelle Einzelbild komplett außerhalb des vorherigen Einzelbildes liegt. Das heißt, zwei aufeinander folgende Einzelbilder haben keine gemeinsamen Bereiche und infolgedessen können keine gemeinsamen Merkmale gefunden werden. Da PTZ-Kameras in der Regel niedrige Bildwiederholfrequenzen haben und die Daten gelegentlich langsam übertragen werden, kann dieses Problem auftreten. Gelöst werden kann es wie oben beschrieben mit einer Kombination der lokalen und der globalen Methode.
- Bei der globalen Methode kann ein ähnliches Problem auftreten, wenn sich das aktuelle Einzelbild von seinem Hintergrundmodell komplett verändert hat, sodass das entsprechende Hintergrundmodell für das aktuelle Einzelbild nicht gefunden werden konnte. Dies ist zum Beispiel dann der Fall, wenn ein Vordergrundobjekt das Bild lückenlos verdeckt, sodass vom Hintergrund nichts mehr zu sehen ist. Für dieses Einzelbild kann dann kein Hintergrundmodell gefunden werden.

4. Erstellung des Hintergrundmodells

- Wenn das aktuelle Bild nicht scharf ist (z.B. wegen einer unpassenden Fokuseinstellung) oder Bewegungsunschärfe enthält, können Merkmale in dem Bild schlecht gefunden werden.
- Weil das gesamte Hintergrundmodell ein großes Bild darstellt, können aus dem Hintergrundmodell viele Merkmale gefunden werden. Dies kann sich bei der globalen Methode negativ auswirken, sodass für ein gefundenes Merkmal des aktuellen Einzelbildes mehrere Merkmale des Hintergrundmodells gefunden werden und sich viele falsche Paare bilden lassen. In der Regel wirken sich falsche Paare negativ auf die Schätzung von PTZ-Parametern aus. Ein weiterer Nachteil der globalen Methode ist, dass die Zuordnung der Merkmale sehr zeit- und berechnungsaufwendig im Vergleich zu der lokalen Methode ist.
- Bei der Anwendung der globalen Methode muss das Hintergrundmodell eine möglichst geringe Verzerrung, die durch Abbildung vom 3D Objekt auf 2D Hintergrundmodell entsteht, aufweisen. Im Falle einer starken Verzerrung des Hintergrundmodells können die Merkmale, wie oben erwähnt, nicht zugeordnet werden.

Abbildung 4.15 gibt einen Überblick der Schritte zur Berechnung der Kamerasichtichtung für das neue Einzelbild F_i .

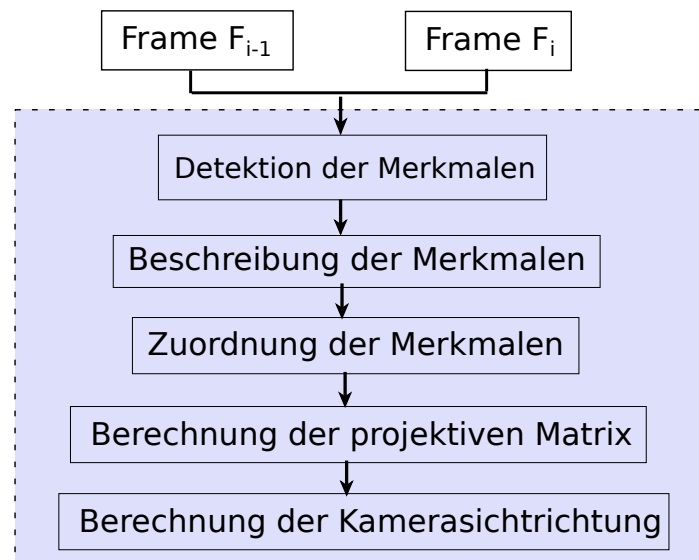


Abbildung 4.15.: Schritte zur Berechnung der Kamerasichtichtung

Der erste Schritt ist die Detektion der Merkmale. Nachdem ein Einzelbild mit den Matrizen, die bei Kamerakalibrierung erstellt worden sind, entzerrt wurde, werden mithilfe eines Detektors seine Merkmale ermittelt. Unter einem **Merkmal** eines Bildes versteht man die Position eines Punktes im Bild, der eine besonderen Eigenschaft aufweist und daher in einem anderen Bild wiedergefunden werden kann. Als nächstes werden die sogenannten Deskriptoren der ermittelten Merkmale berechnet. Die Deskriptoren beschreiben die Merkmale, sodass sie in einem anderen Bild verglichen bzw. identifiziert werden können. Sind die Merkmale zweier Einzelbilder durch Deskriptoren beschrieben worden, können die Keypoints der beiden Einzelbilder einander zugeordnet werden. Das heißt, die Deskriptoren der Merkmale werden miteinander verglichen und zwei Merkmale gelten als ein Paar, wenn deren Deskriptoren sich ähneln. Durch diesen Vorgang können mehrere Merkmale abfallen, wenn nur wenige Deskriptoren einander zugeordnet werden konnten. Nach dieser Zuordnung ist für die zugeordneten Merkmale bekannt, wo sie sich im jeweiligen Bild befinden. Abbildung 4.16 zeigt ein Beispiel für Zuordnung der Merkmale von zwei Bildern.



Abbildung 4.16.: Merkmale und deren Zuordnung

Als Detektor der Merkmale wurde die in OpenCV bereits implementierte Klasse `SURF()` zur Extraktion von SURF-Merkmalen [BTVGo6] verwendet. Bei dieser Funktion soll man einen Wert dem sogenannten Parameter `hessianThreshold` zuweisen, der bestimmt, wie ausgeprägt die Eigenschaften der zu ermittelnden Merkmale sein sollen. Abhängig von diesem Wert werden unterschiedlich viele Merkmale in einem Bild gefunden. In diesem Beispiel hat der Parameter den Wert 500 und es wurden 215 und 269 Merkmale bei den Bildern detektiert. Die ermittelten Merkmale werden dann durch den SURF-Deskriptor, der in OpenCV durch die Klasse `DescriptorExtractor::create("SURF")` implementiert ist, beschrieben. Für die Zuordnung der Merkmale sind in OpenCV verschiedene Verfahren (`FlannBased`, `BruteForce`,

4. Erstellung des Hintergrundmodells

BruteForce-L1, BruteForce-Hamming, BruteForce-HammingLUT) jeweils durch eine Klasse implementiert. Die Klasse `DescriptorMatcher::create("FlannBased")` ist bei diesem Beispiel für die Zuordnung verwendet und 215 Merkmale konnten einander zugeordnet werden.

Als nächstes wird die projektive Transformation für die beiden Einzelbilder mithilfe der zugeordneten Merkmale berechnet. Die projektive Transformation ermöglicht es, die relative Orientierung der Kamera zwischen den beiden Bildern zu bestimmen. Als Ergebnis dieser Berechnung erhält man eine 3×3 Matrix, mit der man die beiden Bilder übereinander legen kann. Die in OpenCV implementierte Funktion `findHomography()` wird dafür verwendet und als Ergebnis ist folgende Matrix für die beiden Bilder erhalten:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \begin{bmatrix} 0.866996 & -0.184487 & 357.228 \\ 0.146757 & 0.9473 & -28.3178 \\ -0.000155782 & 3.98289e - 06 & 1 \end{bmatrix},$$

Mit der berechneten projektiven Matrix kann man die beiden Bilder von der Abbildung 4.16 zusammensetzen. Für die Zusammensetzung wird die Funktion `warpPerspective()` verwendet. Abbildung 4.17 zeigt die Zusammensetzung der beiden Bilder. Um zu sehen, ob die Bilder zusammenpassen, werden sie transparent aufeinander gelegt. Um die neue Kamerasichtichtung bzw. den Drehwinkel aus den beiden Bildern zu berechnen, wird das aktuelle Bild durch die berechnete projektive Matrix nicht komplett transformiert, sondern nur der Mittelpunkt des Bildes. Das heißt, der Mittelpunkt des neuen Einzelbildes wird mit der projektiven Matrix multipliziert, um herauszufinden, wo er durch die Kameradrehung transformiert wurde. Da die Größe des Bildes 736×544 ist, sehen der Mittelpunkt P_{src} und der projektiv transformierte Punkt P_{dst} wie folgt aus:

$$P_{dst} = H \cdot P_{src} = H \cdot \begin{bmatrix} x_{src} \\ y_{src} \\ 1 \end{bmatrix} = H \cdot \begin{bmatrix} 368 \\ 277 \\ 1 \end{bmatrix} = \begin{bmatrix} 662.425 \\ 305.254 \\ 1 \end{bmatrix}$$

Damit erhält man die Position, wo der Mittelpunkt des neuen Einzelbildes im vorherigen Einzelbild ist. Aus der Kamerasichtichtung des vorherigen Einzelbildes kann man die Kugelkoordinaten (θ, ϕ) aller Punkte des Einzelbildes berechnen, wie die Punkte auf 3D Objekt projiziert werden. Das heißt, für den transformierten Mittelpunkt kann man auch die

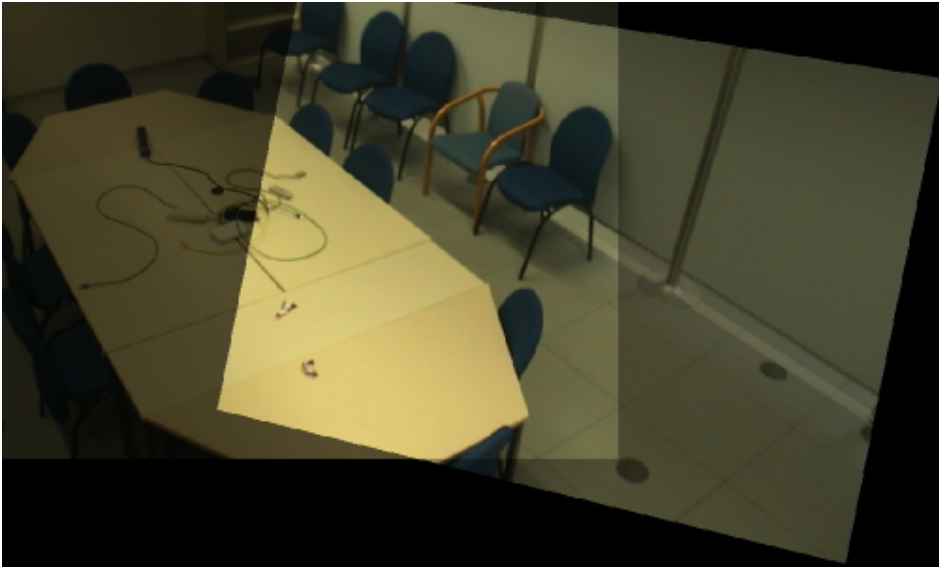


Abbildung 4.17.: Transparente Zusammensetzung von zwei Bildern

Kugelkoordinaten (θ, ϕ) berechnen und diese definieren dann die neue Kamerasichtichtung des aktuellen Einzelbildes.

Weil in dieser Arbeit die lokale Methode verwendet wurde, wurde mithilfe der geschätzten projektiven Transformation zwischen zwei aufeinander folgende Einzelbilder die Kamerasichtichtung des neuen Einzelbildes berechnet. Die Merkmale und die Beschreibungen des aktuellen Einzelbildes wurden gespeichert und bei der Berechnung des nächsten Einzelbildes erneut verwendet.

Bei der Verwendung der globalen Methode muss das Hintergrundmodell zwar auch registriert werden. Man benötigt jedoch nicht die projektive Transformation für die Berechnung des aktuellen Einzelbildes, weil das Hintergrundmodell und das aktuelle Einzelbild keine projektive Beziehung haben. Man muss hierfür einen Algorithmus entwickeln, der aus den gefundenen Merkmalen im Hintergrundmodell den Mittelpunkt des aktuellen Einzelbildes entsprechenden Punkt im Hintergrundmodell berechnet und abhängig von der verwendeten Kartenprojektion für diesen Punkt die Winkel (θ, ϕ) ermitteln kann.

Beseitigung der falsch zugeordneten Paare

Wenn man Abbildung 4.16 genau betrachtet, erkennt man, dass manche Merkmale nicht ihren entsprechenden Merkmalen in dem anderen Bild zugeordnet sind. Zurückzuführen ist dies darauf, dass manche Merkmale keine ausreichende Qualität bzw. nicht genügend ausgeprägte Kennzeichen haben. Dies führt dazu, dass sie ähnlich beschrieben und nicht eindeutig zugeordnet werden. Daher werden falsche Paare gebildet. Sie sind in diesem Bild dadurch erkennbar, dass die Linien, die die Paare verbinden, sich mit den Verbindungslinien anderer Paare mit großen Winkeln schneiden. Um die falschen Zuordnungen zu reduzieren, wird die Hough-Transformation angewendet.

Mit der Hough-Transformation können Geraden durch Punkte in einem dualen Raum, der durch zwei Parameter der Geraden aufgestellt ist, dargestellt werden (siehe Abschnitt 2.2.4). Für die Linien der Abbildung 4.16 wurde eine Hough-Transformation mit einem Dualraum erstellt. Als Parameter wurden der Winkel α und der Abstand d ausgewählt, wobei α der Winkel zwischen der Normale der Gerade und der horizontalen Achse ist, und d den Abstand vom Ursprung zum Lotfußpunkt auf der Gerade angibt. Der Ursprung liegt in der Mitte des Bildes und somit können die Geraden höchstens $\sqrt{(\frac{w}{2})^2 + (\frac{h}{2})^2}$ vom Ursprung entfernt sein, wobei w die Bildbreite und h die Bildhöhe angibt. Der berechnete Winkel einer Gerade kann zwischen 0° und 180° liegen. Abbildung 4.18 stellt diese Hough-Transformation dar.

Aus dem Bild sieht man, dass die meisten Geraden einen Winkel im Bereich zwischen 175° und 180° haben. Es gibt auch mehrere Geraden, deren Winkel unter 170° liegt. Geraden, die kleine Winkel haben, verbinden falsch zugeordnete Paare. Sie müssen entfernt werden, weil sie eine unkorrekte Schätzung der Kamerasichtrichtung verursachen können. Zur Entfernung dieser Geraden werden zunächst alle Winkel der Geraden bestimmt und anschließend wird der Winkel, der am häufigsten vorgekommen ist, gesucht. Zum Schluss werden Paare, deren Winkel von dem am häufigsten vorgekommenen Winkel deutlich abweichen, entfernt. Es gab bei dem Beispiel insgesamt 215 Paare und der Winkel 178° wurde 74-mal und somit am häufigsten gefunden. Abbildung 4.19 darstellt alle Winkel und deren Häufigkeiten.

Geraden, deren Winkel von diesem häufigst vorgekommenen Winkel um 10° abweichen, sind gelöscht. Es sind dann nur 186 Paare übrig geblieben. Abbildung 4.20 zeigt die Linien mit zwei unterschiedlichen Farben. Die rote Linien sind Linien, deren Paare durch diesen

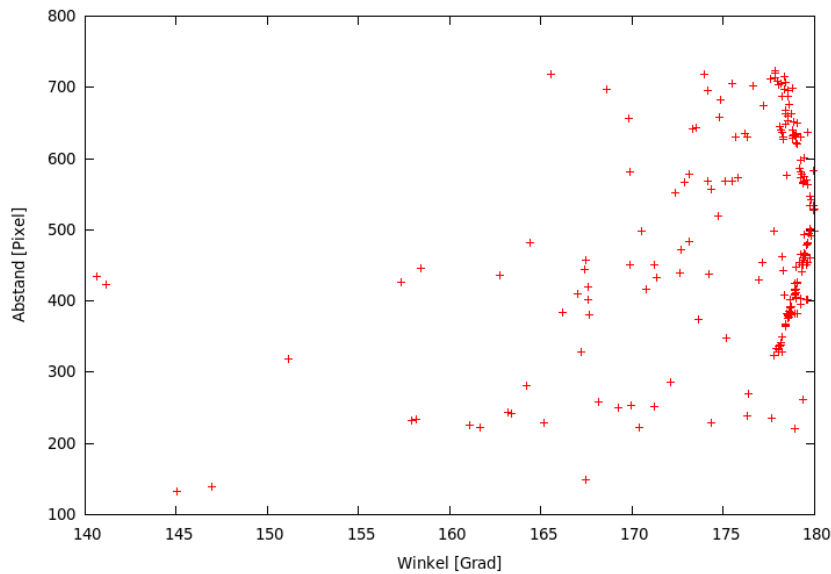


Abbildung 4.18.: Darstellung der Geraden durch eine Hough-Transformation im Dualraum Winkel und Abstand zum Bildmittelpunkt

Vorgang gelöscht sind. Paare die durch grünen Linien verbunden sind, sind übrig geblieben. Man sieht, dass viele der falschen Paaren gelöscht werden konnten.

