



**Evaluierung konkurrierender Datenterme für das  
Stereokorrespondenzproblem**

**DIPLOMARBEIT**

zur Erlangung des akademischen Grades

**Diplom-Ingenieur/in**

im Rahmen des Studiums

**Computergraphik & Digitale Bildverarbeitung**

ausgeführt von

**Roland Gross**

Matrikelnummer 0147161

am:

Institut für Softwaretechnik und Interaktive Systeme

Betreuung:

Betreuer: Ao.Univ.Prof. Dipl.-Ing. Mag. rer. nat. Dr. techn. Margrit Gelautz

Mitwirkung: Projektass.(FWF) Dipl.-Ing. Dr. techn. Michael Bleyer

Wien, 23. 09. 2009

\_\_\_\_\_  
(Unterschrift Verfasser/in)

\_\_\_\_\_  
(Unterschrift Betreuer/in)

# Erklärung

Roland Gross  
Köhlgasse 1  
2700 Wiener Neustadt

Hiermit erkläre ich, 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, 23. September 2009*

---

(Unterschrift)

# Abstract

At the moment, many stereo matching algorithms (global and local) only use intensity values for the disparity map calculation. However, recent studies show that the usage of color information can improve the robustness and quality of matching algorithms. Especially images with radiometric differences pose a challenge to these algorithms. Examples of radiometric differences are different exposure times or different lightning conditions. This thesis deals with the question whether the usage of color information and radiometric transforms can give better results than calculations based solely on intensity. Therefore, based on ten datasets, five color spaces (Intensities, RGB, AC1C2, I1I2I3, LUV), five radiometric transforms (Mean, Laplacian of Gaussian, Rank, SoftRank, Bilateral Subtraction), four cost calculation functions (sum of absolute differences, Birchfield Tomasi, normalized cross correlation, hierarchical mutual information) and two optimization approaches (local, global) are evaluated. Furthermore the usage of weights in radiometric transforms, normalized cross correlation and aggregation are evaluated. A comparison of two weight calculation schemes (Yoon, Geodesic) is also included. The results of the performed evaluation show that the usage of color, radiometric transforms and weights significantly improves the quality of the generated disparity maps.

# Zusammenfassung

Im Moment verwenden viele Stereo Matching Algorithmen (global sowie lokal) nur Intensitätswerte um Disparitätsbilder zu berechnen. Aktuelle Studien zeigen jedoch, dass die Verwendung von Farbinformation die Robustheit und Ergebnisqualität der Matchingkostenberechnung erhöhen kann. Dies zeigt sich besonders deutlich an radiometrisch verzerrten Basisdaten. Einfache Beispiele für radiometrische Verzerrungen sind zum Beispiel unterschiedliche Belichtungszeiten oder Beleuchtungsverhältnisse. Die Diplomarbeit beschäftigt sich mit der Frage, ob die Verwendung von Farbinformation und radiometrischen Transformationen bessere Ergebnisse als herkömmliche intensitätsbasierte Ansätze liefern kann. Dazu werden fünf Farbräume (Graustufen, RGB, AC1C2, I1I2I3, LUV), fünf radiometrische Transformationen (Mittelwert, Marr-Hildreth Operator, Rank, SoftRank, Bilaterale Subtraktion), vier Kostenberechnungsverfahren (Summe absoluter Differenzen, Birchfield Tomasi, Normalisierte Kreuzkorrelation, Hierarchische Transinformation) und zwei Optimierungsverfahren (Lokal, Global) in einer Evaluierung auf zehn Testdatensätzen untersucht. Weiters wird untersucht, ob eine Verwendung von Gewichten in radiometrischen Transformationen, normalisierter Kreuzkorrelation und Aggregation zu besseren Ergebnissen führt. Dabei wird auch ein Vergleich zwischen der Gewichtsrechnung nach Yoon und geodätischen Gewichten angestellt. Die Ergebnisse der Evaluierung zeigen, dass die Verwendung von Farbinformation, radiometrischen Transformationen und Gewichten die Ergebnisqualität signifikant erhöht.

# Danksagungen

Ich möchte diese Arbeit meinen Eltern, die immer für mich da sind, widmen.

Besonders bedanken möchte ich mich bei meinem Betreuer Dr. Michael Bleyer, der immer ein offenes Ohr für meine vielen Fragen hatte und ohne dessen Hilfe diese Arbeit nicht möglich gewesen wäre.

Des Weiteren bedanke ich mich bei Prof. Dr. Gelautz für Ihre Unterstützung und die gute Atmosphäre, in der ich die Arbeit erstellen konnte.

# Inhaltsverzeichnis

<b>Erklärung</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Zusammenfassung</b>	<b>iii</b>
<b>Danksagungen</b>	<b>iv</b>
<b>1 Einführung</b>	<b>1</b>
1.1 Motivation und verwandte Arbeiten . . . . .	2
1.2 Aufbau der Arbeit . . . . .	3
<b>2 Stereo Vision</b>	<b>4</b>
<b>3 Farbraumtransformationen</b>	<b>8</b>
<b>4 Radiometrische Transformationen</b>	<b>12</b>
4.1 Radiometrische Verzerrungen . . . . .	12
4.2 Mittelwert Transformation (Mean) . . . . .	14
4.3 Marr-Hildreth Operator (LoG) . . . . .	14
4.4 Rank Transformation (Rank) . . . . .	15
4.5 Soft Rank Transformation (SoftRank) . . . . .	16
4.6 Bilaterale Subtraktion Transformation (BilSub) . . . . .	16
4.7 Gewichtete Transformationen . . . . .	17
<b>5 Kostenberechnung</b>	<b>19</b>
5.1 Differenzmaße . . . . .	20
5.2 Birchfield und Tomasi (BT) . . . . .	21
5.3 Normalisierte Kreuzrelation (NCC) . . . . .	22
5.4 Hierarchische Transinformation (HMI) . . . . .	23

<b>6</b>	<b>Aggregation</b>	<b>27</b>
6.1	Quadratische Aggregation . . . . .	27
6.2	Gewichtete Aggregation . . . . .	28
6.2.1	Gewichte nach Yoon . . . . .	28
6.2.2	Geodätische Gewichte . . . . .	30
<b>7</b>	<b>Disparitätsberechnung</b>	<b>33</b>
7.1	Lokaler Ansatz . . . . .	33
7.1.1	Behandlung von Verdeckungen . . . . .	34
7.2	Globaler Ansatz . . . . .	35
7.2.1	Energiefunktion . . . . .	35
7.2.2	Energieoptimierung mit Graph Cuts . . . . .	37
<b>8</b>	<b>Evaluierung</b>	<b>42</b>
<b>9</b>	<b>Ergebnisse</b>	<b>46</b>
9.1	Farbraum . . . . .	46
9.1.1	Lokaler Ansatz . . . . .	47
9.1.2	Globaler Ansatz . . . . .	49
9.2	Radiometrische Transformation . . . . .	51
9.2.1	Lokaler Ansatz . . . . .	51
9.2.2	Globaler Ansatz . . . . .	56
9.3	Kostenberechnungsmethode . . . . .	59
9.3.1	Lokaler Ansatz . . . . .	59
9.3.2	Globaler Ansatz . . . . .	61
9.4	Fenstergröße, Gewichtsverwendung und Gewichtstyp . . . . .	61
9.5	Beispielhafte Ergebnisse . . . . .	63
9.5.1	Lokaler Ansatz . . . . .	66
9.5.2	Globaler Ansatz . . . . .	66
<b>10</b>	<b>Zusammenfassung</b>	<b>72</b>
	<b>Literaturverzeichnis</b>	<b>76</b>
<b>A</b>	<b>Bilder</b>	<b>81</b>
<b>B</b>	<b>Abkürzungen und Übersetzungen</b>	<b>83</b>

# Abbildungsverzeichnis

1.1	Architekturschema . . . . .	3
2.1	Epipolargeometrie . . . . .	5
2.2	Rektifizierung . . . . .	6
2.3	Kostenberechnung . . . . .	7
4.1	Radiometrische Transformationen . . . . .	13
4.2	Laplacian of Gaussian . . . . .	14
4.3	Rank Transformation . . . . .	15
4.4	Gewichtete Transformationen . . . . .	17
5.1	Interpolation bei Birchfield Tomasi . . . . .	22
6.1	Geodätische Distanzberechnung . . . . .	31
7.1	Pixelklassifizierung . . . . .	38
7.2	Beispielgraph . . . . .	39
8.1	Testdaten . . . . .	44
9.1	Farbräume im Vergleich bei lokaler Optimierung . . . . .	48
9.2	Testdaten Farbperformance bei Gewichtsverwendung lokal . . . . .	48
9.3	Farbräume im Vergleich bei globaler Optimierung . . . . .	50
9.4	Testdaten Farbperformance bei Gewichtsverwendung global . . . . .	52
9.5	Radiometrische Transformationen im Vergleich bei lokaler Optimierung . . . . .	52
9.6	Radiometrische Transformation Vergleich ohne Gewichte . . . . .	54
9.7	Radiometrische Transformation Vergleich Farbe . . . . .	54
9.8	Radiometrische Transformation Vergleich Gewichte lokal . . . . .	55
9.9	Gewichtete radiometrische Transformation Vergleich Farbe . . . . .	57
9.10	Radiometrische Transformation im Vergleich bei globaler Optimierung . . . . .	57



## *Abbildungsverzeichnis*

9.11 Radiometrische Transformation Vergleich Farbe . . . . .	58
9.12 Radiometrische Transformation Vergleich Gewichte . . . . .	60
9.13 Gewichtete radiometrische Transformation Vergleich Farbe . . . . .	60
9.14 Kostenberechnungsmethoden im Vergleich bei lokaler Optimierung . .	62
9.15 Vergleich SAD/BT . . . . .	62
9.16 Kostenberechnungsmethoden im Vergleich bei globaler Optimierung .	62
9.17 Fenstergröße im Vergleich bei lokaler Optimierung . . . . .	64
9.18 Gewichtsverwendung im Vergleich bei lokaler Optimierung . . . . .	65
9.19 Vergleich von Fenstergröße bei Gewichtsverwendung . . . . .	65
9.20 Vergleich von Gewichtsverwendung im Aggregationsschritt . . . . .	67
9.21 Beispielhafte Ergebnisse bei lokaler Optimierung . . . . .	68
9.22 Beispielhafte Ergebnisse bei globaler Optimierung . . . . .	70
A.1 Architekturbild . . . . .	82

# Tabellenverzeichnis

3.1	Farbraumtransformationen . . . . .	11
8.1	Verwendete Kombinationen in der Evaluierung . . . . .	43
9.1	Vergleich Farbräume lokal . . . . .	48
9.2	Vergleich Farbräume bei lokaler Optimierung mit Gewichtsverwendung	50
9.3	Vergleich Farbräume global . . . . .	50
9.4	Vergleich Farbräume bei globaler Optimierung mit Gewichtsverwendung	52
9.5	Vergleich radiometrische Transformationen lokal . . . . .	54
9.6	Radiometrische Transformation Vergleich Farbe . . . . .	55
9.7	Radiometrische Transformation Vergleich Gewichte lokal . . . . .	55
9.8	Vergleich radiometrische Transformationen global . . . . .	57
9.9	Radiometrische Transformation Vergleich Farbe . . . . .	58
9.10	Radiometrische Transformation Vergleich Gewichte global . . . . .	60
9.11	Vergleich Kostenberechnungsmethoden lokal . . . . .	64
9.12	Vergleich Kostenberechnungsmethoden global . . . . .	64
9.13	Vergleich Fenstergrößen lokal . . . . .	64
9.14	Vergleich Gewichte lokal . . . . .	65
9.15	Vergleich Gewichte in Aggregation lokal . . . . .	67
9.16	Verwendetes Parameterset bei lokaler Optimierung . . . . .	67
9.17	Lokale Ergebnisse in Prozent . . . . .	69
9.18	Verwendetes Parameterset bei globaler Optimierung . . . . .	69
9.19	Globale Ergebnisse in Prozent . . . . .	71

# Kapitel 1

## Einführung

Computer Stereo Vision ist ein sehr aktives Forschungsgebiet mit einer, für Informatikbegriffe, langen Geschichte. [BF82] definiert Computer Stereo Vision als die Rekonstruktion von dreidimensionalen Charakteristika einer Szene aufgenommen aus verschiedenen Perspektiven von zwei oder mehr Kameras.<sup>1</sup> Die Entwicklungen im Computer Stereo Vision bis 1981 können in [BF82] eingesehen werden. Darauf aufbauend liefert [DA89] einen Überblick für die Fortschritte in den Achtzigerjahren. Aktuelle Entwicklungen können in [BBH03] oder [SS02] nachgelesen werden.

Frühe Ansätze waren, durch begrenzte Rechenleistung, hauptsächlich auf Ergebnisse für spezielle Bereiche (Merkmale) beschränkt, während aktuelle Ansätze durchgehende Ergebnisse, sogenannte „dichte“ Ergebnisse liefern. Üblicherweise wird zwischen lokalen und globalen Methoden unterschieden, wobei generell lokale Methoden ungenauer und schneller, globale Methoden genauer aber langsamer sind. Die hier betrachteten Ansätze beziehen sich auf Stereo Vision mit zwei Kameras.

Zielgruppe dieser Arbeit stellen Wissenschaftlerinnen im Bereich Computer Stereo Vision, Studenten der Computergraphik, Techniker aus dem Bereich Maschinelles Sehen oder der Robotik und interessierte Techniker mit Grundkenntnissen in den Bereichen Computergrafik, Stereo Vision und Farbe dar. Es wird eine grundlegende Einführung gegeben und immer auf weiterführende Literatur verwiesen, jedoch wird ein gewisses Basiswissen vorausgesetzt.

---

<sup>1</sup>Übersetzt aus dem Englischen: „...as the recovery of the three-dimensional characteristics of a scene from multiple images taken from different points of view.“ [BF82, p.1].

## 1.1 Motivation und verwandte Arbeiten

Der Großteil der aktuell eingesetzten Methoden zur Korrespondenzsuche verwendet zur Berechnung der Ergebnisse nur Intensitätswerte und keine Farbinformation (siehe auch Evaluierung aktueller Techniken in [SS02]). Dieses bewusste Verwerfen von Information scheint zunächst unlogisch, beruht aber auf der Tatsache, dass bisher intensitätsbasierte Ansätze ähnliche Ergebnisse wie farbbasierte Ansätze lieferten. Neuere Studien wie zum Beispiel [BCPG08] oder [CC04] zeigen jedoch, dass eine Verwendung von Farbinformation eine Verbesserung der Ergebnisse erwirken kann. Weiters zeigen Studien wie [HS09], dass eine Berücksichtigung von möglichen radiometrischen Verzerrungen in der Stereobearbeitung ebenfalls zu einer Ergebnisverbesserung führt, jedoch je nach verwendeten Kostenberechnungsverfahren unterschiedliche Ansätze zum Einsatz kommen sollten, um optimale Ergebnisse zu erzielen. Darüber hinaus können diverse Aggregationsverfahren, welche zum Beispiel besonders Farbinformation oder Nachbarschaftsbeziehungen berücksichtigen, eingesetzt werden.

Eine Evaluierung von Kostenberechnungsverfahren im Allgemeinen und in Bezug auf radiometrische Transformationen wurde zum Beispiel in [HS07] durchgeführt. Umfassender beschreibt [HS09] ebenfalls Kostenberechnungsverfahren und radiometrische Verzerrungen, sowie radiometrische Transformationen, die verwendet werden können um diesen Verzerrungen entgegenzuwirken. Eine Beschreibung von gewichteter Aggregation der Kosten (siehe Abschnitt 6) findet sich in [YK06] und [HBGR09]. Wobei [YK06] Farbe und Nachbarschaftsbeziehungen verwendet und [HBGR09] als Weiterentwicklung gesehen werden kann, welche Pfade auf der Grundlage von Abstand und Farbe berechnet.

Ziel dieser Arbeit ist die Untersuchung von Kombinationen von Farbraum, radiometrischer Transformation, Kostenberechnungsverfahren, Aggregationsart und Disparitätsberechnungsmethode in Hinblick auf Prozent der richtig berechneten Pixel im Disparitätsbild. Besonderes Augenmerk liegt auf der Verbesserung des sogenannten Daten-terms (siehe Abschnitt 7.2). Weiters grenzt sich diese Arbeit von bestehenden Arbeiten dahingehend ab, als dass die Verwendung von Gewichten nicht auf den Aggregations-schritt beschränkt wird, sondern auch bei radiometrischen Transformationen und normalisierter Kreuzkorrelation, also fensterbasierten Ansätzen, untersucht wird. Um eine Aussage treffen zu können, werden mögliche und sinnvolle Kombinationen experimentell evaluiert und in Relation zu bestehenden Daten gesetzt.

