FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Eine Untersuchung zur Wichtigkeit von Farbe in Stereo-Matching

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

im Rahmen des Studiums

Medieninformatik

eingereicht von

Alexander Haßlacher
Matrikelnummer 0526452

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung
Betreuer/in: ao. Univ. Prof. Mag. DI Dr. Margrit Gelautz
Mitwirkung: Univ.-Ass. Mag. DI Dr. Christoph Rhemann
Univ.-Ass. DI Dr. Michael Bleyer

Wien, 02.03.2012

(Unterschrift Verfasser) (Unterschrift Betreuerin)

Erklärung zur Verfassung der Arbeit

Haßlacher Alexander
Arthaberplatz 10 / Tür 11
A-1100 Wien

Hiermit erkläre ich, Ing. Haßlacher Alexander, BSc., dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, am 02. März 2012

Unterschrift

Kurzfassung

Ziel von Stereo-Algorithmen ist es, korrespondierende Punkte in zwei, auf der horizontalen Ebene verschobenen Bildern zu finden. Dabei wird die Verschiebung von korrespondierenden Pixeln als Disparität bezeichnet. Die Disparitäten geben Rückschluss auf die Entfernung der Bildpunkte zum Kamerasensor, wodurch sich Tiefenkarten erstellen lassen.

Obwohl die Verwendung von Farbbildern als Eingabe für Stereo-Algorithmen eine höhere Qualität der Resultate als bei Grauwertbildern erwarten lässt, wurde in vergangenen Arbeiten gezeigt, dass Farbbilder bei Verwendung von helligkeits-invarianten Stereo-Korrespondenzmaßen sogar einen negativen Einfluss auf die Resultate haben können. Jedoch wurden diese Experimente auf Bildern durchgeführt, welche unter ähnlichen Laborbedingungen und mit derselben Kamera aufgenommen wurden. Daher stellt sich die Frage, ob diese Erkenntnisse auch für andere Datensätze Gültigkeit haben.

Das Ziel dieser Diplomarbeit ist es, die Einsichten vergangener Arbeiten auf einem neuen Datensatz und mit unterschiedlichen Kameras zu verifizieren. Dazu wurden in dieser Arbeit Stereobildpaare mit unterschiedlichen Kameras aufgenommen und Referenzlösungen in Form von Disparitätskarten unter Einsatz von strukturiertem Licht erzeugt.

Anschließend wurden die aufgenommenen Stereobildpaare verwendet, um Disparitätskarten mit einem etablierten Stereo-Algorithmus zu berechnen. Zur Berechnung der Disparitätskarten wurden verschiedene Farbräume sowie Korrespondenzmaße verwendet. Die berechneten Disparitätskarten wurden dann mit den Referenzlösungen verglichen und die daraus erhaltenen Fehlerraten ausgewertet.

Die Auswertungen bestätigen die Einsichten vergangener Arbeiten: Zum einen hat Farbe in radiometrisch unverzerrten Bildregionen wenig positiven Einfluss auf die Resultate. Im Gegensatz dazu können radiometrisch insensitive Korrespondenzmaße auch in diesen Regionen das Resultat verbessern. Zum anderen wurde in dieser Arbeit bestätigt, dass radiometrisch insensitive Korrespondenzmaße auf Grauwertbildern bessere Ergebnisse liefern als auf Farbbildern.

Abstract

The aim of stereo-algorithms is to find corresponding points in two horizontally displaced images. The displacement between these corresponding pixels is called disparity. The disparities vary over the image according to the scene depth, which makes it possible to infer depth maps.

When using color images as input for stereo algorithms one would expect higher quality results than when using gray-scale images. However, previous experimental studies found that the use of color images together with brightness-invariant stereo correspondence measures can even have a negative influence on the results. However, these experiments were conducted on a dataset that comprises of images that where captured with the same camera under similar lighting conditions. This raises the question whether these findings can be directly applied to images obtained with different cameras and illumination conditions.

Therefore, the aim of this master thesis is to verify the insights of previous works through a new dataset that comprises of images recorded with multiple different cameras. To this end, we first captured stereo image pairs with a variety of different cameras. For the captured stereo pairs we computed reference solutions, in the form of disparity maps, by using a structured light technique.

Afterwards, the captured stereo image pairs were used to compute disparity maps with a state-of-the-art stereo matching algorithm. The disparity maps were computed in different color spaces and a variety of correspondence measures were used. Finally, these disparity maps were compared with the reference solutions to obtain error rates.

The analysis of the error rates confirms the major insights of previous works: using color in radiometric undistorted image regions only has a minor positive influence on the results. In contrast, radiometric insensitive correspondence measurements can improve the result in these regions. Furthermore, this work confirms that radiometric insensitive correspondence measurements applied on gray-scale images provide better results than if applied on color images.

Inhaltsverzeichnis

1. EINLEITUNG 1

1.1 MOTIVATION 1
1.2 BEITRAG 2
1.3 METHODISCHES VORGEHEN 2
1.4 ORGANISATION 3

2. GRUNDLAGEN 4

2.1 STEREO-MATCHING – EINE ÜBERSICHT 4
2.1.1 Korrespondenzproblem 5
2.1.2 Epipolargeometrie 7
2.1.3 Rektifizierung 9
2.1.4 Triangulation 12
2.1.5 Stereo-Matching 13
2.2 STRUKTURIERTES LICHT 21
2.2.1 Streifenprojektionsmethode 22
2.2.2 Binäre Streifenprojektion 23
2.2.3 Graycode-Muster Methode 24
2.2.4 Phasen-Shift Methode 25
2.2.5 Line-Shift Methode 27
2.2.6 Space-Time Methode 28
2.3 DIE ROLLE VON FARBE BEIM STEREO-MATCHING 29
2.3.1 Radiometrisch verzerrte Bildregionen 31
2.3.2 Absolute Farbdifferenzen 33
2.3.3 Mutual Information 34
2.3.4 Normalized Cross-Correlation 35
2.3.5 Census 35

3. IMPLEMENTIERUNG 37

3.1 ERSTELLUNG VON STEREO-TIEFENKARTEN MIT STRUKTURIERTEM LICHT ... 37
3.1.1 Erstellung der Eingabedaten 38
3.1.2 Konvertierung 39
3.1.3 Versuche mit mehreren Kameras 40
3.1.4 Rektifizierung 41
3.1.5 Gray-Codes 42
3.1.6 Binarisierung 43
3.1.7 Berechnung der Sichtunterschiede (View Disparities) 45
3.1.8 Berechnung der Beleuchtungsunterschiede (Illumination Disparities) ... 47
3.1.9 Kombination der Disparitätskarten 51
3.2 VORGEHENSWEISE ZUR DATENEVALUIERUNG 52

4. EXPERIMENTE UND RESULTATE 54

4.1 BERECHNUNG DER DISPARITÄTSKARTEN 54
4.2 QUALITATIVE FEHLERMETRIKEN 55
4.3 VISUALISIEREN VON FEHLERN 56
4.4 FARBBASIERTE MESSUNGEN 57
4.5 RADIOMETRISCH UNEMPFINDLICHE MESSUNGEN 59
4.6 VERWENDUNG UNTERSCHIEDLICHER KAMERAS 62

5. SCHLUSSFOLGERUNG 66

LITERATURVERZEICHNIS 68

ABBILDUNGSVERZEICHNIS 72

1. Einleitung

1.1 Motivation

Ziel von Stereo-Algorithmen ist es, korrespondierende Punkte in zwei, auf der horizontalen Ebene verschobenen Bildern zu finden. Dabei wird die Verschiebung von korrespondierenden Pixeln als Disparität bezeichnet. Über die Disparitäten jedes einzelnen Pixels lässt sich anschließend die Entfernung eines Objektes im Bild ermitteln und es entstehen Stereo-Tiefenkarten.

In den vergangenen Jahren kam es zu einem gestiegenen Interesse an Stereo-Korrespondenzalgorithmen. Dies liegt unter anderem an der Existenz von Datenbanken die einen Vergleich solcher Algorithmen mit Hilfe von Referenzlösungen ermöglichen. Zum Erstellen solcher Referenzlösungen kann beispielsweise strukturiertes Licht verwendet werden. Dabei werden korrespondierende Pixel in einem Bildpaar durch einen Binärcode eindeutig gekennzeichnet, sodass eine Übereinstimmungsberechnung in beiden Bildern automatisch erfolgen kann. Dieser Binärcode wird durch die Projektion von Lichtmustern auf die Szene erstellt. Die daraus entstehenden Korrespondenzen dienen schließlich als Tiefeninformationen, mit denen weitergearbeitet werden kann.

Zur Erstellung solcher Tiefenkarten und deren korrespondierenden Pixeln in natürlichen Bildern (ohne strukturiertem Licht) gibt es verschiedenste Methoden. Viele dieser Algorithmen arbeiten mit dem Vergleich von Farb- oder Helligkeitsmerkmalen der korrespondierenden Pixel. Mehrere Ausarbeitungen von [4, 7, 22, 27, 28] zeigen, dass Farbe die Leistung solcher Algorithmen gegenüber dem Vergleich von Graustufenbildern erhöht. In der Arbeit von [3] wurde dieses Ergebnis bestätigt. Jedoch wurde auch gezeigt, dass Farbinformationen bei Verwendung von Korrespondenzmaßen wie beispielsweise Mutual Information (MI), Normalized Cross-Correlation (NCC) oder Census sogar einen negativen Einfluss haben können.

1.2 Beitrag

In dieser Arbeit wird mit „real“ aufgenommenen Bildern unter sich ändernden Bedingungen gearbeitet. Die Resultate in [3] wurden auf Bildern erzielt, welche unter ähnlichen Bedingungen (Laborbedingungen) und mit derselben Kamera aufgenommen wurden. Um die Ergebnisse von [3] zu überprüfen, soll die Auswertung von [3] nachimplementiert werden. Des Weiteren werden die Test-Sets mit mehreren unterschiedlichen Kameras aufgenommen, um die Aussagen nicht auf eine einzige Kamera zu reduzieren und zu prüfen, ob auch eine Abhängigkeit in Bezug auf den Kamerasensor besteht.

Das Hauptziel, das sich diese Arbeit stellt, ist es nun zu überprüfen, ob die Ergebnisse, welche in der Arbeit von [3] gewonnenen wurden, von der Kamera abhängen oder ob diese allgemein gültig sind.

1.3 Methodisches Vorgehen

Um den Einfluss von Farbinformationen zu testen, soll die Qualität von verschiedenen Korrespondenzmaßen (SAD, MI, NCC, Census) mit und ohne Farbinformation ermittelt werden. Zum Ermitteln dieser Informationen werden Testdaten und Referenzlösungen benötigt. Um die Referenzlösungen zu erstellen, kommen Algorithmen zum Einsatz, welche Tiefenkarten über strukturiertes Licht erstellen [10, 19, 32, 33].

Um diese Tiefenkarten automatisch erstellen zu können, wird eine Software benötigt, welche die Aufnahme der Bildpaare durchführen kann. Eine solche Software wurde teilimplementiert vom Institut zur Verfügung gestellt und auf die benötigten Anforderungen erweitert. Des Weiteren wurde ein Programm implementiert, welches die Erstellung von Stereo-Tiefenkarten mittels strukturiertem Licht nach dem Paper von [33] ermöglicht.

Sind die Disparitätskarten vorhanden, kann durch eine weitere, vom Institut zur Verfügung gestellte Software, die Auswertung und der Vergleich verschiedenster Korrespondenzmaße mit und ohne Farbinformationen berechnet werden.

1.4 Organisation

Der Aufbau dieser Arbeit gliedert sich in vier Hauptteile. In Abschnitt 2 werden Grundlagen aufgeführt, welche sich in die Bereiche „Stereo-Matching – eine Übersicht“, „Strukturiertes Licht“ und „Die Rolle von Farbe beim Stereo-Matching“ gliedern. Abschnitt 3 beschreibt die Implementation zur Durchführung der Auswertung. Dabei wird in Abschnitt 3.1 genauer auf die Erstellung von Stereo-Tiefenkarten mit strukturiertem Licht eingegangen und in Kapitel 3.2 die Vorgehensweise zur Datenevaluierung besprochen. Abschnitt 4 zeigt abschließend die Resultate der Auswertungen.

2. Grundlagen

2.1 Stereo-Matching – eine Übersicht

Unter Stereo-Matching versteht man das Erstellen von Korrespondenzpunkten in einem Bildpaar, dessen zwei Ausgangsbilder (Stereobilder) sich durch eine pixelweise horizontale Verschiebung auf der X-Achse unterscheiden. Die Annahme einer rein horizontalen Verschiebung setzt das Vorliegen von epipolar rektifizierten Bildern voraus (siehe Abschnitt 2.1.3). Um korrespondierende Bildpunkte zwischen diesen zwei Ausgangsbildern finden zu können, sind unterschiedlichste Algorithmen im Umlauf, deren Anzahl jährlich steigt. Dabei liegt das Ziel dieser Stereo-Matching Algorithmen darin, horizontal verschobene, korrespondierende Bildpunkte in den zwei Stereobildern zu finden. Dies wird als Disparität bezeichnet.

Dabei beschreibt die Disparität den Betrag der Verschiebung zwischen einem Pixel im linken Bild zu dessen korrespondierenden Pixel im rechten Bild und ist invers proportional zur Tiefe einer Szene. Unter invers proportional versteht man, dass Korrespondenzen von weiter entfernten Objekten (beispielsweise dem Hintergrund) kleinere Disparitätswerte produzieren und, je näher diese Objekte in den Vordergrund rücken, sich auch deren Disparitätswerte erhöhen. Die Berechnung aller Korrespondenzen für die einzelnen Pixel im linken und rechten Stereobild ergibt eine Disparitätskarte, deren Pixel in Grauwerten codiert sind. Je höher der Grauwert ist, desto näher ist das Objekt an der Kamera und umgekehrt. Bei vielen Stereo-Matching Algorithmen wird eine Disparitätskarte $d(x,y)$ in einem Disparitätsraum (x,y,d) erstellt. Durch die Operation auf zwei Kamerabildern und die Berechnung der Disparitäten für jeden einzelnen Pixel kann eine kompakte Stereo-Tiefenkarte für die linke und/oder rechte Seite eines Stereobildpaares erstellt werden, deren Pixelkoordinaten (x,y) mit den Pixelkoordinaten des jeweiligen Referenzbildes übereinstimmen. Die Disparitätskarte ermöglicht somit einen direkten Bezug zwischen linkem und rechtem Stereobild.

Die Erstellung von Disparitätskarten unter Verwendung unterschiedlichster Algorithmen ist als Stereo-Matching Problem oder Korrespondenzproblem (Abschnitt 2.1.1) bekannt. Dabei liegt das Augenmerk darauf, möglichst viele und

möglichst korrekte Korrespondenzen zwischen den Pixeln der Eingabebilder herzustellen.

Da der Bereich Stereo-Matching ein vielerforschtes und sehr umfangreiches Gebiet darstellt, wird in dieser Arbeit lediglich auf grundlegende Ansatzweisen eingegangen. Dabei beschreiben Abschnitt 2.1.1 – 2.1.4 die für das Stereo-Matching essentiellen Thematiken Korrespondenzproblem, Epipolargeometrie, Rektifizierung sowie Triangulation. Abschnitt 2.1.5 befasst sich mit der Lösung des Korrespondenzproblems, dem Stereo-Matching und den grundlegenden Ansatzweisen.

2.1.1 Korrespondenzproblem

Die Aufnahme eines Bildes lässt sich vereinfacht als das Abbilden einer 3D-Szene in ein zweidimensionales Szenenbild beschreiben, was in Abbildung 2.1 gezeigt wird. Dabei stellt das Objektiv der Kamera das Projektionszentrum und somit den Fokuspunkt der Aufnahme dar. Geometrisch lässt sich dies durch eine Lochkamera veranschaulichen, über deren Fokuspunkt die Aufnahme der 3D-Szene auf ein zweidimensionales Bild (Bildebene) abgebildet wird. Ausgehend vom Projektionszentrum der Kamera lassen sich theoretisch unendlich viele (virtuelle) Strahlen auf die 3D-Szene projizieren, deren Schnittpunkte auf der Bildebene das 2D-Bild ergeben. In der Praxis beschränkt sich die Anzahl der virtuell abgebildeten Strahlen auf die Pixelauflösung der Kamera.

Die umgekehrte Prozedur ist nicht mehr ganz so trivial. Um aus dem 2D-Raum auf den 3D-Raum schließen zu können, bedarf es zumindest zweier 2D-Szenenbilder, welche aus unterschiedlichen Positionen aufgenommen wurden. Diese werden in den weiteren Abschnitten als linkes Bild (L) und rechtes Bild (R) bezeichnet. Schießt

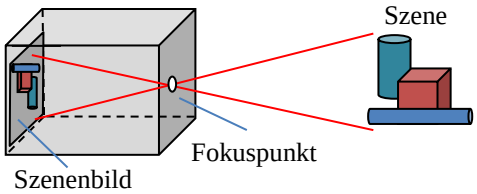


Abbildung 2.1: Aufnahme eines Bildes
Schematische, vereinfachte Darstellung der Aufnahme eines Bildes mittels Lochkamera. Die Lichtstrahlen der Szene kreuzen sich im Fokuspunkt (Projektionszenturm) der Kamera und erzeugen ein 2D-Szenenbild. Abbildung aus [44] nachgebildet.

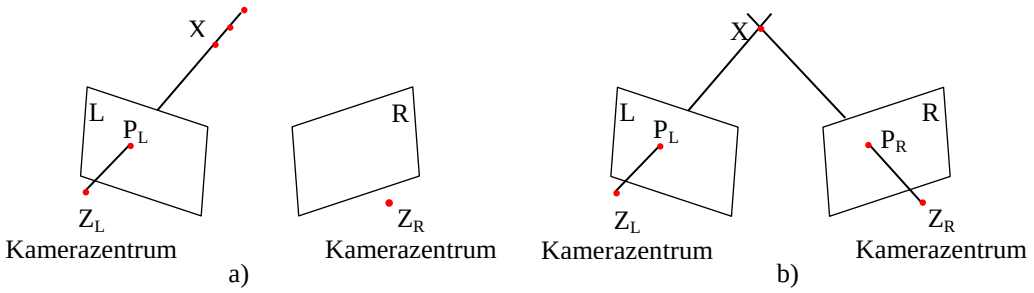


Abbildung 2.2: Das Korrespondenzproblem
a) Anlegen einer Geraden ($Z_L P_L$) zwischen Kamerazentrum Z_L und Szenenpunkt P_L. Auf dieser Geraden liegt der zu findende 3D-Szenenpunkt X. b) Finden des 3D-Szenenpunktes X bei bekanntem korrespondierenden Szenenpunkt P_R durch Schneiden der Geraden $Z_L P_L$ und $Z_R P_R$. Abbildung aus [44] nachgebildet.

man nun beispielsweise vom linken Szenenbild, ausgehend vom linken Projektionszentrum (Z_L), virtuelle Strahlen für jedes einzelne Pixel dieser 2D-Szene (L) in den Raum, so liegen die Raumpunkte der 3D-Objekte auf diesen Projektionsstrahlen. Abbildung 2.2 illustriert die Situation. Betrachtet man einen einzigen Strahl, welcher durch einen Szenenpixel P_L in der linken Szene geht, so muss der 3D-Szenenpunkt X auf diesem Strahl ($Z_L P_L$) liegen. Dabei ist allerdings nicht bekannt, wo genau auf diesem Strahl der Punkt X zu finden ist (Abbildung 2.2a). Um also den 3D-Szenenpunkt finden zu können, muss die projizierte Gerade aus dem linken Bild (L) mit einer äquivalenten projizierten Geraden aus dem rechten Bild (R) geschnitten werden. Angenommen, die beiden korrespondierenden Pixelpositionen von linkem und rechtem 2D-Szenenbild sind bekannt, so lässt sich dieser Schnittpunkt eindeutig bestimmen (Abbildung 2.2b).

Die Problematik besteht nun darin, dass zu einem Szenenpunkt P_L dessen korrespondierender Szenenpunkt P_R gefunden werden muss, wodurch sich wiederum der 3D-Szenenpunkt X errechnen lässt (Abbildung 2.3a). Diese Prozedur wird als Stereo-Matching Problem oder Korrespondenzproblem bezeichnet und beschreibt die Problematik der zweidimensionalen Suche von korrespondierenden Pixeln in einem Stereo Bildpaar. Um diese Suche zu vereinfachen und den Aufwand zu verringern, lässt sich das Korrespondenzproblem mittels der Epipolargeometrie der Szene beschreiben.

2.1.2 Epipolargeometrie

Die Epipolargeometrie hilft, das Korrespondenzproblem zu beschreiben und korrespondierende Bildpunkte zu finden. Dabei unterliegt sie einer fixen Terminologie, wobei die Positionen der Projektionszentren nicht bekannt sein müssen. Wie bereits beschrieben, gibt es für einen Szenenpunkt P_L, durch welchen eine virtuelle Gerade ($Z_L P_L$) aufgezogen wird, mehrere Szenenpunkte P_R, welche diese Gerade schneiden (Abbildung 2.3a). Zieht man nun zwischen den beiden Kamerazentren Z_L und Z_R eine virtuelle Linie auf, so entstehen Ebenen für die jeweiligen Schnittpunkte auf der Projektionsgeraden $Z_L P_L$, welche als Epipolarebenen bezeichnet werden. Die Verbindungslinie zwischen den Kamerazentren wird Baseline genannt, um welche sich alle möglichen Epipolarebenen für eine Szene aufspannen, sofern sich die Kamerazentren nicht ändern. Der Schnittpunkt der Baseline zur Bildebene für die jeweilige Seite stellt die Projektion des Kamerazentrums der gegenüberliegenden Seite dar und wird Epipol (e) genannt. Epipole können sich auch außerhalb der Bildebene befinden. Verbindet man nun den Epipol mit dem Szenenpunkt P_L, so wird diese Gerade als Epipolarlinie für diesen Punkt bezeichnet und zeigt die Projektion des Strahls $Z_R P_R$ der gegenüberliegenden Seite. Die Epipolarlinie lässt sich auch für die rechten Szenenpunkte P_R aufziehen, wobei ersichtlich wird, dass alle möglichen Korrespondenzpunkte der rechten Seite auf dieser Epipolarlinie liegen (Abbildung 2.3b).

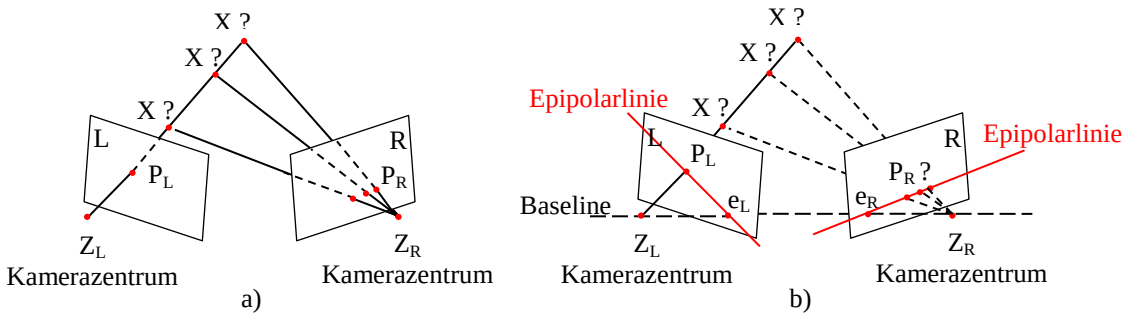


Abbildung 2.3: Epipolargeometrie
a) Für einen Szenenpunkt P_L mit aufgezogener Geraden durch das Kamerazentrum Z_L gibt es mehrere Schnittpunkte, welche aus den zu suchenden, korrespondierenden Szenenpunkten P_R resultieren. b) Aufgespannte Epipolarebenen für die Szenenpunkte P_R mit eingezeichneter Baseline, den zugehörigen Epipolen und Epipolarlinien. Die zu findenden Korrespondenzpunkte von P_R liegen alle auf derselben Epipolarlinie. Abbildung aus [14] nachgebildet.

Durch Ausnützung der Epipolargeometrie vereinfacht sich die zweidimensionale Suche im 2D-Szenenbild auf eine eindimensionale Suche, da lediglich die Epipolarlinie nach den korrespondierenden Bildpunkten abgesucht werden muss, was auch als Epipolarbedingung bezeichnet wird. Die Epipolarbedingung hat zum einen den Vorteil, dass der nun über die Epipolarlinie beschränkte Suchraum die Chance von falsch bestimmten Korrespondenzpunkten verringert, also die Qualität der Tiefenkarten erhöht, und zum anderen sich die Suche der korrekten Korrespondenzpunkte um eine Dimension verringert, was bei der Berechnung der Tiefenkarten (Disparitäten) einen großen Leistungsgewinn bringt. Deshalb sollte die Epipolarbedingung beim Stereo-Matching nach Möglichkeit herangezogen werden [44].

Eine ausführliche Erläuterung zur Epipolargeometrie und deren geometrischen Berechnungen findet sich im Buch „Multiple View Geometry in Computer Vision“ von Richard Hartley und Andrew Zisserman [14], Abschnitt 9.

Epipolare Rektifizierung

Die epipolare Rektifizierung stellt einen Spezialfall der Epipolargeometrie dar. Dabei liegen die beiden 2D-Bildeben *L* und *R* auf derselben horizontalen Achse und die Baseline damit parallel zur X-Achse. Die zugehörigen Epipolarlinien decken sich mit den horizontalen Zeilen (Scanlinien) im Bild (Abbildung 2.4a). Somit liegen die zu findenden, korrespondierenden Pixelwerte auf derselben Y-Koordinate im Bild und die Suche der Korrespondenzpunkte verringert sich wiederum auf eine zeilenweise, eindimensionale Suche auf der horizontalen Scanline (Abbildung 2.4b) [44].