Im allgemein können durch diesen Vorgang nicht alle falsche Paare entfernt werden und können die richtige Paare auch betroffen werden, weil die Kamera keine Verschiebungsbewegungen, sondern Drehbewegungen macht. Das heißt, die richtige Linien werden sich auch mit einander schneiden, wenn die Kamera sich um (genau oder ungefähr) ihre optische Achse dreht. Die PTZ-Kameras drehen sich um ihren optischen Achse, wenn sie Schwenkbewegungen bei dem Neigungswinkel $\theta = 0^\circ$ oder $\theta = 180^\circ$ macht. Wenn die Kamera genau nach unten oder nach oben schaut und sich um vertikalen Achse dreht, werden Bilder aufgenommen, die durch nur Rotationstranformation zusammengesetzt werden können. In diesem Fall können auch die richtige Linien verschiedene Winkel haben und sich mit einander schneiden. Abbildung 4.21 und 4.22 zeigen einen Beispiel dafür. Zwei Bilder, die bei solchen Bewegungen der Kamera aufgenommen sind, werden durch eine Rotation zusammengesetzt. Die Bilder werden neben einander gestellt und deren zugeordneten Keypoints sind durch Linien verbunden. Man sieht, dass auch einige Linien der korrekten Paare sich schneiden.

4. Erstellung des Hintergrundmodells

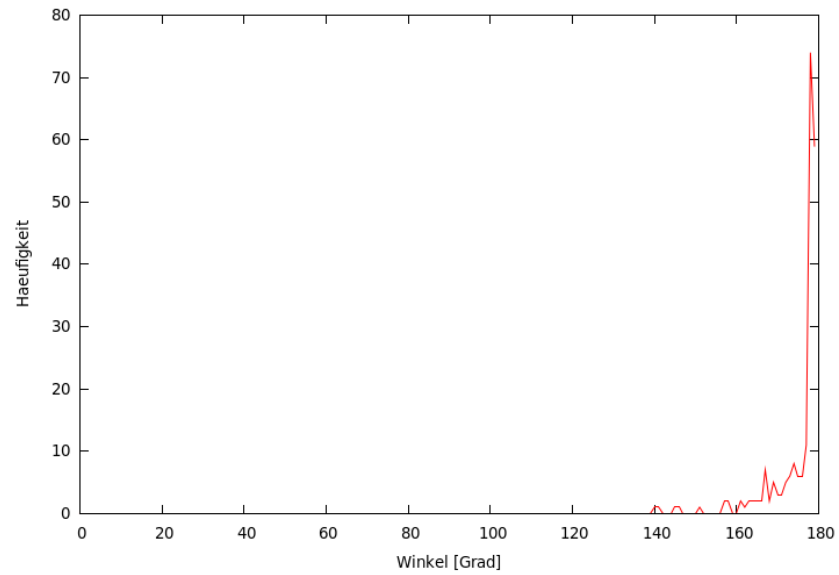


Abbildung 4.19.: Winkel der Linien und deren Häufigkeiten. Meiste Linien haben einen Winkel von 178° .

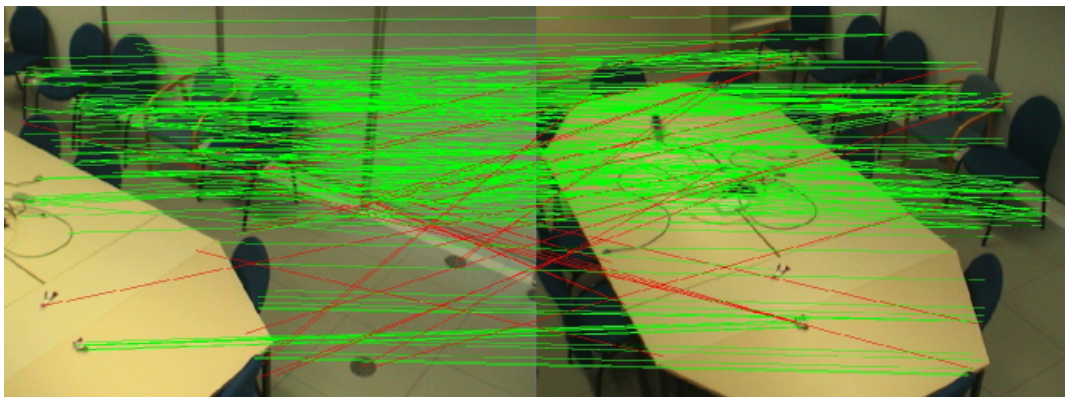


Abbildung 4.20.: eliminierte Paare sind durch rote Linien verbunden.

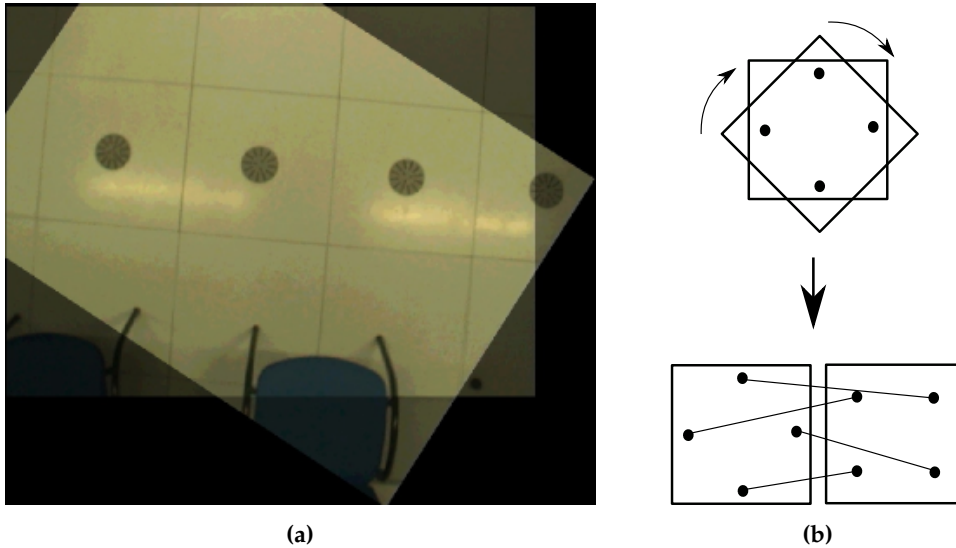


Abbildung 4.21.: Drehung der Kamera um ihre optische Achse (Kamera schaut nach unten und schwenkt)

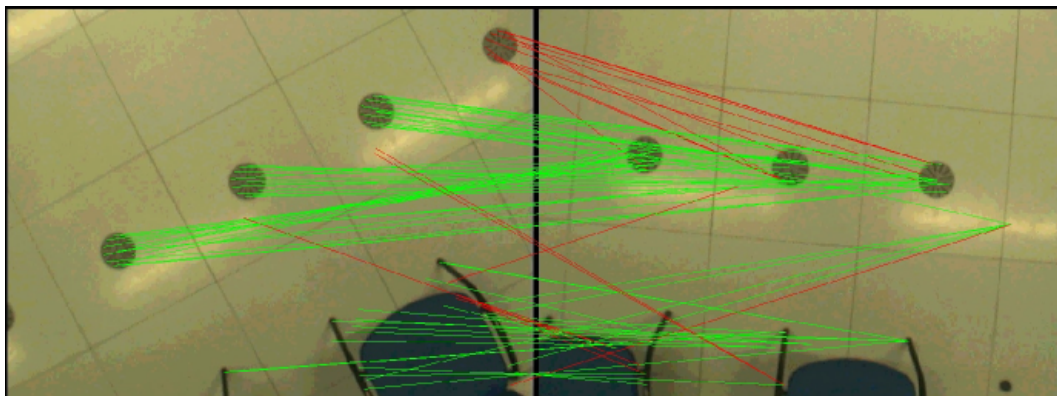


Abbildung 4.22.: Linien, die richtige Paare verbinden, haben verschiedene Winkel und schneiden sich daher auch.

4.4. Berechnung der Zoomstufe

Der dritte Freiheitsgrad der PTZ-Kamera ist die Zoom-Funktion. Eine Änderung des Zoomparameters muss für eine korrekten Abbildung zwischen dem aktuellen Einzelbild und dem Hintergrundmodell erkannt und entsprechend bei den Abbildungen beachtet werden, weil ansonsten ein Einzelbild auf einem zu grossen Bereich im Hintergrundmodell (zu klein bei dem umgekehrten Fall) aktualisiert werden kann.

Über den Zoom wird die Brennweite bzw. Bildwinkel der Kamera auch verändert. Aus den Berechnungen soll der neue Bildwinkel als Ergebnis kommen, weil Abbildung eines Einzelbildes auf das Hintergrundmodell mithilfe des Bildwinkels entsprechend angepasst werden soll. Der Wert der neuen Brennweite ist aber auch wichtig, weil beim Zoomen die Kameramatrix auch aktualisiert werden soll, wenn die Kamerabilder starke Verzeichnungen enthalten und sie mit der Kameramatrix, die durch Kamerakalibrierung erstellt wurde, entzerrt werden sollen. Wird der Zoomfaktor aus den aufgenommenen Bildern geschätzt, ist es für eine korrekten Schätzung wichtig, dass die Zuordnungen der erkannten Merkmale richtige Paare bilden. Das heißt, die Merkmale müssen starke Eigenschaften haben, dass sie eindeutig identifizierbar beschrieben werden können.

Im Folgenden wird ein Ansatz beschrieben, wie man eine Änderung des Zooms erkennen und den neuen Zoomfaktor schätzen kann. Der Ansatz konnte aber aus zeitlichen Gründen in dem Verfahren (bei den Abbildungen zwischen Einzelbild und Hintergrundmodell) nicht integriert werden, sondern nur einzeln getestet.

Bei einem Zoomen unterscheidet man zwischen dem optischen und dem digitalen Zoom. Digitale Zoom bedeutet, dass die Kamera einen Abschnitt des Bildes liefert, der durch eine Skalierung vom aufgenommenen Bild einfach vergrößert oder verkleinert ist. Die Qualität des Bildes wird damit aber auch verschlechtert. Bei dem optischen Zoom ändert sich dagegen der tatsächliche Bildwinkel und die Brennweite der Kamera und die Qualität des Bildes bleibt erhalten. Für die Berechnung der Zoomstufe ist es aber unwichtig, ob die Kamera optische oder digitale Zoom ausführt. Man muss in beiden Fällen einen neuen Bildwinkel berechnen.

Wenn man annimmt, dass das Bild in der Mitte gezoomt wird und die Kamerasichtichtung beim Zoomen gleich geblieben ist, können zwei aufeinander folgende Einzelbilder mit unterschiedlichen Zoomstufen, wie in Abbildung 4.23 gezeigt, dargestellt werden. Bei Aufnahme vom letzten Einzelbild $Frame_{i-1}$ waren die Bildweite der Kamera b_1 und der

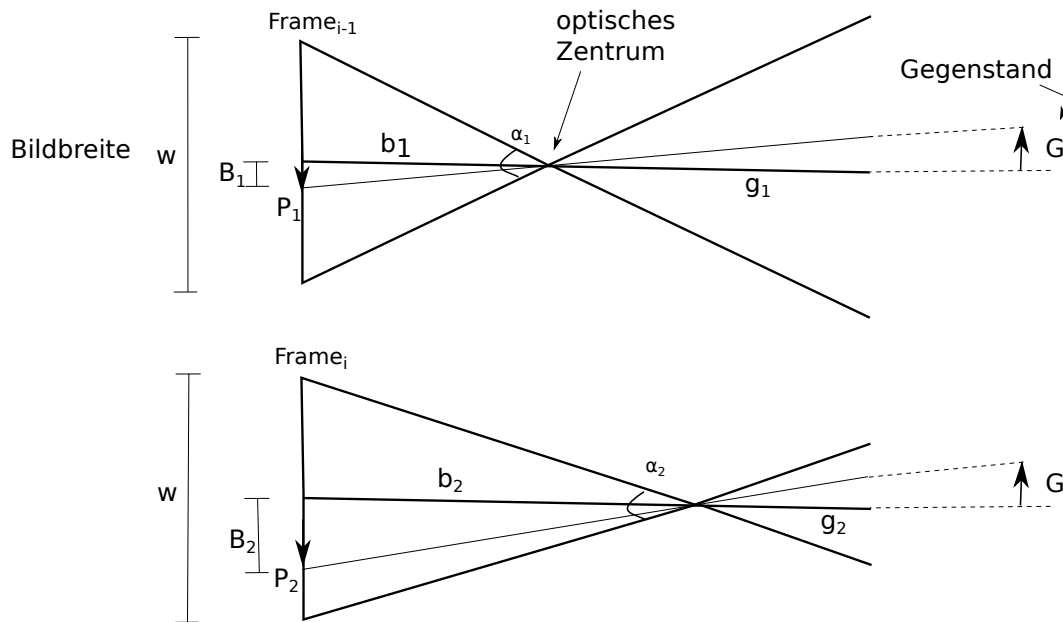


Abbildung 4.23.: aktuelles und vorheriges Einzelbilder mit unterschiedlichen Zoomstufen. Bildweite und Bildwinkel ändern sich beim Zoomen mit. Gegenstand G hat in beiden Einzelbilder unterschiedliche Koordinaten (P_1, P_2) .

Bildwinkel α_1 . Die beide Größen sollten bekannt sein, weil sie für das erste Einzelbild aus der Kameramatrix am Anfang initialisiert sind und bei Zoom für das entsprechenden Einzelbild immer neu berechnet werden. Ein Gegenstand G mit Gegenstandsweite g sei im Einzelbild $Frame_{i-1}$ als ein Merkmal mit B_1 Abstand von Bildmitte erkannt worden. Wir können einen Wert für die Gegenstandsweite z.B. $g_1 = 50 \cdot b_1$ annehmen. Das Einzelbild $Frame_{i-1}$ wird eingezoomt und die Kamera habe nun das aktuelle Einzelbild $Frame_i$ mit der Bildweite b_2 und mit dem Bildwinkel α_2 aufgenommen. Der Gegenstand G sei auch im aktuellen Einzelbild gefunden und habe einen Abstand B_2 von der Bildmitte. Man kann dann folgende Gleichungen schreiben:

$$\left\{ \begin{array}{l} G = B_1 \cdot \frac{g_1}{b_1} \\ G = B_2 \cdot \frac{g_2}{b_2} \\ b_1 + g_1 = b_2 + g_2 \end{array} \right.$$

4. Erstellung des Hintergrundmodells

Aus der ersten Gleichung kann man G berechnen, weil $B_1 = \sqrt{(x_1 - 0,5 \cdot w)^2 + (y_1 - 0,5 \cdot h)^2}$, $b_1 = 0,5 \cdot w / \tan(0,5 \cdot \alpha_1)$ und $g_1 = 50 \cdot b_1$ sind. Dann kann man aus den zweiten und dritten Gleichungen b_2 bestimmen:

$$b_2 = \frac{b_1 + g_1}{1 + \frac{G}{B_2}}$$

wobei $B_2 = \sqrt{(x_2 - 0,5 \cdot w)^2 + (y_2 - 0,5 \cdot h)^2}$. Zum Schluss kann man den neuen Bildwinkel $\alpha_2 = 2 \cdot \arctan(\frac{0,5 \cdot w}{b_2})$ berechnen. Das heißt, man kann für jedes gefundenen Paar von den letzten zwei Einzelbildern einen Bildwinkel berechnen.