## 1.2 Aufbau der Arbeit

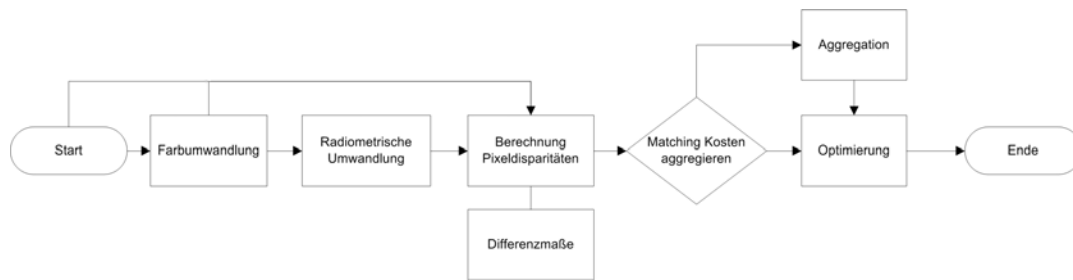


Abbildung 1.1: Schematische Übersicht der Architektur der erstellten Stereosoftware

Die vorliegende Arbeit gliedert sich in 10 Kapitel. Die Gliederung folgt in großen Zügen dem architektonischen Aufbau der verwendeten Stereosoftware, welche in [Abbildung 2.3](#) schematisch dargestellt ist. Einen detaillierten Überblick über die verwendete Architektur und die unterstützten Komponenten bietet [Abbildung A.1](#) im Anhang.

[Kapitel 2](#) gibt einen kurzen Überblick über Stereo Vision und erklärt wichtige Grundlagen die zum Verständnis der Arbeit notwendig sind. [Kapitel 3](#) beschäftigt sich mit Farbe und deren Verwendung in Stereo Vision. [Kapitel 4](#) zeigt mögliche radiometrische Probleme und Ansätze zu deren Vermeidung auf. [Kapitel 5](#) beschreibt verschiedene Methoden zur Kostenberechnung. [Kapitel 6](#) zeigt, wie die berechneten Kosten aggregiert und verfeinert werden können. [Kapitel 7](#) gibt einen Überblick über Verfahren zur Berechnung der Pixeldisparitäten, welche das Ergebnis der meisten Stereoverfahren darstellen. [Kapitel 8](#) beschreibt die verwendete Testumgebung und Einschränkungen, sowie die Testdurchführung. [Kapitel 9](#) zeigt Ergebnisse auf und führt eine Diskussion in Hinblick auf Anwendbarkeit und Relevanz der erzielten Ergebnisse durch. [Kapitel 10](#) gibt eine abschließende Zusammenfassung.

Englische Begriffe werden größtenteils zwecks Lesbarkeit ins Deutsche übersetzt. Begriffe, für die entweder keine brauchbaren Übersetzungen gefunden werden konnten oder eine Übersetzung die Verständlichkeit beeinträchtigen würde, werden im Englischen belassen. In [Anhang B](#) findet sich ein kleines Wörterbuch, das die wichtigsten Begriffe in Deutsch und Englisch auflistet.

# Kapitel 2

## Stereo Vision

Stereo Vision oder stereoskopisches Sehen ist ein Teilgebiet der Computer Vision. Stereo Vision versucht, aus zwei oder mehr Aufnahmen der Umwelt eine 3D Repräsentation abzuleiten. Aufnahmen können von verschiedensten Geräten und Anordnungen dieser Geräte stammen. Ein übliches Beispiel wären zwei parallel, in einem gewissen Abstand angeordnete Foto- oder Videokameras, die der Anordnung der menschlichen Augen nachempfunden sind. Studien wie [BF82], [DA89], [Bro92] oder [SS02] zeigen, dass Stereo Vision ein sehr aktives Forschungsgebiet mit, für Informatikbegriffe, langer Geschichte ist.

[BBH03] beschreibt die grundlegende Basis für Stereo Vision als:

*„The fundamental basis for stereo is the fact that a single three-dimensional physical location projects to a unique pair of image locations in two observing cameras.“*

Also als die Tatsache, dass eine dreidimensionale physikalische Position immer auf ein eindeutiges Paar von Bildpositionen, auf den Bildebenen, zweier auf die Szene gerichteter Kameras abgebildet werden kann. Diese Tatsache kann genutzt werden, um aus zwei oder mehr Kamerabildern die Bildpositionen zu finden, die genau einer dreidimensionalen Position im Raum entsprechen.

Abbildung 2.1 zeigt eine mögliche Anordnung zweier Kameras in einem Stereosystem. Die beiden Kameras mit den optischen Zentren  $O_L$  und  $O_R$  besitzen die selbe Grundlinie  $T$  parallel zur jeweiligen Kamera x-Achse in Entfernung  $f$  der fokalen Länge der Kamera. Ein physischer Punkt  $P$  in Entfernung  $Z$  wird nun jeweils auf den Punkt  $p$  und

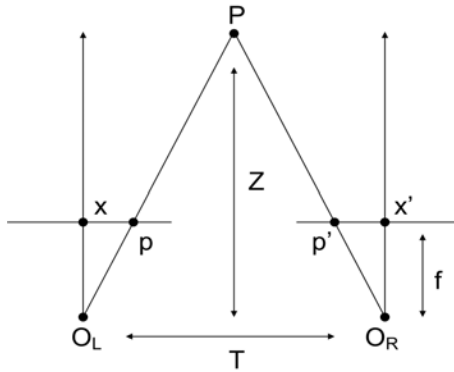


Abbildung 2.1: Epipolargeometrie (Darstellung aus [BBH03])

den dazu korrespondierenden Punkt  $p'$  abgebildet. Die Differenz zwischen den Abständen  $x$  und  $x'$  ergibt dann die sogenannte Disparität  $d$ . Fasst man die Disparitäten für alle Punkte zusammen und speichert sie in einer zweidimensionalen Karte, nennt man diese Karte Disparitätskarte (Disparity Map) oder Disparitätsbild. Aus der Disparität kann mit Hilfe der Formel 2.1 die Tiefeninformation  $Z$  für den Punkt  $P$  abgeleitet werden.

$$Z = f \frac{T}{d} \quad (2.1)$$

Es reicht also, die Disparität zweier Punkte zu bestimmen, um daraus Tiefeninformation gewinnen zu können. Laut [BBH03] sind die primären Probleme im Stereo Vision: Kamerakalibration, Korrespondenzsuche und 3D-Rekonstruktion. Die zwei Gebiete Kamerakalibration und 3D-Rekonstruktion können als gelöst betrachtet werden und entsprechen nicht dem Fokus dieser Arbeit. Korrespondenzsuche bezeichnet die Suche von Punkten  $p$  und  $p'$  im linken und rechten Kamerabild, welche den selben Szenenpunkt darstellen. Abbildung 2.1 zeigt eine idealisierte Version der Wirklichkeit, in der sich korrespondierende Punkte immer auf der gleichen Rasterzeile befinden und somit der eigentliche zweidimensionale Suchraum um eine Dimension eingeschränkt wird. Oftmals entsprechen reale Stereokonfigurationen dieser Annahme nicht, können jedoch unter der Verwendung der sogenannten Epipolareinschränkung modifiziert (rektifiziert) und somit wieder in den eindimensionalen Fall übergeführt werden, wie in Abbildung 2.2 ersichtlich.

Die beiden optischen Zentren  $O_L$  und  $O_R$  bilden mit dem Punkt  $P$  die sogenannte Epipolarebene. Die Epipolarebene schneidet die zwei Kamerabilder in den Epipolarlinien. Die Punkte, an denen die Epipolarlinien die Grundlinie  $T$  kreuzen bezeichnet man als Epipole  $e$  und  $e'$ , und sie entsprechen dem jeweiligen gegenüberliegenden opti-

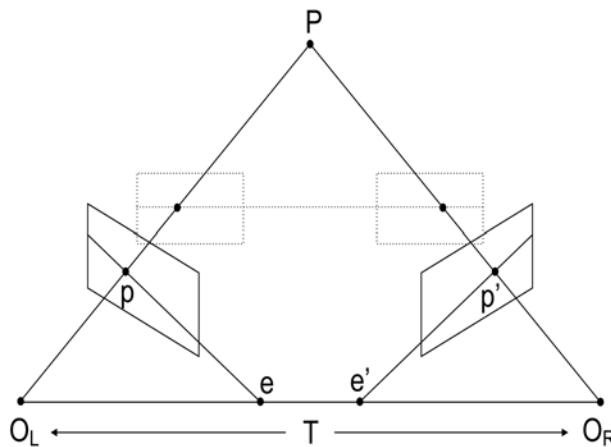


Abbildung 2.2: Rektifizierung (Darstellung aus [BBH03])

schen Zentrum projiziert auf die aktuelle Kamera. Korrespondierende Punkte  $p$  und  $p'$  müssen sich auf den Epipolarlinien befinden. Modifiziert man die Bilder nun so, dass sich korrespondierende Epipolarlinien auf korrespondierenden horizontalen Rasterzeilen befinden, reduziert sich das Problem wieder auf eine eindimensionale Suche in der jeweiligen Rasterzeile.

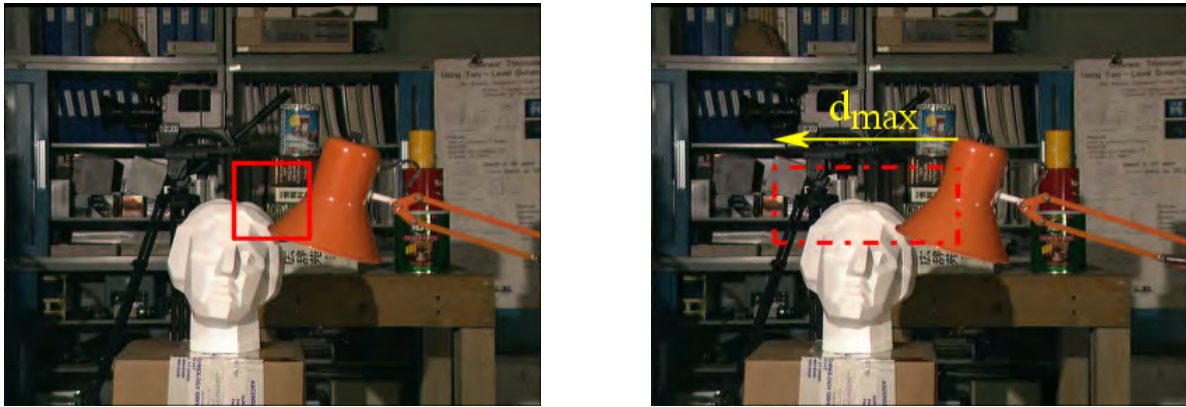
Nach [SS02] kann ein Großteil der heute verwendeten Stereoalgorithmen in vier große algorithmische Blöcke eingeteilt werden:

1. Kostenberechnung
2. Kostenaggregation
3. Disparitätsberechnung und Optimierung
4. Nachbearbeitung der Disparitäten (Verfeinerung)

Im Kostenberechnungsschritt wird für jede Pixelkombination (linkes Bild  $\leftrightarrow$  rechtes Bild) die zugehörige Wahrscheinlichkeit, dass ein Punkt im linken Bild genau dem betrachteten Punkt im rechten Bild entspricht, berechnet. Diese Berechnung erfolgt für alle Punkte bis zu einer definierten maximalen Disparität  $d_{max}$ . Für einfache lokale, fensterbasierte Methoden zeigt Abbildung 2.3 ein Beispiel.

In Abbildung 2.3 wird der Suchbereich im linken Bild fix gehalten, während im rechten Bild pro Durchlauf der Bereich um eine Disparitätsstufe verschoben wird. Pro Disparitätsstufe wird der Grad der Entsprechung des Suchfensters mit dem aktuellen Ver-





(a) Linkes Kamerabild

(b) Rechtes Kamerabild

Abbildung 2.3: Beispiel lokaler, fensterbasierter Kostenberechnung.

gleichsfenster bestimmt und in einem sogenannten Disparitätsraumbild<sup>1</sup> [BI99, YYL93] gespeichert. Das Disparitätsraumbild speichert für jeden Pixel  $(x, y)$  und jede Disparität  $d$  im Suchraum die zugehörige Wahrscheinlichkeit der Entsprechung mit dem Vergleichswert des anderen Bildes. Es stellt also eine dreidimensionale Struktur mit Koordinaten der Form  $DSI(x, y, d)$  dar.

Die Kostenaggregation verwendet diverse Methoden um die berechneten Werte des Disparitätsraumbildes zu summieren oder Durchschnitte zu bilden. Es ergibt sich somit ein durchgehendes, das heißt für alle Pixel eines Bildes definiertes, Disparitätsbild.

Der nächste Schritt ist die Berechnung des eigentlichen Disparitätsbildes (Disparity Map) mittels Auswahl geeigneter Disparitäten je Pixel aus dem Disparitätsraumbild. Diese Auswahl kann einfach (zum Beispiel: Suche der kleinsten Kosten in lokalen Methoden) oder komplex (zum Beispiel: Energieminimierungsproblem in globalen Methoden) ausgestaltet sein.

Abschließend kann noch eine Verfeinerung des berechneten Disparitätsbildes erfolgen, um zum Beispiel Verdeckungen zu behandeln oder Bildrauschen auszugleichen.

---

<sup>1</sup>Disparity space image - DSI

# Kapitel 3

## Farbraumtransformationen

Erster Schritt in der in dieser Arbeit verwendeten Stereosoftware ist die Umwandlung der Eingangsdaten in einen möglichen Farbraum. Üblicherweise wird im Stereo nur mit Intensitätswerten gearbeitet. Dazu werden die drei Kanäle der meist als RGB<sup>1</sup> vorliegenden Bilder in Intensitätswerte transformiert. Bei dieser Transformation von drei Farbkanälen auf einen Intensitätswert scheint intuitiv Information verloren zu gehen. Dennoch findet Farbinformation erstaunlich wenig Verwendung in den aktuellen Stereoverfahren. Ein Grund könnte die Tatsache sein, dass viele Algorithmen auch mit Intensitätswerten gute bis sehr gute Ergebnisse liefern, ohne den zusätzlichen Aufwand der Berechnung in drei Farbkanälen zu rechtfertigen. Trotzdem berichten einige Studien wie [CC04] oder [BCPG08] über eine Verbesserung der Ergebnisse im Zusammenhang mit Verwendung von Farbinformation. Daher soll auch in dieser Arbeit die Rolle von verschiedenen Farbräumen im Stereo untersucht werden. Wichtig scheint vor allem der Einfluss von Farbe auf den sogenannten Datenterm (siehe Abschnitt 7.2). Laut [BCPG08] lässt sich durch die Verwendung von Farbinformation eine Verbesserung gegenüber Intensitäten erzielen, jedoch nicht auf allen untersuchten Testdatensätzen. Die Verbesserung hängt auch von der radiometrischen Beschaffenheit der Eingangsdaten ab (siehe Abschnitt 4).

Um Farbräume ineinander überführen zu können, muss man zuerst die möglichen Farbräume definieren und kategorisieren. Nach [NVP00, p. 1] lassen sich Farbräume in vier große Gruppen einteilen: *Primärsysteme*, *Luminanz-Chrominanz Systeme*, Systeme, die an der *menschlichen Wahrnehmung orientiert* sind, und *Systeme statistisch unabhängiger Komponenten*.

---

<sup>1</sup>Rot-Grün-Blau

*Primärsysteme* beruhen auf der Trichromatenfarbtheorie. Sie besagt, dass es möglich ist, jede, für das menschliche Auge sichtbare, Farbe aus drei Grundfarben zu mischen. Somit definiert sich ein *Primärsystem* über die Auswahl von drei Primärfarben und eines Referenzweißpunktes oder einer Referenzbeleuchtung. Die CIE<sup>2</sup>, als das Normungsinstitut für Farbräume, definiert zwei Standards für Primärsysteme, CIE RGB und CIE XYZ. Dabei wurde der XYZ Farbraum künstlich geschaffen, um einfachere Verarbeitung in diversen Anwendungsfällen zu gewährleisten.