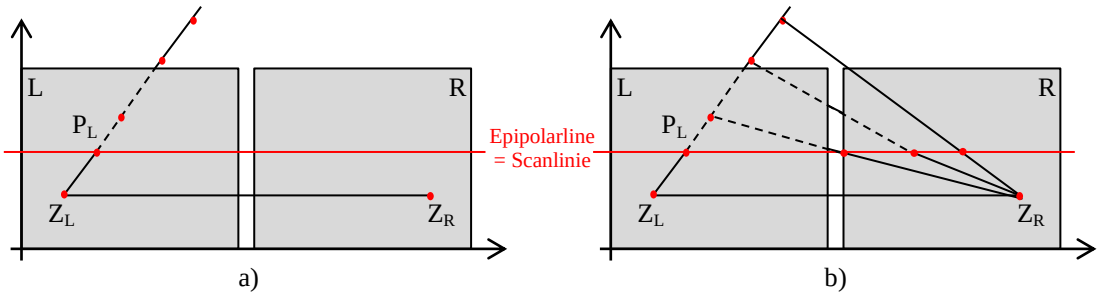


Abbildung 2.4: Epipolare Rektifizierung
a) Die beiden Szenenbilder sind auf horizontaler Achse (X) ausgerichtet. Die Epipolarlinien entsprechen den Scanlinien im Bild und liegen auf der Y-Achse der Bilder. b) Alle möglichen Korrespondenzpunkte für P_R liegen auf der Epipolarline und verringern den Suchaufwand auf eine zeilenweise, eindimensionale Suche. Abbildung aus [44] nachgebildet.

Die Rektifizierung von Stereobildern und verschiedene Möglichkeiten zur Durchführung der Rektifizierung werden im folgenden Abschnitt (2.1.3) genauer erläutert.

2.1.3 Rektifizierung

Unter Rektifizierung versteht man das gegenseitige Ausrichten von Bildpaaren zueinander, sodass Pixel einer Zeile im linken Bild den Pixeln derselben Zeile im rechten Bild entsprechen (epipolare Rektifizierung). Um die Bilder dementsprechend zu positionieren, also die Epipolarlinen achsenparallel zur X-Achse auszurichten, muss eine Transformation zwischen linkem (*L*) und rechtem (*R*) Bild durchgeführt werden (Abbildung 2.5).

Diese Transformation lässt sich mittels Parametern oder Interpolation durchführen. Sind interne (intrinsische) sowie externe (extrinsische) Kameraparameter wie beispielsweise die Brennweite oder Kameraorientierung bekannt, so spricht man von einem parametrischen Kalibrierungsverfahren. Sind diese Parameter nicht bekannt und somit die Kameras nicht kalibriert, muss man mit Pixelkoordinaten der 2D-Szenenbilder arbeiten, was als Interpolationsverfahren bezeichnet wird. Die Rektifizierung an sich unterliegt einer Transformation mittels einer bestimmten Transformationsmatrix, welche für parametrische Verfahren „essential matrix“ und für Interpolationsverfahren „fundamental matrix“ genannt wird.

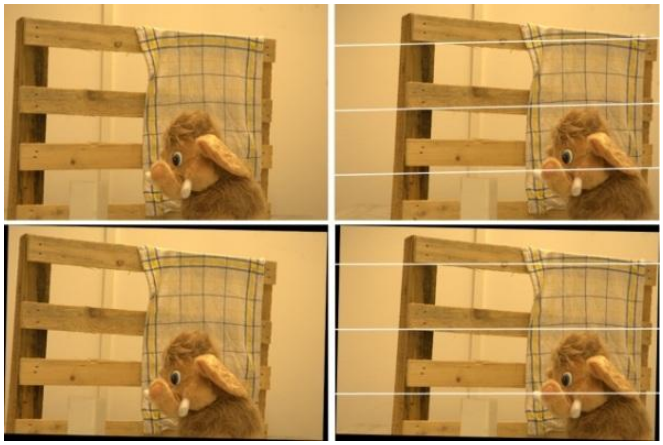


Abbildung 2.5: Prinzip der Rektifizierung
Im oberen Bereich ist ein unrektifiziertes Bildpaar mit eingezeichneten Epipolarlinien zu sehen, unten die rektifizierte Szene der beiden Bilder.

Zur Berechnung der Rektifizierungsparameter über die verschiedenen Transformationsmatrizen und die daraus resultierende epipolare Ausrichtung an die X-Koordinaten sind unterschiedliche Herangehensweisen möglich. Gute Referenzen zu parametrischen Verfahren sowie Interpolationsverfahren liefern die Arbeiten von Z. Zhang [39, 41]. Der Erhalt der Daten für die Berechnungsmatrizen wird im Folgenden kurz erläutert.

Parametrische Verfahren

Der Vorteil des parametrischen Verfahrens liegt darin, dass die Geometrie des Bildes modelliert wird und somit eine sehr hohe Genauigkeit erzielt werden kann. Für parametrische Verfahren gibt es wiederum verschiedene Herangehensweisen, welche sich in die Extrema photogrammetrische Kalibrierung sowie Selbstkalibrierung einteilen lassen. Während erstere sehr aufwändig ist und zusätzliche, teure Hardware benötigt, sind Selbstkalibrierungsverfahren zwar sehr flexibel, jedoch noch nicht ausreichend erforscht [39]. Viele andere Techniken siedeln sich zwischen diesen beiden Extremen an, wobei eine effiziente Methode von Z. Zhang [39] auf einfache Weise sehr gute Ergebnisse liefert.

Um die Parameter (intrinsische, extrinsische) der Kamera zur extrahieren, wird ein simples Schachbrettmuster herangezogen, welches beispielsweise über einen handelsüblichen Drucker ausgedruckt werden kann (Abbildung 2.6). Anschließend werden mindestens zwei Bilder (je mehr Bilder, desto genauer wird die Ausrichtung)

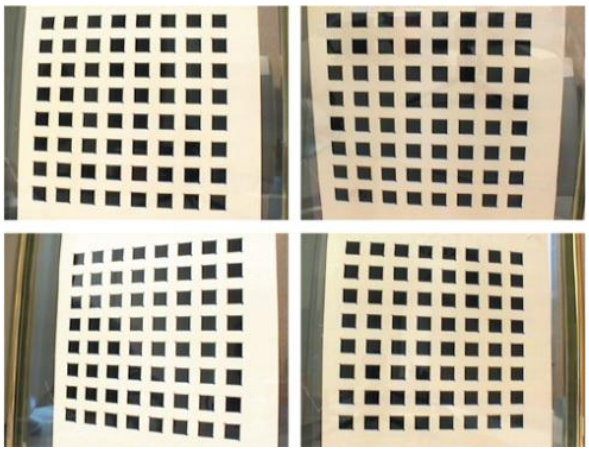


Abbildung 2.6: Parametrische Rektifizierung
Veranschaulichung der Schachbrettmuster und Detektion von Eckpunkten in diesen Mustern. Abbildung aus [39] entnommen.

dieses Musters in unterschiedlichen Orientierungen geschossen. Über eine Feature-Detektion lassen sich Feature-Punkte (Eckpunkte des Schachbrettmusters) detektieren und über die folgende Berechnung der Transformationsmatrix (essential matrix) die Rektifizierung der Stereobilder durchführen.

Verschiedenste Varianten dieser Methode sind im Umlauf, da sie flexibel ist und dazu gute Ergebnisse liefert. Diese Form der Rektifizierung ist auch in einigen Bildverarbeitungsprogrammen wie beispielsweise MATLAB® (Kamera Calibration Toolbox) der Firma MathWorks® bereits implementiert. Eine genaue Vorgehensweise zur Rektifizierung mittels parametrischer Verfahren lässt sich in der Arbeit von [39] nachlesen.

Interpolationsverfahren

Das Interpolationsverfahren basiert darauf, die Transformation über so genannte Passpunkte durchzuführen. Passpunkte sind markante Punkte, welche in beiden Bildern vorkommen (zumeist Ecken oder Kannten) und können beispielsweise über die Methoden von [40] aus den Stereobildpaaren extrahiert werden (Abbildung 2.7). Über die Passpunkte können dann wiederum die Kameraparameter berechnet werden und somit die Transformationsmatrix (fundamental matrix) erstellt werden. Aufgrund des Zusammenhanges zwischen mindestens sieben Passpunkten in den Bildern und der kalkulierten Transformationsmatrix lassen sich dann alle anderen Bildpunkte interpolieren und die Bildpaare gegenseitig ausrichten. Je mehr Passpunkte in den Bildern gefunden werden, desto genauer lässt sich die Rektifizierung der

Abbildung 2.7: Rektifizierung durch Interpolation
Visualisierung der passiven Rektifizierung. Rote Streifen zeigen die Verschiebung sowie die Ausrichtung zum zweiten Stereobild.

Ausgabebilder durchführen. Zur Berechnung der Rektifizierung über die fundamental matrix sind wiederum mehrere Varianten im Umlauf. Einige davon werden in der Arbeit von [40] beschrieben und ausgewertet.

Der Vorteil dieses Verfahrens ist, dass keinerlei Kenntnis über die Aufnahmegeometrie notwendig ist und dementsprechend auch die Kamera nicht explizit kalibriert werden muss. Allerdings weisen Interpolationsverfahren im Generellen eine geringere Genauigkeit gegenüber parametrischen Verfahren auf.

2.1.4 Triangulation

Um abschließend eine Rekonstruktion der 3D-Szene aus den Stereobildpaaren zu ermöglichen, kommt Triangulation zum Einsatz. Die Triangulation stellt eine sehr einfache Methode zur Bestimmung der Lage von Objekte in der 3D-Szene dar, sofern die korrekten Korrespondenzpunkte für die Stereobildpaare bekannt sind (Abbildung 2.8). Da die Berechnung der Lage von 3D-Szenenobjekten für die Auswertungen dieser Arbeit nicht relevant sind, soll dies hier nur kurz erörtert werden.

Durch Bildung von Ähnlichkeiten der Dreiecke lassen sich zwei Äquivalenzgleichungen aufstellen, über die die Berechnung der wahren Tiefe des 3D-Objektes im Raum (P_{3D}) möglich ist. Dabei besteht eine direkte Abhängigkeit zwischen der Tiefe Z und der Disparität d (invers proportional), welche sich über Strahlensätze herleiten lässt. Die beiden Dreiecke entsprechen:

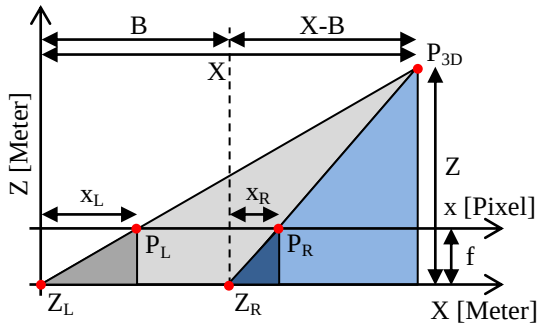


Abbildung 2.8: Prinzip der Triangulation
Schematische Darstellung des Triangulationsprinzips für einen 3D-Szenenpunkt. Abbildung aus [44] nachgebildet.

$$\frac{X}{Z} = \frac{x_L}{f} \quad \text{und} \quad \frac{X-B}{Z} = \frac{x_R}{f} \tag{2.1}$$

Z beschreibt die (zu berechnende) Tiefe in Metern und X die X-Koordinate in Metern. Beide sind nicht bekannt. B ist der Abstand der beiden Kamerazentren Z_L und Z_R zueinander und f die fokale Länge, also die Brennweite der Kamera. x_L und x_R sind die korrespondierenden, bekannten Bildkoordinaten des Kamerabildes. Durch Umformung der Gleichungen ergibt sich:

$$X = \frac{Z \cdot x_L}{f} \quad \text{und} \quad X = \frac{Z \cdot x_R}{f} + B \Rightarrow \frac{Z \cdot x_L}{f} = \frac{Z \cdot x_R}{f} + B \tag{2.2}$$

Somit fällt die Unbekannte X weg und Gleichung (2.2) lässt sich nach der Tiefe Z für jeden einzelnen Bildpunkt im Stereobildpaar bestimmen:

$$Z = \frac{B \cdot f}{x_L - x_R} = \frac{B \cdot f}{d} \tag{2.3}$$

Die Differenz von x_L und x_R ist die Disparität des betrachteten Punktes und kann als d angeschrieben werden. Durch diese einfache Umformung wird die Berechnung eines 3D-Szenenpunktes durchführbar und somit eine Rekonstruktion der gesamten Szene möglich [44].

2.1.5 Stereo-Matching

Wie bereits zu Beginn dieses Kapitels erörtert, ist eine Vielzahl an Stereo-Matching-Algorithmen zur Erstellung von Stereo-Tiefenkarten im Umlauf. Eine Teilproblematik zur Lösung des Stereo-Matching-Problems (Korrespondenz-problems, Abschnitt 2.1.1) stellen Verdeckungen dar. Außerdem müssen die Algorithmen gewisse Bedingungen erfüllen, wie in den folgenden Abschnitten näher erläutert wird.

Das Verdeckungsproblem

Das Verdeckungsproblem beschreibt die Problematik von Verdeckungen in Stereobildpaaren beim Finden von Korrespondenzpunkten. Eine Korrespondenz zwischen einem Pixel im linken Bild und einem Pixel im rechten Bild lässt sich nur berechnen, sofern der jeweilige Pixel in beiden Stereobildern sichtbar ist. Dies ist

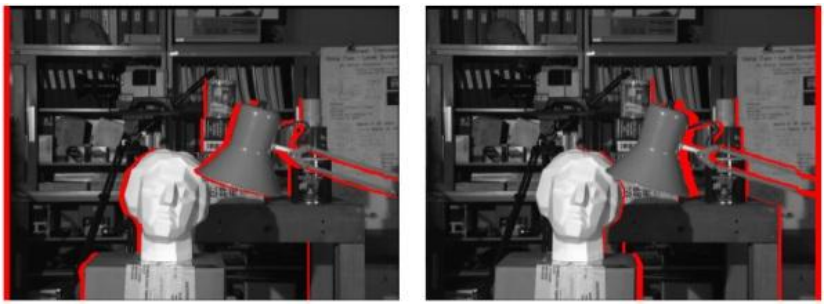


Abbildung 2.9: Verdeckungen
Verdeckungen (in rot eingezeichnet) in linken und rechten Stereobild. Abbildung aus [44] entnommen.

jedoch nicht immer gegeben, da für ein linkes Bild zum Teil Bildpunkte sichtbar sind, welche im rechten Bild nicht vorkommen, also von einem anderen Objekt verdeckt werden, und umgekehrt. Der Pixel ist somit nur in einem der beiden Stereobilder sichtbar, was es schwierig macht, eine korrekte Disparität für diesen Pixel zu berechnen. Verdeckungen entstehen nahe von Objektkanten, wie in Abbildung 2.9 gezeigt.

Abbildung 2.10 zeigt eine schematische Darstellung des Verdeckungsproblems. In Abbildung 2.10 a) ist der Punkt P_1 für beide Kameras sichtbar, was eine Korrespondenzfindung möglich macht. Abbildung 2.10 b) und c) zeigen verdeckte Bildpunkte einer Ansicht, für welche die Erstellung eines Disparitätswertes problematisch ist. P_2 bzw. P_3 wird nur von jeweils einer Kamera gesehen und ist in der zweiten Ansicht nicht vorhanden, da er von einem Vordergrundobjekt verdeckt wird. Hierbei wird von halbverdeckten Bereichen gesprochen. Diese Bereiche treten vor allen in der Nähe von Objektkanten auf (Abbildung 2.10 d).

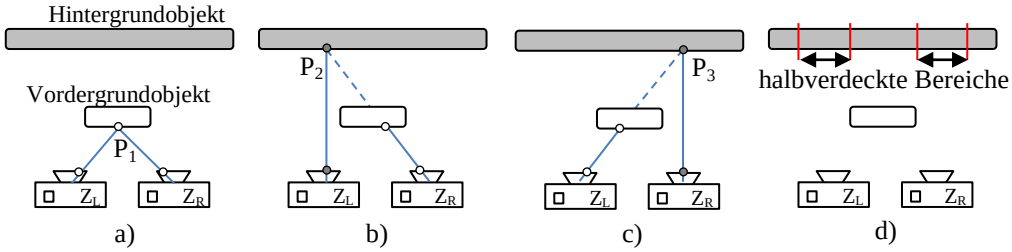


Abbildung 2.10: Das Verdeckungsproblem
Schematische Darstellung einer Szene mit einem Hintergrundobjekt und einem Vordergrundobjekt. a) Hier ist der Pixel P_1 in beiden Kameras sichtbar. b) Pixel P_2 wird nur von der linken Kamera gesehen, für die rechte Kamera ist dieser verdeckt. c) Pixel P_3 wird nur von der rechten Kamera gesehen, für die linke Kamera ist dieser verdeckt. d) Verdeckungen entstehen in der Nähe von Unstetigkeiten. Abbildung aus [44] nachgebildet.

Verdeckungen stellen kritische Problematiken dar und kommen in jeder komplexen Szene vor. Zur Lösung des Verdeckungsproblems gibt es wiederum unterschiedliche Herangehensweisen. Diese sind, je nach Verfahren zur Lösung des Stereo-Matching-Problems, methodenbezogen und unterschiedlich komplex.

Bedingungen

Zur Lösung des Stereo-Matching-Problems sind ein paar Bedingungen zu betrachten, welche die Lösung ermöglichen bzw. vereinfachen. Im Nachfolgenden sind die wichtigsten dieser Bedingungen kurz angegeben:

a) Eindeutigkeitsbedingung

Diese Bedingung hilft dabei, das Verdeckungsproblem besser handhabbar zu machen. Sie besagt, dass zu einem Pixel in einem der beiden Stereobilder es mindestens einen Korrespondenzpunkt im zweiten Bild geben muss, sofern keine Verdeckung vorliegt. Bei Abbildung 2.10 b) oder c) trifft diese Bedingung nicht zu, und somit kann davon ausgegangen werden, dass es sich um eine Verdeckung handelt.

b) Epipolarbedingung

Die Epipolarbedingung bezieht sich auf die Epipolargeometrie und besagt, dass Korrespondenzen in linken und rechten Bild auf derselben Scanlinie zu liegen haben. Diese Bedingung wurde in Abschnitt 2.1.3 bereits behandelt.

c) Photokonsistenzbedingung

Diese Bedingung sagt aus, dass korrespondierende Pixel typischerweise dieselbe Intensität bzw. Farbe haben müssen. Bei Bildrauschen, diversen Artefakten oder unterschiedlichen Belichtungen der beiden Stereobildpaare kann es zu Problemen kommen. Um diese Fälle zu handhaben, bedarf es einer weiteren Bedingung, der Ähnlichkeitsbedingung.

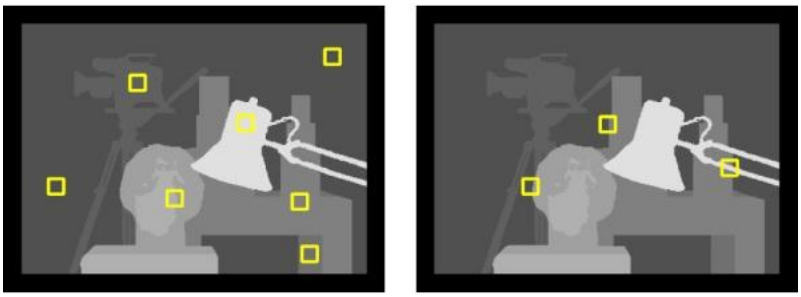
d) Ähnlichkeitsbedingung

Unter Berücksichtigung von Photokonsistenz- und Epipolarbedingung würde sich bereits ein Matching-Algorithmus erstellen lassen. Würde man nach dem Schema vorgehen, dass Korrespondenzen zwischen linkem und rechtem Bild auf derselben

Y-Achse zu finden sind (also eine epipolare Rektifizierung vorliegt) und die Pixel gleiche Farb- bzw. Intensitätswerte aufweisen, könnte man bereits eine Disparitätskarte durch einfache Suche der korrespondierenden Pixel erstellen. Allerdings berücksichtigt keine der beiden Bedingungen, dass gleiche oder ähnliche Farb- bzw. Intensitätswerte in einer Scanlinie mehrfach vorkommen können. Dies führt zwar zu richtigen Matches, jedoch sind diese an einer falschen Stelle und produzieren somit falsche Disparitäten. Um dies zu minimieren, bedarf es der Ähnlichkeitsbedingung.

Sie sagt aus, dass Pixel an ähnlichen Bildkoordinaten auch ähnliche bzw. gleiche Disparitätswerte aufweisen müssen. Betrachtet man die gelben Fenster in Abbildung 2.11 a), so geht die Bedingung davon aus, dass die Disparitätswerte innerhalb eines jeweiligen Fensters überall ähnlich bzw. gleich sind. Diese Bedingung trifft allerdings speziell in Randregionen der Disparitätskarte nicht zu und kann dort zum sogenannten „Foreground Fattening Problem“ führen (Abbildung 2.11 b). Auf diese Problematik wird etwas später in diesem Kapitel noch kurz eingegangen.

Unter Voraussetzung dieser vier Hauptbedingungen, und speziell der Ähnlichkeitsbedingung, lassen sich Algorithmen in lokale und globale Methoden einteilen.



a) b)

Abbildung 2.11: Ähnlichkeitsbedingung
a) zeigt Bereiche in der Disparitätskarte, welche in linker und rechter Karte gleich bzw. ähnlich sein müssen (im jeweiligen Fenster). In b) sind Bereiche eingezeichnet, wo die Bedingung nicht zutrifft. Abbildung aus [44] entnommen.

Globale und lokale Methoden

a) Lokale Methoden

Lokale Methoden werden zumeist auch fensterbasierte Methoden genannt und arbeiten mit Korrelation. Sie basieren darauf, Pixelfarbwerte in den Stereobildpaaren zu vergleichen. Dabei wird anstatt eines pixelweisen Vergleichs ein Fenster auf den zu vergleichenden Pixel gelegt, welches zentral zum Referenzpixel ausgerichtet wird. Dasselbe Fenster wird auch auf dem zweiten Bild des Paares positioniert, sodass die beiden Fenster korrelieren. Das Korrespondenzfenster wird anschließend entlang der horizontalen, korrespondierenden Epipolarlinie verschoben und für jede mögliche Position eine Minimierungsfunktion berechnet. Durch Bilden der absoluten Differenzen der einzelnen Pixel der beiden Fenster sowie der anschließenden Aufsummierung der Differenzfarbwerte lässt sich die Position des Pixels bestimmen, welches mit höchstmöglicher Wahrscheinlichkeit zum Suchpixel korrespondiert (Abbildung 2.12) und somit einen korrekten Match ergibt.

Das in Abbildung 2.12 dargestellte Verfahren benötigt allerdings sehr viel Rechenzeit, da für jede pixelweise Verschiebung die Summen berechnet werden müssen. Dabei spielt die Größe des Suchfensters ebenfalls eine bedeutende Rolle (siehe [44]). Eine Variante zur Lösung dieser Problematik sind so genannte „Sliding Windows“.

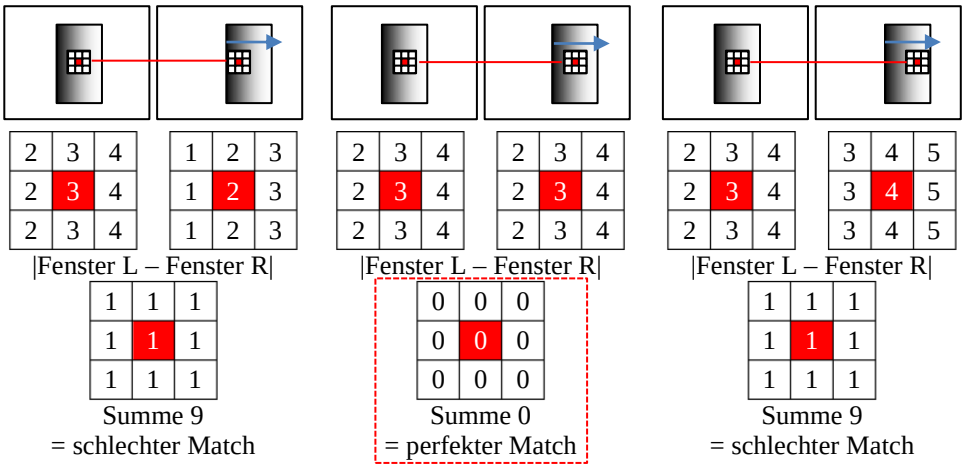


Abbildung 2.12: Fensterbasierte Methoden

Schematische Darstellung von lokalen oder fensterbasierten Methoden. Das Suchfenster wird über die Szene geschoben (horizontal) und die kleinste Summe kann als perfekter Match und somit als Korrespondenz zwischen linkem und rechtem Pixel gesehen werden.