Abbildung 4.24 zeigt zwei Einzelbilder. Das erste (linke) Bild stellt das eingezoomte Einzelbild vom rechten dar. Man sieht aus der Abbildung, dass mehrere falsch zugeordnete Paare durch Eliminierung mit Hough-Transformation nicht entfernt sind. In diesem Fall sollten die zugeordneten Paare verbindenden Linien zwar nicht parallel zu einander sein, aber sie sollten sich nicht schneiden, weil es bei den Einzelbildern angenommen ist, dass es nur gezoomt wird. Der Bildwinkel war bei dem rechten Einzelbild $41,5^\circ$ und gesucht ist der Bildwinkel für das eingezoomte Einzelbild. Die oben erläuterte Berechnungen sind bei diesen zwei Einzelbildern ausgeführt. Abbildung 4.25 zeigt den berechneten Bildwinkel für jedes Paars vom zweiten Einzelbild. Die berechnete Bildwinkel haben unterschiedliche

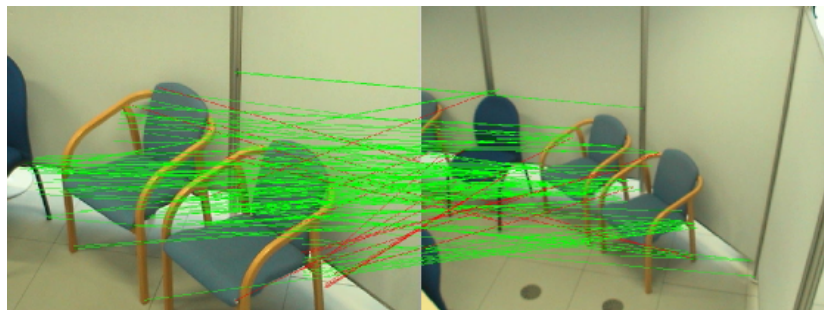


Abbildung 4.24.: Das rechte Bild (mit Bildwinkel $41,5^\circ$) wird im linken eingezoomt. Durch rote Linien verbundene Paare sind als falsche Paare erkannt und eliminiert.

Werte und sind sehr schwanklich. Der Grund hierfür ist, dass manche Paare nicht richtig bzw. aus nicht zusammen gehörenden Merkmalen gebildet sind. Die verwendete Annahmen können auch eine mögliche Ursache dafür sein. Insgesamt für 120 Paare sind Bildwinkel berechnet. Der Mittelwert liegt bei $57,5^\circ$ und der Medianwert bei $63,7^\circ$. Zur Beseitigung der falschen Paare können die beiden Bilder anderes positioniert werden, als neben einander zu stellen. In Abbildung 4.26a sind die zugeordnete Paare auf einem Bild durch Linien

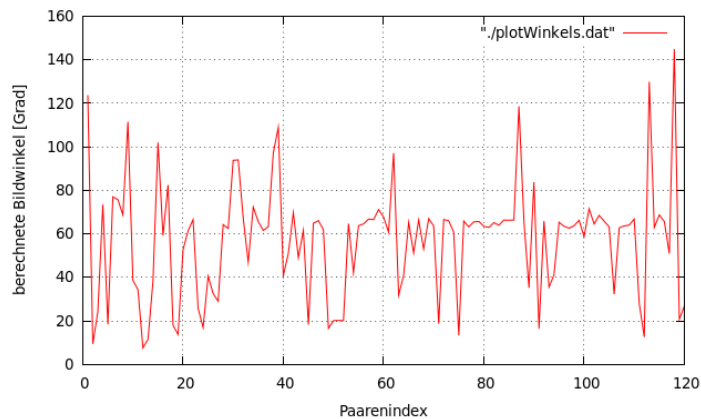


Abbildung 4.25.: Für gefundene 120 Paare jeweils ist der neue Bildwinkel berechnet. Durch falsche Paare verursachte Schwankungen sind sichtbar.

verbunden. Hier sollten sich die Linien eigentlich mit einander nicht schneiden. Sie müssten wie Lichtstrahlen von innen nach aussen gerichtete sein, wenn passende Merkmale einander zugeordnet gewesen wären. Aus dem Bild ist erkennbar, dass viele falsche Paare einander zugeordnet sind.

Für solche Darstellung, dass die Linien wie Lichtstrahlen verlaufen, soll die Hough-Transformation mit anderen Eigenschaften der Linien verwendet werden, um die falsche Paare zu beseitigen. Alternativ sollte hierfür ein anderen Verfahren, das die Lichtstrahlen besser erkennt, zur Eliminierung der falschen Paare verwendet werden. Um ein anderen Verfahren verwenden zu können, muss man für zwei Einzelbilder erkennen können, dass Zoom-Funktion stattgefunden ist, damit zur Beseitigung der falschen Paare nicht die Hough-Transformation, sondern das Verfahren bei Zoom verwendet wird.

Eine Möglichkeit die falsche Paare zu eliminieren ist, dass man bei allen Paaren die Quotient $\frac{B_1}{B_2}$ berechnet und die Paare, deren Quotient größer als 1 sind, eliminiert, wenn meisten Quotienten weniger als 1 sind oder im anderen Fall umgekehrt. Als Beispiel wurden die 120 Paare, die nach Hough-Transformation aus den obigen zwei Einzelbildern geblieben sind, nach diesem Ansatz geprüft und 34 Paare konnten eliminiert werden. Sowohl die gebliebene als auch die eliminierte Paare sind in Abbildung 4.26b durch rote und grüne Linien gezeichnet. Man sieht, dass hierdurch auch nicht alle falsche Paare eliminiert werden. Abbildung 4.27 zeigt eine transparente Zusammensetzung beider Einzelbilder. Die projektive Transformation wurde für die Zusammensetzung aus den gefundenen Paaren erstellt. Auf

4. Erstellung des Hintergrundmodells

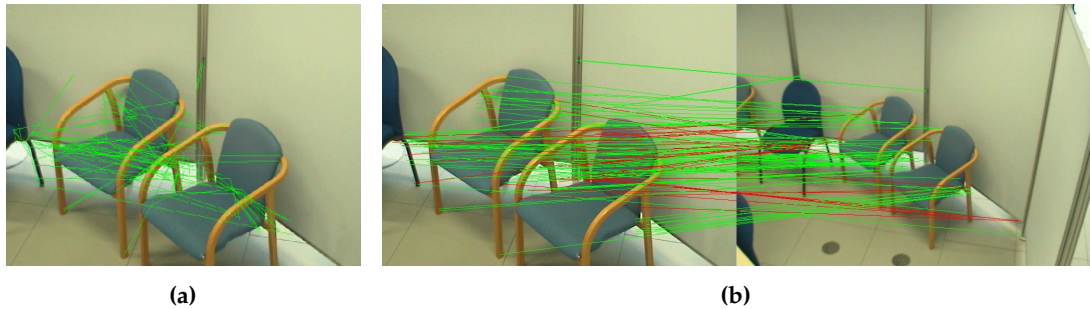


Abbildung 4.26.: (a) Paare sind durch Linien in einem Bild verbunden, (b) Eliminierung der falschen Paare mit der Quotient $\frac{B_1}{B_2}$

dem Bild sieht auch die Wirkung der falschen Paare, dass das eingezoomte Einzelbild im zusammengesetzten Bild kein Rechteck darstellt.

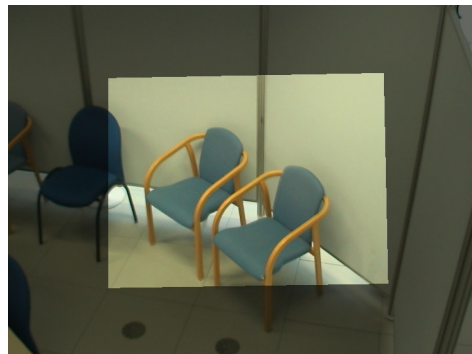


Abbildung 4.27.: Zusammensetzung mit projektiven Transformation. Das eingezoomte Bild ist nicht rechteckig wegen der falsch zugeordneten Paare.

Die projektive Transformation zwischen dem aktuellen und dem letzten Einzelbild gibt auch die Möglichkeit, den neuen Zoomfaktor und den neuen Bildwinkel zu schätzen. Für die Berechnung der projektiven Transformation kann die Funktion `findHomography()` aus OpenCV verwendet werden. Wie schon oben erwähnt, wurden für die zwei Einzelbilder die

projektive Transformation H mithilfe der zugeordneten Paare berechnet. Man kann mit dieser projektiven Transformation die vier Eckpunkten des letzten Einzelbildes transformieren.

$$\begin{cases} P'_1 = H \cdot P_1(0,0) \\ P'_2 = H \cdot P_2(w,0) \\ P'_3 = H \cdot P_3(0,h) \\ P'_4 = H \cdot P_4(w,h) \end{cases}$$

Dann wurden für die Paare (P'_1, P_1) , (P'_2, P_2) , (P'_3, P_3) und (P'_4, P_4) jeweils der neue Bildwinkel, wie oben beschrieben, berechnet. Als Ergebnis sind folgende Werte bekommen: $\alpha_2 = 65,9^\circ, 64,3^\circ, 64,6^\circ$ und $63,8^\circ$. Man sieht, dass die Ergebnisse hier sich nicht so viel unterscheiden, wie bei letzter Berechnung.

Auf diese Weise kann man außerdem auch erkennen, ob im aktuellen Einzelbild im Vergleich zu dem letzten Einzelbild gezoomt ist. Es ist im Allgemein wichtig erkennen zu können, ob Zoomparameter bei dem aktuellen Einzelbild geändert ist. Man kann z.B. dann den neuen Zoomfaktor besser schätzen, indem man die falsch zugeordneten Paare mit einem anderen Verfahren eliminiert. Für die Erkennung kann man die vier Eckpunkte mit der berechneten Homographie-Matrix H , wie oben beschrieben, transformieren. Wenn die Zoomstufe beider Einzelbilder sich unterscheiden, sollen die transformierte vier Punkte entweder alle vier innerhalb oder alle vier ausserhalb des Bildes sein.

Hier sei nochmal erwähnt, dass es angenommen ist, dass bei Zoomen die andere zwei Parameter der PTZ-Kamera sich nicht ändern. Wenn neben Zoomparameter gleichzeitig Schwenke- oder/und Neigeparameter sich zwischen zwei Einzelbilder auch geändert sind, braucht man eine andere Methode den neuen Zoomfaktor zu berechnen. In [SD10] und [SP06] sind verschiedene Methoden beschrieben, wie man Zoomfaktor schätzen kann.

Zum Vergleich sind andere Einzelbilder, bei denen mehrere SURF-Merkmalen gefunden sind, in folgenden Abbildungen 4.28 dargestellt. Die Bilder sind mittels Homographie-Matrizen transparent über einander gelegt. Man sieht, die projizierte Bilder ein Rechteck bilden. Das zeigt, dass die Zuordnung der Paare besser war als zuvor. Bei diesen Bildern hätten die obere Berechnungen des Bildwinkels möglicherweise nicht so viele unterschiedliche Resultate ergeben. Die Berechnungen konnten für diese Bilder nicht durchgeführt werden, weil der Bildwinkel vom vorherigen Einzelbild (der initiale Bildwinkel, der in dem obigen Beispiel mithilfe der Kamerakalibrierung geschätzt wurde) für die Berechnung benötigt wird. Dieser Winkel konnte aus der Kamera, mit der die Bilder erstellt sind, nicht bekommen werden. Für

4. Erstellung des Hintergrundmodells



Abbildung 4.28.: Zusammensetzung der Bilder mit vielen erkannten Merkmalen

die Berechnung des neuen Zoomfaktors ist es wichtig, mit passenden Merkmalen Paare zu bilden bzw. mit einem geeigneten Verfahren die falsch zugeordnete Paare zu eliminieren. Um ein anderes Verfahren zur Eliminierung zu verwenden, braucht man Methode, die Änderung des Zoomfaktors bei dem aktuellen Einzelbild zu erkennen.

Die Wahrscheinlichkeit, dass ein falscher Winkel für das aktuellen Einzelbild geschätzt wird, ist nicht wenig, weil die Änderung des Zoomfaktors zwischen zwei Einzelbildern klein sein kann und durch Zuordnung der Merkmale falsche Paare gebildet werden. Deshalb ist es wichtig, eine Methode zur Überprüfung zu finden, ob die berechnete Bildwinkel richtig ist bzw. den Fehler der berechneten neuen Bildwinkel zu überprüfen.

Beim Einzoomen kann das aktuelle Einzelbild zu groß für den entsprechenden Bereich im Hintergrundmodell sein, wenn der neue Zoomfaktor größer als den Zoomfaktor ist, der für die Erstellung des Hintergrundmodells verwendet ist. In diesem Fall sollte das aktuelle Einzelbild verkleinert werden (siehe Abschnitt 4.2.3).

4.5. Sphärische Projektion

Wenn für ein Einzelbild seine PTZ-Parameter ermittelt sind, kann man es auf das Hintergrundmodell abbilden. In diesem Abschnitt wird erklärt, wie Einzelbilder durch eine sphärische Projektion auf das Hintergrundmodell abgebildet werden.

Man kann die Abbildung eines Einzelbildes auf das Hintergrundmodell in zwei Schritte teilen: Projektion des Einzelbildes auf die Oberfläche einer Kugel und Abbildung der Kugeloberfläche auf das Hintergrundmodell. Für den zweiten Schritt wurden verschiedene Kartenprojektionen erprobt. Der Grund hierfür ist es, dass für die Ermittlung der Kamerasichtichtung eine globale Methode vorgesehen war und die Einzelbilder durch die erprobten Kartenprojektionen verzerrt wurden. Dadurch waren registrierte Merkmale schwierig wieder zu finden. Durch die unterschiedlichen Projektionen wurde versucht, die Verzerrung der Einzelbilder auf das Hintergrundmodell zu vermindern, sodass die Merkmale des aktuellen Einzelbildes auch im Hintergrundmodell wieder gefunden werden können und somit die PTZ-Parameter geschätzt werden können.

4.5.1. Projektion eines Einzelbildes auf eine Kugel

Nachdem die Kamerasichtichtung für das aktuelle Einzelbild berechnet ist, wird das Einzelbild auf die Innenseite einer Kugel mit Radius R projiziert, wie in Abbildung 4.29 dargestellt. Das Einzelbild wird vom Projektionszentrum im Abstand R' verschoben, sodass die vier Ecken des Einzelbildes die Kugel berühren.

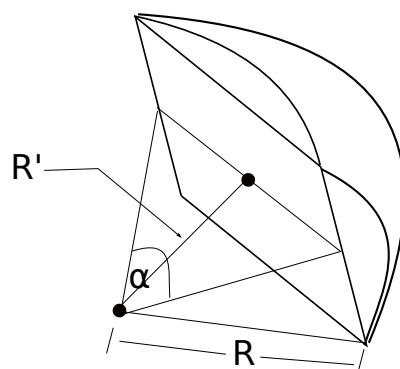


Abbildung 4.29.: Projektion eines Einzelbildes auf eine Kugel mit Radius R

4. Erstellung des Hintergrundmodells

Der Radius R der Sphäre wird aus dem geschätzten horizontalen Bildwinkel wie folgt berechnet:

$$R = \sqrt{\left(\frac{w}{2 \cdot \sin\left(\frac{\alpha}{2}\right)}\right)^2 + \left(\frac{h}{2}\right)^2}$$

wobei w die Bildbreite und h die Bildhöhe sind. Der Abstand R' vom Zentrum zur Bildmitte folgt aus:

$$R' = R \cdot \cos\left(\arcsin\left(\frac{\sqrt{w^2 + h^2}}{2 \cdot R}\right)\right)$$

Um das Bild auf die Sphäre zu projizieren, werden alle Pixel des Einzelbildes nacheinander mithilfe der ermittelten Kamerasichtrichtung (θ, ϕ) transformiert. Abbildung 4.30 zeigt den allgemeinen Vorgang dieser Transformation.

