

Reihe 10

Informatik/
Kommunikation

Nr. 853

Dipl.-Ing. Peter Seibold,
Erkelenz

Visuelle Odometrie in Echtzeit für ein Fluggerät



FernUniversität in Hagen
Schriften zur Informations-
und Kommunikationstechnik

FernUniversität in Hagen

DISSERTATION

Visuelle Odometrie in Echtzeit für ein Fluggerät

Der Fakultät für Mathematik und Informatik der FernUniversität in Hagen zur
Erlangung des akademischen Grades eines

Dr.-Ing.
eingereichte Dissertation
von
Dipl.-Ing. Peter Ulrich Seibold
aus
Balingen

Gutachter: 1. Prof. Dr.-Ing. M. Gerke
2. Prof. Dr.-Ing. J. Horstmann

Datum der mündlichen Prüfung: 23.12.2016

Danksagungen

Die vorliegende Dissertation ist während meiner Forschungsarbeiten im Lehrgebiet Prozesssteuerung und Regelungstechnik der Fakultät für Mathematik und Informatik der FernUniversität in Hagen entstanden.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. Gerke, der meine Forschungsrichtung anregte und begleitete. Mit Akribie widmete er sich meiner Arbeit und gab mir zahlreiche sehr nützliche Hinweise. Bei einem herben Rückschlag bezüglich der Versuchsergebnisse ermutigte er mich einen anderen Weg zu beschreiten.

Herrn Prof. Dr.-Ing. Horstmann danke ich für die Übernahme des Zweitgutachtens. Auch er gab mir in einem ausführlichen Gespräch einige Hinweise, die meine Arbeit abrundeten.

Den Kollegen des Lehrgebiets danke ich für die Hilfe beim Umbau und der Ansteuerung des XY-Tisches, bei der Unterstützung zu Messfahrten und den vielen Gesprächen.

Mein herzlicher Dank gilt meiner Lebensgefährtin Monika, die mich unermüdlich antrieb die Ergebnisse zu Papier zu bringen, geduldig meine monatelange geistige Abwesenheit ertrug und meine orthografischen Kenntnisse deutlich erweiterte.

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Peter Seibold,
Erkelenz

Nr. 853

Visuelle Odometrie in Echtzeit für ein Fluggerät



FernUniversität in Hagen
Schriften zur Informations-
und Kommunikationstechnik

Seibold, Peter

Visuelle Odometrie in Echtzeit für ein Fluggerät

Fortschr.-Ber. VDI Reihe 10 Nr. 853. Düsseldorf: VDI Verlag 2017.

224 Seiten, 212 Bilder, 26 Tabellen.

ISBN 978-3-18-385310-6, ISSN 0178-9627,

€ 76,00/VDI-Mitgliederpreis € 68,40.

Für die Dokumentation: Bildregistrierung mit Stereokamera – Visuelle Odometrie – 3D-Positionsbestimmung – Normierte Kreuzkorrelation (NCC) – Optischer Fluss – Subpixel

Die vorliegende Arbeit wendet sich an Wissenschaftler und Ingenieure im Bereich der Bildverarbeitung. Sie befasst sich mit der Ermittlung der Bewegung und Position eines unbemannten Fluggerätes mit einer Flughöhe zwischen 0,6 m und 100 m. Die Bilderfassung erfolgt durch eine Stereokamera mit kleiner Basisbreite. Zwei Methoden – optischer Fluss und gebietsbasierte Verfahren (block matching) – werden auf Verwendbarkeit zur Bewegungsanalyse geprüft und verglichen. Durch die kleine Basisbreite der Kamera sind besondere Anforderungen an die Subpixelgenauigkeiten gestellt, die eingehend evaluiert werden. Um Echtzeitbedingungen einzuhalten, werden einige Vereinfachungen eingeführt wie z. B. Reduzierung der Bilinearkoeffizienten für die Rektifizierung und situationsabhängige Bildverkleinerungen. Die Hardwareplattformen CPU, FPGA, DSP und GPGPU werden auf Eignung untersucht

Bibliographische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter <http://dnb.ddb.de> abrufbar.

Bibliographic information published by the Deutsche Bibliothek

(German National Library)

The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at <http://dnb.ddb.de>.

Schriften zur Informations- und Kommunikationstechnik

Herausgeber:

Wolfgang A. Halang, Lehrstuhl für Informationstechnik

Herwig Unger, Lehrstuhl für Kommunikationstechnik

FernUniversität in Hagen

© VDI Verlag GmbH · Düsseldorf 2017

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-385310-6

Inhaltsverzeichnis

Abkürzungen	IX
Kurzfassung	X
1 Einleitung	1
1.1 Luftschiff: Entstehung und Eigenschaften	1
1.2 Aufbau der Bilderfassung	6
1.3 Stand der Technik	7
1.4 Aufbau der Arbeit	11
2 Kameramodell und Bewegung	14
2.1 Modell der Lochkamera	14
2.1.1 Geometrische Transformationen	15
2.1.2 Kalibrierung bei unverzerrter Abbildung	20
2.1.3 Geometrie mehrerer Abbildungen einer Szene	22
2.1.4 Rektifizierung	26
2.2 Modell der Linsenkamera	27
2.3 Fotosensoren auf Halbleiterbasis	28
2.4 Abbildungsfehler	29
2.4.1 Abbildungsfehler durch Objektive	29
2.4.1.1 Vignettierung	29
2.4.1.2 Verzeichnung	31
2.4.2 Abbildungsfehler durch Sensoren	34
2.4.2.1 Rauschen	34
2.4.2.2 Blooming	35
2.4.2.3 Smear-Effekt	35
2.4.2.4 Moiré-Effekt (Unterabtastung)	35
2.5 Behebung der Verzeichnungen	37
2.5.1 Beschreibung radialer und tangentialer Verzeichnung	37
2.5.2 Kalibrierung und Rektifizierung bei verzerrter Abbildung	38
2.6 Ermittlung der Bewegung	41
2.6.1 Stereogeometrie	41
2.6.2 Translation	44
2.6.3 Rotation	46
2.6.4 Geschwindigkeitsbestimmung	48
2.7 Zusammenfassung Kapitel 2	48
3 Bewegungsabschätzung und optischer Fluss	49

3.1	Einleitung	49
3.2	Voraussetzungen und Probleme bei der Bewegungserfassung	51
3.2.1	Korrespondenzproblem	51
3.2.2	Blendenproblem	52
3.2.3	Beleuchtungsänderung	53
3.2.4	Verdeckung	54
3.3	Klassifizierung der Bewegungsabschätzungsmethoden	56
3.3.1	Differentielle Methoden	57
3.3.2	Gebietspaarung	57
3.3.2.1	SAD	57
3.3.2.2	SSD	57
3.3.3	Energiebasierte Methoden	58
3.3.4	Phasentechniken	58
3.3.5	Merkmalsbasierte Methoden	58
3.3.5.1	Harris Ecken Detektor	59
3.4	Optischer Fluss, differenzielles Verfahren	60
3.4.1	Smoothness Constraint nach B. K. Horn und B. G. Schunck	63
3.4.2	Flusskonstanz nach Lucas&Kanade	65
3.4.3	Vergleich der Verfahren nach Horn&Schunck und Lucas&Kanade	66
3.4.4	Gaußpyramide	67
3.5	Optischer Fluss, Korrelationsmethode	71
3.5.1	ZNCC	71
3.5.2	NCC	74
3.5.3	Eigenschaften von NCC und ZNCC	75
3.6	Zusammenfassung Kapitel 3	77
4	Vergleich der Verfahren nach Lucas&Kanade und NCC	78
4.1	Reichweite der Erkennung	78
4.2	Helligkeitsempfindlichkeit	80
4.3	Rotationsempfindlichkeit	81
4.4	Berechnungszeit	88
4.5	Subpixelgenauigkeit	88
4.5.1	Subpixelgenauigkeit beim differentiellen Verfahren nach Lucas&Kanade	88
4.5.2	Subpixelgenauigkeit bei der NCC	93
4.5.2.1	Verbesserung durch Abzug des Mittelwertes	97
4.5.2.2	Interpolation über Parabel	103
4.5.2.3	Interpolation über Gaußkurve	106
4.5.2.4	Interpolation über Geraden	106
4.5.2.5	Vorzeichenabhängige Interpolation	107

4.5.2.6	Interpolation durch Bildvergrößerungen	109
4.5.2.7	Neue Methode über 'inverse Parabel'	109
4.5.2.8	Pixel-Locking	113
4.5.3	Vergleich der Subpixelfunktionen	116
4.6	Zusammenfassung Kapitel 4	122
5	Versuchsergebnisse	123
5.1	Virtuelle Welt	123
5.2	XY-Tisch	128
5.2.1	Ergebnisse	130
5.2.1.1	Gerader Pfad	130
5.2.1.2	Kreisbahn	131
5.3	Flugaufnahmen mit Monokamera	134
5.4	Auswertung von Stereoaufnahmen	138
5.4.1	Kameradaten	138
5.4.1.1	Daten nach Herstellerangaben	139
5.4.1.2	Eigene Messungen	139
5.4.2	Erfassungsbereiche für Geschwindigkeiten und Abstände	142
5.4.3	Prüfung der Kalibrierung und Rektifizierung über Abstandsrechnungen	144
5.4.4	Prüfung der Bewegungsberechnung	149
5.4.5	Verfahren für schnelle Berechnung bei realitätsnahen Versuchen	156
5.4.5.1	Ausgleich des Intensitätsverlaufs	157
5.4.5.2	Vereinfachung der bilinearen Interpolation	157
5.4.5.3	Verringerung der Präzision der Bilinearkoeffizienten	158
5.4.5.4	Reduzierung der Rektifikation auf benutzte Bereiche	159
5.4.5.5	Suchbereich für die Bewegungserfassung einschränken	160
5.4.5.6	Bildverkleinerung	161
5.4.5.7	Nachbearbeitung: Ausschließen deutlich falsch berechneter Werte	164
5.4.6	Realitätsnahe Versuche	167
5.5	Zusammenfassung Kapitel 5	176
6	Implementation in Hardware	177
6.1	Berechnungszeiten	177
6.2	Maschinennahe Programmierung	178
6.3	Parallelisierung	178
6.4	Hardwareprozessoren	179
6.4.1	DSP	180
6.4.2	FPGA	180
6.4.3	GPGPU	181
6.5	Zusammenfassung Kapitel 6	185

7 Zusammenfassung und Ausblick	186
Anhang	191
A.1 Lochkamera: Belichtungszeit und Unschärfe durch Beugung	191
A.2 Nützliche Eigenschaften digitaler Filter	192
A.3 Summentabellen	197
A.4 Berechnung der NCC über die Schwarzsche Ungleichung	199
A.5 Synchronisations- und Bildratenbestimmung der Stereokamera	202
Literaturverzeichnis	205

Abkürzungen

3D	dreidimensional
DGPS	Differential Global Positioning System
DSP	Digital Signal Processor
FPGA	Field Programmable Gate Array
fps	Bildrate (frames per second)
GFLOPS	Giga Floating Point Operations Per Second
GPGPU	General Purpose Computation on Graphics Processing Unit
GPS	Global Positioning System
GPU	Graphics Processing Unit
IMU	Inertial Measurement Unit
NCC	Normalized Cross Correlation
NCCa	Normalisierte Kreuzkorrelation approximiert
NED	North – East – Down
SAD	Sum of Absolute Differences
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localization and Mapping
SSD	Sum of Squared Differences
SURF	Speeded Up Robust Features
UAS	Unmanned Aerial System
YPR	Yaw-Pitch-Roll
ZNCC	Zero-mean Normalized Cross Correlation

Kurzfassung

Unbemannte autonome Kleinfluggeräte gewinnen in den letzten Jahren zunehmend an Bedeutung, da durch Beobachtungen aus der Luft sowohl große Areale überwacht werden können als auch Detailansichten aus verschiedenen Perspektiven ermöglicht werden. Die Miniaturisierung der Sensoren und Aktuatoren erlaubt, leichte Fluggeräte mit niedrigem Energiebedarf zu erstellen. Durch diese Sensoren wird, sofern deren Taktrate und Präzision ausreichen, um die Bewegung des Fluggerätes zu erfassen und zu regeln, der autonome Flug – z. B. entlang eines vorgegebenen Pfades – ermöglicht. GPS (Global Positioning System), ein in der Robotik häufig eingesetztes System, liefert die horizontale absolute Position, allerdings ohne zusätzliche Infrastruktur nur mit Genauigkeiten im unteren Meterbereich und die auf ein fiktives Rotationsellipsoid bezogene Höhe auf etwa 10 m genau. Mit der Wiederholungsrate von ca. 1/s ist GPS für eine dynamische Bewegungsregelung nicht ausreichend schnell. Deshalb werden andere Sensoren hinzugezogen. Ein Inertialsensor (IMU) erfasst schnelle Bewegungen sehr gut und ist für eine Kurzzeitregelung geeignet. Da dieser Sensor Beschleunigungen misst, werden Strecken über zweifache Integration ermittelt. Dies ergibt allerdings schon nach kurzer Zeit, bedingt durch einen sich veränderten Gleichanteil im Messwert, quadratisch anwachsende Fehler, die durch weitere Sensoren kompensiert werden müssen. Dazu eignen sich u. a. Kameras, die klein und leichtgewichtig sind und unter Nutzung der visuellen Odometrie zur Verbesserung der Bewegungserfassung führen. Zudem kann mit Stereokameras sehr genau die Höhe des Fluggerätes über dem darunterliegenden Terrain bestimmt werden. Die Daten der drei Sensoren (GPS, IMU und Kameras) werden an ein vorhandenes Flugsystem über ein Kalmanfilter fusioniert.

Die vorliegende Dissertation befasst sich mit der visuellen 3D-Odometrie, um die Position – damit auch die Höhe über Boden – und die momentane Geschwindigkeit eines unbemannten Fluggerätes (UAS, Unmanned Aerial System) zu bestimmen. Als Fluggerät wird in dieser Arbeit ein Blimp (Luftschiff ohne Gerüst) der FernUniversität in Hagen verwendet.



Bild 1: Blimp der FernUniversität in Hagen

Die besonderen Anforderungen sind gegeben durch die geringe Nutzlast, die eingeschränkte Energieverfügbarkeit bei einem Kleinflugschiff und die Echtzeitbildverarbeitung. Um permanente zeitaufwendige Nachkalibrierungen zu vermeiden, kommt eine rigide Stereokamera mit kleiner Basisbreite und geringer Masse zum Einsatz.

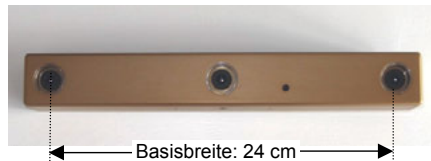


Bild 2: Eingesetzte Stereokamera „Bumblebee“ der Firma Point Grey

Zur Bestimmung des Abstandes zu einer Szene nehmen zwei Kameras diese Szene gleichzeitig auf. Über Bildregistrierungsmethoden wird die Position eines identischen Szenenausschnitts in beiden Aufnahmen bestimmt und über Triangulation der Abstand zur Szene ermittelt. Für die Bewegungsbestimmung werden in zwei nacheinander mit einer Kamera aufgenommenen Bildern identische Szenenausschnitte gesucht. Mit der bekannten Höhe lassen sich die Verschiebung der Kamera und damit die Position des Fluggerätes berechnen. Da die Zeiten der Kameraaufnahmen bekannt sind, kann daraus die Geschwindigkeit des Fluggerätes bestimmt werden. Für die Orientierung der Kameras, die ebenfalls in die Berechnung eingeht, werden Winkeldaten aus der IMU herangezogen.

Durch die für große Abstände sehr geringe Basisbreite der Stereokamera von 24 cm ist das Sichtfeld der eingesetzten Kameras ein Kompromiss zwischen maximal auflösbarer Höhe und maximal erfassbarer Geschwindigkeit. Deshalb werden Verfahren der Bildregistrierung untersucht, die durch Subpixelgenauigkeit eine höhere Auflösung ermöglichen. Hierzu zählen Gradientenverfahren, die ohnehin zunächst nur für Subpixelabstände definiert

sind, oder flächenbasierte Verfahren. Zwei Verfahren, die Gradientenmethode nach Lucas & Kanade und ein flächenbasiertes Verfahren, die Kreuzkorrelation, werden auf Tauglichkeit mit realen Aufnahmen geprüft. Wichtige Eigenschaften sind hierbei die maximal erfassbare Translation, die Genauigkeit, die Empfindlichkeit auf Rotationen und der Einfluss von Belichtungsunterschieden.

Eingehend wird die Subpixelgenauigkeit, u. a. hier mit nachrichtentechnischen Methoden, evaluiert. Zur Subpixelbestimmung im Zusammenhang mit der Kreuzkorrelation wird eine Funktion – hier 'inverse Parabel' genannt – eingeführt. Ziel ist es, eine Subpixelgenauigkeit kleiner als 0.3 px zu erreichen. Mit realen Abbildungen werden die gewählten Verfahren auf ihre Genauigkeiten bezüglich der erfassten Bewegungen überprüft.

Die Echtzeitverarbeitung stellt in diesem Zusammenhang eine besondere Herausforderung dar. Deshalb werden mehrere Maßnahmen vorgestellt, die die Verarbeitung der großen Mengen an Daten vereinfachen, ohne an erforderlicher Präzision zu verlieren. Dazu zählen Gaußfilter mit Ganzzahlwerten und Divisor in Zweierpotenz, Reduzierung der Bilinearkoeffizienten für die Rektifizierung auf wenige Bit Genauigkeit, Einschränken der Gebiete für die Korrespondenzanalyse durch Ausschnitte und Bildverkleinerungen.

Die Parallelisierbarkeit der Algorithmen und deren Einsatz in verschiedene Hardwareplattformen werden erläutert.

1 Einleitung

1.1 Luftschiff: Entstehung und Eigenschaften

Noch vor ca. 150 Jahren wurde die Realisation eines lenkbaren Luftschiffes als unmöglich angesehen. So zitierte Ludwig Boltzmann in seinem Aufsatz „Über Luftschiffahrt“ [1] von 1894 das wissenschaftliche Umfeld so:

„Allein wie vor Gauss die Lösung des Problems der Kreistheilung, so misslang auch bisher die Herstellung des lenkbaren Luftschiffes, so dass das Problem in bedenklicher Weise in Misscredit kam, ja grosse Theoretiker sich sogar zur Ansicht hinneigten, seine Lösung sei unmöglich.“

Doch gleich im Anschluss kommt L. Boltzmann zu folgender Erkenntnis:

„Die Unrichtigkeit der alten Formeln wurde klar erwiesen und ich glaube, Ihnen den Beweis liefern zu können, dass die Lösung des Problems nicht nur möglich ist, sondern aller Wahrscheinlichkeit nach schon in kurzer Zeit gelingen wird.“

Am 13. August 1898 erteilte das Kaiserliche Patentamt Herrn Ferdinand Graf von Zeppelin das Patent Nummer 98580 für einen „Lenkbaren Luftfahrzeug mit mehreren hintereinander angeordneten Tragkörpern“.

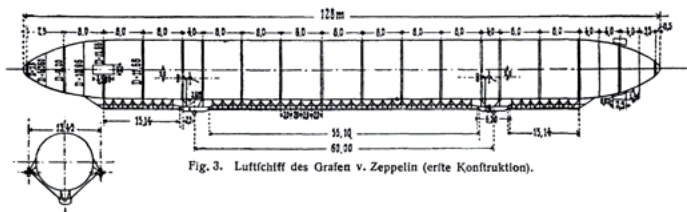


Bild 3: Luftschiff des Grafen von Zeppelin

Quelle: <https://de.wikipedia.org/wiki/Datei:L-Luftschiff1.png>

In dieser Patentschrift ging es im Wesentlichen um:

- schlankes, starres Gerippe aus Duraluminium, bestehend aus Ringen und Längsträgern
- Gasraum, aufgeteilt in viele zylindrische Zellen
- Steuerungsmöglichkeit mit Hilfe von Höhen- und Seitenrudern
- zwei getrennte, fest mit dem Gerippe verbundene Gondeln
- Vortrieb durch Propeller, montiert auf Höhe des größten Luftwiderstandes
- Möglichkeit, mehrere solcher Schiffe wie Zugwaggons aneinander zu koppeln

Schon 1910 flog das erste Passagierluftschiff in Baden-Baden. Von nun an gerieten Luftschiffe in das allgemeine Interesse der Bevölkerung, auch wenn der Luxus eines Fluges nur wenigen ermöglicht wurde. Begonnen hatte das Ausnutzen des statischen Auftriebs nach dem Prinzip „leichter als (kalte) Luft“ mit den Versuchen der Gebrüder Montgolfier, die 1793 den ersten Heißluftballon aufsteigen ließen. Die Technik wird detailliert im Buch des Zeitgenossen Barthélemy Faujas De Saint-Fond [2] beschrieben.

Die zivile Luftschiffahrt erfuhr ihren Schicksalsschlag 1937 mit dem Brand der Hindenburg LZ129. Das Ende der Zeppeline kam 1940, weder das Militär noch die Propaganda sahen einen weiteren Nutzen in Luftschiffen und deshalb wurden die beiden letzten großen Luftschiffe LZ127 und LZ131 vernichtet.

Heute werden bemannte Luftschiffe überwiegend zu Rundflügen und für Werbezwecke eingesetzt.

Andere Anwendungen finden langsam ihren Einzug. Denn Beobachtungen aus der Luft gewinnen zunehmend an Bedeutung, da sowohl große Areale überwacht werden können als auch Detailansichten unter verschiedenen Winkeln ermöglicht werden und dabei die gegenseitige Beeinflussung von Beobachtungsgerät und Betrachtungsobjekt in der Regel nur gering ist. Die Anwendungsgebiete sind vielfältig. Hier können auf ziviler Seite die Überprüfung technischer Systeme wie z. B. die Erkennung und Kontrolle von Hochspannungsleitungen [3], [4] und Gleisen, Zählen von Tierherden [5], geologische Untersuchungen und bei hoheitlichen Aufgaben die Unterstützung sicherheitstechnischer Anforderungen bei Großveranstaltungen, Grenzschutz und Katastrophenszenarien wie z. B. Waldbrände, Überflutungen aufgezählt werden. In der militärischen Nachsorge hilft die Betrachtung aus der Höhe, große Gebiete auf Kontaminierung abzusuchen und zu säubern. An Projekten zur Waldbrandbekämpfung [6] und Minenräumung [7] war die FernUniversität in Hagen beteiligt.

Während Hubschrauber oder Flugzeuge viel Energie benötigen, um eine Aerodynamik aufzubauen, damit sie über dem Boden bleiben, hat ein Luftschiff, weil es leichter ist als Luft, also die Kräfte der Aerostatik nutzt, den Vorteil, mit wenig Antrieb eine erheblich höhere Missionsdauer ohne Energienachschub zu erreichen und es stürzt selbst bei einem Antriebsausfall nicht ab. Flugzeuge überfliegen zu schnell ein genauer und länger zu untersuchendes Gebiet, während Hubschrauber mit ihrem Rotorabwind gerade die sensiblen Flächen, die es zu beobachten gilt, stören oder sogar zerstören. Ein weiterer Vorteil eines Luftschiffes ist seine weitgehende

Geräuschlosigkeit, die den Einsatz in urbanen Gebieten erleichtert. Allerdings ist die Zuladung im Verhältnis zum Volumen klein. So trägt ein m^3 Helium in Luft bei Normalatmosphäre etwa 1 kg an Last. Ein großes Luftschiff kann zwar eine hohe Nutzlast tragen, hat aber den Nachteil, dass das sturmfeste und vandalismussichere Parken eine große Halle erfordert und der Transport zu einem entfernten Einsatzort sehr aufwendig ist.

Die Bedingungen der langen Verweildauer in der Luft, des Langsamfluges und des nahezu unhörbaren Antriebs erfüllt ein Prallluftschiff mit Elektroantrieb. Ein Prallluftschiff, auch Blimp genannt, ist ein Luftschiff, dessen Gestalt, im Gegensatz zum Starrluftschiff, nicht durch ein Innenskelett, sondern durch die Form der Hülle und den Innendruck aufrechterhalten wird. Dadurch lässt es sich im leeren Zustand falten und leicht transportieren. Der Einsatz in weit entfernte Gebiete ist dadurch problemlos möglich. Andererseits weist ein kleines Luftschiff, Bild 4, eine wesentlich größere Dynamik auf als ein großes Luftschiff, Bild 5.



Bild 4: Blimp der FernUniversität in Hagen

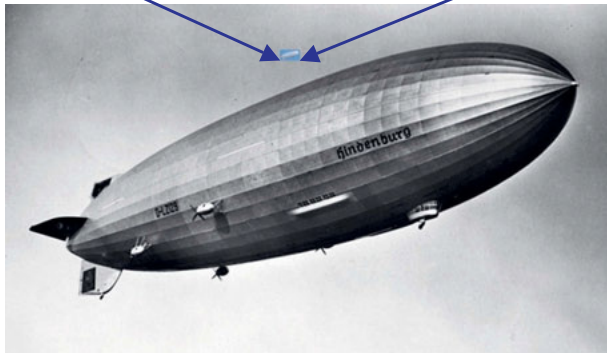


Bild 5: Luftschiff Hindenburg im Größenvergleich zum Blimp der FernUniversität

Quelle: <http://www.alt-deutschland.com/0119Luftschiff1937gross.jpg>

Tabelle 1: Gegenüberstellung des Blimps der FernUniversität in Hagen zum Luftschiff Hindenburg

	FernUni Blimp	Hindenburg (aus [8])
Länge	9 m	245 m
Max. Durchmesser	2.5 m	41 m
Volumen	24 m ³	200 000 m ³
Füllgas	Helium	Wasserstoff
Eigenmasse	17 kg	119 t
Nutzlast	6 kg	96 t
Max. Geschwindigkeit	45 km/h	137 km/h

Grob kann aus Tabelle 1 das Verhältnis der seitlichen Querschnittsflächen zu 500 und das Verhältnis der Massen zu 10000 errechnet werden. Die Windkraft F_w ist proportional zur Fläche. Mit $F_w = mb$, m ist die Masse, kann das Verhältnis der Beschleunigungen b , die die beiden Luftschiffe bei einer Windböe erfahren würden, zu 20 errechnet werden. Das heißt, würde die Hindenburg bei einer Windböe innerhalb einer Sekunde um einen Meter versetzt werden, so würde der Blimp bei derselben Windböe um 20 m zur Seite gehen. Dies könnte bei vertikalen Böen für den Blimp sogar fatal sein. Aus dieser Betrachtung ergibt sich eine besonders hohe Anforderung an die Manövrierfähigkeit und die Positionsregelung des Blimps. Hierzu werden geeignete Sensoren und deren Datenauswertung benötigt.

Im Gegensatz zu landbasierten Fahrzeugen haben Fluggeräte Freiheitsgrade mit großen Auslenkungen. Dies sind drei translatorische Achsen (vorwärts/rückwärts, seitlich und auf/ab) und drei rotatorische Achsen (gieren, rollen und nicken). Diese sechs Achsen zu beherrschen, ist eine Herausforderung für den fernsteuernden Piloten. Denn um Areale zu beobachten, müssen vorgegebene Pfade abgeflogen oder Positionen eingehalten werden. Durch seitliche Luftströmungen oder Aufwinde sind diese Aufgaben sehr ermüdend und schwer zu bewältigen. Erschwerend kommt hinzu, dass die genaue Position des Luftschiffs aus großer Entfernung für den Piloten am Boden kaum erkennbar ist. Der Pilot muss deshalb durch Regelsysteme mit geeigneten Sensoren und Aktuatoren entlastet werden.

Am Lehrgebiet Prozesssteuerung und Regelungstechnik der FernUniversität in Hagen wird an bodengestützten Robotern und Sensoren, luftgestützten Systemen und deren Kombination zu einem Verbund geforscht. In jüngerer Zeit wird ein spezieller Luftschiff-Typ (Blimp) als Flugroboter eingesetzt. Zielsetzung ist dabei einerseits die Erforschung von Sensor- und Regelungssystemen für einen semi-autonomen Blimp, um ihn als Flugroboter zu nutzen, und andererseits die Kopplung mit den vorhandenen bodengebundenen Robotern oder Sensoren zu einem effektiven Tandem-System zu realisieren.

Ziel dieser Arbeit ist, einen Beitrag zur autonomen Navigation eines Blimps zu erbringen.

Zusammenfassung

Die Luftschiffahrt begann Ende des 18. Jahrhunderts und ruhte weitgehend ab Beginn des zweiten Weltkriegs.

Die Vorteile eines Luftschiffs sind geringe Beeinflussung des Betrachtungsobjektes und geringe Energieumsetzung. Im Gegensatz zu Landfahrzeugen hat ein Luftschiff sechs Freiheitsgrade, die vom Piloten oder einer Regelung beherrscht werden müssen. Dadurch, dass das Luftschiff der FernUniversität klein ist, kann es leicht transportiert werden, ist dadurch aber auch sehr windempfindlich, wodurch die Steuerung erschwert wird.

1.2 Aufbau der Bilderfassung

Zwei nach unten gerichtete Kameras nehmen die Szene unterhalb des Fluggerätes auf. Die mit linker und rechter Kamera gleichzeitig aufgenommenen Bilder liefern den Abstand zum Boden. Die Verschiebung von korrespondierenden Bildpunkten zwischen dem aktuellen und dem vorhergehenden linken Bild wird über Bildverarbeitung ermittelt.

Mit dem Abstand, den errechneten Verschiebungswerten und den Winkeldaten aus der IMU wird der zurückgelegte Pfad berechnet, s. Bild 6 und Bild 7.

Da die Aufnahmen in definierten Zeitabschnitten erfolgen, können dann auch die momentanen Geschwindigkeitsvektoren bestimmt werden.

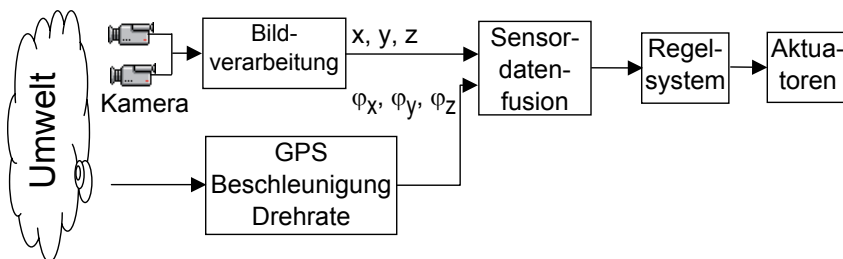


Bild 6: Sensoren – Verarbeitung – Regelung – Aktuatoren

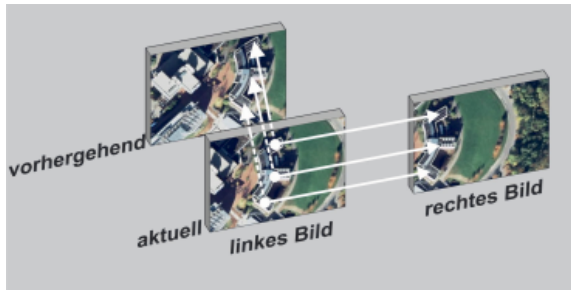


Bild 7: Korrespondenzsuche vom linken aktuellen Bild zum rechten aktuellen Bild und zum linken vorhergehenden Bild

1.3 Stand der Technik

Die multisensorgestützte autonome Navigation von unbemannten Fluggeräten (UAS, unmanned aircraft systems) rückt seit einigen Jahren zunehmend in das Forschungsinteresse der Robotik. Insbesondere die Miniaturisierung von Sensorsystemen erlaubt hier neue technische Ansätze, die bisher auf Flugdrohnen mit geringer Zuladung nicht implementierbar waren.

Das Standardsystem zur Lokalisierung im Außenbereich, das GPS, wird als Basissystem in den meisten mobilen Robotern, die sich außerhalb von Gebäuden bewegen, zur Positionsbestimmung benutzt. Da GPS nur die absolute Position zu einem fiktiven Rotationsellipsoid (Erde) angibt, bleibt der Abstand eines Flugsystems, hier des Blimps, zum Boden zunächst unbekannt. Zudem ist der GPS-Empfang alleine wegen Verfügbarkeitslücken zu unsicher. Die Wiederholfrequenz ist für eine ausreichend schnelle Regelung zu gering und Positionsangaben mit Genauigkeiten besser als 3 m sind nur mit zusätzlicher Infrastruktur (DGPS)[9] erreichbar.

Für kurzzeitige relative Positionsbestimmungen werden zusätzlich Inertialsysteme (IMU, Inertial measurement unit) eingesetzt. IMUs erfassen die Beschleunigung in drei zueinander orthogonalen Achsen und die Drehrate in drei Achsen. Die relative Position wird durch zweifache Integration der Beschleunigungen ermittelt. Diese Integrationen führen nach einiger Zeit zu großen Abweichungen [10].

Neben Inertialsensoren und GPS können Kameras die Multisensorregelungen ergänzen. Bildauswertungen sind bei autonomen Fluggeräten insbesondere dann nützlich, wenn es darum geht, langsam eine genaue Route abzufliegen oder gar eine Position zu halten, und besonders geeignet, weil Kameras sich

durch geringes Gewicht, kleine Leistungsaufnahme und minimales Volumen auszeichnen.

Für die Positionsbestimmung der Flugnavigation über Bildverarbeitung werden folgende Methoden unterschieden [11]:

- **Beobachtung vom Boden (On-ground vision)**

Mit mehreren Kameras am Boden wird die Position des Fluggerätes bestimmt.

- **Visuelle Odometrie (Visual odometry)**

Über Bildsequenzen wird die kumulierte Position errechnet. Dies ist das Prinzip, das in dieser Arbeit vorgestellt wird.

- **Auf ein Ziel bezogene Navigation (Target relative navigation)**

Hier wird die Position relativ zu einem feststehenden Ziel (z. B. Landplatz) oder auch einem bewegten Ziel (z. B. Fahrzeug) erfasst.

- **Gelände oder Landmarken bezogene Navigation (Terrain/landmark relative navigation)**

Bekannte Geländeaufnahmen oder Landmarken werden mit den momentanen Aufnahmen verglichen und damit die absolute Position bestimmt. Dies kann die visuelle Odometrie unterstützen, indem hin und wieder die absolute Position neu bestimmt wird. Ein Verfahren hierzu wird weiter unten näher beschrieben.

- **Simultane Schätzung der Bewegung und der Struktur (Concurrent estimation of motion and structure)**

Ein bekanntes Verfahren ist die simultane Lokalisierung und Kartenerstellung (engl. SLAM). Hier wird die Position zu einem Merkmal geschätzt und in eine Karte eingetragen. Mit zunehmender Anzahl der geschätzten Positionen wird die eigene Positionsbestimmung genauer.

- **Bio-inspirierter optischer Fluss (Bio-inspired optic flow navigation)**

Ein Beispiel ist die Insektennavigation, die über den optischen Fluss stattfindet, der mit der Nähe zu einer Szene zunimmt und damit eine Abstandsschätzung ermöglicht.

Das Problem der Navigation wird in der Roboterforschung oft im Zusammenhang mit gleichzeitiger Lokalisation und Kartenerstellung, engl. „simultaneous localization and mapping“ (SLAM) [12], [13], [14], [15], erwähnt. Die dort vorgeschlagenen Verfahren sind meist für geschlossene Räume

entwickelt und reduzieren die Lokalisation auf eine zweidimensionale Umgebung. Das Verfahren wird überwiegend für landbasierte Fahrzeuge benutzt, findet aber zunehmend Eingang im Zusammenhang mit unbemannten Fluggeräten.

Prinzipiell werden bei SLAM sowohl Positionen von Merkmalen und die jeweilige Position der Kamera geschätzt. Die Schätzungen werden umso genauer, je öfter dieselben Merkmale aus verschiedenen Blickwinkeln wiedergefunden werden. Dabei werden die Schätzungen zumeist über ein Kalmanfilter zusammengeführt. Zusammen mit anderen Sensoren kann dann auch eine Monokamera zu einer verbesserten Positionsbestimmung genommen werden. In [16] wird in einem simulierten Modellhubschrauberflug ein mittlerer quadratischer Fehler des SLAM-Verfahrens von 2 m erreicht. Die Flugdauer war 35 s bei einer Strecke von etwa 10 m.

In [17] werden Aufnahmen bei einem 90 m Flug bei der Integration der SLAM und IMU Daten in ein Kalmanfilter eine Genauigkeit von 6 m erreicht.

Über visuelle Odometrie durch eine Monokamera mit Unterstützung von INS-Daten und barometrischer Höhenbestimmung wird in [18] bei einem Flug in 70 m Höhe auf einer 280 m langen Strecke eine maximale Positionsabweichung von 35 m erreicht. Allerdings ist nach 10 s des 70 s-Fluges der Fehler schon bei 30 m. Da mit der barometrischen Höhenbestimmung der Abstand zu von der Kamera beobachteten Objekten nur grob geschätzt werden kann, wurde eine flache Ebene vorausgesetzt.

Im Zusammenhang mit Luftschiffen sind nur wenige Arbeiten veröffentlicht, die sich mit Bildverarbeitung beschäftigen. Dazu zählen die Dissertation von M. Fach [19], Untersuchungen an der Universität Chemnitz [20] [21], die Dissertation von La Gloria [22] und einige Arbeiten am Institute of Systems and Robotics der Universität Coimbra, Portugal, die sich in der Dissertation von L. Mirisola wiederfinden [23].

M. Fach [19] berechnet mit Aufnahmen einer Monokamera den Abstand der Kamera zur Szene. Über Harris-Detektion werden Merkmale verfolgt und mit Hilfe der aus einer IMU gewonnenen Daten der Abstand berechnet. Über die Genauigkeit des optischen Verfahrens sind kaum Ergebnisse angegeben. Bei realen Versuchen wird nur auf die IMU eingegangen.

Die Arbeit von S. Sünderhauf et al. [21] bezieht sich auf ein Luftschiff. Durch Simulation ist gezeigt worden, dass mit Hilfe eines Kalmanfilters die Fusion von IMU-Daten mit Bildauswertungen die Positionsbestimmung verbessert. Für die Bildverarbeitung wurde eine Monokamera angenommen, die sich über Kurzeit-SLAM-Methode selbst lokalisierte. Hierbei wurden die Positionen der detektierten Landmarken mit ihren Positionsunsicherheiten als gegeben vorausgesetzt. Der Einsatz einer Stereokamera wurde ausgeschlossen, da die Basisbreite der Kamera mit etwa einem Meter veranschlagt wurde.

Ebenfalls mit Simulation bei Modellierung des Luftschiffs und der Kamera kommt La Gloria [22] zu theoretischen Ergebnissen, wobei eine praktisch schwer realisierbare Stereobasisbreite von 2 m vorausgesetzt wurde. Die Detektion von Landmarken durch die Kamera wurde als gegeben vorausgesetzt. Der simulierte Flug in 200 m Höhe und über eine Strecke von 2 km ergab einen Fehler von 10 m.

Mit Hilfe einer Monokamera, dem SURF-Algorithmus [24] für die Bildregistrierung und den Winkeldaten eines Inertialsystems ermittelt L. Mirisola [23] die Bewegung eines Luftschiffs. Hierbei wird über die projektive Transformation jede Bildebene nach Norden und senkrecht zum Lot ausgerichtet. Kurzzeitig kann damit der Flugpfad des Luftschiffs nachgebildet werden. Allerdings war die Höhenbestimmung ungenau, so dass der Pfad zwar in der Form, aber nicht in der Skalierung, richtig wiedergegeben wurde. Trotzdem verbesserte die Fusion der Kameradaten mit den Messwerten des Inertialsystems über ein Kalmanfilter die Ergebnisse.

Ein Ansatz, die absolute Position zu bestimmen, ist in einer im Zusammenhang mit der vorliegenden Arbeit erstellten Masterarbeit zu finden [25]. Hier werden über den SIFT-Algorithmus [26] georeferenzierte Luftaufnahmen des Landes Nordrhein-Westfalen mit Aufnahmen von Google Earth verglichen. Ist die Position grob bekannt, kann mit dieser Methode über einen georeferenzierten Kartenausschnitt und einer momentanen Bildaufnahme der Kamera des Luftschiffs die absolute Position des Luftschiffs bestimmt werden. Unter der Annahme einer Kamera mit einem horizontalen Öffnungswinkel von 48° bei einer Auflösung von 1280 px ergeben sich bei dem Verfahren Positionsabweichungen von ± 20 px, entsprechend ± 0.6 m bei 50 m Höhe.

Ein ähnlicher Ansatz ist in [27] zu finden.

In dieser Arbeit werden Algorithmen untersucht, die schnell und genau sind und weitgehend unabhängig vom Bildinhalt eine vorhersagbare Berechnungszeit benötigen. Deskriptorenmethoden wie SIFT und SURF haben dann Vorteile, wenn die zu vergleichenden Aufnahmen in der Perspektive, Skalierung und Rotation stark unterschiedlich sind. Die Berechnungen sind aufwendiger und vom Bildinhalt abhängig, da besondere Merkmale gesucht werden.

Durch die schnelle Bildfolge und die Gleichzeitigkeit bei Stereoaufnahmen sind diese Einflüsse bei flächenbasierten Methoden durch unterschiedliche Blickwinkel gering. Eine Monookularkamera wird ausgeschlossen, da die genaue Höhe über Boden, deren Kenntnis im Landeanflug besonders wichtig ist, damit nicht genau erfasst werden kann.

1.4 Aufbau der Arbeit

In diesem **Kapitel** wurde auf die Geschichte der Luftschiffahrt und kurz auf die Dynamik des kleinen Luftschiffs der FernUniversität in Hagen im Vergleich zur Dynamik eines Großluftschiffs eingegangen. Die Berechnung der Höhe erfolgt mit zwei gleichzeitig aufgenommenen Abbildungen, während die Bewegung mit einer zusätzlichen und zeitversetzten Aufnahme ermittelt wird. Die Fusion der Ergebnisse in das Multisensorsystem wurde grafisch dargestellt und ist nicht Teil dieser Arbeit. Bisherige Arbeiten, insbesondere im Zusammenhang mit Luftschiffen, wurden erläutert.

Im folgenden **Kapitel 2** werden Grundlagen zur Geometrie einer Kamera bezüglich der Position, Abbildungen, linsenbedingten Verzeichnungen und deren Behebung beschrieben. Weitere Abbildungsfehler durch Objektive und Sensoren, die die Ergebnisse der Bildregistrierung beeinflussen, werden erklärt. Die geometrischen Verhältnisse bei Stereokameraanordnungen und deren Überführung in eine scheinbar planparallele Anordnung durch Rektifizierung werden modelliert und ermöglichen aus Bildpunkten auf Weltpunkte zu schließen, um Bewegungen und Abstände zu ermitteln.

Das **3. Kapitel** widmet sich der Bildregistrierung. Es geht um die Frage: „Wie hat sich ein homologer¹ Bildpunkt bewegt?“ Zur Beantwortung werden Verfahren klassifiziert und kurz vorgestellt, um sich dann auf zwei Methoden – Gradientenverfahren nach Lucas&Kanade und Kreuzkorrelation – zu konzentrieren. Durch die Umformung der Gleichungen für die Kreuzkorrelation werden mit Hilfe von Summentabellen zügige Berechnungen ermöglicht. Die Verfahren werden in ihrer Funktion mit ihren Vor- und Nachteilen verglichen, um dann anhand von Versuchen mit Abbildungen im folgenden Kapitel eine Entscheidung zu treffen.

Das **Kapitel 4** stellt u. a. anhand realer Luftaufnahmen die Verfahren hinsichtlich ihrer Bewegungserfassung bei Beleuchtungsunterschieden, Rotationen und großen Translationen dar. Die Subpixelgenauigkeit, die aufgrund der geringen Basisbreite der Stereokamera sehr wichtig ist, wird eingehend untersucht. Über eine eigene Methode (Bildverkleinerungen) werden exakte Bildverschiebungen erzeugt, um die Subpixelgenauigkeit zu prüfen. Dabei wird eine neue Funktion eingeführt, die die Subpixelgenauigkeit signifikant erhöht.

Mit diesen Ergebnissen werden in **Kapitel 5** Versuche durchgeführt. Die ersten Versuche finden in reproduzierbaren Umgebungen, d. h. in einer virtuellen Welt und mit Hilfe eines XY-Tischs statt.

Flugaufnahmen mit einer Monokamera werden anhand von Mosaikbildern überprüft.

Eingehend wird die Bilderfassung der Stereokamera hinsichtlich der Kalibrierung und Rektifizierung auf Genauigkeit verifiziert. Es werden Verfahren und Vereinfachungen zur schnelleren Berechnungen von Strecken, Geschwindigkeiten und Höhen eingeführt und auch gezeigt, wie Verdeckungen detektiert und überbrückt werden können. Den Abschluss der Versuche bilden Fahrten entlang von Gebäuden mit Verdeckungen und entlang einer Böschung zur Bestimmung von Geschwindigkeiten und Abständen.

¹ Homologe Bildpunkte sind Bildpunkte in verschiedenen Aufnahmen, die zum selben Szenenpunkt gehören.

Im **Kapitel 6** werden verschiedene Hardwareplattformen auf ihre Eignung für Echtzeitberechnungen untersucht. Dabei wird auf die Parallelisierbarkeit des gewählten Verfahrens zur Bildregistrierung eingegangen.

Nach der **Zusammenfassung** folgt der **Anhang**, in dem Beugungseffekte bezüglich der Lochkamera, Vereinfachungen von Filtern, Summentabellen, NCC über Schwarzsche Ungleichung erklärt und Messverfahren hinsichtlich des Zeitverhaltens der Stereokamera erläutert werden.

Das **Literaturverzeichnis** ist nach der Reihenfolge des ersten Zitierens sortiert.

Die **Quellen der Abbildungen** sind, sofern es keine eigenen Bilder sind, jeweils darunter angegeben.

2 Kameramodell und Bewegung

Eine Kamera bildet eine dreidimensionale Umwelt auf eine zweidimensionale Ebene ab. Üblicherweise erfolgt die Beleuchtung durch sichtbares Licht.

Mit vielen in kurzen Zeitabständen gespeicherten Aufnahmen lassen sich Bewegungsabläufe festhalten. Üblich sind bei kommerziellen Kameras etwa 30 Bilder pro Sekunde. Extremwerte sind z. B. die Aufnahme eines Pflanzenwachstums mit einem Bild pro Stunde oder ballistische Untersuchungen mit 10^8 Bildern pro Sekunde.

2.1 Modell der Lochkamera

Das einfachste Kameramodell wird durch die Lochkamera dargestellt.

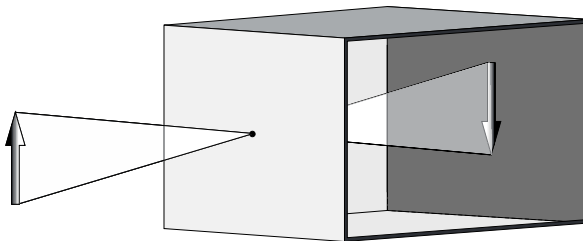


Bild 8: Lochkamera

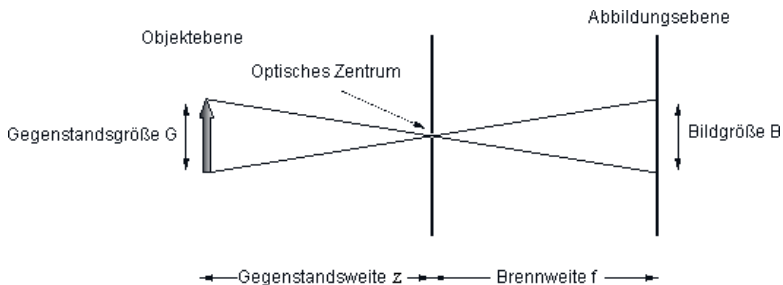


Bild 9: Lochkamera im seitlichen Aufriss

Das optische Zentrum der Lochkamera ist die Stelle des Lichteinlasses.

Die Brennweite f ist der Abstand vom optischen Zentrum zur Abbildungsebene. Die Gegenstandsweite z ist der Abstand vom Gegenstand zum optischen Zentrum.

2.1.1 Geometrische Transformationen

Ziel ist, aus den Bildkoordinaten eines Punktes den korrespondierenden Punkt in Weltkoordinaten zu finden. Während durch Projektion die Überführung einer 3D-Szene in ein 2D-Bild möglich ist, kann aus einer 2D-Position nur der Strahl, auf dem sich der korrespondierende 3D-Punkt befindet, bestimmt werden. Die Bestimmung des Abstands z_c wird in einem späteren Abschnitt erläutert.

Bei der Kamerakalibrierung wird zwischen der inneren (intrinsischen) und äußeren (extrinsischen) Geometrie der Kamera unterschieden. Während die extrinsischen Parameter die Position und Orientierung der Kamera im Raum angeben, beschreiben die intrinsischen Parameter den Zusammenhang zwischen Kamera- und Bildkoordinatensystem. Im Folgenden werden die Indizes wie nachstehend benutzt:

- b: Bildkoordinaten, Bezug ist die Bildebene (Sensor).
- c: Kamerakoordinaten x_c, y_c, z_c , Bezug ist das Projektionszentrum.
 x_c, y_c sind parallel zu x_b, y_b .
 z_c zeigt in Richtung der Szene.
- w: Weltkoordinaten. Sie liegen beliebig im Raum.

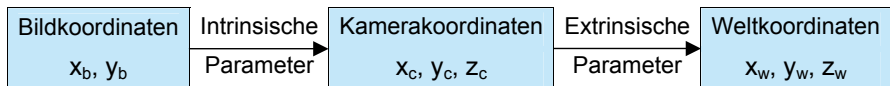


Bild 10: Beziehung zwischen Kameraparametern und den verschiedenen Koordinatensystemen

Im Gegensatz zu Bild 9 ist im Folgenden die Bildebene in Positivlage, d. h., sie liegt zwischen dem optischen Zentrum und der Objektebene. Die Positivlage ändert bis auf Vorzeichen nichts an der Geometrie. Alle Lichtstrahlen gehen durch das optische Zentrum. Dort liegt der Ursprung des orthogonalen Kamerakoordinatensystems. Die z-Achse des Kamerakoordinatensystems geht am Bildhauptpunkt senkrecht durch die Bildebene in Richtung Objektebene und ist die optische Achse. Die Brennweite f ist der Abstand zwischen dem optischen Zentrum und dem Bildhauptpunkt. Die x-Achse und y-Achse des Bildkoordinatensystems sind parallel zu den entsprechenden Kamerakoordinaten. Ein vom Weltpunkt P_w ausgehender Lichtstrahl geht

durch das optische Zentrum und bildet dabei einen Punkt auf der Bildebene ab, s. Bild 11.

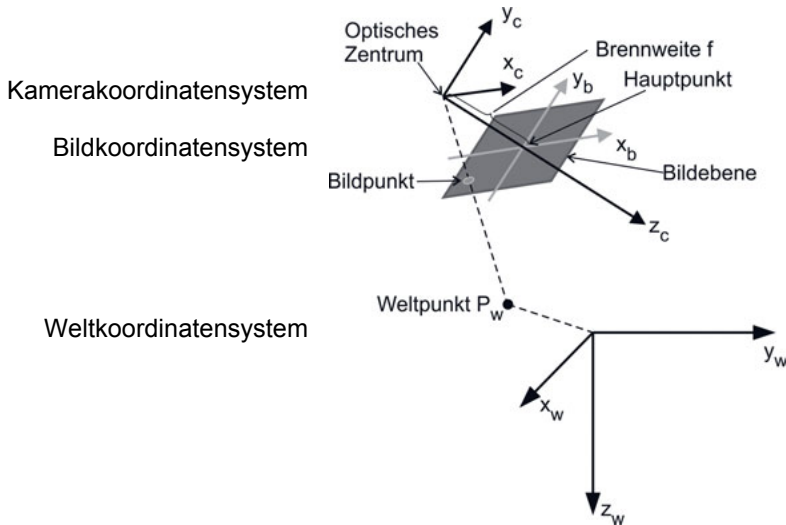


Bild 11: Projektion des Punktes P_w auf die Bildebene einer Lochkamera

Zunächst seien die intrinsischen Parameter anhand der Zentralperspektive, Bild 12, betrachtet.

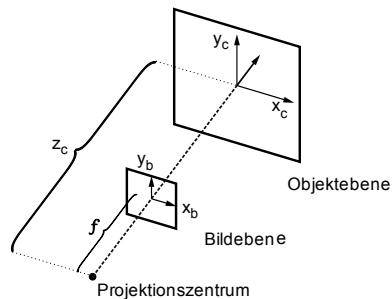


Bild 12: Zentralperspektive

Über die Strahlensätze sind die Verhältnisse der Abbildung gegeben mit

$$x_b = \frac{f_x x_c}{z_c} ; y_b = \frac{f_y y_c}{z_c} \quad (2.1)$$

Unter Verwendung homogener Koordinaten ergibt sich

$$\begin{pmatrix} x_b \\ y_b \\ 1 \end{pmatrix} = \frac{1}{z_c} \begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} \quad (2.2)$$

In der Regel sind die Pixel eines Sensors gleichverteilt und es gilt für die Brennweiten:

$$f_x = f_y = f \quad (2.3)$$

Im Allgemeinen befindet sich das optische Zentrum nicht in der Abbildungsmitte. Gleichung (2.2) wird dann zu

$$\underbrace{\begin{pmatrix} x_b \\ y_b \\ 1 \end{pmatrix}}_{X_b} = \frac{1}{z_c} \underbrace{\begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix}}_K \underbrace{\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}}_{X_c} \quad (2.4)$$

wobei c_x , c_y jeweils der Versatz des optischen Zentrums von der Bildmitte ist. K ist die Kamerakalibriermatrix der linearen unverzerrten Abbildung [28] [29]. Um die Position eines Punktes in Kamerakoordinaten (oder auch Weltkoordinaten) aus einem Bildpunkt zu gewinnen, sind Rückschlüsse von Bildkoordinaten auf Kamerakoordinaten erforderlich. Die Umkehrung der Gleichung (2.4) ist allerdings nicht eindeutig, wenn nur die Bildkoordinaten bekannt sind, denn es kann zunächst ohne Kenntnis von z_c nur der Strahl bestimmt werden, auf dem sich der Szenenpunkt befindet.

$$X_c = z_c K^{-1} X_b \quad (2.5)$$

Mit

$$K^{-1} = \begin{pmatrix} \frac{1}{f} & 0 & -\frac{c_x}{f} \\ 0 & \frac{1}{f} & -\frac{c_y}{f} \\ 0 & 0 & 1 \end{pmatrix} \quad (2.6)$$

folgt:

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = z_c \begin{pmatrix} \frac{x_b - c_x}{f} \\ \frac{y_b - c_y}{f} \\ 1 \end{pmatrix} \quad (2.7)$$

Über die extrinsische Kamerakalibrierung kann eine Beziehung zwischen einer beliebigen Kameraposition und -orientierung und einem Weltkoordinatensystem modelliert werden.

Da sich die Kamera auf einem Luftschiff befindet, ist es sinnvoll, das an die Luftfahrt angelehnte Koordinatensystem zu nehmen. Nach den Luftfahrtnormen DIN 9300 und ISO 1151 sind die Koordinaten rechtwinklig, wobei die x-Achse nach vorne, die y-Achse nach rechts und die z-Achse nach unten weist. Im englischen Sprachraum wird dieses Koordinatensystem als NED-System (**N**orth – **E**ast – **D**own) bezeichnet.

Für die Rotation ist in der Luftfahrtnorm festgelegt, dass zuerst um die z-Achse, dann um die y-Achse und schließlich um die x-Achse im Uhrzeigersinn (Rechte-Hand-Regel) gedreht wird. Die Drehreihenfolge wird auch mit YPR (Yaw-Pitch-Roll) benannt, s. Bild 13.

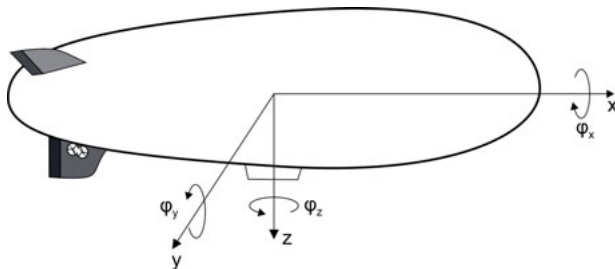


Bild 13: Koordinatensystem in NED –Darstellung

Gierwinkel φ_z , Nickwinkel φ_y und Rollwinkel φ_x in YPR-Darstellung

Die extrinsische Kamerakalibrierung ergibt sich aus der Koordinatentransformation von Weltkoordinaten in Kamerakoordinaten und ist gegeben durch:

$$X_c = R_{wc} X_w + T \quad (2.8)$$

X_c und X_w sind die Koordinaten desselben 3D-Punktes. T ist die Translation vom Weltkoordinatensystem zum Kamerakoordinatensystem.