*Luminanz-Chrominanz Systeme* definieren eine Luminanz Komponente und zwei Chrominanz Komponenten. Erste Unterkategorie sind solche Systeme, die in der menschlichen Wahrnehmung uniform sind. Die Euklidische Distanz in den beiden Primärsystemen RGB und XYZ verhält sich nicht der Wahrnehmung des menschlichen Betrachters entsprechend. Daher hat die CIE die zwei uniformen Systeme  $L^*u^*v^*$  und  $L^*a^*b^*$  eingeführt, wobei  $L^*$  die Helligkeit repräsentiert und  $(u^*, v^*)$  und  $(a^*, b^*)$  Chrominanzkoordinaten sind [CIE86]. Weiters kann man Systeme für den Einsatz im Fernsehen unterscheiden. Hier ist besonders die Trennung zwischen Luminanz und Chrominanz für eine Darstellung wichtig. Der amerikanische NTSC Standard verwendet hier das YIQ System, und die europäische Fernsehnorm greift auf das YUV System zurück. Nächste Kategorie von Luminanz-Chrominanz Systemen sind Systeme, die auf der Gegensatzlichkeitstheorie von Farben basieren. Diese besagt, dass Farbe ins Gehirn über eine achromatische und zwei chromatische Komponenten übertragen wird. Beispiele sind das AC1C2 System oder das  $(wb, rg, by)$  System.

An der *Wahrnehmung orientierte Systeme* versuchen die subjektive menschliche Wahrnehmung von Farbe zu quantifizieren und bedienen sich dabei oftmals der Komponenten Intensität (I), Farbton (F) und Sättigung (S). Nach [NVP00, p.2] unterscheidet man zwischen Polarkoordinatensystemen und IFS Systemen. Polarkoordinatensysteme verwenden zur Adressierung die Komponenten Helligkeit, Chrominanz und Farbton, während IFS Systeme zum Beispiel geometrisch als Dreieckssysteme oder Pyramiden mit sechseckiger Grundfläche definiert sein können (siehe auch [Shi95]).

*Systeme statistisch unabhängiger Komponenten* versuchen nicht-korrelierte Komponenten für die Beschreibung des Farbraumes zu finden und damit mögliche Abhängigkeiten, die in den anderen Farbräumen zwischen den Komponenten auftreten können, auszuschließen. Ein Beispiel findet sich im I1I2I3 System [OKS80], das Otha mittels der Karhunen-Loeve Transformation erstellt.

---

<sup>2</sup>Commission Internationale de l'Eclairage

Üblicherweise liegen Farbbilder im RGB Farbraum vor und können mittels mathematischer Transformationen in die entsprechenden anderen Farbräume übergeführt werden. Tabelle 3.1 aus [CC04, p.2] zeigt eine Auswahl von Formeln zur Farbraumtransformation.

Nach der optionalen Transformation in den gewünschten Farbraum kann eine Bearbeitung in Bezug auf radiometrische Probleme erfolgen. Dies wird im nächsten Abschnitt behandelt. Hier sei noch auf [ST97] verwiesen, welche Farbräume und vor allem die physischen Zusammenhänge mit der menschlichen Wahrnehmung im Detail erläutern.

Tabelle 3.1: Beispiele für Farbraumtransformationen aus [CC04, p.2].

Name	Definition
GREY	$0.299 * R + 0.587 * G + 0.114 * B$
XYZ	$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
$L^* u^* v^*$	$L^* = \begin{cases} 116(Y/Y_w)^3 - 16 & \text{wenn } Y/Y_w > 0.01 \\ 903.3 Y/Y_w & \text{sonst} \end{cases}$ $u^* = 13L^*(u' - u'_w \text{ mit } u' = \frac{4X}{X+15Y+3Z})$ $v^* = 13L^*(v' - v'_w \text{ mit } v' = \frac{9Y}{X+15Y+3Z})$ <p><math>X_w, Y_w, Z_w</math> sind die Weissreferenz Komponenten</p>
$L^* a^* b^*$	$a^* = 500(f(X/X_w) - f(Y/Y_w))$ $b^* = 200(f(Y/Y_w) - f(Z/Z_w))$ $f(x) = \begin{cases} x^{\frac{1}{3}} & \text{wenn } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{sonst} \end{cases}$
$AC_1C_2$	$\begin{pmatrix} A \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
$YC_1C_2$	$\begin{pmatrix} Y \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
HSI	$I = \frac{R+G+B}{3}, S = 1 - 3 \frac{\min(R,G,B)}{R+G+B}$ $H = \begin{cases} \arccos H_1 & \text{wenn } B \leq G \\ 2\pi - \arccos H_1 & \text{sonst} \end{cases}$ $H_1 = \frac{(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)+(G-B)}}$
$I_1I_2I_3$	$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
$H_1H_2H_3$	$\begin{pmatrix} H_1 \\ H_2 \\ H_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -\frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$

# Kapitel 4

## Radiometrische Transformationen

Nächster Schritt in der Pipeline der hier verwendeten Stereosoftware stellt die radiometrische Bearbeitung der Eingangsbilder dar. Um die in den Eingangsbildern üblicherweise vorhandenen radiometrischen Verzerrungen zu reduzieren, können eine Reihe von Transformationen angewendet werden. Alle folgenden Ausführungen beziehen sich auf Intensitätswerte. Die Erweiterung auf Farbe erfolgt durch Anwendung der Transformationen auf jeden Farbkanal.

### 4.1 Radiometrische Verzerrungen

Radiometrische Verzerrungen beschreiben Störungen in den Bildern, die von diversen Ursachen ausgehen können. Stereotestdaten, welche unter kontrollierten Laborbedingungen erstellt wurden, unterscheiden sich von realen Stereodaten oftmals durch die Stärke und Anzahl solcher Störeffekte. Radiometrische Verzerrungen führen dazu, dass tatsächlich korrespondierende Pixeldaten nicht mehr idente Intensitäten oder Farbwerte enthalten und daher in der Korrespondenzsuche nicht gefunden werden. Daher ist es im Interesse guter Stereoergebnisse solche Störungen möglichst abzuschwächen. [HS07, p.3] nennt sechs solcher radiometrischen Probleme mit Stereobildern. Die globale Helligkeit kann sich linear (gain change) oder nicht linear (gamma change) ändern. Weiters kann sich die Helligkeit lokal ändern, zum Beispiel Reflexionen oder Glanzlichter auf glatten Oberflächen. Die Möglichkeit von Bildrauschen, durch zum Beispiel schlechte Sensoren, ist ebenfalls immer gegeben. Weiters können unterschiedliche Belichtungszeiten auftreten, sowie gänzlich unterschiedliche Beleuchtungsverhältnisse in Bezug auf Typ und Position der Lichtquelle.

Im Folgenden findet sich eine Auflistung von möglichen Transformationen um einer oder mehreren dieser Verzerrungen entgegenzuwirken. Abbildung 4.1 zeigt eine Übersicht der betrachteten radiometrischen Transformationen.

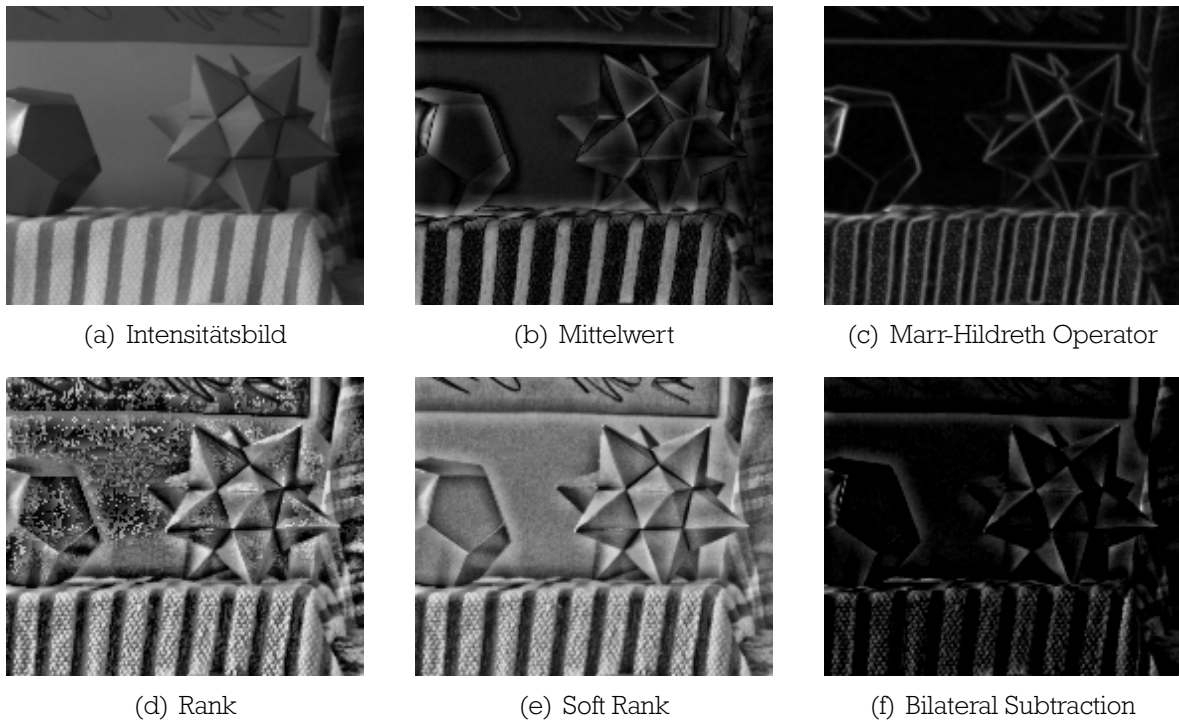


Abbildung 4.1: Anwendung verschiedener Transformationen auf einen Teil eines Testbildes (Moebius).

Die folgenden Transformationen sind filterbasierte Transformationen und können somit als separierter Vorverarbeitungsschritt durchgeführt werden. Um Verwirrungen vorzubeugen ist anzumerken, dass die beiden Verfahren Hierarchische Transformation und normalisierte Kreuzkorrelation im nachfolgenden Kapitel 5 behandelt werden. Es gilt jedoch zu beachten, dass die Auswahl des nachfolgenden Kostenberechnungsverfahrens ebenfalls einen Einfluss auf die Ergebnisse hat. Nicht alle Transformationen können mit allen Kostenberechnungsverfahren kombiniert werden, und nicht in allen Fällen ergibt sich zwangsläufig eine Verbesserung der Ergebnisse. Die Auswahl und Untersuchung von geeigneten und sinnvollen Kombinationsmöglichkeiten stellt auch ein Ziel dieser Arbeit dar.

## 4.2 Mittelwert Transformation (Mean)

Die Mittelwert Transformation (siehe Abbildung 4.1(b)) subtrahiert von jedem Pixel die gemittelten Werte seiner Umgebung. Die so durchgeführte Hintergrundsubtraktion entfernt lokale Abweichungen [HS09]. Wie Formel 4.1 zeigt, ergibt sich der neue transformierte Wert  $I_t$  als Subtraktion der Summe der Werte einer definierten quadratischen Umgebung  $N_p$  dividiert durch die Anzahl der Umgebungswerte.

$$I_t(p) = I(p) - \frac{1}{N_p} \sum_{q \in N_p} I(q) \quad (4.1)$$

## 4.3 Marr-Hildreth Operator (LoG)

Der Marr-Hildreth Operator oder Laplacian of Gaussian (siehe Abbildung 4.1(c)) ist ein spezieller Laplacefilter, der durch die Anwendung des Laplace Operators auf eine Gaussfunktion erstellt wird [MH80]. Die Berechnung erfolgt nach Formel 4.2.

$$I_t = I \otimes K_{LoG} \quad K_{LoG}(x,y) = -\frac{1}{\pi\sigma^4} \left(1 - \frac{x^2 + y^2}{2\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}\right) \quad (4.2)$$

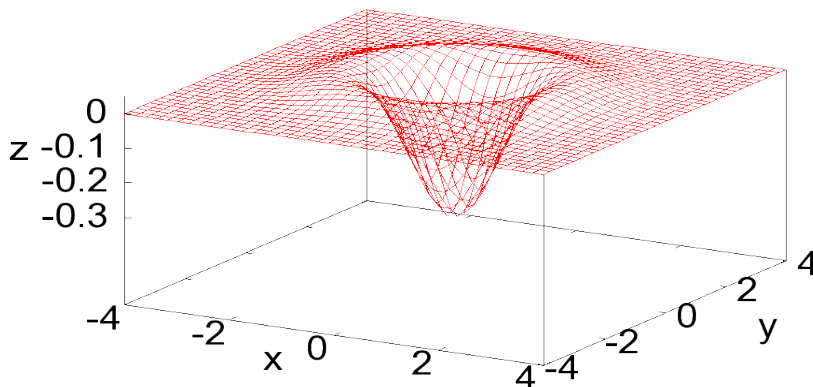


Abbildung 4.2: Laplace Funktion angewendet auf eine 2D Gaussfunktion.



Abbildung 4.2 zeigt eine zweidimensionale Darstellung mit  $\sigma = 0.97$ . Wie ersichtlich, ähnelt die Form der eines Sombreros und wird daher auch oft als Mexican Hat bezeichnet. Erfolgt eine Transformation von Bildwerten beispielsweise mittels einer Faltung mit einem 5x5 LoG Kernel mit  $\sigma = 1$ , führt dies wie in [HS07] beschrieben zu einer Glättung von Störungen und entfernt Verschiebungen in Intensitätswerten.

## 4.4 Rank Transformation (Rank)

Die Rank Transformation (siehe Abbildung 4.1(d)) wurde ursprünglich in [ZW94] vorgeschlagen um die Robustheit von fensterbasierten Methoden gegenüber Ausreißern zu verbessern. Die Rank Transformation ist eine lokale, nicht parametrisierte Transformation und als solche nur von der Reihenfolge der Intensitätswerte abhängig [BBH03]. Dies führt zu einer höheren Toleranz bei radiometrischen Verzerrungen, welche die Reihenfolge der Intensitätswerte nicht ändern. Wie Abbildung 4.3 zeigt, ersetzt die Rank Transformation den Wert eines Pixel mit der Summe der Pixelwerte in einer lokalen Umgebung, deren Intensitätswerte geringer als der Vergleichspixel sind. Im Beispiel in Abbildung 4.3 sind zwei Pixel kleiner als der Vergleichswert fünf, und somit ist das Ergebnis der Rank Transformation Zwei. Formel 4.3 zeigt die Definition der Rank Transformation. Wobei die Funktion  $T$  1 als Funktionswert liefert, wenn die enthaltene Bedingung erfüllt ist, und 0, falls die Bedingung nicht erfüllt ist.

$$I_t(p) = \sum_{q \in N_p} T[I(q) < I(p)] \quad (4.3)$$

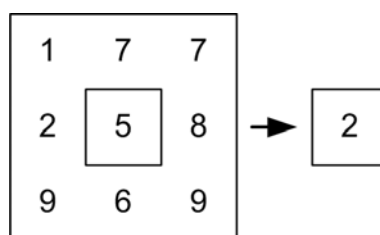


Abbildung 4.3: Beispiel einer Rank Transformation

## 4.5 Soft Rank Transformation (SoftRank)

Die Soft Rank Transformation stellt eine Erweiterung der Rank Transformation dar (siehe 4.4). Wie in Abbildung 4.1(d) ersichtlich, ist die Rank Transformation anfällig für Störungen in Bereichen mit geringer Texturierung. Die Soft Rank Transformation definiert einen linearen Übergangsbereich zwischen 0 und 1 für Bildbereiche in unmittelbarer Nähe zueinander und liefert so bessere Ergebnisse in texturarmen Regionen (siehe Abbildung 4.1(e)). Die Berechnung erfolgt nach Formel 4.4, wobei der Grenzwert  $t$  empirisch zu ermitteln ist. Diese Arbeit verwendet, wie in [HS09] empfohlen den Wert 8.

$$I_t(p) = \sum_{q \in N_p} \min(1, \max(0, \frac{I(p) - I(q)}{2t} + \frac{1}{2})) \quad (4.4)$$

## 4.6 Bilaterale Subtraktion Transformation (BilSub)

Die letzte behandelte Transformation ist Hintergrundsubtraktion mittels bilateraler Filterung<sup>1</sup> [ACM04] (siehe Abbildung 4.1(f)). Laut [TM98a] summiert der zweiseitige Filter Nachbarwerte gewichtet nach Nähe und Farbsimilarität. Dies führt zu einer Weichzeichnung ohne Kantenunschärfen zu erzeugen. Die zusätzliche Hintergrundsubtraktion entfernt lokale Ausreißer. Die Berechnung erfolgt nach Formel 4.5. Die Nachbarschaft  $s$  definiert die Stärke des Weichzeichnungseffektes und wird im Speziellen durch den Parameter  $\sigma_s$ , der in dieser Arbeit und [HS09] auf 3 gesetzt ist, festgelegt. Die Berücksichtigung der radiometrischen Distanz  $r$  soll ein zu starkes Weichzeichnen über Texturgrenzen (Kanten) hinweg verhindern. Der Parameter  $\sigma_r$  legt auch hier die Stärke des Effektes fest und ist in dieser Arbeit und [HS09] auf 20 gesetzt.

$$I_t(p) = I(p) - \frac{\sum_{q \in N_p} I(q) e^s e^r}{\sum_{q \in N_p} e^s e^r} \quad s = -\frac{(q-p)^2}{2\sigma_s^2} \quad r = -\frac{(I(q) - I(p))^2}{2\sigma_r^2} \quad (4.5)$$

---

<sup>1</sup>Background subtraction by bilateral filtering

## 4.7 Gewichtete Transformationen

Mit Ausnahme der bilateralen Subtraktion arbeiten alle bisher genannten Transformationen auf quadratischen Fenstern (Nachbarschaftsbereichen). Dies hat zur Folge, dass besonders bei Disparitätsgrenzen (Kanten, Unstetigkeitsstellen) Störeffekte, wie zum Beispiel Verwischung, auftreten. Um diese Effekte abzuschwächen, nennt [YK06] ein breites Spektrum von Ansätzen, die von adaptiven Fenstergrößen, über die Selektion von verschiedenen Fensterformen, hin zu gewichteten Fenstern gehen. Hier soll auf den gewichteten Ansatz eingegangen werden.