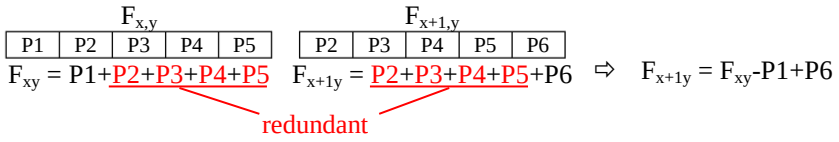


Abbildung 2.13: Sliding Windows

Prinzip der Sliding-Windows Technik. Da Pixelwerte beim Verschieben des Fensters redundant berechnet würden (aufwändig), lässt sich dieser Mehraufwand durch Umformung minimieren. Für quadratische Suchfenster sind zuerst die Zeilen und anschließend die Spalten zu berechnen [27]. Abbildung aus [44] nachgebildet.

Diese können den Rechenaufwand aufgrund von Redundanzgegebenheiten im Suchfenster minimieren und sind auch nicht von der Fenstergröße abhängig. Somit ermöglichen sie eine Echtzeitimplementierung dieser Algorithmen (Abbildung 2.13).

Fensterbasierte Methoden bringen allerdings auch einige Probleme mit sich. Matching-Fehler treten vor allem dort auf, wo es beispielsweise an Textur fehlt (untexturierte Regionen, Apertur-Problem). Dies betrifft zumeist Hintergründe wie einfarbige Wände und dergleichen. Weiters können Fehler bei sich wiederholenden Texturen auftreten (repetitive Bereiche). Um diese Probleme zu minimieren, ist die Größe der gewählten Fenster ein Anhaltspunkt. Je größer die Fenster gewählt werden, desto geringer werden die genannten Probleme. Große Fenster bringen allerdings ein weiteres Problem mit sich: das Foreground Fattening Problem.

Das Foreground Fattening Problem stellt eine Problematik bei zu groß gewählten Matching-Fenstern dar. Fensterbasierte Methoden nutzen die Ähnlichkeitsbedingung, die besagt, dass Pixeldisparitäten in einem Fenster ähnlich bzw. gleich sind. Im Bereich von Kanten in der Disparitätskarte trifft dies jedoch nicht zu. Da im

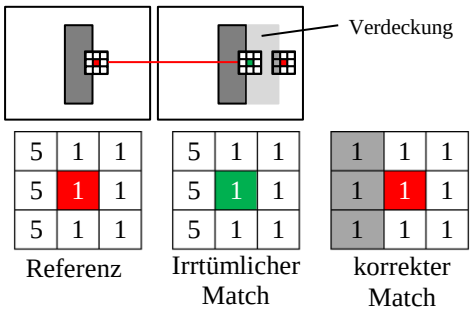


Abbildung 2.14: Das Foreground Fattening Problem

Schematische Darstellung eines irrtümlichen Matches in Randbereichen. Für den Algorithmus ist der Match im zweiten Bild korrekt (grün), jedoch aufgrund der Verdeckung eigentlich falsch, was falsche Disparitätswerte in diesem Bereich produziert.

Referenzbild Pixelwerte nicht sichtbar sind (also verdeckt sind), welche im zweiten Bild jedoch vorhanden sind, kommt es zu dieser Problematik. Der Algorithmus vergleicht die Datenwerte und kommt zu dem Schluss, dass der Match im zweiten Bild korrekt ist (Abbildung 2.14, grün), da die Summe des Matching-Fensters ähnlich bzw. gleich ist. In Wirklichkeit jedoch ist der korrekte Pixelwert horizontal weiter verschoben als der irrtümlich korrekte Match, da die Verdeckung miteinbezogen werden muss. Aus diesem Grund werden in diesen verdeckten Bereichen (Randbereichen) ohne Fehlerkorrektur falsche Disparitätswerte geliefert. Um dies zu minimieren, helfen wiederum kleine Matching-Fenstergrößen, welche jedoch die Probleme in untexturierten und repetitiven Bereichen verstärken [44].

Prinzipiell gibt es allerdings keine Fenstergröße, die beide Problemfälle optimal behandeln kann. Allerdings sind Methoden vorhanden, um diese Problematiken zu verringern. Adaptive Fenster (adaptive Windows) versuchen, die Vorteile von großen und kleinen Fenstern zu kombinieren und somit die Fehler zu minimieren. Das Prinzip besteht darin, ein passables Fenster mit individueller Größe für jeden Pixel zu erstellen. Fusiello [48] und Hirschmüller [17] stellen in ihren Arbeiten zwei gute Methoden vor, welche die Fensteradaption durchführen.

b) Globale Methoden

Globale Methoden basieren auf einer Kostenfunktion, welche die optimalen (minimalen) Kosten einer Disparitätskarte berechnet. Die Ähnlichkeitsbedingung ist nun nicht mehr fensterbezogen (wie bei den lokalen Methoden), sondern wird über einen Ähnlichkeitsterm definiert.

Die Kostenfunktion stellt die Kernberechnung von globalen Methoden dar. Sie versucht, eine Energieminimierung während der Berechnung der Disparitäten durchzuführen und somit den geringsten globalen Fehler zu finden. Die Definition der Kostenfunktion hat folgende Form:

$$E(D) = E_{data}(D) + E_{smooth}(D) \tag{2.4}$$

Die Funktion unterteilt sich wiederum in einen Datenterm, welcher unter Zuhilfenahme der Photokonsistenz-Bedingung die Farbunterschiede für jeden Pixel berechnet. Der zweite Term stellt den Ähnlichkeitsterm dar, dessen Aufgabe es ist,

Ähnlichkeiten in den Nachbarschaften der zu untersuchenden Pixel zu finden. *D* definiert in Gleichung (2.4) die Disparitätskarte des linken oder rechten Stereobildes. Die beiden Terme lassen sich typischerweise in folgender Form anschreiben:

$$E_{data}(D) = \sum_{p \in I} \rho(p,q)$$

$$E_{smooth}(D) = \sum_{(p,q) \in \mathcal{N}} s(d_p, d_q) \qquad (2.5)$$

Im Datenterm beschreibt *I* das Set aller Pixel in der jeweiligen Ansicht. $\rho(p,q)$ ist die Funktion zur Berechnung von Intensitäts- bzw. Farbunterschieden von Pixel *p* zu Pixel *q*. $\mathcal{N}$ steht im Ähnlichkeitsterm für ein Set von Nachbarschaften und $s(d_p,d_q)$ beschreibt die Ähnlichkeitsfunktion für die Disparitäten von *p* und *q*.

Zur Berechnung der Kostenfunktion steht eine breite Auswahl an Algorithmen zur Verfügung. Dabei sind dynamische Programmierung (zB. [5, 15]), Belief Propagation (zB. [9, 26]) oder Graph-Cuts (zB. [21, 23]) bei der Lösung des Energie-Minimierungsproblems behilflich, und Methoden wie die Berechnung von absoluten Differenzen, Mutual Information [18], Normalized Cross-Correlation und Census [38] helfen bei der Beschreibung des Datenterms. Letztere werden in den Abschnitten 2.3.2 bis 2.3.5 noch etwas genauer behandelt, da sie für die Auswertung in dieser Arbeit relevant sind.

Da der Bereich Stereo-Matching ein sehr weitreichendes und vielerforschtes Gebiet darstellt, lässt er sich, mit seinen vielseitigen Aspekten und Facetten, im Rahmen dieser Arbeit nur überblicksmäßig beschreiben. In diesem Kapitel wurde versucht, einen generellen Einblick in die Thematik zu geben und grundlegende Bereiche von Stereo-Matching zu beschreiben. Will man sich mit der Thematik genauer befassen, so kann die Webseite von [43] empfohlen werden, deren Projekte und Publikationen sich ausführlich mit der Thematik Stereo-Matching beschäftigen.

2.2 Strukturiertes Licht

Der Begriff strukturiertes Licht kommt aus dem Bereich Stereoskopie und beschreibt die Eigenschaften der Projektion von Licht auf eine Oberfläche. Projiziert man Licht auf eine dreidimensionale Oberfläche, so entstehen perspektivische Verzerrungen an Objekten, welche mit Hilfe von Triangulationstechniken rückgerechnet und so Tiefeninformationen aus einem zweidimensionalen Stereobild erhalten werden können.

Lichtprojektionstechniken gehören nicht zur typischen Art von Stereo-Matching Algorithmen. Sie stellen eine andere Methodik vor, wie das Korrespondenzproblem effizient gelöst werden kann. Prinzipiell gibt es verschiedene Ansätze der Projektion von Licht. Die Projektion von statischen Mustern wie beispielsweise Gitter [20] oder Punkte [25] sowie Lichtschnittverfahren [35] sind nur einige dieser Varianten. Die Projektion von parallelen Streifenmustern stellt eine weitere, weitverbreitete Variante dar und ist auch unter der Bezeichnung des codierten Lichtansatzes [36] bekannt. Sie ermöglicht es, eine relativ einfache und exakte Auffindung von 3D-Koordinaten eines Objektes über dessen Oberfläche zu erhalten und ist deshalb in häufiger Verwendung. In den Abschnitten 2.2.1 bis 2.2.6 soll der codierte Lichtansatz und einige Varianten dazu erörtert werden.

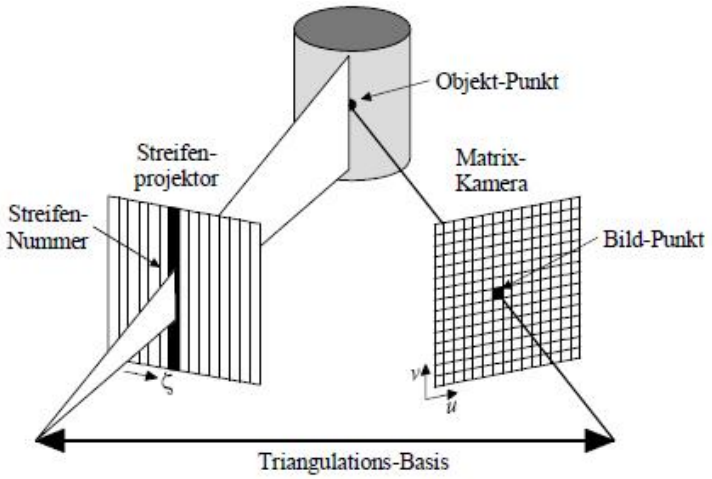


Abbildung 2.15: Prinzip des Streifenprojektionssensors
Projektion des Streifenmusters auf die Szene und Abfotografieren der Szene mit einer Kamera. Abbildung aus [37], S. 10 entnommen.

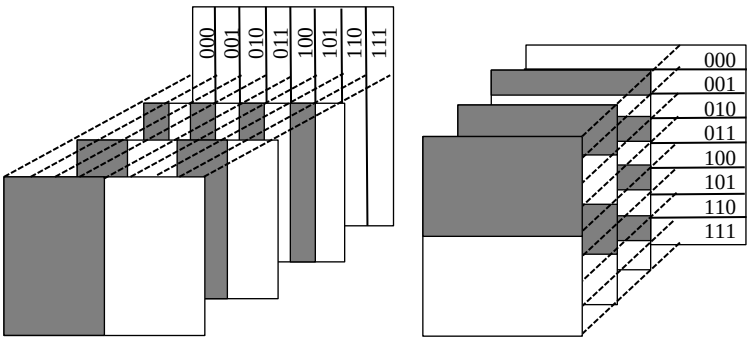


Abbildung 2.16: Streifenprojektion mit Binärmustern
Veranschaulichung der Streifenprojektion mittels binären Codemusters. Links ein vertikales Muster mit zugehörigem Spaltencode, rechts ein horizontales Muster mit zugehörigem Zeilencode. Die Kombination dieser sechs Bit Codewerte ergibt theoretisch einen eindeutigen Code für jeden Pixel.

2.2.1 Streifenprojektionsmethode

Der Projektionsmethode liegt das Beleuchten einer Szene mittels unterschiedlichen Streifenmustern zu Grunde. Dabei werden über einen Videoprojektor einfache Streifen auf eine aufzunehmende Szene projiziert und mit einer Kamera abfotografiert (Abbildung 2.15).

Für jedes unterschiedliche Streifenmuster muss jeweils ein Bild gemacht werden, woraus sich eine sequentielle Folge von unterschiedlichen Helligkeitswerten für einen Bildpunkt in der Szene ergibt (Abbildung 2.16).

Ein Streifenmuster besteht prinzipiell aus parallelen, sich in der Breite ändernden Streifen, die eine hohe hell-dunkel Grenze aufweisen (meist schwarz-weiß Muster). Um eine eindeutige Codierung eines Bildpunktes zu erhalten, müssen Streifenmuster nicht nur in horizontaler, sondern auch in vertikaler Richtung projiziert und aufgenommen werden. Würde man eine Projektion nun wie ein Binärmuster (Abbildung 2.17) auf eine Szene projizieren und jeweils ein Bild machen, könnte man über die Kombination aller Bilder und somit über verschiedene Helligkeitswerte, jedem einzelnen Bildpunkt in dieser Szene einen eindeutigen Code zuordnen und somit dessen Position eindeutig bestimmbar machen.

Dieses Konzept der Projektion von simplen, kleiner werdenden Streifenmustern für ein Bild mit 1024 Spalten und 768 Zeilen ermöglicht eine eindeutige Codierung eines Bildes in logarithmischer Größe und ist mit einem Binärcode vergleichbar.

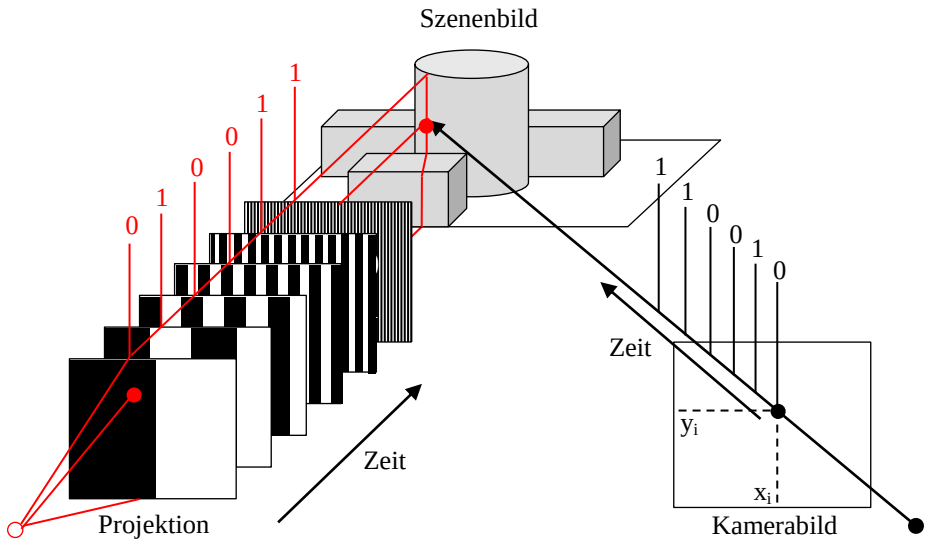
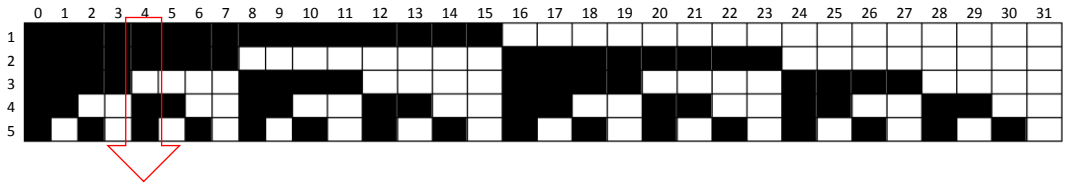


Abbildung 2.17: Streifencodierung
Veranschaulichung der Projektion binärer Streifenmuster und die Generierung eines einzigartigen Codes für ein betrachtetes Pixel auf der Kamerabildebene. Abbildung aus [13], S. 51 nachgebildet.

2.2.2 Binäre Streifenprojektion

Durch die Projektion eines Binärcodes auf die Szene lässt sich das Bild in *log2(n)* und somit 10 aufgenommenen binären Streifenbildern beschreiben. Ist das gesamte Set abfotografiert, so müssen die Bilder binarisiert werden. In der Binarisierung wird bestimmt, ob ein Pixel beleuchtet oder unbeleuchtet ist und dementsprechend wird dem Pixel der Codewert 1 oder 0 zugeordnet. Anschließend kann der entstehende Binärcode für einen Pixel in einen Dezimalwert umgewandelt werden und somit, in der Theorie, jedes einzelne Pixel mit einem eindeutigen Code versehen werden (Abbildung 2.18).



MSB 0 x 16
0 x 8
1 x 4 = 4
0 x 2
LSB 0 x 1

Abbildung 2.18: Binärcode Sequenz
Veranschaulicht eine 5 Bit Binärcode Sequenz auf 32 Spalten (oben). Decodierungsfehler im beispielsweise höchstwertigsten Bit (MSB) erzeugen 2^{n-1} (hier maximal $2^4 = 16$) Codewertfehler.

Ein großer Nachteil der binären Methode ist allerdings die hohe Fehleranfälligkeit. Ein Decodierungsfehler im höchstwertigsten Bit des Codes würde beispielsweise 2^{n-1} ($2^9 = 512$) betragen, was der Hälfte des Codierungsbereichs, also einem Datenfehler von 512, entspricht.

Streifenmuster sind im Allgemeinen in ihrer Auflösung und somit in ihrer Genauigkeit beschränkt, da die Streifenbreite von der Auflösung der Beleuchtungsquelle und somit des Projektors abhängt. Um die Auflösung zu erhöhen und somit Subpixelgenauigkeit zu erhalten, werden verschiedenste Varianten kombiniert. Im Folgenden werden überblicksmäßig die am häufigsten verwendeten Methoden vorgestellt, die zum einen mit der Problematik des Decodierungsfehlers besser umgehen und zum anderen Subpixelgenauigkeit erzielen können.

2.2.3 Graycode-Muster Methode

Die Graycode Methode leitet sich vom Graycode Algorithmus ab und wurde 1953 von Frank Gray entwickelt und patentiert [13]. Sie ermöglicht es, eine eindeutige Codierung eines Bildes ebenfalls in logarithmischer Größe durchzuführen und einen eventuell auftretenden Codierungsfehler zu minimieren. Projiziert man einen Graycode, äquivalent zur binären Variante, auf die Szene, so lässt sich ein Szenenbild wiederum in *log2(n)* Streifenbildern beschreiben. Abbildung 2.19 zeigt den Unterschied zwischen Graycodes und Binärcodes. Nach der Binarisierung der aufgenommenen Bilder kann wiederum die Umwandlung in Dezimalwerte erfolgen. Dazu muss der Graycode zuerst in Binärdaten umgewandelt werden, was mittels XOR-Operation durchgeführt wird. Die Umwandlung in Dezimalwerte erfolgt anschließend wiederum aus den Binärdaten des Graycodes (Abbildung 2.20).

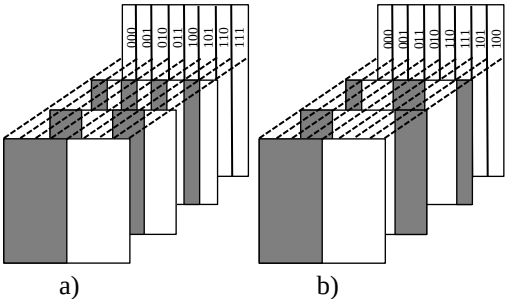


Abbildung 2.19: Sequenzvergleich
Vergleich zwischen Binärcode Sequenz a) und Graycode Sequenz b).

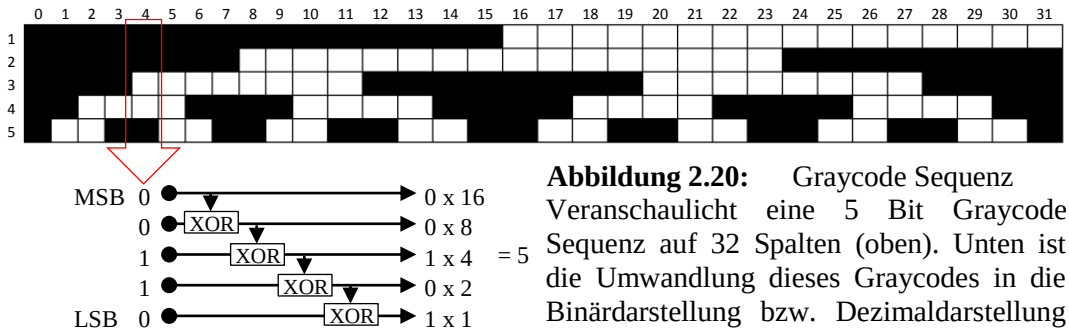
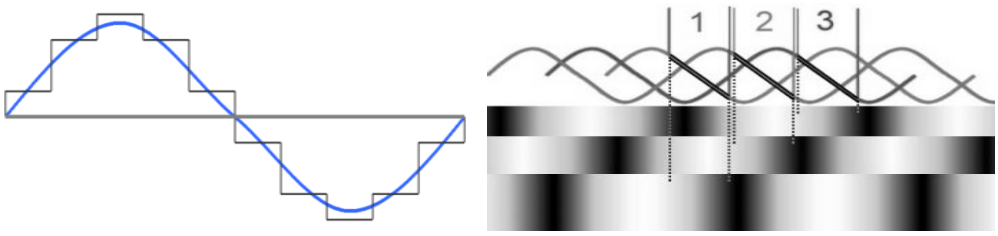


Abbildung 2.20: Graycode Sequenz
Veranschaulicht eine 5 Bit Graycode Sequenz auf 32 Spalten (oben). Unten ist die Umwandlung dieses Graycodes in die Binärdarstellung bzw. Dezimaldarstellung dargestellt. Abbildung wurde aus [13], S. 52 nachgebildet.

Vorteile dieses Verfahrens stellen nun die einfache Decodierung der einzelnen Pixel sowie die Robustheit gegenüber verschiedensten Texturen (mit Ausnahme von transparenten sowie stark reflektierenden Objektoberflächen) dar. Weiters besitzt das System eine hohe Robustheit gegenüber Decodierungsfehlern, da der Unterschied zwischen den aufeinanderfolgenden Codewerten nur maximal ein Bit beträgt und der Datenfehler somit maximal eins ist. Durch die codebedingte, effiziente Fehlerminimierung ermöglicht diese Variante somit eine einfache wie gute Rekonstruktion von statischen Objektoberflächen.

2.2.4 Phasen-Shift Methode

Die Phasen-Shift Methode oder auch Sinus-Methode basiert auf dem Ansatz von Graustufenauflösungen (Abbildung 2.21a). Dieses Verfahren ermöglicht es theoretisch, ein Pixel in einem Bild, welches mit einer Graustufenrampe horizontal und vertikal beleuchtet wird, bereits eindeutig zu beschreiben. Da in der Graustufenauflösung allerdings eine Limitierung vorliegt, werden fortlaufende Graustufenrampen eingesetzt, welche als Schwingungen oder Wellen angesehen



a) b)

Abbildung 2.21: Graustufenauflösung einer Sinuswelle
Schematisierte Abbildung der Graustufenauflösung über eine Sinuswelle a), Abbildung von phasenverschobenen Sinusschwingungen im Graustufen-Prinzip zur eindeutigen Identifikation von Bildpunkten b). Abbildung aus [49] entnommen.

werden können (Abbildung 2.21b). Wie in der Abbildung ersichtlich, reicht jedoch eine einzelne Sinuswelle nicht aus, um ein Pixel eindeutig zu identifizieren, da Codewerte pro Welle mehrmals auftreten. Deshalb werden mehrere Sinuswellen phasenverschoben projiziert und jeweils mit einer Kamera aufgenommen. Horizontal und vertikal angewendet wird wiederum eine eindeutige Identifikation eines Pixels möglich. Je nach Auflösung der Grauwerte bzw. der Anzahl der verwendeten, phasenverschobenen Sinusschwingungen reduziert sich auch die Anzahl der notwendigen Bilder für eine eindeutige Identifikation eines einzelnen Pixels. Somit werden mit dieser Methode auch potentiell weniger Aufnahmen als bei der Graycode-Methode benötigt. Außerdem wird durch die Projektion von mehreren Sinusschwingungen unterschiedlicher Phase eine sehr viel höhere Auflösung möglich. Durch Nutzung unterschiedlicher Frequenzen der Schwingungen lassen sich weitere Verbesserungen bei der Pixelcodierung erzielen.

Sehr oft wird diese Methode auch in Kombination mit der Graycode-Methode verwendet, um den Nachteil der Doppeldeutigkeiten der Sinus-Methode zu minimieren. Durch die Genauigkeit der Graycode-Methode in Kombination mit der Phasen-Shift-Methode lassen somit sich sehr gute Ergebnisse im Subpixel-Bereich erzielen (Abbildung 2.22).

Positiv bei der kombinierten Variante sind die geringe Anfälligkeit gegenüber Reflexionen und Nichtlinearitäten in der Szene und die höhere Auflösung bei weniger Bildsequenzen. Allerdings gestalten sich die Erstellung der strukturierten Binärbilder und die Berechnung der Pixelwerte als sehr komplex, was als Nachteil gesehen werden kann.

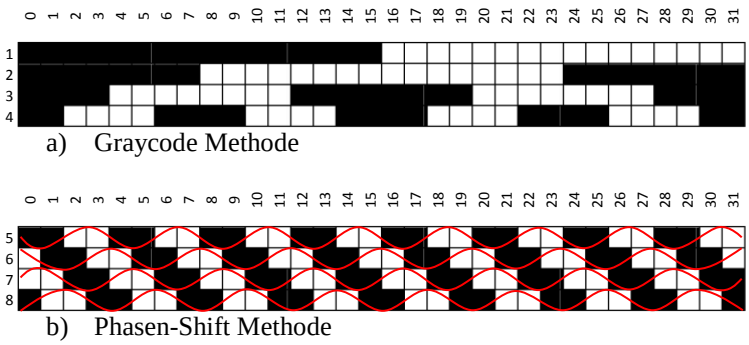


Abbildung 2.22: Phasen-Shift Methode
32 Bit Codesequenz eines kombinierten Phasen-Shift Verfahrens. Graycode Sequenz (a) und zusätzliche Phasen-Shift Sequenz (b). Abbildung wurde aus [13], S. 52 nachgebildet.