Die Rotation R_{wc} vom Weltkoordinatensystem zum Kamerakoordinatensystem ist:

$$R_{wc} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_x & \sin \varphi_x \\ 0 & -\sin \varphi_x & \cos \varphi_x \end{pmatrix} \begin{pmatrix} \cos \varphi_y & 0 & -\sin \varphi_y \\ 0 & 1 & 0 \\ \sin \varphi_y & 0 & \cos \varphi_y \end{pmatrix} \begin{pmatrix} \cos \varphi_z & \sin \varphi_z & 0 \\ -\sin \varphi_z & \cos \varphi_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.9)$$

Mit der invertierten Gleichung (2.8) erhält man den Bezug von Kamerakoordinaten zu Weltkoordinaten:

$$X_w = R_{wc}^T X_c - R_{wc}^T T \quad (2.10)$$

Gleichung (2.8) ist allgemeingültig. Mit ihr lassen sich die Kamerakoordinaten in das für die Regelung wichtige Volumenzentrum [30] oder in den Rotationspunkt [10] des Luftschiffes transformieren. Weiterhin kann die Kamera mit Hilfe der aus der IMU gewonnenen Winkel scheinbar lotrecht positioniert werden, um Trajektorien und Geschwindigkeiten in Bezug auf den Boden, z. B. am Ort der Bodenstation, zu berechnen.

Während die Koordinatentransformation der Kamera zum Volumenzentrum nur einmal berechnet wird, muss die weitere Transformation zur Bodenstation bei jeder Bewegung des Luftschiffs neu berechnet werden.

Diese Berechnung kann wesentlich verkürzt werden, da die Dynamik des Blimps, auch bei Windböen, träge genug ist, um für die Rotationen kleine Winkel anzusetzen.

Die Winkelgeschwindigkeiten des Blimps der FernUniversität sind maximal:

- Gieren 20°/s
- Nicken 2.5°/s
- Rollen 2.5°/s

Wobei das Rollen durch eine Regelung auf $\pm 3^\circ$ eingeschränkt ist.

Bei einer Bildrate von 10 oder mehr Bildern pro Sekunde sind die größten Winkel zwischen zwei Aufnahmen kleiner als ein Zehntel der oben angegebenen Zahlen. Somit kann mit geringem Fehler angenommen werden, dass $\sin \varphi = \varphi$ und $\cos \varphi = 1$ ist, (φ in rad). Der sich ergebende Fehler für x_w , y_w , z_w bei gleichen Längen der Luftschiffkoordinaten x_L , y_L , z_L ist dann kleiner $2 \cdot 10^{-5}$.

Die Koordinatentransformation von einem Punkt X_L zu einem Punkt X_w in homogener Matrix dargestellt, sieht folgendermaßen aus:

$$X_w = \begin{pmatrix} 1 & \varphi_x \varphi_y - \varphi_z & \varphi_y + \varphi_x \varphi_z & t_x \\ \varphi_z & \varphi_x \varphi_y \varphi_z + 1 & \varphi_y \varphi_z - \varphi_x & t_y \\ -\varphi_y & \varphi_x & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_L \\ y_L \\ z_L \\ 1 \end{pmatrix} \quad (2.11)$$

Die Gesamtkoordinatentransformation von Bildkoordinaten mit bekanntem z_c zu den Weltkoordinaten der Bodenstation ist:

$$X_b, z_c \rightarrow X_c \rightarrow X_L \rightarrow X_w$$

In Bild 14 sind die Kamera-, Luftschiff- und die Weltkoordinaten (X_c , X_L , X_w) mit den Richtungen der z-Achsen gekennzeichnet.

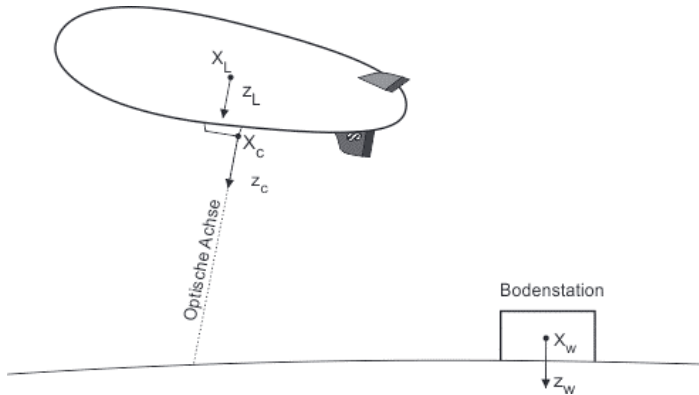


Bild 14: Position des Luftschiffs mit den beteiligten z-Achsen

2.1.2 Kalibrierung bei unverzerrter Abbildung

Ausgehend von Gleichung (2.8) ist der Zusammenhang zwischen Bild- und Weltkoordinaten so herzustellen:

$$sX_b = PX_w \quad (2.12)$$

wobei P die Projektionsmatrix, die sowohl die intrinsischen wie die extrinsischen Parameter enthält, ist:

$$P = K(R | T) \quad (2.13)$$

s ist ein Skalierungsparameter.

Gleichung (2.12) wird geschlossen dargestellt, indem P durch 12 Parameter ersetzt wird:

$$\begin{pmatrix} x_b s \\ y_b s \\ s \end{pmatrix} = \begin{pmatrix} L_1 & L_2 & L_3 & L_4 \\ L_5 & L_6 & L_7 & L_8 \\ L_9 & L_{10} & L_{11} & L_{12} \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (2.14)$$

Ein Vielfaches von X_w würde denselben Bildpunkt X_b erzeugen, deshalb kann $L_{12} = 1$ gesetzt werden.

s wird dann zu:

$$s = L_9 x_w + L_{10} y_w + L_{11} z_w + 1 \quad (2.15)$$

Gleichung (2.14) wird nun umformuliert zu:

$$\begin{aligned} x_b &= \frac{L_1 x_w + L_2 y_w + L_3 z_w + L_4}{s} \\ y_b &= \frac{L_5 x_w + L_6 y_w + L_7 z_w + L_8}{s} \end{aligned} \quad (2.16)$$

Durch eine weitere Umstellung erhält man:

$$\begin{aligned} x_b &= L_1 x_w + L_2 y_w + L_3 z_w + L_4 - L_9 x_b x_w - L_{10} x_b y_w + L_{11} x_b z_w \\ y_b &= L_5 x_w + L_6 y_w + L_7 z_w + L_8 - L_9 y_b x_w - L_{10} y_b y_w + L_{11} y_b z_w \end{aligned} \quad (2.17)$$

Stehen nun n Punktepaare $(x_{bi}, y_{bi}) \leftrightarrow (x_{wi}, y_{wi}, z_{wi})$, $i = 1, 2, \dots, n$ zur Verfügung, kann ein überbestimmtes lineares Gleichungssystem aufgestellt werden:

$$\underbrace{\begin{pmatrix} x_{wl} & y_{wl} & z_{wl} & 1 & 0 & 0 & 0 & 0 & -x_{bl}x_{wl} & -x_{bl}y_{wl} & -x_{bl}z_{wl} \\ 0 & 0 & 0 & 0 & x_{wl} & y_{wl} & z_{wl} & 1 & -y_{bl}x_{wl} & -y_{bl}y_{wl} & -y_{bl}z_{wl} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{wn} & y_{wn} & z_{wn} & 1 & 0 & 0 & 0 & 0 & -x_{bn}x_{wn} & -x_{bn}y_{wn} & -x_{bn}z_{wn} \\ 0 & 0 & 0 & 0 & x_{wn} & y_{wn} & z_{wn} & 1 & -y_{bn}x_{wn} & -y_{bn}y_{wn} & -y_{bn}z_{wn} \end{pmatrix}}_A \begin{pmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11} \\ L \end{pmatrix} = \begin{pmatrix} x_{bl} \\ y_{bl} \\ \vdots \\ x_{bn} \\ y_{bn} \\ b \end{pmatrix} \quad (2.18)$$

Die optimale Lösung L^* lässt sich mit der Methode der kleinsten Quadrate über die Pseudoinverse von A bestimmen zu:

$$L^* = (A^T A)^{-1} A^T b \quad (2.19)$$

Aus den L-Parametern lassen sich die intrinsischen Parameter c_x , c_y , f_x , f_y und die extrinsischen Parameter R und T berechnen [29].

2.1.3 Geometrie mehrerer Abbildungen einer Szene

Montagebedingt sind die beiden Bildebenen der Stereokamera nicht auf einer Ebene und haben unterschiedliche Orientierungen. Dies führt zu aufwendigen Berechnungen der Zuordnung von Bildpunkten zu Weltpunkten bei der Stereogeometrie. Um dies zu vermeiden, kann über die Epipolargeometrie eine scheinbar planparallele Anordnung der Bildebenen errechnet werden.

Die Epipolargeometrie beschreibt die Beziehung zwischen Abbildungen derselben Szene durch mehrere Kameras und hängt nur von den intrinsischen und extrinsischen Parametern, also der relativen Lage der Kameras, ab. In dieser Arbeit wird diese Beziehung ausschließlich für ein Zweikamerasystem betrachtet, obwohl die Epipolargeometrie für multiple Kameraanordnungen gültig ist [31].

An Bild 15 seien die Verhältnisse der Epipolargeometrie an einem Beispiel gezeigt.

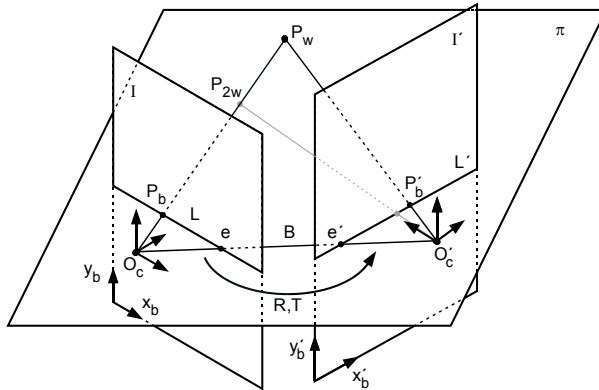


Bild 15: Epipolargeometrie

Legende:

- Bezeichnungen mit Hochstrich, z. B. I' , beziehen sich auf die rechte Kamera.
- I : Bildebene
- x_b, y_b : Bildkoordinaten
- O_c : Optisches Zentrum, Ursprung des Kamerakoordinatensystems
- L : Epipolarlinie
- e : Epipol
- B : Basislinie
- π : Epipolarebene
- R, T : Rotationsmatrix, Translationsvektor
- P_b : Bildpunkt
- P_w : Weltpunkt
- P_{2w} : Zweiter Weltpunkt auf der Linie O_c - P_w

Die beiden Kameras, deren relative Position durch die Rotation R und Translation T gegeben ist, sind einander zugewandt. Der Weltpunkt P_w , dessen Strahl durch das optische Zentrum O_c geht, wird als Punkt P_b auf der Bildebene I der linken Kamera abgebildet. Ziel ist es, den korrespondierenden Bildpunkt P'_b auf der Bildebene I' der rechten Kamera mit möglichst wenig Aufwand zu finden.

Da jeder Punkt auf der Linie O_c-P_w in demselben Punkt P_b abgebildet wird, ist der Schluss von P_b auf P'_b mehrdeutig, was an der Abbildung des Punktes P_{2w} , der in der Bildebene I denselben Punkt P_b erzeugt, aber in der Bildebene I' nicht mehr in P'_b abgebildet ist, deutlich wird.

Dadurch, dass alle Strahlen der abgebildeten Punkte durch die optischen Zentren O_c und O'_c gehen, lässt sich eine Ebene definieren, auf der alle Strahlen, die ausgehend von einem Bildpunkt P_b alle diesbezüglich möglichen Weltpunkte in der Bildebene I' abbilden. Zwei der unendlich vielen Strahlen sind in Bild 15 ausgehend von P_w und P_{2w} gezeigt. Die gesuchte Ebene, die Epipolarebene π , ist durch die drei bekannten Punkte, die optischen Zentren O_c , O'_c und den Bildpunkt P_b , bestimmt. Die Schnittlinien der Epipolarebene mit den Bildebenen I und I' sind die Epipolarlinien L und L'. Jeder Bildpunkt auf der Epipolarlinie L hat seinen korrespondierenden Bildpunkt an unbekannter Stelle auf der Epipolarlinie L'. Dies ist äquivalent mit der Tatsache, dass die Linie, die auf P_b abgebildet wird, zur Linie L' korrespondiert.

Damit ist der Suchbereich für den korrespondierenden Bildpunkt zum eindimensionalen Problem geworden. Alle Weltpunkte auf einer bestimmten Epipolarebene, solange von den Kameras sichtbar, sind in den Bildebenen auf den korrespondierenden Epipolarlinien zu finden. Sind die Weltpunkte nicht auf dieser Ebene, so gibt es eine neue Epipolarebene mit neuen Epipolarlinien. Alle Epipolarebenen haben die Basislinie B gemeinsam und hängen damit von nur einem Parameter ab. Damit gehen alle Epipolarlinien durch die Epipole e bzw. e'.

Den Rückschluss von einem Punkt P_b der einen Bildebene auf die Epipolarlinie L' der zweiten Kamera erhält man über die Fundamentalmatrix F:

$$L' = FP_b \quad (2.20)$$

Die Fundamentalmatrix ist die algebraische Darstellung der Epipolargeometrie und errechnet sich aus Rotation R, Translation T und den Kamerakalibriermatrizen der beiden Kameras K und K'[32][33]:

$$F = K'^{-T} [T]_x RK^{-1} \quad (2.21)$$

Die Translationsmatrix nimmt hierbei folgende schiefsymmetrische Form an:

$$[T]_x = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix} \quad (2.22)$$

Der Ausdruck

$$E = [T]_x R \quad (2.23)$$

ist die Essentialmatrix und enthält die extrinsischen Kameraparameter (Translation und Rotation), also Informationen zu den relativen Positionen der beiden Kameras zueinander. Die Fundamentalmatrix enthält zusätzlich Informationen über die innere Geometrie.

Bild 16 und Bild 17 zeigen die Verhältnisse bei einer in Bezug zur linken Kamera versetzten und gedrehten rechten Kamera. Die Suche beschränkt sich nur auf die korrespondierende Epipolarlinie. Man kann die Suche noch weiter einschränken, indem die gefundenen Positionspaare als Bezug genommen werden. Denn wird nun ein weiterer korrespondierender Bildpunkt auf derselben Epipolarlinie gesucht, so befindet sich dieser in der gleichen Richtung wie im linken Bild. D. h., ein Punkt, der weiter rechts auf derselben Epipolarlinie im linken Bild ist, wird dann, solange er sichtbar ist, im rechten Bild auf der korrespondierenden Epipolarlinie auch weiter rechts gefunden.

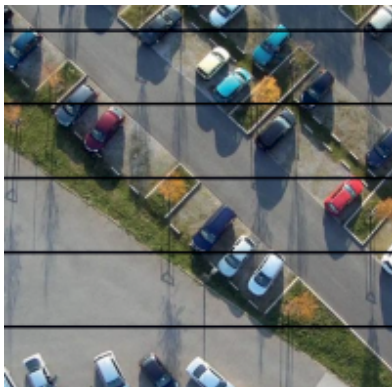


Bild 16: Bild der linken Kamera mit einigen Epipolarlinien

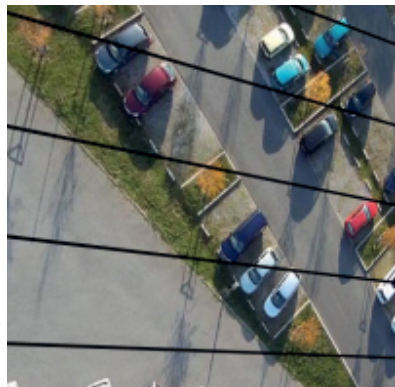


Bild 17: Bild der rechten Kamera mit korrespondierenden Epipolarlinien

2.1.4 Rektifizierung

Ziel der Rektifizierung ist, beide Kamerabilder so zu transformieren, als wären die optischen Achsen parallel, beide Abbildungen auf einer gemeinsamen Ebene liegend und die x-Achsen sich deckend. Dadurch wird erreicht, dass korrespondierende Bildpunkte nur noch auf einer gemeinsamen Linie gesucht werden müssen.

In [34] wird ein Verfahren beschrieben, um zwei Kamerabilder zu rektifizieren. Vorausgesetzt ist, dass die Kalibrierungsparameter der beiden Kameras bekannt sind und beide Kameras dieselben intrinsischen Werte haben, also $K_1 = K_2 = K$ ist.

Die Originalprojektionsmatrizen sind:

$$\begin{aligned} P_{o1} &= K_1 [R_1 | -R_1 C_1] \\ P_{o2} &= K_2 [R_2 | -R_2 C_2] \end{aligned} \quad (2.24)$$

Ziel ist, dass beide Kameras scheinbar dieselbe Orientierung R haben und sich in der Lage nur um den Abstand der beiden optischen Zentren C_1 und C_2 unterscheiden. Damit gilt für die neuen Projektionsmatrizen:

$$\begin{aligned} P_{n1} &= K [R | -RC_1] \\ P_{n2} &= K [R | -RC_2] \end{aligned} \quad (2.25)$$

Für die Rotation

$$R = \begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix} \quad (2.26)$$

gelten folgende Forderungen:

- Die neue x-Achse ist parallel zur Basislinie:

$$r_1 = \frac{C_1 - C_2}{\|C_1 - C_2\|} \quad (2.27)$$

- Die neue y-Achse ist senkrecht zur neuen x-Achse,

$$r_2 = k \times r_1 \quad (2.28)$$

wobei k ein Einheitsvektor ist, der z. B. aus dem z-Einheitsvektor des linken Bildes gebildet wird, wenn dieser als Basis genommen und nur die z-Achse des rechten Bildes verändert wird.

- Die neue z-Achse (hier dann des rechten Bildes) ist orthogonal zu der x- und y-Achse:

$$r_3 = r_1 \times r_2 \quad (2.29)$$

Damit lassen sich Transformationen H_1 , H_2 berechnen, die Originalbildpunkte X_{bo} in rektifizierte Bildpunkte X_{brect} versetzen:

$$\begin{aligned} X_{1brect} &= H_1 X_{1bo} \\ X_{2brect} &= H_2 X_{2bo} \end{aligned} \quad (2.30)$$

2.2 Modell der Linsenkamera

Eine Lochkamera ist wegen der geringen Auflösung durch Beugung und der langen Belichtungszeiten für Filmaufnahmen nicht geeignet, s. Anhang Kap. A.1, obwohl sie eine unbegrenzte Tiefenschärfe hat, verzerrungsfrei ist und damit viel Rechenzeit ersparen würde.

Die Linsenkamera ist für Aufnahmen bewegter Szenen wesentlich geeigneter als die Lochkamera, da sie wegen ihrer größeren Lichtdurchlässigkeit bei gleichen Verhältnissen kürzere Belichtungszeiten, wie sie bei Filmaufnahmen erforderlich sind, zulässt.

Bei den folgenden Betrachtungen wird vorausgesetzt, dass das Licht monochromatisch sei und damit die Systeme dispersionsfrei sind.

Für dünne Linsen wird angenommen, dass der Lichtstrahl beim Übergang von Luft nach Glas und wiederum zur Luft nur in der mittleren Ebene der Linse gebrochen wird.

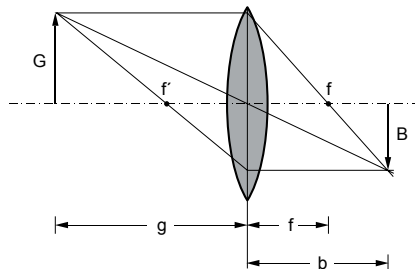


Bild 18: Lichtbrechung bei dünnen Linsen

Bild 18 zeigt einige wichtige Lichtstrahlen:

- Strahlen, die parallel zur optischen Achse verlaufen, gehen durch den Brennpunkt.
- Strahlen durch den optischen Mittelpunkt werden nicht gebrochen.
- Strahlen durch den gespiegelten Brennpunkt f' verlaufen auf der Bildseite parallel.

Die Anwendung der Strahlensätze führt zur Abbildungsgleichung:

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{g} \quad (2.31)$$

Analog gilt:

$$\frac{B}{G} = \frac{b-f}{f} \quad (2.32)$$

Da der Brennpunkt für eine Linse unverändert bleibt, muss bei einer anderen Gegenstandsweite die Bildweite angepasst werden oder eine gewisse Unschärfe akzeptiert werden.

Die verwendete Kamera hat eine feste Brennweite. Eine Unschärfe wird erst ab Abständen kleiner 0.6 m wahrgenommen. Geringe Unschärfe bei kurzen Abständen ist für die Genauigkeit der hier vorgestellten Verfahren ohne Bedeutung.

2.3 Fotosensoren auf Halbleiterbasis

Halbleitersensoren haben nahezu andere lichtempfindliche Medien verdrängt. Die Eigenschaften der digitalen Kameras überragen bei Weitem die der Analogkameras. Dazu gehören Kompaktheit, Aufnahmen unter schlechten Lichtverhältnissen, schnelle Verfügbarkeit der Aufnahmen und einfache Nachbearbeitung. Maschinelles Sehen wäre ohne Digitaltechnik erheblich erschwert.

Die Funktionsweise und das Rauschverhalten von CCD- und CMOS-Sensoren sind detailliert in [35] und [36] dargestellt.

In dieser Arbeit wird letztlich eine Stereokamera mit CCD-Sensor verwendet, da sie sich durch niedrigen Dunkelstrom, geringeres Rauschen, fehlendes sogenanntes „Fixed Pattern Rauschen“ und einen Global-Shutter auszeichnet. Das „Fixed Pattern Rauschen“ erzeugt für jeden Bildpunkt einen zufälligen

aber immer gleichen additiven Intensitätszuwachs. Durch einen „Global-Shutter“ werden alle Bildpunkte gleichzeitig aufgezeichnet

2.4 Abbildungsfehler

Abbildungsfehler entstehen einerseits durch das optische System, andererseits durch die eingesetzten Sensoren und haben zum Teil erheblichen Einfluss auf die Messgenauigkeiten.

2.4.1 Abbildungsfehler durch Objektive

Für die Messtechnik sind Abbildungsfehler relevant, da sie erhebliche Messungenauigkeiten erzeugen können. Auf Fehler, die zur Unschärfe, besonders am Bildrand, führen [37], weil Lichtstrahlen vom selben Objektpunkt verschiedene Wege durch das Objektiv nehmen, wird hier nicht eingegangen, da sie einer Tiefpassfilterung nahe kommen und zu vernachlässigende Ungenauigkeiten ergeben. Dazu zählen sphärische Aberration, Koma und Bildfeldwölbung.

Zwei Effekte, Vignettierung und Verzeichnung, sind im Zusammenhang mit dieser Arbeit einflussreich und werden im Folgenden näher erläutert.

2.4.1.1 Vignettierung

Bildpunkte am Rande der Bildfläche sind dunkler als die Punkte in der Nähe der optischen Achse. Dieser Randlichtabfall wird Vignettierung genannt. Er beruht zum großen Teil darauf, dass seitliche Lichtstrahlen eine geringere wirksame Fläche der Blende oder des Linsensystems passieren können. Hier ist der Lichtabfall in grober Näherung proportional zur vierten Potenz des Kosinus des Winkels zur optischen Achse [38]. Andere Vignettierungsursachen wie Richtungsempfindlichkeit der CCD-Sensoren oder durch Hindernisse, wie Gegenlichtblenden, kommen hinzu.

Da die Vignettierung von der Brennweite und der Blendenzahl abhängt, ist es schwierig eine allgemeingültige Formel zu finden.

Zur Messung des Intensitätsverlaufs wurde eine weiße, matte Fläche im freien Feld an einem trüben und damit weitgehend schattenfreien Tag aufgenommen. Das Vignettierungstestfoto mit der Bumblebeekamera von Point Grey hat als niedrigste Intensität den Wert von 54 % bezogen auf den Maximalwert, s. Bild 19.

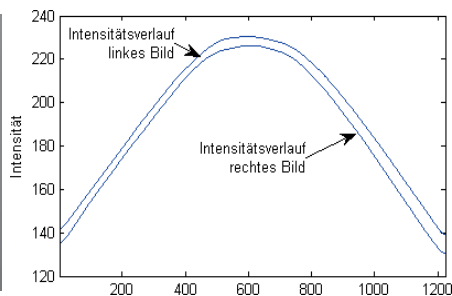
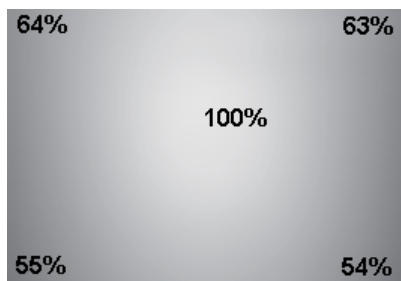


Bild 19: Links: Foto einer Szene mit konstantem Grauwert. Die Zahlen geben die relative Helligkeit in Bezug auf den hellsten Punkt des Bildes an. Das Bild wurde mit der linken Kamera Bumblebee von Point Grey aufgenommen. Die rechte Grafik stellt jeweils den Helligkeitsverlauf der mittleren Zeile für beide Kameras dar.

Der Intensitätsverlauf, Bild 20 und Bild 21, ist für beide Objektive sehr gekrümmt, so dass die Abbildung eines identischen Szenenausschnitts, die sich z. B. im linken Bild in der Mitte und im rechten Bild am Rand befindet, starke Helligkeitsunterschiede aufweist. Die Intensitätsmaxima befinden sich entgegen der Erwartung nicht in der Bildmitte (Zeile 425, Spalte 612), sondern bis zu 137 Pixel daneben, s. Tabelle 2.

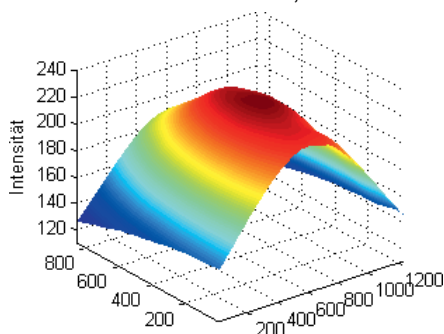


Bild 20: Intensitätsverlauf des linken Objektivs der Bumblebeekamera

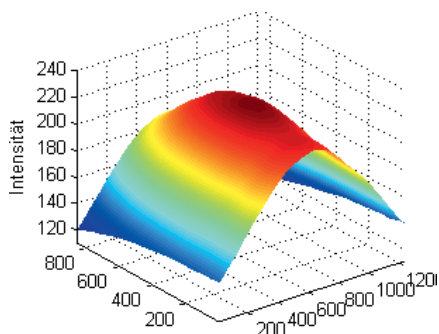


Bild 21: Intensitätsverlauf des rechten Objektivs

Tabelle 2: Intensitätsmaximum/-minimum und Ort für linkes und rechtes Bild

Intensität	Linkes Bild	Rechtes Bild
Niedrigster Wert	125	115
Bei Zeile	850	850
Bei Spalte	1224	1224
Höchster Wert	234	228
Bei Zeile	288	344
Bei Spalte	597	602

Der Einfluss der unterschiedlichen Intensitäten ist insbesondere zwischen linker und rechter Kamera so groß, dass für die Stereobildauswertung nur intensitätsunabhängige Bildregistrierungsverfahren genommen werden können, während der Intensitätsverlauf innerhalb derselben Kamera geringeren Einfluss auf die Bildregistrierung hat.

Versuche innerhalb des 8 Bit-Zahlenraumes, die Intensitäten über eine Tabelle auszugleichen, bringen zwar eine deutliche Verbesserung der Bildregistrierung, aber die Kontrastverminderung in der wichtigen Bildmitte führt zu ungenaueren Ergebnissen bei der Abstandsbestimmung großer Abstände.

2.4.1.2 Verzeichnung

Verzeichnungen erlauben zwar eine Zuordnung eines Szenenpunktes zu einem Bildpunkt, aber sie lassen ohne weitere Kenntnisse keine genaue Angabe über die Position des Szenenpunktes zu. Es wird zwischen dominanten radialen und meist kleinen tangentialen Verzeichnungen unterschieden.

Bei radialen Verzeichnungen entsteht, je nachdem, ob die Blende auf der Gegenstandsseite oder Abbildungsseite der Linse ist, eine Tonnen- oder Kissenverzeichnung [37]. Dort ist die Richtung des Abbildungspunktes vom optischen Zentrum richtig, aber die Entfernung ist zu kurz oder lang, s. Bild 22. Es ergibt sich dann eine Tonnen- oder Kissenverzeichnung, s. Bild 23.

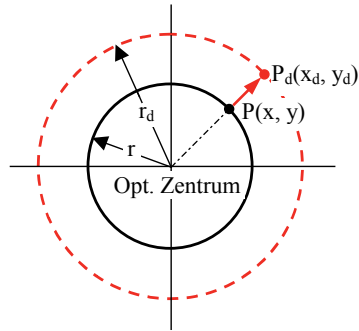


Bild 22: Radiale Verzeichnung, der verzeichnete Punkt ist um $r_d - r$ radial verschoben

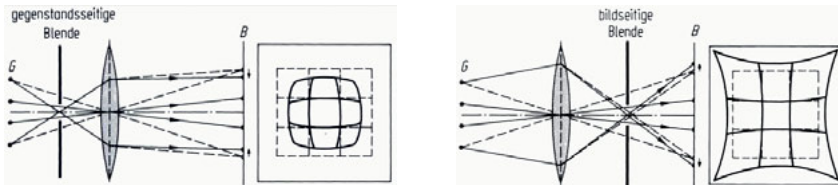


Bild 23: Tonnen- und Kissenverzeichnung (Quelle: [37])

Ein Beispiel einer realen Aufnahme mit einer Bumblebeekamera ist in Bild 24 gezeigt. Sie weist eine Tonnenverzeichnung auf. Die Abweichungen am Bildrand sind zum Teil über 50 px, s. Bild 25.



Bild 24: Originalbild, tonnenverzeichnet

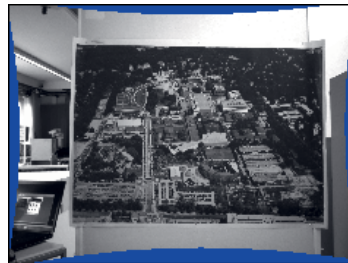


Bild 25: Gleiches Bild nach Kalibrierung, die blauen Randbereiche sind unbesetzte Pixel.

Ein einfacher Ansatz ist, anstatt einer Korrekturmatrix eine Gleichung für radiale Entzerrung, die nach dem quadratischen Term abgebrochen wurde, zu nehmen [39]:

$$x' = \frac{x}{1 + k|x|^2} \quad (2.33)$$

Die tangentielle Verzeichnung, s. Bild 26, entsteht zum großen Teil aus der Dezentrierung des Linsensystems, ist meist klein, da sie fast immer weitgehend vom Kamerahersteller behoben wird.

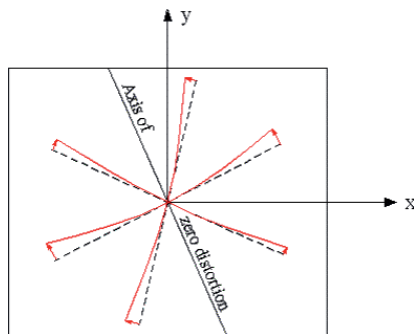


Bild 26: Tangentiale Verzeichnung

Quelle : Analytical Photogrammetry, Lecture notes Ferris State University, 2000

Zur Darstellung der Tangentialverzeichnung, s. Bild 27, die bei üblichen Kameras nicht erkennbar ist, wird einer kalibrierten Abbildung, Bild 28, eine sehr große tangentielle Verzeichnung zugefügt.



Bild 27: Ausschließlich tangential verzeichnete Abbildung



Bild 28: Gleiche Abbildung nach Kalibrierung

In Bild 29 ist die Kombination von radialer (entlang eines Radius) und von tangentialer Verzeichnung (entlang der Tangente) eines Bildpunktes dargestellt.

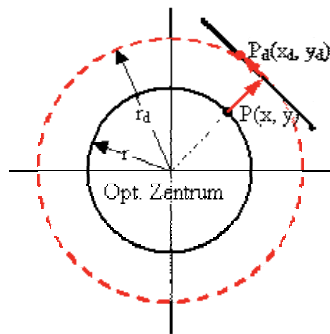


Bild 29: Radiale und tangentiale Verzeichnung

Da in dieser Arbeit die Subpixelgenauigkeit eminent wichtig ist, werden Verzeichnungen weitgehend durch Kalibrierung behoben, s. Kapitel 2.5.

2.4.2 Abbildungsfehler durch Sensoren

Zu den optischen Fehlern des Objektivs kommen noch die Fehler des Sensors hinzu.

2.4.2.1 Rauschen

Dunkelrauschen ist ein in Halbleitern erzeugtes thermisches Rauschen und wird bei längerer Belichtung im Bild sichtbar. Durch Mittelung mehrerer Aufnahmen derselben Szene [35] und durch Subtraktion eines Dunkelbildes kann Rauschen vermindert werden. Dies ist in der Regel nur bei wenig beleuchteten Szenen erforderlich.

Ortsfestes Rauschen ist insbesondere bei kleinen CMOS-Sensoren dominant. Es entsteht hauptsächlich durch die hohe aber von Pixel zu Pixel leicht unterschiedliche Verstärkung, die notwendig ist, um Ladungen in Spannungen umzuwandeln. Es kann durch bestimmte Filtertechniken in der Sensorelektronik stark vermindert werden [40].

Rauschen hat bei guten Belichtungen der Tageslichtflüge des Luftschiffs kaum Einfluss.

2.4.2.2 Blooming

Blooming entsteht durch lokale Überbelichtungen. Hier werden überschüssige Ladungen zu den für den Ausleseprozess gekoppelten Bildelementen weiter transportiert.

2.4.2.3 Smear-Effekt

Der Smear-Effekt tritt während des Ausleseprozesses auf. Die nicht vollkommen abgedunkelten Auslesezellen werden beim Ausleseprozess mitbelichtet und werden als Strich im Bild sichtbar.

2.4.2.4 Moiré-Effekt (Unterabtastung)

Der Moiré-Effekt entsteht durch Unterabtastung, weil die Pixeldichte nicht ausreicht, um alle Helldunkelstellen abzubilden. Dies fällt besonders bei periodischen engen Mustern auf. Der Effekt kann nicht durch Nachbearbeitung behoben werden, da durch die Unterabtastung Informationen verloren gehen [39][41].

Bei Unterabtastung können Muster entstehen, die in Richtung und Abstand nicht mehr der Szene entsprechen, s. Bild 30.

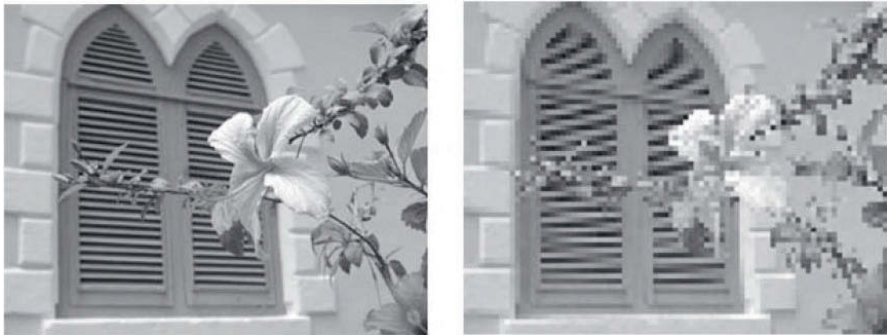


Bild 30: Links: ausreichend abgetastetes Bild, rechts: stark unterabtastetes Bild (Quelle: [41])

Das abgetastete Bild $g_s(x)$ ist in erster Näherung die Multiplikation der Bildfunktion $g(x)$ mit einer „Nagelbrettfunktion“ δ – einem zweidimensionalen Dirac-Kammfilter – entsprechend dem Pixelabstand [39].

$$g_s(x) = g(x) \sum_{m,n} \delta(x - r_{m,n}) \quad (2.34)$$

Eine genauere Beschreibung der Abtastfunktion wird erreicht, wenn man die endlichen Abmessungen der Abtastfläche des Bildsensors berücksichtigt.

In Bild 31 ist eine Szene zu sehen, welche durch ein Gitter aufgenommen wurde. Das Gitter approximiert die „Nagelbrettfunktion“. Im rekonstruierten Bild werden die Quantisierungseffekte sichtbar.

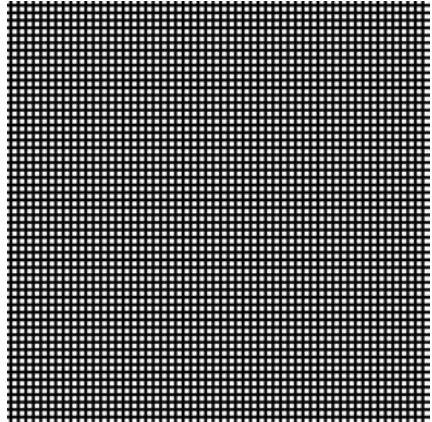
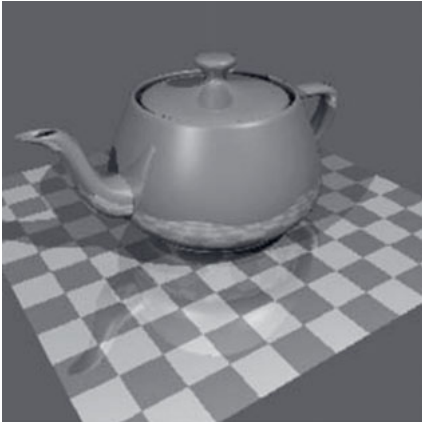


Bild 31: Abtastung Links: Szene,

rechts: Abtastgitter (Quelle: [41])

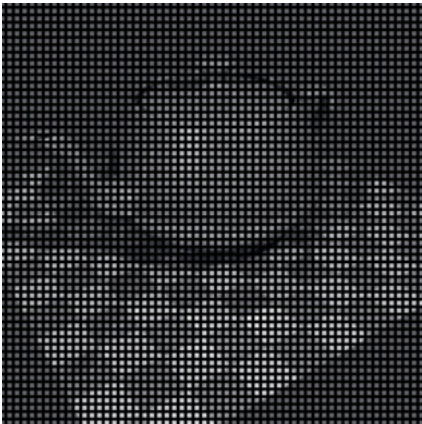


Bild 32: Abtastung. Links: Szene durch Abtastgitter gesehen,
rechts rekonstruierte Abbildung (Quelle: [41])

Unterabtastung entsteht auch beim Bildverkleinern, wenn ungefiltert z. B. jeder zweite Bildpunkt der Zeilen und Spalten herangezogen wird, um ein auf ein Viertel verkleinertes Bild zu erstellen.

2.5 Behebung der Verzeichnungen

2.5.1 Beschreibung radialer und tangentialer Verzeichnung

In Kap. 2.4.1 wurden die Abbildungsfehler, die sich durch Objektive ergeben können, gezeigt. Hierbei ist die dominante Verzerrung durch Linsen die radiale Verzerrung, insbesondere bei kleinen Brennweiten [29][42].

Die Beziehung zwischen den verzerrten (engl. **distorted**) Kamerakoordinaten x_{cd} , y_{cd} und den verzerrten Bildkoordinaten x_{bd} , y_{bd} in der Ebene $z_c = 1$ ist:

$$\begin{pmatrix} x_{bd} \\ y_{bd} \end{pmatrix} = \begin{pmatrix} fx_{cd} + c_x \\ fy_{cd} + c_y \end{pmatrix} \quad (2.35)$$

Da die Verzerrung mit der Entfernung (Radius r) zum Hauptpunkt zunimmt, ist es sinnvoll, diese Abhängigkeit zu benutzen:

$$r = \sqrt{x_c^2 + y_c^2} \quad (2.36)$$

Zwischen den verzerrten Kamerakoordinaten und den unverzerrten Kamerakoordinaten existiert eine Funktion F :

$$\begin{pmatrix} x_{cd} \\ y_{cd} \end{pmatrix} = F(r) \begin{pmatrix} x_c \\ y_c \end{pmatrix} \quad (2.37)$$

$F(r)$ kann durch eine Taylorentwicklung angenähert werden:

$$F(r) = \sum_{i=0}^{\infty} a_i r^i \quad (2.38)$$

Da die Verzerrung symmetrisch zum Hauptpunkt ist, entfallen Terme mit ungerader Potenz. Als realistisch wird eine Genauigkeit von 1/50 Pixel angesehen [43], deshalb wird Gleichung (2.38) in der Regel nach der 4. Potenz abgebrochen [42]. Nur bei extremen Verhältnissen, wie sie durch Fischaugenlinsen entstehen, wird noch die 6. Potenz hinzugefügt.

Gleichung (2.37) wird zu:

$$\begin{pmatrix} x_{cd} \\ y_{cd} \end{pmatrix} = (1 + d_1 r^2 + d_2 r^4) \begin{pmatrix} x_c \\ y_c \end{pmatrix} \quad (2.39)$$

Die tangentiale Verzerrung ist unabhängig von der radialen Verzerrung und wird durch folgende Beziehung beschrieben:

$$\begin{pmatrix} x_{cd} \\ y_{cd} \end{pmatrix} = \begin{pmatrix} d_3(2x_c y_c) + d_4(r^2 + 2x_c^2) \\ d_3(r^2 + 2y_c^2) + d_4(2x_c y_c) \end{pmatrix} + \begin{pmatrix} x_c \\ y_c \end{pmatrix} \quad (2.40)$$

Die Zusammenfassung von radialer und tangentialer Verzerrung führt dann zu:

$$\begin{pmatrix} x_{cd} \\ y_{cd} \end{pmatrix} = (1 + d_1 r^2 + d_2 r^4) \begin{pmatrix} x_c \\ y_c \end{pmatrix} + \begin{pmatrix} d_3(2x_c y_c) + d_4(r^2 + 2x_c^2) \\ d_3(r^2 + 2y_c^2) + d_4(2x_c y_c) \end{pmatrix} \quad (2.41)$$

Sind die unverzerrten Kamerakoordinaten x_c , y_c gesucht, kann in erster Näherung folgende Gleichung benutzt werden, um dann iterativ auf genauere Werte zu kommen.

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} \approx \begin{pmatrix} \tilde{x}_c \\ \tilde{y}_c \end{pmatrix} = \frac{1}{1 + d_1 r^2 + d_2 r^4} \left[\begin{pmatrix} x_{cd} \\ y_{cd} \end{pmatrix} - \begin{pmatrix} d_3(2x_c y_c) + d_4(r^2 + 2x_c^2) \\ d_3(r^2 + 2y_c^2) + d_4(2x_c y_c) \end{pmatrix} \right] \quad (2.42)$$

\tilde{x}_c und \tilde{y}_c sind die angenäherten Werte für die korrekten Koordinaten x_c und y_c . Das rekursive Substituieren von x_c und y_c mit \tilde{x}_c und \tilde{y}_c im rechten Teil der Gleichung (2.42) führt nach einigen Iterationen zu einer ausreichenden Approximation.

2.5.2 Kalibrierung und Rektifizierung bei verzerrter Abbildung

Auf Bildpunkte bezogen wird Gleichung (2.42) umgeformt zu:

$$\begin{pmatrix} x_{bd} - x_b \\ y_{bd} - y_b \end{pmatrix} = \begin{pmatrix} d_1(x_b - c_x)r^2 + d_2(x_b - c_x)r^4 + d_3(2x_w y_w f_x) + d_4(r^2 + 2x_w^2)f_x \\ d_1(y_b - c_y)r^2 + d_2(y_b - c_y)r^4 + d_3(r^2 + 2y_w^2)f_y + d_4(2x_w y_w f_y) \end{pmatrix} \quad (2.43)$$

Mit verschiedenen Welt- und Bildpunktpaaren lässt sich dann wieder ein überbestimmtes Gleichungssystem aufstellen, das dann die Verzerrungsparameter d_1 bis d_4 liefert. Es sind mehr als 1000 Iterationen erforderlich [29], um eine ausreichende Genauigkeit zu erlangen.

Aus den errechneten Parametern lassen sich Matrizen aufstellen, die die Pixel der verzeichneten Abbildung in eine unverzeichnete Abbildung transformieren. Mit der Rektifizierung nach Kapitel 2.1.4, S. 26 erhält man dann eine scheinbar frontal parallele Stereokameraanordnung mit unverzeichneten Bildern.

Durchgesetzt haben sich allerdings andere Methoden der Kalibrierung, die den Kalibrierungsprozess weitgehend geschlossen ausführen. Hier seien die drei wichtigsten Verfahren nach Tsai [44], Heikkila & Silven [42] und Zhang [45] aufgezählt. Ein Vergleich einiger Methoden ist in [46] zu finden.

In der vorliegenden Arbeit wird ein Verfahren, welches eine Kombination der Methode nach Zhang und Heikkila ist und von Jean-Yves Bouguet [47] umgesetzt wurde, benutzt.

Der Ablauf der Kalibrierung und der Rektifizierung sei an Bild 33 bis Bild 38 erklärt und erfolgt in diesen Schritten:

- Gleichzeitige Aufnahme eines Schachbrettmusters mit bekannten Maßen entsprechend Bild 33 mit linker und rechter Kamera aus mindestens drei verschiedenen Positionen.
- Bestimmen der Kalibrierungswerte für linke und rechte Kamera mit Hilfe der Eckenpositionen. Ergebnis ist eine Liste von Entzerrungsparametern.
- Bestimmen der Rektifizierungswerte für die Stereokombination. Zu den Entzerrungsparametern werden die Translations- und die Rotationswerte hinzugefügt. Diese Werte wurden mit Hilfe eines Matlabprogramms¹ von Jean-Yves Bouguet [47] erstellt.
- Jedem Bildpunkt im rektifizierten Bild wird ein Punkt im Originalbild zugeordnet. Dabei werden unganzzahlige Positionen im Originalbild gefunden, die durch bilineare Interpolation aus ganzzahligen Pixelpositionen gewonnen werden. Als Ergebnis dienen Matrizen mit vier bilinearen Koeffizienten für jeden rektifizierten Bildpunkt, die Bilder entsprechend Bild 34 und Bild 37 erzeugen.

Um die Pixel außerhalb der Rechtecke (Bild 34 und Bild 37) nicht erst zu berechnen, werden die effektiven Parameter nach einem eigenen Verfahren extrahiert und sortiert².

¹ Matlab® ist eine Programmierumgebung für Mathematik der Firma The Mathworks inc.

² P. Seibold: <http://www.mathworks.com/matlabcentral/fileexchange/47167-recalculate-rectification-indices-for-rectangle-image>

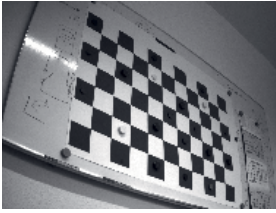


Bild 33



Bild 34

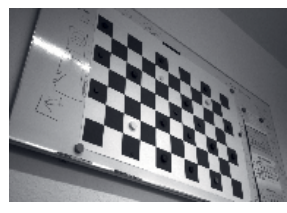


Bild 35

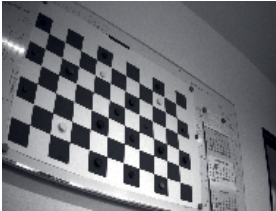


Bild 36



Bild 37



Bild 38

Bild 33: linkes Originalbild

Bild 34: linkes Bild kalibriert und rektifiziert. Das große Rechteck (rot) ist für das größtmögliche rektifizierte Bild. Das kleinere Rechteck (grün) ist der Rahmen für den größtmöglichen gemeinsamen Bildausschnitt.

Bild 35: linkes Bild kalibriert und rektifiziert mit größtmöglichem gemeinsamem Rechteckausschnitt

Bild 36 - Bild 38: wie vorhergehende Abbildungen, allerdings jeweils für das rechte Bild der Stereokamera.

Die Tabellen der Bilinearkoeffizienten sind sehr umfangreich, da zur Berechnung jeden Bildpunkts vier Gleitkommawerte erforderlich sind. Die Datenmenge kann reduziert werden, indem die Genauigkeit der Koeffizienten reduziert wird. Eine weitere Reduzierung, was auch den Rechenaufwand drastisch verkleinert, ist der Verzicht auf die Interpolation, d. h. direkt den benachbarten Pixelwert zu nehmen. In Bild 39 ist die Fehlerverteilung bei 7 Bit Genauigkeit im Vergleich zur vollen Gleitkommagenauigkeit bei einer typischen Abbildung dargestellt. 11010 Pixel (1 %) weisen einen nur um eins zu großen Wert auf und 6370 Pixel (0.6 %) sind um eins zu klein. 1023020 (98.4 %) von insgesamt 1040400 Pixel sind richtig. Bei Anwendung der Nächster-Nachbarn-Interpolation sieht das Ergebnis zunächst desolat aus.

64 % der Werte sind falsch. In Kapitel 5.4.5.2, S. 157 wird diese Methode trotzdem untersucht.

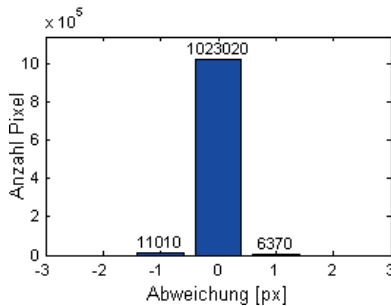


Bild 39: Pixelfehler bei 7 Bit Genauigkeit für die Interpolation im Vergleich zu Gleitkommagenauigkeit

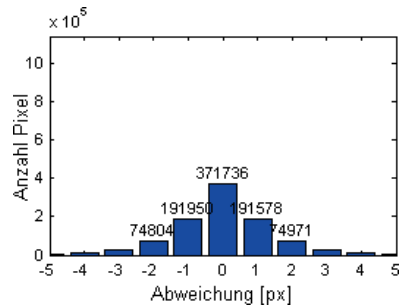


Bild 40: Pixelfehler bei Anwendung der Nächster-Nachbarn-Interpolation im Vergleich zu Gleitkommagenauigkeit

2.6 Ermittlung der Bewegung

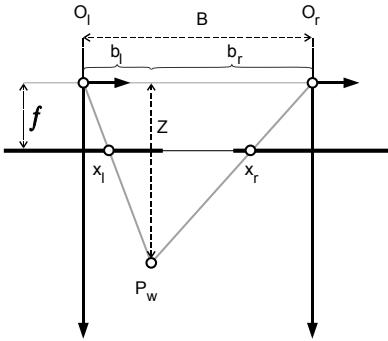
Bei der Ermittlung der Bewegung durch eine auf das Flugsystem montierte Kamera wird davon ausgegangen, dass die darunter liegende Szene starr ist. Dies ist in der Regel der Fall bei Beobachtungen aus großer Höhe, da bewegte Objekte in der Szene dann klein in Bezug zur Gesamtfläche sind und damit geringen Einfluss in Bezug auf einen größeren Bildausschnitt haben.

2.6.1 Stereogeometrie

Während GPS eine ungenaue Höhe über einem fiktiven Rotationsellipsoid angibt und ein Barometer ständig neu kalibriert werden muss und nur die relative Höhe misst, haben Höhenbestimmungen mit Kameras den Vorteil, dass damit die absolute Höhe über Grund bestimmt werden kann.

Diese Tiefeninformationen des sichtbaren Teils einer Szenenoberfläche können über Triangulation mit Hilfe einer Stereokamera gewonnen werden.

Durch die Rektifizierung braucht nur noch der einfache Fall der frontal parallelen Anordnung zweier gleicher Kameras betrachtet werden. Beide Kameras haben parallele optische Achsen, die in Richtung der Szene, die in Bild 41 unten ist, zeigen, und beide optische Zentren sind auf derselben zu den optischen Achsen orthogonalen Basislinie.



Bezeichnungen:

O_l, O_r : Optisches Zentrum links, rechts

B : Basisabstand (Abstand der Kameras)

b_l, b_r : Abstand links, rechts

f : Brennweite

z : Abstand des Punktes P_w

x_l, x_r : x-Lage des Bildpunktes links und rechts in Bezug zu den optischen Zentren

P_w : Weltpunkt

Bild 41: Tiefeninformation über Triangulation mit frontal parallelen Kameras

Mit zwei Kamerakoordinatensystemen, die jeweils ihren Ursprung in den optischen Zentren O_l und O_r haben, lässt sich über die Strahlensätze der Abstand z eines Weltpunktes P_w in z-Richtung der Kamerakoordinatensysteme bestimmen.

Mit

$$\frac{f}{x_l} = \frac{z}{b_l} \quad \text{und} \quad \frac{f}{-x_r} = \frac{z}{b_r} \quad (2.44)$$

$$B = b_r + b_l$$

wird der Abstand Z zu:

$$z = \frac{Bf}{x_l - x_r} \quad (2.45)$$

$$d = x_l - x_r \quad (2.46)$$

Die Differenz d der abgebildeten x-Positionen eines Weltpunktes auf beiden Bildebenen wird Disparität genannt.

Mit bekanntem Pixelabstand p auf dem Sensor kann die Brennweite f_{px} in Pixeln – dies ist bei der Kalibrierung üblich – angegeben werden:

$$f_{px} = \frac{f}{p} \quad (2.47)$$

Damit errechnet sich z mit der Disparität d_{px} , ebenfalls in Pixel, zu:

$$z = \frac{Bf_{px}}{d_{px}} \quad (2.48)$$

Eine Verdopplung der Basisbreite oder der Brennweite führt zu einer Verdopplung der Disparität d bei gleichen Abständen z . Um große Abstände aufzulösen, sind entweder große Brennweiten, große Basisbreiten oder eine hohe Subpixelgenauigkeit erforderlich. Die Erfassung des Nahbereichs wird durch die abnehmende Bildschärfe und eingeschränkte maximale Disparitäten der benutzten Kamera auf 0.6 m begrenzt.

In Bild 42 wird die gemessene Distanz in Abhängigkeit von der Disparität der Stereokamera Bumblebee gezeigt. Beide Teilkameras sind kalibriert und rektifiziert und damit scheinbar frontal parallel.

Die Kenndaten der Kamera sind:

- Basisbreite B : 0.2396 m
- Brennweite f_{px} : 1401 px , entspricht 5.3 mm bei einem Sensorpixelabstand von $3.75 \mu\text{m}$
- Auflösung rektifiziert: 850 x 1224 px

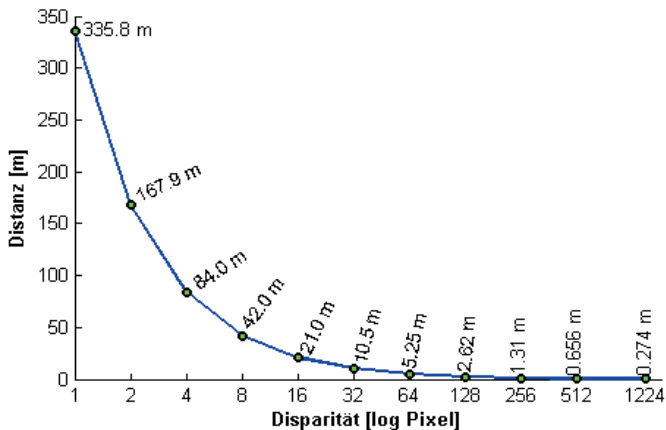


Bild 42: Gemessene Distanz bei verschiedenen Disparitäten, Brennweite 5.3 mm, Stereobasisbreite 24 cm

Wie stark der prozentuale Fehler mit zunehmender Disparität abnimmt, wird in Bild 43 gezeigt.

Bei der gängigen Missionsflughöhe des Luftschiffs von 42 m ist die Disparität 8 Pixel. Beträgt der maximale Disparitätsmessfehler ± 0.3 Pixel, liegt die

tatsächliche Flughöhe zwischen 40.5 m und 43.6 m. Während des Landeanflugs ist der prozentuale Messfehler deutlich geringer. Aus dem Fehlertrichter, Bild 44, ist bei 1.3 m Flughöhe der Messfehler von nur ± 1.5 mm entnehmbar. Die Genauigkeit der Höhenabschätzung steigt mit der Subpixelgenauigkeit, die in Kapitel 4.5 behandelt wird.

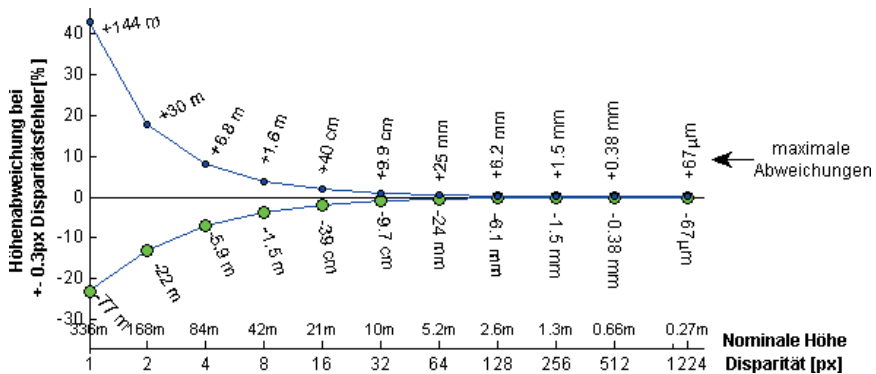


Bild 43: Fehlertrichter bei verschiedenen Disparitäten und einem Disparitätsfehler von ± 0.3 Pixel, Brennweite 5.3 mm, Stereobasisbreite 24 cm

2.6.2 Translation

Die Translation ist die Bewegung des Sensors der Kamera in x- und y-Richtung ausgehend vom optischen Zentrum. Die Bewegung wird aus der Verschiebung zweier aufeinanderfolgender Bilder, die der Sensor aufgenommen hat, errechnet mit Verfahren, die in Kap. 3 vorgestellt werden. Gesucht ist also, wie viel Meter Bewegung einer Bildverschiebung von wie viel Pixeln entspricht.