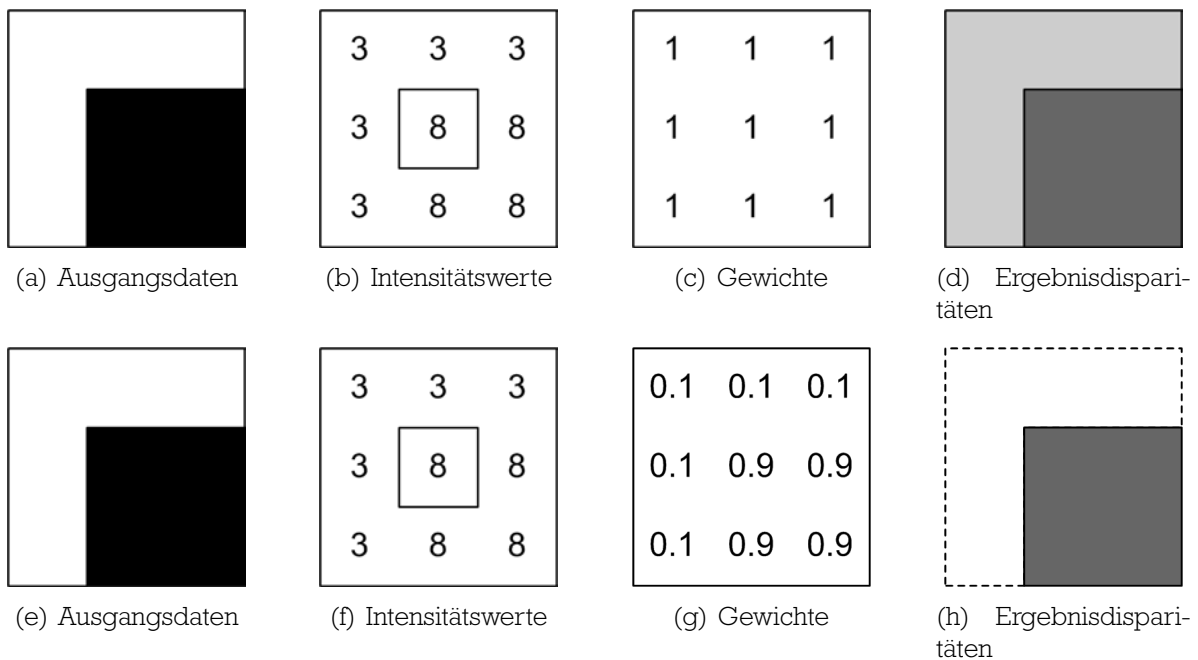


Abbildung 4.4: Beispiel einer Mittelwerttransformation ohne/mit Gewichtung.

Abbildung 4.4(a) visualisiert einen möglichen Bildausschnitt und Abbildung 4.4(b) zeigt beispielhafte Intensitätswerte dieses Ausschnittes. Wie ersichtlich besteht ein Disparitätsunterschied im Fenster, der von ungewichteten Verfahren nur schwer erfasst werden kann. Für die Berechnung der Werte des zentralen Pixels sollten eigentlich nur die schwarz markierten Felder im linken unteren Bereich verwendet werden. Durch die implizite Annahme bei ungewichteten Methoden, dass alle Pixel gleich stark in das Endergebnis eingehen sollen (Abbildung 4.4(c)), ergibt sich eine ungewollte Verwischung. Abbildung 4.4(d) zeigt diesen Sachverhalt beispielhaft. Der „L-förmige“ Bereich links oben sollte eigentlich weiß bleiben, durch die Fenstergröße und die uniforme Berück-

sichtigung aller Pixel im Fenster kommt es jedoch zu Verwischungen. Bedient man sich gewichteter Transformationen, das heißt, die Berechnungsmethoden werden um die Multiplikation mit den berechneten Gewichten erweitert, kann dies vermieden werden, wie in Abbildung 4.4(e) bis Abbildung 4.4(h) ersichtlich. Abbildung 4.4(h) visualisiert wieder beispielhaft Ergebnisdisparitäten. Durch die gewichteten Fenster kommt es nicht zu Verwischungen. Der „L-förmige“ Bereich links oben bleibt weiß. Natürlich ergibt sich durch die Multiplikation mit den Gewichten ein erhöhter Rechenaufwand. Weiters ist die Berechnung der Gewichte je Pixel und Disparitätsstufe durchzuführen, kann jedoch als Vorverarbeitungsschritt realisiert werden. Erfolgt die Berechnung der Gewichte in einem Vorverarbeitungsschritt, so führt dies natürlich zu einer Beschleunigung der Disparitätsbildberechnung, jedoch auf Kosten von Speicherplatz. Bei einer angenommenen Bildgröße von beispielsweise 450 x 375 Pixel für linkes und rechtes Bild, einer Fenstergröße von 29x29 Pixel und der Annahme, dass die Gewichte als vier Byte float gespeichert werden sollen, ergibt sich schon ein Speicherbedarf von ungefähr 1.1 Gigabyte. Bei größeren Fenstergrößen oder Stereokonfigurationen mit mehreren Kameras müssen jedenfalls Speichermanagementstrategien zum Einsatz kommen. Die Berechnung der Gewichte kann nach verschiedenen Verfahren erfolgen und wird im Abschnitt 6 detailliert beschrieben.

Ursprünglich in [YK06] oder [HBGR09] wurden die Gewichte bei der Aggregation der ermittelten Kostenwerte herangezogen. Diese Arbeit verwendet den gewichteten Ansatz, jedoch auch bei der Berechnung von radiometrischen Transformationen. Hierbei ist zu beachten, dass die Transformationen bei Verwendung von Gewichten nicht mehr als Vorverarbeitungsschritt durchgeführt werden können, sondern im Schritt der Kostenberechnung für jeden Pixel und jede Disparität separat ausgeführt werden müssen. Dies ist notwendig um zu gewährleisten, dass sowohl im rechten als auch im linken Bild die selben aggregierten Gewichte verwendet werden.

# Kapitel 5

## Kostenberechnung

Alle passiven Stereo Vision Verfahren müssen eine Möglichkeit besitzen, die Ähnlichkeit von Bildpositionen zu messen um damit mögliche Korrespondenzen finden zu können. Üblicherweise werden diese Ähnlichkeitswerte als Kosten (matching costs) bezeichnet. In den meisten Systemen und Algorithmen erfolgt die Berechnung der Kosten für alle Pixel und alle Disparitäten.

Die Kostenberechnung erfolgt über Funktionen. Diese Kostenfunktionen können nach [HS09] in die drei Typen parametrische Kosten, nicht-parametrische Kosten und Transinformation<sup>1</sup> eingeteilt werden.

Parametrische Kostenfunktionen verwenden die tatsächlichen Werte der Pixel für ihre Berechnungen und sind deshalb anfällig für nicht konstante Pixelwerte im linken und rechten Bild, verursacht durch Beleuchtungsänderung, Belichtungsänderung oder andere Störfaktoren (wie zum Beispiel Ausreißer oder Rauschen). Beispiele für parametrische Kostenfunktionen sind: Absolute Differenz, Birchfield Tomasi, normalisierte Kreuzkorrelation oder die in Kapitel 4 behandelte Mittelwerttransformation, LoG Transformation und bilaterale Subtraktion.

Um diesem Verhalten entgegenzuwirken, wurden nicht-parametrische Kostenfunktionen eingeführt. Nicht-parametrische Kosten verwenden nur die relative Reihung der Pixelwerte (Intensitäts- oder Farbwerte) zur Kostenberechnung und sind daher robust gegen alle radiometrischen Veränderungen, die diese Reihung beibehalten [HS09]. Beispiele für nicht-parametrische Kostenfunktionen sind die in Kapitel 4 behandelte Rank und SoftRank Transformation.

---

<sup>1</sup>Mutual Information

Etwas entkoppelt von den beiden letztgenannten Typen ist Transinformation zu sehen. Hier wird versucht, die komplexen radiometrischen Zusammenhänge zwischen Bildern statistisch zu modellieren und daraus Kosten abzuleiten.

In den meisten Systemen werden die berechneten Kosten in einem sogenannten Disparitätsraumbild<sup>2</sup> zur weiteren Verarbeitung abgelegt. Das Disparitätsraumbild speichert für jeden Pixel  $(x,y)$  und jede Disparität  $d$  im Suchraum die zugehörige Wahrscheinlichkeit der Entsprechung mit dem Vergleichswert des anderen Bildes. Es stellt also eine dreidimensionale Struktur mit Koordinaten der Form  $DSI(x,y,d)$  dar.

Um über die Ähnlichkeit von Bildpunkten sprechen zu können, muss man zuerst definieren, wie diese gemessen werden soll. Der übliche Fall ist die Berechnung der Kosten als Differenz der Intensitätswerte. Besonders im Fall von vorhandener Farbinformation muss diese Differenz aber gesondert betrachtet werden. Dies führt zu erweiterten Definitionen von Differenzmaßen.

## 5.1 Differenzmaße

Erfolgen alle Berechnungen auf Intensitätswerten, ohne die Berücksichtigung von Farbe, dann kann die absolute Differenz der beiden Werte als einfachstes und schnellstes Differenzmaß herangezogen werden, wie Formel 5.1 zeigt.

$$C_I(p,d) = |I_L - I_R| \quad (5.1)$$

Im Unterschied dazu stehen bei der Verwendung von Farbe mehrere Möglichkeiten der Differenzberechnung zur Verfügung.

Die L1 Differenz oder Manhattan Distanz für zwei Pixel ist definiert als die Summe der absoluten Differenzen der Farbwerte (siehe Formel 5.2, welche dies beispielhaft für den RGB Farbraum zeigt). Hauptunterschied zur Euklidischen Distanz ist die Tatsache, dass keine diagonalen Strecken im Farbraum zurückgelegt werden können. Das übliche Beispiel ist ein Taxi in Manhattan, das durch das rechtwinkelige Straßennetz fahren muss und nicht diagonal durch die Häuser fahren kann.

---

<sup>2</sup>DSI - Disparity Space Image

$$d_p(r, g, b) = |r_L - r_R| + |g_L - g_R| + |b_L - b_R| \quad (5.2)$$

Die L2 Differenz oder Euklidische Distanz ist definiert als die Wurzel aus der Summe der Quadrate der Differenzen der Farbwerte (siehe Formel 5.3, welche dies beispielhaft für den RGB Farbraum zeigt).

$$d_p(r, g, b) = \sqrt{(r_L - r_R)^2 + (g_L - g_R)^2 + (b_L - b_R)^2} \quad (5.3)$$

Weitere beispielhafte Differenzmaße wären Minimumsuche, Maximumsuche oder Medianbildung. Wobei bei der Minimumsuche das Minimum aus der Differenz der einzelnen Farbkoordinaten gesucht wird, bei der Maximumsuche das Maximum ermittelt wird und die Medianbildung den Median der Differenzen bestimmt.

Jedes der genannten Differenzmaße kann in den folgenden Kostenfunktionen angewendet werden, um Kosten und in Folge ein Disparitätsraumbild zu erstellen. Da die Funktionsdefinition je verwendetem Farbraum variiert, wird zur Vereinfachung der Darstellung in den folgenden Formeln immer nur die Intensität  $I$  angegeben. Für die Farbversion ist eine Ersetzung aller  $I$  durch die entsprechende Funktion  $d_p$  vorzunehmen.

## 5.2 Birchfield und Tomasi (BT)

Eine Erweiterung der SAD stellt die abtastungsinsensitive Summe absoluter Differenzen von Birchfield und Tomasi dar [BT98]. Diese Kostenfunktion berechnet die absolute Distanz zwischen den Extremwerten der linearen Interpolation des betrachteten Pixel und dessen direktem linken und rechten Nachbarn (siehe Formel 5.4). Die Funktionen  $\min(\cdot)$  und  $\max(\cdot)$  liefern das jeweilige Minimum oder Maximum der enthaltenen Parameter.  $I_L(\cdot)$  und  $I_R(\cdot)$  entsprechen dem Intensitätswert des Arguments.  $d$  beschreibt die Disparität.  $I^-(p)$  und  $I^+(p)$  beschreiben respektive den interpolierten Intensitätswert der Pixel  $p$  zwischen den Positionen  $p - d$ ,  $(p - d) - 1$  und  $(p - d) + 1$ . Wie Abbildung 5.1 zeigt, ergeben sich durch die diskreten Sensoren in Kameras Abtastungsunterschiede der eigentlichen kontinuierlichen Intensitätswerte / Farbwerte der Wirklichkeit. Um

diesen Effekt zu reduzieren, interpolierten Birchfield und Tomasi zwischen den Abtastpunkten im rechten Kamerabild und vergleichen mit den ermittelten Werten im linken Kamerabild.

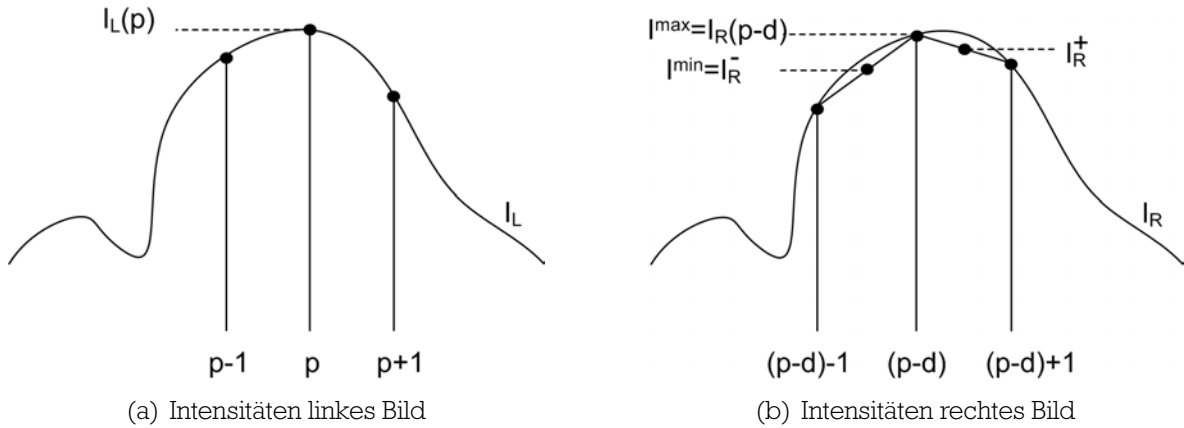


Abbildung 5.1: Interpolation der Intensitätswerte bei Birchfield Tomasi [BT98, p.2].

$$C_t(p, d) = \min(A, B) \quad (5.4)$$

$$A = \max(0, I_L(p) - I_R^{\max}(p-d), I_R^{\min}(p-d) - I_L(p))$$

$$B = \max(0, I_R(p-d) - I_L^{\max}(p), I_L^{\min}(p) - I_R(p-d))$$

$$I^{\min}(p) = \min(I^-(p), I(p), I^+(p)) \quad I^-(p) = \frac{I(p - [1 \ 0]^T) + I(p)}{2}$$

$$I^{\max}(p) = \max(I^-(p), I(p), I^+(p)) \quad I^+(p) = \frac{I(p + [1 \ 0]^T) + I(p)}{2}$$

### 5.3 Normalisierte Kreuzrelation (NCC)

Normalisierte Kreuzkorrelation<sup>3</sup> ist das statistische Standardverfahren zur Ähnlichkeitsberechnung [BBH03, p.995]. Formel 5.5 zeigt die Berechnung.  $p$  bezeichnet die Position des zentralen Pixel, während  $q$  die Position eines Pixel in der lokalen Umgebung  $N_p$  darstellt.  $I_L(\cdot)$  bezeichnet die Intensität des Arguments im linken Bild und  $I_R(\cdot)$  bezeichnet die Intensität des Arguments im rechten Bild.  $\bar{I}_L(\cdot)$  und  $\bar{I}_R(\cdot)$  bezeichnen den Mittelwert des Arguments.  $d$  bezeichnet die Disparität. Wie in Formel 5.5 ersichtlich,