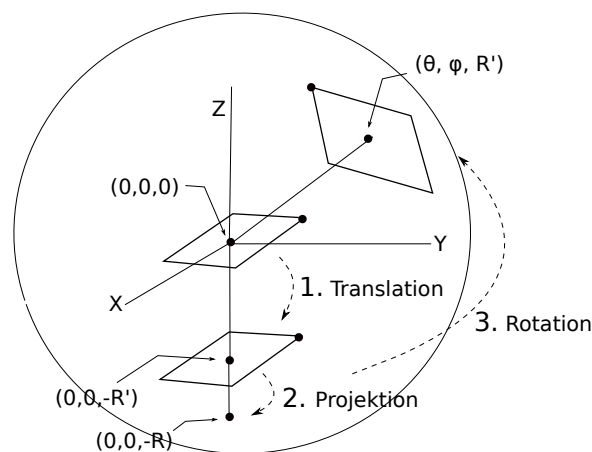


Abbildung 4.30.: Abbildung eines Einzelbildes auf Sphäre

Für diese Transformation müssen alle Pixel des Einzelbildes erst kartesische 3D Koordinaten haben. Das heißt, die Bildkoordinaten jedes Pixels müssen in den dreidimensionalen kartesischen Raum transformiert werden. Dafür wird das Einzelbild erst um R' in Richtung $-Z$ im dargestellten Koordinatensystem verschoben. Jedes Pixel des Einzelbildes bekommt dadurch eine dritte Koordinate $-R'$. Ein Pixel mit Bildkoordinaten $(x; y)$ bekommt dadurch die 3D Koordinaten $(x - \frac{w}{2}; -y + \frac{h}{2}; z)$. Als nächstes werden für alle Pixel die Koordinaten auf der Sphäre berechnet. Dafür wird für jedes Pixel eine Gerade gezogen, die vom Ursprung des Koordinatensystems zu dem berechnenden Pixel läuft und der Schnittpunkt dieser Gerade

mit der Sphäre wird berechnet. Abbildung 4.31 stellt die Projektion eines Pixels $P_1(x_1, y_1, z_1)$ auf den Punkt $P_2(x_2, y_2, z_2)$ der Sphäre dar. Da die Dreiecke $\Delta(OP_1P_4)$ und $\Delta(OP_2P_3)$ ähnlich

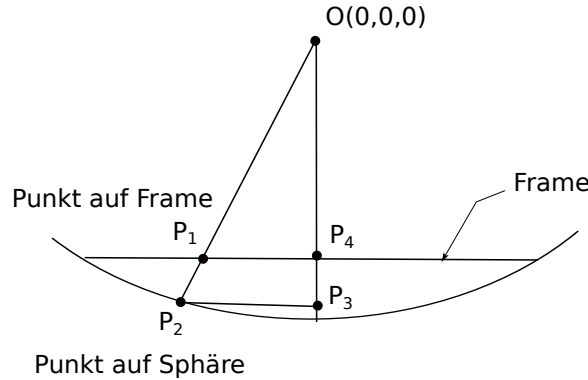


Abbildung 4.31.: Projektion eines Pixels auf die Sphäre

sind, ist die Z-Koordinate des Pixels auf der Sphäre gleich $z_2 = -\frac{R \cdot R'}{\sqrt{R^2 + x_1^2 + y_1^2}}$

Aus der Geradengleichung gilt für P_2 :

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \lambda \cdot \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$$

Da z_2 berechnet ist, gilt $z_1 = -R'$. Dadurch lässt sich λ mit $\lambda = \frac{z_2}{z_1}$ ermitteln. Somit kann man auch $x_2 = \lambda \cdot x_1$ und $y_2 = \lambda \cdot y_1$ bestimmen. Auf dieser Weise werden alle Pixel auf die Sphäre projiziert. Der nächste Schritt ist die Rotation aller Pixel um den Winkel (θ, ϕ) . Dafür kann eine Rotationsmatrix R erstellt und für alle Pixel verwendet werden:

$$R = R_\phi \cdot R_\theta = \begin{pmatrix} \cos(\phi') & -\sin(\phi') & 0 \\ \sin(\phi') & \cos(\phi') & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta') & -\sin(\theta') \\ 0 & \sin(\theta') & \cos(\theta') \end{pmatrix}$$

wobei $\phi' = \phi - \frac{\pi}{2}$ und $\theta' = \pi - \theta$ sind. R_θ und R_ϕ sind Rotationsmatrizen um die X- und Z-Achsen.

Bei der Rotation R spielt die Reihenfolge von R_θ und R_ϕ keine Rolle.

4.5.2. Größe des Hintergrundmodells

Die Größe des Hintergrundmodells hängt von der Länge des berechneten Radius' R der Sphäre ab. Die Breite w und die Höhe h des Hintergrundmodells werden wie folgt berechnet:

$$\begin{cases} w = 2\pi \cdot R \\ h = \pi \cdot R \end{cases}$$

Wenn die Größen berechnet sind, kann das Hintergrundmodell erstellt werden. Die im folgenden Abschnitt vorgestellte Kartenprojektionen haben alle die gleiche Größe. Es gibt aber auch Kartenprojektionen für Sphären, die Hintergrundmodelle anderer Größen benötigen (siehe Abschnitt 2.3).

4.5.3. Abbildungen der Sphäre auf das Hintergrundmodell

In den letzten Abschnitten wurden erklärt, wie alle Pixel eines Einzelbildes auf eine Sphäre projiziert und das Hintergrundmodell erstellt wird. Nun können die Pixel weiter auf das Hintergrundmodell abgebildet werden.

Transformation zwischen kartesischen und Kugelkoordinaten Für die Abbildung werden Kugelkoordinaten der Pixel benötigt und diese können aus deren kartesischen Koordinaten berechnet werden. Wie die Transformation zwischen den beiden Koordinatensystemen definiert ist, wird in Abschnitt 2.2.1 gezeigt. Bei der Implementierung muss man aber beachten, dass Rundungsfehler bei Berechnungen auftreten können. Das heißt, die Transformation kann nach den beschriebenen Formeln nicht ausgeführt werden. Der Winkel $\theta \in [0, \pi]$ wird nach $\theta = \arccos(\frac{z}{R})$ berechnet. Wenn das abzubildende Pixel in den Pol der Kugel projiziert würde ($\sin \theta = 0$), wird dem Winkel $\phi = \pi$ zugewiesen, ansonsten ($\phi \in [0, 2\pi]$) lautet die Berechnung für den Winkel $\phi = \arccos(\frac{x}{R \cdot \sin \theta})$. Weil für $\arccos()$ kein Wert im Bereich $]\pi, 2\pi]$ definiert ist, sollte man auch überprüfen, ob die y Koordinate des Pixels negativ ist. Ist dies der Fall, wird dem Winkel $\phi = 2\pi - \phi$ zugewiesen. Wegen Rundungsfehlern sind die Werte von y manchmal nicht Null sondern z.B. 10^{-8} . Deswegen wird bei der Überprüfung statt $(y < 0, 0)$ die Bedingung $(y < 10^{-4})$ verwendet.

Rektangulärprojektion

Bei der Rektangulärprojektion (eng. *Equirectangular projection*) ist das Hintergrundmodell vertikal im Bereich $[-\frac{\pi}{2}, \frac{\pi}{2}]$ und horizontal im Bereich $[-\pi, \pi]$ definiert. Abbildung 4.32 zeigt das Hintergrundmodell bei Rektangulärprojektion. Die Kugelkoordinaten (θ, ϕ) aller Pixel werden als kartesische Koordinaten auf das Hintergrundmodell entlang der horizontalen und vertikalen Achse abgebildet. Wurden für ein Pixel bei der Projektion auf die Sphäre Kugelkoordinaten (θ_1, ϕ_1) berechnet, wird das Pixel mit diesen Kugelkoordinaten auf das Hintergrundmodell abgebildet. Bei einem großen Winkel θ nimmt jedoch die Verzerrung

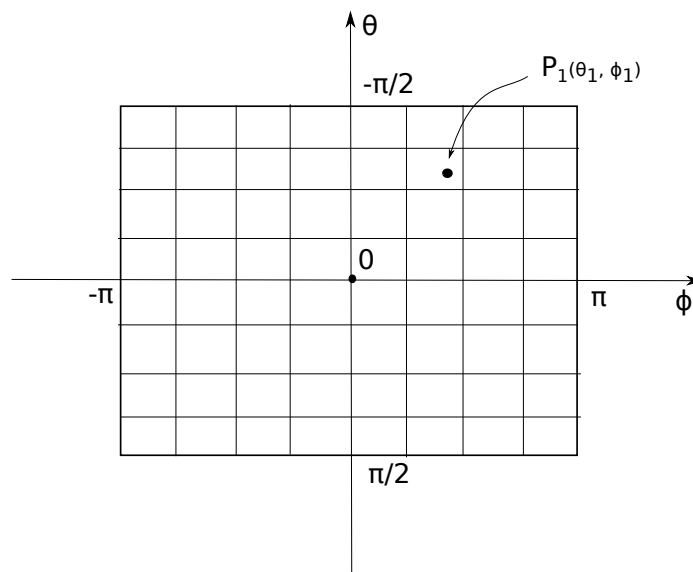


Abbildung 4.32.: Hintergrundmodell bei Rektangulärprojektion

durch diese Abbildung stark zu. Wenn z.B. der Winkel $\theta = \frac{\pi}{2}$ oder $\theta = -\frac{\pi}{2}$ ist, wird das Pixel entlang der gesamten horizontalen Achse gestreckt.

Für die Abbildungen wurde die Methode `remap()` von OpenCV verwendet. Man kann mit dieser Methode ein Bild *src* auf ein anderes Bild *dst* abbilden. Dafür müssen Abbildungen der X- und Y-Koordinaten für jedes Pixel des Zielbildes *dst* bestimmt werden:

$$dst(x, y) = src(map_x(x, y), map_y(x, y))$$

wobei $map_x(x, y)$ die Abbildung für die X-Achse und $map_y(x, y)$ die Abbildung für die Y-Achse definiert.

4. Erstellung des Hintergrundmodells

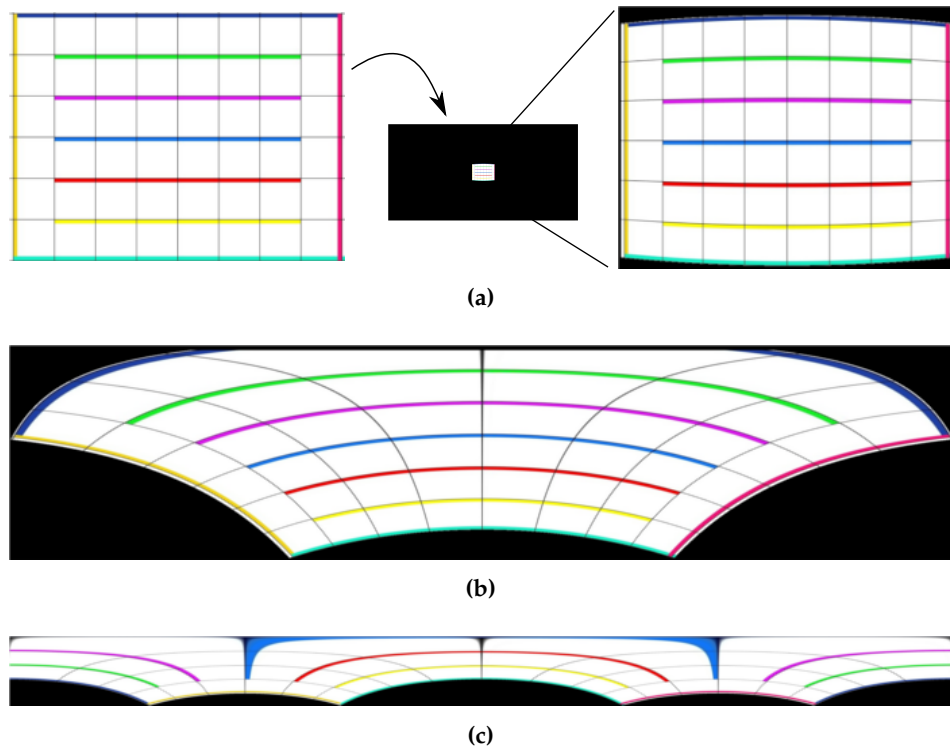


Abbildung 4.33.: Abbildungen eines Beispielmusters auf das Hintergrundmodell bei verschiedenen θ Winkeln: (a) $\theta = 0^\circ$, (b) $\theta = 70^\circ$, (c) $\theta = 90^\circ$

Das heißt, $map_x(x, y)$ gibt für ein Pixel $src(x, y)$ des Quellbildes an, welcher horizontale Wert das entsprechende Pixel des Zielbildes haben wird. Die Abbildung $map_y(x, y)$ gibt analog dazu den vertikalen Wert an. Die beiden Abbildungen $map_x()$ und $map_y()$ werden durch zwei $h_{dst} \times w_{dst}$ Matrizen definiert, wobei h_{dst} die Höhe des Zielbildes und w_{dst} die Breite des Zielbildes sind.

Das heißt, um ein aktuelles Einzelbild der Kamera auf das erstellte Hintergrundmodell abzubilden, müssen zwei $h_{HG} \times w_{HG}$ Matrizen erstellt werden, wobei (h_{HG}, w_{HG}) die Größe des Hintergrundmodells ist. Um die beiden Abbildungen zu beschreiben, muss als nächstes für jedes Pixel des Hintergrundmodells die entsprechenden Werte der beiden Abbildungen berechnet werden.

Diese Berechnung ist sehr rechenaufwendig, da das Hintergrundmodell sehr groß ist. Deswegen wurden zwei Algorithmen entwickelt, mit denen für das aktuelle Einzelbild die maximalen und minimalen Winkel von θ und von ϕ ermittelt werden können. Eine weitere

Überlegung wäre, durch die Umkehrabbildung, also vom Hintergrundmodell auf das Einzelbild, die beiden Matrizen zu berechnen. Diese Lösung führt jedoch zu dem Problem, dass man den Bereich des aktuellen Einzelbildes im Hintergrundmodell nicht vollständig berechnen kann, weil die Fläche des Hintergrundmodells größer ist als die Fläche der Sphäre. Sind für ein Einzelbild die entsprechenden Bereiche im Hintergrundmodell nicht vollständig gefüllt, kann man bei der Projektion eines Einzelbildes einer anderen Kamerasichtichtung auf leere Pixel im Hintergrundmodell treffen.

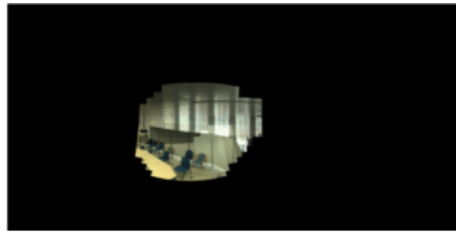


Abbildung 4.34.: Abbildungen mehrerer Einzelbilder

Um die Verzerrungen zu verringern wurden auch andere Kartenprojektionen erprobt.