2.2.5 Line-Shift Methode

Das Prinzip der Line-Shift Methode basiert auf einer anderen Kombination mit der Graycode-Methode. Dabei werden zusätzlich zu der Graycode-Sequenz noch gesonderte Bildfolgen geschaltet, welche nur jede sechste Linie im Bild beleuchtet haben. Die Anzahl der beleuchteten Linien lässt sich prinzipiell variieren, wobei eine geringe Linienanzahl in einer höheren Fehlertoleranz (speziell in steilen Regionen) resultiert, jedoch mehr projizierte Linien weniger Aufnahmen zur eindeutigen Pixelbestimmung benötigen. Das Beleuchten jeder sechsten Linie stellt dazu einen guten Kompromiss zwischen Aufnahmen und Fehleranfälligkeit dar und wird häufig angewendet (Abbildung 2.23).

Über die Graycode-Sequenz lässt sich ein Pixel somit bis auf die zweite Projektionslinie auflösen, identifizieren und anschließend in Kombination mit dem Line-Shifting bis in den Subpixel-Bereich eindeutig bestimmen. Durch diese Kombination ist somit auch eine Zuordnung zwischen unterschiedlichen Kameras über die Projektorzeile bzw. Projektorspalte möglich, was bei der Phasen-Shift Methode nicht durchführbar ist.

Vorteilig bei dieser Methode ist, dass beispielsweise die Bildkoordinaten des Projektors zur Berechnung der Pixelwerte nicht berücksichtigt werden müssen. Die Bestimmung aller Pixel in den Stereobildern kann ausschließlich über die

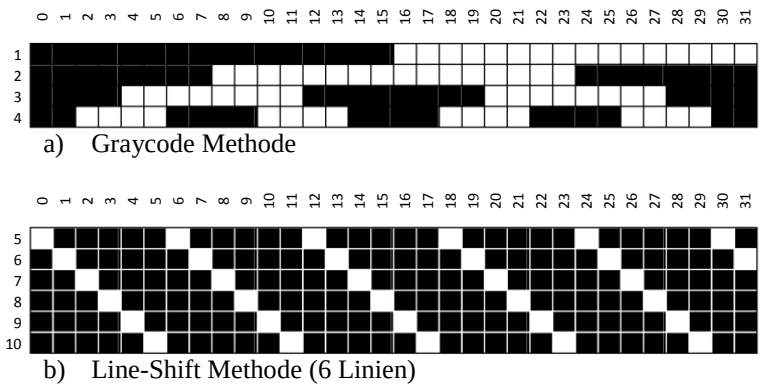


Abbildung 2.23: Line-Shift Methode
Schematische Darstellung einer Projektion von 32 Spalten. Graycode Sequenz a) in Kombination mit einer Line-Shift Sequenz b). Subpixelgenauigkeit lässt sich hierbei bereits mit sechs Projektionen erreichen. Abbildung aus [13], S. 57 nachgebildet.

Beobachtungskameras erfolgen. Dies ist vor allem bei nichtkalibrierten Projektionssystemen von Vorteil. Allerdings können ein kalibrierter Projektor und die zusätzliche Auswertung der Pixelwerte zu einer Steigerung der Zuverlässigkeit führen. Nachteil dieser Methode ist ein sehr hoher Ressourcenverbrauch, da mit vielen temporären Bildern gearbeitet werden muss. Zusammenfassend lässt sich diese Methode allerdings, besonders bei Verwendung mehrerer Kameras, als sehr flexibel und robust beschreiben.

2.2.6 Space-Time Methode

Das Space-Time Verfahren bildet quasi eine Erweiterung des Line-Shift Verfahrens und nutzt dieselbe Codemuster Sequenz. Die Auswertung der einzelnen Pixel findet dabei im Bildraum der Kamera statt, indem die Intensitätsmaxima der einzelnen Pixel zu diskreten Zeitpunkten betrachtet werden. Dies hat den Vorteil, dass der hohe Ressourcenverbrauch des Line-Shift Verfahrens erheblich reduziert wird. Über die Auswertung der Gray-Code Sequenz lässt sich die Codierung wiederum bis zur zweiten Projektionslinie auflösen und durch die bekannte Position der Pixel im Streifenmuster kann für jedes Kamerapixel das korrespondierende Pixel in der Projektor-Bildebene bestimmt werden (Abbildung 2.24).

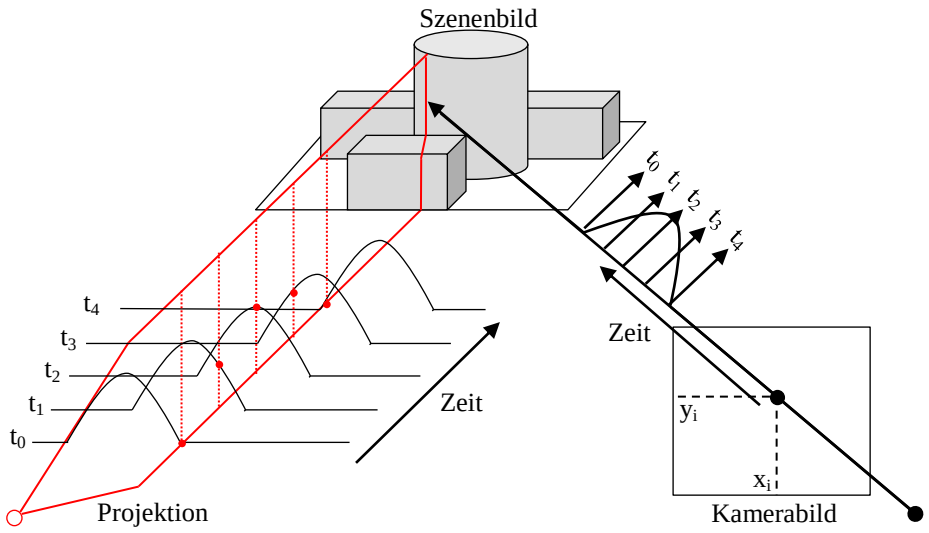


Abbildung 2.24: Space-Time Methode
Schematische Darstellung der Space-Time Methode. Bestimmung korrespondierender Pixel zwischen Kamera- und Projektorbild mittels Zeitmessung. Abbildung aus [13], S. 62 nachgebildet.

Durch die Nutzung einer flächenmäßigen Projektion von Streifenmuster-Sequenzen lässt sich durch diese Methode eine effiziente Auswertung der Codesequenz zum Vorteil eines sehr geringen Ressourcenverbrauchs durchführen.

Zusammenfassung

Die zuvor genannten Varianten sind häufig verwendete Methoden und bringen jede für sich ihre Vor- und Nachteile. Für jede Variante sind eine Vielzahl an Fehlerkorrekturen und Verbesserungsmöglichkeiten im Umlauf, welche die Algorithmen noch effizienter machen. Wie zuvor angeführt kommen zumeist auch Kombinationen von unterschiedlichen Methoden zum Einsatz, um Nachteile einzelner Varianten zu umgehen. Eine umfangreiche Auflistung und Beschreibung zu allen genannten Varianten sowie weitere Methoden finden sich in den Dissertationen von [13] und [30], welche auch die Grundlagen zu den, in Kapitel 2.2 angeführten Methoden geliefert haben.

2.3 Die Rolle von Farbe beim Stereo-Matching

Farbe spielt beim Stereo-Matching eine bedeutende Rolle. In vielen Arbeiten [4, 7, 22, 27, 32] wurde bereits gezeigt, dass Farbe beim Finden von korrespondierenden Pixeln hilfreich ist und bessere Ergebnisse liefern kann als Matching-Verfahren auf Graustufenbildern. Durch den Einsatz von Farbmessungen lassen sich Zweideutigkeiten, im Gegensatz zu den Graustufenbildern, effizienter ausschließen und Fehlklassifizierungen verringern, was ein einfaches Beispiel erläutern soll.

Angenommen man hat zwei Stereobilder mit jeweils einem blauen Pixel im linken und rechten Bild. Im rechten Bild gibt es zusätzlich noch ein rotes Pixel. In diesem Fall ist es relativ einfach, die Korrespondenz zwischen linkem und rechtem Bild zu ermitteln, da lediglich nach dem blauen Pixel im rechtem Bild gesucht werden muss (Abbildung 2.25a). Erstellt man aus den Beispielbildern jetzt jedoch Graustufenbilder, so erscheinen rote und blaue Pixel mit demselben Intensitätsgrauwert. Nun lässt sich nicht mehr so einfach herausfinden, welches der beiden Pixel im rechten Bild das korrespondierende Pixel des linken Bildes ist (Abbildung 2.25b). Es entstehen Doppeldeutigkeiten, wodurch die Berechnung aufwändig wird und sich auch negativ auf die Performance auswirkt. Dies trifft

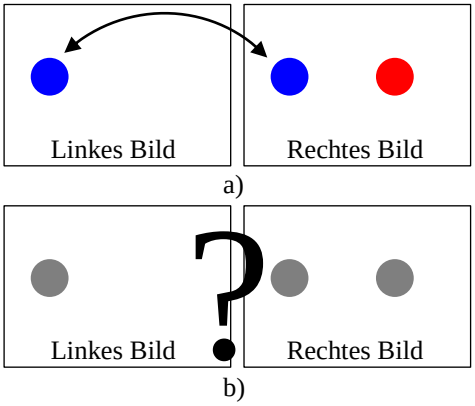


Abbildung 2.25: Problematik Stereo-Matching
Schematische Darstellung der Problematik zwischen Farb- und Graustufenbildern beim Stereo-Matching. Abbildung aus [42], S. 5 nachgebildet.

allerdings nicht auf alle Algorithmen zu. Farb-Korrespondenzberechnungen basieren zumeist auf Algorithmen, welche absolute Farbdifferenzen (AD, Abschnitt 2.3.2) zur Erstellung der Korrespondenzpunkte verwenden. Um nun zu zeigen, welchen Einfluss Farbe gegenüber intensitätsbasierten Auswertungen hat, können radiometrische Daten der Stereobildpaare zur Visualisierung verwendet werden. Durch Erstellung einer radiometrischen Karte (Radio-Map), welche radiometrische Verzerrungen zwischen linkem und rechtem Stereobild zeigt, lassen sich die Vorteile von Farbe bildlich darstellen.

Unter radiometrischen Verzerrungen versteht man die Unterschiede (unterschiedliche Intensitätswerte), welche zwischen linkem und rechtem Stereobildpaar auftreten. Gründe hierfür können geringfügige Änderungen in den Beleuchtungen der beiden Bilder sowie verschiedene Belichtungszeiten oder die Verwendung unterschiedlicher Kamerasensoren sein. Neben anderen Auswertungen wird in der Arbeit von [3] gezeigt, dass Matching-Fehler bei Verwendung von absoluten Differenzen zur Beschreibung des Datenterms speziell dort auftreten, wo auch hohe radiometrische Verzerrungen vorhanden sind. Dabei zeigen die Autoren von [3] jedoch auch, dass der Einsatz von Farbbildern und im speziellen von Farbräumen mit getrennten Farb- sowie Helligkeitskanälen diese Fehler minimiert, was den großen Vorteil von Farbe gegenüber Graustufenbildern darstellt. Allerdings ist die Fehleranfälligkeit beim Einsatz von absoluten Differenzen somit auch abhängig von der Häufigkeit der radiometrischen Verzerrungen in den Bildern, was als Nachteil gesehen werden

muss. Da die Erstellung der Radio-Maps auch für die Auswertung in dieser Arbeit relevant ist, wird dies in Abschnitt 2.3.1 nochmals genauer erläutert.

Im Gegensatz dazu wird in den Arbeiten von [3] und [16] weiters angeführt, dass Farbinformationen bei der Verwendung von radiometrisch insensitiven Korrespondenzalgorithmen wie Mutual Information (MI, Abschnitt 2.3.3), Normalized Cross-Correlation (NCC, Abschnitt 2.3.4) oder Census (Abschnitt 2.3.5) negative Einflüsse auf die Ergebnisse haben kann. Diese Algorithmen arbeiten durchwegs besser auf intensitätsbasierten Messungen und machen somit die Vorteile von Farbmessungen irrelevant.

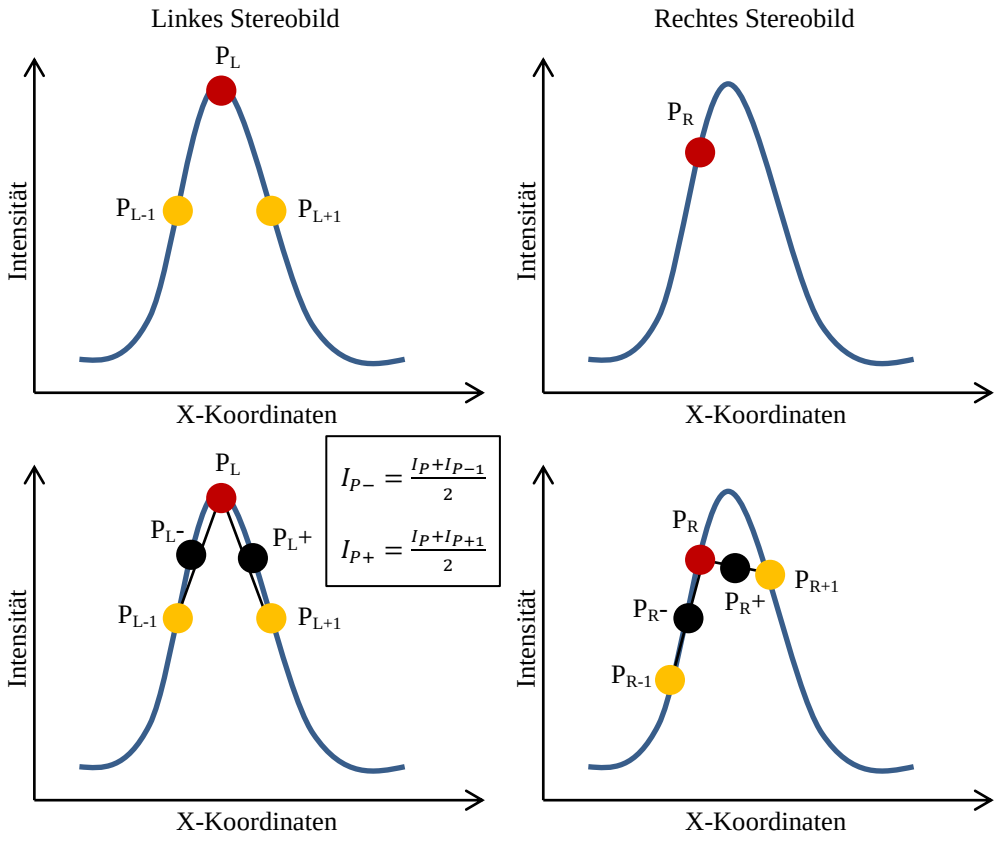
2.3.1 Radiometrisch verzerrte Bildregionen

Zur Extraktion von radiometrisch verzerrten Bereichen sind die zwei Initialbilder notwendig. Über die Ground-Truth Disparitätskarte lassen sich die Positionen der Pixel des linken und rechten Bildes bestimmen:

$$RV(x,y) = \sum |I_L(x,y) - I_R(x + d_L(x,y), y)| \tag{2.6}$$

Durch Bilden der absoluten Differenz zwischen der Pixelintensität an Position (x,y) des linken Bildes $I_L(x,y)$ und der korrespondierenden Pixelintensität auf der rechten Seite $I_R(x+d,y)$, welches über die Ground-Truth Disparitätskarte d des linken Bildes erhalten wird, lässt sich die radiometrische Verzerrung für diesen Pixel bestimmen, welche im Optimalfall null sein sollte. Die Intensität eines Bildpunktes kann als Kurvenfunktion dargestellt werden.

Durch die Aufnahme mit einer Kamera wird diese Intensitätskurve abgetastet (gesampelt) und den kontinuierlichen Intensitätswerten diskrete Werte zugeordnet. Durch diese Zuordnung für beide Stereobilder erhalten korrespondierende Pixel P_L im linken und P_R im rechten Bild meist leicht unterschiedliche Intensitätswerte (Abbildung 2.26), was auch sehr oft zu falschen Übereinstimmungen beim Matching führt. Um dieses Problem zu umgehen und somit eine bessere Differenzberechnung zum Erhalt der radiometrischen Karte zu bekommen, wird in dieser Arbeit eine Methode von [2] eingesetzt, welche durch Bildung von Mittelwerten den wahrscheinlichsten Datenwert bestimmt.



$$\rho(P_L, P_R) = \min(|P_L - P_R|, |P_{L-} - P_R|, |P_{L+} - P_R|, |P_L - P_{R-}|, |P_L - P_{R+}|)$$

Abbildung 2.26: Extrakton radiometrischer Daten
Schematische Darstellung der Extraktion radiometrischer Daten aus den Stereobildpaaren. Pixel im linken Bild müssen auf Intensitätsebene nicht exakt dem gesuchten Pixel auf der rechten Seite entsprechen (oben). Durch Bildung der Mittelwerte und Auswertung der minimalen Differenz kann die beste Intensität für den entsprechenden Pixel gefunden werden. Abbildung aus [2] nachgebildet.

Die Bildung der Mittelwerte erfolgt über die jeweiligen Nachbarn des zu betrachtenden Pixels auf der X-Achse für das linke und rechte Bild. Durch Bilden der absoluten Differenz zwischen den beiden korrespondierenden Pixeln sowie dessen Mittelwert-Pixeln kann der minimale Datenwert bestimmt werden und somit der beste Intensitätswert ermittelt werden (Abbildung 2.27). Ausgeführt auf alle Pixel des Bildes erhält man eine Karte mit den berechneten Datenkosten. Radiometrische Verzerrungen zwischen linkem und rechtem Bild (helle Bereiche in den Abbildungen) treten dann auf, wenn beispielsweise geringe Änderungen in der Beleuchtung der beiden Bilder vorkommen (Abbildung 2.27b). Die roten Bereiche in Abbildung 2.27b zeigen die Verdeckungen zwischen linkem und rechtem Bild welche aufgrund der unterschiedlichen Ansichten zustande kommen. Die Radio-Map

a) b)

Abbildung 2.27: Radiometrische Verzerrungen
a) linkes Bild eines Testsets aus dieser Arbeit mit zugehöriger Disparitätskarte. b) Berechnete Datenkosten wobei rot markierte Pixel die verdeckten Bereiche darstellen und helle Pixelwerte die Bereiche mit hohen Kosten in der Berechnung. Existieren keine radiometrischen Verzerrungen, so würde b) komplett in schwarz dargestellt sein.

wird in Abschnitt 4 als Gewichtungskarte von Fehlern in radiometrisch reinen sowie verzerrten Bereichen herangezogen, über welche die Leistung von radiometrisch sensitiven sowie insensitiven Algorithmen visualisiert werden kann (siehe [3] oder [43]).

2.3.2 Absolute Farbdifferenzen

Die Berechnung von Differenzen zur Datenkostenerstellung stellt eine oft genutzte Methode dar. Dabei ist die Erstellung von absoluten Differenzen eine vielverwendete und gute Methode für das Stereo-Matching. Durch pixelweise Subtraktion der einzelnen Bildpunkte für das linke und rechte Bild und das anschließende Bilden der Summe über den Betrag der Datenwerte hat die Berechnung folgende Form:

$$\rho_{AD}(p, q - d_p) = \sum_{1 \leq i \leq 3} |\sigma_i(p) - \sigma_i(p - d)| \tag{2.7}$$

$\sigma_i(p)$ gibt den Wert des Pixels an Koordinate p für den i-ten Farbkanal zurück. d ist der Datenwert aus der Ground-Truth Disparitätskarte. Wird die Berechnung auf Graustufenbildern durchgeführt, fallen die Farbkanäle weg und die Gleichung vereinfacht sich auf:

$$\rho_{AD}(p, p - d) = |I_p - I_{p-d}| \tag{2.8}$$

Neben der Berechnung der absoluten Differenzen kann auch noch eine abgeflachte Form genannt werden, welche zusätzlich noch bessere Ergebnisse speziell in verdeckten Regionen liefert. Die Abflachung basiert auf einer Minimierungsfunktion zwischen dem Disparitätswert und einem selbstdefinierten Parameter k und liefert somit eine erhöhte Chance, durch Betrachten der umliegenden, nichtverdeckten

Pixelwerte einen korrekten Disparitätswert ermitteln zu können. Die Berechnung kann in folgender Form angeschrieben werden:

$$\rho_{AD}(p, p-d) = min(|I_p - I_{p-d}|, k) \tag{2.9}$$

Nachteilig dabei wirkt sich allerdings der Parameter k aus, da dieser mehrere Werte annehmen kann und es außerdem nicht trivial ist, einen optimalen Wert für jeden betrachteten Pixel zu finden (siehe auch [3] oder [43]).

2.3.3 Mutual Information

Im Gegensatz zu den Methoden der absoluten Differenzen stellt Mutual Information, neben der Kreuzkorrelation und Census, eine radiometrisch unempfindliche Messung dar. Berechnungen, welche nicht von der Radiometrie des Bildes abhängen zeigen zumeist eine erhöhte Qualität der Resultate.

Mutual Information ist eine globale Matching-Methode und arbeitet auf Basis einzelner Pixelwerte. Die Problematik bei dieser Methode liegt nun darin, dass zur Berechnung einer korrekten Disparitätskarte eine vorhandene Disparitätskarte notwendig ist. Um dieses Problem zu lösen, wird zu Beginn eine initiale Disparitätskarte erstellt, was beispielsweise über absolute Differenzen durchgeführt werden kann. Anschließend lassen sich die Datenwerte der Mutual Information Berechnung über diese Disparitätskarte annähern und daraus eine neue, genauere Disparitätskarte erstellen. Somit kann der Aufbau der Methode als eine Art Hierarchie, ausgehend von der initialen Disparitätskarte zur finalen Disparitätskarte, gesehen werden (Abbildung 2.28). Arbeitet man bei der Auswertung wiederum auf

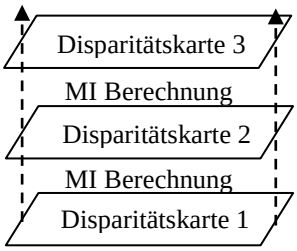


Abbildung 2.28: Mutual Information
Schematischer Aufbau der Mutual Information Berechnung. Ausgehend von einer initialen Disparitätskarte (1) wird nach oben hin die korrigierte Disparitätskarte (3) erstellt.

Farbbildern, so wird die Prozedur für jeden einzelnen Farbkanal durchgeführt und die Resultate abschließend aufsummiert. Nachteilig bei der Mutual Information Berechnung ist, wie zuvor erwähnt, einerseits das Henne-Ei-Problem bezüglich der initialen Disparitätskarte und andererseits, aufgrund des globalen Ansatzes, die Handhabung von lokalen, radiometrischen Änderungen. Als Vorteil kann jedoch die schnelle Berechnung der Disparitätskarte aufgrund des hierarchischen Aufbaus gesehen werden. Die genaue Berechnung der Methode wird in einer Arbeit von [18] erläutert.

2.3.4 Normalized Cross-Correlation

Die normalisierte Kreuzkorrelation ist, genauso wie Census, eine fensterbasierte Methode (Abschnitt 2.1.5) und wie folgt definiert:

$$\rho_{ZNCC}(p, p-d) = \sum_{1 \leq i \leq 3} \frac{\sum_{q \in W_p} [\sigma_i(q) - \overline{\sigma}_i(p)][\sigma_i(q-d) - \overline{\sigma}_i(p-d)]}{\sqrt{\sum_{q \in W_p} [\sigma_i(q) - \overline{\sigma}_i(p)]^2 \sum_{q \in W_p} [\sigma_i(q) - \overline{\sigma}_i(p) \sigma_i(q-d) - \overline{\sigma}_i(p-d)]^2}} \tag{2.10}$$

Dabei stellt W_p das quadratische Fenster dar, welches zentriert auf Pixel p liegt, und $\overline{\sigma}_i(p)$ berechnet den Mittelwert des i-ten Farbkanals über alle Pixel in diesem Fenster. Für intensitätsbasierte Berechnungen vereinfacht sich Gleichung (2.10) auf:

$$\rho_{ZNCC}(p, p-d) = \frac{\sum_{q \in W_p} (I_q - \overline{I_p})(I_{q-d} - \overline{I_{p-d}})}{\sqrt{\sum_{q \in W_p} (I_q - \overline{I_p})^2 \sum_{q \in W_p} (I_{q-d} - \overline{I_{p-d}})^2}} \tag{2.11}$$

Also Vorteil bei der normalisierten Kreuzkorrelation ist deren Robustheit gegenüber radiometrischen Verzerrungen zu nennen, da durch das Bilden der Mittelwerte in den einzelnen Fenstern die Intensitätswerte normalisiert werden. Probleme treten allerdings, wie in Abschnitt 2.1.5 erläutert, in Randbereichen auf (siehe auch [3] oder [43]).