<sup>3</sup>Normalized Cross Correlation



findet zusätzlich eine Normalisierung mit dem Mittelwert statt, die NCC laut [BBH03, p.995] relativ unempfindlich gegen radiometrische Verzerrungen macht. Weiters ist es nach [HS09, p.8] die optimale Methode um Gaußsches Rauschen zu behandeln. Ein Problem von NCC stellt aber die Tatsache dar, dass Ausreißer sehr stark die Berechnung beeinflussen. Dies macht sich besonders bei Kanten (Tiefenunstetigkeiten) bemerkbar und führt zu starkem Verwischen der Kanten.

$$C_{NCC}(p, d) = \frac{\sum_{q \in N_p} (I_L(q) - \bar{I}_L(p))(I_R(q-d) - \bar{I}_R(p-d))}{\sqrt{\sum_{q \in N_p} (I_L(q) - \bar{I}_L(p))^2 \sum_{q \in N_p} (I_R(q-d) - \bar{I}_R(p-d))^2}} \quad (5.5)$$

## 5.4 Hierarchische Transinformation (HMI)

Eine weitere pixelweise Kostenberechnung stellt die Berechnung der Transinformation (MI)<sup>4</sup> [VWI95] dar. MI versucht die komplexen Zusammenhänge zwischen Bildern explizit zu modellieren. Wie in [Hir08] beschrieben, ist MI nicht oder nur gering betroffen von Effekten der Aufnahme- oder Beleuchtungsveränderung, wodurch sie besonders für radiometrisch verzerrte Bilddaten geeignet scheint. Laut [Hir08] definiert sich Transinformation aus den Entropien (dem Informationsgehalt) zweier Bilder, sowie aus der Verbindungsentropie nach der Formel 5.6.

$$MI_{I_1, I_2} = H_{I_1} + H_{I_2} - H_{I_1, I_2} \quad (5.6)$$

Die Entropie wird aus der Wahrscheinlichkeitsverteilung  $P$  der entsprechenden Bilder ermittelt, nach Formel 5.7 für die Entropie und nach Formel 5.8 für die Verbindungsentropie.

$$H_I = - \int_0^1 P_I(i) \log P_I(i) di \quad (5.7)$$

$$H_{I_1, I_2} = - \int_0^1 \int_0^1 P_{I_1, I_2}(i_1, i_2) \log P_{I_1, I_2}(i_1, i_2) di_1 di_2 \quad (5.8)$$

---

<sup>4</sup>Mutual Information

Nach [Hir08] ist die Verbindungsentropie  $H_{I_1, I_2}$  für ähnliche Bilder niedrig, da ein Bild aus dem anderen vorhergesagt werden kann und somit geringe Information vorhanden ist. Dies wiederum führt zu hoher MI. Für Stereoberechnungen muss ein Bild, entsprechend des Disparitätsbildes, mit den Disparitäten  $d$  transformiert werden, damit sich entsprechende Pixel in beiden Bildern an der selben Position befinden und die MI berechnet werden kann. Somit muss  $I_1 = I_L(p)$  und  $I_2 = I_R(p - d)$  sein.

Formel 5.6 wird auf das gesamte Bild angewendet und benötigt daher die gesuchten Disparitäten als Vorbedingung. Dies erschwert die Verwendung von MI [Hir08]. [KKZ03] schlägt zur Lösung dieses Problems die Verwendung eines iterativen Ansatzes vor. Im ersten Durchlauf werden zufällige Disparitäten für die Berechnung herangezogen, welche dann zu neuen Disparitäten führen, die im nächsten Durchlauf als Startwerte verwendet werden. Diese schrittweise Verfeinerung führt schon nach einer geringen Anzahl von Durchläufen zu stabilen Ergebnissen. Durch die hohe Anzahl an Pixel ist auch mit falschen (zufälligen) Disparitäten eine gute Vorhersage der Wahrscheinlichkeitsverteilung möglich. Durch die wiederholte Verwendung der gesamten Bildinformation steigert sich jedoch die Laufzeit des Algorithmus. [Hir08] schlägt daher einen hierarchischen Ansatz vor. Im ersten Durchlauf werden Berechnungen nur mit einem Achtel oder Viertel der Auflösung durchgeführt. Das erhaltene Disparitätsbild wird vergrößert und im nächsten Durchlauf als Eingangsparameter verwendet. Dieser Vorgang wiederholt sich, bis eine stabile Lösung gefunden wird. Laut [Hir08] benötigt man hierfür im Durchschnitt drei Iterationen.

[KKZ03] haben des Weiteren mittels Taylor Expansion gezeigt, dass die Berechnung der Verbindungsentropie als Pixelsumme realisiert werden kann. Das Ergebnis zeigt Formel 5.9. Die Verbindungsentropie ist die Summe von Datentermen, die von entsprechenden Intensitäten eines Pixels  $p$  abhängig sind.

$$H_{I_1, I_2} = \sum_p h_{I_1, I_2}(I_{1p}, I_{2p}) \quad (5.9)$$

Nach [Hir08] werden die Datenterme aus den Verbundwahrscheinlichkeitsverteilungen  $P_{I_1, I_2}$  korrespondierender Intensitäten berechnet. Formel 5.10 zeigt die Berechnung, wobei  $n$  der Anzahl der korrespondierenden Pixel entspricht. Nach [KKZ03] erfolgt dann eine Faltung mit einer 2D Gaussfunktion  $g(i, k)$ , welche einer Parzen Schätzung<sup>5</sup> entspricht. [Hir08] sieht hier die Verwendung eines 7x7 Kernels als ausreichend.

---

<sup>5</sup>Parzen estimation

Der Logarithmus ist für 0 nicht definiert und sollte durch einen sehr kleinen Wert ersetzt werden [Hir08]. Der Datenterm  $h_{I_1, I_2}$  stellt im Wesentlichen ein 2D Array dar, auf das mittels der ermittelten Intensitäten (Farbwerte) zugegriffen wird. Im Fall der Verwendung von Farbe müssen drei dieser Arrays (für jeden Farbkanal im RGB Fall) erstellt werden.

$$h_{I_1, I_2}(i, k) = -\frac{1}{n} \log(P_{I_1, I_2}(i, k) \otimes g(i, k)) \otimes g(i, k) \quad (5.10)$$

Die Berechnung der Verbundwahrscheinlichkeitsverteilung korrespondierender Intensitäten erfolgt nach Formel 5.11. Wobei die Funktion  $T$  1 als Funktionswert liefert, wenn die enthaltene Bedingung erfüllt ist und 0 liefert, falls die Bedingung nicht erfüllt ist.

$$P_{I_1, I_2}(i, k) = \frac{1}{n} \sum_p T[(i, k) = (I_{1p}, I_{2p})] \quad (5.11)$$

Wie Formel 5.13 zeigt, erfolgt die Berechnung der Entropien  $H_1$  und  $H_2$  analog zur Berechnung der Verbindungsentropie, wobei jedoch darauf zu achten ist, dass die Wahrscheinlichkeitsverteilung nicht über die gesamten Bilder  $I_1$  und  $I_2$  berechnet wird, sondern nur über korrespondierende Teile der Bilder, da sonst Verdeckungen nicht berücksichtigt werden [Hir08]. Dies kann erreicht werden, indem man über die korrespondierenden Zeilen und Spalten der Verbundwahrscheinlichkeitsverteilung, mittels  $P_{I_1}(i) = \sum_k P_{I_1, I_2}(i, k)$ , summiert.

$$H_I = \sum_p h_I(I_p) \quad (5.12)$$

$$h_I(i) = -\frac{1}{n} \log(P_I(i) \otimes g(i)) \otimes g(i) \quad (5.13)$$

Aus obigen Überlegungen ergibt sich die Definition von MI nach Formel 5.15.

$$MI_{I_1, I_2} = \sum_p mi_{I_1, I_2}(I_{1p}, I_{2p}) \quad (5.14)$$

$$mi_{I_1, I_2}(i, k) = h_{I_1}(i) + h_{I_2}(k) - h_{I_1, I_2}(i, k) \quad (5.15)$$

Die MI kann wiederum als Kostenfunktion herangezogen werden (siehe Formel 5.17).

$$C_t(p, d) = -mi_{I_L, I_R}(I_{Lp}, I_{Rq}) \quad (5.16)$$

$$q = p - d \quad (5.17)$$

Zusammenfassend muss eine HMI Implementierung zuerst alle Korrespondenzen, die durch ein initiales Disparitätsbild definiert sind, ermitteln. Dieses Disparitätsbild kann entweder ein zufälliges Disparitätsbild im ersten Durchlauf darstellen oder ein verfeinertes, vergrößertes Disparitätsbild eines vorangegangenen Durchlaufes sein. Aus den Korrespondenzen kann die Wahrscheinlichkeitsverteilung  $P$  mittels Formel 5.11 ermittelt werden. Die Größe von  $P$  ist konstant. Nach Faltung mit einem Gaußkernel und Logarithmieren ergeben sich die Kosten  $C_t(p, d)$  des aktuellen Durchlaufes, welche wiederum zur Erstellung des neuen Disparitätsbildes verwendet werden können. Bis auf das Disparitätsbild werden keine Informationen in den Durchläufen weitergeleitet um die Akkumulation von möglichen Fehlern zu vermeiden.

# Kapitel 6

## Aggregation

Der nächste (optionale) Schritt auf dem Weg zu einem Disparitätsbild ist die Aggregation. Bei der Aggregation wird für jeden Pixel auf jeder Disparitätsstufe die unmittelbare Nachbarschaft untersucht und auf Grund der Nachbarschaftsbeziehungen die Stärke, mit der er in die Disparitätsberechnung eingeht, festgelegt. Es ist festzuhalten, dass üblicherweise die Aggregation im lokalen Fall notwendig ist um dichte Disparitätsbilder zu erhalten, während im globalen Fall durch die explizierte Modellierung im sogenannten Smoothness-Term ein vorheriger Aggregationsschritt nicht zwingend erforderlich ist.

### 6.1 Quadratische Aggregation

Bei der quadratischen Aggregation wird über einen fixen, zweidimensionalen Fensterbereich aggregiert. Dabei wird die Gewichtung der einzelnen Pixel nicht variiert. Dies hat den Vorteil, dass die Aggregation mittels schneller Methoden der Summenübergabe implementiert werden kann (wie zum Beispiel in [MMHM02] gezeigt). Ein großer Nachteil tritt jedoch an Kanten (Disparitätsgrenzen) auf, da es hier zu Verwischungen kommt.

Die Berechnung bei einfacher quadratischer Aggregation erfolgt nach Formel 6.1 [HS09, p.7]. Dabei wird über alle Pixel  $q$  einer quadratischen Nachbarschaft  $N_p$  eines Pixels  $p$  summiert.  $d$  ist die Disparität, und der korrespondierende Pixel eines linken Pixels  $p$  ergibt sich als  $p - d$ .

$$C_I(p, d) = \sum_{q \in N_p} |I_L(q) - I_R(q - d)| \quad (6.1)$$

Das Hauptproblem stellt die Wahl einer geeigneten Fenstergröße dar. Das Fenster sollte groß genug sein, um in Bereichen unzureichender Texturierung genügend Information über die unmittelbare Nachbarschaft zu bieten. Zugleich sollte das Fenster aber klein genug sein, um keine Pixel, die sich auf unterschiedlichen Disparitäten befinden, aufzunehmen. Eine Reihe von Ansätzen versucht das Problem zu behandeln, wie zum Beispiel [OK91], [BVZ98], [Vek02] oder [Vek03], die adaptive Fenster verwenden, oder [FRT97], die aus verschiedenen vordefinierten Fenstern passend auszuwählen versuchen. Nach [YK06, p.1] haben alle diese Methoden den entscheidenden Nachteil, dass die Form des Aggregationsfensters nicht beliebig angepasst werden kann und somit nicht alle möglichen Fälle abgedeckt werden können.

## 6.2 Gewichtete Aggregation

Ein anderer Ansatz ist die Verwendung von Gewichten in den Aggregationsfenstern. Durch diesen Ansatz kann die Form des Fensters beliebig angepasst werden. Die Mehrzahl dieser Methoden basiert auf Farbsegmentierung der Eingangsbilder (zum Beispiel [WKSX04], [YK06] oder [HBGR09]). Dabei besteht die implizite Annahme, dass Pixel, welche die selbe Farbe haben, auch die selbe Disparität haben. In den folgenden zwei Abschnitten sollen die beiden Methoden von [YK06] und [HBGR09] genauer vorgestellt werden, welche auch in dieser Arbeit implementiert sind.

### 6.2.1 Gewichte nach Yoon

Nach [YK06] sollen nur jene Nachbarpixel bei der Aggregation verwendet werden, welche sich auf der selben Disparität wie der untersuchte Pixel befinden. Um eine Pixelabgrenzung zu erhalten, bedient sich [YK06] der sogenannten Gestaltgruppierung, besonders der beiden Gruppierungstypen Farbähnlichkeit und räumliche Nähe. Aus diesen beiden Eigenschaften werden die Gewichte für die Aggregationsmasken berechnet. Die Gewichtung eines Pixels richtet sich nach der Stärke der Gruppierung. Je ähnlicher sich Pixel sind, desto stärker gehen sie in die Berechnungen ein. Selbiges

gilt für die Nachbarschaft beziehungsweise Nähe von Pixel. Basierend auf diesen Annahmen kann die Gewichtung eines Pixels  $w(p, q)$  in einem Aggregationsfenster nach Formel 6.2 berechnet werden.

$$w(p, q) = f_s(\nabla c_{pq}) * f_p(\nabla g_{pq}) \quad (6.2)$$

Wobei  $f_s(\nabla c_{pq})$  die Gruppierung nach Ähnlichkeit und  $f_p(\nabla g_{pq})$  die Gruppierung nach Nachbarschaft darstellt.  $\nabla c_{pq}$  beschreibt die Ähnlichkeit zweier Pixel als Farbdifferenz im jeweiligen Farbraum, und  $\nabla g_{pq}$  beschreibt die Nachbarschaftsbeziehung als euklidische Distanz. Der Definition von [YK06] folgend kann man die beiden Terme  $f_s(\nabla c_{pq})$  und  $f_p(\nabla g_{pq})$  wie in Formel 6.3 und Formel 6.4 anschreiben und daraus die finale Definition (siehe Formel 6.5) der Gewichte erhalten.

$$f_s(\nabla c_{pq}) = \exp\left(-\frac{\nabla c_{pq}}{\gamma_c}\right) \quad (6.3)$$

$$f_p(\nabla g_{pq}) = \exp\left(-\frac{\nabla g_{pq}}{\gamma_p}\right) \quad (6.4)$$

$$w(p, q) = \exp\left(-\left(\frac{\nabla c_{pq}}{\gamma_c} + \frac{\nabla g_{pq}}{\gamma_p}\right)\right) \quad (6.5)$$

$\gamma_c$  stellt eine benutzerdefinierte Konstante dar und wird in [YK06] mit fünf angesetzt.  $\gamma_p$  ist ebenfalls eine Konstante und definiert sich als der Radius des verwendeten Aggregationsfensters. Die Gewichtung eines Pixels ist daher umgekehrt proportional zu der Farbdifferenz zwischen dem Pixel und dem Pixel im Zentrum des untersuchten Rechtecks, sowie zu seinem Abstand vom Zentrum. Unter Verwendung der eben genannten Formeln kann somit die Dissimilarität  $E(p, p_d)$  zwischen einem Pixel  $p$  und einem Pixel  $p_d$  nach Formel 6.6 berechnet werden.

$$E(p, p_d) = \frac{\sum_{q \in N_p, q_d \in N_{p_d}} w(p, q) w(p_d, q_d) e(q, q_d)}{\sum_{q \in N_p, q_d \in N_{p_d}} w(p, q) w(p_d, q_d)} \quad (6.6)$$

$p_d$  und  $q_d$  bezeichnen hier die korrespondierenden Pixel im rechten Bild, wenn die Pixel  $p$  und  $q$  im linken Bild die Disparität  $d$  haben.  $e(q, q_d)$  bezeichnet die „rohen“ Kosten, die nach einer Methode in Abschnitt 5 berechnet wurden. Auffallend ist die

Tatsache, dass sowohl Gewichte des linken als auch des rechten Bildes Verwendung finden. [YK06] begründet dies damit, dass durch die Verwendung beider Gewichte Pixel mit ähnlichen Disparitäten wie die Pixel im Zentrum der Vergleichsrechtecke bevorzugt werden. Ein weiterer wichtiger Grund ist die ordnungsgemäße Behandlung von Verdeckungen. Die ermittelten Gewichte können in einer Datenstruktur gespeichert werden, falls die Berechnung in einem Vorverarbeitungsschritt erfolgt, oder direkt verwendet werden.