4. Erstellung des Hintergrundmodells

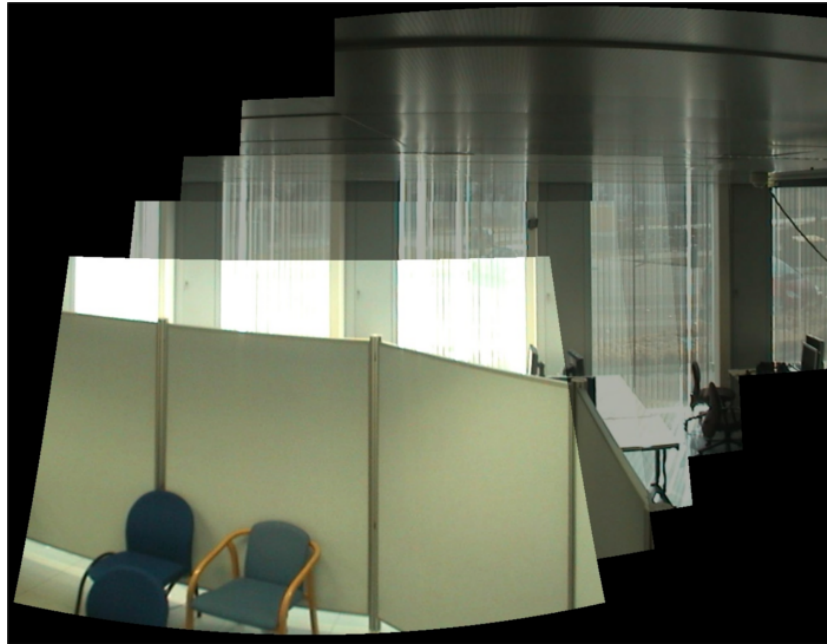


Abbildung 4.35.: Abbildungen mehrerer transparenter Einzelbilder

Sinusoid

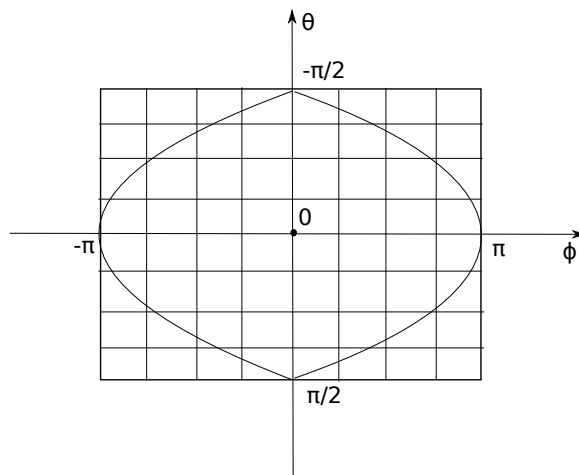


Abbildung 4.36.: Hintergrundmodell bei Sinusoid

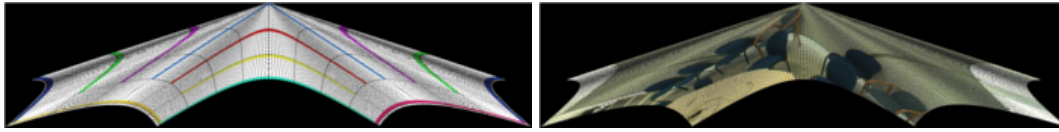


Abbildung 4.37.: Abbildungen mit der Kartenprojektion Sinusoid bei $\theta = 90^\circ$

Segment

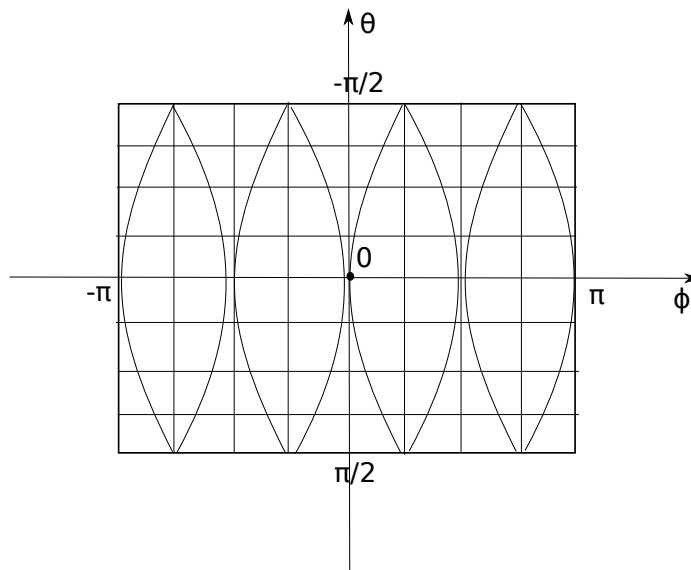


Abbildung 4.38.: Hintergrundmodell bei der Segment-Kartenprojektion



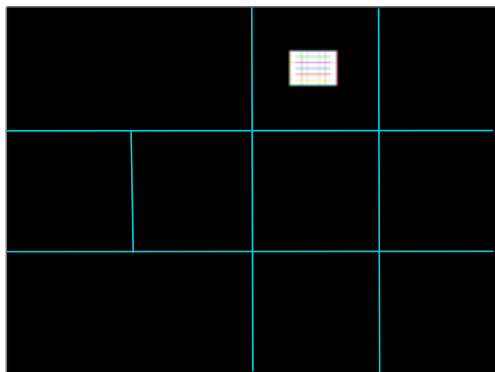
Abbildung 4.39.: Abbildungen mit der Kartenprojektion Segment bei $\theta = 90^\circ$

4.6. Kubische Projektion

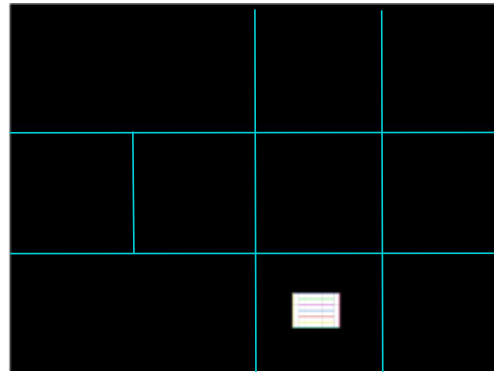
		4	
0	1	2	3
		5	

Abbildung 4.40.: Hintergrundmodell bei der kubischen Projektion

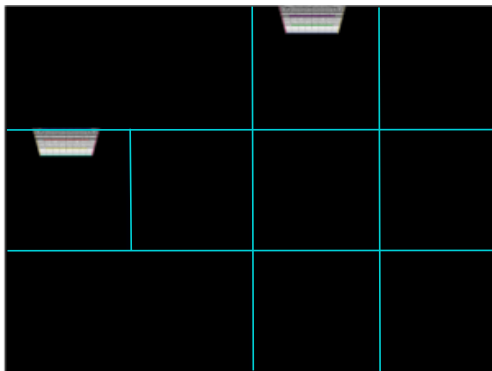
Im Vergleich zur sphärischen Projektion gibt es bei der kubischen Projektion keine Verzerrungen in den Bereichen der Pole.



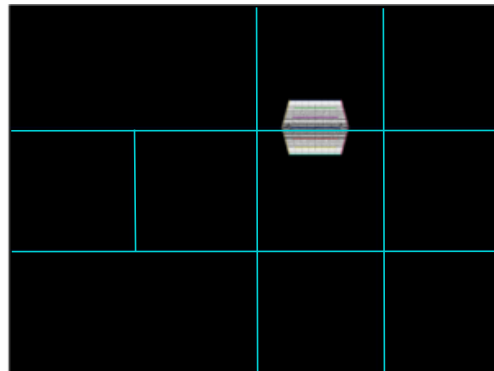
$\theta = -90^\circ, \phi = 0^\circ$



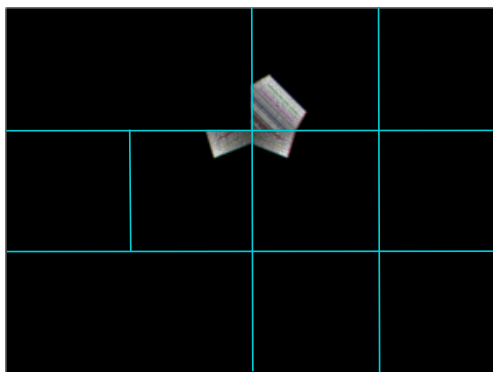
$\theta = 90^\circ, \phi = 0^\circ$



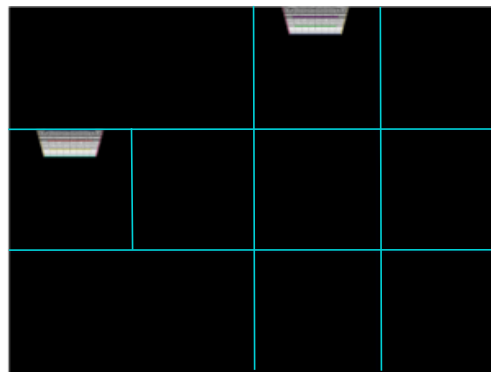
$\theta = -45^\circ, \phi = 180^\circ$



$\theta = -45^\circ, \phi = 0^\circ$



$\theta = -45^\circ, \phi = 45^\circ$



$\theta = -45^\circ, \phi = 180^\circ$

Abbildung 4.41.: Abbildungen der kubischen Projektion eines Musterbildes

4. Erstellung des Hintergrundmodells

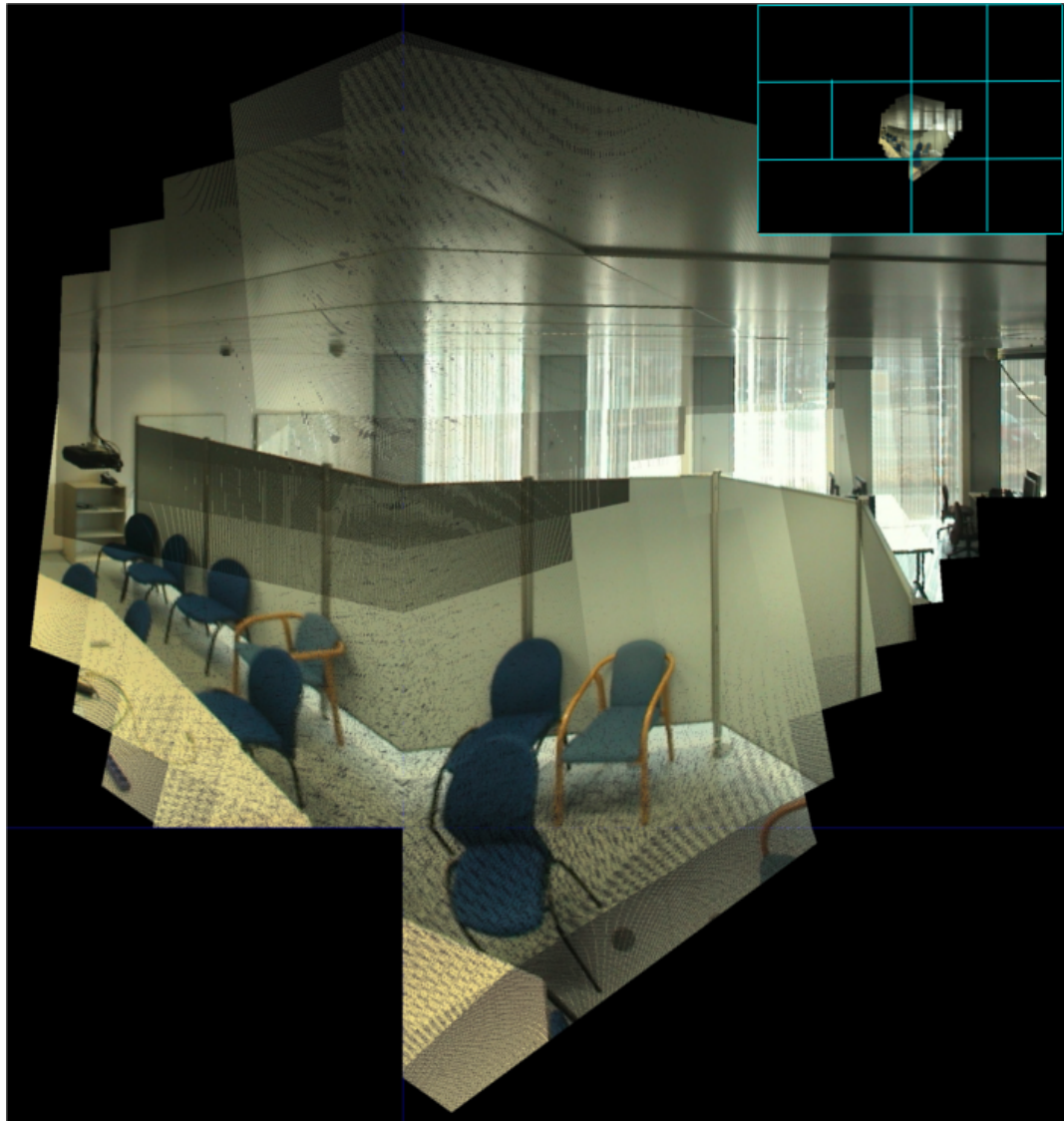


Abbildung 4.42.: Abbildungen bei der kubischen Projektion

5. Deblurring

Im Rahmen der vorliegenden Arbeit wird auf Bewegungsunschärfe im Detail eingegangen. Bewegungsunschärfe ist eine spezielle Art von Bildunschärfe, die durch die Bewegung der Kamera, des Objekts während der Belichtungszeit und durch Ein- und Auszoomen entsteht. In der Regel ist Bewegungsunschärfe unerwünscht.

Auf der Hardware-Ebene versucht man dem Problem mittels spezieller Sensoren und optisch-stabiler Linsen entgegenzuwirken. Auf der Software-Ebene gibt es eine Reihe von Verfahren, die durch Schätzung des Faltungskerns (eng. point spread function, PSF) das originale Bild wiederherstellen. Unter "Motion Deblurring" versteht man den Wiederherstellungsprozess eines Bildes, der aus dem unscharfen Bild mithilfe eines gegebenen Blurring-Kernels das ideale Bild wiederherstellt.

In der Bildverarbeitung wird ein so genanntes Bildentstehungsmodell verwendet, das den Prozess der Entstehung eines unscharfen Bildes aus einem idealen Bild mathematisch beschreibt. Dieses Modell wird in vielen Deblurring-Verfahren eingesetzt. Das Bildentstehungsmodell sieht aus wie folgt:

$$g(y, x) = h(y, x) \otimes f(y, x) + n(y, x)$$

wobei $f(y, x)$ das ideale Bild, $g(y, x)$ das unscharfe Bild, $h(y, x)$ die PSF bzw. der Faltungskern, $n(y, x)$ zusätzliche Störungsfunktion und \otimes Faltung ist.

Das ideale Bild wird bei der Aufnahme durch das Faltungskern $h(y, x)$ gefaltet, infolgedessen entsteht ein unscharfes Bild. Dies bedeutet, dass die Unschärfe hauptsächlich durch den Effekt von PSF ($h(y, x)$) entsteht. Die zusätzliche Störungsfunktion $n(y, x)$ stellt andere Faktoren, wie Sensorrauschen, dar, die zu dem gesamten Effekt der Unschärfe beiträgt. In Abbildung 5.1 ist zu sehen, wie ein Bild mittels PSF gefaltet und unscharf wird. Daher muss für die Wiederherstellung des idealen Bildes die Umkehrung der Faltungsoperation Entfaltung verwendet werden. Mithilfe der Entfaltung wird das ideale Bild aus dem unscharfen Bild wiederhergestellt. Die PSF ist notwendig zur Wiederherstellung des idealen Bildes.



Abbildung 5.1.: Links: Faltung des idealen Bildes mit PSF. Rechts: Entfaltung des unscharfen Bildes mit PSF [Wik11c]

5.1. Entfaltungsfilter

In diesem Abschnitt werden drei Entfaltungsfilter vorgestellt, die zum Verstehen des in der vorliegenden Arbeit vorgeschlagenen Verfahrens nötig sind. Unter dem Namen *Inverse-Filtering* bekannte Methode liegt den anderen beiden Methoden zugrunde. Daher wird sie zuerst beschrieben. Danach wird der Wiener-Filter und der Richardson-Lucy Algorithmus vorgestellt, die in dem vorgeschlagenen Lösungsansatz eingesetzt werden können.

5.1.1. Inverse-Filtering

Ein Bild $f(y, x)$ wird durch einen Faltungskern bzw. durch die PSF $h(y, x)$ gefaltet und das Ergebnis ist das Bild $g(y, x)$.

$$g(y, x) = h(y, x) \otimes f(y, x)$$

Aus dem Faltungssatz gilt dann:

$$G = H \cdot F$$

wobei G, H, F jeweils die Fourier-Transformierte von $g(y, x), h(y, x), f(y, x)$ sind. Die Fourier-Transformierte einer PSF nennt man eine OTF (eng. optical transfer function).

Wenn die PSF bekannt ist, kann man aus der Fourier-Transformierte das entfaltete (geschärfte) Bild wie folgt berechnen:

$$\tilde{f} = \mathcal{F}^{-1}(F) = \mathcal{F}^{-1}\left(\frac{G}{H}\right)$$

wobei \tilde{f} das wiederhergestellte Bild, \mathcal{F}^{-1} die inverse Fourier-Transformation ist.

Inverse-Filtering verwendet also die inverse Fourier-Transformation. Der Nachteil des *Inverse-Filtering* ist es, dass zusätzliche Störungen verstärkt werden, wenn die Fourier-Transformierte der PSF H klein ist. Dies lässt sich durch das Bildentstehungsmodell erklären. Verwenden man den Faltungssatz, so bekommt man:

$$G = H \cdot F + N$$

Wenn man diese Fourier-Transformierte G in obiges G einsetzt, ergibt sich für das wiederhergestellte Bild:

$$\tilde{f} = \mathcal{F}^{-1} \left(\frac{G}{H} \right) = \mathcal{F}^{-1} \left(\frac{H \cdot F + N}{H} \right) = \mathcal{F}^{-1} \left(F + \frac{N}{H} \right)$$

Aus dieser Gleichung wird ersichtlich, dass bei kleinem H die zusätzliche Störung N vergrößert wird. Da unscharfe Bilder meiste zusätzliche Störungen enthalten, wird *Inverse-Filtering* in der Regel nicht oft verwendet.