2.3.5 Census

Census ist ebenfalls eine lokale, also fensterbasierte Methode und arbeitet anhand der Berechnung der Hamming-Distanz. Ein Fenster wird auf dem jeweiligen Pixel p im linken und dem korrespondierenden Pixel q im rechten Bild gelegt und ausgewertet.

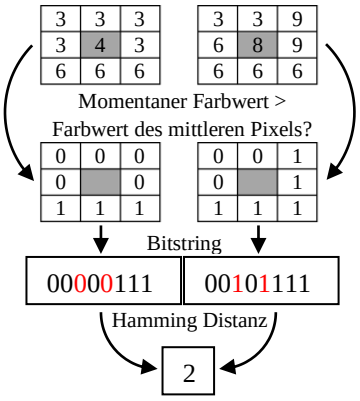


Abbildung 2.29: Census Berechnung
Schematische Veranschaulichung der Census Berechnung. Abbildung aus [3] nachgebildet.

Dabei basiert die Auswertung, ausgehend vom betrachteten, mittig platzierten Pixelwert, darauf, dass alle Pixelwerte, welche größer dem betrachteten Pixelwert sind, mit 1 gekennzeichnet werden und alle kleineren Werte mit 0. Anschließend wird, von links nach rechts, oben nach unten aus dem jeweiligen Fenster ein Bitstring erstellt und die Census-Kostenberechnung über die Auswertung der Positionen der beiden Strings, also der Hamming-Distanz, durchgeführt (Abbildung 2.29). Arbeitet man auf Farbbildern, so wird die Prozedur für jeden Farbkanal einzeln durchgeführt und abschließend die Kosten für alle drei Farbkanäle aufsummiert. Vorteil bei der Census-Methode stellt die erhöhte Robustheit in Randbereichen dar sowie auch eine erhöhte Unempfindlichkeit bei leicht variierenden Beleuchtungen der Stereobildpaare. Ein Problem bei dieser Methode stellen allerdings wiederum untexturierte Bereiche dar, nachdem dort die erstellten Bitstrings ähnlich sind und durch Bildrauschen in diesen Bereichen vermehrt Fehler auftreten können, was sich auf die Berechnung der Hamming-Distanz auswirkt. Weitere Informationen über die Census-Variante findet sich in der Arbeit von [38].

Die in den Abschnitten 2.3.1 bis 2.3.5 aufgezählten Algorithmen finden in der Auswertung für diese Arbeit Verwendung und basieren auf der Vorgehensweise aus der Arbeit von [3]. Die Auswertungen selbst werden in Abschnitt 4 – Experimente und Resultate gezeigt und erläutert.

3. Implementierung

In dieser Arbeit sollen die in [3] durchgeführten Farbevaluierungen auf einen neuen Datensatz wiederholt werden. Ein Ziel dabei ist es zu überprüfen ob die Ergebnisse in [3] kameraspezifisch sind, was in dieser Arbeit durch die Aufnahme von vier Sets mit unterschiedlichen Kameras durchgeführt wird. Zum anderem wird geprüft, ob beispielsweise veränderte Umgebungsbedingungen (wie Beleuchtung) einen Einfluss auf die Ergebnisse haben.

In den folgenden Abschnitten wird gezeigt, wie die Farbevaluierung durchgeführt wurde. Dabei befasst sich Abschnitt 3.1 mit der Erstellung von Stereobildpaaren mit zugehöriger Referenzlösung (Ground-Truth). Abschnitt 3.2 befasst sich mit der Vorgehensweise zur Datenevaluierung von [3] und Abschnitt 4 erklärt die Experimente, die durchgeführt wurden, sowie deren Resultate.

3.1 Erstellung von Stereo-Tiefenkarten mit strukturiertem Licht

Ziel dieses Abschnittes ist es, Stereobildpaare mit Referenzlösungen (Ground-Truth) in Form von Disparitätskarten zu erzeugen. Dabei sollen reale Bildpaare von komplexen, physischen Szenen generiert werden. Zur Erzeugung der Stereobilder wurde eine Kamera auf einer automatisch gesteuerten Schiene befestigt, welche eine horizontale Verschiebung der Kamera ermöglicht. Zur Aufnahme kommt eine, im Zuge einer Praktikumsarbeit erstellte Software zum Einsatz, welche die Kamera und die Schiene automatisiert ansteuert.

Zur Erstellung von Referenzlösungen in Form von Disparitätskarten wurde die Arbeit von [33], welche strukturiertes Licht verwendet, nachimplementiert. Strukturiertes Licht (Abschnitt 2.2) ermöglicht es, Disparitätskarten effizient und stabil zu berechnen. Dabei wird die Szene mit Streifenmustern durch einen Projektor beleuchtet. Zur Aufnahme der Stereobildpaare wurden zwei verschiedene Kameras (Canon EOS-1D Mark III, Canon EOS-5D Mark II – Abbildung 3.3) auf der automatisch gesteuerten Schiene montiert, welche die Kamera horizontal um 15 Zentimeter verschieben kann. Aufgrund der durch den Videoprojektor projizierten Lichtsequenzen lässt sich nachfolgend jedem Pixel ein eindeutiger Code zuordnen.

Übereinstimmungen im zweiten Bild können dann durch Suche nach diesem Code einfach gefunden werden.

Wird jeder Pixel in den Stereobildern von dem Projektor angeleuchtet, so lässt sich ein konsistentes Ground-Truth Set vollautomatisch erstellen. Da jedoch eine komplette Ausleuchtung der Szene durch den Projektor kaum möglich ist, kann es zu Abschattungen kommen, welche sich als Fehler in den Berechnungen wiederspiegeln. Um diese Abschattungen und somit undefinierte Pixelwerte weiter zu minimieren, ist es möglich und sinnvoll, mehrere Beleuchtungsquellen zu verwenden. In dieser Arbeit wird allerdings lediglich eine Beleuchtungsquelle in Form eines Videoprojektors verwendet.

In den Abschnitten 3.1.1 bis 3.1.9 wird nun näher darauf eingegangen, wie mit Stereobildpaaren sehr genaue Tiefenkarten mittels strukturiertem Licht erstellt werden können.

3.1.1 Erstellung der Eingabedaten

Zur Erstellung der Eingabedaten wurde ein Labor eingerichtet und die notwendige Hardware angeschafft. Ein Videoprojektor dient zur Projektion der Streifenmuster (strukturiertes Licht) auf die Szene. Des Weiteren wurde eine Schiene auf einem Stativ montiert, welche sich automatisiert in horizontaler Richtung verschieben lässt. Auf der Schiene können unterschiedliche Kameramodelle angebracht werden. Durch die sequentielle Aufnahme von Bildern und dessen horizontal verschobenen Korrespondenzbildern lassen sich Stereobildpaare mit nur einer Kamera erstellen (Abbildung 3.1).

a) b) c) d)

Abbildung 3.1: Laboraufbau
a) zeigt den Laboraufbau mit der Kamera auf der automatisierten, horizontal beweglichen Schiene sowie den Videobeamer als Beleuchtungsquelle. b) zeigt die Ansicht auf eine aufzunehmende Szene. c) stellt die Ansteuerung mittels eigens entwickelter Software dar und d) zeigt eine Streifensequenz einer Szene, welche aufgenommen wird.

Durch eine im Zuge eines Projektpraktikums erstellte Software ist es möglich, Kamera sowie Schiene und Videoprojektor automatisiert anzusteuern und die Aufnahmen zu machen. Um jedes einzelne Pixel in einer Szene eindeutig codieren zu können, bedarf es nun einer sequentiellen Projektion von Streifenmustern. Der einfachste Fall, um dies durchzuführen, wäre eine Sequenz binärer Streifenmuster, welche mit immer kleiner werdenden Streifen bis zu einer Streifenbreite von einem Pixel das Eingabebild codiert (Abbildung 3.2). Ein Nachteil dieser Methode ist, dass *n* Bilder benötigen werden, wobei *n* für die Breite des Eingabebildes in Pixel steht [1, 8, 29]. Ein weiterer Nachteil der binären Methode ist die, in Abschnitt 2.2.2 beschriebene, hohe Fehleranfälligkeit. Deshalb wurde für diese Arbeit die Projektion von Graycodes (siehe Abschnitt 2.2.3) herangezogen, welche die Anzahl der notwendigen Bilder auf *log2(n)* und somit 10 Bilder für ein 1024 Pixel breites Bild beschränkt.

Um die Stereobildpaare und zugehörige Ground-Truth Disparitätskarten zu erstellen, bedarf es nun einer Sequenz von Aufnahmen der Szene. Dazu gehören die beiden linken und rechten Initialbilder, welche die Szene ohne Streifenmuster zeigen, sowie die jeweils 10 Aufnahmen der mit Graycodes beleuchteten Szene zur Berechnung der linken und rechten Stereo-Tiefenkarte. Die beiden Initialbilder sowie die zu berechnenden Tiefenkarten werden im folgenden Stereo-Set genannt. In der Implementierung von [33] werden insgesamt 164 Aufnahmen benötigt, um ein Stereo-Set zu generieren. Warum genau 164 Eingabebilder notwendig sind, wird in Abschnitt 3.1.6 erläutert.

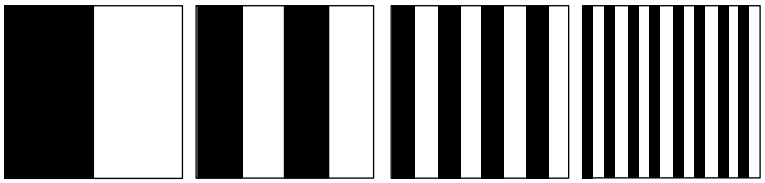


Abbildung 3.2: Binärsequenz
Schematische Darstellung einer Binärsequenz.

3.1.2 Konvertierung

Um Verluste bei der Komprimierung in verschiedenste Bildformate zu verhindern, wurden alle Bilder eines Stereo-Sets immer mit der höchstmöglichen Auflösung sowie im kameraeigenen RAW-Format aufgenommen. Da die kameraeigenen RAW-

Formate für MATLAB allerdings nicht lesbar sind, muss zu Beginn der Berechnungen eine Umwandlung in das MATLAB-interne RAW-Format „Portable Pixmap (PPM)“ durchgeführt werden. Dies lässt sich mittels Freeware Software DCRAW [46] einfach veranlassen.

3.1.3 Versuche mit mehreren Kameras

Ein Ziel dieser Arbeit waren Experimente mit mehreren Kameras. Es sollte herausgefunden werden, ob sich unterschiedliche Kamerasensoren auf die Resultate auswirken. Dazu wurden dieselben Stereo-Sets mit zwei unterschiedlichen Kameratypen (Canon EOS-1D Mark III, Canon EOS-5D Mark II - Abbildung 3.3) aufgenommen und ausgewertet. Es soll herausgefunden werden, ob sich unterschiedliche Kamerasensoren und somit unterschiedliche Aufnahmemethoden in irgendeiner Weise auf die Experimente und deren Resultate auswirken, oder ob keine Kameraabhängigkeit besteht.

Um die Stereobildpaare, welche mit unterschiedlichen Kameras aufgenommen wurden, vergleichen zu können, muss bei der Aufnahme der Sets beachtet werden, dass für alle Kameras möglichst gleiche Initialbilder erstellt werden können. Dabei wurde spezielles Augenmerk darauf gelegt, dass Parameter wie Belichtungszeiten, Brennweite oder Blende für alle Aufnahmen gleich sind.

Bezeichnung: **EOS-1D Mark III** **Typ:** CMOS-Sensor (28,1 x 18,7 mm) **Pixel effektiv:** 10,10 Megapixel **Pixel gesamt:** 10,70 Megapixel **Seitenverh:** 3:2 **Tiefpassfilter:** integriert **Farbfilter:** RGB-Primärfarben **Bildprozessor:** 2 x DIGIC III **Formfaktor:** 1,3	**Bezeichnung:** **EOS-5D Mark II** **Typ:** CMOS-Sensor (36 x 24 mm) **Pixel effektiv:** 21,10 Megapixel **Pixel gesamt:** 22,00 Megapixel **Seitenverh:** 3:2 **Tiefpassfilter:** integriert **Farbfilter:** RGB-Primärfarben **Bildprozessor:** DIGIC 4 **Formfaktor:** 1,0

Abbildung 3.3: Verwendete Kameratypen
Kameradaten sowie Bilder aus [45] entnommen.

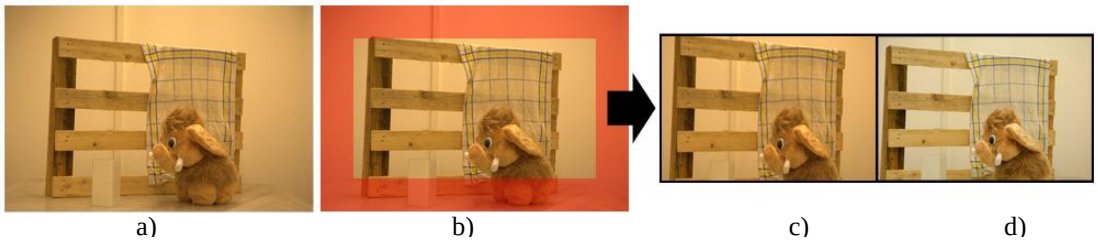


Abbildung 3.4: Cropping von Bildern
Originalbild der Canon 5D a) und der Canon 1D d). Die Bildgröße der 5D ist um einiges größer als die der 1D bedingt durch den unterschiedlichen Formfaktor von 1 zu 1,3. Um das Bild der 5D auf das der 1D anzupassen, muss das 5D-Bild zuerst um den Formfaktor beschnitten werden (Zoom mit Faktor 1,3) b) und anschließend auf die Bildgröße der 1D verkleinert werden c).

Ein weiterer wichtiger Punkt ist der so genannte Kamera-Formfaktor (Crop-Factor), welcher sich von Kamera zu Kamera unterscheiden kann. Der Formfaktor definiert das Längenverhältnis der flächenmäßigen Bildgröße. Vereinfacht gesagt bedeutet das, dass Kameras mit einem Formfaktor von 1 (hier Canon 5D) bei gleichem Kameraobjektiv-Einstellungen einen größeren Bildausschnitt einer Szene liefern als Kameras mit einem Faktor von 1,3 (hier Canon 1D). Um Vergleiche durchführen zu können, müssen die Bilder der Canon 5D vor der Weiterverarbeitung zuerst „beschnitten“ werden, um annähernd denselben Bildausschnitt der Canon 1D zu erhalten (Abbildung 3.4).

3.1.4 Rektifizierung

Da die bewegliche Schiene im Labor, auf welcher die Kameras montiert sind, nicht exakt genug ausgerichtet werden kann, sodass eine Zeile des linken Bildes exakt derselben Zeile des rechten Bildes entspricht, müssen alle Bildpaare der Sets rektifiziert werden.

Die Rektifizierung ermöglicht das gegenseitige Ausrichten der Stereobildpaare zueinander, sodass alle Pixel einer Zeile im linken Bild in derselben Zeile des rechten Bildes gefunden werden können (Abbildung 3.5). In dieser Arbeit wurde die Rektifizierung mittels Interpolationsverfahren (passives Verfahren, Abschnitt 2.1.3) implementiert. Für die Rektifizierung in dieser Arbeit wurden die Methoden von [24, 47, 48] herangezogen. Eine detailliertere Beschreibung der Rektifizierung ist in Abschnitt 2.1.3 zu finden.

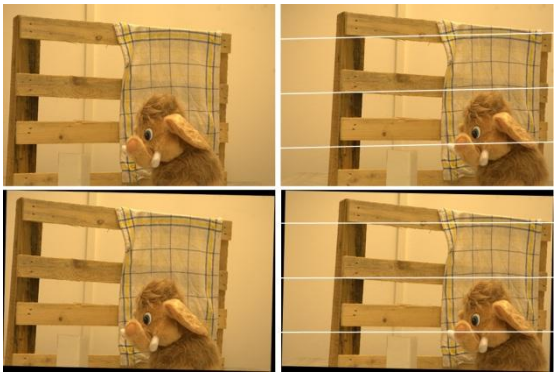


Abbildung 3.5: Rektifizierung
Darstellung einer rektifizierten Szene. Oben die unrektifizierte Originalszene mit eingezeichneten Epipolarlinien (oben, rechts), Unten die rektifizierte Szene, Epipolarlinien ausgerichtet (unten, rechts).

Durch die Rektifizierung verringert sich auch der weiterführende Berechnungsaufwand, da die Suche nach korrespondierenden Bildpunkten nur zeilenweise durchgeführt wird. Aufgrund von Speicherproblemen werden bei der Rektifizierung in dieser Arbeit die Bilder um 50 Prozent verkleinert.

3.1.5 Gray-Codes

Wie bereits in Abschnitt 2.2.3 erläutert besteht das zu projizierende Graycode Muster aus versetzten schwarzen und weißen Streifen, welche eine eindeutige Codierung jedes einzelnen Pixels ermöglichen. Wird ein solches Muster auf eine Oberfläche projiziert, so lässt sich jeder Pixel mit bereits *log2(n)* Mustern eindeutig identifizieren, wobei *n* für die horizontale Pixelauflösung des Videoprojektors steht. Ein weiterer Vorteil von Graycodes ist, dass sich der Code benachbarter Pixel lediglich um ein Bit ändert (Abbildung 3.6). Dadurch bewirken kleine Fehler in der Binarisierung (Abschnitt 3.1.6) keine großen Änderungen des Codes [31] (siehe auch Abschnitt 2.2.3).

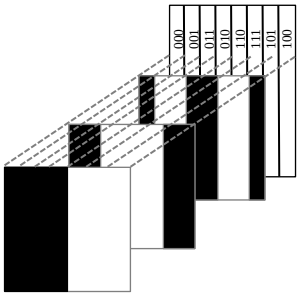


Abbildung 3.6: Graycode Sequenz
Schematische Darstellung einer Graycode Sequenz. Jeder Codewert ändert sich lediglich um ein Bit pro Muster.

Im konkreten Fall bedeutet dies nun, dass für ein Bild mit einer Auflösung von 1024x768 Pixel die Projektion von 10 vertikalen sowie 10 horizontalen Streifenmustern ausreicht, um jedes einzelne Pixel im Bild eindeutig zu identifizieren und somit dessen Position (im Folgenden als u und v Koordinaten bezeichnet) zu bestimmen (Abbildung 3.7). In weiterer Folge muss dann lediglich ausgewertet werden, ob ein Pixel im Bild beleuchtet wurde oder nicht, was im folgenden Abschnitt genauer erklärt wird.

3.1.6 Binarisierung

Unter der Binarisierung versteht man das Umwandeln aufgenommener Bilder, welche mit einem Streifenmuster beleuchtet wurden, in Binärbilder. Dazu werden alle eingelesenen Bilder in Grauwertbilder umgewandelt und anschließend ausgewertet, ob ein Pixel beleuchtet ist, oder nicht. In der Theorie bedeutet dies, dass man den Mittelwert des Grauwertbildes berechnet und die Pixel, je nachdem ob diese größer oder kleiner dieses Mittelwertes sind, dementsprechend auf weiß (beleuchtet) oder schwarz (unbeleuchtet) setzt. Die Berechnung des Mittelwertes bezeichnet das Rückstrahlvermögen einer nicht selbstleuchtenden Oberfläche und wird auch Albedo genannt.

Durch die Projektion der Streifenmuster entstehen in der Praxis allerdings erhöhte Helligkeiten in der Nähe von hellen Bildbereichen innerhalb des Projektors oder durch Reflexionen in der Szene, welche Fogging-Artefakte genannt werden und die Berechnung der Albedo problematisch machen. In [33] werden diese Probleme mittels zusätzlicher Projektion der invertierten Streifenmuster verringert. Allerdings schlägt sich dies in einer Verdopplung der aufzunehmenden Bildsequenzen nieder.

Abbildung 3.7: Graycode Aufnahmesequenz
Darstellung einer kompletten Aufnahmesequenz für das linke Stereobild mit vertikalem Streifenmuster.

Des Weiteren wurde in der Arbeit von [33] erkannt, dass die Bestimmung von Pixeln mit hohem, variierendem Rückstrahlvermögen weiterhin ein Problem bleibt. Um dies zu umgehen und Fehlklassifizierungen weiter zu minimieren, werden alle Streifenbilder mit zwei unterschiedlichen Belichtungszeiten aufgenommen. Dies hat eine weitere Verdoppelung der aufzunehmenden Bilder zur Folge. Somit ergibt sich eine Gesamtzahl von 164 Bildern pro Set (jeweils für die linke und rechte Ansicht - 10 Streifenbilder horizontal, vertikal, invertiert mit zwei verschiedenen Belichtungen sowie die Initialbilder für die linke und rechte Seite), um eine effiziente Erstellung der Disparitätskarten zu garantieren.

Um nun ein Streifenbild (beispielsweise horizontal) zu binarisieren, wird für beide Belichtungen *B1*, *B2* die Differenz des Streifenmusters *H1* und dessen Inversen $H1_{inv}$ gebildet und Differenzbilder *D1*, *D2* erstellt. Dadurch erhalten unbeleuchtete Pixel einen negativen und beleuchtete Pixel einen positiven Farbwert. Anschließend werden die Absolutwerte der Differenzbilder *D1* und *D2* verglichen und der Pixelwert mit dem höheren Absolutwert in einem der beiden Bilder in ein finales Differenzbild *D* zusammengeführt.

Liegen nun einzelne Pixelwerte von *D* innerhalb einer gewissen Threshold (in dieser Arbeit zwischen -16 und +16), so werden diese Pixel als „unbekannt“ gekennzeichnet. Diese undefinierten Pixelwerte treten vor allem in schattigen Bereichen sowie Regionen mit hohen Reflexionen und schiefen Winkeln auf und lassen sich bei der Aufnahme nur schwer vermeiden (graue Pixel in Abbildung 3.8).

Zudem werden Werte unterhalb des Thresholds als unbeleuchtet (schwarz) bzw. oberhalb als beleuchtet (weiß) gekennzeichnet (Abbildung 3.8). Anschließend wird das codierte Bild noch gesäubert, indem fehlklassifizierte Pixel speziell an den

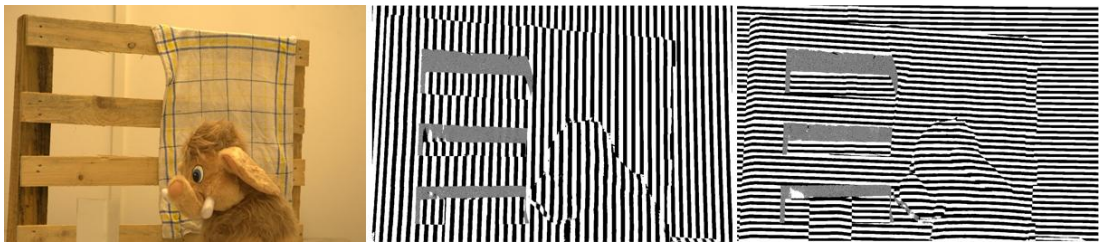


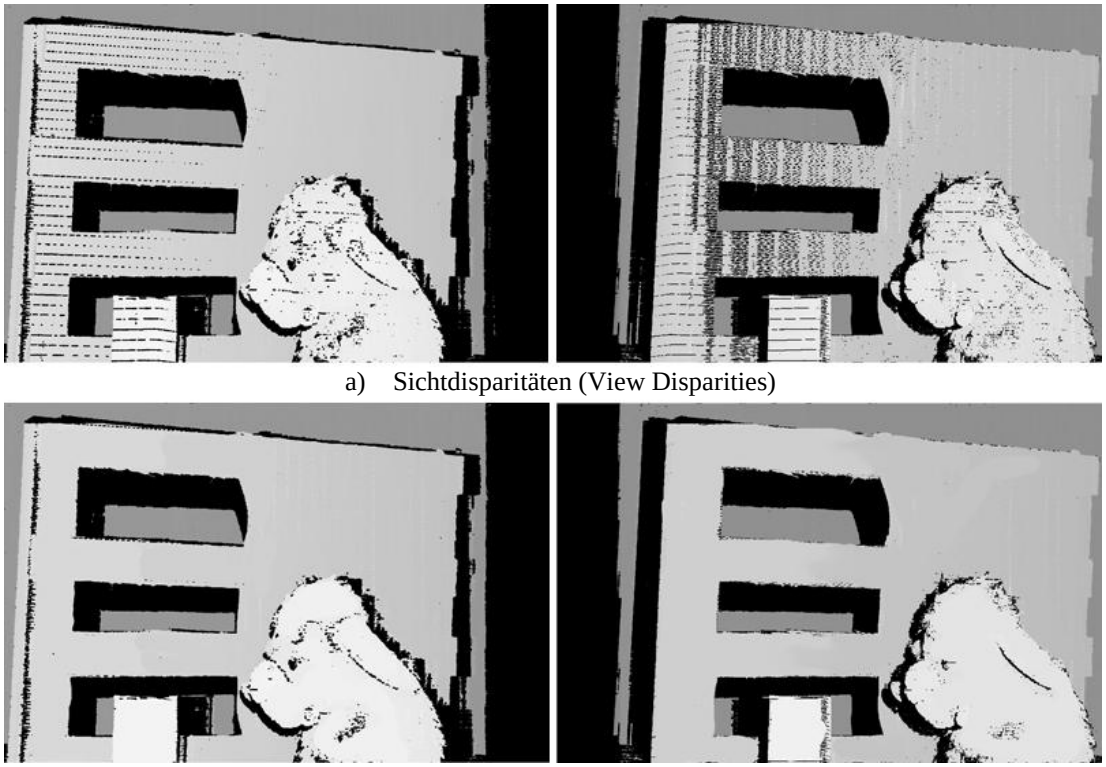
Abbildung 3.8: Codierung
Codierung der einzelnen Pixelwerte auf der Ebene sieben der 10 Bilder für die linke Seite. Graue Bereiche stellen die als „unbekannt“ gekennzeichneten Pixelwerte dar.

binarisierten Streifenrändern korrigiert und kleinere Löcher geschlossen werden. Diese Operation führt man für alle 80 streifencodierten, linken und rechten Bilder aus und erhält ein konsistentes Set von insgesamt 40 gesäuberten Graycode-Bildern (*uv*-Koordinaten).