## 6.2.2 Geodätische Gewichte

[HBGR09] verfolgt einen ähnlichen Ansatz wie [YK06], verwendet zur Berechnung der Gewichte im Aggregationsfenster aber geodätische Distanz. Dabei wird geodätische Distanz als der kürzeste Pfad im Farbraum zwischen einem Pixel  $p$  im Aggregationsfenster und dem Mittelpunkt  $c$  des Fensters (siehe Formel 6.7) definiert.

$$D(p, c) = \min_{P \in P_{p,c}} d(P) \quad (6.7)$$

In Formel 6.7 beschreibt  $P_{p,c}$  alle möglichen Pfade zwischen Pixel  $p$  und  $c$ . Wobei ein Pfad  $P$  die Sequenz aller räumlich direkt benachbarten Punkte (in Achtkonnektivität)  $p_1, p_2, \dots, p_n$  festlegt. Die Kosten eines solchen Pfades werden nach Formel 6.8 berechnet.

$$D(p) = \sum_{i=2}^{i=n} d_C(p_i, p_{i-1}) \quad (6.8)$$

$d_C()$  misst die Farbdifferenz im jeweiligen Farbraum und könnte für den RGB Farbraum zum Beispiel als Euklidische Distanz implementiert sein (siehe Formel 6.9).

$$d_C(p, q) = \sqrt{(r_p - r_q)^2 + (g_p - g_q)^2 + (b_p - b_q)^2} \quad (6.9)$$

In diesem Zusammenhang definiert [HBGR09] die Eigenschaft der *Konnektivität*, welche besagt, dass die geodätische Distanz zwischen einem Pixel  $p$  und dem Mittelpunkt  $c$  niedrig ist, wenn ein Pfad existiert, auf dem die Farbvarianz gering ist. Mit der Annahme, dass Pixel mit geringer geodätischer Distanz auf der selben Disparität wie der



Mittelpunkt  $c$  liegen und damit geringe geodätische Distanz in hohe Gewichte übergehen soll, ergibt sich Formel 6.10.

$$w(p, c) = \exp\left(-\frac{D(p, c)}{\gamma}\right) \quad (6.10)$$

$\gamma$  in Formel 6.10 stellt eine benutzerdefinierte Konstante dar, welche die Stärke der Segmentierung angibt und in [HBGR09] auf zehn gesetzt ist. Formel 6.10 findet dann Anwendung bei der Berechnung der Dissimilaritäten in Formel 6.11.

$$m(c, d) = \sum_{p \in W_c} w(p, c) \cdot f(p, p - d) \quad (6.11)$$

$W_c$  in Formel 6.11 definiert das rechteckige Aggregationsfenster um das Zentrum  $c$ , und die Funktion  $f()$  definiert wieder die „rohen“ Kosten, die nach einer Methode in Abschnitt 5 berechnet wurden, ähnlich Formel 6.6.

Die eigentliche Berechnung der geodätischen Distanzen und somit Kosten erfolgt nach der Methode von [Bor86]. Jedem Pixel  $p$  im Fenster werden die Kosten  $C(p)$  zugewiesen. Zu Beginn betragen die Kosten des mittleren Pixels  $c$  Null, und die Kosten der umliegenden Pixel werden auf einen konstant großen Wert gesetzt. Im ersten Durchlauf des Algorithmus werden die Pixel des Fenster zeilenweise durchlaufen und nach Formel 6.12 aktualisiert. Wobei  $P_k$  ein Kernel ist, der aus dem Pixel  $p$  selbst, seinem linken, linken oberen, oberen und rechten Nachbarn besteht. Abbildung 6.1(a) zeigt dies grafisch.

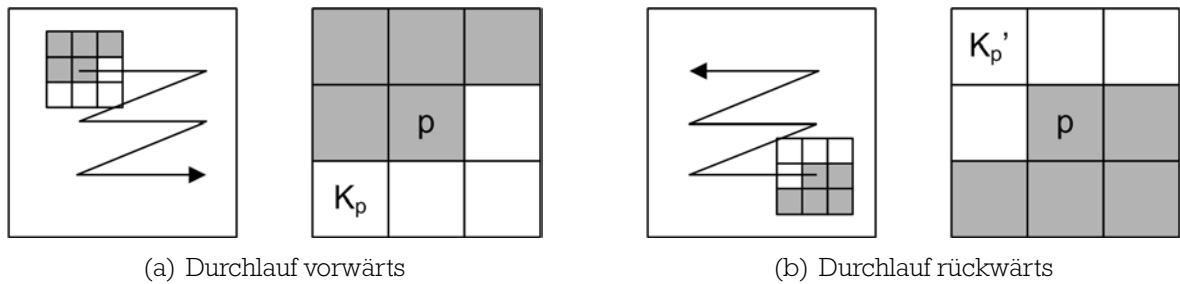


Abbildung 6.1: Approximation geodätischer Distanz aus [HBGR09, p.3].

$$C(p) = \min_{q \in P_k} C(q) + d_C(p, q) \quad (6.12)$$

Nach dem ersten Durchlauf wird das Fenster in umgekehrter Reihenfolge erneut, diesmal mit Kernel  $K'(p)$ , durchlaufen und die Kosten aktualisiert (siehe Abbildung 6.1(b)). Dieses Vorgehen wird iterativ fortgesetzt, bis eine akzeptable Schätzung für die geodätischen Distanzen erreicht ist. Nach [HBGR09] wird dies schon nach drei Iterationen erreicht.

# Kapitel 7

## Disparitätsberechnung

Nächster Schritt in der in dieser Arbeit verwendeten Stereosoftware ist die Berechnung des Disparitätsbildes aus dem DSI<sup>1</sup>. Man unterscheidet zwei Berechnungsansätze: lokale Methoden und globale Methoden. Nach [BBH03, p.995] kann man lokale Methoden in fensterbasierte, gradienten-basierte und merkmalsbasierte Ansätze einteilen. Beispielhafte globale Methoden sind Dynamic Programming, Graph Cuts oder Belief Propagation. Alle Methoden erzeugen als Ausgabe ein Disparitätsbild mit Disparitäten  $d$ , die jeden Pixel im Referenzbild in das Zielbild überführen und mit denen Tiefeninformation berechnet werden kann (siehe Abschnitt 2). Die in dieser Arbeit verwendeten Methoden zur Disparitätsberechnung werden im folgenden beschrieben. Für den lokalen Ansatz wurde WTA implementiert, während der Vertreter der globalen Ansätze eine Implementierung des Graph Cuts Algorithmus ist.

### 7.1 Lokaler Ansatz

Lokale Methoden berücksichtigen nur lokale Eigenschaften und machen eine Reihe impliziter Annahmen. Ein Beispiel ist die Annahme, dass sich die Disparität innerhalb der betrachteten Umgebung eines Pixels nicht ändert. Diese Annahme kann zutreffen, muss es aber nicht immer. Besonders an Objektkanten (Tiefenunstetigkeitsstellen) kann es dadurch zu ungewünschten Effekten, wie Verwischen der Kanten, kommen. Auch die Eindeutigkeit der Zuordnung von Pixel und Verdeckungen bereiten oft Probleme.

---

<sup>1</sup>Disparity Space Image

Gradienten-basierte Ansätze wie beispielhaft [LK81], welche Funktionen minimieren, oder merkmals-basierte Ansätze wie [BT99] oder [DA89], welche nicht Intensitäten, sondern Merkmale vergleichen, werden hier nicht behandelt.

Nach der Kostenberechnung und Aggregation wird im in dieser Arbeit verwendeten lokalen Ansatz für jeden Pixel die Disparität mit den niedrigsten Kosten gewählt. Es wird eine lokale Optimierung nach dem Prinzip „winner-takes-all“ (WTA) durchgeführt. Durch die vorangegangene Aggregation führt dies zu guten Ergebnissen, wobei jedoch das in [SS02, p.4] beschriebene Problem der Uneindeutigkeit auftritt. Die Eindeutigkeit der Zuordnung von Bildpunkten kann nur für das Referenzbild gewährleistet werden, während Bildpunkte des anderen Bildes möglicherweise mehreren Punkten zugeordnet werden. Um dieses Problem zu verbessern, müssen weitere Überprüfungen wie zum Beispiel Links-Rechts Konsistenzcheck und Verdeckungsüberprüfung durchgeführt werden.

### 7.1.1 Behandlung von Verdeckungen

Die Behandlung von Verdeckungen im lokalen Fall besteht aus einem Links-Rechts Konsistenzcheck, einem Füllschritt und einer Nachbearbeitung der Ergebnisse. Der Link-Rechts Konsistenzcheck berechnet zuerst das Disparitätsbild mit dem linken Bild als Referenzrahmen und danach ein Disparitätsbild mit dem rechten Bild als Referenzrahmen. Im nächsten Schritt wird für jedes Pixel des linken Disparitätsbildes verglichen, ob die selbe Disparität wie im korrespondierenden Pixel des rechten Disparitätsbildes vorliegt. Im negativen Fall wird das Pixel als ungültig markiert. Nach [HBGR09] werden dadurch sowohl verdeckte Bereiche als auch ungültige Disparitäten detektiert. Die gefundenen ungültigen Pixel werden im nächsten Schritt mit neuen Disparitätswerten belegt. Für jeden ungültigen Pixel wird der nächste gültige linke und rechte Nachbarpixel gesucht. Die neue Disparität des ungültigen Pixel ergibt sich dann aus dem Minimum der Disparitätswerte der gefundenen Nachbarpixel.

Diese Füllstrategie führt nach [HBGR09] zu horizontaler Streifenbildung, und daher werden die Ergebnisse in einem dritten Schritt nachbearbeitet. Die ungültigen Pixel werden mit einem Weichzeichnungsfilter gefiltert, wobei im Fall von gewichteter Aggregation ein gewichteter Median Filter zum Einsatz kommt. Dieser Filter verwendet die Gewichte, die im Aggregationsschritt ebenfalls Verwendung fanden, also entweder Gewichte nach Yoon oder geodätische Gewichte. Im Fall quadratischer Aggregation

kommt ein einfacher Median Filter (5x5 Pixel) ohne Gewichtung zum Einsatz. Beide Filter arbeiten nur auf den als ungültig markierten Pixel und beeinflussen daher das Ergebnis in den unverdeckten Bereichen nicht.

## 7.2 Globaler Ansatz

Im Unterschied zu lokalen Methoden kommen bei globalen Methoden auch nicht-lokale Nebenbedingungen zur Anwendung. Dies hat den Vorteil, dass die Methoden in Bereichen, die lokalen Methoden Probleme bereiten, bessere Ergebnisse liefern. Besonders in Bereichen mit geringer Texturierung wirkt sich der globale Ansatz vorteilhaft aus. Die zusätzliche Betrachtung von Nebenbedingungen erhöht natürlich andererseits die Rechenzeit. Dies hat zur Folge, dass globale Methoden in der Regel nicht für den Echtzeiteinsatz geeignet sind.

### 7.2.1 Energiefunktion

Wie in [SS02] ausgeführt, sind die meisten globalen Methoden als Energieminimierungssystem<sup>2</sup> formuliert. Ziel ist es eine Disparitätsfunktion  $d$  zu finden, welche die globale Energie, wie in Formel 7.1 definiert, minimiert.

$$E = E_{data} + \lambda E_{smooth} + E_{occlusion} \quad (7.1)$$

Der Datenterm  $E_{data}$  entspricht im Wesentlichen dem Ergebnis lokaler Methoden, das heißt, er beschreibt die Güte der punktweisen Korrespondenzsuche, die Kosten. Wie schon ausgeführt, ergeben sich die Kosten aus der Differenz der korrespondierenden Punkte im jeweiligen Farbraum. Die Kosten werden üblicherweise in einem Disparitätsraumbild<sup>3</sup> gespeichert (siehe Abschnitt 2). Erweitert man Formel 7.1, ergibt sich somit ein Datenterm wie in Formel 7.2 gezeigt. Hierbei bezeichnet  $P$  alle Pixel im Referenzbild. Alle Pixel  $p \in P$  sollen einer Disparität  $d_p$  zugeordnet werden, und die Funktion  $m$  berechnet die Kosten zwischen Pixel  $p$  und dem korrespondierenden Pixel  $p - d_p$ . Dazu wird meist das Disparitätsraumbild, das die Kosten enthält, herangezogen.

---

<sup>2</sup>Energy-minimization framework [Ter86]

<sup>3</sup>DSI - Disparity Space Image

$$E_{data} = \sum_{p \in P} m(p, p - p_d) \quad (7.2)$$

Wie schon erwähnt, beschäftigt sich diese Arbeit hauptsächlich mit der Frage, wie der Datenterm für Stereoberechnungen verbessert werden kann. Er stellt also eine wichtige Komponente dar.

Über den Smoothness-Term  $E_{smooth}$  können nicht-lokale Eigenschaften, wie zum Beispiel Glattheit modelliert werden. Es gibt eine Reihe von üblichen Funktionen, die Verwendung finden können. Wie jedoch in [Vek99] gezeigt, kann das Lösen solcher Minimierungsprobleme NP-vollständig sein. Die meisten globalen Methoden nehmen daher gewisse Vereinfachungen vor, um in vertretbarer Zeit zu einer Lösung zu kommen. Ein Beispiel aus [SS02] ist in Formel 7.3 zu sehen.

$$E_{smooth}(d) = \sum_{(x,y)} \rho(d(x,y) - d(x+1,y)) + \rho(d(x,y) - d(x,y+1)) \quad (7.3)$$

Hier werden nur benachbarte Bildpunkte in den Berechnungen berücksichtigt, wobei  $\rho$  eine monoton wachsende Funktion von Disparitätsunterschieden ist. Ein weiteres Beispiel wäre die Dynamic Programming Methode, die nur zeilenweise optimiert und somit keine zweidimensionale Suchaufgabe lösen muss. Der in dieser Arbeit verwendete Ansatz entspricht einem modifizierten Potts Model und behandelt benachbarte Bildpunkte in Vier-Konnektivität. Formel 7.4 zeigt den verwendeten Ansatz.

$$E_{smooth} = \sum_{(p,p') \in N} \begin{cases} 0 & \text{if } d_p = d_{p'} \\ P_1 & \text{if } |d_p - d_{p'}| = 1 \\ P_2 & \text{otherwise} \end{cases} \quad (7.4)$$

$P_1$  und  $P_2$  stellen definierte Strafwerte dar, die vergeben werden, falls die Disparität von benachbarten Pixel unterschiedlich ist. Wobei  $P_1$  eine kleine Strafe darstellt und nur in einem ein Pixel breiten Bereich um den betrachteten Pixel vergeben wird.  $P_2$  (wobei immer  $P_2 > P_1$  gilt) bestraft größere Disparitätssprünge, wie sie zum Beispiel an Kanten auftreten. Nachdem die globale Energiefunktion definiert ist, kann eine Reihe von Algorithmen herangezogen werden um das Minimierungsproblem zu lösen.

Der Occlusion-Term  $E_{occlusion}$  vergibt eine Strafe  $P_{occlusion}$  für verdeckte Pixel  $p$ . Nach [KZ01] kann er wie in Formel 7.5 angeschrieben werden.

$$E_{occlusion} = \sum_{p \in P} P_{occlusion} \cdot T(|N_p| = 0) \quad (7.5)$$

Wobei der Term  $|N_p| = 0$  genau die verdeckten Pixel repräsentiert.  $N_p$  repräsentiert jene Menge von Pixel, für die korrespondierende Pixel gefunden wurden. Die Funktion  $T$  liefert 1 als Funktionswert, wenn die enthaltene Bedingung erfüllt ist, und 0, falls die Bedingung nicht erfüllt ist.

Nach [BBH03, p.995] sind die wichtigsten Vertreter globaler Methoden: Dynamic Programming, Graph Cuts, Intrinsische Kurven, nichtlineare Diffusion und Methoden ohne Korrespondenzverwendung. Die Methoden, die intrinsische Kurven [TM96] [TM98b] verwenden, konvertieren Epipolarlinien in den sogenannten Intrinsischen Kurven Raum und können so das zweidimensionale Suchproblem, das normalerweise bei globalen Methoden auftritt, auf ein Nachbarschaftssuchproblem reduzieren, wobei jedoch Mehrdeutigkeiten entstehen, die mittels einer anderen Methode aufgelöst werden müssen. Nichtlineare Diffusion, wie in [SS96] oder [Sha93] beschrieben, aggregiert Pixel auf Grundlage eines lokalen Diffusionsprozesses. Belief Propagation, wie zum Beispiel [SSZ02], versucht, Disparitäten über Nachrichten in einer speziellen Netzwerkstruktur zu berechnen. Methoden ohne Korrespondenzverwendung, wie [FL95] oder [SK99], versuchen mittels Modellbildung und Deformation Disparitäten zu erzeugen. Auf die Methode Graph Cuts wird in Folge in Abschnitt 7.2.2 genauer eingegangen, da diese Methode in vielen Stereosystemen Anwendung findet und auch in dieser Arbeit verwendet wird.