5.1.2. Wiener Filter

Der Wiener Filter wurde von Norbert Wiener vorgeschlagen. Das ideale Bild wird durch die folgenden Gleichung geschätzt:

$$\begin{aligned} \tilde{f} &= \mathcal{F}^{-1} \left(\frac{H^2}{H^2 + k} \cdot \frac{G}{H} \right) \\ &= \mathcal{F}^{-1} \left(\frac{H^2}{H^2 + k} \cdot \frac{H \cdot F + N}{H} \right) \end{aligned}$$

wobei k eine nicht negative Konstante ist. Die Konstante k ergibt sich aus dem Verhältnis des quadratischen Spektrums von Störung N und Bild F . Das heißt, durch Minimierung des quadratischen Fehlers $\min ||f(y, x) - \tilde{f}(y, x)||^2$ wird versucht, ein möglichst ähnliches Bild wie das ideale wiederherzustellen.

Bei einem großen Wert für k wird die Störung N unterdrückt. Bei $k = 0$ ist der Wiener Filter gleich wie *Inverse-Filtering*. Das heißt, *Inverse-Filtering* ist ein Spezialfall des Wiener Filters.

5.1.3. Richardson-Lucy Algorithmus

Der Richardson-Lucy Algorithmus ist ein iteratives Verfahren. Die Schätzung des idealen Bildes wird Schritt für Schritt verfeinert. Die Methode wird oft angewendet.

Bei der Methode bleibt sowohl die lokale als auch die globale Energie erhalten, d.h. das erstellte Bild oder ein Teil davon wird nicht heller oder dunkler.

Die Schätzung wird mit der folgenden Formel durchgeführt:

$$\begin{cases} \forall y, x : \tilde{f}_0(y, x) := c, & c \in \mathfrak{R} \setminus 0 \\ \tilde{f}_{n+1} = \tilde{f} \left(\check{h} \otimes \frac{g}{h \otimes \tilde{f}_n} \right), & n > 0 \end{cases}$$

wobei \check{h} ein 180° gedrehter Faltungskern ist.

Nach jeder Iteration wird überprüft, ob der Fehler klein genug ist. In diesem Fall wird das Verfahren nicht weiter durchgeführt. Eine große Anzahl von Iterationen bringt nicht unbedingt bessere Resultate, denn Störungen können bei weiteren Iterationen verstärkt werden.

Ein Nachteil der Methode ist der sogenannte "Ringing Effekt". Das sind ringähnliche Artefakte auf dem Bild, die abhängig von den Iterationsschritten auftauchen.

5.2. Eigenes Verfahren

Die oben beschriebene Entfaltungsmethoden benötigen die PSF. Die PSF kann aus dem optischen Fluss bestimmt werden, um die Bewegungsunschärfe der von der PTZ-Kamera aufgenommenen Bilder zu reduzieren. Unter dem optischen Fluss eines Bildes versteht man ein Vektorfeld, das für jedes Pixel des Bildes seine Bewegungsrichtung und seine Geschwindigkeit auf der Bildebene angibt.

Man kann z.B. die Methode `calcOpticalFlowFarneback()` der OpenCV-Bibliothek verwenden, um den optischen Fluss eines Bildes zu bestimmen. Die Methode nimmt zwei Einzelbilder als Eingabe und berechnet für jedes Pixel den optischen Fluss zwischen den Bildern. Abbildung 5.2 zeigt ein Beispiel, das den optischen Fluss eines Einzelbildes darstellt. Hier wurde der Wert des optischen Flusses an jedem 16-ten Pixel gezeichnet.

Die Methode verwendet intern das Farneback-Verfahren [Far03] zur Berechnung des optischen Flusses. Aus dem berechneten optischen Fluss kann man entnehmen, in welche Richtung und wie weit sich ein beliebiges Pixel bewegt hat.

Als nächstes kann man das Bild in Blöcke von bestimmten Größen (z.B. 32×32 Pixel) unterteilen und für alle Pixel jedes Blockes den durchschnittlichen optischen Fluss berechnen. Aus

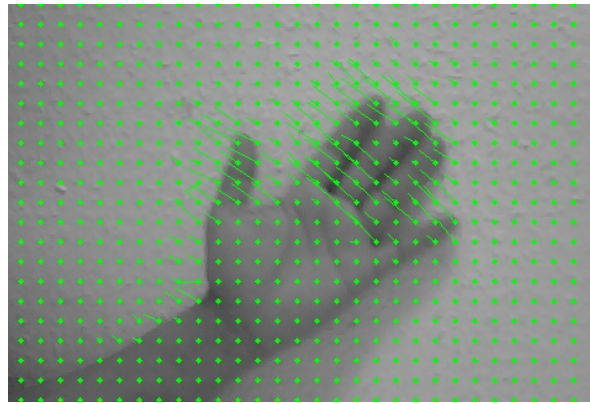


Abbildung 5.2.: Bestimmung des optischen Flusses

diesen Werten kann die PSF eines Blockes bestimmt werden. Die Unterteilung in Blöcken hilft unterschiedliche Bewegungen im Bild zu erfassen und gleichzeitig die Anfälligkeit gegenüber Rauschen zu vermindern. Mit Hilfe einer der oben beschriebenen Entfaltungsfiler (z. B. Wienerfilter) werden alle Blöcke dann entfaltet, wobei die Block-PSFs eingesetzt werden.

6. Zusammenfassung und Ausblick

Das Ziel der vorliegenden Arbeit war die Entwicklung eines Verfahrens zur Hintergrundsubtraktion mit PTZ-Kameras. Zu diesem Zweck wurde ein sphärisches Hintergrundmodell erstellt, das die Umgebung der PTZ-Kamera vollständig umfasst.

Die fehlerfreie Schätzung der PTZ-Parameter ist notwendig, damit das gesamte Bild korrekt auf das Hintergrundmodell projiziert wird.

Zur Ermittlung der Kamerasichtrichtung wurde die lokale Methode verwendet, die das aktuelle Einzelbild registriert und mit dem vorherigen Einzelbild vergleicht.

Ursprünglich war geplant, aus dem Vergleich des aktuellen Einzelbildes und dem gesamten Hintergrundmodell die Kamerasichtrichtung zu bestimmen. Dies führte zu unvorhergesehenen Problemen. Zum einen entstanden dadurch mehr falsche Paare als bei der lokalen Methode, weil zu viele Merkmale aus dem Hintergrundmodell gefunden wurden und das Zuordnen von Merkmalen fehleranfällig und rechenintensiv war.

Zum anderen konnte das Hintergrundmodell nicht verzerrungsfrei erstellt werden. Dies erschwerte das Wiederfinden von Merkmalen des aktuellen Einzelbildes im Hintergrundmodell. Um das Problem mit Verzerrung zu verringern, wurden verschiedene Kartenprojektionen getestet.

Mit kubischer Projektion konnten Einzelbilder mit den wenigsten Verzerrungen auf das Hintergrundmodell abgebildet werden. Die lokale Methode ist ihrerseits nicht problemlos (siehe Abschnitt 4.3).

Unter Berücksichtigung der Nachteile der lokalen Methode gelangt man zu dem Fazit, dass es besser ist, die beiden Methoden zu kombinieren.

Ein anderes wichtiges Thema bei der Schätzung von PTZ-Parametern war die Beseitigung von falschen Paaren. Das Vorhanden von vielen falschen Paaren führte dazu, dass Einzel-

bilder nicht auf den vorhergesehenen Positionen im Hintergrundmodell landeten. Bei der Erkennung von falschen Paaren hat sich die Hough-Transformation als effektiv erwiesen.

Das Zoomen ändert die Position des Nodalpunktes und verursacht somit eine Parallaxenverschiebung. Über die damit verbundenen Probleme kann man im Abschnitt 4.2.1 nachlesen. In dieser Arbeit wurde ein Verfahren zur Erkennung der Zoomfaktoränderung und Bestimmung des neuen Zoomfaktors vorgeschlagen (siehe Abschnitt 4.4). Dieses Verfahren konnte, aus zeitlichen Gründen, mit dem sphärischen Hintergrundmodell nicht getestet werden.

Damit eine Hintergrundsubtraktion mit dem sphärischen Hintergrundmodell durchgeführt werden kann, wurden zwei Abbildungen nötig. Eine bildet Einzelbilder mit ermittelten Kamerasichtrichtungen (Winkel) auf das Hintergrundmodell ab. Die zweite Abbildung ermöglicht einer Kamerasicht (Winkel oder Einzelbild) entsprechendes Bild aus dem ganzen Hintergrundmodell zu bekommen. Warum die zweite Abbildung mit Hilfe der ersten bzw. die erste mit Hilfe der zweiten realisiert werden kann, wird im Abschnitt 4.5.3 erklärt.

Die erste Abbildung ist sehr rechenintensiv, da im schlimmsten Fall für alle Pixel des Hintergrundmodells geprüft werden muss, ob sie zum aktuellen Einzelbild gehören oder nicht.

Um den Rechenaufwand zu verringern, wird aus dem ermittelten Kamerasichtwinkel ein Bereich im gesamten Hintergrundmodell berechnet $((\theta_{min}, \theta_{max}), (\phi_{min}, \phi_{max}))$ innerhalb dessen nach dem Hintergrundmodell gesucht wird, das dem aktuellen Einzelbild entspricht.

Zudem wurde ein lokales Deblurring-Verfahren vorgeschlagen, das den optischen Fluss von Einzelbildern zur Ermittlung der Degradationsfunktion verwendet (siehe Kapitel 5).

Eines der wichtigsten Ergebnisse der vorliegenden Arbeit ist, dass die PTZ-Parameter sich in Umgebungen (z.B. offene Bereiche mit weiter Sicht) genauer berechnen lassen, in denen Bildregistrierungsverfahren wie SURF viele Merkmale finden können.

Ausblick

In weiteren Arbeiten sollten künftig die Einsatzmöglichkeiten und das Potenzial des in dieser Arbeit vorgeschlagenen Deblurring-Verfahrens untersucht werden.

Um den Zeitaufwand des Hintergrundsubtraktionsverfahren in PTZ-Kameras zu verringern, könnte man die Möglichkeit untersuchen, einige aufwendige Algorithmen mittels CUDA auf Grafikkarten zu realisieren.

Zudem sollte man die Möglichkeit untersuchen das Hintergrundmodell nicht auf 2D zu projizieren sondern ein 3D Modell der Szene zu benutzen. Dadurch könnte der Rechenaufwand der Abbildungen zwischen aktuellem Einzelbild und dem Hintergrundmodell verringert und das Problem mit Verzerrungen durch Projektion auf 2D beseitigt werden.

A. Hintergrundsubtraktion

Um Vordergrundobjekte aus Videodaten zu segmentieren, wird die Technik der Hintergrundsubtraktion verwendet. Die Videodaten kann man sich als eine Bildsequenz vorstellen, die aus mehreren Bildern (Frames) besteht. Die Frames der Videodaten werden nacheinander bearbeitet und ihre Vordergrundobjekte und Hintergrundbereiche werden bei jedem Frame separat bestimmt. Zur Segmentierung der Vordergrundobjekte bei den Frames benötigt die Technik einen bereitgestellten Hintergrund. Die Grundidee dieser Technik ist es, durch den Vergleich des aktuell zu bearbeitenden Frames mit dem Hintergrund zu bestimmen, ob ein Pixel zum Vordergrund oder zum Hintergrund gehört. Ein einfaches Verfahren zur Hintergrundsubtraktion kann durch folgende Ungleichung beschrieben werden:

$$|Frame_i - Hintergrund| > Schwellenwert$$

Von jedem Pixel des aktuellen Frames $Frame_i$ wird ein entsprechendes Pixel im Hintergrundbild subtrahiert. Ein Pixel gehört zum Vordergrund, wenn die absolute Differenz über einem bestimmten Schwellenwert liegt. Somit wird jedes Pixel eines Frames entweder als Hintergrund- oder als Vordergrundpixel klassifiziert. Der Schwellenwert legt also fest, wie stark ein Pixel von seinem entsprechenden Pixel im Hintergrund abweichen darf, um als Vordergrund bzw. als Hintergrund bestimmt zu werden. Unter der Subtraktion zweier Pixel versteht man die Subtraktion deren Farbwerte. Graue Bilder lassen sich einfach von einander subtrahieren, weil deren Pixel als Farbwert einfache Skalare von Größe 0 bis 255 haben. Farbbilder haben dagegen Pixel, deren Farbwerte sich durch eine Kombination mehrerer Komponenten darstellen lassen. Daher gibt es mehrere Möglichkeiten, Farbbilder voneinander zu subtrahieren. Eine häufig verwendete Variante ist, die Farbbilder vor der Subtraktion in grauen umzuwandeln und sie dann von einander zu subtrahieren. Andere Varianten hierfür sind die verschiedenen komponentenweisen Subtraktionen. Als Beispiel seien zwei RGB-Farbpixel $P_1 = (R_1, G_1, B_1)$ und $P_2 = (R_2, G_2, B_2)$ gegeben. Dann könnte deren Subtraktion z. B. wie folgt definiert werden:

A. Hintergrundsubtraktion

- die Summe der absoluten Differenzen der Farbkomponenten von zwei Pixeln.

$$D(P_1, P_2) = |R_1 - R_2| + |G_1 - G_2| + |B_1 - B_2|$$

- die Summe der quadratischen Differenzen der Farbkomponenten von zwei Pixeln.

$$D(P_1, P_2) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2}$$

- das Maximum der absoluten Differenzen der Farbkomponenten von zwei Pixeln.

$$D(P_1, P_2) = \max\{|R_1 - R_2|, |G_1 - G_2|, |B_1 - B_2|\}$$

Nun wird klar, warum der absolute Wert der Farbkomponenten bei der Subtraktion verwendet wird. Bei der Hintergrundsubtraktion ist nur die Stärke der Abweichung der Farbwerte von Pixeln interessant und nicht die Vorzeichen.

Ein Beispiel für die Hintergrundsubtraktion wird in Abbildung A.1 gezeigt. Das linke Bild ist das vorgefertigte Hintergrundbild. Dieses wird vom aktuell bearbeitenden Bild (mittlere) subtrahiert. Das Ergebnis ist ein Differenzbild (rechts), das die Segmentierung der Vordergrundobjekte ermöglicht.



Abbildung A.1.: Links: Hintergrundbild, Mitte: Einzelbild, Rechts: Differenzbild. Quelle: [hgS]

Damit die Vordergrundobjekte erfolgreich segmentiert werden können, müssen sie sich aufgrund ihrer Farbe bzw. ihres Intensitätswerts vom Hintergrund abheben. Auf dem Differenzbild kann man sehen, in welchen Bereichen und in welchem Ausmaß das aktuell zu

bearbeitende Frame vom Hintergrund abweicht. Je heller eine Region ist, umso größer ist ihre Abweichung vom Hintergrund.

Ein Binärbild weist dagegen jedes Pixel dem Hintergrund oder Vordergrund zu und kann mit der oben beschriebenen Ungleichung erstellt werden. Um die Vordergrundobjekte aus dem aktuell zu bearbeitenden Bild herauszubekommen, verwendet man das Binärbild als eine Maske. Daher wird das Binärbild auch Vordergrundmaske genannt. In Abbildung A.2 sind Binärbilder bei unterschiedlichen Schwellenwerten aus dem obigen Differenzbild erstellt. Für



Abbildung A.2.: Links: Differenzbild, Mitte: Binärbild bei einem großen Schwellenwert, Rechts: Binärbild bei einem kleinen Schwellenwert. Quelle: [hgS]

vollständige Erkennung eines Vordergrundobjekts spielt die Auswahl des Schwellenwerts wichtige Rolle. Wählt man ihn zu groß, werden manche Bereiche der Vordergrundobjekte, die dem Hintergrund ähneln, nicht als Vordergrund erkannt (das mittlere Bild der Abbildung A.2). Wählt man dagegen ihn zu klein, werden auch die Bereiche als Vordergrund erkannt, die vom Hintergrund geringfügig abweichen. In diesem Fall kann das Subtraktionsbild neben Schatten auch Rauschen und Störungen enthalten, weil die Pixelwerte des Hintergrundes nicht immer gleich bleiben, auch wenn kein neues Vordergrundobjekt aufgetreten ist. Solche Änderung des Pixelwertes kann durch vielen Sachen verursacht sein. Zum einen ist das empfundene Licht des Bildsensors einer Kamera bei gleichen Szenen nicht immer 100% gleich und zum anderen kann sein, dass der dem Pixel entsprechende Punkt infolge einer Lichtänderung unterschiedlich beleuchtet wird. Im rechten Bild der Abbildung A.2 sieht man, dass neben dem Schatten auch einige Pixel als Vordergrund er-

A. Hintergrundsubtraktion

kannt sind. Zum Entfernen solcher Störungen gibt es unterschiedliche Verfahren. Sie werden Nachbearbeitungsverfahren genannt. Abbildung A.3 zeigt ein Beispiel dafür. Mithilfe eines Medianfilters wurden die Störungen des linken Binärbildes eliminiert.