Durch die jeweils 10 horizontalen sowie 10 vertikalen binären Graycode-Bilder für die linke und rechte Seite lässt sich nun für jedes einzelne Pixel eines Bildes mit der Größe 1024x768 ein eindeutiger Code zuordnen und dieses klassifizieren. Da aber in der Praxis Kameraauflösungen viel höher sind als die durchschnittlichen Projektor-Auflösungen, verursacht dies weitere Probleme in der Klassifizierung. Der Videobeamer projiziert Streifenmuster in einer Auflösung von 1024x768 Pixel auf die Szene, die Kamera nimmt jedoch beispielsweise die Bilder in einer Auflösung von 2048x1536 Pixel auf. Das bedeutet, dass ein Pixel am Videobeamer vier Pixel im Kamerabild entspricht. Somit müsste man die Graycode Bilder weiters interpolieren, um doppelte Codewerte zu vermeiden. Aufgrund der ohnehin hohen Qualität der berechneten Sichtunterschiede (View Disparities), welche im nachfolgenden Abschnitt erläutert werden, wurde auf die Interpolation in der Implementierung für diese Arbeit jedoch verzichtet.

3.1.7 Berechnung der Sichtunterschiede (View Disparities)

Bei der Berechnung der Sichtunterschiede geht es nun darum, für einen Pixel im linken Bild den korrespondierenden Pixel im rechten Bild zu finden. Durch die vorangegangene Rektifizierung lässt sich dies nun auf eine eindimensionale Suche auf der jeweiligen Scanline einschränken und effizient berechnen (Abbildung 3.9). Gesucht wird nach den 20-Bit Graycode-Werten (die Kombination der vertikalen und horizontalen Graycode-Werte), welche in der Binarisierung zuvor berechnet wurden. Sind zwei korrespondierende Pixel im jeweils linken und rechten Bild gefunden, so wird die absolute Differenz der X-Koordinaten der beiden Pixel zueinander gebildet und man erhält die Verschiebung (Disparität) für dieses Pixelpaar. Diese Verschiebung wird dann in der ViewDisparity-Map an der jeweiligen X und Y Koordinate abgespeichert.



a) Sichtdisparitäten (View Disparities)

b) Gesäuberte Sichtdisparitäten

Abbildung 3.9: Sichtdisparitäten
Oben sind die berechneten Sichtdisparitäten nach der Binarisierung der beiden Stereobilder zu sehen a). Unten dieselben Bilder nach der Säuberung und Fehlerkorrektur b). Schwarze Bereiche zeigen undefinierte Codewerte.

Führt man diese Prozedur für jeden Pixel des Bildes aus, so erhält man eine vollständige ViewDisparity-Map mit den resultierenden Verschiebungen zwischen linkem und rechtem Bild. Die Berechnung muss für das linke und rechte Stereobildpaar gesondert durchgeführt werden, wodurch sich als Ergebnis die ViewDisparity-Maps d_{LR} und d_{RL} ergeben (Abbildung 3.9 a). In der Praxis wirft dieses einfache Konzept jedoch einige Probleme auf, auf die geachtet werden muss. Dabei ist das größte dieser Probleme die teilweise Verdeckung von Bildbereichen im linken bzw. rechten Stereobildpaar, welches sich jedoch nicht vermeiden lässt (Abbildung 3.9 – schwarze Bereiche). Ein weiteres Problem sind Abschattungen der Streifenmusterprojektionen, wodurch in machen Bildbereichen keine Codewerte generiert werden können. Allerdings lassen sich Abschattungen minimieren, indem man mehrere Projektoren benutzt und die Disparitätskarten für jeden dieser Projektoren berechnet. Weiters sind Fehlberechnungen bzw. Ausreißer bei der Differenzberechnung der Pixel im linken und rechten Bild problematisch, welche in

dieser Arbeit nach Berechnung einer Disparitäts-Zeile allerdings erkannt und als fehlerhaft klassifiziert werden.

Abschließend werden die beiden Sicht-Disparitätskarten d_{LR} und d_{RL} noch gesäubert, was in dieser Arbeit mittels morphologischem Closing durchgeführt wird. Eine unabhängige Gegenprobe (Cross-Check) der berechneten Werte ergibt schließlich die konsistenten, korrigierten Sichtunterschiede für das Stereobildpaar eines Sets. Dieser Cross-Check erfolgt jeweils unabhängig voneinander für die linke Disparitätskarte $d_{LR}(x,y,d)$ sowie die rechte Disparitätskarte $d_{RL}(x,y,d)$, indem überprüft wird, ob der Disparitätswert d auf Position $d_{LR}(x,y) = d_{RL}(x+d,y)$ ist. Pixel, welche den Cross-Check nicht erfüllen, werden als undefiniert gekennzeichnet. Nach der Überprüfung sind nur noch die Disparitätswerte vorhanden, welche in beiden Kameraansichten sichtbar sind und von mindestens einer Lichtquelle beleuchtet werden (Abbildung 3.9b).

3.1.8 Berechnung der Beleuchtungsunterschiede (Illumination Disparities)

Unter der Berechnung der Beleuchtungsunterschiede sind die Disparitäten gemeint, welche zwischen Projektorgeometrie und den aufgenommenen Bildern (Kamerageometrie) ermittelt werden können. Diese Unterschiede lassen sich mit den bereits berechneten Sichtdisparitäten und der Streifenmuster berechnen. Eine Kalibrierung der Beleuchtungsquelle kann hierbei entfallen, da die zuvor berechneten, binären Graycode-Werte in direktem Zusammenhang mit den Bildkoordinaten stehen und jedes Kamera-Beleuchtungspaar somit als unabhängige Quelle von Stereo-Differenzen dient, welche über die Sichtdisparitäten registriert werden. Wie zuvor bereits erwähnt, können die Sichtunterschiede als projektive Tiefe und somit als dreidimensionale Rekonstruktion der Szene aufgefasst werden, was eine Lösung der Projektionsmatrix M zur Erstellung der Beleuchtungs-unterschiede für jede Beleuchtungsquelle möglich macht.

Die Projektionsmatrix M für Beleuchtungsquelle 0 und der linken Ansicht der Stereobildpaare sieht dann folgendermaßen aus:

$$P \cong M_{0L} \cdot S \tag{3.1}$$

P repräsentiert hierbei die Graycode-Werte u_{L0} und v_{L0} sowie dessen X- und Y-Koordinaten im Beleuchtungsmuster und kann als homogener 2D-Szenenpunkt $P = [u_{0L}\ v_{0L}\ 1]^T$ beschrieben werden. Den homogenen 3D-Szenenpunkt stellen die Sichtunterschiede dar, welche hier mit S gekennzeichnet sind. Da die Sicht-Disparitäten in direktem Zusammenhang mit der Beleuchtungsquelle stehen und über deren Pixelwerte die projektive Tiefe repräsentieren, lässt sich S als $S = [x\ y\ d\ 1]^T$ darstellen, wobei die projektive Tiefe d für die linke Ansicht als $d = d_{LR}(x, y)$ geschrieben werden kann. M_{0L} ist die unbekannte 4x3 Projektionsmatrix der Beleuchtungsquelle 0 und der linken Ansicht, welche in weiterer Folge nun gelöst werden muss:

$$P = M_{0L} \cdot S \quad \cong \quad \begin{bmatrix} u_{0L} \\ v_{0L} \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} m_{12} m_{13} m_{14} \\ m_{21} m_{22} m_{23} m_{24} \\ m_{31} m_{32} m_{33} m_{34} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} \quad \cong \quad \begin{bmatrix} u_{0L} \\ v_{0L} \\ 1 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} \tag{3.2}$$

Da m_{34} lediglich einen Skalierungsfaktor darstellt, kann $m_{34} = 1$ gesetzt werden. Wenn nun m_1, m_2, m_3 als die drei Zeilen der Projektionsmatrix M_{0L} bezeichnet werden, ergeben sich folgende zwei Gleichungen deren elf unbekannte Einträge für jedes Pixel, für welches Sichtunterschiede und Beleuchtungs-Codes bekannt sind, gelöst werden müssen:

$$\begin{aligned} u_{0L} m_3 S &= m_1 S \\ v_{0L} m_3 S &= m_2 S \end{aligned} \tag{3.3}$$

Eine effiziente und stabile Methode dieses Gleichungssystems stellt hierbei die Berechnung der kleinsten Quadrate [12] dar, sofern Beleuchtungs-Codes und Sichtunterschiede richtig sind. Da allerdings bereits eine kleine Anzahl an Ausreißern das Ergebnis stark verfälschen kann, muss zusätzlich eine Kontrolle für diese eingebaut werden.

Hierzu iteriert man die Berechnung der kleinesten Quadrate und kalkuliert über die ergebende Projektionsmatrix die Pixelwerte für die Beleuchtungsunterschiede. Über einen Threshold werden dann Ausreißer der berechneten Werte herausgefiltert und die Iteration mit den korrigierten Werten nochmals durchgeführt. Durch die

Verringerung der Threshold nach jeder Iteration lassen sich effektiv gravierende Ausreißer ausfiltern und eine korrekte Projektionsmatrix berechnen. Im Paper von [33] wurde festgestellt, dass bereits vier Iterationen der Prozedur ausreichen, um eine sehr gute Anpassung der Projektionsmatrix zu erhalten.

Durch das erneute Anwenden der Least-Squares Methode mit der angepassten Projektionsmatrix lässt sich anschließend Gleichung (3.3) effizient lösen was als Resultat die Beleuchtungsunterschiede d_{0L} ergibt. Diese Prozedur führt man nun für beide Stereo-Ansichten sowie alle Lichtquellen durch und erhält die Beleuchtungs-Sets d_{iL}, d_{iR} wobei i hier die Anzahl der Lichtquellen definiert. In diesen Sets sind nun alle Pixelwerte vorhanden, welche in linker und rechter Ansicht sichtbar sind und somit auch die halbverdeckten Punkte enthalten.

In dieser Arbeit wird die oben beschriebene Prozedur wie folgt durchgeführt: In einer Vorverarbeitung werden all jene Pixel der Sichtunterschiede S gesucht, deren Pixelwert nicht undefiniert ist. In einer $Kx3$ Matrix (S_M) wird anschließend für alle definierten Werte deren X- und Y-Koordinate sowie der eigentliche Pixelwert (projektive Tiefe, d) abgespeichert. Dabei steht K für die Anzahl aller sichtbaren, definierten Pixelwerte:

$$S_M = \begin{bmatrix} x_1 & y_1 & d_1 \\ x_2 & y_2 & d_2 \\ \vdots & \vdots & \vdots \\ x_K & y_K & d_K \end{bmatrix} \tag{3.4}$$

Im selben Zuge werden auch die Werte der Graycodes ausgelesen und in einer Kx5 Matrix A_M gespeichert. Dabei werden wiederum die X- und Y-Koordinaten sowie der Pixelwert der Sichtunterschiede d gespeichert. Zusätzlich werden auch noch die Pixelwerte der jeweils horizontalen und vertikalen Graycodes (u, v) mitgespeichert:

$$A_M = \begin{bmatrix} x_1 & y_1 & d_1 & u_1 & v_1 \\ x_2 & y_2 & d_2 & u_2 & v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_K & y_K & d_K & u_K & v_K \end{bmatrix} \tag{3.5}$$

Hier gilt es zu beachten, dass undefinierte Graycode-Werte in der Matrix A_M den Wert 0 erhalten. Aus der Matrix S_M werden dann Ausreißer herausgefiltert und alle

Graycode-Werte, welche dieselben Positionskoordinaten *(x, y)* in beiden Matrizen S_M und A_M enthalten, in der Matrix S_M hinzugefügt:

$$S_M = \begin{bmatrix} x_1 & y_1 & d_1 & u_1 & v_1 \\ x_2 & y_2 & d_2 & u_2 & v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_K & y_K & d_K & u_K & v_K \end{bmatrix} \tag{3.6}$$

Somit erhält man zwei *Kx5* Matrizen, wobei in der Matrix A_M alle relevanten Pixelwerte enthalten sind und die Matrix S_M nur die definierten Pixelwerte enthält. Mit S_M wird anschließend die Projektionsmatrix berechnet (Gleichung (3.3)).

Da die Least-Squares Methode eine definierte Berechnungsmatrix zur Berechnung der 3x4 Projektionsmatrix voraussetzt, muss S_M in eine *Kx11* Matrix S_{PM} umgewandelt werden. Dies erfolgt wie folgt:

$$S_{PM} = \begin{bmatrix} x_1 & y_1 & d_1 & 1 & 0 & 0 & 0 & 0 & -x_1 \cdot u_1 & -y_1 \cdot u_1 & -d_1 \cdot u_1 \\ 0 & 0 & 0 & 0 & 1 & x_1 & y_1 & d_1 & -x_1 \cdot v_1 & -y_1 \cdot v_1 & -d_1 \cdot v_1 \\ x_2 & y_2 & d_2 & 1 & 0 & 0 & 0 & 0 & -x_2 \cdot u_2 & -y_2 \cdot u_2 & -d_2 \cdot u_2 \\ 0 & 0 & 0 & 0 & 1 & x_2 & y_2 & d_2 & -x_2 \cdot v_2 & -y_2 \cdot v_2 & -d_2 \cdot v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_K & y_K & d_K & 1 & 0 & 0 & 0 & 0 & -x_K \cdot u_K & -y_K \cdot u_K & -d_K \cdot u_K \\ 0 & 0 & 0 & 0 & 1 & x_K & y_K & d_K & -x_K \cdot v_K & -y_K \cdot v_K & -d_K \cdot v_K \end{bmatrix} \tag{3.7}$$

Anschließend kann die Projektionsmatrix berechnet werden. Um nun Ausreißer zu detektieren, wird Gleichung (3.3) berechnet und man erhält die neuen Pixelwerte (projektive Tiefe) für die Matrix S_M. Die neu berechneten Disparitätswerte werden dann mit einer Threshold verglichen und alle Daten größer dieser Threshold aus der Matrix S_M entfernt. Iteriert man diesen Vorgang mit der jeweiligen Verkleinerung der Threshold, so erhält man eine angepasste Projektionsmatrix S_{PM}.

Zur finalen Berechnung der Beleuchtungsunterschiede wird anschließend Gleichung (3.3) mittels der erhaltenen, gesäuberten Projektionsmatrix S_{PM} kalkuliert und die Pixelwerte in das jeweilige Beleuchtungs-Set (d_{xL} bzw. d_{xR}) gespeichert, was die Matrix für die Sichtunterschiede ergibt (Abbildung 3.10).

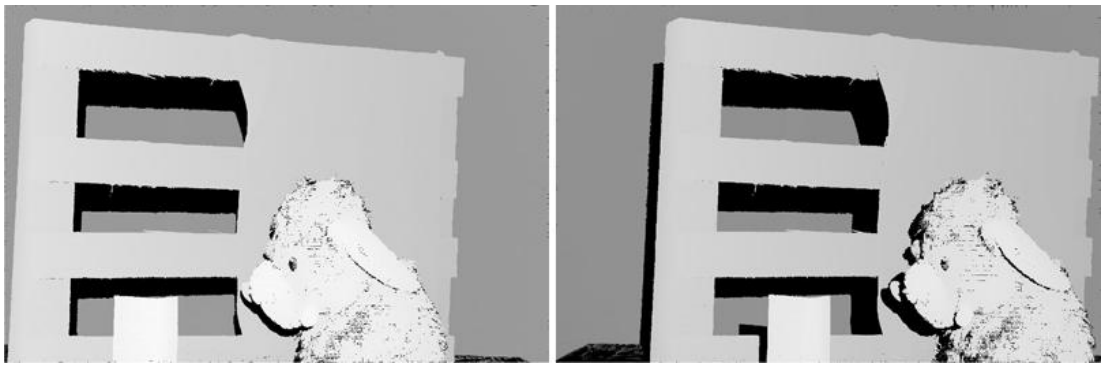


Abbildung 3.10: Beleuchtungsdisparitäten
Berechnete Beleuchtungsdisparitäten für linkes und rechtes Stereobild. Schwarze Bereiche zeigen undefinierte Codewerte.

3.1.9 Kombination der Disparitätskarten

Abschließend werden die zwei Sichtdisparitäten (d_{LR}, d_{RL}) mit den $2N$ Beleuchtungsdisparitäten (d_{iL}, d_{iR}) kombiniert, wobei N für die verwendeten Beleuchtungsquellen steht. Dies wird für die linke und rechte Ansicht getrennt durchgeführt, wobei für jedes Pixel, welches mehr als eine Disparitätsschätzung hat, ein robuster Mittelwert erstellt wird. Undefinierte Pixel in allen Disparitäten werden weiterhin als undefiniert in das Disparitäts-Set geschrieben:

$$d_{Left} = \frac{1}{N+1}\left(d_{LR} + \sum_i^N d_{iL}\right)$$

$$d_{Right} = \frac{1}{N+1}\left(d_{RL} + \sum_i^N d_{iR}\right) \tag{3.8}$$

Das Ergebnis dieser Operation sind die zwei Disparitäts-Sets d_{Left} bzw. d_{Right}, welche anschließend noch auf Konsistenz geprüft werden müssen:

$$d_L(x,y) = d_{Right}(x + d_{Left}(x,y), y)$$

$$d_R(x,y) = d_{Left}(x - d_{Right}(x,y), y) \tag{3.9}$$

Sind die Unterschiede der zwei Disparitätskarten d_{Left} und d_{Right} nur geringfügig, so werden diese anhand der Gleichung (3.9) zu den finalen Disparitäts-Sets d_L bzw. d_R zusammengefügt und man erhält exakte Paare von Disparitätskarten entsprechend der linken und rechten Stereo-Ansicht (Abbildung 3.11).

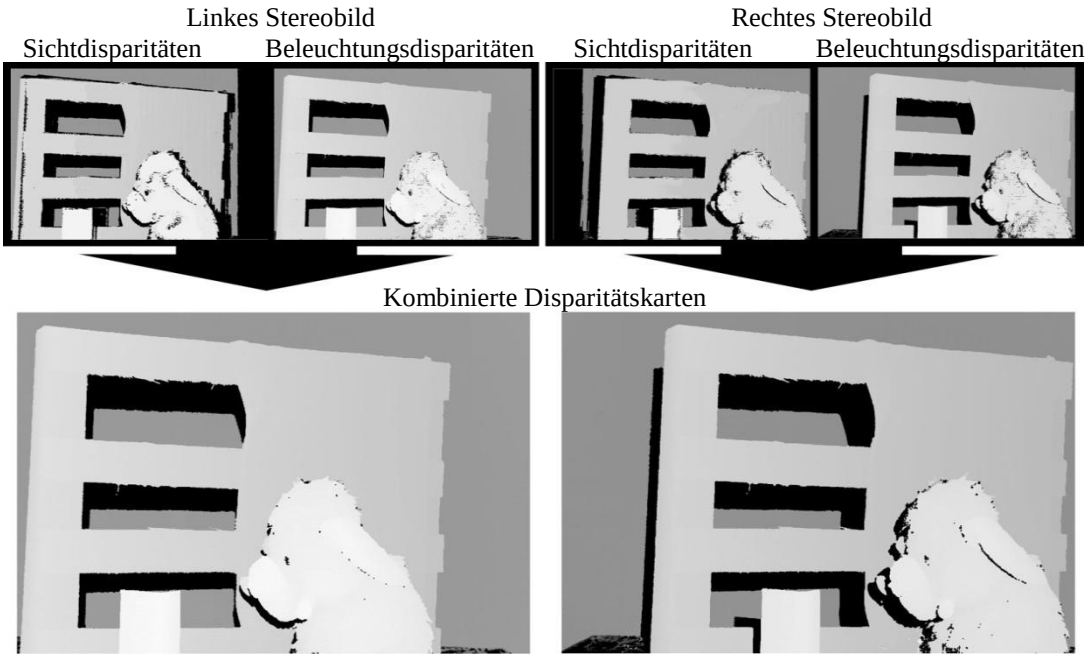


Abbildung 3.11: Kombination von Disparitätskarten
Kombinieren der Sicht- sowie Beleuchtungsdisparitäten und Erhalten der finalen Disparitätskarte für linkes und rechtes Stereobild. Schwarze Bereiche sind undefinierte Codewerte.

Um die konsistenten Disparitätskarten für die weitere Auswertung vorzubereiten, werden diese auf ein Viertel ihrer Größe verkleinert (bis zu diesem Punkt wurde mit der vollen, rektifizierten Bildgröße gearbeitet), da die meisten Stereo-Auswertungsalgorithmen nur mit kleineren Bildgrößen arbeiten können. Zur Vorbereitung für die weitere Evaluierung werden abschließend noch kleinere Löcher in den finalen Disparitätskarten gefüllt, was in dieser Arbeit wiederum mittels morphologischem Closing durchgeführt wird. Hierbei kommt es vor, dass größere Löcher von undefinierten Pixeln übrig bleiben, was speziell in stark reflektierenden Bereichen oder Bereichen mit sehr geringer Albedo der Fall ist. Da sich dies aber nicht vermeiden lässt, sollten solche Fälle beim Erstellen der Szene nach Möglichkeit von vorn herein vermieden werden.

3.2 Vorgehensweise zur Datenevaluierung

In Abschnitt 3.1 wurden Stereobilder mit korrespondierender Ground-Truth Tiefenkarte generiert. Nun gilt es herauszufinden, wie wichtig Farbinformationen für Stereo-Matching Algorithmen aus diesem Datensatz sind. Dazu wird nach der in [3] beschriebenen Vorgehensweise vorgegangen: Zuerst werden Disparitätskarten mit

dem Stereo Matching Algorithmus [5] unter Verwendung unterschiedlicher Farbräume und Korrespondenzmaße von den Stereobildern berechnet. Anschließend (siehe Kapitel 4) werden diese Disparitätskarten mit den in Kapitel 3.1 berechneten Ground-Truth Disparitätskarten verglichen und die Fehlerraten analysiert.

Dabei werden neun Farbsysteme zur Berechnung der Disparitätskarten verwendet: die Primärsysteme (RGB, XYZ), den Luminanz-Chrominanz Systeme (LUV, LAB, AC1C2, YC1C2) und die statisch unabhängigen Komponentensysteme (I1I2I3, H1H2H3). Die verwendeten Korrespondenzmaße sind absolute Farbdifferenzen, Mutual Information, Normalized Cross-Correlation sowie Census und eine Auswertung über die radiometrischen Daten (Abschnitt 2.3.1). Zudem werden intensitätsbasierte Messungen mittels Grauwertbildern durchgeführt. Die Ergebnisse werden im folgenden Abschnitt 4 erläutert.

4. Experimente und Resultate

In dieser Arbeit soll geprüft werden, ob die Aussagen von [3] auf dem in Kapitel 3 erstellten Datensatz nachvollziehbar sind. In diesem Abschnitt werden die dazu notwendigen Experimente erläutert. Dabei beschreibt Abschnitt 4.1 die Berechnung von Disparitätskarten mit unterschiedlichen Farbräumen und Korrespondenzmaßen auf den Test-Sets und Abschnitt 4.2 erläutert die Fehlermetriken, welche verwendet werden, um die Qualität der Disparitätskarten zu messen. Abschnitt 4.3 zeigt, wie sich die Disparitätsfehler in den verschiedenen Farbräumen verhalten. Abschnitte 4.4 bis 4.6 zeigen das Verhalten von Fehlerraten in Bezug auf die unterschiedlichen Korrespondenzmaße sowie in Bezug auf unterschiedliche Kameras.