## 7.2.2 Energieoptimierung mit Graph Cuts

Graph Cuts bezeichnen eine gängige Methode um Energiefunktionen, wie Formel 7.1, zu minimieren. Graph Cuts erzielt gute Ergebnisse und findet häufig Anwendung (zum Beispiel [ZVB99], [IG98], [KZ01], [RC98], [Vek99]). Aus diesen Gründen wird es auch in dieser Arbeit besprochen und verwendet. Hauptunterschied zur auch häufig zum Einsatz kommenden Dynamischen Programmierung besteht darin, dass Graph Cuts die volle zweidimensionale Konnektivität miteinbezieht und nicht eindimensional (zeilenbasiert) optimiert.

Das Optimierungsproblem kann als Klassifizierungsproblem betrachtet werden. Ziel ist es, eine optimale Klassifizierung (zum Beispiel nach Intensitäten, Disparitäten, Regionen), basierend auf verrauschten (unsicheren) Messungen, für alle Pixel zu finden. Dazu wird ein spezialisierter Graph als Repräsentation der Energiefunktion erstellt und so minimiert, dass die minimale Energie genau einem Minimalschnitt durch den Graphen entspricht. Dieser Minimalschnitt kann sehr effizient mit sogenannten Max-Flow Algorithmen berechnet werden (siehe Max-Flow-Min-Cut Theorem [FF62]).

Abbildung 7.1 zeigt eine beispielhafte Klassifizierung im Stereo Vision Kontext. Abbildung 7.1(a) zeigt ein Intensitätsbild mit einer Menge von Pixel  $P$ , welche jeweils Intensitäten  $I_p$  haben. Abbildung 7.1(b) zeigt die zugehörige Klassifikation  $L$  mit  $L_p \in \{0, 1, 2\}$ . Im Stereo entsprechen diese Werte den Disparitäten. Die dicken Linien in Abbildung 7.1(b) visualisieren Klassifikationsdiskontinuitäten zwischen benachbarten Pixel.

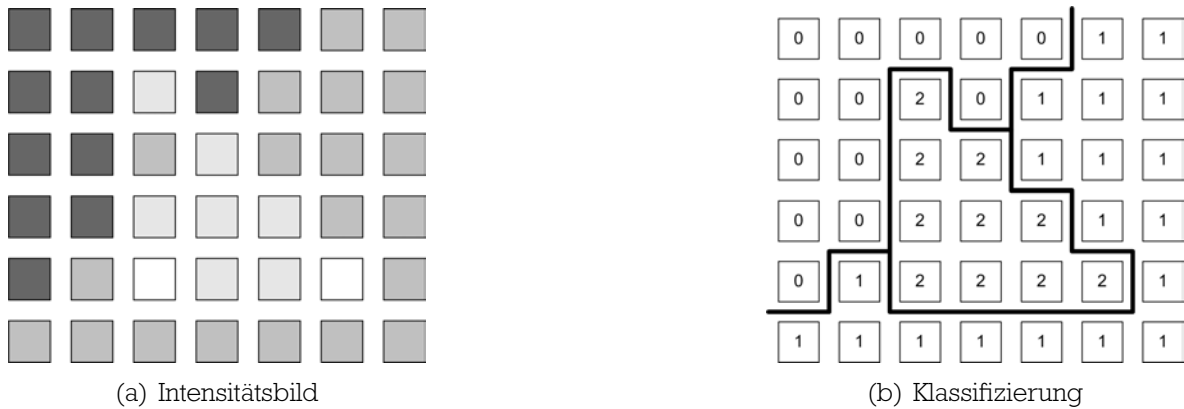


Abbildung 7.1: Klassifizierungsbeispiel aus [KB01, p.3].

Nun kann Formel 7.1 mit Hilfe des Klassifizierungsansatzes wie in Formel 7.6 beschrieben werden [GPS89].

$$E(L) = \sum_{p \in P} D_p(L_p) + \sum_{(p,q) \in N} V_{p,q}(L_p, L_q) \quad (7.6)$$

Wobei  $L = \{L_p | p \in P\}$  eine Klassifikation von Bild  $P$  ist,  $D_p(L_p)$  einen Datenterm und  $V_{p,q}(L_p, L_q)$  den Smoothness-Term darstellt, sowie  $N$  ein Nachbarschaftsbereich ist.

Unter Verwendung von Formel 7.6 ist es nun möglich das Klassifizierungsproblem als Graph zu modellieren. Sei  $G = \langle \mathbf{v}, \boldsymbol{\varepsilon} \rangle$  ein gerichteter, gewichteter Graph bestehend aus einer Menge  $\mathbf{v}$  von Knoten und einer Menge  $\boldsymbol{\varepsilon}$  von gewichteten Kanten. Üblicherweise



entsprechen die Knoten den Pixel. Weiters können Graphen spezielle Knoten enthalten, im Stereokontext üblicherweise die sogenannte Quelle  $s$  und Senke  $t$ . Diese speziellen Knoten repräsentieren im Stereokontext die möglichen Klassifizierungen, die den Pixel zugewiesen werden können. Abbildung 7.2(a) zeigt einen einfachen Graphen mit einer Quelle und einer Senke, der dazu verwendet werden kann die Minimierungsaufgabe von Formel 7.6 zu lösen.

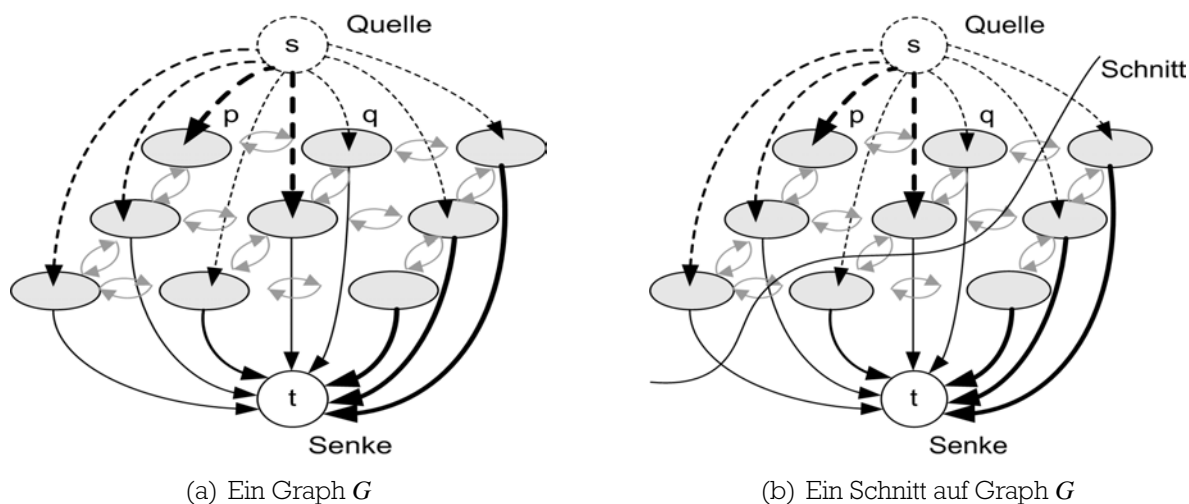


Abbildung 7.2: Einfaches Graphenbeispiel aus [GPS89].

Nach [KB01, p.5] kann die Feinstruktur dieses Graphen je nach Ansatz variieren, der grundlegende Aufbau als zwei- oder dreidimensionaler Netzgraph bleibt jedoch erhalten. Dies ergibt sich aus der Tatsache, dass üblicherweise die Pixel als Knoten repräsentiert werden. Im Gegensatz dazu repräsentieren die Kanten Gewichte oder Kosten, wobei darauf zu achten ist, dass die Kosten einer gerichteten Kante  $(p, q)$  von den Kosten der entgegengesetzten Kante  $(q, p)$  abweichen können. Nach [KB01, p.5] kann man die Kanten in zwei Typen einteilen, in sogenannte  $n$ -Kanten und  $t$ -Kanten.  $n$ -Kanten verbinden benachbarte Pixel und modellieren die Nachbarschaftsbeziehungen im Bild. Zugeordnete Kosten entsprechen den Strafkosten, die vergeben werden, falls Diskontinuitäten auftreten. Somit entsprechen  $n$ -Kanten genau dem Term  $V_{p,q}$  (dem Smoothness-Term) aus Formel 7.6.  $t$ -Kanten verbinden Knoten (Pixel) mit Quelle und Senke. Die Kosten, die einen Pixel und die Quelle oder Senke verbinden, entsprechen den Strafkosten, die anfallen, falls der Pixel einer entsprechenden Klassifikationsklasse zugeordnet wird. Diese Kosten entsprechen dem Term  $D_p$  (dem Datenterm) aus Formel 7.6.

Hat man einen Graphen konstruiert, kann man den minimalen Schnitt des Graphen berechnen und somit die Energie minimieren. Ein sogenannter *s/t Schnitt*  $C$  auf einem Graphen mit einer Quelle  $s$  und einer Senke  $t$  teilt den Graphen in zwei disjunkte Subgraphen  $S$  und  $T$ , wobei die Quelle  $s$  in  $S$  und die Senke  $t$  in  $T$  enthalten ist. Abbildung 7.2(b) zeigt einen beispielhaften Schnitt. In der kombinatorischen Optimierung ergeben sich die Kosten eines Schnittes  $C = \{S, T\}$  als die Kosten der Kanten  $(p, q)$ , wenn  $p \in S$  und  $q \in T$ . Dabei gilt zu beachten, dass die Kosten „gerichtete“ Kosten darstellen, da auch die Kanten gerichtet (von  $S$  nach  $T$ ) sind. Somit ist der Minimalschnitt eines Graphen der Schnitt mit den minimalen Kosten aus allen betrachteten Schnitten.

Um dieses Problem zu lösen besagt die kombinatorische Optimierung, dass man den *s/t Minimalschnitt* finden kann, indem man den maximalen Fluss im Graphen von der Quelle zur Senke findet. Ein gerne verwendetes Beispiel beschreibt den maximalen Fluss als jene Menge Wasser, welche von der Quelle in die Senke fließen kann, falls man Kanten als gerichtete Rohre mit einer Kapazität, welche den Kantengewichten entspricht, betrachtet. Das Theorem von Ford und Fulkerson [FF62] besagt, dass der maximale Fluss von einer Quelle  $s$  in eine Senke  $t$  eine Menge von Kanten durchläuft und sättigt, welche den Graphen genau in zwei disjunkte Mengen  $\{S, T\}$  von Knoten trennt, die einem Minimalschnitt entsprechen. Daraus ergibt sich, dass ein Minimalschnitt (die Kosten des Minimalschnittes) genau dem maximalen Fluss (Wert des maximalen Flusses) durch den Graphen entspricht und somit Methoden zur Berechnung von maximalen Flüssen angewendet werden können.

Nach [KB01, p.7] können Methoden zur maximalen Flussberechnung in einem Graphen in Goldberg-Tarjan „push-relabel“ Methoden [GT88] und Ford-Fulkerson „augmenting paths“ Methoden [FF62] kategorisiert werden.

Ein Beispiel für eine Methode vom Typ Ford-Fulkerson findet sich in [Din70]. Es wird ein Fluss durch noch nicht saturierte Pfade von der Quelle in die Senke propagiert, bis der maximale Fluss im Graphen  $G$  erreicht ist. Üblicherweise speichert die Methode dazu Informationen über die Verteilung des aktuellen Flusses  $f$  von  $s$  nach  $t$  durch die Kanten von  $G$  in einem Residualgraphen  $G_f$ . Die Struktur von  $G_f$  entspricht jener von  $G$ , aber die Kapazität einer Kante in  $G_f$  stellt die Restkapazität der entsprechenden Kante in  $G$  unter der schon verbrauchten Flussmenge dar. Zu Beginn besteht kein Fluss im Graph ( $f = 0$ ), und die Kapazitäten in den Graphen  $G$  und  $G_f$  sind ident. In jeder neuen Iteration sucht die Methode nach dem kürzesten Pfad  $s \rightarrow t$  entlang noch nicht vollständig saturierter Kanten in  $G_f$ . Falls ein Pfad gefunden wird, wird zuerst der maxi-

mal mögliche Fluss  $df$ , der zumindest eine der im Pfad enthaltenen Kanten vollständig saturiert, berechnet. Danach werden die Restkapazitäten im Pfad um  $df$  verringert, während die Kapazitäten der gegenläufigen Kanten um  $df$  erhöht werden. Somit erhöht jeder Durchlauf den totalen Fluss von der Quelle in die Senke um  $f = f + df$ . Der maximale Fluss ist erreicht, sobald einer der Pfade  $s \rightarrow t$  eine vollkommen saturierte Kante in  $G_f$  kreuzt.

Goldberg-Tarjan Methoden speichern nicht einen gültigen Fluss während ihrer Laufzeit, sondern speichern Knoten, die eine geringe Schätzung für die Distanz zur Senke entlang nicht vollständig saturierter Pfade angeben. Die Methoden versuchen dann Fluss an diese Knoten zu verteilen. Üblicherweise wird diese Verteilungsaktion von aktiven Knoten mit großer geschätzter Distanz zur Senke aus gestartet. Eine weitere Möglichkeit wäre eine FIFO<sup>4</sup> Auswahl. Die Distanzen wachsen so kontinuierlich an, während Kanten durch eine Verteilungsaktion saturiert werden. Fluss, der nicht weitergegeben werden kann, staut sich so gegebenenfalls bis zur Quelle zurück.

Somit kann mittels der Berechnung des minimalen Schnittes eines Graphen die optimale  $\alpha$ -Expansion gefunden werden, welche das angesprochene Minimierungsproblem löst.  $\alpha$ -Expansion in diesem Kontext beschreibt das Finden jener Klassifizierung  $L'$  mit den niedrigsten Kosten, die sich von der Anfangsklassifizierung  $L$  durch ein Subset von Label unterscheidet, welche in das Label  $\alpha$  übergegangen sind. Wobei die neue Klassifizierung  $L'$  dann zur aktuellen Klassifizierung  $L$  wird, falls die Menge der dem Label  $\alpha$  zugewiesen Pixel sich erhöht hat. Löst man dieses Subproblem iterativ, führt dies zu einer guten Approximation des Optimums der globalen Kosten.

Der in dieser Arbeit verwendete Ansatz basiert auf dem Graph-Cut  $\alpha$ -Expansion Algorithmus von [ZVB99]. Zur Minimierung der definierten Energiefunktion aus Formel 7.1 wird das Markov Random Field Framework von [SZS<sup>+</sup>08] verwendet. Ebenfalls erfolgt eine Behandlung von Verdeckungen.

---

<sup>4</sup>First In First Out

# Kapitel 8

## Evaluierung

Für die Evaluierung wurde ein Stereoframework erstellt. Abbildung A.1 im Anhang zeigt den schematischen Aufbau der Anwendung. Besonderes Augenmerk wurde auf leichte Erweiterbarkeit und Abdeckung einer Vielzahl von Kombinationsmöglichkeiten, besonders in der Variation des Datenterms, gelegt. Um die Tests automatisiert durchzuführen wurde eine zweite Anwendung entwickelt, die die gewünschten Tests durchführt und das Stereoframework entsprechend aufruft. Tabelle 8 zeigt eine Auflistung der in der Evaluierung untersuchten Kombinationen, wobei in der Evaluierung jede Parameterkombination mit jeder anderen Parameterkombination getestet wurde.

Alle verwendeten Testdatensätze stammen aus der Middlebury Ground Truth Datenbank ([SS02], [SS03], [SP07], [HS07]). Abbildung 8.1 zeigt alle in der Evaluierung verwendeten Testdaten. Es ist jeweils das linke Bild und die zugehörige Ground Truth<sup>1</sup> abgebildet. Die entnommenen Daten wurden in Gruppen nach Jahr der Erstellung eingeteilt, um einerseits einen besseren Überblick zu geben, andererseits gezielt die Daten voneinander abzugrenzen. Besonders die neueren Datensätze stellen höhere Anforderungen an die Algorithmen als die älteren Datensätze, und dies führt auch zu höheren Fehlerraten bei den Ergebnissen zu diesen Datensätze. In Summe wurden zehn Datensätze für diese Evaluierung ausgewählt. Die verwendete Disparitätssuchtiefe beträgt 16 Pixel für Tsukuba, 32 Pixel für Venus, 64 Pixel für Teddy und Cones, sowie 80 Pixel für den Rest der Daten. Aus den in Tabelle 8 genannten Kombinationsmöglichkeiten und den zehn Datensätzen ergeben sich, unter Berücksichtigung von Redundanzen, knappe 14000 Kombinationen, die in diese Evaluierung einfließen.