Aus der Brennweite f und dem Abstand z zur betrachteten Szene lässt sich die Beziehung zwischen der Strecke x_b auf der Bildebene und der Strecke x_c in Kamerakoordinaten herstellen .

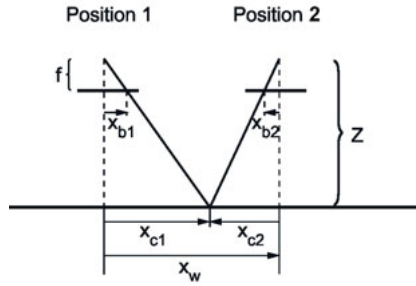


Bild 44: Geometrie bei Translation

Zunächst ist die Kamera an der Position 1. Es gilt:

$$x_{b1} = \frac{f x_{c1}}{z} \quad (2.49)$$

Nach der Translation ist die Kamera an der Position 2. Für denselben Szenenpunkt ergibt sich:

$$x_{b2} = \frac{f x_{c2}}{z} \quad (2.50)$$

Die Strecke x_w berechnet sich nun zu:

$$x_w = x_{c1} - x_{c2} = \frac{z}{f} (x_{b1} - x_{b2}) \quad (2.51)$$

Mit der Brennweite f_{px} in Pixel kann direkt aus den Pixelwerten auf der Bildebene x_{b1px} und x_{b2px} auf die Strecke geschlossen werden:

$$x_w = \frac{z}{f_{px}} (x_{b1px} - x_{b2px}) \quad (2.52)$$

Letztlich ist die Geometrie für die Translation gleich der Stereogeometrie. Die Translation x_w entspricht der Basisbreite B aus Gleichung (2.48).

Durch Zusammenführen der Gleichung für den Abstand (2.48) in die Gleichung für die Translation (2.52) ergibt sich, dass die Translation unabhängig von der Brennweite ist. Das heißt, Veränderungen der Kalibrierung bezüglich der Brennweite beeinflussen nicht die Messgenauigkeit für die Translation x_w :

$$x_w = \frac{B}{d_{px}}(x_{b1_{px}} - x_{b2_{px}}) \quad (2.53)$$

Ein Beispiel für eine Translation ist mit Bild 45 gegeben.

Die Translation von 4.5 px in x-Richtung und 18.8 px in y-Richtung wurde nach der Gradientenmethode von Lucas und Kanade berechnet, wobei im mittleren Fenster der Durchschnitt aller Flusswerte gebildet wurde.

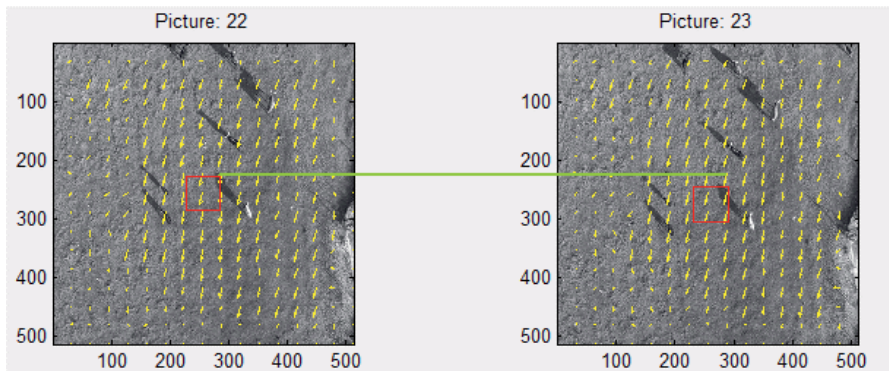


Bild 45: Ermittlung der Translation mit Hilfe des optischen Flusses nach Lucas&Kanade. Links: Bild mit Berechnungsfenster, rechts: nachfolgendes Bild mit positioniertem Fenster, nur jeder 32. Flussvektor ist gezeigt

2.6.3 Rotation

Mit einer Kamera kann auch die Rotation um die optische Achse bestimmt werden.

Die Rotation wird in ähnlicher Weise wie die Translation ermittelt. Allerdings werden hierzu zwei Fenster benötigt, um dann den Winkel zu berechnen. In Bild 47 sind in der linken Abbildung zwei kleine Fenster gezeigt, für die der optische Fluss zwischen beiden Abbildungen berechnet wird. An den Flussvektoren kann die Rotation gut erkannt werden.

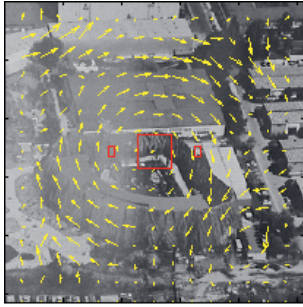


Bild 46: Erstes Bild

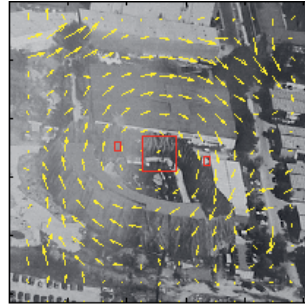


Bild 47: Nachfolgendes Bild

Ermittlung der Rotation mit Hilfe des optischen Flusses nach Lucas&Kanade. Mit den Abständen der beiden kleinen Fenster im rechten Bild in x- und y-Richtung wird der Winkel berechnet.

Um die Rotation zu berechnen, wird die Verschiebung der kleinen Fenster ermittelt.

Der Winkel φ berechnet sich aus den Abständen A_x , A_y , die sich im zweiten Bild ergeben, s. Bild 47 und Bild 48. Im ersten Bild, Bild 46, ist der Abstand gesetzt zu $A_y = 0$.

$$\varphi = \arctan\left(\frac{A_y}{A_x}\right) \quad (2.54)$$

Bild 48 ist ein Ausschnitt aus Bild 47 mit den eingezeichneten Parametern zu Gleichung (2.54).

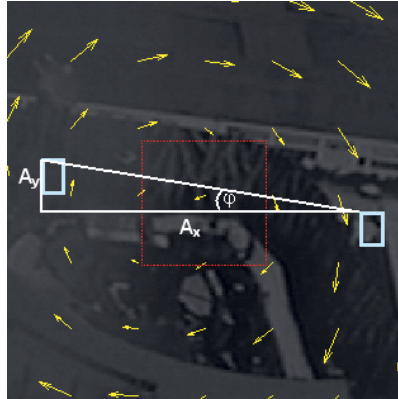


Bild 48: Ausschnitt aus Bild 47 zur Bestimmung der Rotation um die optische Achse

2.6.4 Geschwindigkeitsbestimmung

Die momentanen Geschwindigkeiten v_x , v_y werden aus den momentanen Translationen in x- und y-Richtung x , y und der Bildrate fps bestimmt zu:

$$\begin{aligned} v_x &= x \cdot fps \\ v_y &= y \cdot fps \end{aligned} \quad (2.55)$$

2.7 Zusammenfassung Kapitel 2

Mit dem Lochkameramodell können die Beziehungen zwischen Weltpunkten und Bildpunkten beschrieben werden. Über die Epipolargeometrie ist die Geometrie mehrerer Bilder einer Szene beschrieben. Sie ermöglicht die Rektifizierung zweier Kameras, was den späteren Korrespondenzsuchprozess wesentlich vereinfacht. Da die Lochkamera eine zu geringe Auflösung und eine zu hohe Belichtungszeit hat, wird die Linsenkamera eingeführt. Diese hat wiederum andere Nachteile, insbesondere die Vignettierung und die Verzeichnung. Die Verzeichnung lässt sich weitgehend über Kalibrierung beheben. Hinzu kommen noch Effekte des Halbleiterbildsensors, die in dieser Anwendung nicht dominant sind.

Nachdem die Stereokamera kalibriert und rektifiziert ist, können aus Bildpunkten die Bewegung und der Abstand berechnet werden. Allerdings müssen in jeweils zwei Bildaufnahmen die korrespondierenden Punkte gefunden werden. Verfahren hierzu werden im folgenden Kapitel vorgestellt.

3 Bewegungsabschätzung und optischer Fluss

3.1 Einleitung

Unter optischem Fluss versteht man nach B. Horn und B. Schunck [48] die Verteilung der scheinbaren Geschwindigkeiten der Bewegung von Helligkeitsmustern in einem Bild.

Schon 1867 beschreibt H. Helmholtz im „Handbuch der Physiologischen Optik“ [49], wie der optische Fluss vom Menschen wahrgenommen wird, folgendermaßen:

„Wenn wir vorwärts gehen, so bleiben die Gegenstände, welche sich am Wege ruhend befinden, hinter uns zurück; sie gleiten in unserem Gesichtsfelde scheinbar an uns vorbei, und zwar in entgegengesetzter Richtung, als wir fortschreiten. Entferntere Gegenstände thun dasselbe, aber langsamer, während sehr entfernte Gegenstände, wie die Sterne, ruhig ihren Platz im Gesichtsfelde behaupten, so lange wir die Richtung unseres Körpers und Kopfes beibehalten. Es ist leicht ersichtlich, dass die scheinbare Geschwindigkeit der Winkerverschiebungen der Gegenstände im Gesichtsfelde hierbei ihrer wahren Entfernung umgekehrt proportional sein muss, so dass aus der Geschwindigkeit der scheinbaren Bewegung sichere Schlüsse auf die wahre Entfernung gemacht werden können.“

J. Gibson illustriert den „Fluss der bildlichen Wahrnehmung“ 1950 [50] mit Pfeilen und spricht hierbei von Geschwindigkeit und Richtung. In Bild 49 zeigt er, wie eine Person – im Flugzeug sitzend – eine Szene (Flugplatz) wahrnimmt, wenn sie sich nach rechts blickend vorwärtsbewegt. Die Szene gibt die oben zitierte Beschreibung von H. Helmholtz wieder.

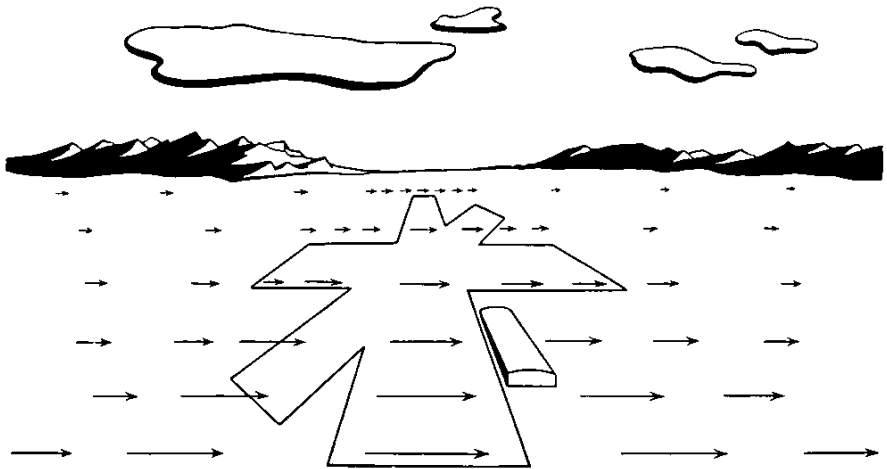


Bild 49: Wahrnehmung des Bewegungsfeldes (Quelle: [50])

Im Gegensatz zu Menschen und Tieren, die stereoskopisch ihre Umwelt wahrnehmen, nutzen insbesondere Insekten den optischen Fluss zur Navigation. So können Bienen z. B. einen Tunnel in etwa mittig durchfliegen ohne anzustoßen [51].

Hierbei ist zwischen Bewegungsfeld und optischem Fluss zu unterscheiden. B. Jähne [39] definiert das Bewegungsfeld als Projektion der Bewegungen einer 3D-Szene auf die Bildebene, während der optische Fluss den Grauwertfluss in der Bildebene wiedergibt [39][52]. In Bild 50 wird deutlich, dass der Grauwertfluss nicht mit dem Bewegungsfeld übereinstimmen muss. Beide Teilbilder zeigen jeweils eine Kugel mit Lambertscher Oberfläche (Strahlungsdichte richtungsunabhängig). Obwohl sich die linke Kugel, Bild 50a, dreht und das Bewegungsfeld eine Rotation um die senkrechte Achse abbildet, ist kein optischer Fluss feststellbar. Die Helligkeitswerte nacheinander aufgenommener Bilder unterscheiden sich nicht. Im Gegensatz dazu ist bei der rechten Szene, Bild 50b, durch die Beleuchtungsänderung – hier dreht sich eine Lichtquelle um die senkrechte Achse – ein optisches Flussfeld vorhanden, obwohl die Kugel steht. Es ist also ersichtlich, dass nicht jede Bewegung zu einem optischen Fluss führt und nicht jede Grauwertänderung auf einer Bewegung beruht.

Die Schlussfolgerung, je mehr Textur eine Szene aufweist, umso mehr stimmt in der Regel der optische Fluss mit dem Bewegungsfluss überein, ist richtig [52].



Bild 50: Bewegung und resultierender optischer Fluss an einer Kugel gezeigt,
a) Die Kugel bewegt sich: kein optischer Fluss
b) Die Beleuchtung bewegt sich: optischer Fluss

Quelle: B. K. Horn, Robot Vision, MIT Press, 1986

Klarer wird das Verhalten des optischen Flusses, wenn eine Ebene betrachtet wird, die senkrecht zur optischen Achse steht.

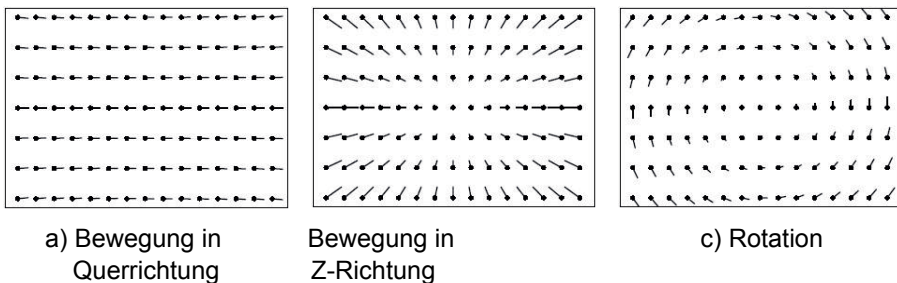


Bild 51: Optischer Fluss einer Szene, die senkrecht zur optischen Achse ist.
Quelle: H. A. Mallot, Sehen und die Verarbeitung visueller Information, Vieweg Verlag, 2001

3.2 Voraussetzungen und Probleme bei der Bewegungserfassung

3.2.1 Korrespondenzproblem

Das Grundproblem der Bildregistrierung ist, einen Abbildungspunkt eines Szenenpunktes in einer zweiten Abbildung wiederzufinden.

Aus Bild 52 ist zu erkennen, dass die Zuordnung der Bildpunkte der gleichen Objektpunkte verschiedener Bilder oft nur geschätzt werden kann. Die richtige Einschätzung ist umso wahrscheinlicher, wenn sich korrespondierende Bildpunkte in etwa in der gleichen Position befinden. Dies gilt insbesondere, wenn sich die betrachteten Gebiete nicht durch andere Merkmale (z. B. Farbe, Größe oder Form) unterscheiden.

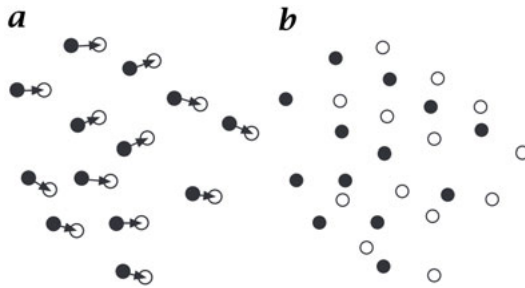


Bild 52: Korrespondenzproblem. Die dunklen Kreise sind von der ersten Aufnahme. Die hellen Kreise geben die Position im 2. Bild wieder. Die Teile sind weder durch Form, Größe noch Farbe zu unterscheiden. In a) wird angenommen, dass die nahen Nachbarn zueinander gehören. In b) ist keine Zuordnung erkennbar. (Quelle: [39])

3.2.2 Blendenproblem

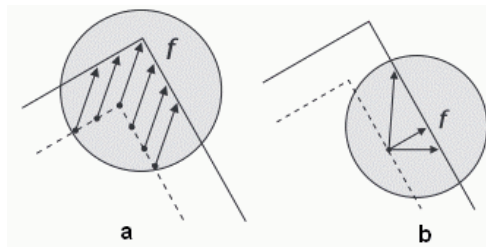


Bild 53: Blendenproblem: In Bild a kann der Fluss aus der Umgebung bestimmt werden. In Bild b kann nur die vertikale Komponente des Flusses bestimmt werden. (Quelle: [53])

Das Blendenproblem ist ein Spezialfall des Korrespondenzproblems. In der Regel wird zur Ermittlung des optischen Flusses nur ein kleines Gebiet, wie durch eine Blende betrachtet, einbezogen. In Bild 53, Teil a kann der optische Fluss für die Ecke ermittelt werden und damit auch für das Gebiet in der Nähe der Ecke. Im Teilbild b kann beim betrachteten Ausschnitt nicht bestimmt werden, wie ein Punkt sich bewegt hat. Der Verschiebungsvektor bleibt bis auf seine vertikale Komponente unbekannt. Der Fluss nach der differentiellen Methode kann nur in Richtung des Gradienten bestimmt werden.

Bei periodischen Mustern, z. B. Bild 54, kann auch selbst die vertikale Komponente zum Objekt nicht bestimmt werden. Erst wenn die Blende so weit geöffnet ist, dass Randbereiche des Gitters mit einbezogen werden, wird die Verschiebung bestimmbar.

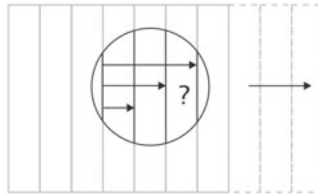


Bild 54: Periodisches Muster erlaubt bei Verschiebungen keine eindeutige Zuordnung (Quelle: [39])

Bei Isointensitätswerten gibt es keinen Gradienten, somit ist der optische Fluss nicht bestimmbar. Das Problem kann umgangen werden, indem die lokalen Operationen auf ein größeres Gebiet mit Struktur erweitert werden. Allerdings wird dann die Auflösung des optischen Flusses geringer.

3.2.3 Beleuchtungsänderung

Die Grundvoraussetzung für die Berechnung des optischen Flusses nach Horn&Schunk oder Lucas&Kanade ist, dass die Bildpunkte eines Objektpunktes annähernd dieselbe Intensität aufweisen.

Die meisten Beleuchtungsänderungen sind multiplikativ, d. h., sie bewirken eine prozentuale Intensitätswertänderung, Kontraständerung genannt. Ursache hierfür kann die Vignettierung, s. Kapitel 2.4.1, Belichtungszeit- oder Blendenänderung oder Eigenschaften des Flugsystems sein. Änderungen des Außenlichts (z. B. durch Wolken) sind hier nicht relevant, da diese Prozesse im Vergleich zur Bildfolge langsam ablaufen. Die anderen genannten Helligkeitsänderungen sind nicht determiniert und bewirken Fehler in der Berechnung des optischen Flusses.

Additive Intensitätsänderungen entstehen im Sensor, wenn sich z. B. trotz geschlossenem Verschluss der Kamera Ladungen ansammeln, s. Kapitel 2.4.2, oder verschiedene Kameras benutzt werden.

Lichtspiegelungen führen zu lokalen und vom Blickwinkel abhängigen Übersteuerungen, die keine Bestimmung des optischen Flusses zulassen.

In Kapitel 4.2 wird gezeigt, wie sich Beleuchtungsänderungen auswirken.

3.2.4 Verdeckung



Linkes Stereobild



Rechtes Stereobild

Bild 55: Verdeckung bei Aufnahmen aus verschiedenen Positionen

Bild 55 zeigt eine Szene, bei welcher ein Hintergrundobjekt (Baum) durch ein Objekt im Vordergrund (Mauer) bei einer Aufnahme aus einer anderen Position verdeckt wird. Hier ist der Baumstamm im linken Bild, kleines Quadrat, noch sichtbar, während er im rechten Bild verdeckt wird. Es gibt also keine Korrespondenz der betreffenden Bildpunkte und führt zu Fehlinterpretationen.

Die Ansätze dieses Problem zu erfassen, sind sehr aufwendig, da es einerseits schwierig ist, alle Verdeckungen zu erkennen, andererseits werden nicht vorhandene Verdeckungen als solche erkannt [54].

Ein gängiges Verfahren, Verdeckungen zu detektieren, ist in [55] vorgestellt. Hier werden zunächst entlang einer Epipolarlinie die zum ersten Bild I_1 korrespondierenden Pixel im zweiten Bild I_2 gesucht und danach die zum zweiten Bild korrespondierenden Pixel ermittelt. Sind beide Korrespondenzen gleich – linke Pfeile in Bild 56 – ist die Berechnung vermutlich richtig. Bei unterschiedlichen Korrespondenzen – rechte Pfeile – wird von einer Verdeckung ausgegangen.

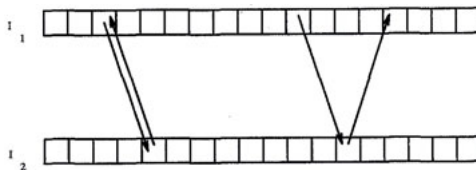


Bild 56: Konsistente und inkonsistente Korrespondenzen (Quelle [55])

Die bekannten Verfahren führen zu einer Vervielfachung des Rechenaufwandes. Deshalb wird in dieser Arbeit eine vereinfachte Methode vorgestellt.

Die meisten Verdeckungen sind klein im Vergleich zum großen Hintergrund. Wird ein größerer Bildausschnitt genommen, verringert sich der Einfluss der Verdeckung und der Hintergrund gewinnt bei der Bildregistrierung ein größeres Gewicht, so dass es wahrscheinlicher wird, dass nur der Hintergrund in die Berechnung eingeht.

Hierzu wird zunächst im rektifizierten Bild mit voller Auflösung über die Korrelation ein Bildausschnitt aus der Mitte des linken Bildes im rechten Bild gesucht (Bild 57 und Bild 58) und dann mit derselben Fenstergröße in Pixel, also beide Male z. B. 32 px Kantenlänge, in auf ein Viertel Kantenlänge verkleinerten Bildern wiederholt (Bild 59 und Bild 60). Die Disparitätssuche läuft in beiden Fällen sehr schnell, da nur in einem Bildstreifen gesucht werden muss. Sind die Disparitäten um mehrere Pixel unterschiedlich, wird von einer Verdeckung ausgegangen und die Translationsberechnung erfolgt auch in verkleinerten Bildern, allerdings dann ungenauer.

Allerdings kommt es trotzdem zu Fehlberechnungen. Durch die Trägheit des Fluggerätes sind Beschleunigungen beschränkt. Sobald diese aufgrund von Berechnungen der Bildverarbeitung überschritten werden, kommen ersatzweise vorhergehende Werte für die Translation zum Tragen.

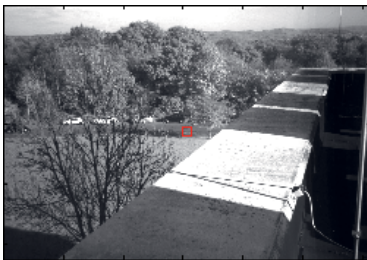


Bild 57: Zu suchendes Teilbild bei Verdeckung im linken Stereobild mit voller Auflösung



Bild 58: Gefundenes Teilbild bei Verdeckung im rechten Stereobild mit voller Auflösung

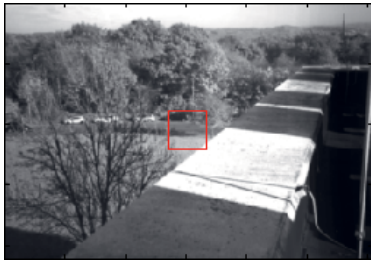


Bild 59: Zu suchendes Teilbild bei Verdeckung im linken auf 1/4 verkleinertes Stereobild

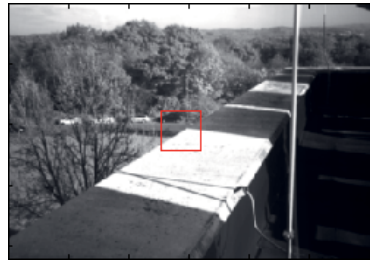


Bild 60: Gefundenes Teilbild bei Verdeckung im rechten auf 1/4 verkleinertes Stereobild

3.3 Klassifizierung der Bewegungsabschätzungsmethoden

Zur Klassifizierung der Verfahren zur Bildregistrierung sind verschiedene Ansätze veröffentlicht. Hier werden zwei oft zitierte Artikel erwähnt, die Abhandlungen von J. Barron et al. [56] und B. Zitová mit J. Flusser [57].

B. Zitová et al. [57] teilen die Methoden in die zwei Gruppen gebietsbasierte und merkmalsbasierte Techniken auf.

Zu den gebietsbasierten Methoden zählen sie unter anderem die normalisierte Kreuzkorrelation, s. Kapitel 3.5, und Phasentechniken, s. Kapitel 3.3.4.

Bei merkmalsbasierten Methoden wird angenommen, dass es korrespondierende Merkmale in den Bildern gibt. Dies wird Kapitel 3.3.5 näher erklärt.

J. Barron et al. [56] teilen die Methoden zur Bildregistrierung in vier Gruppen ein:

1. Differentielle Methoden
2. Gebietspaarung (engl. Matching)
3. Energiebasierte Techniken
4. Phasentechniken

3.3.1 Differentielle Methoden

Die beiden bekanntesten differentiellen Methoden nach Horn & Schunk und Lucas&Kanade werden detailliert in Kapitel 3.4 vorgestellt.

3.3.2 Gebietspaarung

Bei der Gebietspaarung (engl. Matching) wird ein Maß angewendet, welches je nach Verfahren bei kleinstem oder größtem Ergebnis Gebiete mit der größten Übereinstimmung findet. Als Verfahren sind SAD (sum of absolute differences), SSD (sum of squared differences) und NCC (normalized cross correlation) geläufig.

3.3.2.1 SAD

Hier wird die Summe der absoluten Differenzen zwischen einem Teilbild (Template) T und einem zu durchsuchenden Bild I gebildet. Das Teilbild T liegt dann vermutlich an der Stelle x,y im Bild I, wenn diese Summe den kleinsten Wert hat.

$$SAD(x, y) = \sum_{i=1}^n \sum_{j=1}^m |I(x+i, y+j) - T(i, j)| \quad (3.1)$$

3.3.2.2 SSD

Die Summe der quadrierten Differenzen ist ähnlich wie die Summe der absoluten Differenzen. Nur wird anstatt des Absolutwertes das Quadrat gebildet.

$$SSD(x, y) = \sum_{i=1}^n \sum_{j=1}^m (I(x+i, y+j) - T(i, j))^2 \quad (3.2)$$

Ein Vergleich bezüglich SAD, SSD, NCC und weiterer Verfahren ist in einer Dissertation [58] durchgeführt. Dort ergibt sich, dass NCC, insbesondere bei multiplikativen Helligkeitsschwankungen, besonders robust ist. Zu ähnlichen Ergebnissen kommt A. Giachetti [59], der zusätzlich feststellt, dass SSD und NCC in Rechenzeitaufwand und Fehler sich ähnlich verhalten. Die normalisierte Kreuzkorrelation (NCC) wird in Kapitel 3.5 vorgestellt.

3.3.3 Energiebasierte Methoden

Ein oft zitiertes Verfahren ist von D. Heeger [60] entwickelt worden. Die Bilder müssen mit 12 Gaborfiltern (Gaußbandpässe, Bewegungsenergiefilter) gefiltert werden. Aus den Ergebnissen werden dann nach der Methode der kleinsten Fehlerquadrate die geschätzten Verschiebungen ermittelt. Gaborfilter werden u. a. für Fingerabdruckvergleiche und Gesichtserkennungen verwendet. Aufgrund des erwarteten Rechenaufwandes und des eingeschränkten Bereiches auf Verschiebungen im Subpixelbereich wird dieses Verfahren hier nicht weiter in Betracht gezogen.

3.3.4 Phasentechniken

Nach B. Jähne [39] enthält die Phase der Fouriertransformation eines Bildes mehr Informationen zum Bild als die Amplitude der Fouriertransformation.

Zwei Bilder seien um u, v in x - und y -Richtung verschoben:

$$I_2(x, y) = I_1(x - u, y - v) \quad (3.3)$$

Durch Anwendung der Fouriertransformation auf Gleichung (3.3) erhält man:

$$\mathcal{F}_2(\omega_x, \omega_y) = e^{-j(\omega_x u + \omega_y v)} \mathcal{F}_1(\omega_x, \omega_y) \quad (3.4)$$

Die fouriertransformierten Bilder I_1 und I_2 unterscheiden sich also nur durch eine Phasenverschiebung. Diese kann über das Kreuzleistungsspektrum berechnet werden.

$$\frac{\mathcal{F}_1(\omega_x, \omega_y) \mathcal{F}_2^*(\omega_x, \omega_y)}{|\mathcal{F}_1(\omega_x, \omega_y) \mathcal{F}_2^*(\omega_x, \omega_y)|} = e^{-j(\omega_x u + \omega_y v)} \quad (3.5)$$

Die Stelle des Maximums der Rücktransformation des Kreuzleistungsspektrums ergibt die geschätzte Translation (u, v) [61].

Die Vorteile dieser Technik sind die geringe Rauschempfindlichkeit und Resistenz gegen Beleuchtungsänderungen [62].

3.3.5 Merkmalsbasierte Methoden

Bei merkmalsbasierten Methoden wird angenommen, dass es korrespondierende Merkmale in den Bildern gibt. Diese können

ausgezeichnete Punkte, Endpunkte oder Zentren bestimmter Linien oder Gewichtsschwerpunkte verschiedener Gebiete sein. Eine verbreitete Methode ist der Harris Ecken Detektor [63], der zunehmend vom SIFT-Algorithmus [26] und SURF-Algorithmus [24] abgelöst wird.

3.3.5.1 Harris Ecken Detektor

Ecken sind im Idealfall rechtwinklig und haben einen hohen Kontrast zur Umgebung. Die Eigenschaft von Ecken ist, dass sie bei Positionsänderungen unabhängig von der Bewegungsrichtung hohe Intensitätsänderungen verursachen, während z. B. Linien bei Verschiebungen in Linienrichtungen keine Änderungen verursachen.

Ausgehend von der etwas umgeschriebenen Gleichung (3.2), wobei T in Gleichung (3.6) ein um u und v verschobener Ausschnitt aus demselben Bild ist, folgt:

$$SSD_H(u, v) = \sum_{i=1}^n \sum_{j=1}^m \left(I(x_i, y_j) - I(x_i + u, y_j + v) \right)^2 \quad (3.6)$$

Der zweite Term der Gleichung (3.6) kann durch eine Taylorreihenentwicklung approximiert werden zu:

$$I(x + u, y + v) \approx I(x, y) + uI_x(x, y) + vI_y(x, y) \quad (3.7)$$

I_x , I_y sind die jeweiligen Ableitung nach x oder y.

So wird SSD_H zu

$$SSD_H(u, v) \approx \sum_{i=1}^n \sum_{j=1}^m \left(uI_x(x_i, y_j) + vI_y(x_i, y_j) \right)^2 = (u \ v) A \begin{pmatrix} u \\ v \end{pmatrix} \quad (3.8)$$

mit

$$A = \sum_{i=1}^n \sum_{j=1}^m \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \quad (3.9)$$

Die Eigenwerte λ_1 , λ_2 des Strukturensors A lassen Rückschlüsse auf die Intensitätsverteilung von Bildregionen an der Stelle x,y zu.

Es sind drei Fälle zu unterscheiden:

- $\lambda_1 \approx \lambda_2 \approx 0$: Es liegt keine Struktur vor, es handelt sich hier um eine homogene Region.
- $\lambda_1 \gg 0, \lambda_2 \approx 0$: Es liegen eine Kante oder Rampe vor.
- $\lambda_1 \geq 0, \lambda_2 \gg 0$: Eine zweidimensionale Struktur, z. B. eine Ecke, liegt vor.

Um nun korrespondierende Ecken in zwei Bildern und damit die Verschiebung zu ermitteln, kann ein Verfahren angewandt werden, wie es von Z. Zhang et al. in [64] vorgestellt wird. Es wird zunächst um eine Ecke in Bild 61a an der Stelle x_1, y_1 ein Fenster (Korrelationsfenster genannt) definiert. Nachdem in Bild 61b ein Suchfenster um das Gebiet der gesuchten Korrespondenz gelegt ist, wird um alle Punkte im Suchfenster die Korrelation mit dem Korrelationsfenster gebildet. Der Ort mit der besten Korrelation ist dann die Stelle des gesuchten Punktes. Liegt eine ungenaue Vermutung vor, so muss das Fenster vergrößert werden.

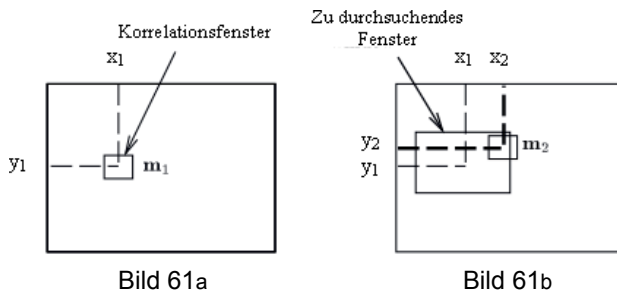


Bild 61: Fenster zur Findung korrespondierender Ecken

3.4 Optischer Fluss, differenzielles Verfahren

Die Grundannahme für die Berechnung des optischen Flusses nach dem differenziellen Verfahren ist, dass ein Szenenpunkt immer – auch nach einer Bewegung – seine Intensität beibehält (engl. brightness consistency).

Zunächst sei der eindimensionale Fall, also z. B. eine Bildzeile, für den optischen Fluss betrachtet. Die Bildzeile habe zum Zeitpunkt t eine Intensitätsfunktion $I(x,t)$. Zu einem späteren Zeitpunkt $t+\Delta t$ ist die Funktion um Δx verschoben, s. Bild 62. Es gilt dann:

$$\underbrace{I(x,t)}_{G(x)} = \underbrace{I(x + \Delta x, t + \Delta t)}_{F(x)} \quad (3.10)$$

Die Ableitung $F'(x)$ der Funktion $F(x)$ an der Stelle x wird angenähert durch:

$$F'(x) \approx \frac{G(x_1) - F(x_1)}{\Delta x} \quad (3.11)$$

Damit kann die Verschiebung Δx geschätzt werden zu:

$$\Delta x \approx \frac{G(x_1) - F(x_1)}{F'(x_1)} \quad (3.12)$$

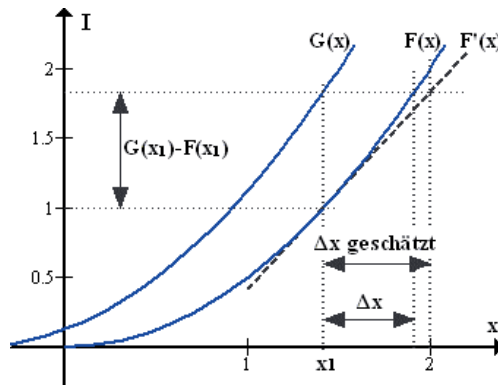


Bild 62: Differenzielle Methode zur Bestimmung des optischen Flusses für den eindimensionalen Fall

Für den zweidimensionalen Fall gilt:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (3.13)$$

Kleine Verschiebungen nach einer Zeit dt um dx und dy können durch eine Taylorreihenentwicklung um den Punkt (x, y, t) dargestellt werden:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \varepsilon \quad (3.14)$$

ε fasst die Terme höherer Ordnung zusammen. Diese werden in der Regel zur Berechnung des optischen Flusses vernachlässigt, d. h., es wird annähernd lineares Verhalten in der Umgebung von (x, y, t) vorausgesetzt. Aus den Gleichungen (3.13) und (3.14) folgt dann:

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \quad (3.15)$$

Dies ist gleich:

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad (3.16)$$

Meist wird Gleichung (3.16) vereinfacht geschrieben:

$$I_x u + I_y v + I_t = 0 \quad (3.17)$$

Die Gleichung (3.17) wird im englischen Sprachraum „Optical Flow Constraint Equation“ genannt und sagt aus, dass die Intensität eines abgebildeten Szenenpunktes sich bei Verschiebung und auch zeitlich nicht ändern darf.

I_x, I_y sind die partiellen Ableitungen der Intensitätswerte in x- und y-Richtung; u und v sind die gesuchten Flussvektoren in x- und y-Richtung. Sie werden auch Geschwindigkeitsvektoren genannt, weil sie sich aus der Ableitung des Weges nach der Zeit errechnen.

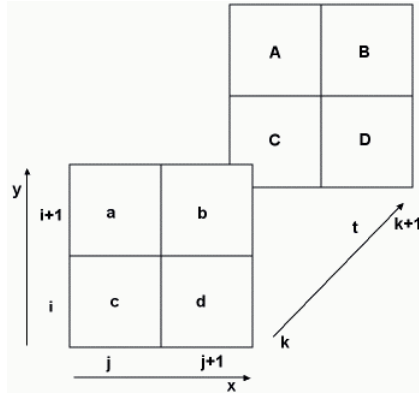


Bild 63: Vierpixelausschnitte aus zwei Abbildungen

I_t und die partiellen Ableitungen I_x, I_y können, wie unten gezeigt, innerhalb zwei aufeinanderfolgender Bilder, s. Bild 63, gebildet werden:

$$I_x = \frac{1}{4} (I_{i,j+1,k} - I_{i,j,k} + I_{i+1,j+1,k} - I_{i+1,j,k} + I_{i,j+1,k+1} - I_{i,j,k+1} + I_{i+1,j+1,k+1} - I_{i+1,j,k+1}) \quad (3.18)$$

$$I_y = \frac{1}{4} (I_{i+1,j,k} - I_{i,j,k} + I_{i+1,j+1,k} - I_{i,j+1,k} + I_{i+1,j,k+1} - I_{i,j,k+1} + I_{i+1,j+1,k+1} - I_{i,j+1,k+1}) \quad (3.19)$$

$$I_t = \frac{1}{4} \left(I_{i,j,k+1} - I_{i,j,k} + I_{i+1,j,k+1} - I_{i+1,j,k} + I_{i,j+1,k+1} - I_{i,j+1,k} + I_{i+1,j+1,k+1} - I_{i+1,j+1,k} \right) \quad (3.20)$$

Anschaulicher dargestellt, sehen obige Gleichungen folgendermaßen aus:

$$I_x = \frac{1}{4} (d - c + b - a + D - C + B - A) \quad (3.21)$$

$$I_y = \frac{1}{4} (a - c + b - d + A - C + B - D) \quad (3.22)$$

$$I_t = \frac{1}{4} (C - c + A - a + D - d + B - b) \quad (3.23)$$

Alleine mit der Gleichung (3.17) können die Werte von u und v nicht ermittelt werden. Es müssen weitere Einschränkungen (engl. constraints) gefunden werden. Die beiden bekanntesten Ansätze, der von Horn&Schunk und der von Lucas&Kanade, werden hier vorgestellt.

3.4.1 Smoothness Constraint nach B. K. Horn und B. G. Schunck

B. K. Horn und B. G. Schunck stellen fest, dass bei unabhängiger Verteilung der Helligkeitsmuster wenig Hoffnung auf die Berechnung des optischen Flusses für alle Bildpunkte besteht [48]. Deshalb werden Annahmen, die auf die meisten Szenen zutreffen, getroffen. Hierzu gehört, dass Gegenstände undurchsichtig und endlich groß sind und sich fest bewegen oder deformieren. Mit diesen Annahmen kann davon ausgegangen werden, dass benachbarte Punkte sich ähnlich bewegen und damit das Flussfeld sich nur langsam ändert.

Das Minimieren der Summe der quadrierten partiellen Ableitungen erfüllt diese Bedingung:

$$\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \rightarrow 0 \quad (3.24)$$

Da nicht angenommen werden kann, dass die Funktionen (3.17) und (3.24) exakt gleich null sind, wird ein Energiefunktional E^2 gebildet, welches durch Minimierung die Werte für den optischen Fluss liefert.

$$E^2 = \iint E_I^2 + \alpha^2 E_S^2 dx dy \quad (3.25)$$

mit

$$E_I = I_x u + I_y v + I_t \quad (3.26)$$

und

$$E_S = \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \quad (3.27)$$

Entsprechend der Variationsrechnung nach Euler-Lagrange [37] lässt sich das Integral aus Gleichung (3.25) auflösen zu:

$$\begin{aligned} I_x^2 u + I_x I_y v &= \alpha^2 \nabla^2 u - I_x I_t \\ I_x I_y u + I_y^2 v &= \alpha^2 \nabla^2 v - I_y I_t \end{aligned} \quad (3.28)$$

Der Regularisierungsparameter α dient dazu, Fehler, die durch Quantisierung und Rauschen entstehen, zu gewichten. Die Laplace Operatoren $\nabla^2 u$ und $\nabla^2 v$ können durch folgende Gleichungen approximiert werden:

$$\nabla^2 u \approx k (\bar{u}_{i,j,k} - u_{i,j,k}) \text{ und } \nabla^2 v \approx k (\bar{v}_{i,j,k} - v_{i,j,k}) \quad (3.29)$$

\bar{u} und \bar{v} sind lokale Mittelwerte, während k ein Proportionalitätsfaktor ist.

Die Gleichungen (3.28) können damit umgeschrieben werden zu:

$$\begin{aligned} (\alpha^2 + I_x^2) u + I_x I_y v &= (\alpha^2 \bar{u} - I_x I_t) \\ I_x I_y u + (\alpha^2 + I_y^2) v &= \alpha^2 \bar{v} - I_y I_t \end{aligned} \quad (3.30)$$

Daraus kann nach dem Gauß-Seidel-Verfahren [37] eine iterative Methode zur Bestimmung der Flusswerte gebildet werden:

$$\begin{aligned} u^{n+1} &= \bar{u}^n - \frac{I_x [I_x \bar{u}^n + I_y \bar{v}^n + I_t]}{(\alpha^2 + I_x^2 + I_y^2)} \\ v^{n+1} &= \bar{v}^n - \frac{I_y [I_x \bar{u}^n + I_y \bar{v}^n + I_t]}{(\alpha^2 + I_x^2 + I_y^2)} \end{aligned} \quad (3.31)$$

Zu Beginn sind \bar{u}^n und \bar{v}^n unbekannt und werden als null angenommen.

Mit den Gleichungen (3.31) lassen sich die Flussvektoren für jedes Pixel errechnen.

3.4.2 Flusskonstanz nach Lucas&Kanade

Im selben Jahr – 1981 – wie Horn&Schunck entwickelten Lucas&Kanade [65] einen anderen Ansatz zur Bestimmung des optischen Flusses. Dieser Ansatz bezieht sich ebenfalls auf die Optical Flow Constraint Equation (3.17). Die Unterbestimmtheit wird hier durch die Annahme gleicher Bewegungen innerhalb eines kleinen Szenenausschnitts, also des gleichen Flusses u, v innerhalb eines kleinen Bildausschnitts, gelöst. Gleichung (3.17) wird dann für diesen Bereich, der quadratisch mit der Kantenlänge m um das betreffende Pixel ist, erweitert zu:

$$\underbrace{\begin{pmatrix} I_x(1,1) & I_y(1,1) \\ I_x(1,2) & I_y(1,2) \\ \vdots & \vdots \\ I_x(x,y) & I_y(x,y) \\ \vdots & \vdots \\ I_x(m,m) & I_y(m,m) \end{pmatrix}}_A \underbrace{\begin{pmatrix} u \\ v \end{pmatrix}}_U = - \underbrace{\begin{pmatrix} I_t(1,1) \\ I_t(1,2) \\ \vdots \\ I_t(x,y) \\ \vdots \\ I_t(m,m) \end{pmatrix}}_b \quad (3.32)$$

Das Gleichungssystem ist überbestimmt.

Ziel ist es,

$$\|AU - b\|^2 \quad (3.33)$$

zu minimieren.

Aus

$$AU = b \quad (3.34)$$

wird die Normalgleichung [37] gebildet:

$$U = (A^T A)^{-1} A^T b \quad (3.35)$$

Mit Gleichung (3.32) wird dies zu:

$$U = \left(\begin{array}{cc} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{array} \right)^{-1} \left(\begin{array}{c} \sum I_x I_t \\ \sum I_y I_t \end{array} \right) \quad (3.36)$$

Wenn beide Eigenwerte λ_1 und λ_2 von $A^T A$ ungleich null sind, ist $A^T A$ invertierbar und der optische Fluss U kann errechnet werden. Sind z. B. beide Eigenwerte null, so ist der betreffende Bereich in der Abbildung homogen und der optische Fluss kann nicht errechnet werden.

Es ist sinnvoll, bei der Berechnung die Ableitungen in der nahen Umgebung des betreffenden Pixels, für welches der optische Fluss ermittelt wird, höher zu gewichten und den Einfluss der entfernteren Ableitungen zu verringern. Man führt deshalb eine Gewichtsfunktion W ein [66]. Gleichung (3.36) wird dann zu:

$$U = \left(\begin{array}{cc} \sum W^2 I_x^2 & \sum W^2 I_x I_y \\ \sum W^2 I_x I_y & \sum W^2 I_y^2 \end{array} \right)^{-1} \left(\begin{array}{c} \sum W^2 I_x I_t \\ \sum W^2 I_y I_t \end{array} \right) \quad (3.37)$$

Es kann also hierzu z. B. mit $m = 5$, s. Gleichung (3.32), ein Gaußfilter genommen werden:

$$W = \frac{1}{4096} \begin{pmatrix} 1 & 14 & 34 & 14 & 1 \\ 14 & 196 & 476 & 196 & 14 \\ 34 & 476 & 1156 & 476 & 34 \\ 14 & 196 & 476 & 196 & 14 \\ 1 & 14 & 34 & 14 & 1 \end{pmatrix} \quad (3.38)$$

Zur schnellen Berechnung der Filterung sind Überlegungen aus dem Anhang, Kapitel A.2, hilfreich.

3.4.3 Vergleich der Verfahren nach Horn&Schunck und Lucas&Kanade

Die Lucas&Kanade-Methode ist ein lokales Verfahren und liefert ein nicht so dichtes Flussfeld, wie es bei dem globalen iterativen Verfahren nach Horn&Schunck möglich ist [67]. Der Fluss ist in kleiner Umgebung bei Lucas&Kanade konstant, allerdings auch weniger rauschempfindlich. Ein Ansatz, die Vorteile beider Verfahren zu kombinieren, ist in [67] zu finden.

Nur wenige relevante Vergleiche [68], [69], [70], [56], [71], [72], [73], [74] verschiedener Verfahren wurden veröffentlicht. Hierbei wurden zumeist synthetische Bilder (z. B. Yosemite Park Sequence) benutzt, so dass Einflüsse auf Bilder realer Szenen unberücksichtigt blieben und damit die Bedingungen

zur Gewinnung des optischen Flusses leicht erfüllt wurden (frei von Beleuchtungsschwankungen, Reflexionen und Moiréeffekten digitaler Aufnahmen). Die Ergebnisse der vorliegenden Untersuchungen sind schwer vergleichbar, da unterschiedliche Bewertungskriterien für die Güte der Berechnungen zum optischen Fluss herangezogen wurden. In fast allen Untersuchungen wurde das Gradientenverfahren von Lucas & Kanade mit einbezogen. Dabei bewährte sich dieser Algorithmus durch Effizienz und Genauigkeit. Keiner der Tests untersuchte allerdings bisher die Eignung auf Koppelnavigation für Fluggeräte. Eigene Versuche zeigen an realen Bildern doch Schwächen der differenziellen Methoden, die zum Teil auf den im folgenden Abschnitt beschriebenen Einschränkungen beruhen.

3.4.4 Gaußpyramide

Da die differentiellen Methoden annähernd Linearität voraussetzen, lassen sich, streng genommen, nur Verschiebungen kleiner als ein Pixel berechnen. Bei Versuchen mit realen Bildern, s. Kapitel 4.1, S.78, ergab sich die maximal richtig berechenbare Verschiebung zu 1.16 Pixel. Diese Einschränkung lässt sich durch Bildverkleinerungen in Grenzen beheben.

Für jede Bildgröße gilt, dass Verschiebungen von einem Pixel, aber nicht viel mehr, berechnet werden können. Wird die Verschiebung z. B. zweier auf halbe Breite und Höhe verkleinerten Bilder zu U errechnet, so ist die Verschiebung dann im Originalbild $2U$.

Zur Berechnung werden Bildverkleinerungen in mehreren Eben durchgeführt. Dies führt zur s. g. Gaußpyramide, s. Bild 64.

Auch wenn die Pyramide mit beliebigen Skalierungsfaktoren erstellt werden kann, ist es einfacher, bei jeder Stufe die Bildgröße auf halbe Kantenlänge zu verkleinern, denn dann ist keine Interpolation notwendig, weil nur jedes zweite Pixel zur Verkleinerung herangezogen wird.

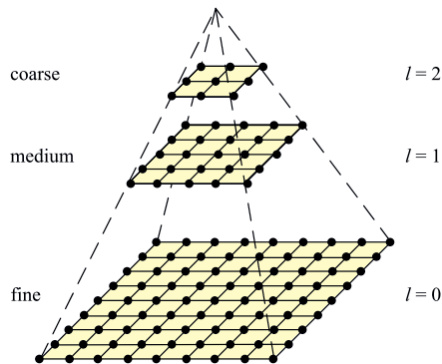


Bild 64: Übliche Gaußpyramide. Jede höhere Ebene l hat die halbe Kantenlänge, also ein Viertel der Pixel der vorhergehenden Ebene (Quelle: [75])

Um das Abtasttheorem nicht zu verletzen, wird jedes Bild vor der Verkleinerung tiefpassgefiltert. Das Gaußfilter ist wegen der Eigenschaft, eine maximale Flankensteilheit ohne Überspringen bei gleichzeitig schmalem Filter im Ortsbereich zu haben, besonders geeignet, siehe auch Anhang A.2. Die übliche Gaußpyramide wird entwickelt vom Originalbild, Ebene $l=0$, bis zu einer Ebene $l=n$, indem jeweils jeder vierte Pixelwert über ein Gaußfilter gebildet wird.











Pyramide für ein Bild	Ebene	Pyramide für darauf folgendes Bild
 32 x 32	4	
 64 x 64	3	
 128 x 128	2	
 256 x 256	1	
 512 x 512	0	

Bild 65: Gaußpyramiden für zwei aufeinander folgende Bilder. Die Bilder der Ebene 0 sind die Originalbilder. Die anderen Ebenen werden durch ein 5 x 5-Gaußfilter und sukzessive Elimination jedes zweiten Pixels in x- und y-Richtung gebildet.

Mit der Gaußpyramide lassen sich also größere Pixelversätze errechnen. Gilt zum Beispiel, dass in einem Bildpaar je Ebene der Versatz bis zu zwei Pixel berechnet werden kann, so gilt Tabelle 3 für die Ebenen nach Bild 65:

Tabelle 3: Maximal errechenbarer Gesamtversatz bei zwei Pixel Versatz pro Bildebene der Gaußpyramide.

Ebene	Maximaler Versatz
4	2
3	$2*2+2 = 6$
2	$2*6+2 = 14$
1	30
0	62

Allgemein gilt:

$$\text{Maximaler Gesamtversatz} = u(2^n - 1) \quad (3.39)$$

u ist der maximale errechenbare Versatz in einer Bildebene, n die Anzahl der Ebenen.

Die Anzahl der Ebenen ist begrenzt durch die Subpixelgenauigkeit, da sich die Fehler bis zur Ebene 0 potenzieren. Eigene Versuche haben ergeben, dass ab Ebene 5 bei 512 x 512 px großen Originalaufnahmen – entspricht einem 16 x 16 px Bild in Ebene 5 – die Berechnungen ungenau werden. Dies wird auch in [76] bestätigt und ist aus Bild 65 ersichtlich, denn dort sind schon kaum Details erkennbar und homogene Bildanteile nehmen zu.



Bild 66: Vergrößerung des Bilds der Ebene 4 aus Bild 65 oben links

Zu beachten ist, dass für eine schnelle Berechnung die Außenmaße vom Originalbild (Ebene 0) bis zur vorletzten Ebene jeweils ganzzahlig durch zwei teilbar sein müssen, da sonst bei der Bildverkleinerung interpoliert werden

müsste und die Zuordnung der Flusswerte von Ebene zu Ebene schwieriger wäre.

In Kapitel 4.1 wird gezeigt, welcher Versatz in realen Bildern pro Ebene errechnet werden kann.

3.5 Optischer Fluss, Korrelationsmethode

Im Gegensatz zur Gradientenmethode, die für jedes Pixel den optischen Fluss bestimmt, wird beim Region-Based Matching nur die Verschiebung eines Gebietes ermittelt. Ein Verfahren ist die Kreuzkorrelation. Ein Teilbild der ersten Aufnahme wird in einer zweiten Aufnahme gesucht, indem dort für alle möglichen Positionen die Kreuzkorrelation gebildet wird. Die Position mit der maximalen Korrelation ist dann die wahrscheinlichste Position desselben Szenenausschnittes.

Ein Beispiel ist in den Bild 67a und b zu finden, in denen das Teilbild jeweils schwarz umrahmt ist. Die Kamera hat sich über dem Boden nach rechts bewegt. Deshalb ist das Teilbild mit der maximalen Korrelation im nachfolgenden Bild weiter links zu finden.



Bild 67a: Bild und gesuchtes Teilbild



Bild 67b: Nachfolgendes Bild und gefundenes Teilbild

3.5.1 ZNCC

Um die Empfindlichkeit gegen Helligkeitsschwankungen sowohl in additiver als auch in multiplikativer Form zu vermindern, werden die Werte der sich überlappenden Flächen normalisiert und von allen Pixeln der Flächenmittelwert abgezogen. Man erhält die mittelwertfreie normalisierte Kreuzkorrelation (engl. ZNCC, Zero mean Normalized Cross Correlation) [77][78].

$$ZNCC(u, v) = \frac{\sum_{x,y} [I(u+x, v+y) - \bar{I}(u, v)][T(x, y) - \bar{T}]}{\sqrt{\sum_{x,y} [I(u+x, v+y) - \bar{I}(u, v)]^2} \sqrt{\sum_{x,y} [T(x, y) - \bar{T}]^2}} \quad (3.40)$$

hierbei ist $\text{ZNCC}(u,v)$ der normierte und mittelwertfreie Korrelationskoeffizient an der Stelle u,v .

$I(x,y)$ ist der Helligkeitswert an der Stelle x,y des zu durchsuchenden Bildes.

$T(x,y)$ ist der Helligkeitswert an der Stelle x,y des gesuchten Teilbildes (template).

\bar{I} und \bar{T} sind jeweils die Helligkeitsmittelwerte in den durch die Teilbildgröße N_x , N_y vorgegebenen Umgebungen.

Zur schnelleren Berechnung kann die Gleichung (3.40) weiter aufgelöst werden.

Für T lässt sich schnell und einmalig das mittelwertfreie Bild berechnen. Der Normierungsunterterm für das Teilbild wird zusammengefasst zu:

$$\tau(x,y) = T(x,y) - \bar{T} \quad (3.41)$$

Für den Mittelwert \bar{T} und \bar{I} gilt:

$$\bar{T} = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} T(x,y) \quad (3.42)$$

$$\bar{I}(u,v) = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} I(u+x, v+y) \quad (3.43)$$

$\bar{I}(u,v)$ muss für jedes u,v neu berechnet werden.

Zur besseren Übersicht werden die unabhängigen Variablen u, v, x, y , die in Gleichung (3.40) explizit aufgeführt werden, im Folgenden nicht mitgeführt. Die Berechnungen sind angelehnt an einer Veröffentlichung von K. Briechele und U. Hanebeck [79].

Der Zähler Z der ZNCC lässt sich so umformen:

$$\begin{aligned} Z &= \sum_{x,y} [I - \bar{I}] [T - \bar{T}] = \sum_{x,y} [I - \bar{I}] \tau \\ &= \sum_{x,y} I\tau - \bar{I}\tau = \sum_{x,y} I\tau - \bar{I} \sum_{x,y} \tau \end{aligned} \quad (3.44)$$

Da Folgendes gilt,

$$\begin{aligned}
\sum_{x,y} \tau &= \sum_{x,y} T - \sum_{x,y} \left(\frac{\sum_{x,y} T}{N_x N_y} \right) = \sum_{x,y} T - \left(\frac{\sum_{x,y} T}{N_x N_y} \right) \sum_{x,y} 1 \\
&= \sum_{x,y} T - \left(\frac{\sum_{x,y} T}{N_x N_y} \right) N_x N_y = 0
\end{aligned} \tag{3.45}$$

vereinfacht sich Gleichung (3.44) zu:

$$\begin{aligned}
Z &= \sum_{x,y} I \tau \\
Z(u, v) &= \sum_{x,y} I(u + x, v + y) \tau(x, y)
\end{aligned} \tag{3.46}$$

Wird das Bild I in x- und y-Richtung zu Bild I' gespiegelt, entspricht obige Gleichung einer Faltung:

$$Z(u, v) = \sum_{x,y} I'(u - x, v - y) \tau(x, y) \tag{3.47}$$

Somit können durch Fouriertransformation und Rücktransformation die nicht normierten Korrelationskoeffizienten ermittelt werden.

$$Z(u, v) = \mathcal{F}^{-1} \left[\mathcal{F}(I') \mathcal{F}^*(\tau) \right] \tag{3.48}$$

Hierbei ist \mathcal{F}^{-1} die inverse Fouriertransformation und \mathcal{F}^* die konjugiert komplexe Fouriertransformation. Durch die Fouriertransformation kann die Anzahl der Multiplikationen, insbesondere bei relativ großen Teilbildern, erheblich reduziert werden. Sind $M_x = M_y = M$ die Kantenlängen des zu durchsuchenden Bildes und $N_x = N_y = N$ Kantenlängen des Teilbildes, so erfordert die direkte Faltung nach Gleichung (3.47)

$$N^2 (M - N + 1)^2 \tag{3.49}$$

Multiplikationen und mit Fouriertransformationen nach Gleichung (3.48)

$$12M^2 \log_2 M \tag{3.50}$$

Multiplikationen [77]. Mit $N = 32$ und $M = 512$ ist das Verhältnis der Anzahl der Multiplikationen von Gleichung (3.49) zu Gleichung (3.50) $2.4 \cdot 10^8$ zu $2.8 \cdot 10^7$.