4.1 Berechnung der Disparitätskarten

Wie in Abschnitt 3.2 angesprochen, werden nun für jedes der vier in Kapitel 3 beschriebenen Test-Sets Disparitätskarten berechnet. Pro Test-Set werden 36

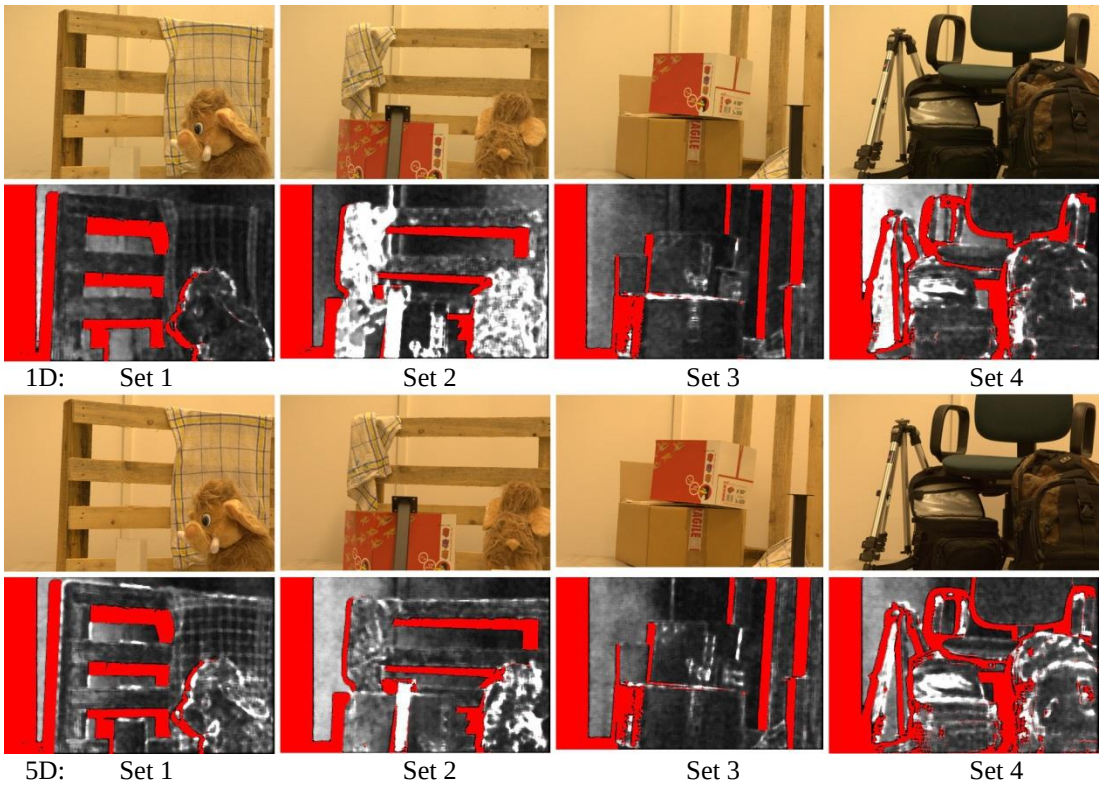


Abbildung 4.1: Radiometrische Karten
Die Abbildung zeigt die Karten mit den radiometrischen Verzerrungen für die, in dieser Arbeit verwendeten Test-Sets. Oben sind die Karten für die vier Sets der Canon 1D mit zugehörigem linken Initialbild dargestellt, unten die Karten für der Canon 5D. Rote Bereiche kennzeichnen die Verdeckungen für die linke Ansicht und helle Bereiche zeigen hohe radiometrische Verzerrungen.

Disparitätskarten erstellt – je eine für jede Kombination aus Farbraum und Korrespondenzmaß (siehe Abschnitt 3.2 für eine Auflistung der Farbräume und Korrespondenzmaße). Wie in [3] werden zudem für jedes Bildpaar die radiometrisch verzerrten Regionen nach der in Abschnitt 2.3.1 vorgestellten Methode extrahiert. In Abbildung 4.1 sind die vier verschiedenen Sets der beiden Kameras und deren radiometrisch verzerrte Regionen dargestellt.

4.2 Qualitative Fehlermetriken

Um die Qualität der berechneten Disparitätskarten zu messen, müssen diese mit der Referenzlösung mittels einer Fehlermetrik verglichen werden. In dieser Arbeit werden drei Fehlermessungen E_V, E_D, E_C verwendet. Dabei definiert E_V eine Fehlermessung über alle Fehler im Bild und wird nach der Methode von [32] berechnet (Gleichung (4.1)). E_D zeigt die Fehler in radiometrisch verzerrten Regionen, welche aus den zuvor berechneten Radiometrie-Bildern erhalten werden (Gleichung (4.2)). Im Gegensatz dazu gibt E_C den prozentualen Fehler in den Regionen an, die nicht von radiometrisch verzerrten Bereichen betroffen sind (Gleichung (4.3)). Die Berechnung der Fehler sieht folgendermaßen aus:

$$E_V(\mathrm{d}) = \frac{1}{V}\sum\left|d_p - d_p{}'\right| > \varepsilon \tag{4.1}$$

$$E_D(\mathrm{d}) = \frac{\sum_{p\in V} T[|d_p - d_p{}'| > 1]\cdot\omega_p}{\sum_{p\in V}\omega_p} \tag{4.2}$$

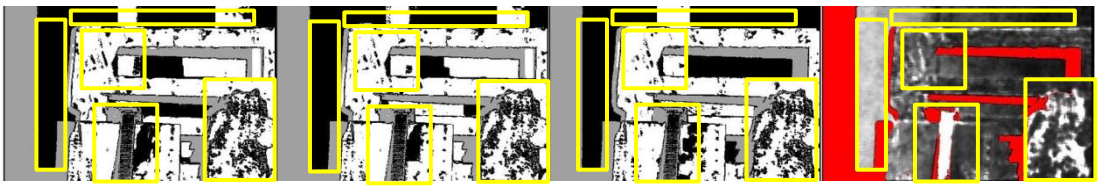
$$E_C(\mathrm{d}) = \frac{\sum_{p\in V} T[|d_p - d_p{}'| > 1]\cdot(1-\omega_p)}{\sum_{p\in V}(1-\omega_p)} \tag{4.3}$$

V ist die Menge aller sichtbaren Pixelwerte im Bild. d_p ist die Ground-Truth Disparität von Pixel p und $d_p{}'$ ist die berechnete Disparität an Pixel p. ε definiert einen Toleranzfaktor für die Disparitäten und wird auf 1.0 gesetzt. T ist eine Indikatorfunktion welche 1 zurückgibt, wenn das Argument richtig ist oder 0 wenn nicht. ω_p stellt eine Gewichtung dar, welche zwischen 0 (schwarze Pixel) und 1 (weiße Pixel) liegt und sich direkt aus dem zuvor berechneten Radiometrie-Bild (Abschnitt 2.3.1) ergibt.

4.3 Visualisieren von Fehlern

Zur Visualisierung der Fehler wird die absolute Differenz der jeweiligen Disparitätskarte mit dessen zugehöriger Ground-Truth Karte an jedem Pixel berechnet. Ein Fehler tritt dann auf, wenn die Differenz eines Pixelwertes größer eins ist. In Abbildung 4.2 wird dies anhand der Farbräume Grey, RGB und LUV für das Set 2 – 5D aus dieser Arbeit gezeigt. Dabei stellen die grauen Bereiche Fehler in verdecken Regionen dar und die schwarzen Bereiche Fehler in den sichtbaren Regionen. In [3] konnte beobachtet werden, dass Fehler in den Berechnungen der unterschiedlichen Farbräume nicht willkürlich verteilt sind. Sie treten meist dort auf, wo auch hohe radiometrische Verzerrungen vorkommen (Abbildung 4.2, gelbe Boxen). Bezugnehmend auf das farbbasierte Matching, welches in Abschnitt 4.4 genauer erläutert wird, lässt sich im Set 2 – 5D erkennen, dass Farbe durchaus Fehler reduziert. Auffallend hierbei ist allerdings, dass speziell der Hintergrund für farbbasiertes Matching in allen vier Test-Sets dieser Arbeit als Fehler deklariert wurde. Dies dürfte an den fehlenden Texturen bzw. an der gleichbleibenden Farbe des Hintergrundes liegen, wodurch der Matching-Algorithmus aus [3] Datenwerte in diesen Bereichen produziert, dessen Abweichung in Bezug auf die Ground-Truth Disparitätskarte größer eins sind (siehe auch Abschnitt 2.1.5).

Test-Set 2 – Canon 5D

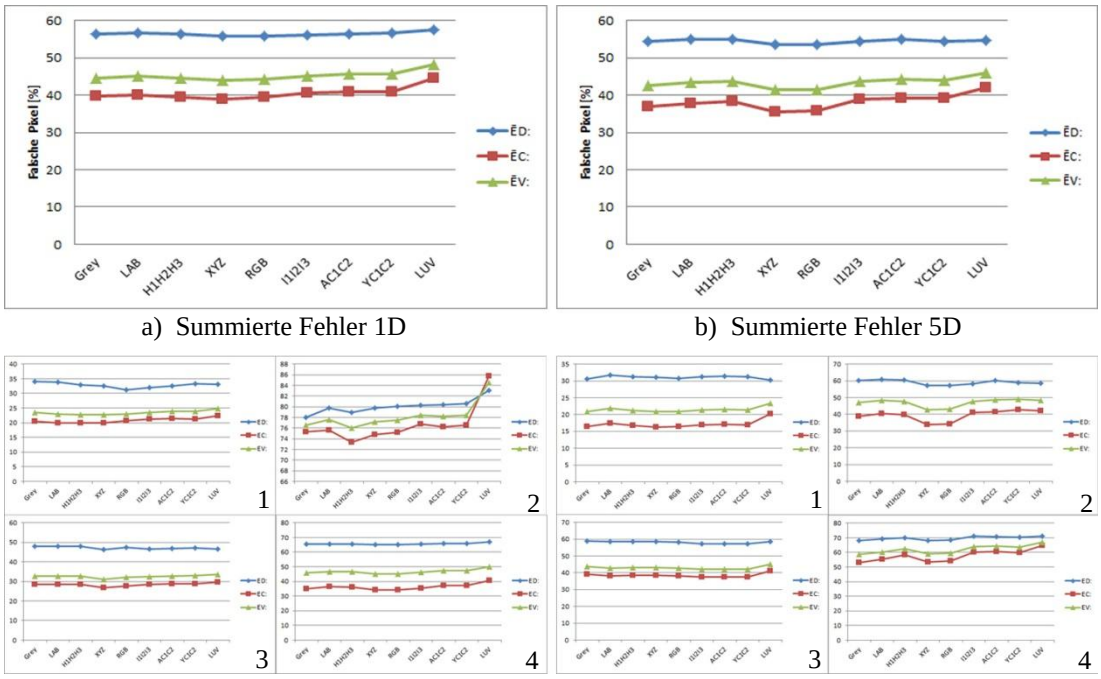


a) Grey b) RGB c) LUV d) RadioMap

Abbildung 4.2: Disparitätsfehler
Visualisierung der berechneten Disparitätsfehler für a) intensitätsbasiertes, b) RGB-basiertes und c) LUV-basierte Farbmatching. Graue Bereiche zeigen Fehler in verdeckten Regionen, schwarze Bereiche Fehler in sichtbaren Regionen. Fehler treten meist immer in den Bereichen mit hohen radiometrischen Verzerrungen auf, welche in d) visualisiert sind (helle Bereiche), die roten Bereiche in d) zeigen die Verdeckungen für die jeweilige Ansicht.

4.4 Farbbasierte Messungen

In diesem Abschnitt wird die Performance der neun verschiedenen Farbräume genauer untersucht. Dazu werden Disparitätskarten in den neun Farbräumen mit SAD als Korrespondenzmaß berechnet und deren Fehlerkarten ausgewertet. Abschnitt 4.3 zeigt, dass Farbe bei Matching-Algorithmen einen Vorteil bringt. Unter Betrachtung von Abbildung 4.2 kann eine Verringerung der Matching-Fehler bei Farbvergleichen und im speziellen im LUV-Farbraum beobachtet werden. Außerdem ist auffallend, dass Fehler speziell in Bereichen mit hohen radiometrischen Verzerrungen auftreten. Durch Aufsummieren der Fehlerraten aus den jeweils vier Test-Sets mit den zwei unterschiedlichen Kameras lässt sich dies visualisieren. $\bar{E}_D$ zeigt die Fehler in radiometrisch verzerrten Bereichen und $\bar{E}_C$ Fehler in radiometrisch unverzerrten Bereichen. $\bar{E}_V$ zeigt den mittleren Fehler über alle getesteten Sets (Abbildung 4.3).



a) Summierte Fehler 1D b) Summierte Fehler 5D

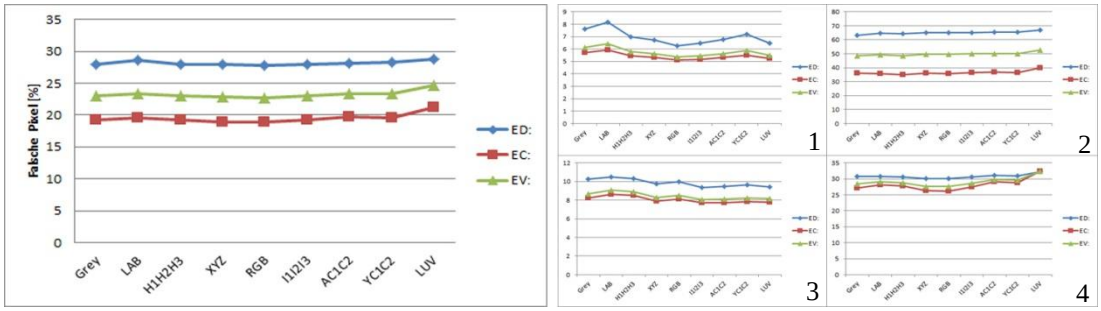
c) Einzelfehler 1D d) Einzelfehler 5D

Abbildung 4.3: Vergleich der farbbasierten Messungen
Abbildung a) und b) zeigen die summierten Auswertungen der vier Test-Sets für die jeweiligen Kameras. c) und d) zeigen die Einzelfehlersummen auf den jeweils vier Testsets. Alle Berechnungen wurden auf neun verschiedenen Farbräumen durchgeführt. $\bar{E}_D$ zeigt die Fehler in radiometrisch verzerrten Bereichen, $\bar{E}_C$ die Fehler in radiometrisch unverzerrten Bereichen und $\bar{E}_V$ den mittleren Fehler über alle Sets.

Vergleicht man die Aussagen aus [3] nun mit den hier verwendeten Auswertungen, so lässt sich aus Abbildung 4.3 heraus nicht bestätigen, dass Farbe weniger Fehler produziert als Graustufenauswertungen. Unter Betracht der Summenfehler von Abbildung 4.3 a) und b) zeigt sich keine deutliche Verbesserung der Fehlerrate in radiometrisch verzerrten Bereichen ($\bar{E}_D$). Es zeigt sich eher eine Zunahme der Fehlerrate speziell im LUV-Farbraum. Betrachtet man jedoch die Einzelauswertungen der Test-Sets so zeigt sich zumindest in Abbildung 4.3 1c) und 3c) sowie Abbildung 4.3 1d) und 2d) eine geringfügige Verbesserung der Fehlerrate $\bar{E}_D$. Warum Abbildung 4.3 2c) eine derartige Verschlechterung der Fehlermessungen im LUV-Farbraum hervorruft, lässt sich allerdings nicht erklären. Eine weitere Erkenntnis aus [3], nämlich, dass Farbe in radiometrisch reinen Bereichen eines Bildes wenig Einfluss hat, lässt sich aus beiden Vergleichsmessungen herauslesen, da der mittlere Fehler $\bar{E}_V$ dasselbe Verhalten aufweist wie die Fehlerrate in radiometrisch unverzerrten Bereichen (Abbildung 4.3, $\bar{E}_C$ und $\bar{E}_V$).

Warum in den hier verwendeten Sets keine konkreten Aussagen über die Verbesserungen beim farbbasierten Matching getroffen werden können, liegt vermutlich daran, dass alle vier Test-Sets sehr viele Fehler speziell im Bereich des Hintergrundes produzieren. Dies könnte möglicherweise daran liegen, dass der Hintergrund weniger Textur aufweist, wodurch die Korrespondenzsuche sehr schwer wird. Eine weitere Erklärung könnte sein, dass sich bei den Aufnahmen der Sets die Belichtungszeit zwischen linkem und rechtem Bild geringfügig geändert hat. Zu Testzwecken und der Visualisierung wurde der Hintergrund bei der Fehlerberechnung manuell entfernt. Für die Aufnahmen mit der Canon 1D erhält man Abbildung 4.4.

Unter Betracht der summierten Fehlerraten für die vier Sets (Abbildung 4.4 a)) zeigen sich wiederum kaum Verbesserungen in Bezug auf Farbmessungen. Die um ca. 50% verringerte Fehlerraten im Vergleich zu Abbildung 4.3 a) ist auf die herausgenommenen Hintergrundfehler zurückzuführen. Betrachtet man allerdings Abbildung 4.4 1b) und 3b), so lässt sich eine Fehlerverringerung bei den Farbmessungen erkennen, was wiederum die Aussagen aus [3] bestärkt. Abbildung 4.4 2b) und 4b) hingegen zeigen eine Erhöhung der Fehler. Ähnliche Ergebnisse



a) Summierte Fehler 1D b) Einzelfehler 1D

Abbildung 4.4: Farbbasierte Messungen ohne Hintergrund
Durch das manuelle Entfernen des Hintergrundes aus den Auswertungen wird speziell in den Einzelauswertungen 1b) und 3b) eine Fehlerverringerung erkennbar. 2b) und 4b) hingegen zeigen wiederum eher eine Erhöhung des Fehlers speziell im LUV-Farbraum.

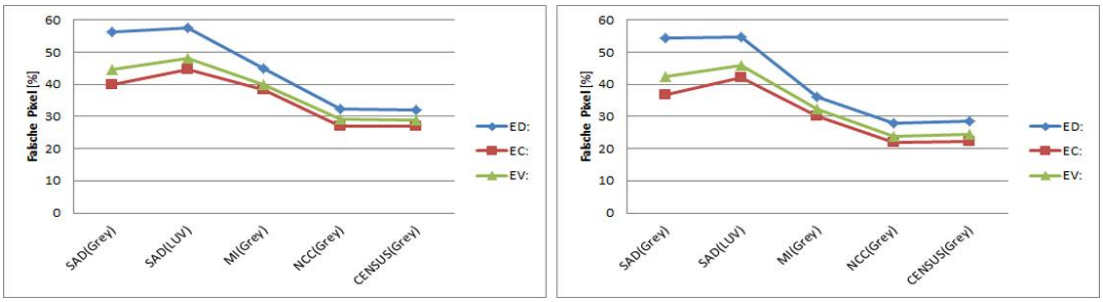
zeigen auch die Auswertungen auf dem Set, welches mit der Canon 5D aufgenommen wurde.

Prinzipiell lässt sich somit in dieser Arbeit und mit den hier verwendeten Test-Sets eine Verbesserung bei Farbmessungen nur durch Abbildung 4.4 1b) und 3b) bestätigen. Ein möglicher Grund dafür könnten sich ändernde Bedingungen bei den Aufnahmen der Test-Sets sein. Im Gegensatz zu den Auswertungen und den verwendeten Aufnahmen aus [3], welche alle unter konstanten Bedingungen aufgenommen wurden, liegt die Vermutung nahe, dass dies der ausschlaggebende Punkt für die schlechten Ergebnisse sein könnte. Wie auch bereits erwähnt, ist die Disparitäts-Berechnung unter Zuhilfenahme von absoluten Differenzen (Abschnitt 2.3.2) abhängig von den radiometrischen Verzerrungen (Abschnitt 2.3.1) der Stereobildpaare und somit von beispielsweise geringfügigen Beleuchtungsänderungen, welche bei den Aufnahmen aufgetreten sein könnten.

Um die Problematiken der Radiometrie in den Bildern zu minimieren, werden im folgenden Abschnitt Auswertungen auf radiometrisch insensitiven Korrespondenzmaßen durchgeführt.

4.5 Radiometrisch unempfindliche Messungen

Da Farbe, laut der Arbeit von [3], einen Vorteil in radiometrisch verzerrten Regionen bringt, dies jedoch mit den Test-Sets in dieser Arbeit nur schwer nachzuvollziehen ist, sollen nun weiters Algorithmen getestet werden, welche unempfindlich gegenüber solchen Regionen sind. Als Vergleichsmessungen werden, äquivalent zu



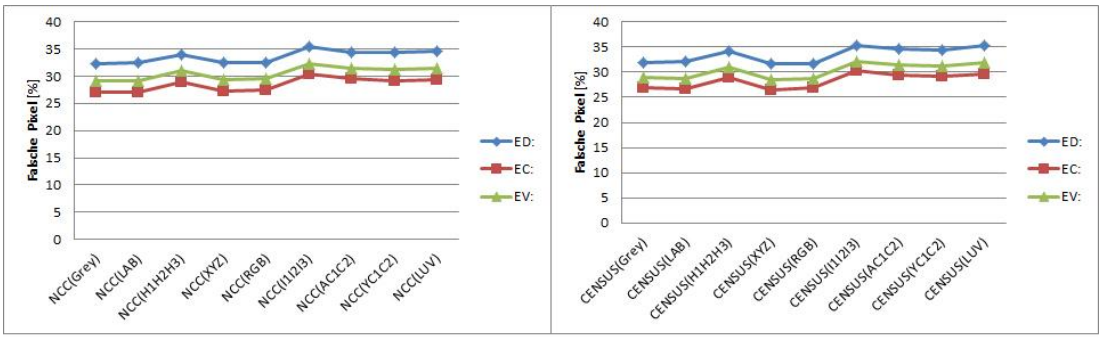
a) Sets 1D b) Sets 5D

Abbildung 4.5: Farbbasierte vs. radiometrisch insensitive Messungen
Gegenüberstellung von farbbasierter Messung (Grey, LUV mittels Summe der absoluten Differenzen) zu radiometrisch insensitiven Messungen (Grey mittels Mutual Information, Normalized Cross Correlation und Census). Abbildung a) und b) zeigen Auswertungen der vier Test-Sets für die jeweiligen Kameras.

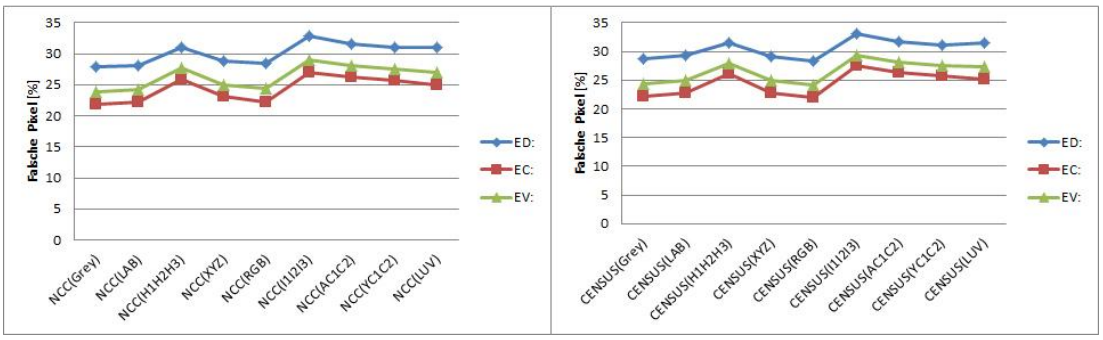
[3], Messungen für Graustufenbilder sowie dem LUV-Farbraum mit der Summe der absoluten Differenzen (SAD) als Korrespondenzfunktion berechnet und den Korrespondenzberechnungen Mutual Information (MI), Normalized Cross-Correlation (NCC) und Census, angewendet auf Graustufenbildern, gegenüber gestellt. Abbildung 4.5 zeigt die Messungen aus den jeweils vier, in dieser Arbeit verwendeten Test-Sets.

Wie in Abbildung 4.5 erkennbar, bestätigen die Auswertungen die Effektivität von radiometrisch insensitiven Messungen und zeigen deutlich deren verbesserte Erkennungsrate in den Datensets. Vorteil dieser Messungen ist, dass sie nicht nur unempfindlich gegenüber radiometrischen Verzerrungen sind, sondern auch in radiometrisch nichtverzerrten Bereichen die Fehlerrate stark reduzieren. Unter Betrachtung von Abbildung 4.5 ist eine Steigerung der Leistung von ca. 15% für den mittleren Fehler $\bar{E}_V$ erkennbar. Die Auswertung der Daten der Canon 5D ergibt sogar eine Steigerung von ca. 20% gegenüber der Berechnung der absoluten Differenzen.

Äquivalent zu [3] zeigen sich somit bei radiometrisch insensitiven Messungen deutlich verringerte Fehlerraten in den Test-Sets. Somit kann eine Kernaussage aus [3], nämlich dass ein genereller Einsatz von radiometrisch insensitiven Messungen aufgrund der sehr hohen Leistungssteigerung (siehe Census und NCC) den standardisierten Korrespondenzberechnungen wie beispielsweise der Berechnung der absoluten Differenz (SAD) bevorzugt werden sollten, bestätigt werden.



NCC a) Sets 1D CENSUS



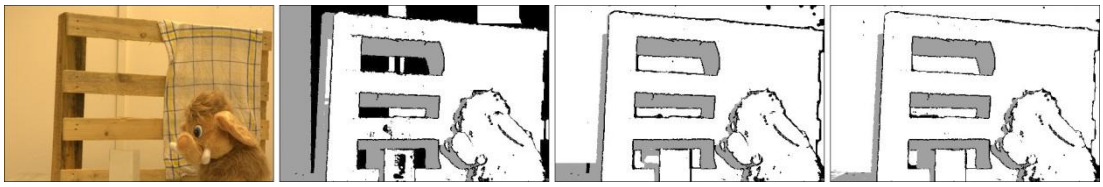
NCC b) Sets 5D CENSUS

Abbildung 4.6: NCC vs. Census
Vergleich von NCC und Census. Beste Ergebnisse werden in den Auswertungen auf Graustufenbildern erzielt. Farbe verschlechtert die Fehlerrate in den meisten Fällen. Abbildung a) und b) zeigen Auswertungen der vier Test-Sets für die jeweiligen Kameras.

Um diese Aussage weiter zu festigen, wird ein Vergleich von Census und NCC über alle neun, hier getesteten Farbräume durchgeführt. Abbildung 4.6 zeigt die Berechnungen in dieser Arbeit.

Dabei wird deutlich, dass radiometrisch insensitive Messungen gegenüber farbbasierten Berechnungen die Fehlerrate senken und Farbe in solchen Berechnungen das Ergebnis sogar noch verschlechtern kann. Betrachtet man Abbildung 4.6, so kann man im Vergleich zu den farbbasierten Messungen (Abbildung 4.3 a und b) eine Verbesserung der Fehlerrate des mittleren Fehlers $\bar{E}_V$ von knapp 20% erkennen. Dies liegt daran, dass radiometrisch insensitive Messungen sehr viel weniger Fehler in wenig texturierten Bereichen produzieren und damit bessere und genauere Disparitätskarten erstellen als farbbasierte Messungen (Abbildung 4.7).