Abbildung A.3.: Links: Binärbild mit Störungen, Rechts: Die Störungen mithilfe eines Medianfilter beseitigt. Quelle: [hgS]

In der Hintergrundsubtraktion versteht man unter Hintergrund nicht unbedingt Objekte, die hinterste aus der Sichtrichtung sind bzw. die vorderste Objekte werden nicht alle als Vordergrund definiert. Ein als Vordergrund segmentiertes Objekt kann bei einer späteren Subtraktion auch als Hintergrund erkannt werden.

Die Änderbarkeit des Hintergrundes ist ein wichtiges Thema bei der Hintergrundsubtraktion. Aus der obigen Ungleichung sieht man, dass nur ein Bild als Hintergrund bei der Subtraktion aller Bildern der Videodaten verwendet wurde. Solche Hintergrundsubtraktion, bei der der Hintergrund vordefiniert ist bzw. gleich bleibt, nennt man statische Hintergrundsubtraktion und das entsprechende Verfahren nicht-adaptives Verfahren. Das Hintergrundbild soll bei solchen Verfahren möglichst geringe Anzahl von Objekten enthalten, die als Vordergrund erkannt werden müssen.

Andererseits ist ein Modell als Hintergrund erwünscht, das sich dynamisch anpassen lässt. Das heißt, der Hintergrund bleibt nicht immer gleich, sondern kann mittels eines Verfahrens nach jeder Subtraktion abhängig von den bereits bearbeiteten Bildern verändert werden. So kann ein als Vordergrund erkanntes Objekt zum Hintergrund werden, wenn es sich z. B. über einen bestimmten Zeitraum nicht bewegt. Die Hintergrundsubtraktionen, bei denen der

Hintergrund abhängig von bearbeiteten Bildern verändert wird, nennt man dynamische Hintergrundsubtraktionen und die entsprechenden Verfahren adaptive Verfahren.

Es gibt verschiedene Verfahren zur Hintergrundsubtraktion. Sie analysieren den Hintergrund auf unterschiedliche Art und Weise und unterscheiden sich im wesentlichen dadurch, wie ihre Hintergrundmodelle aufgebaut sind und wie diese Modelle aktualisiert werden. Unter dem Hintergrundmodell ist ein Bild zu verstehen, das als Hintergrund angenommen wird und nicht unbedingt einem Frame der Videodaten entsprechen muss. Bei adaptiven Verfahren werden mehrere Bilder in einem bestimmten Zeitfenster zur automatischen Anpassung des Hintergrundmodells verwendet. Der Aufbau des Hintergrundmodells wird nach jeder Bearbeitung des aktuellen Frames verändert und das Hintergrundmodell entsprechend aktualisiert. Das heißt, das Hintergrundmodell wird in seinen Regionen verändert. Bei nicht-adaptiven Verfahren ist das Hintergrundmodell ein vorbestimmtes Bild, das in seinen Regionen nicht aktualisiert wird, sondern entweder unverändert bleibt oder durch ein neues Frame komplett ersetzt wird. Das Ziel eines Hintergrundmodells ist richtig den Hintergrund zu schätzen. Deshalb werden in diesem Modell neue Hintergründe und sich wiederholende Ereignisse gelernt, damit aus der Erfahrung entschieden werden kann, ob ein Pixel zum Vordergrund oder zum Hintergrund gehört.

Der Ablauf einer Hintergrundsubtraktion sieht wie folgt aus:

1. **Erstellung des Binärbildes:** pixelweiser Vergleich der absoluten Differenz mit einem Schwellenwert
2. **Nachbearbeitung:** Verbesserung des Binärbildes (z.B. mit einem Medianfilter). Das Ergebnis ist die Vordergrundmaske.
3. **Vordergrundobjekte:** Segmentierung/Extraktion der Vordergrundobjekte mit erstellter Maske im aktuellen Frame
4. **Aktualisierung des Hintergrundmodells:** Aktualisierung des Hintergrundmodells mit entsprechendem Verfahren für die nächste Subtraktion

A.1. Einige Verfahren

- **Differenzbild Verfahren:**

Das Verfahren verwendet das letzte Frame als Hintergrund bei der nächsten Subtrakti-

on. Das heißt, jeweils zwei aufeinander folgende Frames werden verglichen. Da das Hintergrundmodell vordefiniert ist, handelt es sich um ein nicht-adaptives Verfahren. Folgende Ungleichung beschreibt das Verfahren.

$$|Frame_i - Frame_{i-1}| > \text{Schwellenwert}$$

Die Vordergrundobjekte werden schnell (ab der nächsten Subtraktion) als Hintergrund eingestuft, sobald sie sich nicht mehr bewegen. Das Verfahren ist bei einer Änderung des Schwellenwerts sehr empfindlich. Wenn der Schwellenwert groß ausgewählt ist, werden meist nur Randbereiche (in Bewegungsrichtung) der Objekte als Vordergrund erkannt. Das liegt daran, dass die Änderungen bei den Pixel-Werten der inneren Bereiche gering sind, weil die inneren Bereiche meist dieselben oder ähnliche Farben haben. Werden nur kleine Randbereiche als Vordergrund erkannt, kommen sie bei der Nachbearbeitung meistens nicht durch und somit wird das Objekt als Vordergrund nicht erkannt. Dieses Problem tritt auch auf, wenn sich das Objekt langsam bewegt oder die Framerate hoch ist. Der Vorteil dieses Verfahrens ist, dass es einfach zu implementieren ist und wenig Speicherplatz benötigt.

- **Mittelwert Verfahren:**

Bei diesem Verfahren werden n Frames für die Berechnung des Hintergrundmodells verwendet. Die arithmetischen Mittelwerte werden pixelweise bei den n Frames berechnet. Das heißt, die Pixel aller n Frames werden erst addiert und dann wird die so berechnete Summe durch die Anzahl der verwendeten Frames n dividiert. Das Ergebnis-Bild wird dann das Hintergrundmodell.

Die Auswahl der n Frames kann unterschiedlich sein. Werden die ersten n Frames für die Berechnung verwendet, sehen die das Verfahren beschreibende Formeln wie folgt aus:

$$\begin{cases} \text{Mittelwert} = \sum_{k=0}^n \frac{Frame_k}{n} \\ |Frame_i - \text{Mittelwert}| > \text{Schwellenwert} \end{cases}$$

In diesem Fall ändert sich das Hintergrundmodell nicht mehr, wenn es einmal erstellt ist. Das heißt, kein neues Vordergrundobjekt wird im Hintergrund integriert und bei

einer Änderung des Hintergrundes wird sich das Hintergrundmodell nicht anpassen. Bei der Erstellung des Hintergrundmodells sollten möglichst keine Vordergrundobjekte bzw. keine abnormale Szenen mitberechnet werden, weil dies zu einem möglichen Fehler im Hintergrundmodell, der nicht mehr korrigierbar ist, führen kann. Wieviele Frames sich bei der Erstellung des Hintergrundmodells beteiligen soll, spielt eine wichtige Rolle bei der Qualität des Hintergrundes. Abgesehen von der einmaligen Berechnung des Hintergrundmodells sind der Rechenaufwand und der Speicherbedarf des Mittelwert-Verfahrens wie bei dem Differenzbild Verfahren.

Aber wenn die Mittelwerte aus den letzten n Frames berechnet werden, ist das Verfahren ziemlich rechenintensiv und benötigt mindestens ($3 * size(Frame)$) Speicher. Die entsprechenden Formeln lauten dann:

$$\begin{cases} \text{Mittelwert} = \sum_{k=1}^n \frac{Frame_{i-k}}{n} \\ |Frame_i - \text{Mittelwert}| > \text{Schwellenwert} \end{cases}$$

Der Vorteil in diesem Fall ist, dass das Hintergrundmodell aus den letzten n Frames aktualisiert wird.

- **Median-Verfahren:**

Wie bei dem Mittelwert-Verfahren werden bei diesem Verfahren auch n Frames für die Berechnung des Hintergrundmodells verwendet. Die Berechnung unterscheidet sich jedoch dadurch, dass die Pixel, die in allen n Frames die gleiche Position haben, erst nach ihren Farbwerten sortiert werden und das mittlere Pixel aus der Sortierung ein Pixel im Hintergrundmodell wird. Bei den Farbbildern benötigt man für jede Farbkomponente eine extra Sortierung. Zum Beispiel für RGB-Bildern werden für jedes Pixel drei Sortierungen in drei Kanalen jeweils ausgeführt. Um die rechenintensiven Sortierungen zu vereinfachen, gibt es auch Ansätze.

Einer davon ist das in [MS95] veröffentlichte Verfahren. Hier wird der Wert des Median-Pixels nicht durch eine Sortierung genau berechnet, sondern geschätzt. Der geschätzte Wert wird mit dem aktuellen Wert verglichen und der Wert des Median-Pixels wird um eins erhöht, wenn er kleiner ist als der Wert des aktuellen Pixels. In anderem Fall, wenn der geschätzte Median größer ist, wird er um eins reduziert. Ansonsten, wenn die beide Werte gleich sind, bleibt der Median unverändert. Auf diese Weise kann der geschätzte

Wert gegen den Median konvergieren, wenn eine Hälfte der kommenden Werte kleiner und die andere Hälfte größer ist als der geschätzte Wert. Ohne Verwendung solcher Ansätze ist der Speicherbedarf bei diesem Verfahren groß ($n * size(Frame)$).

- **“Running-average”:**

Bei diesem Verfahren wird das Hintergrundmodell mit bestimmten Gewicht des letzten Frames aktualisiert. Folgende Gleichungen beschreiben das Verfahren.

$$\begin{cases} Hintergrund_{i+1} = \alpha \cdot Frame_i + (1 - \alpha) \cdot Hintergrund_i \\ |Frame_i - Hintergrund_i| > Schwellenwert \end{cases}$$

Die Größe des Gewichts wird durch die Variable α geregelt. Sie kann einen Wert zwischen 0 und 1 annehmen. Ist $\alpha = 1,0$, wird das letzte Frame das Hintergrundmodell und das Verfahren ist in diesem Fall ein Differenzbild Verfahren. Ist dagegen $\alpha = 0,0$, ändert sich das Hintergrundmodell nicht bzw. vom letzten Frame wird nichts übernommen. Wie schnell ein Vordergrundobjekt in den Hintergrund übergeht, hängt von der Größe dieser Variable ab. Je größer der Wert ist, umso schneller wird ein Vordergrundobjekt im Hintergrund integriert. Die Variable α nennt man daher auch Lernrate. Eine typische Auswahl dafür ist $\alpha = 0,05$. Hier besteht der neue Hintergrund aus der Summe vom letzten Frame, das mit 5% gewichtet ist, und vom alten Hintergrund, der mit 95% gewichtet ist. Der Vorteil dieses Verfahrens ist, dass das Verfahren nur Speicher einer Framegröße ($size(Frame)$) benötigt.

A.2. Das Verfahren “Gaussian Mixture Model”

Das von Stauffer und Grimson in [SG99] vorgestellte Verfahren “Gaussian Mixture Model” (auf deutsch Gauß’sches Mischungsmodell) zur Hintergrundsabtraktion ist ein bekanntes und häufig eingesetztes Verfahren. Dieses Verfahren unterscheidet sich von den zuvor erklärten Verfahren dadurch, dass es die Häufigkeit bzw. die Wahrscheinlichkeit der Werte jedes Pixels bei der Berechnung des Hintergrundmodells einbezieht. Das heißt, für jedes Pixel werden mehrere Werte gespeichert, die bei den Pixeln in den letzten Frames meistens vorgekommen sind. Die gespeicherten Werte werden dann bei der Berechnung des Hintergrundmodells verwendet. Daher eignet sich das Verfahren sehr gut für Hintergründe, die

sich schnell und oft verändern. Beispiele hierfür sind Blätter und Fahnen, die als Hintergrund erkannt werden sollen, aber sich im Wind bewegen. Dieses Verfahren ist auch stabil gegenüber Beleuchtungsänderungen.

Zur Modellierung des Hintergrundes verwendet das Verfahren die Gauß-Verteilungen. Die Idee dieses Verfahrens ist es, die bereits aufgenommenen Werte X_1, \dots, X_t^1 eines Pixels als eine normalverteilte Zufallsvariable zu betrachten und als eine Gauß-Verteilung darzustellen. Unterschiedliche Werte eines Pixels lassen sich so durch mehrere Gauß-Verteilungen interpretieren. Um die Häufigkeit eines Wertes zu berücksichtigen, sind die Gauß-Verteilungen jeweils mit einem Gewicht versehen. Tritt beispielsweise ein Wert bei einem Pixel in den letzten Frames häufig auf, wird die entsprechende Gauß-Verteilung mit einem hohen Gewicht verstärkt. Jedes Pixel des Hintergrundmodells ist also durch eine Gauß'sche Mischungsverteilung modelliert, die aus mehreren gewichteten Gauß-Verteilungen besteht. Die Gauß-Verteilungen werden durch die drei Größen Gewicht ω , Mittelwert μ und Standardabweichung σ charakterisiert. Somit stellt das Verfahren für jedes Pixel eine Mischung mehrerer Gauß-Verteilungen bzw. eine Wahrscheinlichkeit dafür, dass ein Pixel des Hintergrundmodells den Wert X_t hat, durch folgenden Gleichung dar:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

wobei K ist die Anzahl der im Pixel verwendeten Gauß-Verteilungen, $\omega_{i,t}$ das Gewicht der i -te Gauß-Verteilung zum Zeitpunkt t mit dem Mittelwert $\mu_{i,t}$ und mit der Kovarianzmatrix $\Sigma_{i,t}$. η ist die der i -ten Gauß-Verteilung entsprechende Gauß-Funktion und wird durch folgende Gleichung definiert:

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)}$$

Wegen des Rechenaufwands haben Stauffer und Grimson angenommen, dass es keine Zusammenhänge bei den Zufallsvariablen gibt bzw. die RGB-Farbwerte unabhängig von einander sind und die gleiche Varianz haben. Damit hat die Kovarianzmatrix $\Sigma_{k,t}$ eine einfache Form:

$$\Sigma_{k,t} = \sigma_k^2 \mathbf{I}$$

¹ X_j bedeutet dabei ein Farbwert eines Pixels zum Zeitpunkt i

Basierend auf dieser Annahme können sowohl die Kovarianzmatrix als auch die inverse Kovarianzmatrix einfach erstellt werden. Bei der Auswahl des Parameters K sollen der zur Verfügung stehende Speicher und die Rechenleistung berücksichtigt werden. Bei hohen K werden entsprechend mehrere Gauß-Verteilungen für jedes Pixel benötigt und somit steigt der Rechenaufwand und Speicherbedarf. K wird in der Regel zwischen drei und fünf ausgewählt.