Für die Berechnung des Nenners aus (3.40) ergeben sich die folgenden Vereinfachungen.

Der Ausdruck $\sqrt{\sum_{x,y} [T(x,y) - \bar{T}]^2}$ kann einmalig berechnet werden und gilt dann für alle u,v.

Der andere Term des Nenners ist aufwendiger zu berechnen und lässt sich so umformen:

$$\sum_{x,y} [I(u+x, v+y) - \bar{I}(u,v)]^2 = \sum_{x,y} [I - \bar{I}]^2 = \sum_{x,y} I^2 - 2 \sum_{x,y} I \bar{I} + \sum_{x,y} \bar{I}^2 \quad (3.51)$$

$\sum_{x,y} I^2$ lässt sich für jedes u,v über Summentabellen [80] errechnen, s. auch Anhang A.3.

Mit Hilfe von (3.43) wird der zweite Term von (3.51) umgeformt zu:

$$2 \sum_{x,y} I \bar{I} = 2 \bar{I} \sum_{x,y} I = 2 \frac{1}{N_x N_y} \sum_{x,y} I \sum_{x,y} I = 2 \frac{1}{N_x N_y} \left(\sum_{x,y} I \right)^2 \quad (3.52)$$

Der dritte Term ergibt:

$$\sum_{x,y} \bar{I}^2 = \sum_{x,y} \frac{\left(\sum_{x,y} I \right)^2}{(N_x N_y)^2} = \frac{\left(\sum_{x,y} I \right)^2}{(N_x N_y)^2} \sum_{x,y} 1 = \frac{\left(\sum_{x,y} I \right)^2}{N_x N_y} \quad (3.53)$$

Damit wird der Nenner von Gleichung (3.40) zu

$$\sqrt{\sum_{x,y} I^2 - \frac{1}{N_x N_y} \left[\sum_{x,y} I \right]^2} \sqrt{\sum_{x,y} \tau^2} \quad (3.54)$$

Die jeweiligen Summen I^2 und I für jedes u,v lassen sich auch hier schnell über Summentabellen ermitteln.

Mit diesen Umformungen kann die ZNCC wesentlich zügiger berechnet werden.

3.5.2 NCC

Die ZNCC ist in ihrer Berechnung aufwendig. Daher wird hier die normalisierte Kreuzkorrelation (engl. Normalized Cross Correlation, NCC) vorgestellt. Die Vor- und Nachteile der ZNCC zur NCC werden im nachfolgenden Kapitel (3.5.3) erläutert. Die normalisierte Kreuzkorrelation NCC wird über folgende Gleichung errechnet:

$$NCC(u, v) = \frac{\sum_{x,y} I(x+u, y+v)T(x, y)}{\sqrt{\sum_{x,y} I^2(x+u, y+v)} \sqrt{\sum_{x,y} T^2(x, y)}} \quad (3.55)$$

Der Ort der größten Übereinstimmung ist an der Stelle u, v , an der $NCC(u, v)$ maximal ist. Die linke Wurzel des Nenners kann wieder über Summentabellen errechnet werden. Die rechte Wurzel muss nur einmal berechnet werden.

Trotzdem ist die konventionelle Berechnung der NCC bisher zu aufwendig, um mit einem normalen Prozessor einen Echtzeitbetrieb einer Regelung zu ermöglichen.

In einer Veröffentlichung von S. D. Wei und S. H. Lai [81] wird ein Verfahren vorgestellt, welches die Berechnung mit Hilfe einer sukzessiven Approximation und eines Eliminationsverfahrens wesentlich schneller ablaufen lässt. Dieses Verfahren beruht auf der Schwarzschen Ungleichung:

$$\int_a^b I(x)T(x)dx \leq \sqrt{\int_a^b I^2(x)dx} \sqrt{\int_a^b T^2(x)dx} \quad (3.56)$$

Im Durchschnitt ist diese Methode etwa zehnmal schneller als die direkte Berechnung der NCC. Allerdings sind die Rechenzeiten nicht vorhersehbar und können je nach Bildinhalt die Rechenzeiten der direkten NCC sogar überschreiten. Deshalb ist das Verfahren für die Echtzeitberechnung nicht geeignet.

Das Verfahren wird im Anhang, Kap. A.4, detailliert erklärt.

3.5.3 Eigenschaften von NCC und ZNCC

Unterschiedliche Aufnahmesituationen können die Bildhelligkeit verändern. Dies kann verursacht werden unter anderem durch:

- Unterschiedliche Belichtungszeiten
- Unterschiedliche Blenden

Dadurch ändert sich die Intensität des Teilbildes T in Bezug auf das Bild I um einen Faktor k . Man spricht von multiplikativer Helligkeitsänderung, denn T wird in Gleichung (3.55) dann zu kT . Dies entspricht einer Kontraständerung. Die NCC ist durch ihre Normalisierung invariant zu diesem Faktor, wie aus folgender Gleichung ersichtlich ist:

$$\begin{aligned}
NCC(u, v) &= \frac{\sum_{x,y} I(x+u, y+v) k T(x, y)}{\sqrt{\sum_{x,y} I^2(x+u, y+v)} \sqrt{\sum_{x,y} k^2 T^2(x, y)}} \\
&= \frac{\sum_{x,y} I(x+u, y+v) T(x, y)}{\sqrt{\sum_{x,y} I^2(x+u, y+v)} \sqrt{\sum_{x,y} T^2(x, y)}}
\end{aligned} \tag{3.57}$$

In einigen Fällen, insbesondere bei verschiedenen Kameras, kommen noch additive Intensitätsänderungen hinzu.

Zusammen mit der multiplikativen Änderung wird die Intensität des Teilbildes dann zu:

$$T \rightarrow kT + c \tag{3.58}$$

k ist dann die multiplikative Änderung (Kontrast) und c die additive Änderung (Helligkeit).

Der Übersicht halber wird $[I(u+x, v+y) - \bar{I}(u, v)]$ zu \tilde{I} zusammengefasst. T wird mit $kT+c$ substituiert. Gleichung (3.40) wird dann zu:

$$\begin{aligned}
ZNCC(u, v) &= \frac{\sum_{x,y} \tilde{I}[kT + c - (k\bar{T} + c)]}{\sqrt{\sum_{x,y} \tilde{I}^2} \sqrt{\sum_{x,y} [kT + c - (k\bar{T} + c)]^2}} \\
&= \frac{\sum_{x,y} \tilde{I}[T - \bar{T}]}{\sqrt{\sum_{x,y} \tilde{I}^2} \sqrt{\sum_{x,y} [T - \bar{T}]^2}}
\end{aligned} \tag{3.59}$$

Daraus folgt, dass die ZNCC invariant ist zu additiven und multiplikativen Lichtänderungen.

Die Vorteile der ZNCC gegenüber der NCC sind eher geringfügig. In der Regel zeigen Versuche mit realen Bildern geringe Abweichungen in der Erkennung [82][83] oder sagen aus, dass NCC mit realen Bildern sogar besser ist als ZNCC [59]. Die Rechenzeit ist bei der ZNCC etwa doppelt so groß wie bei der NCC [84].

Allerdings sind die Verhältnisse anders bei Aufnahmen derselben Szene durch verschiedene Kameras. Hier kann der Gleichanteil der Helligkeit selbst beim gleichen Modell desselben Herstellers dominant werden und somit die Anwendung der ZNCC, also z. B. bei Stereoaufnahmen, erforderlich sein.

3.6 Zusammenfassung Kapitel 3

Probleme der Bildregistrierung sind hauptsächlich unterschiedliche Bildintensitäten, das Blendenproblem – insbesondere bei Kanten – und Verdeckung. Die Gradientenmethode zur Bestimmung des optischen Flusses und die Kreuzkorrelation sind die beiden favorisierten Methoden. Vorteil der Gradientenmethode im Vergleich zur Kreuzkorrelation ist die implizite Subpixelgenauigkeit. Der Nachteil ist die globale Intensitätsempfindlichkeit, die die Kreuzkorrelation (NCC oder ZNCC) nicht aufweist. Größere Verschiebungen lassen sich bei dem Gradientenverfahren mit Hilfe von Gaußpyramiden erfassen, aber – wie im folgenden Kapitel gezeigt wird – ist diese Methode stark begrenzt.

4 Vergleich der Verfahren nach Lucas&Kanade und NCC

In diesem Kapitel werden anhand einiger Versuche beide Verfahren verglichen bezüglich:

- Maximal erkennbarer Verschiebung
- Helligkeitsempfindlichkeit
- Rotationsempfindlichkeit
- Berechnungszeit
- Subpixelgenauigkeit

4.1 Reichweite der Erkennung

Der optische Fluss nach der differenziellen Methode erlaubt zunächst nur einen Versatz kleiner oder gleich einem Pixel zu bestimmen. Es kann daher nur ein Versatz innerhalb der Strecke zwischen zwei Pixeln errechnet werden, da z. B. die Ableitung zwischen einem 2. und 3. Pixel negativ zur Ableitung zwischen dem 1. und 2. Pixel sein könnte und damit die Berechnung falsch wäre.

Um maximal berechenbare Verschiebungen der Gradientenmethode ohne Gaußpyramiden zu verifizieren, werden aus mit dem Flugsystem aufgenommenen Filmsequenzen jeweils Bildpaare innerhalb derselben Aufnahme, die um ein, zwei oder drei Pixel zueinander verschoben sind, auf ihren optischen Fluss untersucht. Jeweils 2500 Bildpaare verschiedener Landschaftsszenen wurden untersucht. In Bild 68 sind die Ergebnisse dargestellt.

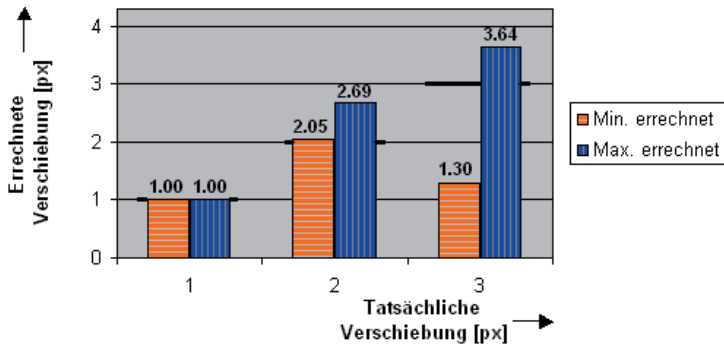


Bild 68: Gegenüberstellung von tatsächlichen Bildverschiebungen zu errechneten Bildverschiebungen bei dem Verfahren nach Lucas&Kanade. Dargestellt sind die errechneten Minimal- und Maximalwerte aus 2500 Bildpaaren.

Aus Bild 68 kann entnommen werden, dass die Berechnung für genau einen Pixel Versatz immer exakt ist (auf 1/100 Pixel genau). Bei zwei Pixel Versatz ist die maximale Abweichung schon 0.69 Pixel und ab drei Pixel Versatz wird das Ergebnis unbrauchbar.

Nimmt man nun an, dass die Genauigkeit bei zwei Pixel Versatz ausreicht, ließe sich mit der Lucas&Kanade-Methode mit Hilfe der Gaußpyramiden nach Kapitel 3.4.4 bei fünf Pyramidenebenen ein Gesamtversatz von bis zu 62 Pixel, s. Gleichung (3.39), errechnen.

Die Abweichungen in Tabelle 4 wurden mit Ausschnitten aus demselben Bild errechnet. Die Realität mit verschiedenen Bildern, zeitlich und räumlich versetzte Aufnahmen der gleichen Szene, ergibt einen deutlich niedrigeren berechenbaren Gesamtversatz. Hierzu wurde aus verschiedenen Filmsequenzen jeweils der optische Fluss aus einem Bild und dem 6. nachfolgenden Bild errechnet. Die ermittelten Versätze wurden mit den Berechnungen der NCC verglichen. Die ersten falschen Werte ergaben sich bei einem Versatz von 36 Pixel, während ab 40 Pixel Versatz die Lucas&Kanade-Methode deutlich versagte, s. Tabelle 4.

Tabelle 4: Gegenüberstellung vom tatsächlichen Bildversatz zum errechneten Bildversatz bei dem Verfahren nach Lucas&Kanade mit verschiedenen Bildern und Pyramidenbearbeitung

Versatz [Pixel] →	≤36	>36
Lucas&Kanade Abweichung von NCC	±<1	unbrauchbar

Bei 5 Pyramidenebenen kann nach Gleichung (3.39) umgekehrt der maximal erfassbare Versatz zu 1.16 Pixel errechnet werden.

Das heißt, in realen Bildern sind mit der Lucas&Kanade-Methode Verschiebungen innerhalb einer Bildebene der Gaußpyramide, die geringfügig über ein Pixel hinausgehen, zunehmend unsicher.

4.2 Helligkeitsempfindlichkeit

Durch Helligkeitsänderungen der beobachteten Szene kann sich die Belichtung (aufgrund der Aufnahmezeit oder der Blende) einer Kamera ändern. Um dies nachzubilden, wurde nach einem Flug über eine Landschaft die Intensität jeden zweiten Bildes durch eine Multiplikation mit 0.95 um 5 % geändert. Wie erwartet, versagt die differentielle Methode dann. Bild 69 zeigt ein aus 29 Bildern zusammengesetztes Mosaikbild¹, in dem die Fehlererkennung deutlich wird, und den chaotischen Fluss bei einem Bildpaar. Helligkeitsänderungen unter 1 % sind gerade noch verträglich bei der Methode nach Lucas&Kanade.

Die NCC ist erwartungsgemäß für Helligkeitsänderungen unempfindlich, s. Bild 70.

¹ Ein Mosaikbild ist ein aus mehreren sich inhaltlich überlappenden Bildern zusammengesetztes Gesamtbild.

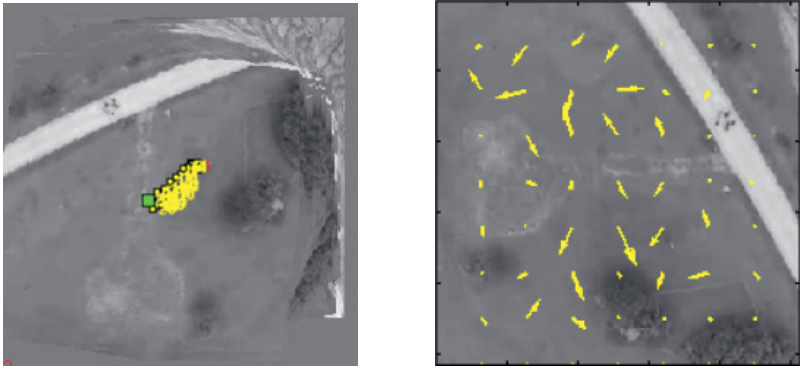


Bild 69: Differentielle Methode bei Helligkeitsänderung um 5 %, Links: Mosaikbild, rechts chaotischer optischer Fluss zwischen zwei Bildern



Bild 70: Mosaikbild bei NCC mit Helligkeitsänderungen
Quelle: Bild 69 und Bild 70, eigene Veröffentlichung [85]

4.3 Rotationsempfindlichkeit

Der zu betrachtende Bereich der Gierwinkel des Experimentierluftschiffes ist maximal $20^\circ/\text{s}$. Mit einer Bildrate der Kamera von 12 Bilder/s sind dies maximal 1.7° pro Bildpaar.

Um die Winkelpräzision zu ermitteln, wurden als Vorlage verschiedene Abbildungen, s. Bild 71 bis Bild 73, 115-mal mit Hilfe von Matlab[®] um jeweils ein Grad um das Zentrum gedreht. Bei jedem Drehwinkel wurden dann 100 Bildpaare für die drei Abbildungen berechnet, indem z. B. für 4° Drehwinkel das erste mit dem fünften Bild, das zweite mit dem 6. Bild usw. gepaart wurde.

Bild 71 bis Bild 73 stellen einige sich ergebende optische Flüsse bei Rotationen dar.

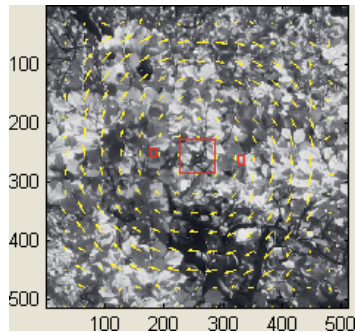


Bild 71: Optischer Fluss bei der Rotation der Abbildung eines Laubbaumes um 4°

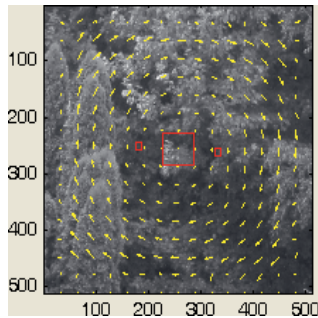


Bild 72: Optischer Fluss bei der Rotation der Abbildung einer Baumgruppe um 4°

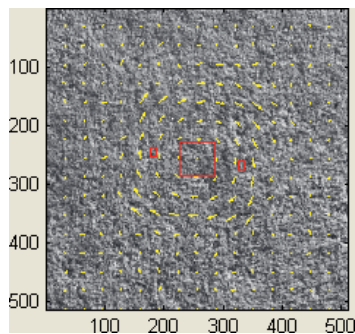


Bild 73: Optischer Fluss bei der Rotation der Abbildung einer Wiese um 8°

Bei der Berechnung des Winkels nach der Methode von Lucas&Kanade, durchgezogene Linie mit der Bezeichnung L&K in Bild 74, Bild 76 und Bild 78, ist der Fehler bis 8° unter 5 %. Ab 9° Rotation werden die Abweichungen erheblich. Dies wird am Ende des Kapitels erläutert.

Die Berechnung mit der NCC zeigt größere Abweichungen und erreicht 5 % Fehler schon bei 3° , bleibt allerdings bis 9° unter 10 %. Ab 10° weichen die berechneten Winkel erheblich vom erwarteten Wert ab.

Bei der Position wird keine Verschiebung erwartet, da die Bilder (scheinbar) exakt um das Zentrum rotiert wurden. Dass dies nicht stimmt, wird später mit Bild 80 genauer untersucht.

Bei der Berechnung nach Lucas und Kanade, Bild 75, Bild 77 und Bild 79, ist die Verschiebung immer kleiner als wenige Hundertstel Pixel und somit vernachlässigbar.

Die NCC ist bezüglich der ermittelten Position sehr empfindlich auf Rotationen und verlässt ab 3° den Subpixelbereich, s. Bild 77.

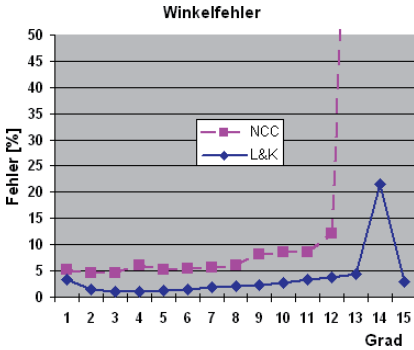


Bild 74: Maximaler prozentualer Winkelfehler bei der Rotation von Bild 71

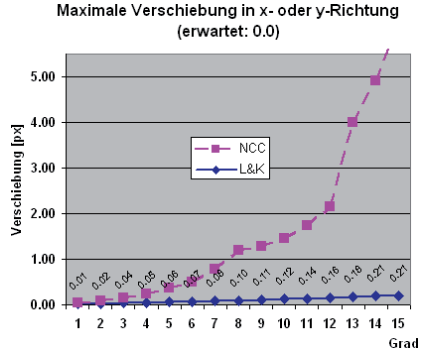


Bild 75: Maximaler Positionsfehler in x- oder y-Richtung bei der Rotation von Bild 71

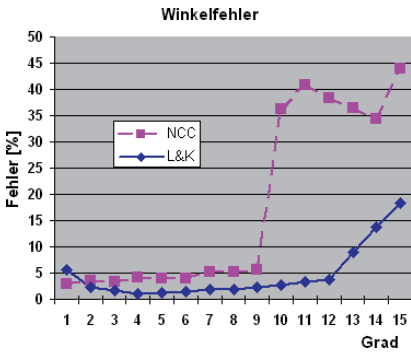


Bild 76: Maximaler prozentualer Winkelfehler bei der Rotation von Bild 72

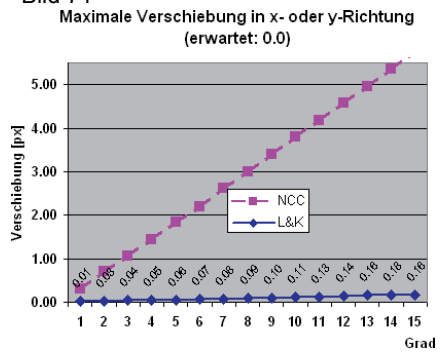


Bild 77: Maximaler Positionsfehler in x- oder y-Richtung bei der Rotation von Bild 72

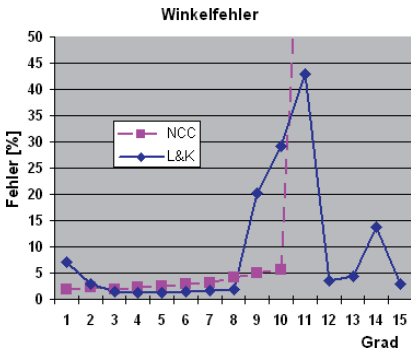


Bild 78: Maximaler prozentualer Winkelfehler bei der Rotation von Bild 73

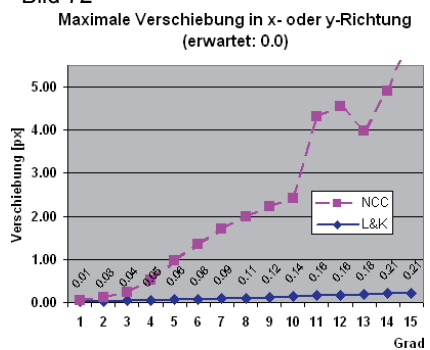


Bild 79: Maximaler Positionsfehler in x- oder y-Richtung bei der Rotation von Bild 73

Zu beachten ist, dass jeweils die größten Abweichungen aus 100 Bildpaaren in den Grafen dargestellt sind. Die Mittelwerte sind allerdings meist nicht viel kleiner. Beispielhaft bei drei Grad Rotation werden in Tabelle 5 die Maximal-, Minimal- und Mittelwerte der Winkel und der Verschiebung in x-Richtung für die Berechnung nach Lucas und Kanade gezeigt. Die Verschiebung in y-Richtung ist erwartungsgemäß sehr ähnlich und wird nicht dargestellt. Der Winkelfehler schwankt um 1 %. Die Verschiebung ist im Subpixelbereich und damit kaum relevant.

Tabelle 5: Berechnete Winkel und x-Positionen für 3° Rotation bei der Berechnung nach Lucas und Kanade

Berechnete Werte ↓	Laubbaum von Bild 71	Bäume von Bild 72	Wiese von Bild 73
Max. Winkel	3.030°	3.049°	3.044°
Min. Winkel	2.996°	2.991°	3.007°
Mittelwert Winkel	3.014°	3.015°	3.021°
Max. x	0.031 px	0.039 px	0.034 px
Min. x	0.022 px	0.019 px	0.025 px
Mittelwert x	0.026 px	0.029 px	0.028 px

Schlechter sind die Verhältnisse bei der Berechnung mit der NCC, denn dort sind die Winkelfehler noch akzeptabel, aber die ermittelte Position verlässt bei Bild 72 den Subpixelbereich, s. Tabelle 6.

Tabelle 6: Winkel und x-Position für 3° Rotation bei der Berechnung mit NCC

Berechnete Werte ↓	Laubbaum von Bild 71	Bäume von Bild 72	Wiese von Bild 73
Max. Winkel	3.105°	3.097°	3.047°
Min. Winkel	2.860°	2.954°	2.944°
Mittelwert Winkel	2.997°	3.032°	2.997°
Max. x	0.084 px	0.234 px	-0.004 px
Min. x	-0.123 px	-1.020 px	-0.161 px
Mittelwert x	-0.067 px	-0.534 px	-0.104 px

Auffällig ist, dass die Verschiebungen nicht um den Nullpunkt variieren, sondern stetig im positiven Bereich mit dem Winkel anwachsen, s. Bild 79 L&K-Linie, was auf einen Fehler bei der Rotation der Bildvorlage deutet, wie auch durch Bild 80 bestärkt wird. Hier wird ein Bild, Bild 71, sukzessiv 100-mal um 5° im Uhrzeigersinn und 100-mal im Gegenuhrzeigersinn gedreht. Bei der Berechnung nach Lucas & Kanade ist das Zentrum der Position bei $x,y = 0.04, 0.05$ bzw. symmetrisch hierzu bei $x,y = -0.05, -0.04$. Erwartet wird $x,y = 0,0$. Hieraus ergibt sich, dass die Rotation der Bildvorlage durch Matlab® nicht genau im Zentrum des Bildes erfolgt, sondern drehrichtungsabhängig etwas versetzt ist.

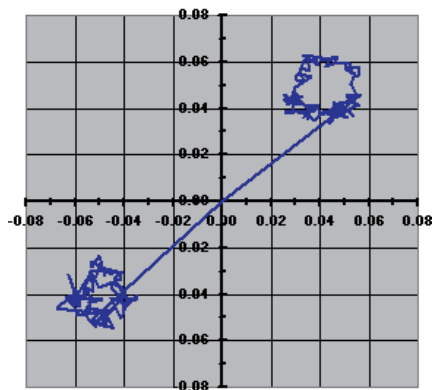


Bild 80: x- und y-Positionen in px bei multiplen Rotationen um 5° im Uhrzeigersinn (rechts im Bild) und Gegenuhrzeigersinn (links im Bild), Berechnung nach Lukas&Kanade

Bei steigendem Winkel werden die optischen Flussvektoren in den vom Rotationszentrum entfernten Bereichen so groß, dass sie mit der Gradientenmethode nicht mehr berechnet werden können. In Bild 81 ist der optische Fluss bei 4° Rotation sehr gleichmäßig bis in die äußeren Bildbereiche, während bei 14° Rotation, wie in Bild 82 gezeigt, der optische Fluss vom Zentrum ausgehend immer ungleichmäßiger wird. Dies ist begründet durch den optischen Fluss, der im Randbereich zu groß wird, um erfasst zu werden.

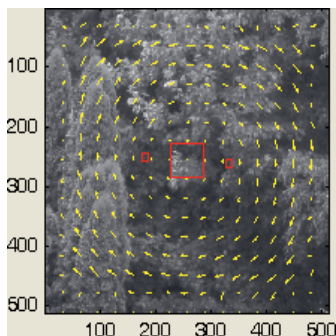


Bild 81: 4° Rotation: Gleichmäßiger optischer Fluss bis in die Randbereiche

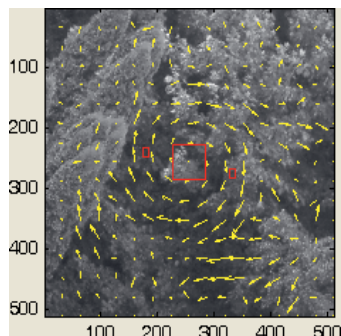


Bild 82: 14° Rotation: Zunehmend chaotischer optischer Fluss vom Zentrum aus

Dass die Position bei der NCC zunehmend mit dem Winkel falsch berechnet wird, liegt an verschiedenen Bildinhalten zwischen dem gedrehten und ungedrehten Bild. So haben bei 5° Rotation in einem 32 x 32 px Fenster die korrespondierenden Pixel am äußeren Rand einen Abstand von mehr als 3 px und tragen damit zur Korrelation kaum bei. Dies zeigt sich auch an abnehmenden Korrelationswerten bei steigenden Winkeln, s. Bild 83.

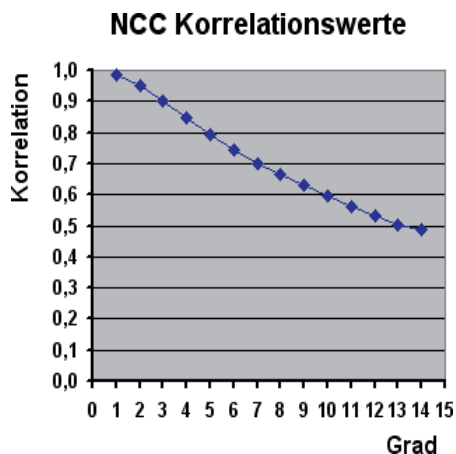


Bild 83: Durchschnittliche Korrelationswerte bei Rotationen zwischen 1° und 14°

Zusammenfassung

Das Resümee ist, dass die NCC ab 2° mit zunehmenden Fehlern auf Rotationen reagiert, während die Berechnung nach Lucas und Kanade bis 9° Positionsfehler im Subpixelbereich liefert. Sollen bei großen Verschiebungen und großen Rotationen genaue Positionen ermittelt werden, so kann hierzu die Kombination beider Verfahren genommen werden, indem zunächst mit der NCC die Position grob bestimmt wird und dann die gewonnenen Bildausschnitte mit der Gradientenmethode nachberechnet werden.

4.4 Berechnungszeit

Die Berechnungszeit ist in Matlab[®] für die Gradientenmethode dreimal höher als für die NCC.

4.5 Subpixelgenauigkeit

Um eine genauere Positionsbestimmung zu erhalten, können noch Bruchteile der Pixelverschiebungen – Subpixel – herangezogen werden. Die Subpixelgenauigkeit ist hier relevant, da dieselbe Kamera für die Translations- und Höhenbestimmung genommen werden soll, denn die Translation bei hohen Geschwindigkeiten und geringen Abständen erfordert eine Abbildung, die wenige Pixel pro abgebildete Wegstrecke beinhaltet, also eine kleine Brennweite, während die Höhenbestimmung bei großen Abständen einige Pixel Disparität für ausreichende Genauigkeit erfordert, also eine große Brennweite oder große Basisbreite der Stereokamera. Bei einer großen Basisbreite der Stereokamera ist das Gewicht zu groß für ein kleines Fluggerät und auch zu instabil, was eine ständige Neukalibrierung erfordert. Damit nur eine Brennweite bei akzeptabler Basisbreite benutzt werden kann, muss die Genauigkeit über Subpixelberechnung erhöht werden.

Welche Subpixelgenauigkeit erwartet werden kann, wird in den folgenden Abschnitten ermittelt.

4.5.1 Subpixelgenauigkeit beim differentiellen Verfahren nach Lucas&Kanade

Zur Untersuchung der Subpixelgenauigkeit wird zunächst ein Ausschnitt T_1 aus einem Bild I_1 in einem Bild I_2 mit Hilfe der NCC gesucht, s. Bild 84 bis Bild 87. Das Teilbild T_2 aus dem Bild I_2 hat die größte Übereinstimmung mit dem Teilbild T_1 aus dem nachfolgenden Bild I_2 .

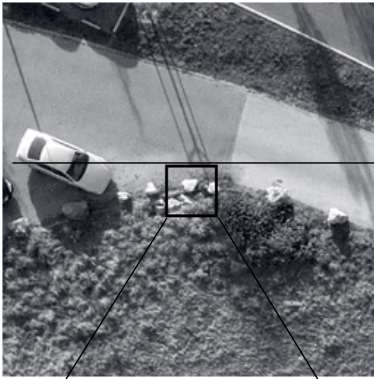


Bild 84: Bild I1 mit gesuchtem Teilbild T1

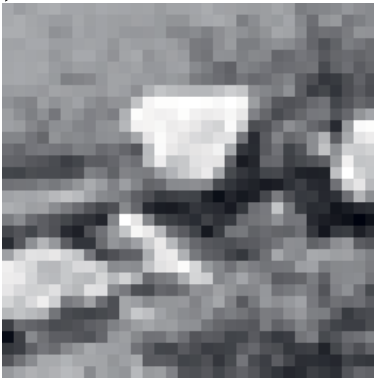


Bild 86: Teilbild T1 eines Bildes I1

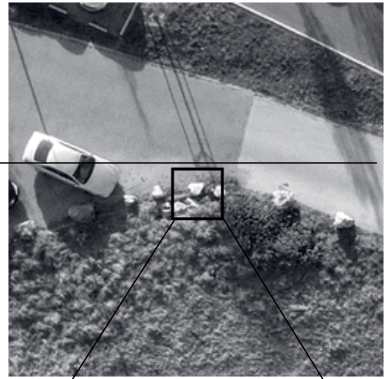


Bild 85: Nachfolgendes Bild I2 mit gefundenem Teilbild T2

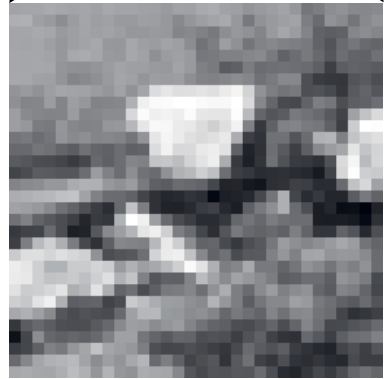


Bild 87: Teilbild T2 des nachfolgenden Bildes I2 mit der größten NCC

Dass die Übereinstimmung nicht perfekt ist, zeigt die Differenzbildung der Teilbilder T1 und T2 (Ableitung I_d), s. Bild 88. Die Darstellung ist dreifach überhöht, wobei schwarz dem Wert 255/3 entspricht und weiß der Null. Die größte Differenz zweier Pixelwerte der ungefilterten Bilder ist hier 72.

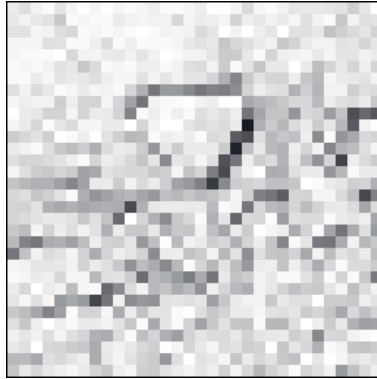


Bild 88: Absolute Differenz der Teilbilder T1 und T2 , (dreifach überhört), Weiß zeigt die kleinsten Differenzen

Dann werden mit dem differentiellen Verfahren die Flusswerte gebildet. Obwohl sich die Kamera nur geringfügig zur stillstehenden Szene bewegt hat und sich daher annähernd gleiche Flusswerte für benachbarte Pixel ergeben müssten, sind die Flusswerte der zweiten Aufnahme in Bezug zur ersten Aufnahme bei kleinem Ableitungsfenster, hier 5 x 5 Pixel groß, wertemäßig breit gestreut, s. Bild 89.

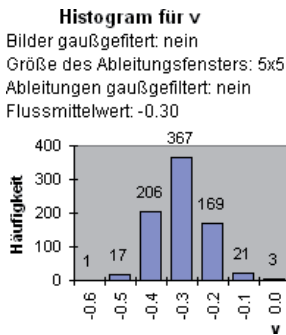


Bild 89: Histogramm für Flusswerte, 5 x 5 Ableitungsfenster

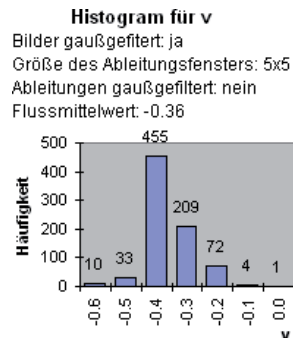


Bild 90: Histogramm für Flusswerte, 5 x 5 Ableitungsfenster, mit Tiefpassfilterung

Die Diagramme, Bild 89 und Bild 90, zeigen die Häufigkeit der Flusswerte v in zehntel Pixelschritten und sind ein Maß für die Unsicherheit der Berechnungsergebnisse.

Im Diagramm von Bild 89 sind die Häufigkeiten breit verteilt zwischen $v = 0.4$ und $v = 0.2$.

Die spatiale Verteilung des Flussfeldes ist nicht glatt, d. h., kleine und große Flusswerte sind über das Bild verteilt, s. Bild 91.

Eine vorhergehende Tiefpassfilterung der Bilder führt zu einer weiteren Verbesserung, denn mehr Flusswerte liegen in der Nähe des Mittelwertes, s. Bild 90. Folgerichtig weisen auch die Ableitungen I_x , I_y , I_t geringere Schwankungen auf.

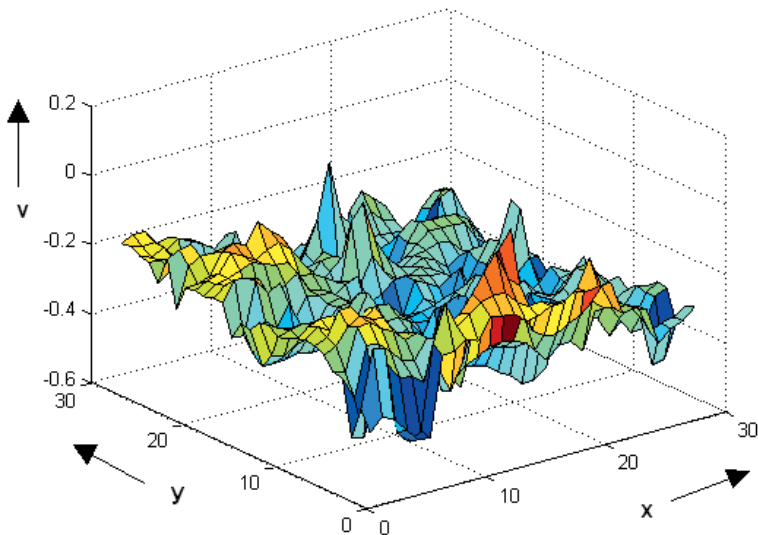


Bild 91: Spatiale Verteilung der Flusswerte v , Ableitungsfenster 5×5 , Bilder und Ableitungen sind nicht tiefpassgefiltert

Nimmt man größere Ableitungsfenster, also geht von der Konstanz der Ableitungen in einem größeren Gebiet aus, was hier der Fall sein sollte, da sich die gesamte Szene gleichförmig in Bezug auf die Kamera bewegt, wird die Streuung der Flusswerte wesentlich geringer, s. Bild 92. Werden das Bild oder die Ableitungen gefiltert, so ist die Streuung noch etwas geringer, s. Bild 93.

Histogramm für v
 Bilder gaußgefiltert: nein
 Größe des Ableitungsfensters: 17x17
 Ableitungen gaußgefiltert: nein
 Flussmittelwert: -0.33

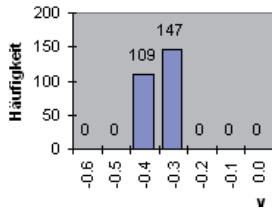


Bild 92: Histogramm für Flusswerte, 17 x 17 Ableitungsfenster, Bilder ungefiltert.

Histogramm für v
 Bilder gaußgefiltert: ja
 Größe des Ableitungsfensters: 17x17
 Ableitungen gaußgefiltert: nein
 Flussmittelwert: -0.37

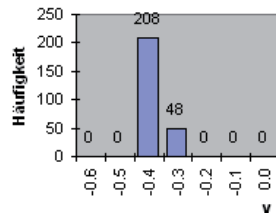


Bild 93: Histogramm für Flusswerte, 17 x 17 Ableitungsfenster, Bilder gaußgefiltert

In Bild 94 ist die Glattheit der Flusswerte zu erkennen. Der absolut kleinste Flusswert -0.28 Pixel ist in Bild 94 an der Stelle $x, y = 1, 16$ und damit links unten vom Zentrum von Bild 86 positioniert. 95 % der Flusswerte liegen im Fenster von -0.30 bis -0.40. Alle Werte liegen im Bereich -0.28 bis -0.40. Die erhöhten Werte beim Maximum ($x, y = 0, 20$, links oben in Bild 94) deuten auf eine Rotation hin, die tatsächlich stattfand, da bei einer Rotation der optische Fluss mit der Entfernung zum Rotationszentrum steigt.

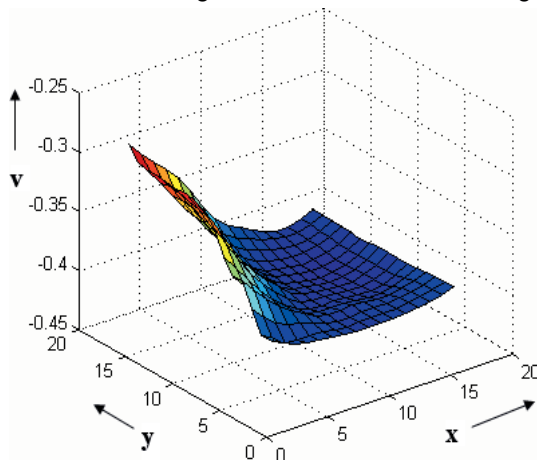


Bild 94: Spatiale Verteilung der Flusswerte v, Ableitungsfenster 17 x 17, Bilder sind tiefpassgefiltert

Kleine Ableitungsfenster erlauben eine größere Auflösung der Flusswerte für einzelne Pixel, sind aber dann ungenauer.

Die erreichbare Genauigkeit bei realen Bildern ist etwa 0.06 Pixel, s. Kapitel 4.5.3. Dies wird auch bestätigt durch eine Arbeit von J. Klapstein [86], der etwa 0.1 Pixel angibt. Ein anderer Artikel [54] erwähnt, ohne weitere Untersuchungen, Genauigkeiten bis zu 1/256 Pixel, allerdings abhängig von Bildqualität, Bildinhalt und dem Berechnungsaufwand.

4.5.2 Subpixelgenauigkeit bei der NCC

Die zweidimensionale Kreuzkorrelation ergibt zunächst nur ganzzahlige Pixelpositionen für die beste Übereinstimmung zweier Bildgebiete an. Im Folgenden werden Ansätze zur Erlangung einer Subpixelgenauigkeit beschrieben und bewertet. Hierzu wird der Korrelationsverlauf in der Umgebung der besten Übereinstimmung analysiert in Bezug auf:

- Größe des Korrelationsfensters
- Bildinhalt
- Helligkeitsmittelwert

Mit diesen Erkenntnissen werden verschiedene Approximationsfunktionen auf ihre Übereinstimmung zum Korrelationsverlauf untersucht. Aus diesen Approximationsfunktionen werden die Subpixelwerte ermittelt.

Um die Subpixelgenauigkeit zu verifizieren, werden aus einem sehr großen Bild, s. Bild 95, 500 jeweils um ein Pixel nach rechts verschobene 4096 x 4096 Pixel große Teilbilder entnommen und auf ein Zweiunddreißigstel verkleinert. Diese verkleinerten Teilbilder – in den Abmaßen 128 x 128 Pixel – sind dadurch jeweils exakt um ein 32tel Pixel zum benachbarten Teilbild verschoben. Mit weiteren Abbildungen, s. Bild 96 und Bild 97, wurde auch so verfahren, um den Einfluss von Bildinhalten auf die Berechnungen auszuschließen. Bei der Bildverkleinerung wird angenommen, dass der Einfluss auf das Korrelationsverhalten gering ist.



Bild 95: Bild, um genaue Verschiebungen zu ermitteln. 15515 x 5917 Pixel
 Quelle: <http://jleporcq.free.fr/BaladesUnes/FumayPlatale/Images/Grandes/FumayPlatale016.jpg>



Bild 96: 2. Bild, um genaue Verschiebungen zu ermitteln. 10000 x 7071 Pixel
 Quelle: http://www.stadt-zuerich.ch/hbd/de/index/hochbau/ausgeschriebene_ww/luggwegstrasse.html



Bild 97: 3. Bild, um genaue Verschiebungen zu ermitteln. 12577 x 8385 Pixel
 Quelle: http://farm9.staticflickr.com/8296/7888915552_30c3b2eedf_o_d.jpg

Ein einfacher Ansatz ist die Interpolation der Korrelationswerte, die nur an ganzzahligen Pixelpositionen bekannt sind, mit einer Parabelfunktion durchzuführen, um damit die Lage des tatsächlichen Maximums besser zu bestimmen [87][88][89][90]. Hierbei wird davon ausgegangen, dass die Korrelationswerte um ihr Maximum normalverteilt und symmetrisch sind. Bild 98 und Bild 99 geben dem groben Anschein nach für die Kreuzkorrelation eines Bildes mit einem Ausschnitt aus demselben Bild eine Gaußkurve wieder, wobei hier nur die Verschiebung in einer Richtung betrachtet wird. Die Breite der glockenähnlichen Kurve ist abhängig von der Größe des Korrelationsfensters, s. Bild 98, und vom Bildinhalt, s. Bild 99.

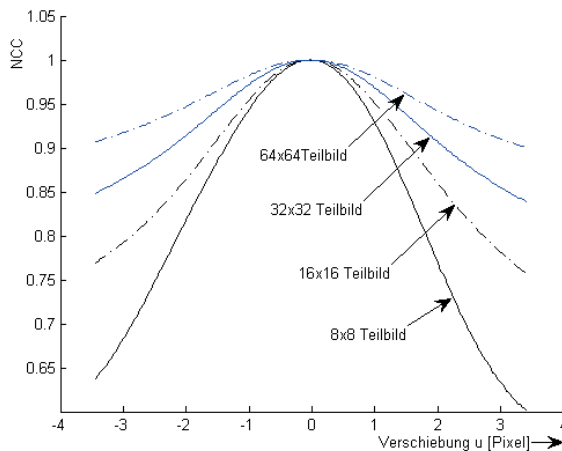


Bild 98: Verlauf der Kreuzkorrelation bei verschiedenen Größen des Teilbildes

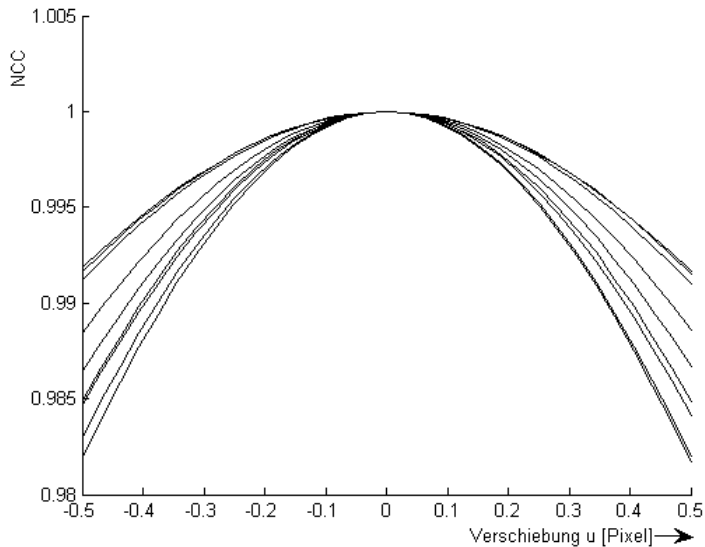


Bild 99: Verlauf der Kreuzkorrelation bei verschiedenen Bildinhalten und konstanter Teilbildgröße

In Bild 100 ist die dreidimensionale Darstellung der Quasi-Autokorrelation in x- und y-Richtung eines Bildausschnittes mit dem Bild, aus dem dieser Ausschnitt entnommen wurde, gezeigt. Deshalb ist hier die Korrelation bei Übereinstimmung sehr dominant und erreicht den größtmöglichen Maximalwert 1.0. Diese Quasi-Autokorrelation ist keine exakte Autokorrelation, weil ein Ausschnitt eines Bildes mit einem Gesamtbild korreliert wird und damit beide Funktionen nicht identisch sind.

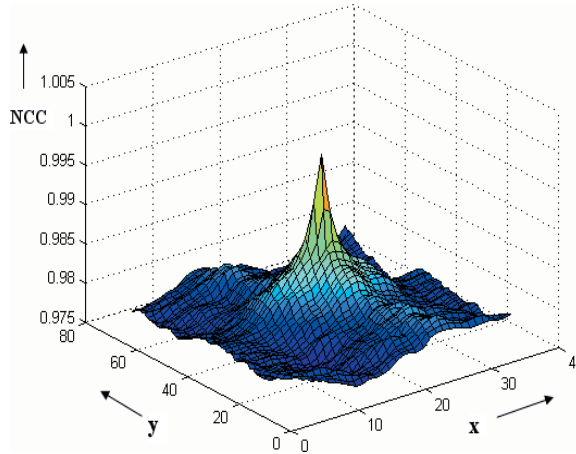


Bild 100: Grafische Darstellung der Korrelationswerte bei einer Korrelation eines Bildausschnittes aus demselben Bild und dem Bild selbst

4.5.2.1 Verbesserung durch Abzug des Mittelwertes

Bei der ZNCC wird für die Berechnung der Korrelation bei jeder Pixelumgebung der Mittelwert neu berechnet. Dies hat den Vorteil, dass die schlecht korrelierten Gebiete deutlich niedrigere Korrelationswerte liefern als es bei der NCC wäre. Das heißt, das Korrelationsmaximum ist bei der ZNCC wesentlich dominanter als bei der NCC. Wenn nur das Maximum gesucht wird, ist dies nicht relevant. Wird Subpixelgenauigkeit gefordert, ist ein dominantes Maximum der Korrelation nützlich. Um nicht aufwendig immer wieder den Mittelwert zu bilden, wird nur der Mittelwert des Teilbildes genommen und sowohl vom Teilbild als auch von dem zu durchsuchenden Gesamtbild abgezogen. Damit wird unter der Annahme, dass das Teilbild und das zu durchsuchende Bild *annähernd* den gleichen Mittelwert haben, erreicht, dass der Mittelwert für beide Bilder in der Umgebung des Maximums nahe bei null liegt.

Es wird hier nur der eindimensionale Fall, eine Bildzeile G , der sich leicht auf den zweidimensionalen Fall übertragen lässt, erläutert. Beim Hinzuzugieren eines Mittelwertes m zur Bildzeile G und eines Ausschnittes T nimmt das Korrelationsintegral folgende Form an:

$$C(\tau) = \int_{-F/2}^{F/2} [T(t) + m][G(t + \tau) + m] dt \quad (4.1)$$

C ist der Korrelationswert an der Stelle τ , G ist das zu durchsuchende Gesamtbild, T ist das Teilbild aus G, t die Ortsvariable, m der Mittelwert.

Da das Teilbild T eine limitierte Breite von F hat, sind damit die Integrationsgrenzen mit dem um τ versetzten Fenster mit $\pm \frac{F}{2}$ festgesetzt.

Das obige Integral (4.1) nimmt ausmultipliziert folgende Form an:

$$C(\tau) = \underbrace{\int_{-F/2}^{F/2} T(t)G(t+\tau)dt}_{\text{Term1}} + \underbrace{\int_{-F/2}^{F/2} mT(t)dt}_{\text{Term2}} + \underbrace{\int_{-F/2}^{F/2} mG(t)dt}_{\text{Term3}} + \underbrace{\int_{-F/2}^{F/2} m^2dt}_{\text{Term4}} \quad (4.2)$$

Im mittelwertfreien Fall, mit $m = 0$, liegt das klassische Kreuzkorrelationsintegral vor, s. Term1 der Gleichung (4.2).

Mit der groben Annahme, dass eine Aufnahme mit strukturiertem Inhalt farbigem Rauschen entspricht, dies wird in Kapitel 4.5.2.7 erhärtet, kann das Korrelationsintegral C für einen Mittelwertabzug m betrachtet werden.

Das unbegrenzte Integral über Rauschen ist null. Angenähert gilt dies auch für ein begrenztes Integral. Da auch die Multiplikation von nicht korrelierten Rauschsignalen wieder ein Rauschsignal ergibt, ist der Term 1 immer angenähert null, außer bei $\tau = 0$, da dann das Teilbild T und der Ausschnitt aus dem Gesamtbild G übereinstimmen. Die Terme 2 und 3 sind immer annähernd null. Der Term 4 liefert unabhängig von τ einen Wert, der quadratisch mit dem Mittelwert m wächst, während der Term1 im Korrelationsmaximum unverändert bleibt. Das heißt, die Dominanz des maximalen Korrelationswertes nimmt mit zunehmendem Mittelwert m ab.

Bild 101a-c verdeutlichen die Verhältnisse der Gleichungen (4.1) und (4.2) bei zusätzlicher Normierung. Bild 101a ist ein Rauschsignal mit Amplituden zwischen 0 und 255, wie es bei üblichen Digitalkameras der Fall ist. Der Mittelwert m ist 127.5, was bei ausgeglichener Ausleuchtung einer Szene gegeben ist. Dieses Rauschsignal, das eine Bildzeile darstellen soll, wird mit einem 256 px breiten Ausschnitt, Bild 101b, aus der Mitte des Rauschsignals korreliert und liefert eine Korrelationsfunktion, die in Bild 101c dargestellt ist.

Der Zahlenwert von Term4 aus Gleichung (4.2) errechnet sich bei einer Fensterbreite F von 256 px und dem Mittelwert von 127.5 zu:

$$\text{Term4} = m^2 \cdot F = 127.5^2 \cdot 256 = 4.16 \cdot 10^6 \quad (4.3)$$

Während das eigentliche Korrelationsintegral, Term1, bei einem Rauschen mit einer Standardabweichung von z. B. $\sigma = 36$ sich errechnet zu [91]

$$Term1 = \sigma^2 \cdot F = 36^2 \cdot 256 = 0.32 \cdot 10^6 \quad (4.4)$$

hat C abseits des Korrelationsmaximums den Wert des Terms 4 und im Korrelationsmaximum den Wert von Term1 plus Term4. Damit ergibt sich das Verhältnis vom maximalen Korrelationswert zu abseitigen Korrelationswerten zu:

$$\frac{Term1 + Term4}{Term4} = 1.08 \quad (4.5)$$

Das heißt, dass der maximale Korrelationswert bei diesem Rauschsignal mit einer Standardabweichung von 36 nur etwa 8 % höher ist als die umgebenden Korrelationswerte.

Bei den gleichen Signalen, aber mittelwertfrei, s. Bild 102a-c, ragt das Korrelationsmaximum dominant aus den umgebenden Korrelationswerten heraus. Bei unbegrenzten Rauschsignalen wäre die Korrelation exakt null außerhalb des Korrelationsmaximums. Bei ortsbegrenzten Rauschsignalen schwankt der Term1 geringfügig um die Nulllinie.

Mit Rauschsignalen ist die Korrelationsgleichung (4.1) für mittelwertbehaftete Signale erläutert worden. Bei Bildern, hier aus einer Bildzeile (Bild 95) entnommen, sind die Signale nicht deterministisch, allerdings wird mit Bild 103a-c und Bild 104a-c gezeigt, dass die Verhältnisse den Rauschsignalen ähnlich sind. Auch hier ragt das Korrelationsmaximum beim mittelwertbehafteten Signal kaum aus der Umgebung heraus, s. Bild 103c, während das mittelwertfreie Bild ein sehr dominantes Korrelationsmaximum liefert, Bild 104c.

Erwartungsgemäß haben die Bereiche in Bild 104c rechts und links vom Maximum eine größere Amplitude als bei Rauschsignalen, da Bildinhalte sich ähneln können und damit nicht mehr vollkommen zufällig sind.

Da die Signale (Bildausschnitte) ortsbegrenzt sind und auch kein reines Rauschen darstellen, wird die Umgebung des Maximums bei größerem Mittelwert immer flacher, s. Bild 105.