Somit können die Aussagen von [3] in dieser Arbeit bestätigt werden. Farbe hilft zwar in radiometrisch verzerrten Bereichen und verringert somit die Fehlerrate, wird aber bei radiometrisch insensitiven Messungen unwichtig.



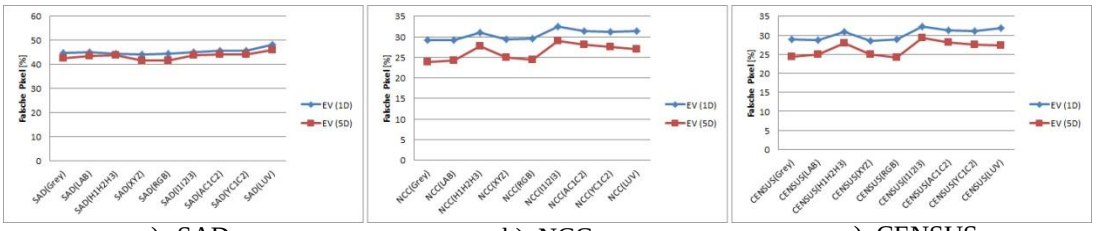
Initialbild SAD (Grey) NCC (Grey) Census (Grey)

Abbildung 4.7: Vergleich SAD zu NCC zu Census
Test-Set 1 der Canon 1D Mark II. Im direkten Vergleich wird klar ersichtlich, dass radiometrisch insensitive Berechnungen auch in Bereichen mit wenig Textur (hier Hintergrund) erheblich weniger Fehler produzieren (schwarze Bereiche) und somit deutlich bessere Ergebnisse liefern.

Zusammenfassend kann gesagt werden, dass sich die Erkenntnisse aus [3] in Bezug auf Farbmessungen für die hier verwendeten Testsets nicht nachvollziehen lassen. Eine deutliche Verringerung der Fehlerrate ist hingegen bei radiometrisch insensitiven Messungen erkennbar, wodurch sich die Aussagen in [3] für Aufnahmen unter sich eventuell ändernden Bedingungen bestätigen lassen. Bei radiometrisch insensitiven Messungen ist weiters auch erkennbar, dass Farbberechnungen zu einer Verschlechterung des Ergebnisses führen. Radiometrisch insensitive Messungen zeigen eine Verringerung der Fehler in radiometrisch verzerrten sowie unverzerrten Bereichen und senken Berechnungsfehler um einen vergleichsweise hohen prozentuellen Anteil (Abbildung 4.6). Ein weiterer Vorteil ist auch, dass diese Berechnungen offensichtlich besser in wenig texturierten Bereichen arbeiten, was darauf zurückzuführen ist, dass deren Anfälligkeit gegenüber Sensorrauschen oder leicht unterschiedliche Belichtungszeiten in den Bildern gegenüber farbbasierten Messungen kaum vorhanden ist.

4.6 Verwendung unterschiedlicher Kameras

Eine weitere Überprüfung dieser Arbeit ist es, ob der Einsatz unterschiedlicher Kameras eine Rolle für Stereo-Korrespondenz Algorithmen spielen. Die ursprüngliche Idee war es, mehrere unterschiedliche Kameras und Kameratypen zu testen. Da jedoch eine große Menge an Bildern, möglichst automatisiert, aufgenommen werden muss und die meisten Hersteller keine Software-Development-Kits zur Verfügung stellen, beschränkt sich diese Arbeit auf zwei Canon Kameras, für welche eine Automatisierungssoftware erstellt werden konnte. Unter Verwendung standen eine Canon EOS-1D Mark III sowie eine Canon EOS-5D Mark II. Die Aufnahmen wurden alle unter denselben Bedingungen (Blende,



a) SAD b) NCC c) CENSUS

Abbildung 4.8: Kameravergleich 1D vs. 5D
Vergleich der beiden verwendeten Kameras Canon 1D – Mark III und Canon 5D – Mark II. Die Visualisierung bezieht sich auf den mittleren Fehler $\bar{E}_V$ ausgewertet auf den hier verwendeten vier Test-Sets. Eine durchschnittliche Verringerung der Fehlerrate, speziell bei radio-metrisch insensitiven Messungen, von ca. 10% - 15% ist erkennbar.

Belichtung) und mit der jeweilig höchsten verwendbaren Kamera-RAW-Auflösung erstellt. Eine Visualisierung der mittleren Fehlerrate für diese beiden Kameras über die hier verwendeten Test-Sets zeigt generell wenig Unterschiede in den Kameras (Abbildung 4.8).

Bei der Auswertung von Farbmessungen mittels absoluter Differenzen als Korrespondenzfunktion zeigen sich kaum Änderungen in der durchschnittlichen Fehlerrate. Beide Kameras verhalten sich ähnlich, wobei die Canon 5D in den Farbräumen XYZ und RGB geringfügig weniger Fehler produziert (Abbildung 4.8 a). Auch bei den Messungen für radiometrisch insensitive Berechnungen zeigen beide Kameras dasselbe Verhalten, wobei die Canon 5D im Allgemeinen um ca. 5% weniger Fehler produziert als die Canon 1D (Abbildung 4.8 b und c). Auffällig ist allerdings, dass zwischen den farbbasierten Messungen und den radiometrisch insensitiven Messungen eine Verbesserung der Fehlerrate von 10% - 15% festgestellt werden kann. Wie bereits erwähnt, liegt die allgemein hohe Fehlerrate für die in dieser Arbeit verwendeten Test-Sets an den vielen Fehlern, welche speziell in den untexturierten Bereichen, also im Hintergrund, produziert werden. Dies liegt daran, dass farbbasierte Messungen Probleme mit diesen Bereichen haben.

Beide Canon Kameras arbeiten mit ähnlichen CMOS-Chips zur Generierung der Bilder (Abbildung 3.3). Während die Canon 1D mit dem CMOS APS-H Sensor arbeitet (29 x 19 mm), welcher einen Formfaktor von 1,3 aufweist, greift die Canon 5D auf einen Vollformat Sensor (36 x 24 mm) mit Formfaktor 1 zurück. Laut Canon haben alle Sensoren ein minimiertes Grundrauschen sowie weißes Rauschen was selbst bei hohen ISO-Werten weder Auflösung noch Bildqualität beeinträchtigt.

Durch einen Tiefpass-Filter vor dem Sensor werden Farbfehler und Moiré-Effekte unterdrückt und ein sehr hoher Dynamikbereich für einen erhöhten Kontrastumfang und optimierter Detailgenauigkeit runden den CMOS-Sensor ab [45].

Aufgrund dieser Ähnlichkeiten beeinträchtigt dies vermutlich die Ergebnisse in dieser Arbeit nur wenig. Was allerdings diese geringfügigen Veränderungen hervorrufen könnte, wäre beispielsweise das Zuschneiden (Cropping – Abschnitt 3.1.2) der Canon 5D Bilder auf die Bildgröße der Canon 1D, welche wegen des Formfaktors von 1,3 um diesen Faktor kleinere Bilder erzeugt. Während sich die Prozedur des Zuschneidens auf allen Bildern gleich verhält ist es allerdings unmöglich, beide Kameras beim Wechseln für die Aufnahmen der Bild-Sets exakt gleich auf die bewegliche Schiene (Abschnitt 3.1) zu montieren. Beim direkten Vergleich der beschnitten Bilder wurde ersichtlich, dass die Bildausschnitte von Canon 1D und Canon 5D nicht exakt dieselben sind. Dies könnte wiederum eine Erklärung für die sich leicht ändernden Datenwerte, speziell in radiometrisch insensitiven Messungen, sein. Dieses Problem lässt sich jedoch mit dem momentanen Setup im Labor kaum beheben.

Ein weiteres Problem könnte durch die Konvertierung der Kamera-RAW-Bilder in ein MATLAB-lesbares Format sein. Die Konvertierung wurde mittels DCRAW [46] durchgeführt, wobei auffiel, dass die Software aus den relativ farbintensiven Bildern der 5D dunklere Bilder erzeugte (Abbildung 4.9). Allerdings macht dies die Software auch für die Bilder der 1D, wodurch das Verhältnis der Farbwerte dasselbe bleiben müsste. Obwohl sich dies kaum auf die Erstellung der Ground-Truth Daten auswirkt (Abschnitt 3.1), kann nicht ausgeschlossen werden, dass die von [3] zur Verfügung gestellte Software damit Probleme hat. Eine Vermeidung dieser Tatsache wäre ein

1D: Original Konvertiert 5D: Original Konvertiert

Abbildung 4.9: Vergleich der dcraw-Umwandlung
Veranschaulichung der Unterschiede bei der Umwandlung der Kamera-RAW-Bilder in ein MATLAB-lesbares Format mittels Freeware Software DCRAW [46]. Ersichtlich ist, dass die Software die Farbe und Helligkeit der Bilder beeinflusst.

direktes Arbeiten auf den Kamera-RAW-Formaten, wobei nicht sichergestellt ist, ob diese auch von der Software aus [3] gelesen werden können.

Ein dritter Aspekt für die Datenänderung könnte die Rektifizierung sein (Abschnitt 2.1.3). Da in dieser Arbeit ein Interpolationsverfahren verwendet wird, also die Ausrichtung der Bilder zueinander mittels Passpunkten durchgeführt wird, könnten sich die rektifizierten Bilder der Canon 1D von denen der Canon 5D geringfügig unterscheiden. Eine Umgehung dieses Problems würde sich vermutlich mittels parametrischer Verfahren bewerkstelligen lassen, was eine Kalibrierung der Kamera und der Beleuchtungsquelle voraussetzt.

Als letzte mögliche Beeinflussung der Ergebnisse kann hier noch die Skalierung der einzelnen Bilder genannt werden. Durch die Rektifizierung ist, aufgrund von zu großen Datenmengen für die Rektifizierungs-Software, eine Verkleinerung der Originalbildgröße um 50% notwendig. Des Weiteren werden die Bilder nach der Erstellung der Ground-Truth Disparitätsdaten nochmals um 75% verkleinert, um mit gängigen Stereo-Algorithmen arbeiten zu können.

Zusammengefasst kann in dieser Arbeit keine genaue Aussage getroffen werden, ob unterschiedliche Kameras die Ergebnisse der berechneten Disparitätskarten mit der Software aus [3] beeinflussen. Da keine erheblichen Verbesserungen zwischen den beiden Kameras festgestellt werden konnten und die zuvor genannten Aspekte ebenfalls zu Problemen führen könnten, bleibt die Frage offen, ob die Auswertungen kameraabhängig sind. Deutlich ersichtlich ist allerdings, dass radiometrisch insensitive Messungen auf Grauwertbildern weniger Fehler produzieren und dadurch bessere Ergebnisse liefern. Dies liegt auch in der Insensitivität gegenüber Beleuchtungsunterschieden, welche auch durch einige der vorhin aufgezählten Probleme auftreten könnten.

5. Schlussfolgerung

In dieser Arbeit wurde die Wichtigkeit von Farbbildern für das Erstellen von Stereo-Tiefenkarten untersucht. Das Hauptziel dieser Arbeit war es, die Erkenntnisse einer vergangenen experimentellen Studie [3] auf einem neuen Stereo-Bilddatensatz zu untersuchen. Dies ist insofern von Bedeutung, da die Experimente in [3] auf Bildern durchgeführt wurden, welche mit ein und derselben Kamera unter ähnlichen Beleuchtungsbedingungen aufgenommen wurden. Diese Diplomarbeit gibt eine Antwort auf die Frage ob sich die in [3] gewonnen Einsichten auch auf anderes Bildmaterial übertragen lassen. Zu diesem Zweck wurde in dieser Arbeit ein neuer Stereo-Bilddatensatz mit zwei unterschiedlichen Kameramodellen aufgenommen. Für jedes Stereo-Bildpaar wurde eine Referenz-Tiefenkarte mit Hilfe von Strukturiertem Licht erstellt. Danach wurden für jedes Bildpaar Stereo-Tiefenkarten berechnet. Die Berechnung wurde, wie in [3], in verschiedenen Farbräumen sowie mit einer Reihe von unterschiedlichen Stereo-Korrespondenzmaßen durchgeführt. Die dadurch erhaltenen Stereo-Tiefenkarten wurden mit den Referenzlösungen verglichen und die Ergebnisse ausgewertet.

Die Experimente bestätigen die Einsicht aus [3], dass in radiometrisch unverzerrten Bildbereichen Farbbilder wenig Einfluss auf das Ergebnis haben. Grauwertbilder erzielen ähnliche Resultate. Im Gegensatz zu [3] kann jedoch nicht bestätigt werden, dass in radiometrisch verzerrten Bildregionen Farbinformation hilfreich sein kann. Auf diesem Datensatz zeigen auch Grauwertbilder eine ähnliche Leistung. Diese Diskrepanz in den Resultaten kann sich zum Teil dadurch erklären lassen, dass die Fehlerraten durch große Fehler in untexturierten Bereichen dominiert werden. Dadurch fallen etwaige kleine Verbesserungen, welche durch Farbbilder entstehen können, kaum ins Gewicht.

Des Weiteren bestätigen die Messungen die Einsicht von [3], dass radiometrisch insensitive Korrelationsmaße wie Mutual Information (MI), Census und Normalized Cross Correlation (NCC), deutlich besser als reine farbbasierte Maße in allen Bildregionen arbeiten. Zudem hat sich gezeigt, dass radiometrisch insensitive Maße auf Grauwertbildern bessere Ergebnisse erzielen als auf Farbbildern. Daher ist es ratsam, radiometrisch insensitive Maße auf Grauwertbildern zu verwenden.

Darüber hinaus wurde in dieser Arbeit die Qualität der beiden verwendeten Kameratypen (Canon EOS-1D Mark III und Canon EOS-5D Mark II) untersucht. Diese Untersuchung brachte jedoch keine eindeutigen Ergebnisse. Zwar lässt sich eine Verringerung der Fehlerrate von ca. 5% bei der Canon 5D erkennen, allerdings konnte sich nicht klären lassen ob dieser Qualitätsgewinn auf die Kamera zurückzuführen ist, da die Bilder der beide Kameras aufnahmebedingt nicht den exakt gleichen Szenenausschnitt zeigen.

Somit kann abschließend zusammengefasst werden, dass radiometrisch insensitive Messungen im Allgemeinen den Farbmessungen vorzuziehen sind. Die zwei unterschiedlichen Kameratypen produzieren ähnliche Ergebnisse und ein Unterschied zwischen den beiden Kameras konnte in dieser Arbeit nicht festgestellt werden.

Literaturverzeichnis

[1] P. Besl, *Active, optical range imaging sensors*.: Jorge L. C. Sanz, Editor, Advanced in Machine Vision, 1-63, 1989.

[2] S. Birchfield and C. Tomasi, *A pixel dissimilarity measure that is insensitive to image sampling*.: PAMI, Vol. 20, 401-406, 1998.

[3] M. Bleyer and S. Chambon, *Does color really help in dense stereo matching*.: 3DPVT, 2010.

[4] M. Bleyer, S. Chambon, U. Poppe, and M. Gelautz, *Evaluation of different methods for using colour information in global stereo matching*.: ISPRS, Vol. 37, 415-422, 2008.

[5] M. Bleyer and M. Gelautz, *Simple but effective tree structures for dynamic programming-based stereo matching*.: VISAPP, Vol. 2, 415-422, 2008.

[6] Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts*.: PAMI, Vol. 23 (11), 1222-1239, 2001.

[7] S. Chambon and A. Crouzil, *Color correlation-based matching*.: Int. J. Robotics Autom., Vol. 20 (2), 78-85, 2005.

[8] B. Curless and M. Levoy, *Better optical triangulation through spacetime analysis*.: ICCV, 987-994, 1995.

[9] P. Felzenszwalb and D. Huttenlocher, *Efficient belief propagation for early vision*.: IJCV, Vol. 70 (1), 41-54, 2006.

[10] T. Frohlinghaus and J. M. Buhmann, *Regularizing phased stereo*.: ICPR, Vol. 1, 451-455, 1996.

[11] A. Fusiello, V. Roberto, E. Trucco, and J. Garibaldi, *Efficient stereo with multiple windowing*.: CVPR, 858-863, 1997.

[12] G. Golub and C. F. Van Loan, *Matrix computation, third edition*.: The Johns Hopkins University Press London, 1996.

[13] J. Gühring, *3D-Erfassung und Objektrekonstruktion mittels Streifenprojektion*.: Institut für Photogrammetrie, Universität Stuttgart, 2002.

[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*.: Cambridge University Press, 2004.

[15] H. Hirschmüller, *Accurate and efficient stereo processing by semi-global matching and mutual information*.: CVPR, Vol. 2, 807-814, 2005.

[16] H. Hirschmüller and D. Scharstein, *Evaluation of stereo matching cost on images with radiometric differences*.: PAMI, Vol. 31 (9), 1582-1599, 2009.

[17] H. Hirschmüller, P. Innocent, and J. Garibaldi, *Real-time correlation-based dtereo vision with reduced border errors*.: IJCV, Vol. 47 (1), 229-246, 2002.

[18] H. Hirschmüller, *Stereo processing by semiglobal matching and mutual information*.: PAMI, Vol. 30 (2), 328-341, 2008.

[19] W. Hoff and N. Ahuja, *Surfaces from stereo: integrating feature matching, disparity estimation and contour detection*.: PAMI, Vol. 11 (2), 121-136, 1989.

[20] G. Hu and G. Stockman, *3-D surface solution using structured light and constraint propagation*.: PAMI, Vol. 11 (4), 390-402, 1989.

[21] V. Kolmogorov and C. Rother, *Minimizing nonsubmodular functions with graph cuts - a review*.: PAMI, Vol. 29 (7), 1274-1279, 2007.

[22] A. Koschan, *Dense stereo correspondence using polychromatic block matching*.: CAIP, Vol. 719, 538-542, 1993.

[23] V. Lempitsky, C. Rother, and A. Blake, *LogCut - Efficient graph cut optimization for markov random fields*.: ICCV, 1-8, 2007.

[24] D. Lowe, *Distinctive image features from scale-invariant keypoints*.: IJCV, Vol. 60 (2), 91-110, 2004.

[25] H.-G. Maas, *Robust automatic surface reconstruction with structured light*.: IAPRS, Vol. 29, 709-713, 1992.

[26] T. Meltzer, C. Yanover, and Y. Weiss, *Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation*.: ICCV, Vol. 1, 428-435, 2005.

[27] K. Mühlmann, D. Maier, J. Hesser, and R. Männer, *Calculating dense disparity maps from color stereo images, an efficient implementation*.: IJCV, Vol. 47 (1-3), 79-88, 2002.

[28] M. Okutomi, G. Tomita, and O. Yoshizaki, *Color stereo matching and its application to 3-D measurement of optic nerve head*.: ICPR, Vol. 1, 509-513, 1992.

[29] K. Pulli, H. Abi-Rached, T. Duchamp, L. G. Shapiro, and W. Stuetzle, *Acquisition and visualization of colored 3D objects*.: ICPR, 11-15, 1998.

[30] J. Salvi Mas, *An approach to coded structured light to obtain three dimensional information*.: ISBN: 84-8458-107-1, 2002.

[31] K. Sato and S. Inokuchi, *Three-dimensional surface measurement by space encoding range imaging*.: J. of Robotic Systems, Vol. 2, 27-39, 1985.

[32] D. Scharstein and R. Szeliski, *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*.: IJCV, Vol. 47 (1-3), 7-42, 2002.

[33] D. Scharstein and R. Szelinski, *High-accuracy stereo depth maps using structured light*.: CVPR, Vol. 1, 195-202, 2003.

[34] N. Shrikhande and G. Stockman, *Surface orientation from a projected grid*.: PAMI, Vol. 11 (6), 650-655, 1989.

[35] E. Trucco, R. B. Fisher, A. W. Fitzgibbon, and D. K. Naidu, *Calibration, data consistency and model acquisition with laser stripers*.: IJCIM, Vol. 11 (4), 293-310, 1998.

[36] F. M. Wahl, *A coded light approach for depth map acquisition*.: Springer Verlag, Mustererkennung, 12-17, 1986.

[37] G. Wiora, *Optische 3D-Messtechnik: Präzise Gestaltvermessung mit einem erweiterten Streifenprojektionsverfahren*. Heidelberg: Ruprechts-Karl-Universität, 2001.

[38] R. Zabih and J. Wood, *Non-parametric local transforms for computing visual correspondence*.: ECCV, Vol. 801, 151-158, 1994.

[39] Z. Zhang, *A flexible new technique for camera calibration*.: PAMI, Vol. 22 (11), 1330-1334, 2000.

[40] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, *A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry*.: Artificial Intelligence Journal, Vol. 78 (1-2), 87–119, 1995.

[41] Z. Zhang, *Determining the epipolar geometry and its uncertainty: a review*.: IJCV, Vol. 27 (2), 161-198, 1998.

[42] M. Bleyer, *Does color really help in dense stereo matching?*: http://www.ims.tuwien.ac.at/research/stereo_matching/presentations/DoesColorReallyHelp.pdf, Zugriff: August 2011.

[43] M. Bleyer, *Research Homepage*.: http://www.ims.tuwien.ac.at/staff_detail.php?ims_id=bleyer, Zugriff: Juli 2011.

[44] M. Bleyer, *Stereo Vision*.: http://www.ims.tuwien.ac.at/teaching_detail.php?ims_id=188.HQK, Zugriff: Oktober 2011.

[45] Canon, *Produkte*.: www.canon.at, Zugriff: Juni 2011.

[46] D. Coffin, *Decoding raw digital photos in Linux*.: http://www.cybercom.net/~dcoffin/dcraw/, Zugriff: Mai 2011.

[47] M. Dailey, *Match two sets of feature descriptors without spatial information*.: http://www.cs.ait.ac.th/~mdailey/home/index.php?action=home, Zugriff: August 2011.

[48] A. Fusiello, *Epipolar rectification toolkit*.: www.sci.univr.it/~fusiello/, Zugriff: April 2011.

[49] Wikipedia, *Structured-light 3D scanner*.: http://en.wikipedia.org/wiki/Structured-light_3D_scanner, Zugriff: Juni 2011.

Abbildungsverzeichnis

Abbildung 2.1: Aufnahme eines Bildes ... 5
Abbildung 2.2: Das Korrespondenzproblem ... 6
Abbildung 2.3: Epipolargeometrie ... 7
Abbildung 2.4: Epipolare Rektifizierung ... 8
Abbildung 2.5: Prinzip der Rektifizierung ... 9
Abbildung 2.6: Parametrische Rektifizierung ... 10
Abbildung 2.7: Rektifizierung durch Interpolation ... 11
Abbildung 2.8: Prinzip der Triangulation ... 12
Abbildung 2.9: Verdeckungen ... 14
Abbildung 2.10: Das Verdeckungsproblem ... 14
Abbildung 2.11: Ähnlichkeitsbedingung ... 16
Abbildung 2.12: Fensterbasierte Methoden ... 17
Abbildung 2.13: Sliding Windows ... 18
Abbildung 2.14: Das Foreground Fattening Problem ... 18
Abbildung 2.15: Prinzip des Streifenprojektionssensors ... 21
Abbildung 2.16: Streifenprojektion mit Binärmustern ... 22
Abbildung 2.17: Streifencodierung ... 23
Abbildung 2.18: Binärcode Sequenz ... 23
Abbildung 2.19: Sequenzvergleich ... 24
Abbildung 2.20: Graycode Sequenz ... 25
Abbildung 2.21: Graustufenauflösung einer Sinuswelle ... 25
Abbildung 2.22: Phasen-Shift Methode ... 26
Abbildung 2.23: Line-Shift Methode ... 27
Abbildung 2.24: Space-Time Methode ... 28
Abbildung 2.25: Problematik Stereo-Matching ... 30
Abbildung 2.26: Extrakton radiometrischer Daten ... 32
Abbildung 2.27: Radiometrische Verzerrungen ... 33
Abbildung 2.28: Mutual Information ... 34
Abbildung 2.29: Census Berechnung ... 36
Abbildung 3.2: Binärsequenz ... 39
Abbildung 3.1: Laboraufbau ... 38
Abbildung 3.3: Verwendete Kameratypen ... 40
Abbildung 3.4: Cropping von Bildern ... 41
Abbildung 3.5: Rektifizierung ... 42
Abbildung 3.6: Graycode Sequenz ... 42
Abbildung 3.7: Graycode Aufnahmesequenz ... 43
Abbildung 3.8: Codierung ... 44
Abbildung 3.9: Sichtdisparitäten ... 46
Abbildung 3.10: Beleuchtungsdisparitäten ... 51
Abbildung 3.11: Kombination von Disparitätskarten ... 52

Abbildung 4.1: Radiometrische Karten 54
Abbildung 4.2: Disparitätsfehler 56
Abbildung 4.3: Vergleich der farbbasierten Messungen 57
Abbildung 4.4: Farbbasierte Messungen ohne Hintergrund 59
Abbildung 4.5: Farbbasierte vs. radiometrisch insensitive Messungen 60
Abbildung 4.7: Vergleich SAD zu NCC zu Census 62
Abbildung 4.7: NCC vs. Census 61
Abbildung 4.8: Kameravergleich 1D vs. 5D 63
Abbildung 4.9: Vergleich der dcraw-Umwandlung 64