Aktualisierung des Hintergrundmodells: Das Hintergrundmodell wird für jedes Frame aktualisiert und die Gauß'sche Mischungsverteilungen aller Pixel werden auch erneut berechnet. Bei der Aktualisierung eines Pixels im Hintergrundmodell werden nicht unbedingt alle ihre Gauß-Verteilungen berücksichtigt. Das heißt, die Gauß-Verteilungen jedes Pixels werden erst jeweils nach ihren Wahrscheinlichkeiten bzw. nach ihren Quotienten $\frac{\omega}{\sigma}$ sortiert und nur die ersten b Gauß-Verteilungen werden bei der Aktualisierung berücksichtigt. Welche Gauß-Verteilung als wahrscheinlich oder unwahrscheinlich angesehen wird, hängt also von dem Quotient der entsprechenden Gauß-Verteilung ab. Eine wahrscheinliche Gauß-Verteilung enthält hier einen Wert, der im aktuellen Frame wahrscheinlich als Hintergrund vorkommen wird. Diese Vermutung folgt aus den vorherigen Werten der Pixel. Eine Gauß-Verteilung mit hoher Wahrscheinlichkeit muss also ein großes Gewicht ω und/oder eine kleine Standardabweichung σ im Vergleich zu den anderen Gauß-Verteilungen des entsprechenden Pixels haben, damit ihr Quotient einen hohen Wert ergibt. Ein großes Gewicht bekommt eine Gauß-Verteilung nur dann, wenn ihrer Mittelwert (mit bestimmter Abweichung) in den letzten Frames häufig vorgekommen ist und als Hintergrund erkannt wurde. Andererseits bleibt die Standardabweichung einer Gauß-Verteilung dann gering, wenn bei einer Erkennung dieser Gauß-Verteilung als Hintergrund der neue Wert des aktuellen Frames von ihrem Mittelwert geringfügig abgewichen ist. Nach dieser Sortierung wird bei jedem Pixel bestimmt, wieviele Gauß-Verteilungen für die Entscheidung verwendet werden und ob die neue Werte der entsprechenden Pixel zum Hintergrund oder Vordergrund gehören. Dazu wird folgende Summe pixelweise in Abhängigkeit eines ausgewählten Schwellenwertes T gebildet.

$$\operatorname{argmin}_b \left(\sum_{k=1}^b \omega_k \geq T \right)$$

Man addiert also bei jedem Pixel die Gewichte der ersten b wahrscheinlichsten Gauß-Verteilungen, sodass die resultierende Summe gerade etwas größer als der Schwellenwert T ist. Wieviele Gewichte dafür addiert werden müssen, hängt stark von der Auswahl des Schwellenwertes T ab. Der Schwellenwert T kann zwischen 0 und 1 liegen. Ist $T = 1$, so

müssen für die obige Bedingung alle Gewichte addiert werden bzw. es werden alle Gauß-Verteilungen bei der Aktualisierung des Hintergrundmodells verwendet. In diesem Fall ist die Wahrscheinlichkeit hoch, dass der neue Wert als Hintergrund erkannt wird. Daher eignet sich große Schwellenwert T gut bei Szenen, in denen der Hintergrund sich ständig verändert. Es besteht aber dabei die Gefahr, dass die Vordergrundobjekte schnell bzw. ab dem nächsten Frame als Hintergrund erkannt werden. Ist dagegen der Schwellenwert zu klein z. B. $T = 0$ ausgewählt, so wird nur eine Gauß-Verteilung verwendet. Damit ein Pixel als Hintergrund erkannt wird, muss in diesem Fall der neue Wert des Pixels nur dem Wert ähneln, den die wahrscheinlichste Gauß-Verteilung darstellt. Man sieht, dass die Anzahl der verwendeten Gauß-Verteilungen in jedem Pixel unterschiedlich hoch sein darf. Die nicht verwendeten Gauß-Verteilungen werden aber nicht gelöscht, weil sie bei der Aktualisierung des nächsten Frames eine Rolle spielen könnten.

Ist für ein Pixel bestimmt, wieviele und welche Gauß-Verteilungen verwendet werden müssen, so kann für einen neuen Wert eines Pixel untersucht werden, ob sich der Wert auf den Vordergrund oder auf den Hintergrund bezieht. Dafür muss man überprüfen, ob der neue Wert einer der ausgewählten Gauß-Verteilungen zugeordnet werden kann. Der neue Wert kann einer Gauß-Verteilung zugewiesen werden, wenn er nicht weit vom Mittelwert der prüfenden Gauß-Verteilung entfernt ist bzw. wenn folgende Ungleichung² gilt:

$$|\text{neuerWert} - \text{Mittelwert}_i| < 2,5 \cdot \sigma_i$$

wobei σ_i die Standardabweichung der prüfenden Gauß-Verteilung ist. Gilt diese Bedingung, gehört der neue Wert zur prüfenden Gauß-Verteilung und das entsprechende Pixel wird als Hintergrund erkannt. Anderenfalls wird das dem neuen Wert entsprechende Pixel Vordergrund. Nach der Feststellung eines Pixels zum Vordergrund oder Hintergrund müssen seine Gauß'sche Mischungsverteilungen für die nächsten Aktualisierungen auch erneut berechnet werden.

Aktualisierung der Gaußschen Mischungsverteilungen: Hinter jedem Pixel sind K wahrscheinlichste Hintergrundwerte mit unterschiedlichen Abweichungen enthalten. Um heraus-

²Bei dieser Ungleichung wurde für den neuen Wert die Größe $2,5 \cdot \sigma$ als erlaubte Abweichung vom Mittelwert der prüfenden Gauß-Verteilung ausgewählt. Bei einer Gauß-Verteilung liegen im allgemein 95,45% aller Messwerte im Bereich der doppelten Standardabweichung $\pm 2 \cdot \sigma$ um den Mittelwert und 99,73% aller Messwerte im Bereich der dreifachen Standardabweichung $\pm 3 \cdot \sigma$ um den Mittelwert. Siehe [Wik11h]

zufinden, ob der neue Wert eines Pixels bekannt in seinen Hintergrundwerten ist, wurde bei der Aktualisierung des Hintergrundmodells bereits untersucht, ob der neue Wert zu einer Gauß-Verteilung des Pixels zugeordnet werden kann. Wenn für den neuen Wert keine Gauß-Verteilung gefunden wird, wird eine neue Gauß-Verteilung für ihn erstellt und die am wenigsten wahrscheinliche Gauß-Verteilung des Pixels wird durch diese ersetzt. Dem Mittelwert der neuen Gauß-Verteilung wird dann der neue Wert als Anfangswert zugewiesen. Der Anfangswert für Gewicht soll klein und der Anfangswert für die Varianz hoch ausgewählt werden, damit der neue Wert bzw. die neue Gauß-Verteilung nicht sofort als Hintergrund erkannt wird.

In anderem Fall, wenn der neue Wert zugeordnet werden könnte, werden die Größen der Mittelwert, die Varianz und das Gewicht bei der entsprechenden Gauß-Verteilung aktualisiert. Nach der Aktualisierung des Gewichtes werden sich aber auch die Gewichte anderer Gauß-Verteilungen des Pixels verändern, weil das Gewicht der zugeordneten Gauß-Verteilung erhöht wird und infolgedessen alle Gewichte normalisiert werden. Um die Gewichte alter Werte im Verhältnis zu neuen Werten zu schwächen und neue Werte zu lernen, wird eine Lernrate α bei der Aktualisierung der Gewichte eingeführt. Das neue Gewicht wird nach folgender Gleichung berechnet:

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha$$

wobei $\omega_{k,t}$ das Gewicht der k -te Gauß-Verteilung zum Zeitpunkt t ist. Mit einer großen Lernrate α wird das alte Gewicht stark geschwächt und infolgedessen werden neue Werte schnell adaptiert.

Die Tabelle A.1 gibt eine Übersicht der Parameter, die das Verfahren Gauß'sche Mischverteilungen verwendet.

Parameter	Bedeutung
K	Anzahl der in einem Pixel verwendeten Gauß-Verteilungen
ω	initiales Gewicht
σ^2	initiale Varianz
α	Lernrate
T	Schwellenwert für Hintergrundtest mit Summe der Gewichte
σ	Koeffizient der erlaubten Abweichung vom Mittelwert

Tabelle A.1.: Parameter des Verfahrens Gauß'sche Mischverteilungen

Literaturverzeichnis

- [Br1] Brutzer, S. and Höferlin, Benjamin and Heidemann, Gunther . Evaluation of Background Subtraction Techniques for Video Surveillance . In *Computer Vision and Pattern Recognition (CVPR)* , pp. 1937–1944 . IEEE , 2011 . (Zitiert auf den Seiten 7 und 12)
- [ADSB05] P. Azzari, L. Di Stefano, A. Bevilacqua. An Effective Real-Time Mosaicing Algorithm Apt to Detect Motion Through Background Subtraction Using a PTZ Camera. In *Proceedings of international Conference on Advanced Video and Signal Based Surveillance*, pp. 511–516. 2005. (Zitiert auf Seite 40)
- [BBZ96] B. Bascle, A. Blake, A. Zisserman. Motion Deblurring and Super-resolution from an Image Sequence. In *In Proceedings of the Fourth European Conference on Computer Vision*, pp. 573–582. Springer-Verlag, 1996. (Zitiert auf Seite 42)
- [BDBH02] A. Bartoli, N. Dalal, B. Bose, R. Horaud. From Video Sequences to Motion Panoramas. In *Proceedings of the Workshop on Motion and Video Computing*, pp. 201–207. 2002. (Zitiert auf Seite 40)
- [BEBV08] T. Bouwmans, F. El Baf, B. Vachon. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. *Recent Patents on Computer Science*, 1(3):219–237, 2008. (Zitiert auf Seite 39)
- [BEN04] M. Ben-Ezra, S. K. Nayar. Motion-Based Motion Deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:689–698, 2004. (Zitiert auf Seite 41)
- [BSK00] K. S. Bhat, M. Saptharishi, P. K. Khosla. Motion Detection and Segmentation using Image Mosaics. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1577–1580. 2000. (Zitiert auf den Seiten 39 und 40)

- [BTVGo6] H. Bay, T. Tuytelaars, L. Van Gool. SURF: Speeded Up Robust Features. In *Proceedings of the ninth European Conference on Computer Vision*, pp. 404–417. 2006. (Zitiert auf Seite 63)
- [Che95] S. E. Chen. QuickTime VR: an image-based approach to virtual environment navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, SIGGRAPH '95*, pp. 29–38. 1995. (Zitiert auf Seite 40)
- [CLo9] S. Cho, S. Lee. Fast motion deblurring. In *ACM Transactions on Graphics (TOG)*, volume 28, p. 145. ACM, 2009. (Zitiert auf den Seiten 7 und 13)
- [CPVo6] R. Cucchiara, A. Prati, R. Vezzani. Advanced video surveillance with pan tilt zoom cameras. In *Proceedings of Workshop on Visual Surveillance at ECCV*. 2006. (Zitiert auf Seite 40)
- [CToo] S. Coorg, S. Teller. Spherical Mosaics with Quaternions and Dense Correlation. *International Journal of Computer Vision*, 37:259–273, 2000. (Zitiert auf Seite 40)
- [Far03] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian conference on Image analysis*, pp. 363–370. Springer-Verlag, 2003. (Zitiert auf Seite 92)
- [Furo9] Furuti, C.A. Map Projections. <http://www.progonos.com/furuti/MapProj/Normal/TOC/cartTOC.html>, 2009. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 8 und 38)
- [GTS⁺10] C. Guillot, M. Taron, P. Sayd, Q. C. Pham, C. Tilmant, J. M. Lavest. Background subtraction adapted to PTZ cameras by keypoint density estimation. In *Proceedings of the British Machine Vision Conference*, pp. 34.1–34.10. BMVA Press, 2010. (Zitiert auf Seite 41)
- [hgS] Hintergrundsubtraktion. <http://wiki.ohm-hochschule.de/toolboxwiki/index.php/Hintergrundsubtraktion>. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 10, 100, 101 und 102)
- [KPK⁺03] S. Kang, J. Paik, A. Koschan, B. Abidi, M. Abidi. Real-time video tracking using PTZ cameras. In *Proceedings of the Sixth International Conference on Quality Control by Artificial Vision*, pp. 103–111. 2003. (Zitiert auf Seite 39)

- [Kö] Köhne, A. and Wößner, M. Kartenprojektionen. <http://www.kowoma.de/gps/geo/Projektionen.htm>. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 34)
- [Luc74] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745–754, 1974. (Zitiert auf Seite 41)
- [MB95] L. McMillan, G. Bishop. Plenoptic Modeling: An Image-Based Rendering System. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, SIGGRAPH '95*, pp. 39–46. 1995. (Zitiert auf Seite 40)
- [MS95] N. McFarlane, C. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8:187–193, 1995. (Zitiert auf Seite 105)
- [par] How To Take And Make A Panoramic Photo. <http://www.great-landscape-photography.com/panoramic-photo.html>. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 8 und 49)
- [Pico04] M. Piccardi. Background subtraction techniques: a review. In *Proceedings of IEEE International Conference on Systems Man and Cybernetics*, volume 4, pp. 3099–3104. 2004. (Zitiert auf Seite 39)
- [Ric72] W. H. Richardson. Bayesian-Based Iterative Method of Image Restoration. *Journal of The Optical Society of America*, 62, 1972. (Zitiert auf Seite 41)
- [Scho05] O. Schreer. *Stereoanalyse und Bildsynthese*. Springer, Berlin, Heidelberg, New York, 1 edition, 2005. (Zitiert auf den Seiten 7 und 30)
- [SD10] K. Sankaranarayanan, J. W. Davis. PTZ camera modeling and panoramic view generation via focal plane mapping. In *Proceedings of the 10th Asian Conference on Computer Vision, ACCV'10*, pp. 580–593. 2010. (Zitiert auf den Seiten 40 und 75)
- [SG99] C. Stauffer, W. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 246–252. 1999. (Zitiert auf den Seiten 15, 39 und 106)
- [SP06] S. N. Sinha, M. Pollefeys. Pan-tilt-zoom camera calibration and high-resolution mosaic generation. *Computer Vision and Image Understanding*, 103(3):170–183, 2006. (Zitiert auf Seite 75)

- [TMO8] R. Trichet, B. Merialdo. Keypoints labeling for background subtraction in tracking applications. In *Proceedings of the 2008 IEEE International Conference on Multimedia and Expo*, pp. 1301–1304. 2008. (Zitiert auf Seite 41)
- [Wik11a] Wikipedia, Die freie Enzyklopädie. Bildwinkel. <http://de.wikipedia.org/w/index.php?title=Bildwinkel&oldid=88488554>, 2011. (Abgerufen: 3. Mai 2011, 16:10 UTC). (Zitiert auf den Seiten 7 und 19)
- [Wik11b] Wikipedia, Die freie Enzyklopädie. Brennweite. <http://de.wikipedia.org/w/index.php?title=Brennweite&oldid=88528899>, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 18)
- [Wik11c] Wikipedia, Die freie Enzyklopädie. Dekonvolution. <http://de.wikipedia.org/w/index.php?title=Dekonvolution&oldid=88729890>, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 10 und 90)
- [Wik11d] Wikipedia, Die freie Enzyklopädie. Hough-Transformation. <http://de.wikipedia.org/w/index.php?title=Hough-Transformation&oldid=85891917>, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 32)
- [Wik11e] Wikipedia, Die freie Enzyklopädie. Kartennetzentwurf. <http://de.wikipedia.org/w/index.php?title=Kartennetzentwurf&oldid=89003763>, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 8 und 37)
- [Wik11f] Wikipedia, Die freie Enzyklopädie. Linsengleichung. <http://de.wikipedia.org/w/index.php?title=Linsengleichung&oldid=87590765>, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 19)
- [Wik11g] Wikipedia, Die freie Enzyklopädie. Mittabstandstreue Azimutalprojektion. http://de.wikipedia.org/w/index.php?title=Mittabstandstreue_Azimutalprojektion&oldid=86520616, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 34)
- [Wik11h] Wikipedia, Die freie Enzyklopädie. Normalverteilung. <http://de.wikipedia.org/wiki/Normalverteilung>, 2011. (Abgerufen: 3. Mai 2011, 16:10 UTC). (Zitiert auf Seite 109)
- [Wik11i] Wikipedia, Die freie Enzyklopädie. Verzeichnung. <http://de.wikipedia.org/w/index.php?title=Verzeichnung&oldid=86363029>, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 24)

- [Wik11j] Wikipedia, The Free Encyclopedia. Equirectangular projection. http://en.wikipedia.org/wiki/Equirectangular_projection, 2011. (Abgerufen: 9. Mai 2011). (Zitiert auf den Seiten 7 und 35)
- [Zha00] Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1330–1334, 2000. (Zitiert auf Seite 51)
- [ZXRH99] Z. Zhu, G. Xu, E. Riseman, A. Hanson. Fast Construction of Dynamic and Multi-Resolution 360° Panorama from Video Sequences. In *Proceedings of International Conference on Multimedia Computing and Systems*, volume 1, pp. 400–406. 1999. (Zitiert auf Seite 40)

Alle URLs wurden zuletzt am 17.03.2008 geprüft.

Erklärung

Hiermit versichere ich, diese Arbeit selbständig verfasst und nur die angegebenen Quellen benutzt zu haben.

(Davaadorj Tsendragchaа)