Rauschen mit Gleichanteil

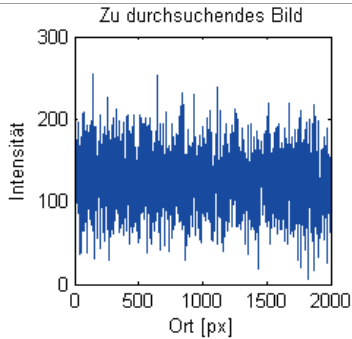


Bild 101a: Mittelwertbehaftetes Rauschen

Rauschen ohne Gleichanteil

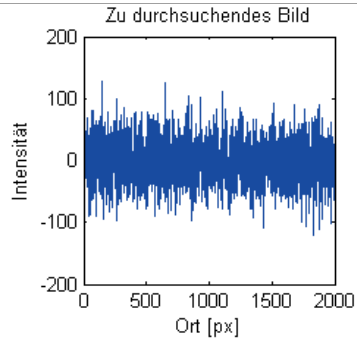


Bild 102a: Mittelwertfreies Rauschen

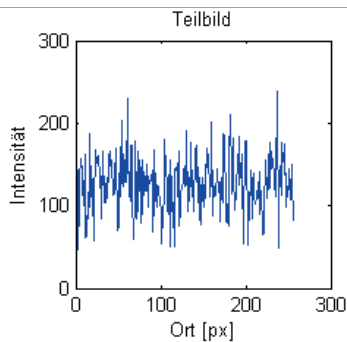


Bild 101b: Teil aus Bild 101a
symmetrisch um 1000 px

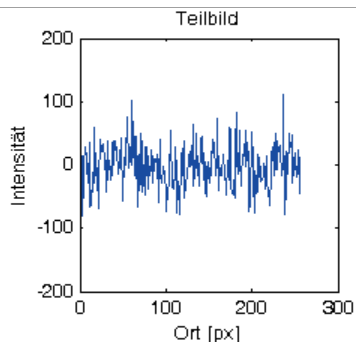


Bild 102b: Teil aus Bild 102a
symmetrisch um 1000 px

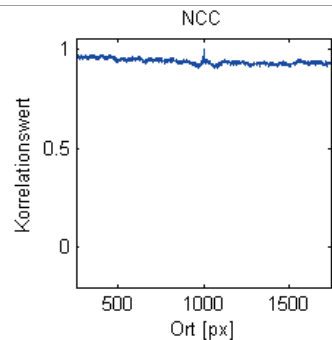


Bild 101c: Korrelationsergebnis der
mittelwertbehafteten Rauschsignale

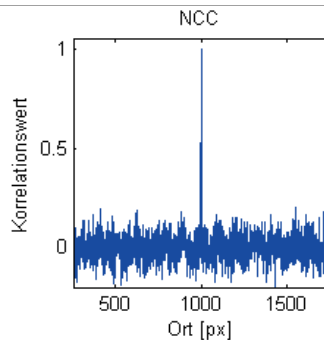
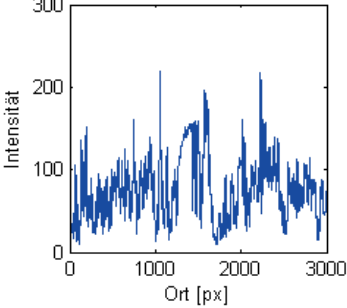
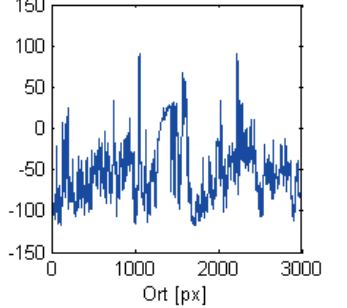
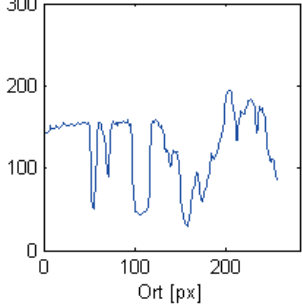
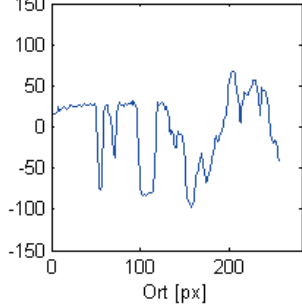
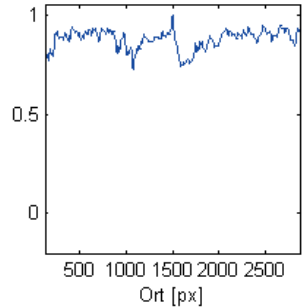
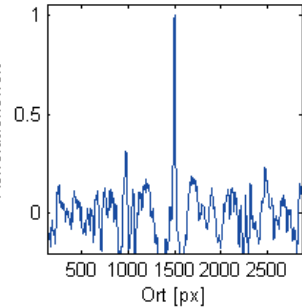


Bild 102c: Korrelationsergebnis der
mittelwertfreien Rauschsignale

<p style="text-align: center;">Bildzeile Intensität von 0 bis 255</p> <p style="text-align: center;">Zu durchsuchendes Bild</p>  <p style="text-align: center;">Bild 103a: Mittelwert<u>behaftete</u> Bildzeile</p>	<p style="text-align: center;">Bildzeile mittelwertfrei</p> <p style="text-align: center;">Zu durchsuchendes Bild</p>  <p style="text-align: center;">Bild 104a: Mittelwert<u>freie</u> Bildzeile</p>
<p style="text-align: center;">Teilbild</p>  <p style="text-align: center;">Bild 103b: Teil aus Bild 103a symmetrisch um 1500 px</p>	<p style="text-align: center;">Teilbild</p>  <p style="text-align: center;">Bild 104b: Teil aus Bild 104a symmetrisch um 1500 px</p>
<p style="text-align: center;">NCC</p>  <p style="text-align: center;">Bild 103c: Korrelationsergebnis der mittelwertbehafteten Bildzeilen</p>	<p style="text-align: center;">NCC</p>  <p style="text-align: center;">Bild 104c: Korrelationsergebnis der mittelwertfreien Bildzeilen</p>

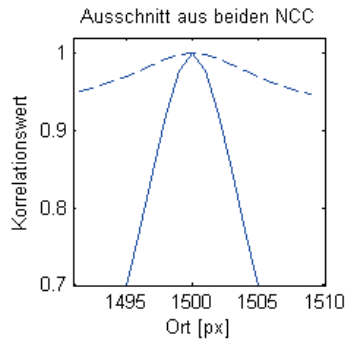


Bild 105: Gestrichelte Linie: Ausschnitt aus Bild 103c, (mit Mittelwert) um das Korrelationsmaximum, durchgezogene Linie: Ausschnitt aus Bild 104c (ohne Mittelwert) um das Korrelationsmaximum

Wie stark der Kurvenverlauf vom Mittelwert abhängt, wird in folgendem Bild (Bild 106) deutlich. Hier ist eine Bildzeile aus dem hellen Wolkenbereich der Bild 95 berechnet. Der Mittelwert der Intensität ist mit 218 sehr hoch. Während die direkt benachbarten Korrelationswerte links und rechts vom Maximum (bei 1499 px und 1501 px), Bild 106, bei der mittelwertfreien Korrelation immerhin um 0.02 niedriger sind, sind bei der mittelwertbehafteten Korrelation die benachbarten Werte nur um $4 \cdot 10^{-6}$ niedriger.

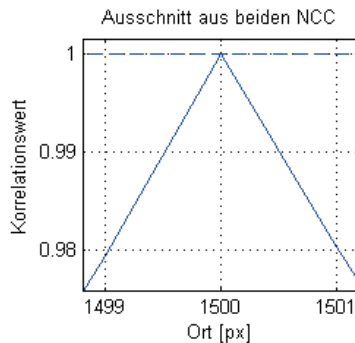


Bild 106: Korrelationsverlauf bei kontrastarmer Bildzeile, gestrichelt: Bildzeile mit Mittelwert, durchgezogen: Bildzeile ohne Mittelwert

Bei der oberen Linie, fast eine Gerade, führen schon sehr kleine Abweichungen in der Umgebung des Korrelationsmaximums zu falschen Interpolationen, die insbesondere dann entstehen, wenn die Bilder zwar dieselbe Szene abbilden, aber aus unterschiedlichen Blickwinkeln

aufgenommen werden und damit leicht unterschiedliche Bildinhalte haben oder wenn mit einfacher Genauigkeit (6 signifikante Stellen) gerechnet wird. Der Mittelwert hat also unerwünschten Einfluss auf die Interpolationen zur Subpixelberechnung, die in den nächsten Kapiteln erläutert werden.

4.5.2.2 Interpolation über Parabel

Wie schon erwähnt, ist die Subpixelgenauigkeit für die Positionsermittlung der Kamera erforderlich, weil die Kameradaten bei gleichen Objektiven für die Translations- und Höhenbestimmung ein Kompromiss zwischen erfassbarer Translation und Genauigkeit der Höhenbestimmung sind.

Im Folgenden werden Methoden zur Subpixelbestimmung erläutert.

Da der Verlauf der Korrelationskurve vom Bildinhalt abhängt, kann keine genaue Funktion dafür angegeben werden. Eine markante Eigenschaft des Korrelationsverlaufs ist die weitgehende Symmetrie in der Nähe des Maximum, die je nach Bildinhalt variiert. Deshalb sind symmetrische Funktionen zur Approximation besonders geeignet.

Die Parabel ist eine symmetrische Funktion. Mit einer gewissen Genauigkeit (diese wird in Kap. 4.5.3 quantifiziert) gilt also die Interpolation über eine Parabelkurve der Form:

$$y = a \cdot x^2 + b \cdot x + c \quad (4.6)$$

Die Position u_{\max} , die den angenäherten größten NCC-Wert und damit die Verschiebung in Subpixelgenauigkeit angibt, errechnet sich aus den NCC-Werten y_1 , y_2 , y_3 – in Bild 107 gezeigt – mit der Scheitelwertsgleichung der Parabel zu:

$$u_{\max} = u_{NCC \max} + \frac{y_1 - y_3}{2y_1 - 4y_2 + 2y_3}$$

mit:

$$\begin{aligned} y_1 &= NCC(u_{NCC \max} - 1, v_{NCC \max}) \\ y_2 &= NCC_{\max} = NCC(u_{NCC \max}, v_{NCC \max}) \\ y_3 &= NCC(u_{NCC \max} + 1, v_{NCC \max}) \end{aligned} \quad (4.7)$$

Für die v-Richtung gilt Entsprechendes.

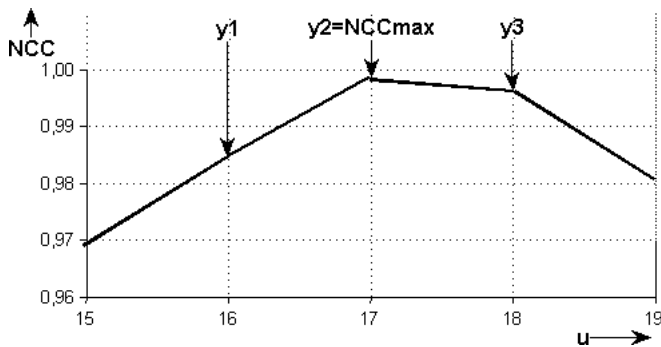


Bild 107: Typischer Verlauf einer Korrelation mit den für die Parabelfunktion relevanten NCC-Werten y_1 , y_2 , und y_3

Beispiel:

Zwei um genau 0.3750 Pixel versetzte Bildausschnitte werden miteinander korreliert. Das Ergebnis ist ein ausgeprägtes Maximum bei $u, v = 17, 17$, s. Bild 108.

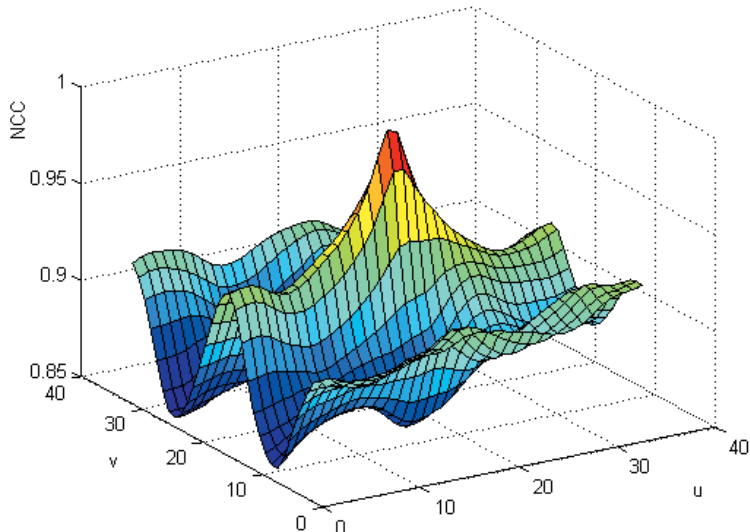


Bild 108: Korrelation eines Bildes mit einem um 0.3750 Pixel versetzten Bildausschnitt aus demselben Bild.

Die Werte der NCC um dieses Maximum (0.9986) bei $u, v = 17,17$ sind in Tabelle 7 gezeigt.

Tabelle 7: Werte der NCC in der Umgebung des Maximums

v	0,9260	0,9340	0,9412	0,9432	0,9420
\downarrow	0,9523	0,9657	0,9782	0,9790	0,9703
	0,9693	0,9849	0,9986	0,9961	0,9803
	0,9635	0,9730	0,9799	0,9749	0,9604
	0,9421	0,9430	0,9427	0,9383	0,9298
$u \rightarrow$					

Mit dem maximalen NCC-Wert (in Tabelle 7 Mitte) und seinen zwei Nachbarn kann die Ausgleichsparabel berechnet werden. In Bild 109 ist die Ausgleichsparabel eingezeichnet.

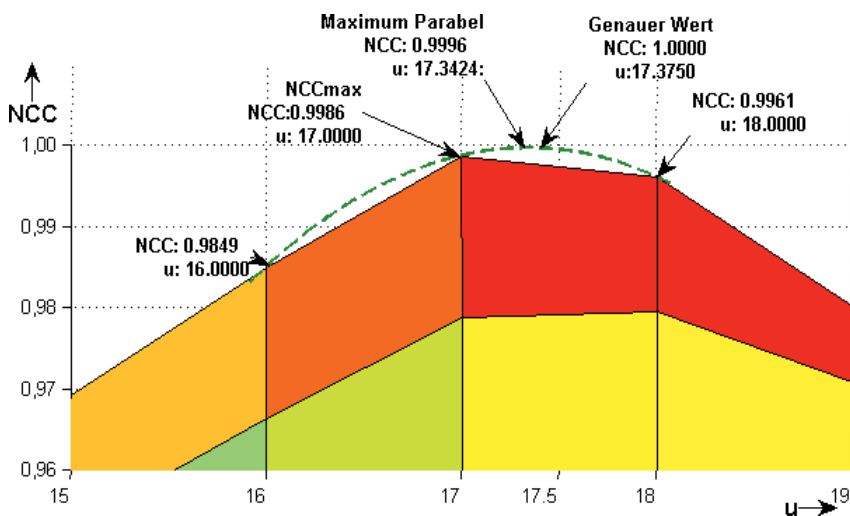


Bild 109: NCC-Werte in Bezug auf die Verschiebung u mit Ausgleichsparabel. Ausschnitt aus Bild 108

Aus den drei NCC-Werten bei $u = 16, 17, 18$ wurde die Ausgleichsparabel berechnet, s. Bild 109. Dadurch ergibt sich ein Parabelmaximum bei $u = 17.3424$. Das wirkliche Maximum, der genaue Ort der Verschiebung, liegt allerdings bei $u = 17.3750$. Die Annäherung über eine Parabel ist recht nah an

der Wirklichkeit und trotzdem weiter verbesserungsfähig. Deshalb sind noch weitere Ansätze untersucht worden.

4.5.2.3 Interpolation über Gaußkurve

Etwas bessere Ergebnisse – weiter unten ausgewertet – erhält man durch die Interpolation über die Gaußkurve [92]. Die Berechnung ist allerdings aufwendiger, da Logarithmen bestimmt werden müssen.

$$y = ae^{-b(x-x_0)^2} \quad (4.8)$$

x_0 gibt die Position u_{\max} des Maximums an und errechnet sich zu:

$$u_{\max} = u_{NCC \max} + \frac{\ln y_1 - \ln y_3}{2 \ln y_1 - 4 \ln y_2 + 2 \ln y_3} \quad (4.9)$$

Die Gaußinterpolation führt im Vergleich zur Parabelinterpolation zu einer leicht verbesserten Positionsbestimmung, wie später noch gezeigt wird.

4.5.2.4 Interpolation über Geraden

Eine sehr einfache Interpolation, die ebenfalls von der Symmetrie der Korrelationskurve ausgeht, ist die Spiegelung einer Geraden in der Umgebung des Korrelationsmaximums. Diese Methode ist in [93] angedeutet.

Eine Gerade wird durch den Punkt mit dem maximalen ZNCC-Wert (y_2) und dem kleinsten Punkt der drei ZNCC-Werte gelegt, s. Bild 110. Im mittleren ZNCC-Wert (hier y_3) wird die Gerade mit der negativen Steigung gespiegelt. Der Subpixelwert u liegt im Schnittpunkt, der sich so berechnen lässt:

$$u = \frac{y_3 - y_1}{2y_2 - 2y_3} \text{ für } y_1 \geq y_3 \quad (4.10)$$

und

$$u = \frac{y_1 - y_3}{2y_1 - 2y_2} \text{ für } y_1 < y_3 \quad (4.11)$$

Für den Korrelationsverlauf in Bild 110 ergibt sich für die Lineare Approximation ein Subpixelwert von 0.3673, während die Parabelfunktion mit 0.2903 deutlich näher am erwarteten Wert von 0.3125 liegt.

Die Interpolation über Geraden wird nicht weiter verfolgt, da sie in der Regel größere Abweichungen der Subpixelwerte im Vergleich zu den anderen vorgestellten Methoden liefert.

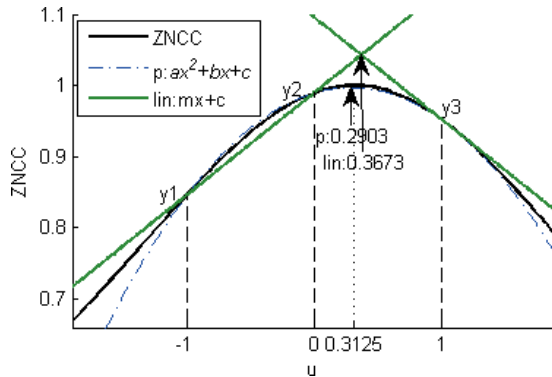


Bild 110: Approximation mit gespiegelten Geraden

4.5.2.5 Vorzeichenabhängige Interpolation

Die Motivation für die Einführung der vorzeichenabhängigen Interpolation sehen S. Masao und O. Masatoshi [94] hauptsächlich im s. g. Pixel-Locking. Hier handelt es sich um auffällige Häufungen von Subpixelwerten bei der Parabelapproximation in der Nähe ganzzahliger Pixelwerte. Bei einer sehr großen Anzahl von zufälligen Bildverschiebungen wäre zu erwarten, dass alle Subpixelwerte gleichverteilt sind. In einem späteren Abschnitt wird tiefer auf das Pixel-Locking eingegangen.

Die Betrachtung des Pixel-Lockings führt S. Masao und O. Masatoshi [94] zu einem weiteren Ansatz, die Subpixelbestimmung zu verbessern. Ausgehend von einer Fehlerkurve bei der Parabelapproximation, wie sie in Bild 111 nachvollzogen wurde, ist zu erkennen, dass der größte Fehler etwa in der Nähe von -0.25 oder +0.25 Pixel liegt. Der Verlauf der Kurve ist prinzipiell mit der Kurvenform, der Lage des Maximums und des Minimums etwa gleich, allerdings sind die Abweichungen vom Bildinhalt abhängig.

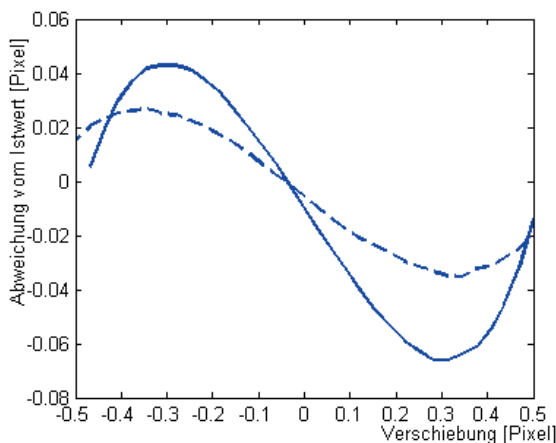


Bild 111: Fehlerverlauf bei Subpixelverschiebungen. Durchgezogene Linie: Ausschnitt aus Bild 96. Gestrichelte Linie: Ausschnitt aus Bild 95.

In [94] wird mit Hilfe der Mittelwertbildungen zweier um ein Pixel nach rechts und links verschobenen Bilder mit sich selbst

$$I_{m1}(u) = \frac{I(u) + I(u+1)}{2} \quad \text{für } -0.5 \leq \Delta u < 0 \quad (4.12)$$

$$I_{m2}(u) = \frac{I(u) + I(u-1)}{2} \quad \text{für } 0 \leq \Delta u \leq 0.5 \quad (4.13)$$

erneut die NCC berechnet. Natürlich braucht die NCC nur für zwei Bildversätze neu errechnet zu werden, denn die Ganzpixelposition ist schon vorher errechnet worden. Für die vertikale Richtung v gilt dasselbe. In den Gleichungen (4.12) und (4.13) muss dann u mit v substituiert werden.

Die Subpixelpositionen werden wieder nach der Parabelscheitelpunktfunktion (4.7) (oder für die Gaußapproximation nach (4.9)) errechnet. Da die Mittelwertbildung genau einer Verschiebung um 0.5 Pixel entspricht, muss bei einer Berechnung für ursprünglich negative Subpixel 0.5 abgezogen werden und bei positiven Subpixel 0.5 hinzugefügt werden.

4.5.2.6 Interpolation durch Bildvergrößerungen

Ähnlich wie durch die Mittelung benachbarter Bildausschnitte kommt man durch Vergrößerung eines Bildes auf feinere Auflösungen [95]. Die Bilder werden hierzu durch bikubische Interpolation auf das 16-Fache vergrößert und dann die gewonnenen NCC-Werte über eine Parabelfunktion interpoliert. Nach den Autoren ist der Subpixelgenauigkeitsgewinn um ca. 40 % höher gegenüber der Verarbeitung der Originalbilder. Da der Rechenaufwand sehr groß ist, wird dieser Ansatz nicht weiter verfolgt.

4.5.2.7 Neue Methode über 'inverse Parabel'

Bilder können nicht nur im spatialen Raum, sondern auch im Frequenzraum betrachtet werden. In Analogie zur Zeitfunktion (z. B. in der Nachrichtentechnik) gilt im spatialen Raum die Ortsfunktion. Das heißt, die Zeit in Sekunden in der Nachrichtentechnik korrespondiert zur Strecke in Pixel in der Bildverarbeitung. Damit können die bekannten Theoreme der Nachrichtentechnik angewendet werden. Im Weiteren werden wegen der Anschaulichkeit nur eindimensionale „Bilder“, also Bildzeilen, betrachtet.

In grober Näherung kann ein Bild als weißes Rauschen angesehen werden. Die Autokorrelation ist dann immer null, außer wenn beide Bilder übereinander liegen.

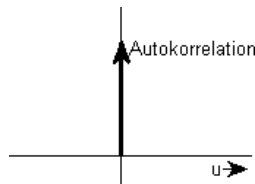


Bild 112: Autokorrelation vom weißen Rauschen

Als Nächstes ist bekannt, dass die spatiale Frequenz eines Bildes begrenzt ist, da die kleinste örtliche Periode durch 2 Pixel, z. B. schwarz und weiß, gegeben ist. Im einfachsten Fall kann die Bandbegrenzung durch eine Multiplikation des Spektrums der Autokorrelationsfunktion mit einem Rechteck dargestellt werden. Die Rücktransformation ist dann eine $\frac{\sin u}{u}$ -Funktion, s.

Bild 113.

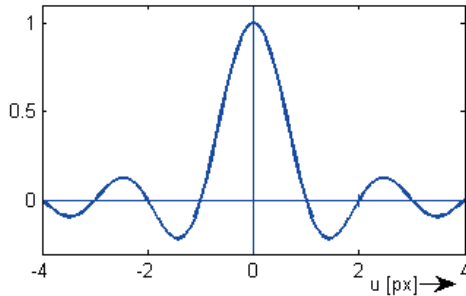


Bild 113: Autokorrelationsfunktion von rechteckbandbegrenztem weißen Rauschen

Wie sich später herausstellt, ist die Approximation durch die $\frac{\sin u}{u}$ -Funktion schon etwas besser als durch eine Parabel.

Die Autokorrelation ist eine Faltung mit derselben, aber zeitinversen, Funktion. Das dazugehörige Spektrum wird Energiedichtespektrum genannt, da es sich entsprechend Gleichung (4.14) mit dem Betragsquadrat des Spektrums eines Signals – hier eines Bildes – errechnet [91].

$$I(u) * I(-u) \rightarrow \left| \mathcal{F}\{I(u)\} \right|^2 \quad (4.14)$$

Betrachtet man das Energiedichtespektrum realer Landschaftsbilder und Luftaufnahmen, so zeigt sich, dass alle einen ähnlichen Verlauf haben, unabhängig davon, ob die Aufnahmen von strukturierten Gebieten (Städte, s. Bild 96) oder natürlichen Landschaften sind. Ein typischer Verlauf des Energiedichtespektrums, der genauer untersucht wird, ist in Bild 114 als durchgezogene Linie dargestellt.

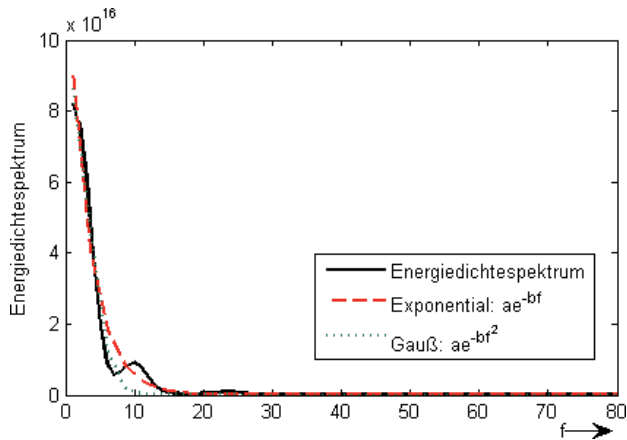


Bild 114: Energiedichtespektrum eines Bildes (s. Bild 95) mit passender Exponential- und Gaußapproximation

Zur Erstellung des Energiedichtespektrums wurden Zeilen aus einem Bild vom Mittelwert (Gleichanteil) befreit, zum Erzeugen einer stetigen Periodizität global mit einer Gaußfensterfunktion multipliziert und dann fouriertransformiert. Die Beträge dieser Fouriertransformierten sind nun quadriert und addiert worden.

Zunächst sieht es so aus, dass eine Bildkorrelation in groben Zügen einer Gaußkurve entspricht, gepunktete Linie in Bild 114. Aus der logarithmischen Darstellung des Energiedichtespektrums, Bild 115, kann allerdings ersehen werden, dass eine Exponentialfunktion eine bessere Approximation als eine Gaußfunktion ist. Die Gaußfunktion weicht schon bei niedrigen Frequenzen sehr schnell vom Energiedichtespektrum ab.

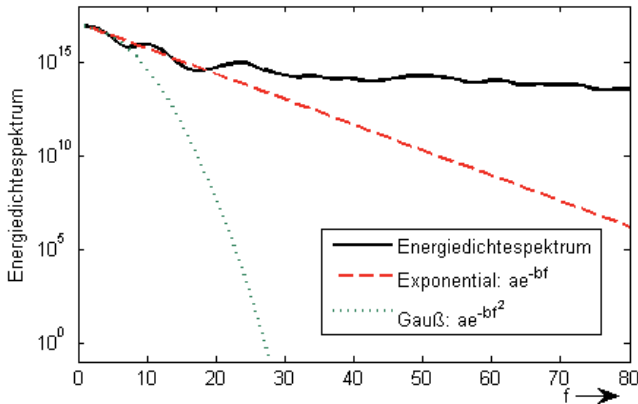


Bild 115: Energiedichtespektrum eines Bildes (s. Bild 95) mit passender Exponential- und Gaußapproximation, wie Bild 114, allerdings mit logarithmischer Koordinate

Da das Energiedichtespektrum aus einem reellen Signal, nämlich einem Bild, gewonnen wurde, ist es eine gerade Funktion. Die Rücktransformation des Doppel-exponentialspektrums [91] ergibt.:

$$e^{-|f|} \rightarrow \frac{1}{(1+x^2)\pi} \quad (4.15)$$

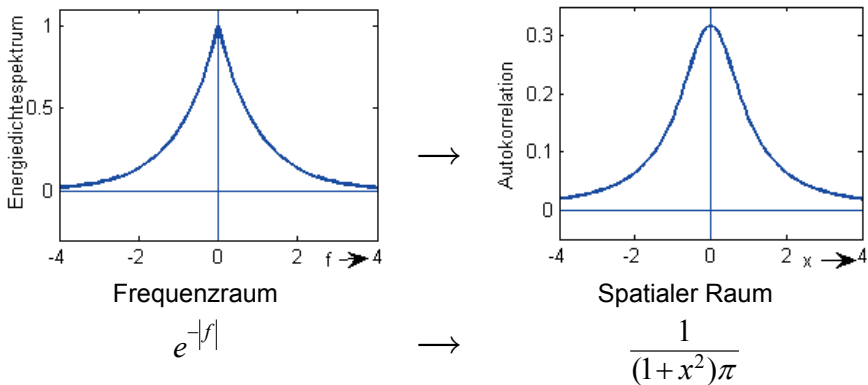


Bild 116: Exponentialfunktion und Rücktransformation

Allgemeiner ausgedrückt, nimmt die Ortsfunktion aus (4.15) diese Form an:

$$\frac{1}{ax^2 + bx + c} \quad (4.16)$$

Diese Funktion wird in dieser Dissertation als 'inverse Parabel' bezeichnet.

Mit drei Punktepaaaren

$$(x_1, y_1), (x_2, y_2), (x_3, y_3) \quad (4.17)$$

mit $x_1 = x_2 - 1$ und $x_3 = x_2 + 1$

lässt sich die Position x_s des Scheitelpunktes bezüglich x_2 errechnen zu:

$$x_s = \frac{y_1 y_2 - y_2 y_3}{4y_1 y_3 - 2y_2 y_3 - 2y_1 y_2} \quad (4.18)$$

4.5.2.8 Pixel-Locking

Pixel-Locking bezeichnet den Effekt, dass bei vielen zufälligen Bildversätzen, deren Subpixelverschiebungen über Approximationsfunktionen berechnet werden, in der Nähe von ganzzahligen Werten dominante Häufungen der berechneten Subpixelwerte vorhanden sind. In Bild 117 sind die Häufungen um die ganzzahligen Werte 0, 1, 2, 3 besonders auffällig.

Pixel-Locking ist eingehend durch S. Masao und O. Masatoshi besprochen [94] und oft zitiert worden.

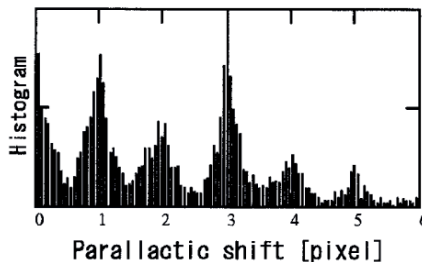


Bild 117: Pixel-Locking. Kumulation der Verschiebungswerte bei ganzen Zahlen nach einer Grafik aus [94]. Die Subpixelwerte wurden über eine Parabelapproximation errechnet.

Werden viele zufällige Bildverschiebungen subpixelgenau berechnet, wird eine Gleichverteilung der Berechnungswerte erwartet. Wenn die Gleichverteilung nicht gegeben ist, liegt ein systematischer Fehler in der Subpixelberechnungsmethode vor.

Die in Bild 117 sehr dominante Häufung um die ganzzahligen Verschiebungen konnte in eigenen Berechnungen in dieser Stärke nicht bestätigt werden. In [96] wird erwähnt, dass der Effekt des Pixel-Lockings besonders stark bei Bildern mit Objekten, die wenige Pixel an Fläche einnehmen, ausgeprägt ist.

Eigene Berechnungen über verschiedene Bilder und insgesamt 46332 Bildpaare führten bei Verschiebungen bei einigen Verfahren der Subpixelbestimmung auch zu einer Dominanz der Häufigkeiten beim ganzzahligen Wert, der in Bild 118a - e bei 0.016 (hell schraffierter Balken) liegt.

Die Häufigkeitsgrafiken sind am besten von links nach rechts zu lesen. Z. B. stellt der schraffierte Balken die Häufigkeit der errechneten Werte, die zwischen -0.016 und +0.016, also um den ganzzahligen Wert null liegen, dar. Bei der Parabelapproximation, Bild 118a, ist das Verhältnis der Häufigkeiten um ± 0.25 px zu den Häufigkeiten um 0 px wie 1300 zu 1900 oder 1:1.5, während nach Bild 117 das Verhältnis etwa 5 ist und damit in dieser Stärke nicht bestätigt ist.

Die Gaußkurvenapproximation, Bild 118b, verhält sich bezüglich der Häufigkeitsverteilung ähnlich wie die Parabelapproximation.

Die anderen Berechnungen sind weitgehend gleichverteilt, wobei die vorzeichenabhängige Approximation, Bild 118e, am ausgeglichensten ist. Die vorzeichenabhängige Approximation wurde hier im Endschrift mit der 'inversen Parabel' durchgeführt.

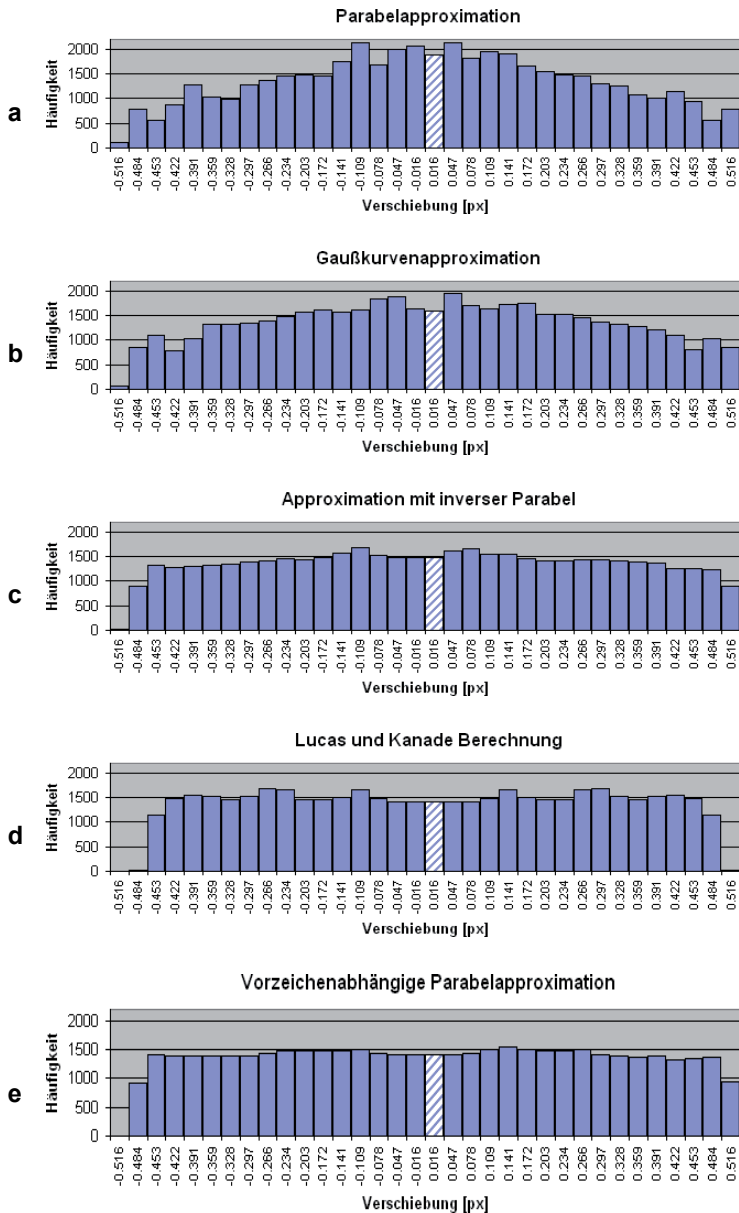


Bild 118a - e: Häufigkeitsverteilung der errechneten Verschiebungen bei verschiedenen Approximationen. Die tatsächlichen Verschiebungen sind gleichverteilt.

4.5.3 Vergleich der Subpixelfunktionen

In den vorhergehenden Unterkapiteln sind Funktionen vorgestellt worden, die den Korrelationsverlauf annähernd nachbilden. Die erhöhte Häufung der Subpixelwerte um ganzzahlige Verschiebungen führte nach [94] zur vorzeichenabhängigen Interpolation.

Die Approximationsfunktionen

- Parabel
 - Gaußfunktion
 - sinc-Funktion
 - 'inverse Parabel'
 - vorzeichenabhängige Interpolation
 - und zusätzlich das Gradientenverfahren nach Lucas&Kanade
- werden in diesem Kapitel auf Subpixelgenauigkeit untersucht und quantifiziert.

In Bild 119 ist die NCC, durchgezogene Linie, im Bereich von -1.5 bis 1.5 Pixel mit den verschiedenen Ausgleichskurven abgebildet. Die Ausgleichskurven wurden so eingepasst, dass sie genau an den Verschiebungen $u = -1, 0$ und $+1$ die NCC-Kurve schneiden oder berühren. Zu erkennen ist, dass alle Kurven im relevanten Bereich von $u = -1$ bis $+1$ zunächst einander ähnlich sind.

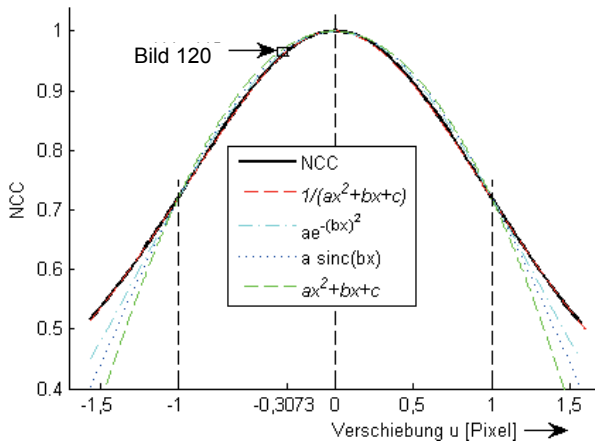


Bild 119: NCC mit verschiedenen Ausgleichsfunktionen, Mittelwert des Teilbildes wurde abgezogen

Doch durch eine Betrachtung im Detail, Bild 120, ist der Unterschied der Approximationen erkennbar. Die Kurven stellen von rechts nach links die NCC und dann die Approximationen mit 'inverser Parabel', mit Gaußkurve, mit sinc-Funktion und mit regulärer Parabel dar. In dieser Reihenfolge ist auch die Genauigkeit zum betrachteten Subpixelwert, hier 0.3073. Die 'inverse Parabel' schmiegt sich wesentlich besser an die NCC als die Gaußapproximation und ist einfacher zu berechnen. Die sinc-Funktion wird nicht weiter betrachtet, da sie aufwendig zu berechnen ist und keine besonders gute Approximation ist.

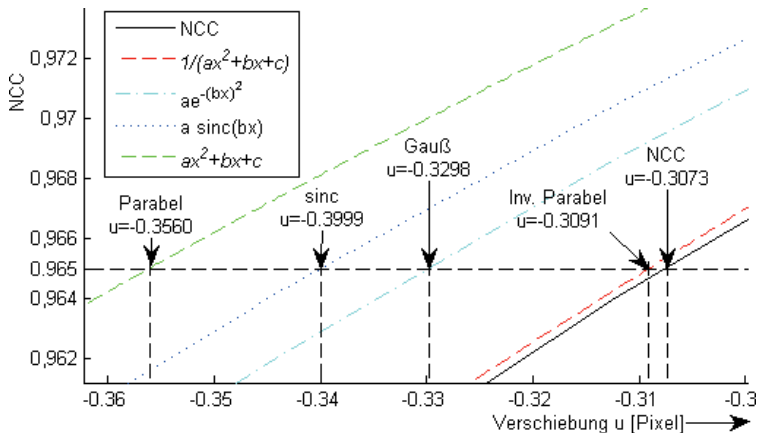


Bild 120: Ausschnitt aus Bild 119

Wird der Mittelwert vom Teilbild bei der Berechnung der NCC nicht von beiden Bildern abgezogen, so ist die 'inverse Parabel' immer noch die bessere Approximation, allerdings nicht mehr so deutlich, s. Bild 121 und Bild 122.

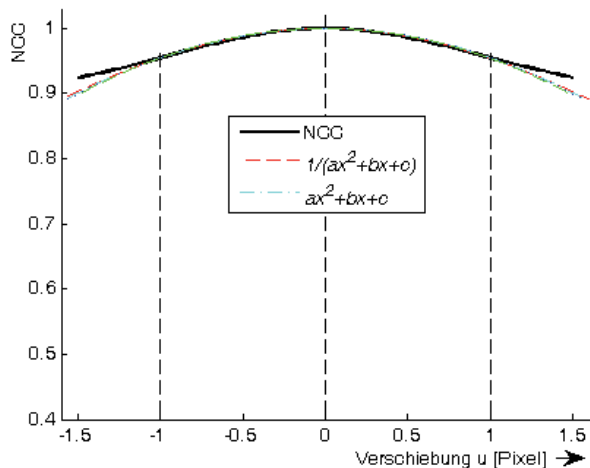


Bild 121: Wie Bild 119, allerdings nicht mittelwertfrei, nur für Parabel- und 'inverse Parabelfunktion', andere Funktionskurven liegen zwischen den Parabelfunktionen

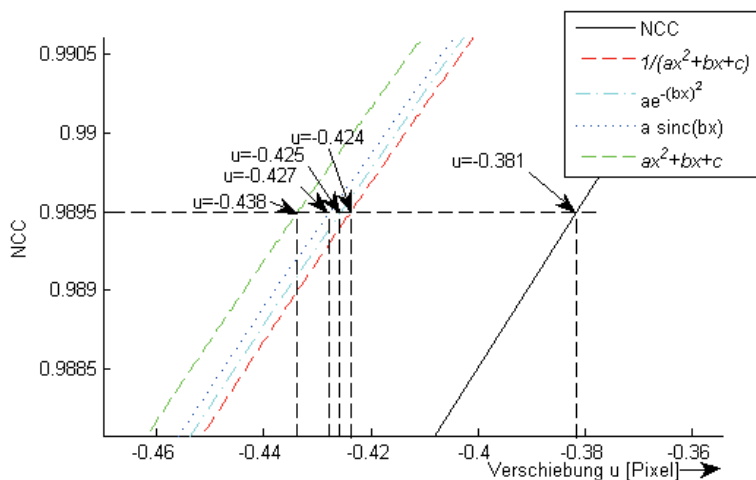


Bild 122: NCC mit verschiedenen Ausgleichsfunktionen, ohne Abzug des Mittelwerts des Teilbildes

Bei einer Verschiebung von -0.2500 Pixel ergibt sich eine Grafik, wie sie in Bild 123 gezeigt ist. Bei der Berechnung der NCC sind nur die NCC-Werte bei $u = -1$, 0 und 1 bekannt. Durch diese drei Punkte gehen die

Approximationskurven, deren Maxima die vermuteten Subpixelpositionen angeben.

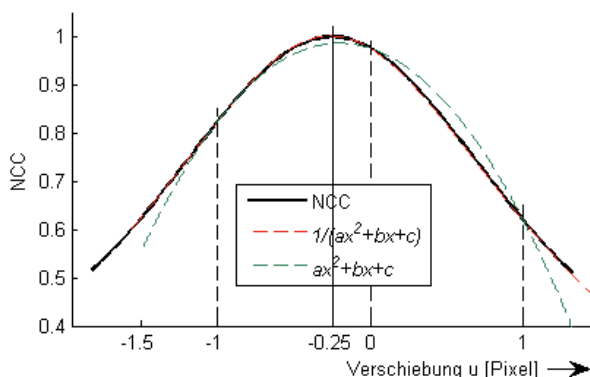


Bild 123: NCC, Approximation mit einfacher und 'inverser Parabel' bei einer Verschiebung um -0.25 Pixel

Aus Bild 124 – ein Ausschnitt aus Bild 123 – kann ersehen werden, dass die 'inverse Parabel' in diesem Fall einen wesentlich genaueren Wert liefert als die reguläre Parabel.

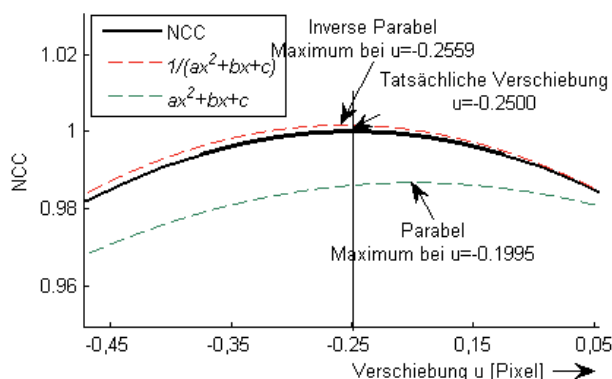


Bild 124: Ausschnitt im Bereich des Maximums aus Bild 123

Die Auswertung von 46332 Bildpaaren, die zwischen -0.5 bis $+0.5$ Pixel um jeweils $1/32$ Pixel verschoben sind, ergibt eine Verbesserung der maximalen Abweichung zwischen der Parabelapproximation und der Approximation durch die 'inverse Parabel' um fast den Faktor 2, s. Tabelle 8. Während bei der Parabelapproximation 99.7 % (entspricht einem 2σ -Wert bei einer Standardabweichung) der Subpixelabweichungen in einem Bereich zwischen -

0.076 und +0.076 liegen, ist bei der 'inversen Parabel' der Fehlerbereich eingeeengt auf ± 0.049 Pixel.

Die vorzeichenabhängige Approximation unterscheidet sich je nach Anwendung der Approximation über Parabel oder 'inverse Parabel' hauptsächlich beim maximalen Fehler (0.044 px), der bei der regulären Parabel 0.058 px ist.

Die Standardabweichung wurde hier aufgeführt, weil dies bei ähnlichen Statistiken üblich ist, hat in diesem Zusammenhang allerdings wenig Bedeutung, da die Fehlerverteilung keiner Glockenkurve mehr ähnelt, s. Bild 125a, b.

Bei der vorzeichenabhängigen Approximation, Bild 125e, und der 'inversen Parabelapproximation', Bild 125d, konzentrieren sich die Fehler um ± 0 px, also nahe der Fehlerfreiheit.

Der Berechnungsaufwand ist bei der Parabel und 'inversen Parabel' gleich gering, mit Abstand gefolgt von der Gaußkurvenapproximation und der vorzeichenabhängigen Approximation. Die Berechnung nach Lucas & Kanade ist die aufwendigste Methode und hier nicht sehr genau. Die Methode wurde in dieser Arbeit auf Schnelligkeit optimiert und kann bezüglich Subpixelgenauigkeit verfeinert werden.

Durch ihre Genauigkeit und einfache Berechnung wird die 'inverse Parabelapproximation' bevorzugt.

Tabelle 8: Subpixelabweichungen bei verschiedenen Approximationen und der Gradientenmethode

Approximations- methode →	Parabel	Gauß	Lucas & Kanade	'Inv. Parabel'	Vorzeichen- abhängige Approximation mit 'inv. Parabel'
Max. absolute Abweichung	0.093	0.070	0.057	0.053	0.044
99.7 % haben eine Abweichung zwischen	± 0.076	± 0.066	± 0.052	± 0.049	± 0.040
Standardabweichung	0.050	0.035	0.021	0.022	0.015

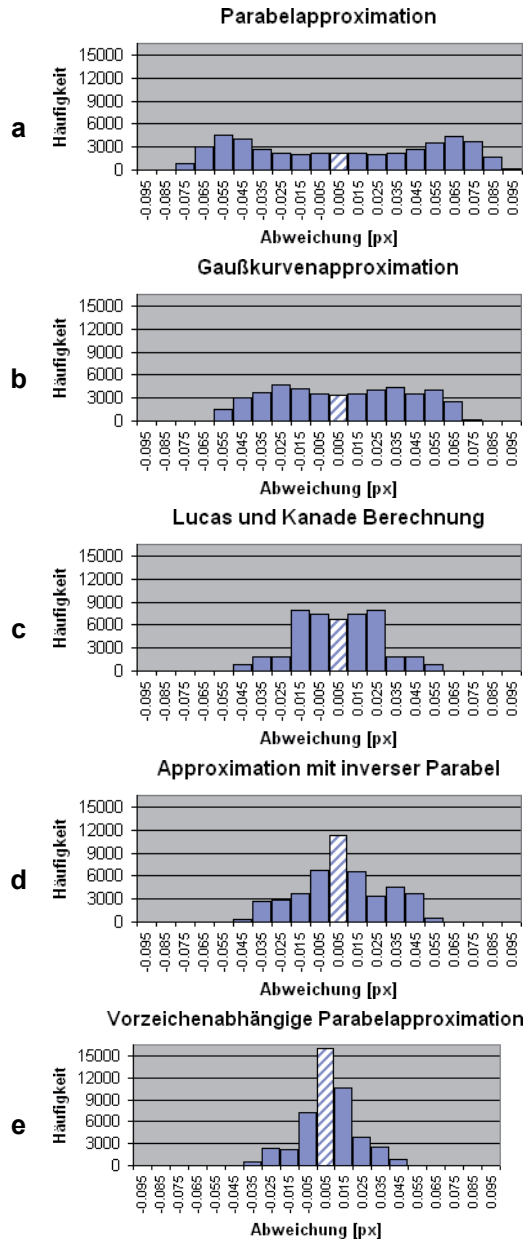


Bild 125a - e: Häufigkeiten der Subpixelberechnungsabweichungen bei verschiedenen Approximationen, die schraffierte Säule stellt Abweichungen zwischen -0.005 und $+0.005$ px um den Sollwert 0 px dar.

4.6 Zusammenfassung Kapitel 4

Die Gradientenmethode nach Lucas&Kanade hat im Vergleich zur normalisierten Kreuzkorrelation einige Schwächen für die Verwendung einer auf einem Fluggerät montierten Kamera, um Positionen zu bestimmen.

Der maximal berechenbare optische Fluss in einer Pyramidenebene ist 1.16 px. Dies ergibt bei 5 Pyramidenebenen eine Reichweite des feststellbaren Versatzes von 36 px

Bei relativ kleinen Helligkeitsunterschieden (5 %) ist der optische Fluss nicht mehr zu ermitteln. Allerdings ist die NCC bei Rotationen schon ab 2° ungenau, während die Gradientenmethode bis 9° gute Ergebnisse liefert.

Eingehend wurde die für diese Arbeit wichtige Subpixelgenauigkeit für die NCC und die Gradientenmethode untersucht. In den Vergleichen wurden für die NCC die Approximation über Parabel, Logarithmus, sinc-Funktion, vorzeichenabhängige Approximation und eine neue Methode – genannt 'inverse Parabel' – einbezogen. Das vielzitierte Pixel-Locking – erhöhte Häufigkeit nicht exakt berechneter Subpixelverschiebungen ganzzahliger Pixelpositionen – konnte nicht in dem dargestellten Ausmaß bestätigt werden. Die Ergebnisse zur Subpixelgenauigkeit sind in untenstehender Tabelle nochmals aufgeführt.

Tabelle 9: Subpixelabweichungen bei verschiedenen Approximationen und der Gradientenmethode

Approximations- methode →	Parabel	Gauß	Lucas & Kanade	'Inv. Parabel'	Vorzeichenabhängige Approximation mit 'inv. Parabel'
Max. absolute Abweichung	0.093	0.070	0.057	0.053	0.044

5 Versuchsergebnisse

Zu grundsätzlichen Untersuchungen sind künstliche Umgebungen oder die Suche im selben Bild sehr nützlich, da das erwartete Ergebnis exakt bekannt ist. Bei diesen Versuchen gehen Artefakte wie Änderung der Beleuchtung, Verzerrungen durch die Optik und andere Beleuchtungsverhältnisse nicht ein. Auf Untersuchungen mit künstlichen Bildern, deren Verhältnisse genau bekannt sind und deshalb in vielen Untersuchungen ohne Praxisbezug herangezogen werden, wird hier nicht eingegangen, sondern es werden drei nicht künstliche Umgebungen, eine virtuelle Welt, ein XY-Tisch und reale Aufnahmen betrachtet.

5.1 Virtuelle Welt

Die virtuelle Welt wird durch die Virtual Reality Modeling Language (VRML) erstellt. VRML ermöglicht, 3-D-Szenen mit Geometrien und verschiedenen Ausleuchtungen aus beliebigen Blickwinkeln zu betrachten. Der Vorteil dieser Vorgehensweise ist, dass die Verhältnisse genau bekannt sind. Der Nachteil ist, dass sie weit entfernt von der Realität ist. Trotzdem lassen sich Verfahren in dieser Umgebung als Grundvoraussetzung validieren.

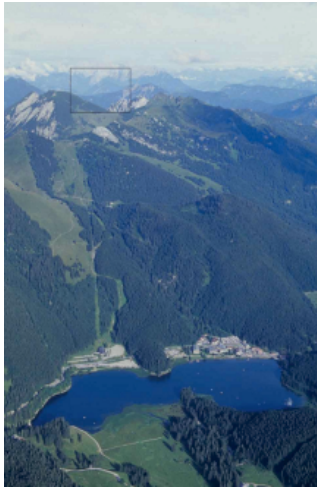


Bild 126: Fotovorlage für die virtuelle Welt

Quelle: www.bergfex.de

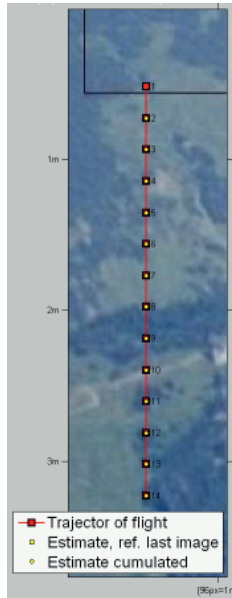


Bild 127: Gerader Flug nach unten mit kumulierter Pfadberechnung

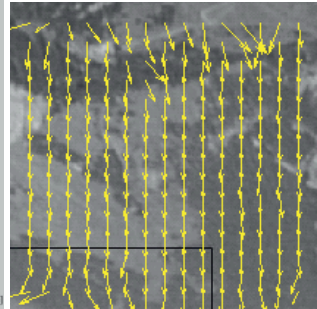


Bild 128: Flussvektoren zwischen zwei Bildern

Bild 127 zeigt einen einfachen Flugpfad innerhalb der virtuellen Welt, die ein Foto einer Landschaft – Bild 126 – im Sichtfeld hat. Das Flussfeld wurde mit der Methode von Lucas&Kanade berechnet. Da nur die Flussvektoren in der Mitte des Vektorfeldes – Bild 128 – zur Berechnung der Verschiebung genommen werden, sind Randeffekte ohne Bedeutung. Die maximale Abweichung vom Sollwert der Verschiebung zwischen zwei Bildern beträgt 0.1 px sowohl in x- als auch in y-Richtung. Der kumulierte Fehler der Pfadstrecke ist 0 px. Die Berechnung der Orientierung, die sich nicht änderte, lieferte einen maximalen Fehler von 0.3° . Während in diesem einfachen nachgebildeten Flug die Fehler vernachlässigbar sind, stellt sich die Situation bei einem stark gekrümmten simulierten Flug über ein teilweise homogenes Landschaftsbild anders dar.

Eine spiralförmige Bahn, in Bild 129 die rote Linie, wird simuliert abgeflogen. An den großen roten Quadraten werden orientierungsrichtige Bildausschnitte generiert. Jeweils zwei aufeinanderfolgende Bilder werden zur Berechnung des optischen Flusses herangezogen. Die errechnete Position in Bezug auf

das vorhergehende Bild ist jeweils durch das kleine gelbe Quadrat wiedergegeben und zeigt meist eine geringe Abweichung zur erwarteten Position. Die kumulierten Positionen, durch gelbe Kreise dargestellt, geben den errechneten Flugpfad, der sichtlich vom tatsächlichen Pfad abweicht, wieder.

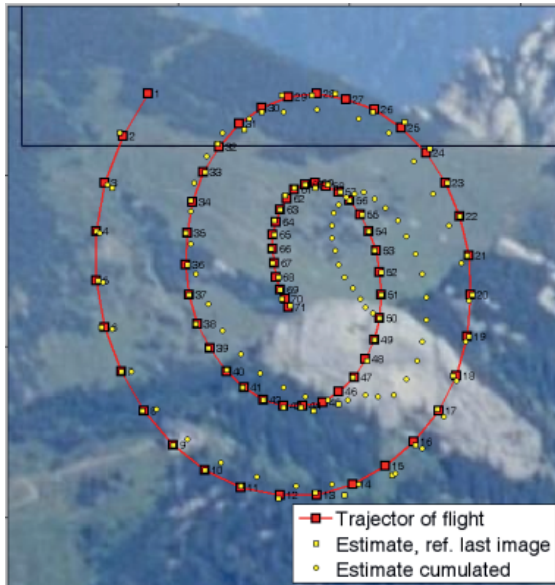


Bild 129: Virtuelle Welt, spiralförmiger Flug

Beispielhaft sei das Ergebnis der Flussvektorberechnung an einem fehlerhaft berechneten Bildpaar gezeigt, s. Bild 130 und Bild 131. Die Fehlberechnungen entstehen hauptsächlich in den höheren Gaußpyramidenebenen, da durch die starke Bildverkleinerungen homogene Bereiche oder auch Kanten entstehen können, deren Verschiebung nicht richtig berechnet werden kann.

Die Flussvektoren am Bildrand, die nicht in das Gefüge der umgebenden Vektoren passen, sind unbedeutend, da nur die Umgebung im Bildzentrum zur Positionsbestimmung gewertet wird. Aber auch hier ist zu erkennen, dass trotz Rotation der Bilder einige Flussvektoren im mittleren Bildbereich nicht zirkular verlaufen, was zu einer Fehlbestimmung der Position und der Orientierung führt, s. Tabelle 10

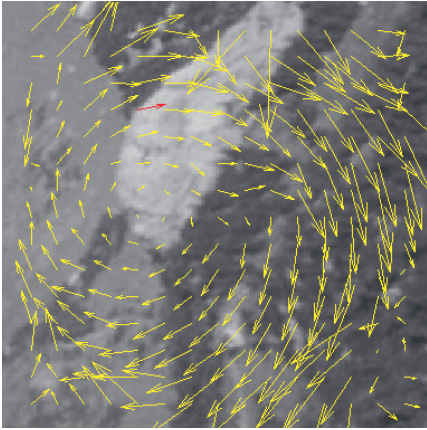


Bild 130: Erste Aufnahme

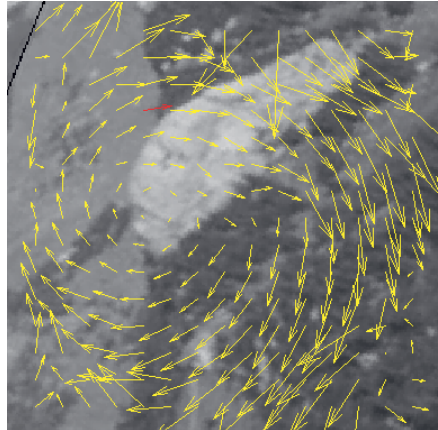


Bild 131: Nachfolgende Aufnahme

Flussvektorfeld bei einem Bildpaar mit fehlerhaften Berechnungsergebnissen

Korrespondierende Bildpunkte werden mit Flussvektoren verknüpft, wie der rote Vektor beispielhaft zeigt. Der Bildpunkt am Pfeilbeginn in der linken Abbildung, Bild 130, korrespondiert im nachfolgenden Bild (rechte Abbildung) mit dem Bildpunkt an der Pfeilspitze. Der Pfeilbeginn in der linken Abbildung und die Pfeilspitze in der rechten Abbildung weisen auf denselben Szenenpunkt hin.

Tabelle 10: Relative Positions- und Orientierungswerte bei einem Bildpaar mit fehlerhaften Berechnungsergebnissen

	realer Wert	berechneter Wert
Position in x-Richtung	-11 px	-6.8 px
Position in y-Richtung	2 px	-0.8 px
Orientierung	15.7°	7.5°

In Bild 132 und Bild 133 sind die Flussvektorfelder im mittleren Bildbereich gleichmäßig und führen zu guten Ergebnissen, die in Tabelle 11 aufgeführt sind.

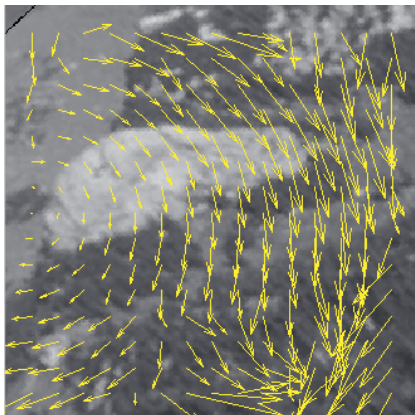


Bild 132: Erste Aufnahme

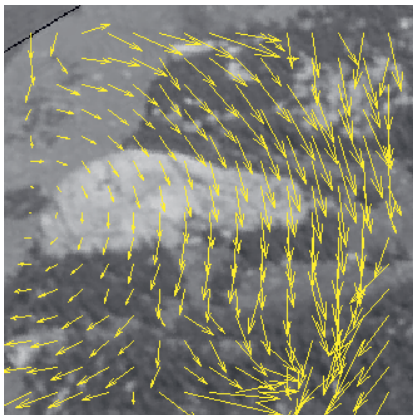


Bild 133: Nachfolgende Aufnahme

Flussvektorfeld zu einem Bildpaar mit richtigen Ergebnissen im mittleren Bildbereich

Die Abweichungen vom realen Wert sind geringfügig, s. Tabelle 11.

Tabelle 11: Relative Positions- und Orientierungswerte bei einem Bildpaar mit guten Berechnungsergebnissen

	realer Wert	berechneter Wert
Position in x-Richtung	-10 px	-10.4 px
Position in y-Richtung	20 px	19.2 px
Orientierung	10.6°	10.9°

Obwohl der berechnete Pfad, in Bild 129 gelbe Kreise, dem Anschein nach deutlich vom realen Pfad (große rote Quadrate) abweicht, sind die kumulierten Fehler im einstelligen Prozentbereich, s. Tabelle 12.

Tabelle 12: Kumulierte Positions- und Orientierungswerte bei einem spiralförmigen Pfad entsprechend Bild 129

	realer Wert	berechneter Wert
Gesamtstreckenlänge	1108 px	1099 px (0.8 % Abweichung)
Entfernung von tatsächlicher Endposition	-	47 px
kumulierte Orientierungswinkel	790°	811° (2.7 % Abweichung)

Dass die Ermittlung von Position und Orientierung über den optischen Fluss prinzipiell funktioniert, ist in zahlreichen Versuchen in der virtuellen Welt erwiesen worden. Dennoch bleibt der Einfluss einer realen Kamera hier unberücksichtigt.

5.2 XY-Tisch

Um reale Bildsequenzen mit einer Kamera zu erstellen und trotzdem nachprüfbare Ergebnisse zu erhalten, wurde ein XY-Tisch angesteuert.

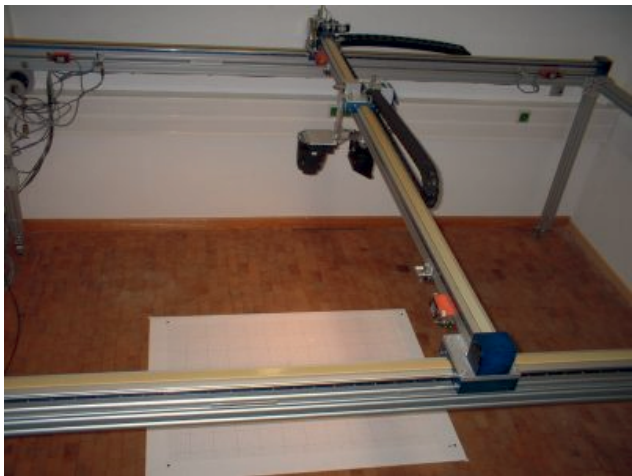


Bild 134: XY-Tisch mit montierter Kamera

Der Tisch hat die Maße 2.2 x 1.2 x 3.1 [B x H x T in m]. Die Laufkatze hat einen Fahrbereich von ca. 2 m x 3 m und wird mit der montierten Kamera auf 0.1 mm genau positioniert. Die Kamera lässt sich um $\pm 170^\circ$ in $0,075^\circ$ -Schritten drehen. Durch Verstellen des optischen Zooms können zusätzlich verschiedene Grundhöhen über Boden nachgebildet werden.

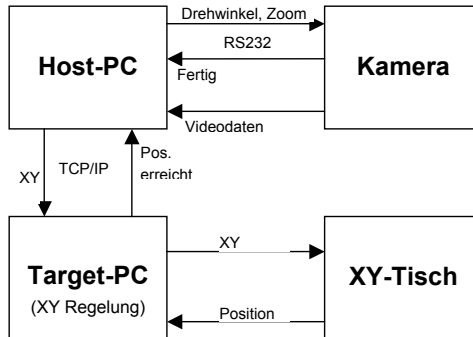


Bild 135: Positionsregelung des XY-Tisches

Die Steuerung erfolgt über einen Host-PC, der einem Target-PC die Sollposition übergibt. Der Target-PC stellt die Position mit einer Regelung, die im 10 ms-Takt erfolgt, ein und meldet das Erreichen der Zielposition dem Host. Die Kamera wird im Zoom und Drehwinkel ebenfalls vom Host-PC gesteuert und gibt eine Ausführungsbestätigung zurück. Dann wird ein Bild gespeichert. Mit dieser Apparatur kann ein vorgegebener Pfad „abgeflogen“ und die Ergebnisse der Bilderauswertung können mit den Sollwerten verglichen werden. Weiterhin kann eine Stereokamera beliebiger Basisbreite nachgebildet werden, indem ein zweiter Pfad, der parallel zum ersten Pfad verläuft, „abgeflogen“ wird. Für Abstandsmessungen wurde die Kamerablickrichtung horizontal ausgerichtet und auf eine Tafel zugefahren. Zusätzlich erfolgt eine qualitative Beurteilung über ein Mosaikbild. Hier werden die einzelnen Bilder der Filmsequenz mit den berechneten Positionen und Winkeln zu einem Gesamtbild (Mosaikbild) zusammengefügt. Fehler der Winkelberechnung werden dann besonders deutlich.

5.2.1 Ergebnisse

5.2.1.1 Gerader Pfad

Eine Filmsequenz über einen geraden Pfad ist der einfachste Fall für die Berechnung der Verschiebungen und wird von beiden Verfahren (Lucas&Kanade und NCC) recht gut bewältigt, wobei das Gradientenverfahren einen bemerkbaren Winkelfehler aufwies, s. Tabelle 13, während die Translationsfehler bei beiden Verfahren gering waren. Bild 136 und Bild 137 zeigen das jeweilige Mosaikbild.

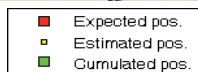
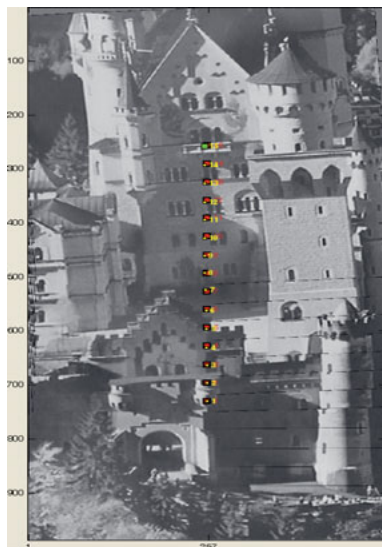


Bild 136: Gerader Pfad,
Berechnung nach Lucas&Kanade

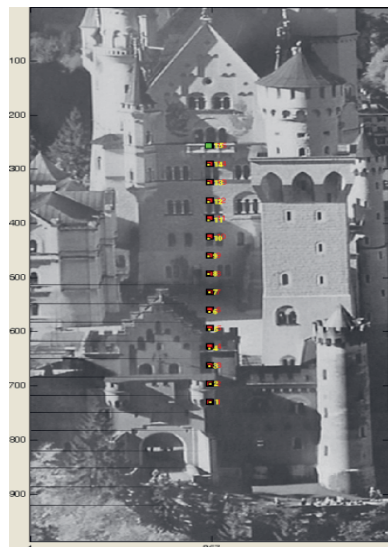


Bild 137: Gerader Pfad,
Berechnung mit NCC

Tabelle 13: Ergebnisse bei der Berechnung nach Lucas&Kanade (L&K), s. Bild 136 und mit NCC, s. Bild 137

Methode	Erwartete kumulierte Werte	Abweichung absolut		Abweichung prozentual	
		L&K	NCC	L&K	NCC
x-Richtung	0 px	3 px	2 px	-	-
y-Richtung	477 px	1 px	1 px	0.2 %	0.2 %
Winkel	0°	3.6°	0.2°	-	-

5.2.1.2 Kreisbahn

Weitaus schwieriger richtig zu berechnen, sind die Verschiebungen bei einer Kreisbahn. Während das Gradientenverfahren, Bild 138, hier hauptsächlich an der Schrittweite (47 px) scheiterte, lieferte die NCC, Bild 139, gute Ergebnisse, siehe Tabelle 14.

Qualitativ kann das Ergebnis auch anhand des Mosaikbildes, Bild 139, beurteilt werden: Die Straße im linken Teil, in der Realität gerade, bleibt weitgehend gerade.

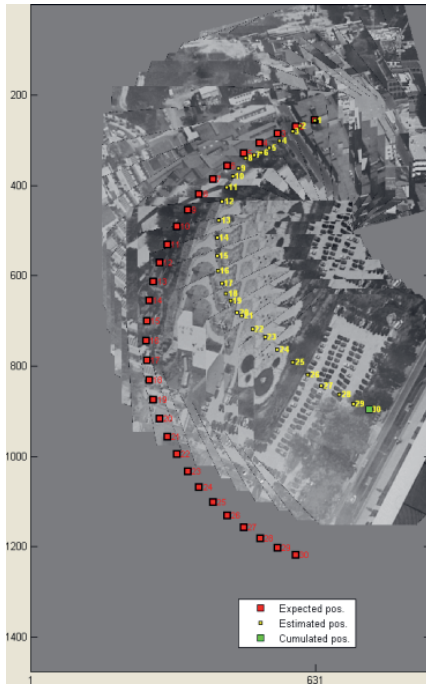


Bild 138: Kreisbahn mit 5° Winkeländerung und großen Schritten, Berechnung nach Lucas&Kanade

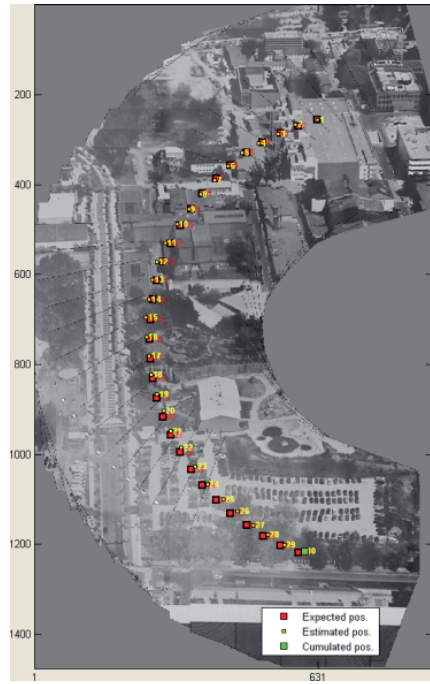


Bild 139: Gleiche Bahn wie links, Berechnung über NCC

Tabelle 14: Kumulierte Strecken und Winkel mit Abweichungen von der Endposition bei der Berechnung über NCC, s. Bild 139

	Kumulierte Absolutwerte	Abweichung vom erwarteten Wert	Fehler
x-Richtung	734 px	17 px	2.3 %
y-Richtung	958 px	2 px	0.2 %
Winkel	145.0°	0.6°	0.4 %

Weitere eigene Vergleiche sind in [85] veröffentlicht.

Die Reichweite der Gradientenmethode mit 36 px ist zu gering für den Einsatz in einem dynamischen Flugsystem. Die Auswirkungen der geringen

Reichweite seien mit den Verhältnissen beim Landeanflug anhand der Geometrie bei der Translation berechnet.

Folgende Daten seien gegeben:

- Brennweite f_{px} : 1401 px (entspricht 5.3 mm)
- Bildrate fps: 12/s
- Höhe über Boden z : 2 m
- Geschwindigkeit beim Landeanflug: 1 m/s (3.6 km/h)

Die Verhältnisse sind in Bild 44 dargestellt.

Aus der Bildrate und der Geschwindigkeit errechnet sich die geflogene Strecke x_w zwischen der Aufnahme an Position 1 und der Aufnahme an Position 2 zu 1/12 m.

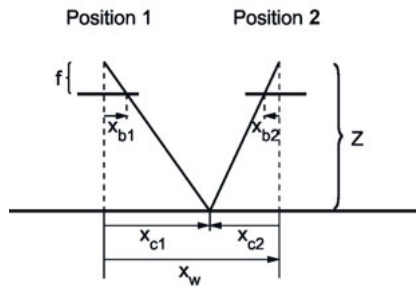


Bild 44 Geometrie bei Translation (aus Kapitel 2.6.2)

Mit der Gleichung für die Translation, sie sei hier wiederholt,

$$x_w = \frac{z}{f_{px}} (x_{b1px} - x_{b2px}) \quad (2.52)$$

errechnet sich die Verschiebung $x_{b1px} - x_{b2px}$ zu 58 px.

Wie das Beispiel verdeutlicht, wird beim Landeanflug – hier mit einer Verschiebung von 58 px – die Reichweite von 36 px der Gradientenmethode deutlich überschritten.

Deshalb wird die Gradientenmethode aus den weiteren Betrachtungen ausgeschlossen.

5.3 Flugaufnahmen mit Monokamera

Zunächst wird für die Aufnahmen während der Flüge mit dem Blimp der FernUniversität eine Hero3 Kamera der Firma GoPro verwendet. Diese Kamera hat eine so stark verzerrende Linse, dass eine Kalibrierung mit üblichen Mitteln erfolglos blieb. Durch geringes Gewicht und internen Speicher für die Aufzeichnung ist sie für erste Tests geeignet. Die ausgewerteten Bildbereiche befinden sich hierbei in der Nähe der weitgehend verzerrungsfreien Bildmitte und Subpixel werden nicht berücksichtigt.

Die Translation und Rotation der Bildsequenzen werden mit der NCC berechnet und die Bilder entsprechend der Berechnungen zu einem Mosaikbild zusammengefügt. Zur Prüfung wird bei kurzen Aufnahmesequenzen das erste Bild in Farbe mit transparentem Mittelteil dem Mosaikbild überlagert. Im transparenten Teil ist dann in schwarz-weiß das letzte dem Mosaikbild zugefügte Bild zu sehen. An den nahtlosen Übergängen ist dann das Gelingen der Berechnung erkennbar.

Bild 140 ist das Mosaikbild, das nach einem kurzen Flug über einen Weg erstellt wird. In der Mitte des Bildes ist der errechnete Flugpfad mit der farbigen Linie dargestellt. Das dunkle Quadrat in der Mitte ist die erste Position, während das grüne Quadrat weiter oben für die letzte Position steht. In Bild 141 sind an den nahtlosen Übergängen vom ersten Bild (in Farbe) zum letzten Bild (in s/w) die richtigen Berechnungen zu erkennen.



Bild 140: Flug über einen Weg.
Bilder übereinandergelegt zu einem
Mosaikbild entsprechend der
berechneten Translation und Rotation.



Bild 141: Überlagertes erstes Bild (in
Farbe) mit transparentem Mittelteil,
welches das letzte Bild (in s/w)
durchscheinen lässt.

Ein weiterer Flug zeigt die Genauigkeit der Winkelmessung nach 86 Aufnahmen. Der Anfangswinkel des mittleren Weges auf der Bild 142 ist 83° . Nachdem sich das Fluggerät mit der Kamera um 25.5° gedreht hat, nimmt der Weg in Bild 143 einen Winkel von 57.5° ein.

Der Winkel der kumulierten Rotationsberechnungen ist 28.5° . Um diesen Winkel wird Bild 143 nach links gedreht und als letztes Bild dem ersten Bild (Bild 142) überlagert, um das Mosaikbild, Bild 145, zusammenzusetzen. Richtig ist eine Rotation um 25.5° . Die Differenz von 3° ergibt bei 85 sequentiellen Bildpaaren einen durchschnittlichen Fehler von 0.035° .



Bild 142: Flug mit starker Rotation,
erstes Bild,
Winkel des mittleren Weges 83°

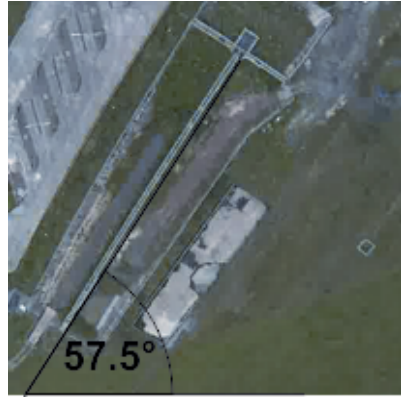


Bild 143: Flug mit starker Rotation,
letztes Bild,
Winkel des mittleren Weges 57.5°

In der Bild 145 ist das erste Bild der Flugaufnahmen (Bild 142) als Farbbild mit transparentem Mittelteil dem letzten Bild (Bild 143) überlagert, wobei das letzte Bild um die errechnete Translation und Rotation zurückversetzt bzw. zurückrotiert wird. Diese Rückpositionierung ist sowohl in Bild 144 als auch im Mittelteil der Bild 145 zu sehen.

Zusätzlich zur Winkelabweichung ist in Bild 145 eine Translationsabweichung erkennbar, die hauptsächlich durch Skalierungsänderung wegen eines geringeren Abstandes zur Szene begründet ist.

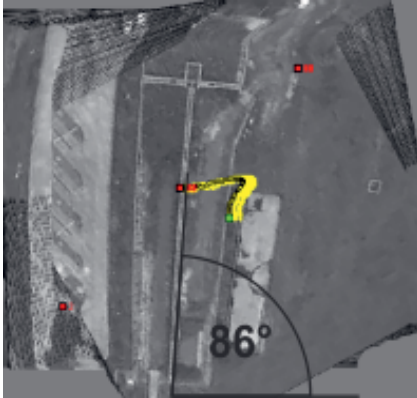


Bild 144: Mit berechneten Winkeln und Translationen überlagerte Bilder, Winkel des mittleren Weges 86° (Sollwinkel 83°), letztes Bild ist obenauf



Bild 145: Erstes Bild als Farbbild mit transparentem Mittelteil dem letzten Bild überlagert

Bei längeren Filmaufnahmen überlappt das erste Bild nicht mehr das letzte Bild, deshalb erfolgt die Kontrolle nur noch nach Augenschein.

Ein Beispiel ist ein Landeanflug auf eine Betonstraße (Bild 146). Die kleinen nummerierten Quadrate sind jeweils in der Mitte der laufenden Aufnahmen und geben den von der Kamera erfassten Flugpfad wieder.

Die Richtigkeit der Berechnung wird aus dem kontinuierlichen Verlauf des von der Kamera aufgezeichneten Flugpfades und den nahtlosen Bildübergängen entnommen.

Erstes Bild mit
überdeckenden Anteilen
der Bilder 2 bis 10

Letztes Bild

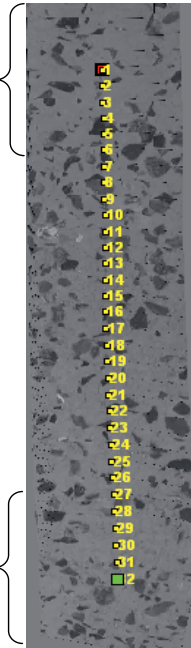


Bild 146: Landung auf einer Betonpiste

5.4 Auswertung von Stereoaufnahmen

5.4.1 Kameradaten

Für Stereoaufnahmen dient die Bumblebeekamera XB3 der Firma Point Grey (Bild 147). Die Kamera hat sich bei anderen Forschungsvorhaben bewährt, allerdings wurde sie aufgrund der relativ kleinen Basisbreite¹ von 24 cm nicht für große Entfernungen eingesetzt, sondern meistens innerhalb von Gebäuden.



Bild 147: Bumblebeekamera der Firma Point Grey

¹ Eine andere Kamera, Cameleon Academic Pack von Optomotive mit variabler Basisbreite, erwies sich als ungeeignet, da u. a. durch Überbelichtungen unbrauchbare Bilder entstanden.

Im Folgenden werden einige Kenndaten und eigene Messungen zur Bumblebee aufgeführt.

5.4.1.1 Daten nach Herstellerangaben

Physische Daten:

- Drei Objektive im Abstand von jeweils 12 cm, hier werden nur die beiden äußeren Objektive benutzt.
- Abmessungen: 277 x 37 x 41.8 [mm]
- Masse: 505 g

Elektrische Daten:

- Betriebsspannung: 12 V
- Leistungsaufnahme: 4 W

Anschlüsse:

- IEEE-1394b fire-wire-Anschluss für Bildübertragung und Kameraeinstellungen
- Mehrzweckanschluss für Triggersignale (unbenutzt)

Sensor- und Objektivdaten:

- Sensor: Sony ICX445 (CCD)
- Sensordiagonale: 6 mm
- Pixelgröße: 3.75 x 3.75 [μm]
- Pixelanzahl: 1280 x 960
- Maximale Bildrate: 16/s
- Nominale Brennweite: 6 mm

5.4.1.2 Eigene Messungen

Synchronisation und Bildratenkonstanz

Um die Synchronisation und Bildratenkonstanz zu ermitteln, wurden Aufnahmen von einem schnell rotierenden Ventilator und blinkenden LEDs erstellt. Die Berechnungsmethoden sind im Anhang Kap. A.5, S. 202 beschrieben.

Die Synchronisation der linken und rechten Kamera ist besser als 15 μs . Hierbei ist die rechte Kamera immer voreilend. Der sich ergebende Abstandsfehler für Geschwindigkeiten unter 50 km/h bleibt kleiner als 0.1 %.

Die Bildrate ist, wie nominal spezifiziert, 12/s im Durchschnitt. Allerdings weichen einige Bildfolgezeiten um bis zu 1.3 ms ab. Da die Abweichung direkt danach jeweils negativ dazu ist, heben sich die Fehler auf.

Intensitätsverlauf

Der Intensitätsverlauf der Bumblebeekamera wurde schon im Kapitel 2.4.1.1, S. 30 dargestellt und sei hier noch einmal gezeigt.

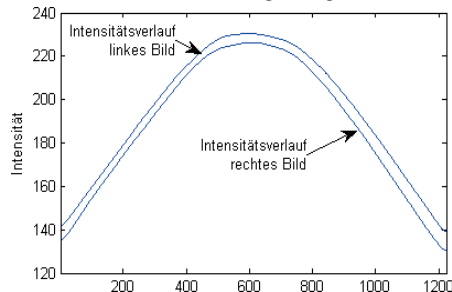


Bild 148: Intensitätsverlauf der Bumblebeekamera

Blooming-Effekt

Ein manchmal störender Effekt der Kamera ist, dass Überbelichtungen einen sogenannten Bloomingeffekt ergeben. In Bild 149 ist deutlich der Effekt durch eine spiegelnde Fläche auf der linken Seite zu erkennen, wodurch sich eine vertikale Linie bildet, die wiederum zu fehlerhaften Positionserkennungen führen kann.



Bild 149: Blooming-Effekt links im Bild

Kalibrierungs- und Rektifizierungsergebnisse

Die aufgeführten folgenden Werte sind Ergebnisse nach der Kalibrierung und anschließender Rektifizierung. Auffällig sind die hauptsächlich durch die Rektifizierung scheinbar veränderte Brennweite und der Ort der optischen Zentren.

Kenndaten der Bumblebee-Kamera nach Kalibrierung und Rektifizierung:

- Gemeinsame Brennweite f_{px} : 1401 px (entspricht 5.3 mm, nominal: 6 mm)
- Optisches Zentrum linke Kamera cc_{lx} , cc_{ly} : +10.5 px, +11.0 px
- Optisches Zentrum rechte Kamera cc_{rx} , cc_{ry} : -10.8 px, +11.0 px
- Nutzbare Pixel: 1224 x 850
- Basisbreite B, Abstand der optischen Zentren: 0.2397 m

Bei beiden Kameras ist der Durchtritt der optischen Achse etwa 11 px von der Bildmitte, s. Bild 150, was für die Bildverarbeitung ohne Einfluss ist, allerdings in die Entfernungsgleichung (5.1) eingeht.

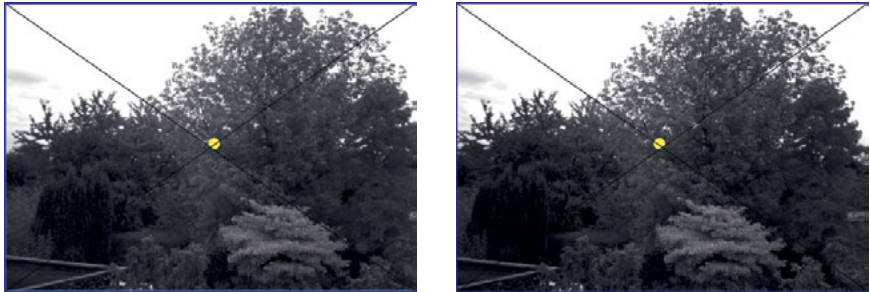


Bild 150: Lage der optischen Zentren (Kreise) im linken und rechten Stereobild

Die Disparität d_{px} bezieht sich auf die optischen Zentren. Damit wird der Abstand z errechnet zu:

$$z = \frac{Bf_{px}}{d_{px} + cc_{rx} - cc_{lx}} \quad (5.1)$$

5.4.2 Erfassungsbereiche für Geschwindigkeiten und Abstände

Mit den vorgenannten Daten und folgenden festgesetzten Daten lassen sich die Erfassungsbereiche für die Abstände und Geschwindigkeiten in Bezug zu einer Szene errechnen. Die Stereobasis sei dabei parallel zur Verbindungslinie Bug-Heck des Luftschiffs, denn dann stehen mit der Bildbreite in Flugrichtung mehr Pixel zur Verfügung, um Translationen zu berechnen.

- Die Bildrate sei 12/s.
- Bildbreite: 1224 px
- Die Subpixelgenauigkeit sei 0.3 px.
- Der maximal zugelassene Fehler sei 10 %.

Tabelle 15: Minimal, normal und maximal erfassbare Flughöhen mit Messfehlern

Fall für Abstand	Abstand	Disparität	Fehler bei 0.3 px Subpixelgenauigkeit
min. Abstand	0.62 m	538 px	0.06 %
normale Flughöhe	50 m	6.7 px	4 %
max. Abstand bei 10 % Fehler	102 m	3.3 px	10 %

Tabelle 16: Minimal und maximal detektierbare Geschwindigkeit bei verschiedenen Flughöhen und dazugehörige Messfehler

Fall	Grenzfall für Geschwindigkeit	Geschwindigkeit	Versatz	Fehler bei 0.3 px Subpixelgenauigkeit
1	min. Geschwindigkeit bei 1 m Flughöhe	0.09 km/h	3 px	10 %
2	min. Geschwindigkeit bei 50 m Flughöhe	4.6 km/h	3 px	10 %
3	max. Geschwindigkeit bei 1 m Flughöhe	17 km/h	560 px	0.1 %
4	max. Geschwindigkeit bei 50 m Flughöhe	866 km/h	560 px	5 %

Zu Fall 1:

Erst bei einem Versatz von 3 px für die Translation ergibt sich die geforderte Genauigkeit von 10 %. Der Messbereich für die minimal erfassbare Geschwindigkeit kann beliebig erweitert werden, indem für die Translationsberechnung ältere Bilder herangezogen werden. Der Fehler nimmt dann proportional mit der Zeitdifferenz zwischen den Aufnahmezeitpunkten ab.

Zu Fall 2:

Wie Fall 1.

Zu Fall 3:

Die maximal erfassbare Geschwindigkeit bei niedriger Flughöhe kann annähernd verdoppelt werden, indem der zu suchende Bildausschnitt nicht wie sonst üblich in der Mitte, sondern am Rand der Abbildung positioniert wird. Um nochmals 10 km/h wird der Messbereich erweitert, wenn linke und rechte Kamera zur Translationsberechnung herangezogen werden, so dass bei 1 m Abstand Geschwindigkeiten bis zu 45 km/h erfasst werden können.

Zu Fall 4:

Der Fehler von 5 % für die maximale Geschwindigkeit bei 50 m Flughöhe ist dominant durch die Unsicherheit bei der Flughöhenbestimmung begründet.

5.4.3 Prüfung der Kalibrierung und Rektifizierung über Abstandsberechnungen

Bei gelungener Kalibrierung und Rektifizierung muss sich bei der Abbildung eines sehr entfernten Objektes eine Disparität nahe null ergeben. Hierzu werden Objekte, deren Abstand zur Kamera bekannt ist, aufgenommen. Die Entfernungen wurden mittels topographischen Karten des Landes Nordrhein-Westfalen ermittelt.

Aus Tabelle 17 ist eine Aufstellung von Disparitätsberechnungen anhand von drei Objekten, die jeweils an verschiedenen Bildpositionen abgebildet sind, dargestellt.

Die Berechnungen an äußeren linken Bildrändern weisen Abweichung vom erwarteten Disparitätswert von etwa 0.5 px auf. Die Abweichung am rechten Bildrand ist mit 0.1 px kleiner. Ist das Objekt in der Bildmitte abgebildet, beträgt die maximale Abweichung 0.08 px, was eine sehr gute Präzision ist, zumal die Strukturen der Abbildungen verwischt wirken, wie aus Tabelle 17 , zu sehen ist.



















Dieser Versuch zeigt, dass die Kalibrierung und Rektifizierung der Stereokamera gut gelungen ist.

Die Randfehler sind kaum wirksam, da als Referenz immer ein Ausschnitt aus der Bildmitte genommen wird und bei großen Abständen der korrespondierende Bildausschnitt auch nahezu in der Bildmitte ist. Die Bildränder kommen erst bei kurzen Abständen oder großen Geschwindigkeiten, d. h. bei großer Disparität, zum Tragen. Z. B. ist bei 1000 mm Abstand die erwartete Disparität 335.83 px. Bei einem Disparitätsfehler von +0.4 px ergibt sich ein für das Fluggerät unbedeutender Abstandsfehler von 1.2 mm. Bei großen Geschwindigkeiten, also beispielsweise einer Verschiebung von 600 px, bewirkt ein Fehler von 0.4 px unabhängig vom Abstand einen Geschwindigkeitsfehler von 0.07 %.

Die Subpixelgenauigkeit ist im mittleren Bildbereich sehr gut und wird in Tabelle 18 mit für ein kleines UAV relevanten Abständen verifiziert.

Die Rektifizierung selbst, d. h. u. a. gleiche horizontale Epipolarlinien auf beiden Abbildungen, ist sehr gut, denn bis auf wenige Ausnahmen (bildstrukturbedingt) sind die berechneten vertikalen Abweichungen vom gesuchten Ausschnitt zum gefundenen Ausschnitt im Subpixelbereich. Zum Bestimmen der Abweichung sind bei 365 Bildpaaren die Disparitätsabweichungen in vertikaler Richtung ermittelt worden. Korrelationen bei Verdeckung oder homogenen Gebieten sind ausgeschlossen. Daraus ergibt sich, dass das rechte Bild um etwa 0.07 px in vertikaler Richtung nach unten versetzt ist. Deshalb ist die Suche entlang einer Bildzeile ausreichend, was den Korrelationsalgorithmus wesentlich schneller ablaufen lässt.

Tabelle 17: Disparität bei weit entfernten Objekten, die jeweils in der Mitte oder an den Rändern abgebildet sind

Objekt	Position im Bild	gesuchtes Teilbild	Abstand [m]	Abstand berechnet [m]	Disparität erwartet [px]	Disparität berechnet und (Abweichung) [px]
Gebäude links			1197	14958	0.28	0.14 (-0.14)
Gebäude mittig			1197	1466	0.28	0.23 (-0.05)
Gebäude rechts			1197	-13820	0.28	-0.02 (-0.30)
Kirche links			1093	7949	0.31	0.13 (-0.18)
Kirche mittig			1093	1533	0.31	0.23 (-0.08)
Kirche rechts			1093	-1316	0.31	-0.21 (-0.56)
Straßenleuchte links			419	549	0.80	0.69 (-0.11)
Straßenleuchte mittig			419	438	0.80	0.77 (-0.03)
Straßenleuchte rechts			419	1139	0.80	0.33 (-0.47)

Für einen weiteren Versuch zur Bestimmung der Genauigkeit bei der Abstandsberechnung ist eine Tafel mit einem großen Foto in Abständen zwischen 1 m und 25 m aufgestellt worden, s. Bild 151 und Bild 152.

Für 40 m und 65 m Abstand wurden Gebäudefassaden abgebildet, s. Bild 153 und Bild 154.

Die Abstände werden durch einen Laserentfernungsmesser mit 0.1 % Genauigkeit bestimmt.

In Tabelle 18 sind die erwarteten und die berechneten Abstände mit prozentualer Abweichung und der absoluten Abweichung der Disparität aufgeführt. Die Abweichung von 0.12 px bei einem Meter Abstand entspricht einem Messfehler von 360 μm .

Für Berechnungen der Abstände zwischen 1 m und 65 m ist die Fehlbestimmung bei diesen Messungen kleiner als 0.4 %. Die Disparität weicht nur um maximal 0.12 px vom erwarteten Wert ab.

Die Abstandsmessungsgenauigkeit ist damit für die Bestimmung der Flughöhe über Boden sehr gut.



Bild 151: Tafel bei 24 m,
rektifiziertes Bild



Bild 152: Tafel bei 2 m,
Bild vor der Rektifizierung



Bild 153: Gebäudefassade bei 40 m



Bild 154: Gebäudefassade bei 65 m

Tabelle 18: Erwartete und berechnete Abstände

Abstand, erwarteter Wert	Abstand, berechneter Wert	Abstand, Abweichung	Abweichung der Disparität vom erwarteten Wert
0.984 m	0.984 m	-0.04 %	0.12 px
1.974 m	1.973 m	-0.01 %	0.02 px
2.976 m	2.975 m	-0.03 %	0.04 px
4.969 m	4.972 m	0.06 %	-0.04 px
7.982 m	7.969 m	-0.17 %	0.07 px
9.995 m	9.968 m	-0.27 %	0.09 px
14.97 m	14.95 m	-0.13 %	0.03 px
19.97 m	19.98 m	0.06 %	-0.01 px
25.01 m	24.98 m	-0.10 %	0.01 px
40.01 m	39.86 m	-0.38 %	0.03 px
64.96 m	64.84 m	-0.18 %	0.01 px

Die Disparitäten in Tabelle 18 werden mit der 'inversen Parabel' berechnet. Ein Vergleich mit der regulären Parabel lieferte nur sehr gering – im Tausendstel Pixelbereich – bessere Werte. Daraus kann geschlossen werden, dass die Anwendung der 'inversen Parabel' bei realen Stereoaufnahmen nur bei der Verkleinerung großer Bildbereiche eine deutliche Verbesserung ergibt.

5.4.4 Prüfung der Bewegungsberechnung

Die Abstandsberechnung wird mit Aufnahmen, die gleichzeitig von der linken und rechten Kamera erstellt werden, ermittelt und liefert gute Ergebnisse. Voraussetzung für die Bewegungsrechnung ist die Abstandsberechnung, da damit aus nacheinander aufgenommenen Bildern von den Translationen in Pixel auf Translationen in metrischen Längen geschlossen wird. Die Fehler der Abstandsberechnung gehen somit in die Fehler der Translation ein.

Die Strecken werden über Koppelnavigation ermittelt, indem die über visuelle Odometrie errechneten inkrementellen Bewegungsinformationen integriert werden.

Um die Gesamtgenauigkeit zu ermitteln, werden Messungen mit drei Szenen, deren Dimensionen bekannt sind, durchgeführt.

Erste Szene

Mit einem Laborwagen und aufmontierter Kamera wird entlang einer Tafel (Bild 155) mit verschiedenen Abständen zur Tafel einmal hin- und hergefahren, so dass die letzte Position nahe bei der ersten Position ist. Nach den Berechnungen für 135 sequentielle und Stereobildpaare, wobei die Translationsschritte entsprechend der Bewegungsrichtung addiert oder subtrahiert werden, müsste das Ergebnis nahe null sein.

Die Positionen der gefundenen kleinen Rechtecke in Bild 156 und Bild 157 werden dem Prüfungsobjekt zugeordnet. Die erwarteten Abstände in x- und y-Richtung werden aus den Zuordnungen ermittelt. Die Positionsabweichungen zwischen den erwarteten Abständen und berechneten Abständen sind in Tabelle 19 aufgeführt.

Nach 2.481 m zurückgelegtem Weg ist der Fehler für die horizontale Richtung 10 mm, entsprechend 0.4 %.



Bild 155: Objekt zur Prüfung der Kamerabewegung

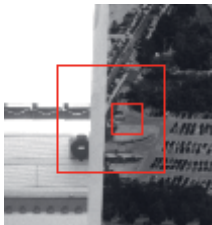


Bild 156: Erste gefundene Position
(kleines Quadrat)

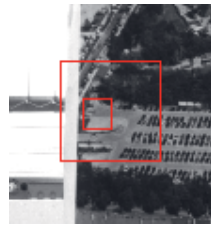


Bild 157: Letzte gefundene Position
(kleines Quadrat)

Tabelle 19: Erwartete Positionsdivergenz und Positionsdivergenz aus Berechnung zwischen erstem und letztem gefundenen Teilbild bei 2.481 m Wegstrecke

	Erwartete Positionsdivergenz	Positionsdivergenz aus Berechnung	Fehler	Fehler bezogen auf 2.481 m
Horizontal	3 mm	13 mm	10 mm	0.4 %
Vertikal	-2 mm	4 mm	6 mm	-

Eine zweite Versuchsfahrt führt zu ähnlich genauen Ergebnissen. Das kleine Quadrat in Bild 159, welches sich an der letzten gefundenen Position befindet, ist gegenüber dem Quadrat der ersten gefundenen Position, Bild 158, nach rechts um 9 mm und etwas nach oben um 3 mm verschoben. Damit stimmen die kumulierten Werte der Positionsdifferenzen aus den Berechnungen mit der wahren Strecke fast überein, siehe Tabelle 20. Somit ist der Fehler bei dem zurückgelegten Weg von 2.672 m etwa 0.07 %.

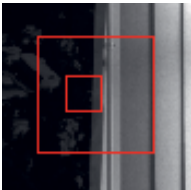


Bild 158: Erste gefundene Position (kleines Quadrat)

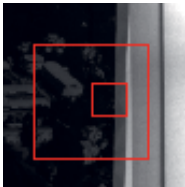


Bild 159: Letzte gefundene Position (kleines Quadrat)

Tabelle 20: Erwartete Positionsdifferenz und Positionsdifferenz aus Berechnung zwischen erstem und letztem gefundenen Teilbild bei 2.672 m Wegstrecke

	Erwartete Positionsdifferenz	Positionsdifferenz aus Berechnung	Fehler	Fehler bezogen auf 2.672 m
Horizontal	9 mm	7 mm	2 mm	0.07 %
Vertikal	3 mm	2 mm	1 mm	-

Zweite Szene

Ein Außenversuch bei größeren Abständen, größeren Strecken und schlecht strukturierten Bildern führt zu etwas größeren Fehlern. Die Kamera ist hier auf einem Stativ montiert und rotiert von links nach rechts mit Blickrichtung entlang einer Hauswand, s. Bild 160 und Bild 161.



Bild 160: Hauswand zur Berechnung einer Strecke, Sicht aus einer Kameraorientierung

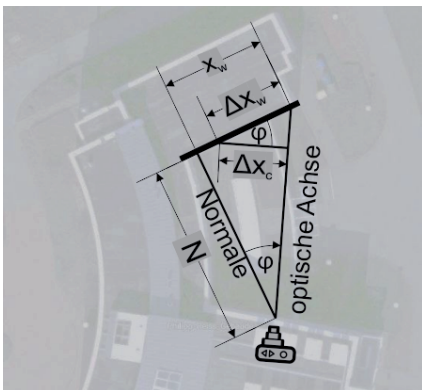


Bild 161: Luftbild mit bezeichneter Hauswand und Kameraposition
(Quelle des unterlegten Luftbilds: Google maps)

Während der Kamerarotation weichen die projizierten Streckenabschnitte Δx_c zunehmend mit dem Blickwinkel φ der Kamera vom gesuchten wahren Streckenabschnitt Δx_w ab (s. Bild 162), deshalb gehen die Blickwinkel in die Streckenberechnung ein:

$$\varphi = \arctan\left(\frac{x_w}{N}\right)$$

$$\Delta x_w = \frac{\Delta x_c}{\cos \varphi} = \Delta x_c \sqrt{1 + \left(\frac{x_w}{N}\right)^2} \quad (5.2)$$



x_w : Kumulierte Strecke

Δx_w : Streckenabschnitt auf der Wand zwischen zwei aufeinanderfolgenden Bildern

Δx_c : Projizierter, von der Kamera wahrgenommener, Streckenabschnitt

φ : Blickwinkel der Kamera, Bezug ist die Normale

N : Länge der Normalen

Bild 162: Trigonometrische Verhältnisse beim Blick auf die Hauswand

Der Abstand N der Normalen zur Kamera wird mit der Stereokamera bestimmt. Der Fußpunkt der Normalen ist am ersten unteren Fenster von Bild 160. Die Strecke x_w auf der Wand des Gebäudes wird sukzessive ermittelt. Ausgehend vom Fußpunkt der Normalen, $\varphi = 0$, wird schrittweise die Strecke x_w mit Gleichungen (5.2) ermittelt, wobei das neue x_w jeweils das vorhergehende $x_w + \Delta x_w$ ist.

Die realen Positionen werden über die Fensterabstände, die genau 1.20 m betragen, bestimmt. Die linke untere Ecke des Quadrates des ersten Teilbildes, Bild 163, ist 0.9 m von der linken Kante des ersten Fensters entfernt und 0.3 m über dem unteren Fenster. Die linke untere Ecke des Quadrates des letzten Bildes ist 0.8 m von der linken Kante des letzten Fensters entfernt und 0.27 m unter dem Fenster.

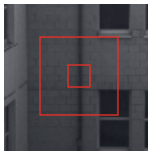


Bild 163: Position des ersten gefundenen Teilbildes (kleines Quadrat)



Bild 164: Position des letzten gefundenen Teilbildes (kleines Quadrat)

Die gesamte wahre Strecke ergibt sich zu 12.5 m, während die berechnete Strecke 12.4 m beträgt, s. Bild 165. Dies entspricht einem Fehler von etwa 0.8 %.

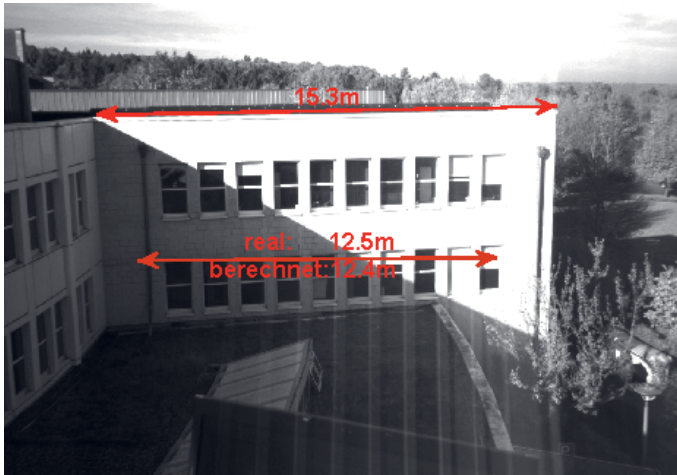


Bild 165: Hauswand mit Maßen zur Berechnung von Strecken

Die teilweise überbelichteten Bildbereiche, die die Bildstruktur reduzierten, siehe Bild 167, führten zu schlechten Korrelationen und damit zu Sprüngen im Abstandsverlauf, siehe Bild 166, denn bei den Abständen um 27 m ergibt ein Disparitätsunterschied von 0.5 px schon eine Abstandsabweichung von 1 m.

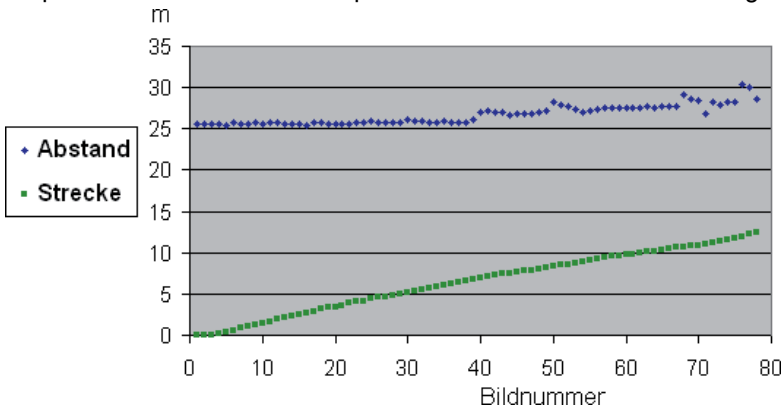
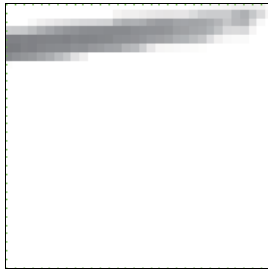
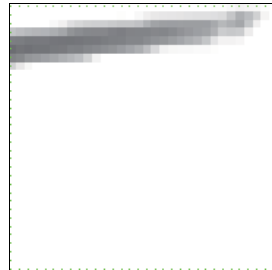


Bild 166: Berechnete Abstände der Hauswand zur Kamera und Streckenverlauf



gesuchtes Teilbild



gefundenes Teilbild

Bild 167: Teilbilder der linken und rechten Kamera nahe dem letzten Stereobildpaar

Dritte Szene

Ein weiterer Versuch mit einer beschatteten Fassade und längerer Gesamtstrecke, Bild 168, führt zu ähnlichen Ergebnissen. Die Kamera rotiert wie beim vorhergehenden Beispiel, allerdings hin und her. Es gelten dieselben Winkelgleichungen. Nachdem die Fassade 5-mal von rechts nach links oder umgekehrt überstrichen wurde, ergab sich eine Gesamtstrecke von 69.60 m. Berechnet wurden 70.79 m, was einem Fehler von 1.7 % entspricht. Jeweils beim Rücklauf, Bildnummer 270 und 403 der Tabelle 21, in die Nähe des Startpunkts, Bildnummer 1, stimmt die wahre Position mit der berechneten Position weitgehend überein.

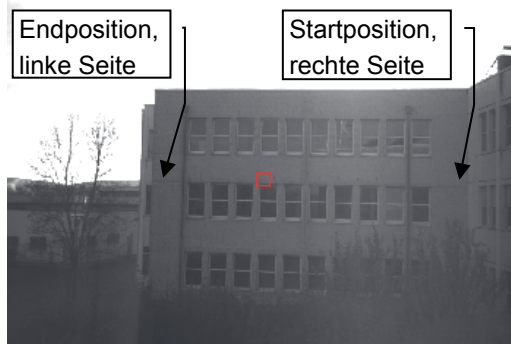
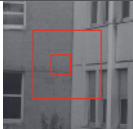
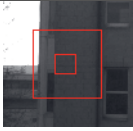
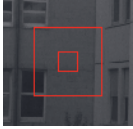
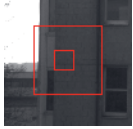
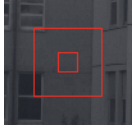
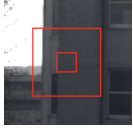


Bild 168: Gebäudefassade zur Abstands- und Streckenberechnung

Tabelle 21: Wahre und berechnete Position und Strecken beim mehrfachen Überstreichen einer Fassade

Position, Bild- nummer		Wahre Position	Berechnete Position	Wahre Strecke vom Nullpunkt, und (von vorhergehender Position)	Berechnete Strecke vom Nullpunkt, und (von vorhergehender Position)
Rechte Seite, 1		0 m	-	-	-
Linke Seite, 104		-14.29	-14.47	14.29 (14.29)	14.47 (14.47)
Rechte Seite, 270		-0.45	-0.53	28.13 (13.84)	28.41 (13.94)
Linke Seite, 347		-14.34	-14.54	42.01 (13.88)	42.43 (14.02)
Rechte Seite, 403		-0.56	-0.29	55.78 (13.77)	56.68 (14.25)
Linke Seite 469		-14.38	-14.41	69.60 (13.82)	70.79 (14.11)

5.4.5 Verfahren für schnelle Berechnung bei realitätsnahen Versuchen

Die bisherigen Versuche waren mit überwiegend flachen Szenen, die auch keine Verdeckung beinhalteten und mit mäßigen Geschwindigkeiten bei weitgehend konstanten Abständen durchgeführt worden, um die Genauigkeit

der Verfahren nachzuweisen. Die angestrebte Echtzeitverarbeitung und die Verdeckungsprobleme erfordern Erweiterungen und Vereinfachung der Verfahren, die in diesem Unterkapitel erläutert werden.

5.4.5.1 Ausgleich des Intensitätsverlaufs

Wie schon in Kap. 2.4.1.1 (Vignettierung) gezeigt, hat die Bumblebeekamera einen stark gekrümmten Intensitätsverlauf.

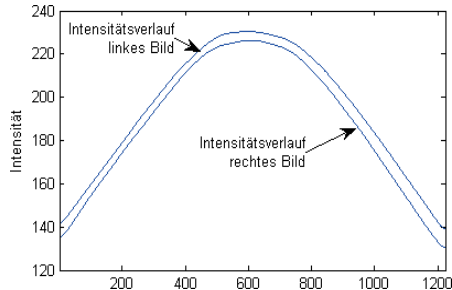


Bild 169: Intensitätsverlauf für linke und rechte Kamera

Versuche innerhalb des 8-Bit-Zahlenraumes, die Intensitäten über eine Tabelle auszugleichen, bringen zwar eine deutliche Verbesserung der Bildregistrierung, aber die Kontrastverminderung in der wichtigen Bildmitte führt zu ungenaueren Ergebnissen bei der Bestimmung großer Abstände mit der Stereokamera. Diese Methode wird deshalb nicht benutzt.

Um die Einflüsse der multiplikativen und additiven Intensitätsunterschiede der linken und rechten Kamera auszugleichen, wird für die weiteren Versuche nur die ZNCC zur Berechnung der Bewegung und des Abstandes benutzt.

5.4.5.2 Vereinfachung der bilinearen Interpolation

Nächster-Nachbar-Interpolation

Ein großer Rechenaufwand ist die für die Rektifizierung erforderliche bilineare Interpolation, die für jeden Bildpunkt vier Multiplikationen und vier Additionen verlangt. Eine drastische Vereinfachung ergibt sich durch die Nächster-Nachbarn-Interpolation, die jedem rektifizierten Bildpunkt einen Bildpunkt aus dem Originalbild zuordnet. Zur ersten Verifizierung wird der Versuch aus Kap. 5.4.3, Tabelle 19 mit der Methode des Nächsten-Nachbarn wiederholt. Der Zeitvorteil ist erheblich, denn in Matlab® wird damit nur 1/5 der Rechenzeit benötigt.

Wie unten stehende Tabelle 22 zunächst vermuten lässt, sind die Abweichungen geringfügig, denn die mittlere Abweichung der absoluten

Disparitäten ist nur 0.10 px. Dies gilt allerdings nur bei gut strukturierten Abbildungen.

Ist die Vorlage mit Bereichen schwacher Struktur, kommt es mit der Nächster-Nachbarn-Interpolation zu erheblichen Fehlberechnungen.

Bei 26 von 469 Korrelationen waren die Abweichungen größer als ein Pixel. Damit ist die Nächster-Nachbar-Interpolation zu ungenau.

Tabelle 22: Abweichungen der Abstände und Disparitäten bei Rektifizierung mit Nächster-Nachbar-Interpolation im Vergleich zur bilinearen Interpolation

Abstand, erwarteter Wert	Abstandsabweichung		Abweichung der Disparität vom erwarteten Wert	
	bei Nächster- Nachbar- Interpolation	bei bilinearer Interpolation	bei Nächster- Nachbar- Interpolation	bei bilinearer Interpolation
0.984 m	-0.08 %	-0.04 %	0.28 px	0.12 px
1.974 m	0.07 %	-0.01 %	-0.12 px	0.02 px
2.976 m	-0.09 %	-0.03 %	0.10 px	0.04 px
4.969 m	-0.08 %	0.06 %	0.05 px	-0.04 px
7.982 m	-0.16 %	-0.17 %	0.07 px	0.07 px
9.995 m	-0.37 %	-0.27 %	0.13 px	0.09 px
14.97 m	-0.40 %	-0.13 %	0.09 px	0.03 px
19.97 m	0.15 %	0.06 %	-0.03 px	-0.01 px
25.01 m	0.28 %	-0.10 %	-0.04 px	0.01 px
40.01 m	2.07 %	-0.38 %	-0.17 px	0.03 px
64.96 m	0.32 %	-0.18 %	-0.02 px	0.01 px
Mittlerer Fehler der Absolutwerte	0.4 %	0.1 %	0.10 px	0.04 px

5.4.5.3 Verringerung der Präzision der Bilinearkoeffizienten

Eine weitere Vereinfachung ist, die Präzision der Bilinearparameter für die Rektifizierung auf 7 Bit Genauigkeit zu reduzieren. Dies hat in der Hardware den Vorteil, dass weniger Daten transportiert werden müssen und je nach

System Ganzzahloperationen schneller ablaufen. Die sich ergebenden Abweichungen der Intensitätswerte der rektifizierten Bildpunkte zur Gleitkommagenauigkeit werden hier nochmal gezeigt, s. Bild 170.

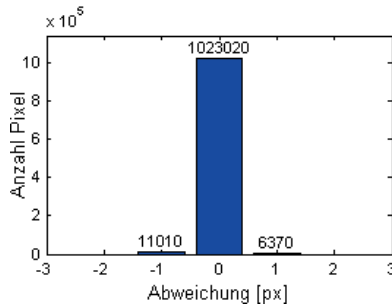


Bild 170: Pixelfehler bei 7 Bit Genauigkeit für die Interpolation im Vergleich zu Gleitkommagenauigkeit

Bei 1600 Verschiebungen in x- und y-Richtung und bei gleich vielen Disparitäten ist – bis auf eine Ausnahme – die maximale absolute Abweichung von der erwarteten Verschiebung oder Disparität, die sich durch die bilineare Interpolation bei voller Präzision im Vergleich zur 7 Bit bilinearen Interpolation ergibt, immer kleiner als 0.006 px und damit vernachlässigbar. Die Ausnahme tritt bei einem fast homogenen Bereich auf, und hier hat zufälligerweise die 7 Bit bilineare Interpolation das richtige Ergebnis geliefert.

5.4.5.4 Reduzierung der Rektifikation auf benutzte Bereiche

Für die Translationsbestimmung wird immer das linke Bild der Stereokamera verwendet. Hier werden nur die Pixel rektifiziert, die in ein maximal großes Rechteck passen, das nur Bildpunkte enthält und damit auch ausschließlich die dafür benötigten Bilinearparameter herangezogen, s. Bild 171.

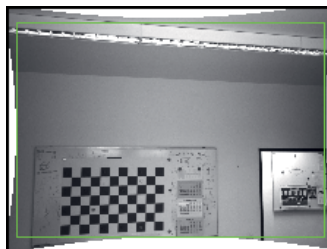


Bild 171: Vollständig rektifiziertes Bild, genutzte Pixel sind im inneren Rechteck

Bei der Abstandsberechnung müssen nur die Bereiche rektifiziert werden, die auch benutzt werden. Da ein Ausschnitt aus dem linken Bild, s. Bild 172, im rechten Bild gesucht wird und das rechte Bild nur zur Abstandsberechnung verwendet wird, muss wegen der gleichen Epipolarlinien nur ein Streifen rektifiziert werden, s. Bild 173.



Bild 172: Linkes Stereobild, Suchfenster

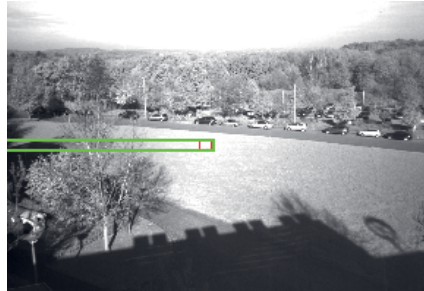


Bild 173: Rechtes Stereobild, zu durchsuchender Streifen

5.4.5.5 Suchbereich für die Bewegungserfassung einschränken

Das Suchen eines Teilbildes im ganzen vorhergehenden Bild ist sinnvoll bei großen Bewegungen (hohe Geschwindigkeiten), würde allerdings bei voller Auflösung des Bildes einen zu großen Rechenaufwand erfordern. Deshalb werden im Folgenden einige Strategien vorgestellt.

Bei geringen Bewegungen wird ein Teilbild (z. B. 32 x 32 px, s. Bild 174) in einem Fenster – großes Quadrat (z. B. 128 x 128 px) in Bild 175 – gesucht. Zusätzlich wird dieses Fenster entsprechend der vorhergehenden Bewegung positioniert, deutlich zu sehen in Bild 175, denn das Fenster ist nicht mehr in der Mitte des Bildes, sondern etwas weiter rechts. Die Kamera bewegte sich in diesem Beispiel von links nach rechts. Das vorhergehende Bild ist in Bild 174 dargestellt.

Die Vorhersage ist in diesem Beispiel sehr gut, denn das gefundene Teilbild, kleines Quadrat in Bild 175, befindet sich recht genau in der Mitte des zu durchsuchenden Bereichs, großes Quadrat in Bild 175.

Bild 174:
Vorhergehendes Bild
mit zu suchendem Teilbild

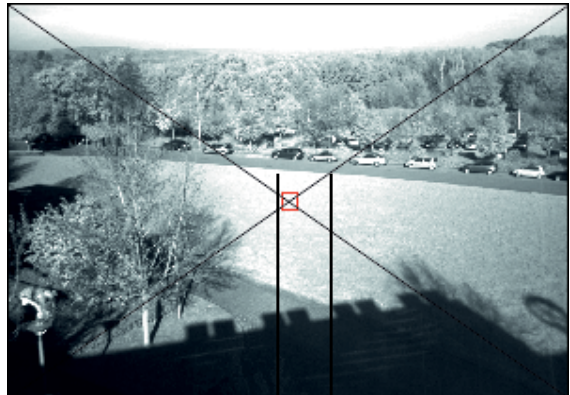
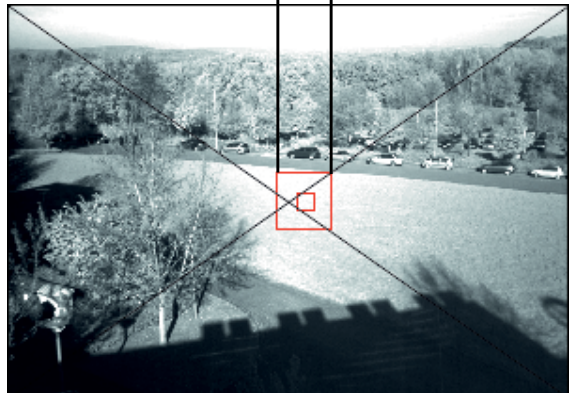


Bild 175:
Aktuelles Bild
mit zu durchsuchendem
Fenster (großes Quadrat)
und gefundenem Teilbild
(kleines Quadrat)



Nachführung des zu durchsuchenden Fensters,
Bewegung der Kamera von links nach rechts

5.4.5.6 Bildverkleinerung

Bei der Verfolgung der vermuteten neuen Position sind einige Sonderfälle zu beachten:

- **Start**

Am Anfang ist die vorhergehende Bewegung unbekannt. Deshalb muss im ganzen Bild gesucht werden. Um einen immensen Rechenaufwand zu vermeiden, werden das aktuelle und das vorhergehende Bild auf 1/8 Kantenlänge verkleinert (s. Bild 178) und der Suchprozess wird zunächst trotz geringerer Auflösung im gesamten Bereich dieser Bilder durchgeführt.

- **Große Geschwindigkeit**

Die Geschwindigkeit oder die Beschleunigung ist so groß, dass die Gefahr, das Teilbild zu verlieren, deutlich wächst. Dann werden auf 1/4 oder 1/8, je nach Größe der Geschwindigkeit, verkleinerte Bilder genommen. In den auf 1/4 verkleinerten Bildern (Bild 177) werden für das zu suchende Teilbild und das zu durchsuchende Fenster die gleichen Kantenlängen wie im Bild der vollen Auflösung genommen. Ein Teilbild mit einer Kantenlänge von 32 px in einem auf ein Viertel verkleinertem Bild (Bild 177) hat bei gleicher Position den gleichen Bildinhalt wie ein Teilbild der Kantenlänge von 128 px im unverkleinerten Bild (Bild 176), allerdings ist die Auflösung verschieden.

Die 1/8 großen Bilder (Bild 178) werden immer komplett mit einem 20 x 20 px großen Teilbild durchsucht.

- **Geringe Flughöhe**

Wenn der Abstand – die Flughöhe des UAS – über dem Boden gering ist, ergeben kleine Bewegungen schon große Verschiebungen in der Abbildung. Es werden unter 15 m Flughöhe 1/4 große Bilder verwendet und unter 6 m Flughöhe werden 1/8 große Bilder benutzt.

- **Homogener Bildbereich**

Ist das zu suchende Teilbild – z. B. durch Überbelichtung oder Unterbelichtung – homogen, wird die vorhergehende Berechnung herangezogen.

Die Homogenität kann durch Addition der absoluten Differenzen benachbarter Pixel des Teilbildes ermittelt werden. Wenn die Summe nahe null ist, gilt das Teilbild als homogen.

Bei Aufnahmen von überflogenen Gebieten entstehen mit den verwendeten Kameras homogene Bildbereiche nur durch Unter- oder Überbelichtung. Deshalb wird in dieser Arbeit nur der Intensitätsmittelwert des Teilbildes, der ohnehin für die Korrelation berechnet werden muss, herangezogen. Ist dieser Mittelwert null, so ist das Bild unterbelichtet. Bei einem Wert von 255 ist das Teilbild überbelichtet.

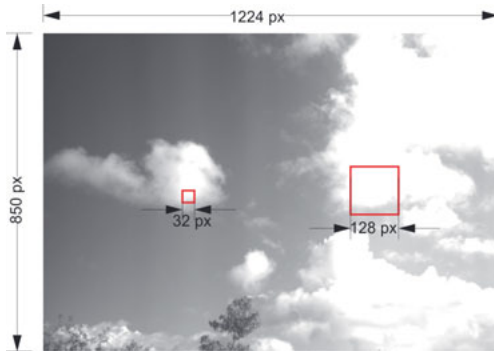


Bild 176: Volle Bildgröße

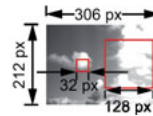


Bild 177:
1/4 Größe

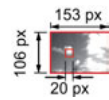


Bild 178:
1/8 Größe

Bild 176 bis Bild 178: Verschiedene Bildauflösungen zur Berechnung der Verschiebungen und Disparitäten in Abhängigkeit von der Translationsgeschwindigkeit und des Abstandes.

Die kleinen Quadrate stellen die Größe für das zu suchende Teilbild dar, während die großen Rahmen die Maße des zu durchsuchenden Fensters wiedergeben.

Die Verkleinerung führt zu größeren Fehlern, die aber bei hohen Geschwindigkeiten prozentual nicht sehr ins Gewicht fallen. Beim Vergleich der Verschiebungen in Pixel bei Verkleinerung der Bilder ergeben sich Werte entsprechend Tabelle 23. Hierbei wird zur Berechnung der Translation bei 1/4 Bildgröße die errechnete Verschiebung in Pixel mit 4 multipliziert und bei 1/8 Bildgröße wird die errechnete Verschiebung in Pixel mit 8 multipliziert.

Tabelle 23: Absolute Fehler der Verschiebungsberechnung bei Verkleinerung der Bilder auf 1/4 und 1/8 Bildgröße

	1/4 Bildgröße Differenz zur vollen Bildgröße	1/8 Bildgröße Differenz zur vollen Bildgröße
Maximale Differenz	0.99 px	1.91 px
Mittelwert der absoluten Differenzen	0.2 px	0.4 px

Wegen des Berechnungsaufwands wird die erste Verkleinerung auf 1/2 Bildgröße mit einem kleinen 3 x 3 Gaußfilter durchgeführt, während die

weiteren Verkleinerungen bis zu 1/8 Bildgröße mit jeweils einem 5 x 5 Gaußfilter vollzogen werden.

Die Berechnungszeit der Kreuzkorrelation ist bei voller und 1/4 Bildgröße gleich, bei ein 1/8 Bildgröße etwas kürzer.

5.4.5.7 Nachbearbeitung: Ausschließen deutlich falsch berechneter Werte

Teilweise durch Verdeckung und seltener durch falsche Korrelationsmaxima sind errechnete Geschwindigkeiten falsch. Diese Geschwindigkeitsabweichungen werden über maximal real vorkommenden Geschwindigkeitszu- oder -abnahmen, die die Dynamik des Fluggerätes zulässt, detektiert.

Die maximal auftretende Beschleunigung a des Blimps durch eine Windböe beträgt 2 m/s^2 . Bei einer Bildrate von $12/\text{s}$ ist das Zeitintervall $\Delta t = 1/12 \text{ s}$. Aus Gleichung (5.3) ergibt sich, dass sich in der Zeit zwischen zwei Bildaufnahmen die Geschwindigkeit des verwendeten Fluggerätes um maximal 0.6 km/h ändert.

$$\Delta v_{\max} = a \Delta t = \frac{2 \text{ m}}{\text{s}^2} \frac{1}{12} \text{ s} = 0.6 \text{ km/h} \quad (5.3)$$

Überschreiten die mit der Bildverarbeitung ermittelten Geschwindigkeitsänderungen den Wert Δv_{\max} , werden vorhergehende Berechnungen herangezogen und die aktuelle Geschwindigkeitsberechnung nach folgender Formel (digitales Tiefpassfilter erster Ordnung) angeglichen:

$$v_f(t) = v_f(t-1) \cdot 0.97 + v(t) \cdot 0.03 \quad (5.4)$$

mit:

Momentane gefilterte Geschwindigkeit: $v_f(t)$

Vorhergehende gefilterte Geschwindigkeit: $v_f(t-1)$

Momentane errechnete Geschwindigkeit: $v(t)$

Liegt keine Geschwindigkeitszu- oder -abnahme von über 0.6 km/h vor, wird die aktuelle Geschwindigkeit übernommen:

$$v_f(t) = v(t) \quad (5.5)$$

Der Einfachheit halber könnte anstatt des Tiefpassfilters, Gleichung (5.4), direkt der alte Filterwert $v_f(t-1)$ genommen werden. Dies führt unter bestimmten Umständen dazu, dass dieser Wert nie verlassen wird und die wahre Geschwindigkeit nicht mehr erfasst wird, insbesondere dann, wenn erste berechnete Werte aufgrund zu geringen Abstandes der Kamera zur betrachteten Szene außerhalb des Erfassungsbereichs der Disparität liegt und damit falsche Abstände berechnet werden. Mit Gleichung (5.4) wird sichergestellt, dass die Berechnung zum wahren Wert der Geschwindigkeit führt. Der Tiefpassfilterkoeffizient von 0.03 ist ein Kompromiss zwischen schnellem Einschwingverhalten und Fehlbestimmung der Geschwindigkeit. Die Sprungantwort des Filters führt nach 75 Bildern zu einer 10 % Abweichung und nach 151 Bildern zu einer 1 % Abweichung.

In Bild 179 ist der stark vereinfachte Ablauf der Bildverarbeitung zur Abstands- und Streckenberechnung gezeigt.

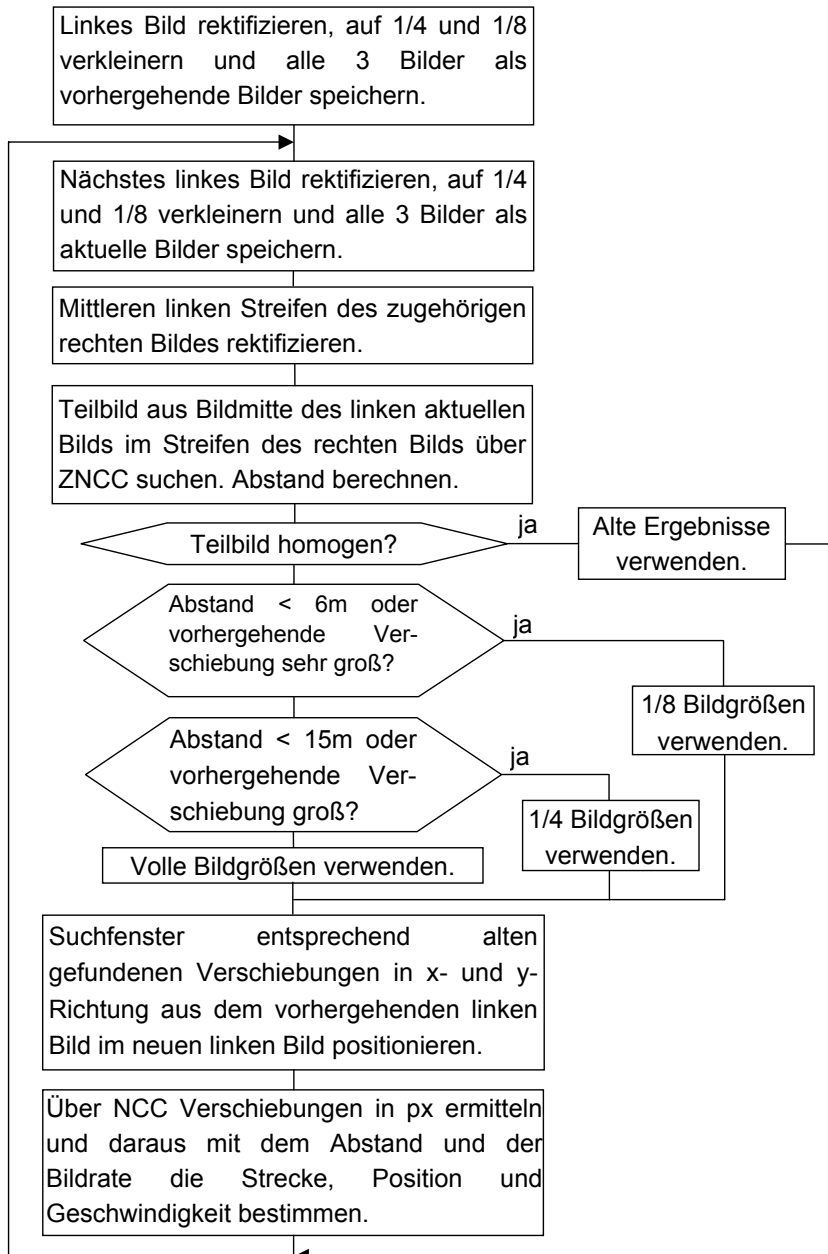


Bild 179: Vereinfachtes Flussdiagramm der Bildverarbeitung

5.4.6 Realitätsnahe Versuche

Die im Folgenden gezeigten Versuche sind mit komplizierteren Szenen durchgeführt worden, als sie bei Flugaufnahmen vorliegen würden, da sowohl Gebäude mit davor parkenden Fahrzeugen und auch die Vegetation überwiegend vertikale Strukturen aufweisen, die durch die Nähe und die seitliche Ansicht Verdeckungen wesentlich häufiger als bei einer Ansicht von oben auftreten lassen. Die Aufnahmen wurden mit einer auf einem Fahrzeug seitlich montierten Kamera erstellt.

Die Genauigkeit des Verfahrens kann über die berechnete Geschwindigkeit geschätzt werden. Gibt es in der Geschwindigkeit Sprünge, so ist vermutlich eine Verdeckung vorhanden oder die Korrelation führte nicht zum erwarteten Ergebnis.

Einige Beispiele zeigen Szenen, die zu Fehlinterpretationen führen. Hierzu gehören Verdeckung und periodische Bildausschnitte.

Mit Bild 180 und Bild 181 ist ein Extremfall gezeigt. Dort sehen linke Kellerfenster und das Kellerfenster daneben gleich aus und zusätzlich verdeckt ein Baumstamm einen Teil der Szene. In diesem Fall war die errechnete Korrelation mit dem falschen Fenster mit 0.99 sehr hoch, so dass auch hierüber keine richtige Diskriminierung möglich wäre.



Bild 180: Translation, zu suchender Bildausschnitt bei Verdeckung



Bild 181: Gefundener Bildausschnitt:
linkes kleines Quadrat
richtige Position: rechtes kleines Quadrat

Ein anderer Fall ist, wenn die Translation, Bild 182 und Bild 183, auf einer anderen Ebene, als es dem errechneten Abstand entspricht, bestimmt wird. In Bild 184 und Bild 185 wird für die Abstandsbestimmung der Laternenmast mit einbezogen und dessen Abstand zu 11.5 m berechnet, während für die

Translation nur die Hauswand im Abstand von 23.2 m verwendet wird. Erschwerend ist zusätzlich die dominierende gerade Kante des Kellerbereichs. Bei der darauffolgenden Aufnahme werden die Abstandsbestimmung und die Translationsberechnung durch den Masten vorgegeben.



Bild 182: Translation, zu suchender Bildausschnitt bei Verdeckung, aktuelles Bild

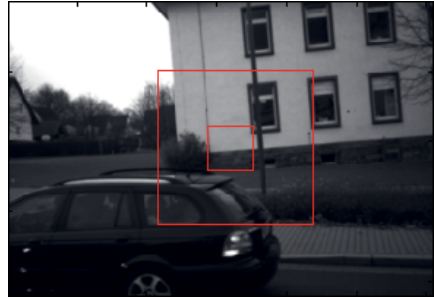


Bild 183: Gefundener Bildausschnitt: Gebäudeteil ohne Laternenmast, vorhergehendes Bild



Bild 184: Linkes Stereobild mit zu suchendem Fenster, (identisch mit Bild 182)



Bild 185: Rechtes Stereobild mit gefundenem Bildausschnitt einschließlich Mast

Die Entfernungsbestimmung wird erschwert, wenn die betrachtete Szene nahezu parallel zur optischen Achse ist und zusätzlich, wie in Bild 187 zu sehen ist, nur im oberen Teil des zu suchenden quadratischen Fensters eine dominante Struktur vorhanden ist. Zur Verdeckungsdetektion werden, wie in Kap. 3.2.4 dargestellt, ein kleiner und ein großer Bildanteil zur Abstandsbestimmung herangezogen, s. kleines und großes Quadrat in Bild 187. Im vorhergehenden linken Stereobild, Bild 186, ist noch etwas

Struktur im linken mittleren Teil des Quadrats zu erkennen, während in Bild 187 nur im oberen Teil des zu suchenden großen quadratischen Fensters eine dominante Struktur vorhanden ist. Dies führt dazu, dass im vorhergehenden Fall, Bild 186, der Abstand mit 13.1 m bestimmt wird und im darauf folgenden Bildpaar der Abstand für das große Quadrat zu 16.6 m und für das kleine Quadrat zu 13 m errechnet wird. Der Widerspruch bei den Abstandsberechnungen im selben Bild wird als Verdeckung detektiert, die allerdings nicht vorhanden ist.



Bild 186: Szene weitgehend parallel zur optischen Achse, mit Struktur im Bildmittelteil des zu suchenden Fensters

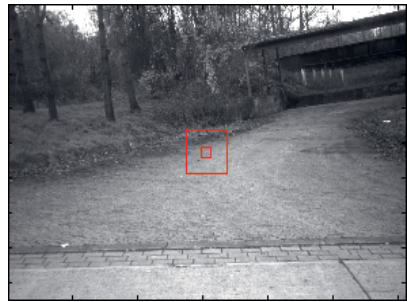


Bild 187: Szene weitgehend parallel zur optischen Achse, wenig Struktur im Bildmittelteil des zu suchenden Fensters

Fahrt entlang eines Gebäudes mit einigen Verdeckungen

Die Aufnahmen sind während einer Fahrt mit etwa 20 km/h entlang eines Gebäudes, welches durch 13 Bäume, Fahnenmaste oder Laternen in kleinen Abschnitten verdeckt ist, erstellt worden. Das Bild 188 gibt den Verlauf abschnittsweise wieder.



Abbildung Nr. 12



Abbildung Nr. 20



Abbildung Nr. 52



Abbildung Nr. 72



Abbildung Nr. 92



Abbildung Nr. 112



Abbildung Nr. 132



Abbildung Nr. 152

Bild 188: Bildausschnitte aus einer Fahrt entlang eines Gebäudes mit einigen Verdeckungen und Angabe der Bildnummer innerhalb der Aufnahmesequenz

In Bild 189 sind die berechneten Verläufe der Abstände und Geschwindigkeiten wiedergegeben. Die Hausfassade ist jeweils etwa 22 m von der Kamera entfernt, während die Fahnenmasten 15 m, die Bäume jeweils 9 m und die beiden Laternen unterschiedlich entfernt sind. Da die Bildrate mit 12 Bildern pro Sekunde konstant ist, gibt die Bildnummer auch proportional die Zeit wieder.

Der Verlauf der Geschwindigkeit weist einige Sprünge auf. Bis auf den Sprung bei Bildnummer 14 – eine falsche Korrelation entlang der Kellerkante – sind die Abweichungen durch Verdeckungen bedingt, da dort die Translationen entweder auf anderen Ebenen als den berechneten Abständen bestimmt werden oder die Verdeckung zu große Bildunterschiede erzeugt, um richtige Korrelationsergebnisse zu erlangen.

Die Welligkeit der Geschwindigkeitskurve mit einer Periodendauer von etwa 417 ms kann eventuell aus Vibrationen des Fahrzeugs erklärt werden.

Die Gebäudebreite ist 65.7 m. Werden alle berechneten Translationen von der linken bis zur rechten Hauskante addiert, ergibt sich eine Strecke von 66.2 m. Dies entspricht einem Fehler von 0.75 %.

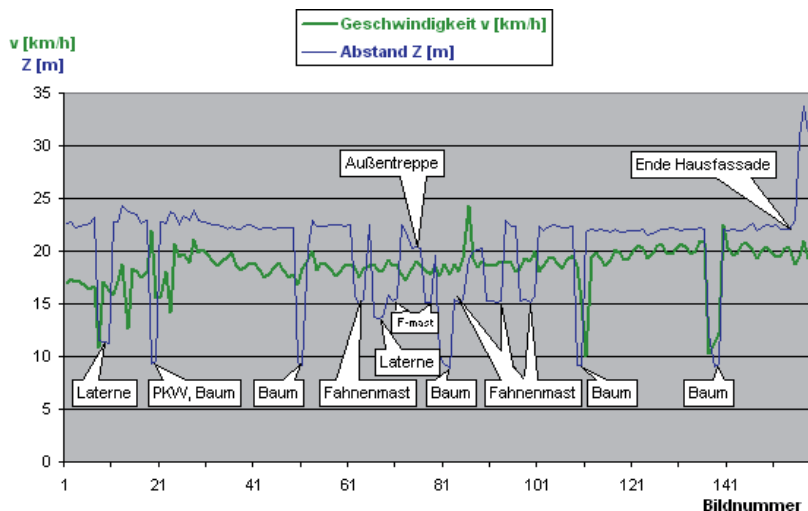


Bild 189: Berechnete Geschwindigkeiten und Abstände mit Bezeichnung der Verdeckungen bei einer Fahrt entlang eines Gebäudes mit einigen Verdeckungen

Bild 190 gibt die nach den Gleichungen (5.4) und (5.5) bedingte Filterung der Geschwindigkeit, die bis auf die überlagerte Welligkeit einigermaßen glatt ist, wieder.

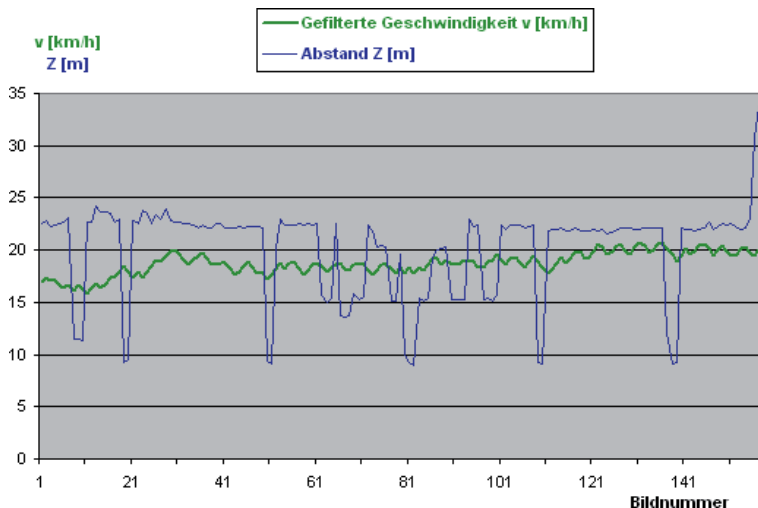


Bild 190: Wie Bild 189, aber mit gefilterten Geschwindigkeiten

Fahrt entlang eines Gebäudes mit vielen Verdeckungen

Bei dieser Fahrt sind viele Verdeckungen durch teilweise erfasste Autodächer, Personen, Laternen und Bäume aufgetreten. Das Bild 191 gibt die Situation wieder. Trotz der Verdeckungen sind meistens, wie dem Bild 192 entnommen werden kann, die Geschwindigkeiten richtig berechnet worden.

In Bild 193 ist die gefilterte Geschwindigkeit gezeigt.



Abbildung Nr. 15

Abbildung Nr. 41

Abbildung Nr. 58

Abbildung Nr. 91



Abbildung Nr. 116

Abbildung Nr. 141

Abbildung Nr. 166

Abbildung Nr. 191

Bild 191: Bildausschnitte aus einer Fahrt entlang eines Gebäudes mit vielen Verdeckungen und Angabe der Bildnummer innerhalb der Aufnahmesequenz

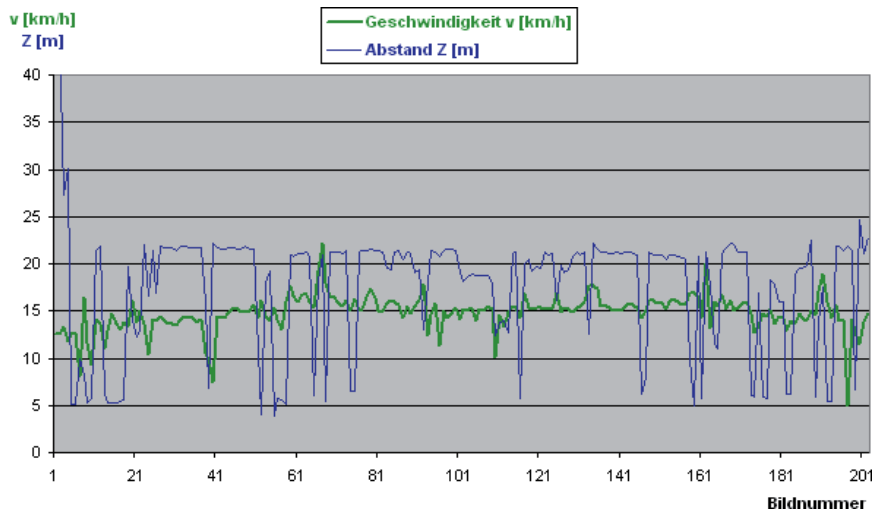


Bild 192: Berechnete Geschwindigkeiten und Abstände bei einer Fahrt entlang eines Gebäudes mit vielen Verdeckungen

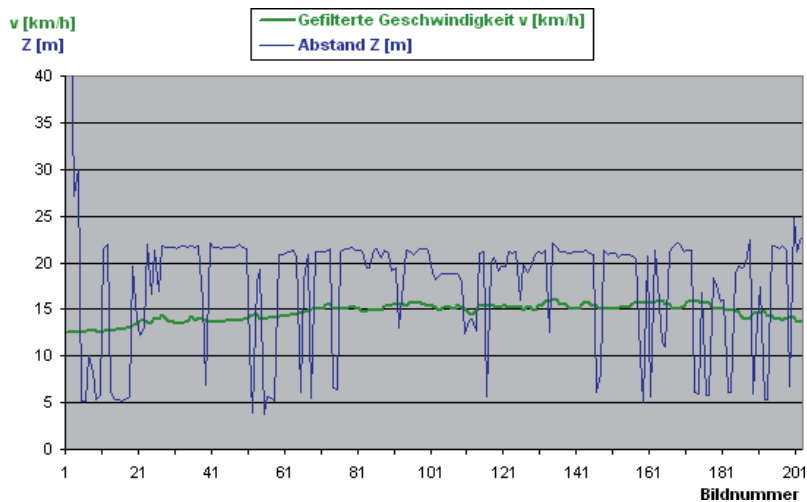


Bild 193: Wie Bild 192, aber mit gefilterten Geschwindigkeiten

Fahrt entlang einer Böschung

Die Fahrt entlang einer Böschung ist durch die fast zur optischen Achse parallelen Flächen mit geringer und periodischer Struktur, Abbildungen Nr. 15 und 105 in Bild 194, den Verdeckungen innerhalb des Buschwerks, z. B. Abbildung Nr. 157, und der Nähe der zur beobachteten Szene bei relativ hohen Geschwindigkeiten in Bezug zum Abstand, siehe Abbildungen Nr. 212 und 256, eine Herausforderung für den Algorithmus. In der zweiten Hälfte der Bildsequenz, ab Abbildung 280, gab es keine Verdeckungsprobleme mehr. Das führte zu einem einigermaßen glatten Geschwindigkeitsverlauf, der unverändert im gefilterten Verlauf, Bild 196, wiederzufinden ist.



Abbildung Nr. 15



Abbildung Nr. 75



Abbildung Nr. 105



Abbildung Nr. 157



Abbildung Nr. 212

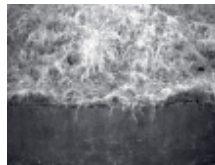


Abbildung Nr. 256

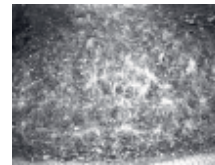


Abbildung Nr. 377



Abbildung Nr. 557

Bild 194: Bildausschnitte aus einer Fahrt entlang einer Böschung mit Angabe der Abbildungsnummer innerhalb der Aufnahmesequenz

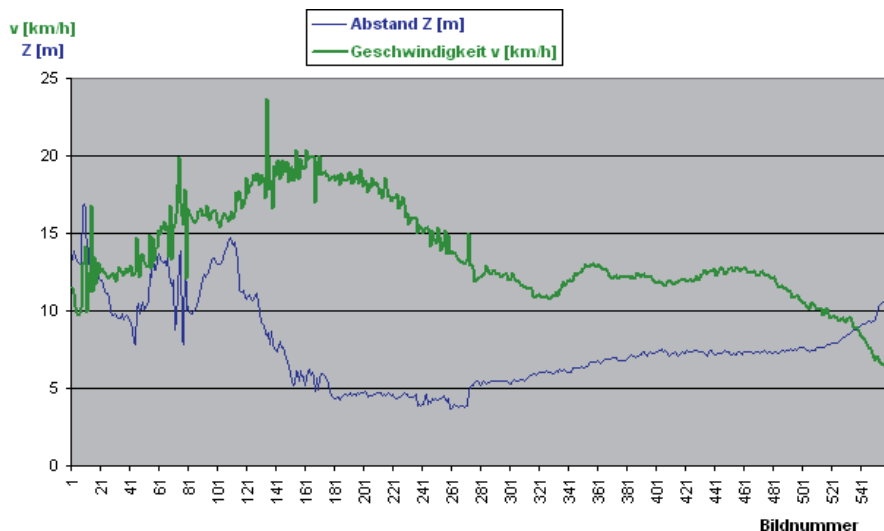


Bild 195: Berechnete Geschwindigkeiten und Abstände bei einer Fahrt entlang einer Böschung

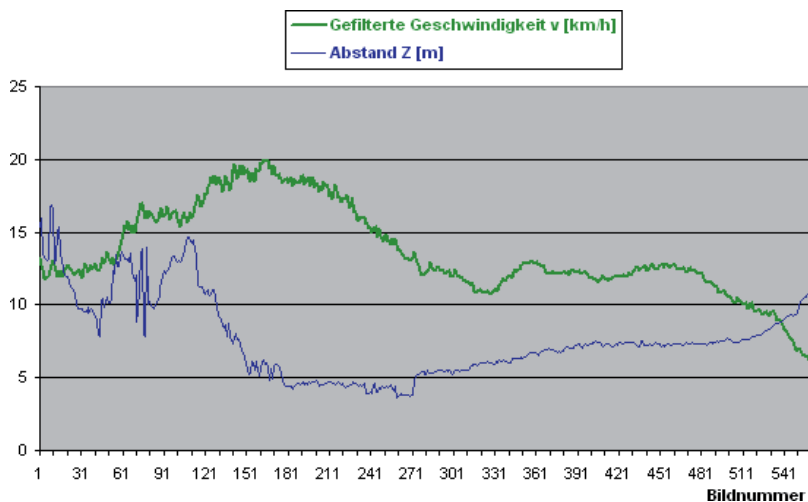


Bild 196: Wie Bild 195, aber mit gefilterten Geschwindigkeiten

5.5 Zusammenfassung Kapitel 5

Über eine virtuelle Welt und einen XY-Tisch wurde die grundsätzliche Funktion der Verfahren nach Lucas&Kanade und der Kreuzkorrelation nachgewiesen. Flugaufnahmen wurden mit einer Monokamera durchgeführt, da die benutzte Hardware für die Stereokamera zu schwer war.

Der Kalibrierungs- und Rektifizierungserfolg der Stereokamera ist gut, auch wenn an den weniger wichtigen Rändern eine Abweichung von bis zu 0.6 px feststellbar ist. Das Ziel, eine Subpixelgenauigkeit von besser als 0.3 px für große Abstände, wurde erreicht.

Die erfassbaren Geschwindigkeiten bei unterschiedlichen Höhen genügen den Anforderungen für das Experimentierluftschiff.

Strecken und Abstände können mit dem vorgestellten Verfahren bei verdeckungsfreien Szenen mit Fehlern besser als 2 % berechnet werden. Geschwindigkeiten und Strecken sind bei Verdeckungen nur gefiltert brauchbar und weisen dann auch kleine Ausschläge auf. Bei Flügen über eine Landschaft sind Verdeckungen wesentlich seltener zu erwarten.

Um eine Echtzeitverarbeitung zu garantieren, wurden Vereinfachungen vorgestellt, die die erforderliche Genauigkeit beibehalten. Hierzu zählen:

- Verringerung der Präzision der Bilinearkoeffizienten
Hierbei erwies sich eine 7-Bit-Genauigkeit als völlig ausreichend, während die Nächsten-Nachbar-Interpolation versagte.
- Reduzierung der Rektifikation auf benutzte Bereiche
- Suchbereich für die Bewegungserfassung einschränken
- Bildverkleinerung je nach Situation

Mit diesen Vereinfachungen werden im folgenden Kapitel verschiedene Plattformen zur Echtzeitberechnung vorgestellt.

6 Implementation in Hardware

6.1 Berechnungszeiten

Angestrebt ist eine Bildrate von 12/s. Dies ergibt eine zur Verfügung stehende Zeit von 83 ms für jede Bildfolge.

Die umfangreichen Berechnungen für die Rektifizierungen und für die Kreuzkorrelationen nehmen auf normalen Arbeitsplatzrechnern zu viel Zeit in Anspruch, um eine Echtzeitfähigkeit zu erreichen, s. Tabelle 24. Zusätzlich ist noch die Leistungsaufnahme der Hardware zu berücksichtigen¹.

Tabelle 24: Rechenzeiten in Matlab® bei einem 64-Bit, 3.3-GHz-Prozessor, zu suchendes Teilbild: 32 x32 px, zu durchsuchendes Fenster: 128 x128 px, Bildmaße rektifiziert: 850 x 1224 px

Prozess	Benötigte Rechenzeit
Ganzes linkes Bild rektifizieren	0.46 s
Linkes Bild auf 1/4 und 1/8 verkleinern	0.06 s
Benötigten Ausschnitt des rechten Bildes rektifizieren	0.002 s
Rechten Ausschnitt auf 1/4 verkleinern	0.003 s
Korrelation für Translation	0.44 s
Korrelation für Abstand	0.09 s
Sonstige Berechnungen	0.001 s
Gesamtzeit	1.06 s

Ein 32-Bit, 3.0 GHz-PC benötigt ungefähr die doppelten o. g. Rechenzeiten. Daraus ist zu ersehen, dass die Prozessdauer sehr plattformabhängig ist.

¹ Der Gedanke, die Bilddaten an die Bodenstation zu senden, dort zu verarbeiten und die Ergebnisse zurück zu senden, scheitert an der hohen erforderlichen Bandbreite bei einer Reichweite von mehreren Kilometern, für die keine Sendegenehmigung zu erhalten ist.

6.2 Maschinennahe Programmierung

Die C-Sprache ist sehr maschinennah. Dadurch werden Befehle wesentlich schneller ausgeführt als in Matlab®.

Der geforderten Berechnungszeit von 83 ms ist durch Programmierung in C++ deutlich näher zu kommen. Auf demselben Arbeitsplatzrechner, wie in Tabelle 24 erwähnt, benötigte die Korrelation für eine Translation statt 440 ms nur noch 68 ms. Dies entspricht einem 6.5-fachen Zeitgewinn.

6.3 Parallelisierung

Ein vielversprechender Ansatz ist das Parallelisieren der Algorithmen. Ein Algorithmus ist dann parallelisierbar, wenn er in voneinander kausal unabhängigen Unteralgorithmen aufgeteilt werden kann. In der Bildverarbeitung sind insbesondere pixel-lokale Operationen für parallele Verarbeitung prädestiniert.

Als Beispiel sei die Rektifizierung des linken Bildes erwähnt. Jedes der ca. 10^6 Pixel des rektifizierten Bildes wird durch bilineare Interpolation aus 4 Pixeln des Originalbildes gewonnen. Die Berechnungen können für jeden rektifizierten Bildpunkt, da sie voneinander unabhängig sind, nebenläufig erfolgen. Diese nebenläufigen Berechnungen sind innerhalb der ersten sechs Prozesse der Tabelle 24 durchführbar. Zusätzlich sind die Prozesse „linkes Bild rektifizieren“ und „linkes Bild verkleinern“ mit „rechtes Bild rektifizieren“, „rechtes Bild verkleinern“ und „Korrelation für Abstand“ parallelisierbar, s. Bild 197.

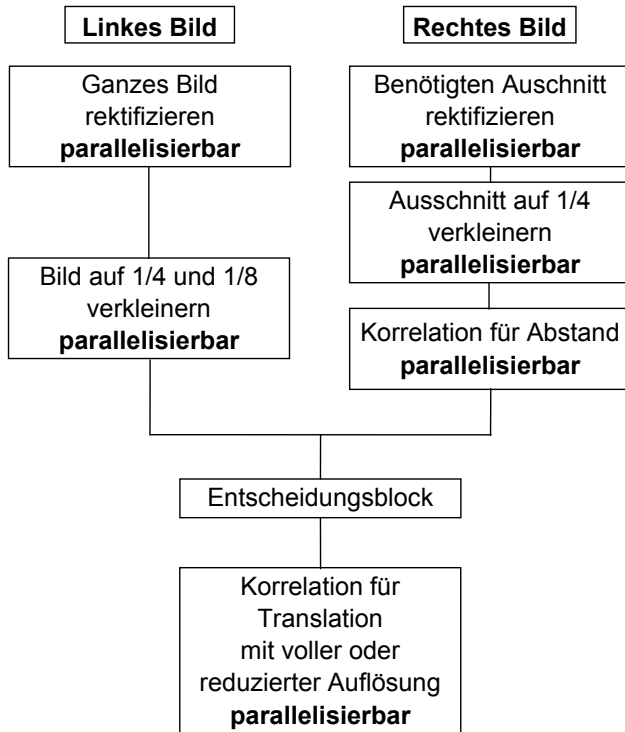


Bild 197: Parallelisieren der Prozesse zur Abstands- und Streckenmessung

6.4 Hardwareprozessoren

Dezidierte Hardware hat neben der Berechnungsgeschwindigkeit, besonders wenn sie zusätzlich Prozesse parallel verarbeitet, den Vorteil, für ein kleines Fluggerät ein geringes Volumen und kleine Masse bei niedrigem Energiebedarf zu haben.

Als Alternativen werden untersucht:

- DSP (Digital Signal Processor)
- FPGA (Field Programmable Gate Array)
- GPGPU (General Purpose Graphics Processing Unit)

Obwohl ein DSP Algorithmen nicht parallel verarbeiten kann, wird er trotzdem in die folgende Betrachtung miteinbezogen.

6.4.1 DSP

Ein DSP verarbeitet Daten bei einer hohen Taktrate nacheinander und führt eine Multiplikation mit zusätzlich einer Addition in einem Taktzyklus durch. Die Programmierung erfolgt meist über die C-Sprache.

Mit einem DSP in Ganzzahlarithmetik wurde die Echtzeitfähigkeit für die Berechnung der NCC nicht annähernd erreicht [97], obwohl hierzu das für serielle Verarbeitung schnelle Verfahren von S. D. Wei und S. H. Lai [81], s. Kap. A.4, implementiert wurde.

Nach Recherchen des Autors ist kein Artikel über das Implementieren einer Bildverarbeitungskorrelation in ein DSP gefunden worden, so dass anzunehmen ist, dass die Korrelation hierfür nicht sinnvoll in einem DSP durchführbar ist.

6.4.2 FPGA

Ein FPGA verarbeitet viele logische Operationen nebenläufig und ist trotz der relativ niedrigen Taktrate sehr schnell. Zur Programmierung sind spezielle Hardwarebeschreibungssprachen entwickelt worden. VHDL (Very High Speed Integrated Circuit Hardware Description Language) und Verilog beschreiben das Verhalten einer Schaltung. Die kompilierten Anweisungen können z. B. in ein FPGA implementiert werden, das dann genau dem beschriebenen Verhalten folgt. Allerdings ist die Programmierung sehr zeitaufwendig und das Finden von Programmierfehlern komplex.

In der Literatur [98], [99] sind erfolgreiche Implementierungen der NCC in ein FPGA erwähnt¹. Der Zeitgewinn wird in [98] mit 80-fach gegenüber einem Arbeitsplatzrechner angegeben. Dies wäre für die Durchführung der Abstands- und Streckenmessung, wie in dieser Arbeit vorgestellt, ausreichend. Allerdings scheint der 80-fache Gewinn sehr hoch angesetzt zu sein, denn in einer anderen Literaturstelle [100] ist der Gewinn mit nur 10-fach angegeben.

¹ Ein eigener Versuch Bildverarbeitungsoperationen in ein FPGA-System einzufügen, scheiterte an der falschen Spezifikation des Bausteins, der anstatt bei 300 MHz nur bei 80 MHz fehlerfrei funktionierte.

6.4.3 GPGPU

Eine GPGPU (General Purpose Graphics Processing Unit) ist im Wesentlichen eine GPU (Graphics Processing Unit), die für andere Anwendungen als nur grafische Darstellungen konzipiert ist. Die GPGPU hat einen erweiterten Befehlssatz und kann einfach über die C-Sprache programmiert werden. Während eine CPU für Programme mit vielen Verzweigungen ausgelegt ist, verarbeitet eine GPU Daten, die sehr schnell einfließen, parallel. Im Unterschied zu einer CPU besitzt eine GPU sehr viele (bis zu mehreren 1000) Recheneinheiten (ALUs), s. Bild 198.

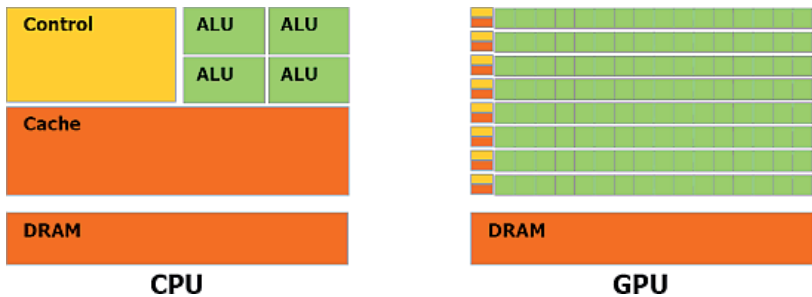


Bild 198: Architektur einer GPU im Vergleich zu einer CPU

Quelle: <http://docs.nvidia.com/cuda/cuda-c-programming-guide/>

Der Einsatz einer GPU zur Berechnung der NCC ist in [101] mit Filmsequenzen, die mit dem Luftschiff der FernUniversität in Hagen aufgenommen wurden, untersucht worden. Die Algorithmen sind in C++ programmiert. Dabei sind die Prozesszeiten sowohl für den Ablauf in einer CPU als auch in einer GPU und der direkte NCC-Algorithmus mit dem Verfahren über die Schwarzsche Ungleichung, s. Gleichung (3.56), verglichen worden, s. Tabelle 25.

Tabelle 25: Durchschnittliche und maximale Rechenzeiten für die NCC auf einer CPU und einer GPU für eine Translation ohne vorhergehende Rektifizierung. 64-Bit, 3.3-GHz-Prozessor, zu suchendes Teilbild: 32 x 32 px, zu durchsuchendes Fenster: 128 x 128 px

Methode	CPU-Zeit		GPU-Zeit	
	Mittelwert	Maximalwert	Mittelwert	Maximalwert
Direkte NCC	68 ms	68 ms	4.7 ms	4.7 ms
NCC über Schwarzsche Ungleichung	3.9 ms	12 ms	2.5 ms	26 ms

Der Vergleich der Berechnungszeit für die Korrelation aus Tabelle 24 (440 ms) und Tabelle 25 (4.7 ms) lässt den Schluss zu, dass die verwendete GPU für die NCC-Berechnungen und die Rektifizierung etwa 100-mal schneller abläuft als in Matlab® und somit die Echtzeitforderung erfüllt werden kann. Allerdings ist zu beachten, dass die Berechnung auf der GPU zusammen mit einem schnellen Arbeitsplatzrechner erfolgt und die Leistungsaufnahme allein der GPU etwa 140 Watt und für das Komplettsystem 240 Watt ist. Damit ist dieses System derzeit nicht direkt auf einem kleinen Fluggerät einsetzbar. Deshalb wird die Implementierung auf eine mobile Plattform untersucht und zukünftig durchgeführt. Ausgewählt ist die Jetson TK1-Karte der Firma Nvidia wegen der Leistungsfähigkeit, der Masse und der Leistungsaufnahme, s. Bild 199.

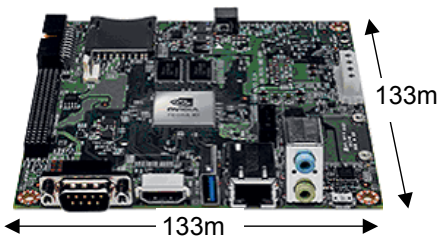


Bild 199: Nvidia Jetson TK1-Karte
 Quelle: <http://www.nvidia.de/object/jetson-tk1-embedded-dev-kit-de.html>

Einige Kerndaten der Nvidia Jetson TK1-Karte:

- Grafikprozessor: 192 CUDA-Recheneinheiten, 158 GFLOPS¹ (single precision)
- CPU: quad-core 2.3 GHz
- Arbeitsspeicher: 2 GB DDR3L 933 MHz RAM
- Datenspeicher: 16 GB Flash
- Leistungsaufnahme: 12.5 W bei 12 V
- Masse: 120 g

Die Leistungsfähigkeit dieser Karte wurde auf der Internetseite <http://arrayfire.com/arrayfire-on-nvidia-tegra-tk1/> getestet. Dort ergab sich für eine Faltung von einem 5 x 5 Filter mit einem 1280 x 720-px-Bild eine Berechnungszeit von 3.6 ms. Dies entspricht $2.5 \cdot 10^9$ Multiplikationen und Additionen pro Sekunde oder einer Multiplikation und Addition in 0.4 ns.

Wird die Zeit von 0.4 ns als Grundlage für eine Abschätzung der gesamten Berechnungszeit für die Abstands- und Translationsbestimmung genommen, so ergibt sich eine Zeit von 4.4 ms, s. Tabelle 26.

Die Zeiten zu den relativ wenigen Berechnungen für Divisionen und Quadratwurzeln können vernachlässigt werden, auch wenn laut Firma NVIDIA [102] Divisionen 9-mal und Wurzelberechnungen 4-mal länger dauern als Multiplikationen.

Nicht berücksichtigt ist hierbei, dass im Verfahren dieser Arbeit wesentlich mehr Daten transportiert werden müssen. Schon alleine für die Rektifizierung müssen alle Bilinearkoeffizienten in der GPU verarbeitet werden. Dies bedeutet einen etwa 4-fachen Datentransport gegenüber der o. g. Faltung, der nach [101], [103] erheblich in die Prozesszeit eingeht.

Trotzdem ist die geforderte Berechnungszeit von 83 ms – entsprechend einer Bildrate von 12/s – zuversichtlich zu erreichen, auch dann, wenn für den Datentransport einige Millisekunden vergehen.

¹ GFLOPS: Giga-Gleitkommaoperationen pro Sekunde

Tabelle 26: Anzahl der Instruktionen und hochgerechnete Dauer für die Berechnung des Abstandes und der Translation der Kamera

Prozess	Anzahl Multiplikation und Additionen	Benötigte Zeit bei 0.4 ns pro Multiplikation und Addition
Linkes und rechtes Bild aus verschachtelten Daten extrahieren	$2.1 \cdot 10^6$	0.84 ms
Ganzes linkes Bild rektifizieren	$4.2 \cdot 10^6$	1.68 ms
Linkes Bild auf 1/4 und 1/8 verkleinern	$1.3 \cdot 10^6$	0.52 ms
Benötigten Ausschnitt des rechten Bildes rektifizieren	$0.083 \cdot 10^6$	0.03 ms
Rechten Ausschnitt auf 1/4 verkleinern	$0.099 \cdot 10^6$	0.04 ms
Korrelation für Translation	$1.9 \cdot 10^6$	0.76 ms
Korrelation für Abstand	$1.2 \cdot 10^6$	0.48 ms
Summe	$11.3 \cdot 10^6$	4.4 ms

6.5 Zusammenfassung Kapitel 6

Die Rechenzeit für den Gesamtablauf in der Matlabumgebung auf einem 64-Bit-Arbeitsplatzrechner mit 3.3-GHz-Prozessor ist 1.06 s.

Ein DSP wird ausgeschlossen, da die Echtzeitfähigkeit für Berechnung der NCC im Zusammenhang mit der Bildverarbeitung nicht belegt ist. Der serielle Prozessablauf eines DSP erschwert die Echtzeitfähigkeit.

Die Programmierung eines FPGA für einen Prototypen wird wegen des immensen Programmierungsaufwandes und der nicht sicheren Einhaltung der Echtzeitanforderung nicht in Betracht gezogen.

Eine GPGPU, beispielsweise von Nvidia, erfüllt alle Anforderungen bezüglich:

- Parallelisierbarkeit: 192 Prozesse können nebenläufig durchgeführt werden
- Leistungsfähigkeit: 158 GFLOPS
(entspricht 6.5 Picosekunden pro Instruktion)
- Abmaße: 13 cm x 13 cm
- Leistungsaufnahme: 13 W

Die Berechnungszeit wurde auf 4.4 ms geschätzt, so dass die Echtzeitberechnung bei einer Bildrate von 12/s, auch wenn der Datentransport nicht berücksichtigt wurde, zuversichtlich erreicht wird.

7 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wird gezeigt, wie mit einer Stereokamera geringer Basisbreite über die visuelle Odometrie die Position und die Geschwindigkeit eines Fluggerätes ermittelt werden kann.

Zwei Verfahren der Bildregistrierung – das Gradientenverfahren nach B. D. Lucas & T. Kanade und die normierte Kreuzkorrelation – sind auf die Eignung bezüglich Translationsbestimmung und Abstandsermittlung untersucht worden.

Es stellt sich heraus, dass Gradientenverfahren trotz ihrer inhärenten Subpixelgenauigkeit ungeeignet für Fluggeräte sind, weil die maximal berechenbare Verschiebung (36 px) bei 512 x 512 px umfassenden Bildern nicht ausreicht, um große Translationen zu ermitteln. Diese ergeben nämlich, dass der Abstand korrespondierender Bildpunkte in zwei aufeinanderfolgenden Aufnahmen die maximal berechenbare Verschiebung überschreitet. Für die Bestimmung niedriger Flughöhen gilt analog, dass die Disparität größer als die maximal berechenbare Verschiebung der gleichzeitig aufgenommenen Bilder ist. Zudem führt die Helligkeitsverteilung innerhalb der Bilder zu erheblichen Fehlern des berechneten optischen Flusses.

Die normierte Kreuzkorrelation ist nur durch ihre Bildgröße beschränkten maximal erfassbaren Verschiebungen und ihre Beleuchtungsrobustheit den Gradientenverfahren für die untersuchte Anwendung überlegen.

Obwohl in der verwendeten Stereokamera gleiche Kameramodule für die linke und rechte Stereoaufnahme verwendet werden, unterscheiden sich die Intensitäten der Bilder derselben Szene sowohl um einen multiplikativen Faktor als auch einen additiven Anteil. Zusätzlich entsteht durch Vignettierung ein bei beiden Kameras unterschiedlicher Helligkeitsverlauf. Durch die Normierung der Kreuzkorrelation und den Abzug der Mittelwerte der betrachteten Bildgebiete wird der Einfluss der multiplikativen und additiven Helligkeitsunterschiede weitgehend behoben.

Um eine ständige Neukalibrierung der Stereokamera zu vermeiden, wird ein rigides Kamerasystem verwendet. Da dieses System auch ein geringes Gewicht haben muss, ist es geometrisch klein und weist dadurch eine geringe Basisbreite von 24 cm auf. Die geringe Basisbreite hat zunächst Nachteile in Bezug auf die Auflösungen bei der Bestimmung großer Abstände, da z. B. mit

dem verwendeten Kamerasystem bei einem Abstand von 50 m eine Disparitätsdifferenz von 1 px schon einen Abstandsunterschied von 18 % ergibt.

Während die Ermittlung großer Translationen eine kleine Brennweite erfordert, weil die hierzu zu berechnenden Verschiebungen sonst den Bildbereich überschreiten, ist für die Berechnung großer Abstände zunächst eine große Brennweite notwendig, da die Disparitäten sonst zu klein und damit ungenau werden. Dieser Konflikt wird durch eine höhere Pixelauflösung behoben, indem Subpixel bestimmt werden.

Die erforderliche Subpixelgenauigkeit der Kreuzkorrelationsberechnung wird durch eine in diesem Zusammenhang neuartige Funktion – hier genannt 'inverse Parabel' – erreicht. Im einfachsten Fall ist die 'inverse Parabel' eine $1/x^2$ -Funktion, die den Korrelationsverlauf, der zunächst nur für ganzzahlige Pixelpositionen bekannt ist, approximiert. Das Maximum dieser Approximationskurve ist im Idealfall genau an der als reelle Zahl berechneten Position des gesuchten Korrelationsmaximums, das die Bildverschiebung bestimmt. Die Dezimalstellen dieser berechneten Position sind die Subpixelwerte.

Die mit der 'inversen Parabel' erreichte Subpixelgenauigkeit bei Bildern, die exakt um ein Mehrfaches von $1/32$ Pixel versetzt sind, ist besser als 0.06 px. Um den berechneten Versatz zu verifizieren, werden in dieser Dissertation große Bildausschnitte, die um das Mehrfache eines Pixels versetzt sind, auf ein $1/32$ verkleinert und jeweils miteinander korreliert. Bei realen Stereoaufnahmen erweist sich jedoch der Gewinn durch die Anwendung der 'inversen Parabel' im Vergleich zur regulären Parabel geringer als erwartet.

Bei der Verarbeitung von realen Aufnahmen mit Abständen zwischen 1 m und 1.2 km zur Szene wird eine Subpixelgenauigkeit von unter 0.13 px erreicht. Hierbei wird immer die präzis kalibrierte Bildmitte zur Disparitätsberechnung herangezogen. Mit der erreichten Subpixelgenauigkeit können die für ein kleines UAV relevanten Abstände der Kamera zur Szene von 1 m zu 65 m mit einer Genauigkeit von $\pm 0.4 \%$ Abstandsabweichung bestimmt werden. Bei 400 m Abstand steigt der Abstandsfehler auf 4.5 % an. Die Genauigkeiten der Abstandsmessungen gehen auch in die Streckenberechnungen ein, die jeweils über Triangulation ermittelt werden.

Um Strecken zu berechnen, sind folgende Versuche durchgeführt worden:

- Im Labor mit exakt bekannten geometrischen Verhältnissen ist die Kamera auf einem Laborwagen montiert und es werden Aufnahmen einer Tafel mit Bildmotiv ausgewertet.

Ergebnis der Berechnung: maximal 0.4 % Abweichung zu den tatsächlichen Strecken (2.5 m und 2.7 m)

- Im Außenbereich wurden Fassaden mit der auf einem Stativ rotierenden Kamera bewertet.

Ergebnis: maximal 1.7 % Abweichung zu den tatsächlichen Strecken (12.5 m und 69.6 m)

- Eine Kamera ist auf einem PKW montiert, der entlang eines 65.7 m langen Gebäudes fährt.

Ergebnis: 0.75 % Längenabweichung.

Die Momentangeschwindigkeit wird mit der Translation und der Bildrate von 12/s innerhalb von 83 ms ermittelt. Die Bildrate und damit die Bildfolgezeit der verwendeten Kamera ist im Durchschnitt sehr genau. Abweichungen der Zeitintervalle zwischen zwei aufeinanderfolgenden Aufnahmen von bis zu 1.28 ms kommen vor, werden aber durch die unmittelbar darauffolgende dazu negative Abweichung vollkommen kompensiert. Deshalb sind die prozentualen Geschwindigkeits- und Translationsmessfehler gleich. Die Geschwindigkeitsmessung wird qualitativ beurteilt, indem Aufnahmen mit der auf einem PKW montierten Kamera entlang einer Böschung und eines Gebäudes erstellt werden. Dabei unterscheiden sich einige Aufnahmen derselben Szene aus unterschiedlichen Blickwinkeln durch Objekte (z. B. Bäume oder Laternenmasten) im Vordergrund, die den Hintergrund abdecken, so stark, dass die Korrelation zu falschen Ergebnissen führt.

Bei diesen realitätsnahen Versuchen sind Verdeckungen mit der eingeführten Erkennungsmethode, die über Bildverkleinerungen größere Szenenausschnitte bei gleicher oder kleinerer Berechnungszeit betrachtet, detektiert worden. Einige nicht detektierte Verdeckungen und auch Fehlberechnungen werden aufgrund der eingeschränkten Dynamik des Luftschiffs ausgeschlossen. In diesen Fällen werden vorhergehende Berechnungen für die Positionsbestimmung herangezogen.

Um Bedingungen für die Echtzeitverarbeitung einzuhalten, werden viele Vereinfachungen eingeführt, die eingehend auf die erforderliche Genauigkeit untersucht sind:

- Reduzierung der Rektifizierung auf benutzte Bereiche
- Reduzierung der Bilinearkoeffizienten auf 7 Bit
- Nachführung des zu durchsuchenden Fensters
- Situationsabhängige Bildverkleinerung

Durch die gute Parallelisierbarkeit des Algorithmus zur Berechnung der Kreuzkorrelation werden in einer in einem PC eingebauten GPGPU Berechnungszeiten für die Translation in 4.7 ms erreicht.

In der vorliegenden Dissertation ist gezeigt, dass durch visuelle Odometrie die relative Position, die momentane Geschwindigkeit und die genaue Höhe über Boden in Echtzeit trotz geringer Basisbreite der Stereokamera mit guter Genauigkeit berechenbar sind. Die Ergebnisse dieser Bildverarbeitung schließen die Lücke zwischen einem schnellen aber nur kurzzeitig genauen Inertialsystem (IMU) und einem GPS-System, welches relativ langsam die Position – allerdings in Weltkoordinaten – liefert. Durch die Fusionierung der IMU-, der Kamera- und GPS-Daten über ein Kalmanfilter wird die Regelung eines autonomen Fluggerätes in der Bewegung und der Navigation genauer. Das in dieser Dissertation vorgestellte Verfahren kann für verschiedene UAVs eingesetzt werden, wobei die Bildrate der Dynamik des Flugsystems jeweils angepasst werden muss.

Der nächste Schritt ist die Umsetzung des Verfahrens in eine kleine GPGPU-Plattform, die die Positionsberechnung in Echtzeit auf dem Luftschiff durchführt.

Da die Gesamtberechnungszeit im einstelligen Millisekundenbereich liegen wird, sind noch weitere Berechnungen innerhalb der Bildfolgezeit von 83 ms möglich. Z. B. könnte das mittlere von drei Objekten der in dieser Dissertation benutzten Bumblebeekamera mit einer erheblich kürzeren Brennweite versehen und damit der schwierige Landeanflug visuell unterstützt werden, indem vorhergehende Aufnahmen des Landeplatzes in Bezug zu momentanen Aufnahmen gebracht werden und somit die genaue absolute Position des Luftschiffs erfasst wird. Ein Verfahren hierzu ist – in Kapitel 1.3 kurz vorgestellt – erarbeitet worden.

Ebenfalls können Probleme bezüglich Verdeckungen besser beseitigt werden, indem die Korrelation an mehreren Stellen der Aufnahme durchgeführt wird und über eine Mehrheitsentscheidung die vermutlich richtige Korrelation für die Abstands- und Translationsberechnung gewählt wird.

Anhang

A.1 Lochkamera: Belichtungszeit und Unschärfe durch Beugung

Dadurch, dass die Öffnung der Lochkamera eine endliche Größe hat, wird ein Punkt als Unschärfekreis abgebildet. Die Öffnung zur Bildschärfeerhöhung beliebig zu verkleinern, ist wegen der Beugung des Lichtes an Kanten begrenzt. Die kritische Blende ist gegeben durch den minimalen Abbildungskreis, der sich aus der Öffnungs- und Beugungsabbildung ergibt [104][105][37]. Der Unschärfekreis wird dann für die Abbildung unendlich entfernter Punkte doppelt so groß wie die Öffnung. Der optimale Durchmesser d ergibt sich aus [106][104] zu

$$d = \sqrt{\frac{\lambda \cdot f}{2}} \quad (\text{A.1})$$

Werden die Wellenlänge $\lambda = 0.5 \text{ } \mu\text{m}$ (grünes Licht) und die Brennweite $f = 10 \text{ mm}$ gewählt, so ergibt sich ein Durchmesser von $d = 50 \text{ } \mu\text{m}$. Unendlich weit entfernte Punkte erscheinen dann als Kreise mit $100 \mu\text{m}$ Durchmesser. Bei einem gängigen CCD-Sensor mit den Abmaßen von $7.2 \text{ mm} \times 9.6 \text{ mm}$ folgt, wenn man eine Radienüberlappung von 30 % zulässt, eine Anzahl von ca. 14000 zu unterscheidenden Bildpunkten. Bei einer Lochkamera mit einem Sensor von $60 \text{ mm} \times 60 \text{ mm}$ und einer Brennweite von 60 mm ergibt sich eine Auflösung von ca. 420000 Bildpunkten. Eine moderne Digitalkamera hat nach Herstellerangaben (Canon SX210) 14×10^6 Bildpunkte. Allerdings ist die effektive Anzahl der Bildpunkte auch dort durch Beugung, Nebensprechen und Linsenverzerrungen geringer, denn Auflösungsmessungen nach ISO 12233 [107] [108] ergeben bei dieser Digitalkamera durchschnittlich 1280 Linienpaare für die Bildhöhe und somit insgesamt 8×10^6 unterscheidbare Bildpunkte.

Durch die kleine Öffnung der Lochkamera dringt sehr wenig Licht ein und geht auf die Blendenzahl B ein. Sie errechnet sich aus Brennweite geteilt durch Öffnungsdurchmesser. Sie ist also bei o. g. Lochkamera $B_L = 200$. Eine übliche Blende bei der Canon SX210 ist $B_c = 5$. Die Belichtungszeit ist

proportional zur Lichtfläche und damit quadratisch zur Blendenzahl. Die Belichtungszeit muss also bei der Lochkamera $\left(\frac{200}{5}\right)^2 = 1600$ mal größer sein.

Damit ergibt sich, dass eine Lochkamera für einen kleinen Sensor eine viel zu geringe Auflösung aufweist und eine große Belichtungszeit hat. Die Lochkamera ist trotz ihrer Verzerrungsfreiheit daher für Bildbearbeitung ungeeignet.

A.2 Nützliche Eigenschaften digitaler Filter

Teilbarkeit durch 2^n , n natürliche Zahl

Das Teilen durch 2^n ist mit digitalen Prozessoren (z. B. CPU oder DSP) erheblich schneller, da es durch Rechtsschieben um n Bits realisiert werden kann.

Ganzzahligkeit

Viele Prozessoren arbeiten mit ganzen Zahlen schneller, daher sind Kommazahlen zu vermeiden.

Separierbarkeit:

Ein Filter W ist dann separierbar, wenn gilt:

$$W = a \cdot V * H^T \quad (\text{A.2})$$

V und H sind Vektoren, a ist ein Skalierungsfaktor.

Dadurch kann eine zweidimensionale Matrix, z. B. ein Bild I , mit zwei einfachen eindimensionalen Operationen zu einer gefilterten Matrix I_g berechnet werden:

$$I_g = I * a \cdot W = (I * V) * H^T \cdot a \quad (\text{A.3})$$

Erst wird das Bild I mit dem eindimensionalen vertikalen Faltungsvektor V spaltenweise gefaltet. Das Ergebnis wird dann zeilenweise mit dem horizontalen Vektor H^T gefaltet.

Die separierbaren Faltungsvektoren für Gleichung (3.38) sind:

$$H^T * V = \begin{pmatrix} 1 & 14 & 34 & 14 & 1 \end{pmatrix} * \begin{pmatrix} 1 \\ 14 \\ 34 \\ 14 \\ 1 \end{pmatrix} \quad (\text{A.4})$$

Meist werden isotrope Filter genommen. Sie bevorzugen keine Richtung und sind symmetrisch. Damit ist $H^T = V$ wie in Gleichung (A.4) dargestellt. Der Skalierungsfaktor a normiert das Filter, damit es eine Gesamtgewichtung von exakt eins erhält. Er wird also aus dem Kehrwert der Summe aller Filterkoeffizienten errechnet.

Mit separablen Filtern der Dimension $k \cdot k$ werden anstatt k^2 nur $2k$ Rechenoperationen benötigt pro Pixel [75].

Gaußfilter

Eine Filterung ist erforderlich, um Rauschen zu vermindern, aber auch bei Unterabtastung von Bildern, wie es schon in Kapitel 2.4.2 gezeigt wurde. Deutlich wird es auch im folgenden Bild.

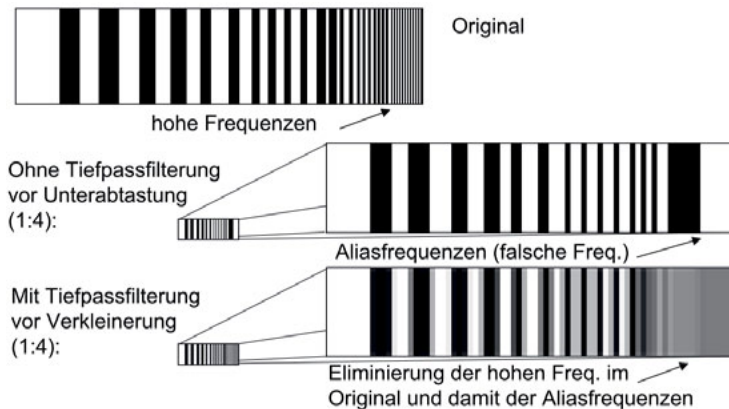


Bild 200: Effekte durch Unterabtastung

Quelle: G. Heising, Digitale Videotechnik, SS 02, TFH Berlin

Im oberen Bild ist das Original mit steigenden Frequenzen zu sehen. Das mittlere Bild (rechts vergrößert) wurde gebildet durch Abtastung jedes vierten Pixels in vertikaler und horizontaler Richtung. Deutlich zu sehen ist die falsche

Abbildung bei hohen Frequenzen. Beim unteren Bild sind die hohen Frequenzen durch Tiefpassfilterung eliminiert.

Ziel einer schnellen Filterung ist es, durch möglichst wenige Rechenoperationen eine gute Filterung zu erhalten.

Wenige Rechenoperationen ergeben sich durch ein Filter mit wenigen Koeffizienten, also ein spatial kleines Filter. Die maximale Filterung wird durch ein möglichst an der Grenzfrequenz steiles Filter erreicht. Aus der Fouriertransformation ist bekannt, dass die Breite der Transformaten $\mathcal{F}(\omega)$ umgekehrt proportional ist zur Breite der Zeitfunktion $f(x)$. Es gilt:

$$f(ax) \rightarrow \mathcal{F}\left(\frac{1}{a}\omega\right) \quad (\text{A.5})$$

Das Gaußfilter ist das Filter, das am kompaktesten sowohl im Orts- als auch im Frequenzbereich ist. Jede andere Ortsbegrenzte Funktion ergäbe einen größeren Einfluss höherer Frequenzen [109]. Dabei wird bei der Diskretisierung zugrunde gelegt, dass Gaußfunktion bei größeren x-Werten null ist. Bei $x = 3\sigma$ ist der Funktionswert ca. 1 % der Amplitude.

Zunächst sei die eindimensionale normierte Gaußfunktion betrachtet:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \rightarrow \mathcal{F}(f(x)) = e^{-\frac{\omega^2\sigma^2}{2}} \quad (\text{A.6})$$

Eine fouriertransformierte Gaußfunktion ist wieder eine Gaußfunktion.

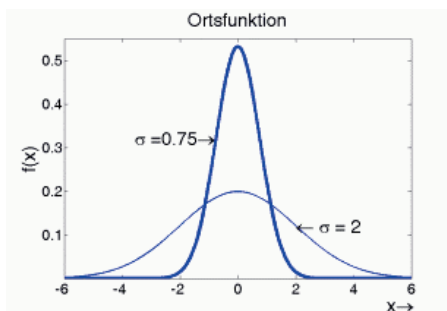


Bild 201: Ortsfunktion zweier Filter mit $\sigma = 0.7$ (dicke Linie) und $\sigma = 2$ (dünne Linie)

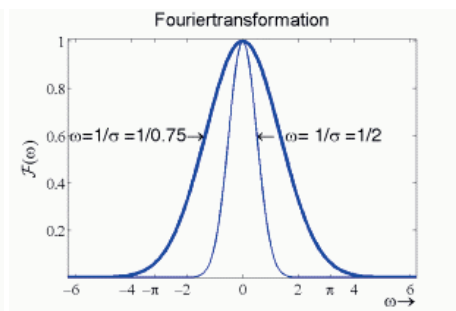


Bild 202: Fouriertransformation der nebenstehenden Funktionen

Die Ortsfunktion und ihre Transformation, Gleichung (A.6), sind für zwei σ -Werte in den Abbildungen 201 und 202 dargestellt. Erwartungsgemäß kann den Grafiken entnommen werden, dass eine breite Ortsfunktion – hier mit $\sigma = 2$ – eine schmale Fouriertransformierte liefert.

Soll ein Bild verkleinert werden, so muss, um Aliasing zu vermeiden, das Abtasttheorem – mindestens 2 Abtastwerte pro Periode der höchsten Frequenz – eingehalten werden. Die höchste Frequenz ergibt sich beim Originalbild zu:

$$f_{\max} = \frac{1}{2px} \quad (\text{A.7})$$

Denn zwei nebeneinanderliegende Pixel unterschiedlicher Intensität ergeben die höchste Dichte der Abtastwerte. Mit der dimensionslosen Bezeichnung px wird dann die höchste Kreisfrequenz zu:

$$\omega_{\max} = \pi \quad (\text{A.8})$$

Wird nun das Bild in x- und y-Richtung halbiert, wird auch die Abtastrate halbiert und alle Frequenzanteile des Originalbildes oberhalb $\pi/2$ müssen durch ein Tiefpassfilter eliminiert werden, um dem Abtastkriterium zu genügen. Ein Rechteckfilter von $-\pi/2$ bis $\pi/2$ im Frequenzbereich wäre ideal. Im Ortsbereich ergibt dies bekannterweise die $\frac{\sin(x)}{x}$ -Funktion, welche langsam

gegen Null konvergiert und deshalb viele Filterkoeffizienten für eine grobe Näherung benötigt. Aus diesem Grunde wird das Rechteckfilter durch ein Gaußfilter approximiert, weil dieses, wie oben erwähnt, sowohl im Orts- als auch im Frequenzbereich kompakt ist.

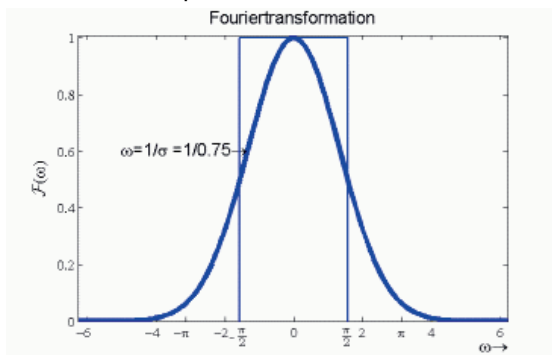


Bild 203: Rechteckfilter und Approximation durch Gaußfilter

Die Approximation wird lediglich durch Variation von σ durchgeführt. Je nach Fehlermaß kommt man zu verschiedenen Ergebnissen. Die Tschebyscheff-Approximation (auch L^∞ -Approximation genannt) ist die Approximation, die für die größte Abweichung den minimalsten Wert liefert [37]. Andere Approximationen, z. B. „least squares“ (auch L_2 - Approximation genannt), können auch durchgeführt werden. Vielfach wird der Wert der Tschebyscheff-Approximation genommen. Er errechnet sich zu [109]:

$$\sigma = 0.75 \quad (\text{A.9})$$

Dieses Filter kann einigermaßen mit fünf diskreten Werten dargestellt werden. Nimmt man z. B. drei diskrete Werte für das Filter, so muss σ viel kleiner gewählt werden, damit das Filter nicht einer Rechteckfunktion mit ihren vielen Nachteilen ähnelt.

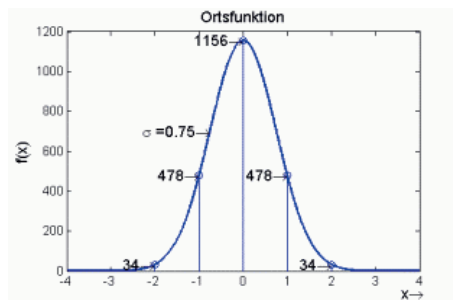


Bild 204: Optimales Filter mit fünf diskreten Werten. Das Filter nach Gleichung (3.38) entspricht in guter Näherung diesem Filter

In Bild 204 ist die Amplitude so gewählt, dass bei den separierten Anteilen nach Gleichung (A.4) der kleinste Wert genau eins ist. Das dazugehörige zweidimensionale Filter ist hier gezeigt:

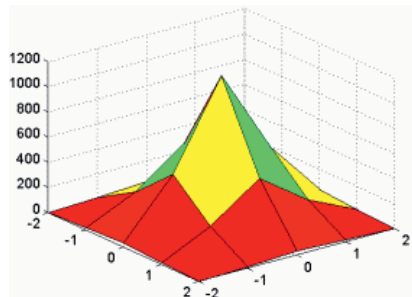


Bild 205: Dreidimensionale Ansicht eines diskreten 5 x 5-Gaußfilters mit $\sigma = 0.75$

Da das Gaußfilter isotrop ist, filtert es in allen Richtungen gleich stark.

Eine Beschreibung und ein eigenes Matlab®-Programm zur Erstellung von Gaußfiltern mit den o. g. Eigenschaften ist im Internet zu finden unter [110]

A.3 Summentabellen

Eine Summentabelle (engl. summed-area table) ermöglicht nach einer einmaligen Aufstellung die schnelle Berechnung der Summe eines beliebigen Rechtecks innerhalb einer Wertetabelle. Im Fall eines Bildes besteht die Wertetabelle $I(x,y)$ aus den Intensitätswerten an den Stellen x,y .

Die Methode wurde zum ersten Mal 1984 in einer Abhandlung von J. Crow [80] dargestellt.

Mit dem Ursprung der Tabelle in der linken unteren Ecke errechnet sich die Summe S der Werte eines Rechtecks, welches den Ursprung beinhaltet und bis zur Stelle u,v reicht, zu:

$$S(u,v) = \sum_{y=1}^v \sum_{x=1}^u I(x,y) \quad (\text{A.10})$$

Um die Summentabelle zu bilden, muss nicht jedes Mal die ganze Summe neu berechnet werden, sondern es kann rekursiv auf vorhergehende Summen zurückgegriffen werden:

$$S(u,v) = I(u,v) + S(u-1,v) + S(u,v-1) - S(u-1,v-1) \quad (\text{A.11})$$

Hierbei ist $S(u,v) = 0$ für $u,v < 1$. Man kann auch einfach die Tabelle I um eine Zeile und Spalte mit Nullen nach unten und links erweitern. Die Summentabelle wird ausgehend vom Ursprung zeilen- oder spaltenweise mit Hilfe der Formel (A.11) gebildet.

Das Bild 206a zeigt die Wertetabelle I für ein 4×4 -Pixel-Bild. Die dazugehörige Summentabelle ist in Bild 206b zu sehen. Die Summe der Intensitätswerte des grauschattierten Gebiets, welches durch u_{\max} und v_{\max} definiert ist, ergibt sich dann zu 45.

	1	2	3	4
V_{max}	5	6	7	8
	9	1	2	3
	4	5	6	7
	U_{max}			
U_{min}				

Bild 206a: Intensitätswerte eines Bildes

	19	33	51	73
V_{max}	18	30	45	63
	13	19	27	37
	4	9	15	22
	U_{max}			
U_{min}				

Bild 206b: Summentabelle zu Bild 206a

Da für alle Rechtecke, die den Ursprung beinhalten, die Wertesumme abgelesen werden kann, können auch Summen für Rechtecke, die vom Ursprung entfernt sind, einfach errechnet werden, indem Summen für anliegende Rechtecke, die dann den Ursprung beinhalten, subtrahiert werden. An einem Beispiel ist dies einfach erklärt.

	1	2	3	4
V_{max}	5	6	7	8
V_{min}	9	1	2	3
	4	5	6	7
	U_{min}			
U_{max}				

Bild 207a: Intensitätswerte eines Bildes. Summe des schraffierten Rechtecks soll errechnet werden.

	19	33	51	73
V_{max}	18	30	45	63
V_{min}	13	19	27	37
	4	9	15	22
	U_{min}			
U_{max}				

Bild 207b: Summentabelle

$$\begin{aligned}
 \sum_{y=v_{min}}^{v_{max}} \sum_{x=u_{min}}^{u_{max}} I(x, y) &= 22 \\
 &= \underbrace{S(u_{max}, v_{max})}_{\Sigma \text{ Gebiet ab Ursprung}} - \underbrace{S(u_{min}-1, v_{max})}_{\Sigma \text{ linker Rand (senkrecht schraffiert)}} - \underbrace{S(u_{max}, v_{min}-1)}_{\Sigma \text{ unterer Rand (waagrecht schraffiert)}} + \underbrace{S(u_{min}-1, v_{min}-1)}_{\text{hinzufragen der doppelt abgezogenen Summe beim Ursprung. (senkrecht und senkrecht schraffiert)}} \\
 &= 73 - 33 - 37 + 19
 \end{aligned}
 \tag{A.12}$$

Die Formel aus Gleichung (A.12) gilt allgemein. Allerdings ist zu beachten, dass die entsprechenden Summanden null zu setzen sind, wenn $u_{\min} - 1$ oder $v_{\min} - 1$ kleiner als die Ursprungskordinaten sind.

Die Anzahl der Additionen errechnet sich aus Folgendem:

Angenommen sei ein großes quadratisches Fenster mit einer Breite von M px und ein kleines Fenster der Breite N px. Für jede Position des kleinen Fensters im großen Fenster soll die Summe aller Werte gebildet werden, so gilt für die direkte Berechnung als Anzahl der Additionen A_d

$$A_d = (M - N + 1)^2 N^2 \quad (\text{A.13})$$

und für die Anzahl A_s bei Berechnung über die Summentabelle

$$A_s = 3(M + 1)^2 + 3(M - N)^2 \quad (\text{A.14})$$

Für $M = 128$ und $N = 32$ ist der Gewinn bei der Berechnung über die Summentabelle 124-fach.

In grober Näherung gilt für den Gewinn:

$$\frac{A_d}{A_s} = \frac{N^2}{6}, \quad M \gg N \quad (\text{A.15})$$

A.4 Berechnung der NCC über die Schwarzsche Ungleichung

S. D. Wei und S. H. Lai [81] haben eine schnelle Berechnungsmethode für die NNC, die auf der Schwarzschen Ungleichung basiert, entwickelt.

Aus der Schwarzschen Ungleichung (A.16) kann entnommen werden, dass ein Gebiet, welches beim rechten Teil der Gleichung einen kleineren Wert liefert als ein anderes Gebiet, welches mit dem linken Teil berechnet wurde, kein Maximum liefern kann.

$$\int_a^b I(x)T(x)dx \leq \sqrt{\int_a^b I^2(x)dx} \sqrt{\int_a^b T^2(x)dx} \quad (\text{A.16})$$

Die beste Übereinstimmung zwischen dem im Bild I und einem gesuchten Teilbild T ist an der Position u, v , bei der die NCC maximal ist. Im Prinzip

können alle u, v nach Gleichung (A.17) durchlaufen werden und dann der größte Wert gesucht werden. Dies ist ein großer Rechenaufwand.

$$NCC(u, v) = \frac{\sum_{x,y} I(x+u, y+v)T(x, y)}{\sqrt{\sum_{x,y} I^2(x+u, y+v)} \sqrt{\sum_{x,y} T^2(x, y)}} \quad (\text{A.17})$$

Der Nenner lässt sich, wie in Kapitel A.3 beschrieben, schnell berechnen. Daher muss der Zähler besonders betrachtet werden.

Der Zähler kann schrittweise approximiert werden über die Schwarzsche Ungleichung [37]:

$$\int_a^b I(x)T(x)dx \leq \sqrt{\int_a^b I^2(x)dx} \sqrt{\int_a^b T^2(x)dx} \quad (\text{A.18})$$

Analog gilt für Summen:

$$\sum_{i=1}^N I_i T_i \leq \sqrt{\sum_{i=1}^N I_i^2} \sqrt{\sum_{i=1}^N T_i^2} \quad (\text{A.19})$$

Werden die Summen feiner unterteilt, gilt folgerichtig:

$$\sum_{i=1}^N I_i T_i \leq \sqrt{\sum_{i=1}^k I_i^2} \sqrt{\sum_{i=1}^k T_i^2} + \sqrt{\sum_{i=k+1}^N I_i^2} \sqrt{\sum_{i=k+1}^N T_i^2} \leq \sqrt{\sum_{i=1}^N I_i^2} \sqrt{\sum_{i=1}^N T_i^2} \quad (\text{A.20})$$

Hier wurde die Summe in zwei Terme aufgeteilt.

Das heißt, je feiner die Unterteilung umso mehr nähert sich die Approximation der wahren Produktsumme des Zählers aus Gleichung (A.17).

Für die normierte Kreuzkorrelation NCC nach Gleichung (A.17) ergibt für den Fall einer Zweiteilung des Teilbildes mit dem mittleren Teil der Gleichung (A.20) eine approximierte normierte Kreuzkorrelation NCCa:

$$NCC(u,v) \leq NCCa(u,v) =$$

(A.21)

$$\frac{\sqrt{\sum_{x=1}^N \sum_{y=1}^k I^2(x+u, y+v)} \sqrt{\sum_{x=1}^N \sum_{y=1}^k T^2(x,y)} + \sqrt{\sum_{x=1}^N \sum_{y=k+1}^N I^2(x+u, y+v)} \sqrt{\sum_{x=1}^N \sum_{y=k+1}^N T^2(x,y)}}{\sqrt{\sum_{x=1}^N \sum_{y=1}^N I^2(x+u, y+v)} \sqrt{\sum_{x=1}^N \sum_{y=1}^N T^2(x,y)}}$$

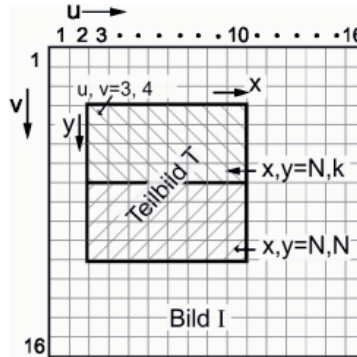


Bild 208: Korrelation eines Ausschnittes von Bild I mit dem Teilbild T nach Gleichung (A.21)

Bild 208 zeigt die Verhältnisse von Gleichung (A.21). Das Teilbild T wird in eine obere und untere Hälfte eingeteilt und mit dem darunterliegenden Bild I (in Bild 208 an der Stelle $u, v = 3, 4$) in grober Näherung korreliert.

Der Zähler von $NCCa$ ist jeweils der obere Grenzwert für den Zähler der NCC . Alle feineren Unterteilungen des Zählerterms von $NCCa$ ergeben kleinere oder gleich große Werte als die NCC und sie nähern sich immer mehr dem Zähler der NCC . Ist schließlich die Unterteilung auf der feinsten Ebene, also auf Pixelgröße, so ist dann die Approximation gleich dem wahren Wert ($NCCa = NCC$).

Das Ziel ist nicht, die NCC für alle u, v auszurechnen, sondern das u, v zu finden, welches die größte NCC hat. Dadurch kann die Berechnung verkürzt werden, indem alle Kandidaten, die eine kleinere approximierte $NCCa$ als die bisher größte $NCCa$ haben, ausgeschlossen werden. Dies wird sukzessiv mit immer höherer Auflösung der Approximation durchgeführt, bis nur noch ein Kandidat – der mit der größten $NCCa$ – übrig bleibt. Damit ist dann auch die NCC für den gefundenen Ort u, v bestimmt.

Das Verfahren ist sehr schnell, hat aber den Nachteil, dass es bei niedrigen Korrelationen deutlich mehr Durchläufe erfordert, somit die Rechenzeit nicht

vorhersagbar ist und in einigen Fällen sogar länger braucht als die direkte Berechnung über die NCC.

A.5 Synchronisations- und Bildratenbestimmung der Stereokamera

Um die Gleichzeitigkeit der Aufnahmen von linker und rechter Kamera festzustellen, wird ein Ventilator mit einer hohen Umdrehungszahl, Bild 209 und Bild 210, benutzt. Die Drehzahl wird mit einer Filmkamera ermittelt, indem der Drehwinkel zwischen zwei aufeinanderfolgenden Bildern unter Berücksichtigung der Unterabtastung bestimmt wird.

Gemessen:

- Umdrehungswinkel zwischen 2 Bildern: $\varphi = 405^\circ$
- Bildrate: $\text{fps} = 16/\text{s}$
- Radius Ventilatormitte zu Markierung in m und px: $r = 0.175 \text{ m}$, $r_{\text{px}} = 180 \text{ px}$
- Brennweite f in px: 1401
- Positionsabweichung p der Markierung in vertikaler Richtung: $p < 0.3 \text{ px}$

Die Positionsabweichung wird mit ZNCC bestimmt. Die maximale Abweichung beträgt 0.3 px, wobei das rechte Bild immer voreilend ist.

Aus diesen Angaben lassen sich die Umlaufgeschwindigkeit v_u der Markierung und der maximale Zeitversatz t zwischen der linken und rechten Aufnahme bestimmen:

$$v_u = 2\pi r \cdot \frac{\varphi}{360} \text{fps} = 19.8 \frac{\text{m}}{\text{s}} \quad (\text{A.22})$$

$$t = \frac{r p}{r_{\text{px}} v_u} \approx 15 \mu\text{s} \quad (\text{A.23})$$

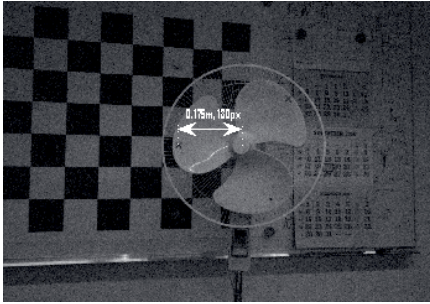


Bild 209: Linkes Bild eines sich drehenden Ventilators

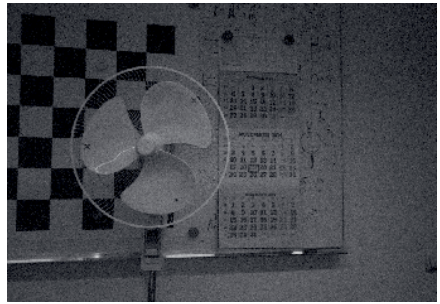


Bild 210: Dazugehöriges rechtes Bild

Fehlerbestimmung:

Es sei angenommen, das Luftschiff bewege sich in Richtung der Kamerabasisachse.

Mit den Gleichungen (2.1) und (2.45) errechnet sich bei einer

- Basisbreite $B = 0.2397 \text{ m}$
- Synchronisationsabweichung $\Delta t = 15 \text{ } \mu\text{s}$
- Geschwindigkeit $v = 50 \text{ km/h}$
- Tatsächliche Höhe $Z_T = 50 \text{ m}$

die gemessene Höhe zu:

$$Z = \frac{B \cdot Z_T}{B + \Delta t \cdot v} = 49.957 \text{ m} \quad (\text{A.24})$$

Die Synchronisation der beiden Kameras ist ausreichend für Geschwindigkeiten unter 50 km/h , denn der Fehler der Höhenbestimmung bleibt unter 0.1% . Andere Fehler, Kamerakalibrierung und Subpixelgenauigkeit, dominieren den synchronisationsbedingten Fehler.

Bildratenbestimmung

Um die Bildrate bzw. die Bildfolgezeit zu bestimmen, wird ein 15-Bit-Binärzähler mit angeschlossenen LEDs betrieben. Die Leuchtdauer der niederwertigsten Leuchtdiode ist $10 \text{ } \mu\text{s}$. Die Kamerabelichtungszeit ist $5 \text{ } \mu\text{s}$. Bild 211 zeigt eine Aufnahme mit einigen leuchtenden Dioden. Aus den Differenzen zwischen zwei benachbarten Binärwerten wird die Bildfolgezeit

berechnet. Dabei wird deutlich, dass Abweichungen vom Sollwert, der 1/12s beträgt, von bis zu 1.28 ms auftreten, s. Bild 212. Allerdings ist die darauffolgende Bildfolgezeitabweichung negativ dazu, so dass sich der Fehler danach wieder aufhebt. Die Abweichungen müssen nicht durch die Kamera bedingt sein. Es kann auch im Zusammenspiel des Rechners mit der Kamera begründet sein.



Bild 211: Aufnahme der Leuchtdioden an einem Binärzähler

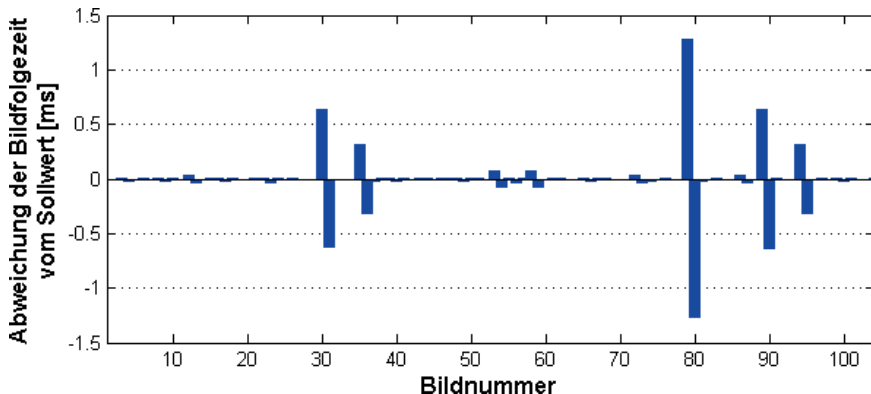


Bild 212: Abweichung der Bildfolgezeit vom Sollwert.

Fehlerbestimmung:

Der Fehler der momentan ermittelten Geschwindigkeit errechnet sich mit:

- Nominale Bildfolgezeit $t = 1/12s$
- Abweichung der Bildfolgezeit $\Delta t = 1.28 \text{ ms}$
- Istgeschwindigkeit v_{ist} , berechnete Geschwindigkeit v_{ber}

ZU:

$$\frac{v_{ist} - v_{ber}}{v_{ist}} = 1 - \frac{t}{t + \Delta t} = 1.5 \% \quad (A.25)$$

Dieser Fehler ist, wie schon erwähnt, von geringer Bedeutung, da er im Folgebild wieder aufgehoben wird.

Literaturverzeichnis

- [1] Boltzmann, Ludwig: Über Luftschiffahrt. Sonderabdruck aus Gesellschaft deutscher Naturforscher und Ärzte, Verhandlungen 1894, Allgemeiner Theil, 1894
- [2] Saint-Fond, Barthélemy Faujas De: Description des expériences de la machine aérostatique de MM. de Montgolfier et de celles auxquelles cette découverte a donné lieu. Buch, Chez Cuchet, 1793:
- [3] Seibold, Peter et al.: Detection of power lines by image processing. Computational Vision and Medical Image Processing, 2013
- [4] Gerke, Michael and Seibold, Peter: Visual Inspection of Power Lines by Unmanned Aerial Systems. International Conference and Exposition on Electrical and Power Engineering, 2014
- [5] Navratil, Florian: Maschinelles Erkennen und Zählen gleichartiger Objekte in Luftaufnahmen. Masterarbeit FernUniversität in Hagen, 2013
- [6] Gerke, Michael: Minen entdecken - Waldbrände erkennen. Homeland Security 2-2009 ISSN 1614-3523, 2009
- [7] N.N.: Detektion, Kartierung, Entschärfung von Landminen. Homeland Security, 2009
- [8] Kleinheins, Peter und Meighörner, Wolfgang: Die großen Zeppeline - Die Geschichte des Luftschiffbaus. Buch, Springer-Verlag, 2005:
- [9] Misra, Pratap and Enge, Per: Global Positioning System: Signals, Measurements & Performance. Buch Lincoln, MA: Ganga-Jamuna Press, 2006
- [10] Jelenčiak, František: Mathematische Modellierung der Flugmechanik eines Luftschiffs sowie Regelung der grundlegenden Fluggrößen zur Navigation. Dissertation, FernUniversität in Hagen, Shaker Verlag, 2014
- [11] Kendoul, Farid: Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems. Journal of Field Robotics, 2012
- [12] Davison, Andrew J. and Molton, Nicholas D.: MonoSLAM: Real-Time Single Camera SLAM. IEEE Trans. Pattern Anal. Mach. Intell., 2007
- [13] Migliore, Davide et al.: Use a single camera for simultaneous localization and mapping with mobile object tracking in dynamic environments. Proceedings of International workshop on Safe navigation in open and dynamic environments application to autonomous vehicles, 2009
- [14] Lange, Sven: Implementierung eines Mono-Kamera-SLAM Verfahrens zur visuell gestützten Navigation und Steuerung eines autonomen Luftschiffs. Diplomarbeit, TU Chemnitz , 2007
- [15] Alcantarilla, P. F. et al.: Optimal Metric SLAM for Autonomous Navigation Assistance. IEEE International Conference on Robotics and Automation, 2009

- [16] Artieda, Jorge, et al.: Visual 3D SLAM from UAVs. Journal of Intelligent and Robotic Systems, 2009
- [17] Caballero, Fernando, et al.: Vision-based odometry and SLAM for medium and high altitude flying UAVs. Unmanned Aircraft Systems, 2009
- [18] Ready, Bryce B. and Taylor, Clark N.: Inertially Aided Visual Odometry for Miniature Air Vehicles in GPS-denied Environments. Journal of Intelligent and Robotic Systems, 2009
- [19] Fach, Martin: Navigation und Regelung eines Luftschiffes mittels optischer, inertialer und GPS Sensoren. Dissertation, Institut für Flugmechanik und Flugregelung, Universität Stuttgart, 2008
- [20] Krause, Thomas et al.: Flugregler für ein autonomes Luftschiff. Autonome Mobile Systeme, 2003
- [21] Sünderhauf, N., Lange, S., Protzel, P.: Using the Unscented Kalman Filter for Mono-SLAM with Inverse Depth Parametrization for Autonomous Airship Control. Control. Proc. of IEEE SSR, 2007
- [22] La Gloria, Nicola: Simultaneous localization and mapping applied to an airship with inertial navigation system and camera sensor fusion. PHD-thesis, University of Padova Italy, 2008
- [23] Mirisola, Luiz Gustavo Bizarro: Exploiting attitude sensing in vision-based navigation, mapping and tracking including results from an airship. Dissertation, Universidade de Coimbra Portugal, 2009
- [24] Bay, Herbert et al.: Speeded-up robust features (SURF). Computer vision and image understanding, 2008
- [25] Roskoni, Ulrich: Positionsbestimmung eines Flugsystems durch Vergleich laufender Luftbildaufnahmen mit georeferenzierten Aufnahmen unter Verwendung von Matlab. Masterarbeit, FernUniversität in Hagen, 2013
- [26] Lowe, David G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision, 2004
- [27] Yol, Aurelien et al.: Vision-based absolute localization for unmanned aerial vehicles. Intelligent Robots and Systems, 2014
- [28] Hartley R. I. and Zisserman A.: Multiple View Geometry in Computer Vision. Buch, Cambridge University Press, 2004
- [29] Azad, P. et al.: Computer Vision: Das Praxisbuch. Buch, Elektor Verlag, 2007
- [30] Ollero, A. and Maza, I.: Multiple Heterogeneous Unmanned Aerial Vehicles. Buch, Springer-Verlag, 2007
- [31] Sato, Jun: Recovering Multiple View Geometry from Mutual Projections of Multiple Cameras. International Journal of Computer Vision, 2006
- [32] Hartley R. and Zisserman A.: Multiple View Geometry in Computer Vision. Buch, Cambridge University Press, 2004

- [33] Prince, S. J. D.: Computer Vision: Models Learning and Inference. Buch, Cambridge University Press, 2012
- [34] Fusiello, Andrea et al.: A compact algorithm for rectification of stereo pairs. Machine Vision and Applications, 2000
- [35] Demant C. et al.: Industrielle Bildverarbeitung: Wie optische Qualitätskontrolle wirklich funktioniert. Buch, Springer, 2011
- [36] Schulz S. D.: Farbbildaufnahme und-wiedergabe unter Infraroteinfluss. Books on Demand GmbH, 2009
- [37] Czichos, Horst und Hennecke, Manfred: HÜTTE -- Das Ingenieurwissen. Buch, Springer, 2008
- [38] Goldman, Dan B and Chen, Jiun-Hung: Vignette and Exposure Calibration and Compensation. The 10th IEEE International Conference on Computer Vision, 2005
- [39] Jähne, Bernd: Digitale Bildverarbeitung. Buch, Springer Verlag Berlin Heidelberg New York, 2005
- [40] Schmidt, Ulrich: Professionelle Videotechnik: Analoge und digitale Grundlagen, Filmtechnik, Fernsehtechnik, HDTV, Kameras, Displays, Videorecorder, Produktion und Studiotchnik. Buch, Springer-Verlag New York Inc., 2007
- [41] Velho, L. et al.: Image processing for computer graphics and vision. Buch, Springer, 2009
- [42] Heikkila, Janne and Silven, Olli: A Four-step Camera Calibration Procedure with Implicit Image Correction. Computer Vision and Pattern Recognition IEEE Computer Society Conference, 1997
- [43] Heikkila, Janne: Geometric Camera Calibration Using Circular Control Points. IEEE Trans. Pattern Anal. Mach. Intell., 2000
- [44] Tsai, Roger Y: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal of Robotics and Automation, 1987
- [45] Zhang, Zhengyou: A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000
- [46] Remondino, F. and Fraser, C.: Digital camera calibration methods: considerations and comparisons. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 2006
- [47] Bouguet, Jean-Yves: Camera calibration toolbox for MATLAB. http://www.vision.caltech.edu/bouguetj/calib_doc, 2004
- [48] Horn, B. K. P. and Schunck, B. G.: Determining optical flow. Technical symposium east, International Society for Optics and Photonics, 1981
- [49] Helmholtz, H. von: Handbuch der Physiologischen Optik. Buch, Leipzig : Leopold Voss, 1867

- [50] Gibson, J. J.: The perception of the visual world. Buch, Houghton Mifflin, 1950
- [51] Srinivasan, M. V. et al.: Honeybee Navigation en route to the Goal: Visual Flight Control and Odometry. Journal of Experimental Biology, 1996
- [52] Verri, A. and Poggio, T.: Motion Field and Optical Flow: Qualitative Properties. IEEE Transactions on Pattern Analysis and Machine Intelligence,
- [53] Jähne, B. et al.: Handbook of Computer Vision and Applications. 2. Signal Processing and Pattern Recognition. Buch, Academic Press, 1999
- [54] Brown, Myron Z. et al.: Advances in Computational Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003
- [55] Fua, Pascal: A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine vision and applications, 1993
- [56] Barron, J. L. et al.: Performance of Optical Flow Techniques. International Journal of Computer Vision, 1994
- [57] Zitová, Barbara and Flusser, Jan: Image registration methods: a survey. Image and Vision Computing, 2003
- [58] Aschwanden, Peter Franz: Experimenteller Vergleich von Korrelationskriterien in der Bildanalyse. Dissertation, ETH Zürich, 1993
- [59] Giachetti, A.: Matching techniques to compute image motion. Image and Vision Computing, 2000
- [60] Heeger, David J.: Optical flow using spatiotemporal filters. International Journal of Computer Vision, 1988
- [61] Ho, Huy Tho and Goecke, Roland: Optical flow estimation using Fourier Mellin Transform. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008
- [62] Brown, Lisa Gottesfeld: A survey of image registration techniques. ACM Computing Surveys, 1992
- [63] Harris, C. and Stephens, M.: A Combined Corner and Edge Detector. Proceedings of the 4th Alvey Vision Conference, 1988
- [64] Zhang, Zhengyou et al.: A Robust Technique for Matching two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. Artificial Intelligence, 1995
- [65] Lucas, Bruce D. and Kanade, Takeo: An iterative image registration technique with an application to stereo vision. Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81), 1981
- [66] Neustaedter, Carman: An evaluation of optical flow using Lucas and Kanade's algorithm. University of Calgary Department of Computer Science, Canada, 2002
- [67] Weickert, Joachim et al.: Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision, 2005:

- [68] McCane, B. et al.: On Benchmarking Optical Flow. Computer Vision and Image Understanding, 2001
- [69] Galvin, B. et al.: Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms. British Machine Vision Conference, 1998
- [70] McCarthy, Chris and Barnes, Nick: Performance of Optical Flow Techniques for Indoor Navigation with a Mobile Robot. Proceedings of the IEEE International Conference on Robotics & Automation, 2004
- [71] Liu, Hongche et al.: Accuracy vs. Efficiency Trade-offs in Optical Flow. Algorithms. Computer vision and image understanding, 1998
- [72] Farneback, Gunnar: Very High Accuracy Velocity Estimation using Orientation Tensors, Parametric Motion, and Simultaneous Segmentation of the Motion Field. In Proc. Eighth International Conference on Computer Vision, 2001
- [73] McCarthy, Christopher D: Performance of Optical Flow Techniques for Mobile Robot Navigation. University of Melbourne, Department of Computer Science and Software Engineering, 2005.
- [74] Thota, Sri Devi et al.: Comparison Between The Optical Flow Computational Techniques. International Journal of Engineering Trends and Technology, 2013
- [75] Szeliski, Richard: Computer Vision : Algorithms and Applications. Buch, Springer-Verlag New York Inc, 2010
- [76] Bouguet, Jean-Yves : Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm. Technical Report, Intel Corporation Microprocessor Research Labs, 2001
- [77] Lewis, J. P.: Fast Normalized Cross-Correlation. Vision Interface, Canadian Image Processing and Pattern Recognition Society, 1995
- [78] Hii, A. J. H. et al.: Fast normalized cross correlation for motion tracking using basis functions. Computer methods and programs in biomedicine, 2006
- [79] Briechle, Kai and Hanebeck, Uwe D.: Template Matching using Fast Normalized Cross Correlation. Proceedings of SPIE Vol. 4387 AeroSense Symposium, 2001
- [80] Crow, Franklin C.: Summed-area tables for texture mapping. SIGGRAPH Comput. Graph., 1984
- [81] Wie, S. D. and Lai, S. H.: Fast Template Matching Based on Normalized Cross Correlation with Adaptive Multilevel Winner Update. IEEE Transactions on Image Processing, 2008
- [82] Fookes C. et al.: Multi-Spectral Stereo Image Matching using Mutual Information. Proceedings of the 3D Data Processing Visualization and Transmission, 2004

- [83] Singh, Sonam and Ganotra, Dinesh : Modifications in Normalized Cross Correlation Expression for Template Matching Applications. International conference on Image Processing, Computer Vision & Pattern Recognition, 2011
- [84] Hirschmuller, Heiko and Scharstein, Daniel: Evaluation of Stereo Matching Costs on Images with Radiometric Differences. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009
- [85] Seibold, Peter: Comparison of Two Methods for Determining the Optical Flow. Proceedings of the 18th International Conference on Process Control, 2011
- [86] Klappstein, Jens: Optical-Flow Based Detection of Moving Objects in Traffic Scenes. Diss., University Heidelberg, Germany, 2008
- [87] Frick, Anatol: Dichte Tiefenschätzung aus Bildpaaren unter Verwendung einer stückweise planaren Szenengeometrie. Diplomarbeit, Universität Kiel, 2008
- [88] Gockel, T.: Interaktive 3D-Modellerfassung mittels One-Shot-Musterprojektion und schneller Registrierung. Buch, Univ.-Verlag Karlsruhe, 2006
- [89] Tian, Qi and Huhns, Michael N.: Algorithms for subpixel registration. Comput. Vision Graph. Image Process., 1986
- [90] Mühlmann, Karsten et al.: Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. International Journal of Computer Vision, 2002
- [91] Lüke, H. D.: Signalübertragung: Einführung in die Theorie der Nachrichtenübertragungstechnik. Buch, Springer-Verlag, 1975
- [92] Nobach, H. and Honkanen, M.: Two-dimensional Gaussian regression for sub-pixel displacement estimation in particle image velocimetry or particle position estimation in particle tracking velocimetry. Experiments in Fluids, Springer journal, 2005
- [93] Shimizu, Masao and Okutomi, Masatoshi: Sub-Pixel Estimation Error Cancellation on Area-Based Matching. International Journal of Computer Vision, 2005
- [94] Shimizu, Masao and Okutomi, Masatoshi: Precise Sub-Pixel Estimation on Area-Based Matching. IEEE International Conference on Computer Vision, 2001:
- [95] Debella-Gilo, Misganu and Käab, Andreas: Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. Journal, Remote Sensing of Environment, 2011
- [96] Fore, L B: Reduction of peak-locking errors produced by Gaussian sub-pixel interpolation in cross-correlation digital particle image velocimetry. Journal, Measurement Science and Technology, 2010
- [97] Baum, Peter: Aufarbeitung eines Algorithmus zur Echtzeitberechnung des optischen Flusses und dessen Implementierung auf einen Digitalsignalprozessor. Masterarbeit, FernUniversität in Hagen, 2011

- [98] Fykse, Egil: Performance Comparison of GPU, DSP and FPGA implementations of image processing and computer vision algorithms in embedded systems. Masterarbeit, Institutt for elektronikk og telekommunikasjon, 2013
- [99] Kwon, Soon and Kim, Hyeon-woo and Ha, Jaewook and Lee, Jang-woo: Real-time Stereo for Embedded Vision System. Bulletin of Networking Computing Systems and Software, 2013
- [100] Gembris, Daniel et al.: Correlation analysis on GPU systems using NVIDIA's CUDA. Journal of real-time image processing, 2011
- [101] Wilhelmi, Erik: Adaption eines Verfahrens zur Berechnung des optischen Flusses mithilfe geeigneter Hardware unter Berücksichtigung von Echtzeit-Anforderungen. Masterarbeit, FernUniversität in Hagen, 2012
- [102] Nvidia Corporation: NVIDIA CUDA Compute Unified Device Architecture, Programming Guide., 2007
- [103] Yang, Zhiyi et al.: Parallel image processing based on CUDA. International Conference on Computer Science and Software Engineering, 2008
- [104] Mahajan, Virendra N.: Optical Imaging and Aberrations, Part II. Wave Diffraction Optics. Buch, SPIE Press Book, 2001
- [105] Mühler, Roland: Physikalische und technische Grundlagen der Fotografie. Vorlesungsskript, Universität Magdeburg, 2000
- [106] Niedrig, Heinz et Al.: Lehrbuch der Experimentalphysik: Optik : Wellen- und Teilchenoptik. Buch, Walter de Gruyter, 2004
- [107] International Organization for Standardization: ISO 12233 Photography -- Electronic still-picture cameras -- Resolution measurements. International Organization for Standardization, 2000
- [108] Ludwig, Michael: Test: Canon PowerShot SX210 IS (Digitalkamera) . Zeitschrift CHIP, 2010
- [109] Bigun, Josef: Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision. Buch, Springer-Verlag New York Inc., 2005
- [110] Seibold, Peter: Gaussian Filter, Determination of integer parameters. <http://www.mathworks.com/matlabcentral/fileexchange/28702-gaussian-filter-determination-of-integer-parameters>, 2010

Lebenslauf

- 21.6.1948 Geboren in Balingen, Baden Württemberg
- 1949 - 1958 Aufgewachsen in Frankreich
- 1968 Abitur, Couvengymnasium Aachen
- 1975 Abschluss als Diplom-Ingenieur, Fachrichtung Allgemeine Elektrotechnik, RWTH Aachen.
- 1975 - 1976 TH Aachen, Institut für Nachrichtengeräte und Datenverarbeitung
Testmustererstellung für digitale Schaltungen.
Wissenschaftliche Hilfskraft und später wissenschaftlicher Mitarbeiter.
- 1976 - 2001 Bosch (davon 4 J. in den USA), Nokia und andere Firmen.
Rundfunk-/Fernsehempfänger, Navigationsgeräte, ATM-Switches.
Als Sachbearbeiter, Gruppenleiter, Abteilungsleiter, Entwicklungsleiter.
Zahlreiche Patente auf dem Gebiet der Rundfunktechnik.
- 2001 - 2007 Selbstständig
Ingenieurbüro und Softwareunternehmen.
Ehrenamtliche Tätigkeiten.
- Ab 2008 FernUniversität in Hagen
Forschung auf dem Gebiet der Bildverarbeitung.

Online-Buchshop für Ingenieure

■ ■ VDI nachrichten

BUCHSHOP

Online-Shops



**Fachliteratur und mehr -
jetzt bequem online recher-
chieren & bestellen unter:
www.vdi-nachrichten.com/
Der-Shop-im-Ueberblick**



**Täglich aktualisiert:
Neuerscheinungen
VDI-Schriftenreihen**



Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

VDI-Berichte und die Reihe **Kunststofftechnik**:

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

Fortschritt-Berichte VDI:

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrichtungen

Newsletter „Neuerscheinungen“:

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriftenreihen bequem per E-Mail

Autoren-Service:

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

Buch- und Medien-Service:

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

VDI nachrichten

BUCHSHOP

www.vdi-nachrichten.com/Der-Shop-im-Ueberblick

Die Reihen der Fortschritt-Berichte VDI:

- 1 Konstruktionstechnik/Maschinenelemente
 - 2 Fertigungstechnik
 - 3 Verfahrenstechnik
 - 4 Bauingenieurwesen
- 5 Grund- und Werkstoffe/Kunststoffe
 - 6 Energietechnik
 - 7 Strömungstechnik
- 8 Mess-, Steuerungs- und Regelungstechnik
 - 9 Elektronik/Mikro- und Nanotechnik
 - 10 Informatik/Kommunikation
 - 11 Schwingungstechnik
- 12 Verkehrstechnik/Fahrzeugtechnik
 - 13 Fördertechnik/Logistik
- 14 Landtechnik/Lebensmitteltechnik
 - 15 Umwelttechnik
 - 16 Technik und Wirtschaft
- 17 Biotechnik/Medizintechnik
- 18 Mechanik/Bruchmechanik
- 19 Wärmetechnik/Kältetechnik
- 20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
 - 21 Elektrotechnik
 - 22 Mensch-Maschine-Systeme
- 23 Technische Gebäudeausrüstung

ISBN 978-3-18-385310-6