---

<sup>1</sup>Stellt das optimale Ergebnis dar. Dient zur Bewertung der Ergebnisse aller untersuchten Methoden.

Tabelle 8.1: Verwendete Kombinationen in der Evaluierung

Farbraum	Intensität (Graustufen), RGB, I1I2I3, AC1C2, LUV
Radiometrische Transformation	Keine, Rank, SoftRank, LoG, Mittelwert, Bilaterale Subtraktion
Kostenberechnungsmethode	SAD, BT, NCC, HMI
Differenzmaß	L1
Aggregation	Keine, Quadratisch, Gewichtet
Optimierungsverfahren	Lokal (WTA), Graph Cuts
Gewichtung radiometrischer Transformation / NCC	Ja, Nein
Gewichtstyp	Yoon, Geodätisch

Die Evaluierung wurde in zwei Stufen durchgeführt. Im ersten Schritt, dem Training, wurden vier Datensätze (namentlich Cones, Art, Moebius, Reindeer) ausgewählt um das notwendige Parametertuning durchzuführen. Ziel des Parametertunings ist es, gute Werte für die Parameter  $P_1$ ,  $P_2$  und  $P_{occlusion}$  (welche jeweils das Verhalten des Graph Cut Algorithmus beeinflussen) zu finden. Wobei  $P_1$  und  $P_2$  dazu dienen, den Datenterm gegenüber dem Smoothness-Term (siehe Abschnitt 7.2.2) auszubalancieren, und  $P_{occlusion}$  den Einfluss von Verdeckungen im Algorithmus regelt. Die Werte für  $P_1$ ,  $P_2$  und  $P_{occlusion}$  werden je nach Datenterm ermittelt. Dazu werden Disparity Maps der Testdatensätze für verschiedene Parameterwerte berechnet und jene Werte gewählt, welche den kleinsten durchschnittlichen Fehler im Vergleich mit den Groundtruth Daten aufweisen. Dieser Schritt stellt einen großen Aufwand dar, ist aber notwendig um die Evaluierung fair zu gestalten. Nach dem Training werden die gefundenen Parameterwerte unverändert gelassen und die eigentliche Evaluierung auf zehn Testdatensätzen durchgeführt.

Das Festhalten der genannten Parameter dient dem Zweck, die Evaluierung fair zu halten indem der Einfluss von in der Untersuchung nicht enthaltenen Größen so gering wie möglich gehalten werden soll. Besonders der Auswahl und dem Festhalten von

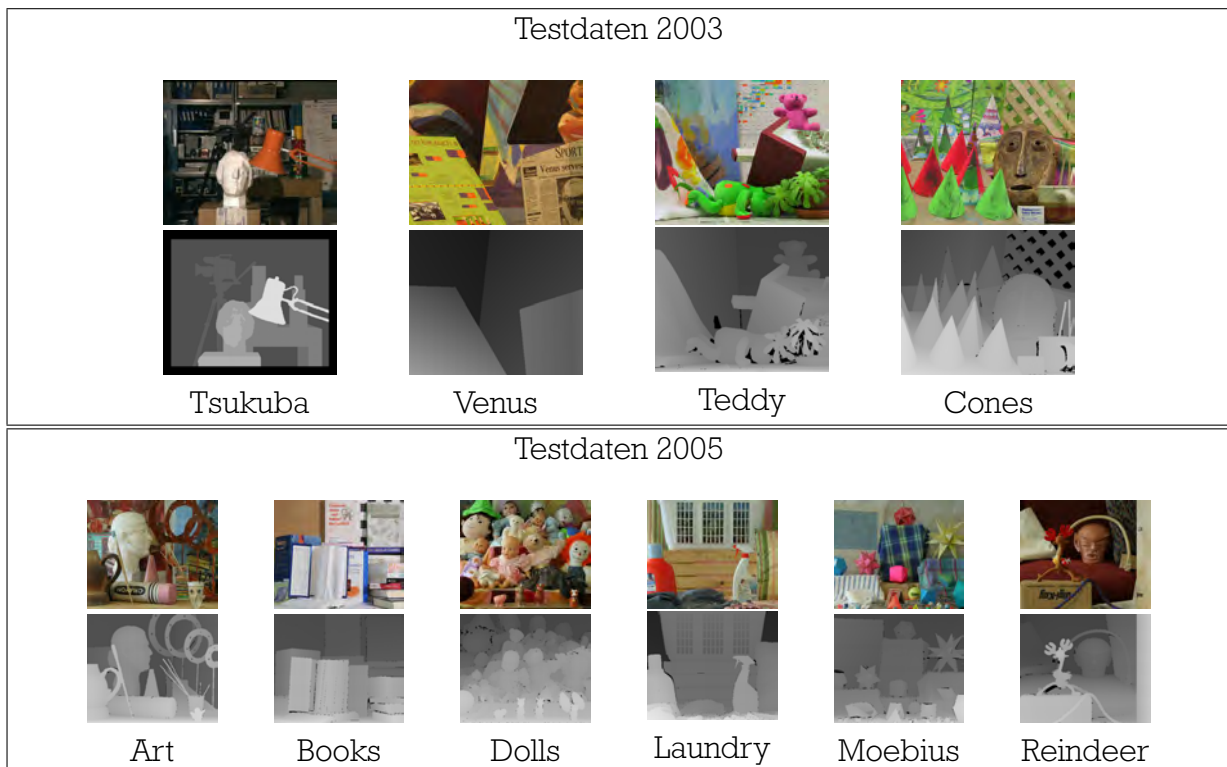


Abbildung 8.1: Linkes Bild und Ground Truth Bild für alle verwendeten Testdaten.

$P_1$ ,  $P_2$  und  $P_{occlusion}$  für die Testläufe kommt eine besondere Bedeutung zu, da dies gewährleistet, dass die Ergebnisse vergleichbar bleiben. In der Evaluierung wird je Testbildpaar das Ergebnis der Berechnungen - das Disparitätsbild - mit den vorhandenen Ground Truth Daten verglichen und die Fehlerraten in den unverdeckten Bereichen, den verdeckten Bereichen und die Gesamtfehlerrate berechnet.

Zum Einen wird der mittlere Prozentsatz jener Pixel, die einen Fehlergrenzwert von einem Pixel (bei der Klassifizierung) in den nicht verdeckten Regionen der Testdaten übersteigen, berechnet. Zum Anderen wird für jedes Testbild die jeweilige Position (Rang) innerhalb der Testdaten in Bezug auf die aktuell betrachtete Größe ermittelt. Zum Beispiel würde im Falle der Farburntersuchung der Farbraum mit der geringsten Fehlerrate die Position eins einnehmen und der Farbraum mit der größten Fehlerrate die höchste Position zugewiesen bekommen.

Danach wird für alle Ränge aller Testdaten der Mittelwert ermittelt. Dieses Ranking orientiert sich an dem Ranking der Middlebury Website für Stereo Algorithmen<sup>2</sup> und soll ein Performance Ranking der untersuchten Kombinationen ermöglichen. In den

<sup>2</sup><http://vision.middlebury.edu/stereo/eval/> Letzte Abfrage: 01.07.2009

Tabellen im Abschnitt 8 ist immer der mittlere Fehler in den unverdeckten Bereichen  $Fehler_{\emptyset}$  und der mittlere Rang  $Rang_{\emptyset}$  dargestellt. Die Sortierung erfolgt nach  $Rang_{\emptyset}$  des Testsets *Alle Daten*. Auch alle Graphen in Abschnitt 8 stellen den mittleren Fehler in den unverdeckten Bereichen dar.

Der implementierte Graph Cuts Algorithmus sowie die lokale Methode (WTA) unterstützen die Behandlung von Verdeckungen.

# Kapitel 9

## Ergebnisse

In diesem Abschnitt finden sich eine Auflistung und Analyse der erzielten Ergebnisse der in Abschnitt 8 beschriebenen Evaluierung. Die Gliederung folgt dem schematischen Aufbau der verwendeten Stereosoftware und den untersuchten Kombinationen. Ergebnisse werden immer im Kontext von lokaler und globaler Optimierung betrachtet. Bei der Analyse liegt der Fokus auf der Untersuchung der Performance in Bezug auf Farbraum, radiometrischer Transformation und Kostenberechnungsmethode. Weiters spielt besonders im lokalen Fall, durch Wegfall des *Smoothness Terms*, die Wahl einer Fenstergröße eine wichtige Rolle und wird daher ebenfalls betrachtet. Abweichend vom lokalen Fall spielt im globalen Fall die Wahl der Fenstergröße eine untergeordnete Rolle, da im Zuge des Trainings die Fenstergröße 29x29 Pixel, also die größte im Test befindliche Fenstergröße, immer die besten Ergebnisse lieferte und somit als konstanter Parameter gewählt wurde. Weiters entfällt die Betrachtung des Aggregationsschrittes im globalen Fall, da hier der *Smoothness-Term* für global-aggregierte Ergebnisse sorgt und daher, für diese Evaluierung, auf eine vorgelagerte (nochmalige) Aggregation der Ergebnisse verzichtet wurde. Die Auswirkungen der Auswahl einer Gewichtsberechnungsfunktion werden beleuchtet, und abschließend wird das jeweils beste Ergebnisbild (Disparitätskarte) je Testbild, sowie die zugehörigen Parameter, dargestellt.

### 9.1 Farbraum

Folgender Abschnitt zeigt alle Ergebnisse in Bezug auf den Farbraum. Für alle folgenden Abbildungen und Tabellen gilt, dass bis auf den Parameter Farbraum alle Para-



meter aus Tabelle 8 variiert werden. Das heißt zum Beispiel, dass in den dargestellten mittleren Werten für den RGB Farbraum alle Ergebnisse aller durchgeführten Parametervariationen enthalten sind, in denen der RGB Farbraum Anwendung fand.

### 9.1.1 Lokaler Ansatz

Abbildung 9.1 zeigt die Performance der untersuchten Farbräume auf den Testdaten. Betrachtet man die Testdaten 2003 (Abbildung 9.1(a)), bestätigt sich die Aussage von [BCPG08, p.65], dass eine Verwendung von Farbe wenig zur Ergebnisverbesserung beiträgt. In dieser Untersuchung ergibt sich eine durchgehende Verschlechterung im Vergleich zum intensitätsbasierten Ansatz. Die Testdaten 2005 (Abbildung 9.1(b)) zeigen kein so klares Bild. In einigen Fällen ist eine Verbesserung, in einigen Fällen eine Verschlechterung bei der Verwendung von Farbinformation festzustellen. Hierbei muss jedoch beachtet werden, dass die gezeigten Ergebnisse alle durchgeführten Kombinationen für den jeweiligen Farbraum enthalten. Hierzu zählt auch zum Beispiel NCC mit einer Fenstergröße von 29x29 Pixel ohne Verwendung von Gewichten. Dieser Fall führt zu sehr groben Verwischungen und einem entsprechenden Anstieg der Fehlerraten. Im Folgenden werden auch die Ergebnisse unter Aussparung von Methoden ohne Gewichtsverwendung präsentiert.

Tabelle 9.1 zeigt eine quantitative Aufschlüsselung der untersuchten Farbräume und Testdaten. Wie schon in Abschnitt 8 erwähnt, erfolgt die Auflistung in Anlehnung an das Ranking der Middlebury Website für Stereo Algorithmen. Es ist ersichtlich, dass auch unter Einbeziehung aller Testdatensätze eine Verwendung von Farbinformation sinnvoll erscheint.

Abbildung 9.2 zeigt die selben Ergebnisse wie Abbildung 9.1, jedoch unter Aussparung von Ergebnissen ohne Gewichtsverwendung. Durch die Aussparung werden Ergebnisse mit starken Verwischungen, bedingt durch große Fenstergrößen, weitestgehend gefiltert, und es zeigt sich ein klareres Bild in Bezug auf die Farbräume. Alle untersuchten Farbräume liefern bessere Ergebnisse für die Testdaten 2005. Die besten Ergebnisse liefern der Luminanz-Chrominanz Farbraum, AC1C2 und der statistisch-unabhängige Farbraum I1I2I3. Auch dieses Ergebnis deckt sich mit den Ergebnissen von [BCPG08].

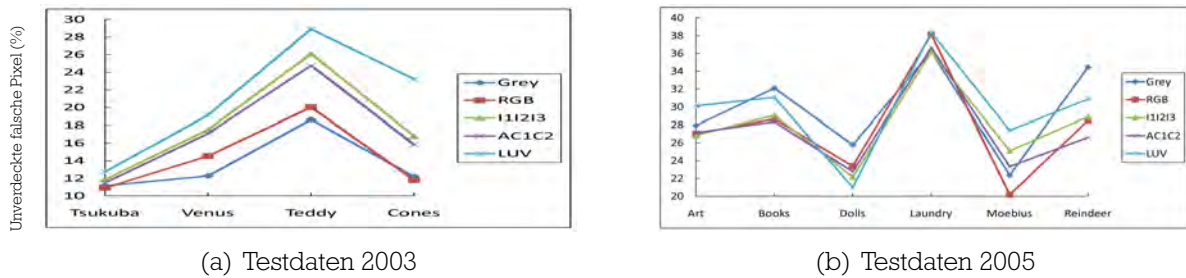


Abbildung 9.1: Vergleich der Gesamtperformance der Farbräume auf den Testdaten bei lokaler Optimierung.

Tabelle 9.1: Farbräume im Vergleich bei lokaler Optimierung.

Farbraum	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
RGB	2.1 <sub>1</sub>	22.3 <sub>1</sub>	1.5 <sub>1</sub>	14.3 <sub>2</sub>	2.5 <sub>3</sub>	27.6 <sub>2</sub>
AC1C2	2.6 <sub>2</sub>	23.4 <sub>3</sub>	3.0 <sub>3</sub>	17.2 <sub>3</sub>	2.3 <sub>1</sub>	27.5 <sub>1</sub>
Grey	2.9 <sub>3</sub>	23.3 <sub>2</sub>	1.5 <sub>1</sub>	13.5 <sub>1</sub>	3.8 <sub>4</sub>	29.8 <sub>4</sub>
I1I2I3	3.0 <sub>4</sub>	24.0 <sub>4</sub>	4.0 <sub>4</sub>	18.0 <sub>4</sub>	2.3 <sub>1</sub>	28.0 <sub>3</sub>
LUV	4.4 <sub>5</sub>	26.2 <sub>5</sub>	5.0 <sub>5</sub>	20.9 <sub>5</sub>	4.0 <sub>5</sub>	29.8 <sub>4</sub>

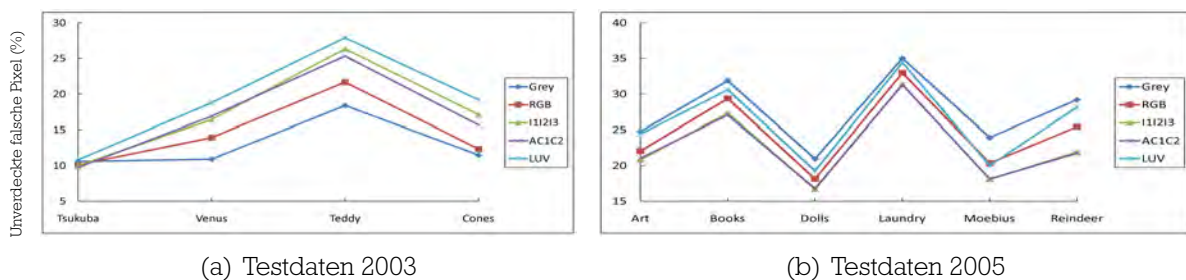


Abbildung 9.2: Gesamtperformance der Farbräume auf den Testdaten, bei Betrachtung nur jener Ergebnisse mit Gewichtsverwendung.

Betrachtet man die mittleren Fehlerraten in den unverdeckten Bereichen unter der Aussparung von Ergebnissen ohne Gewichtsverwendung für die Testdaten 2005, ergibt sich für Intensitäten ein Wert von 27.62 Prozent, während der beste Farbraum (I1I2I3) 22.69 Prozent erzielt. Dies entspricht einer Verbesserung von 17.9 Prozent. Unter Betrachtung aller Daten ergibt sich eine Verbesserung um 6.2 Prozent zwischen Intensität und dem besten Farbraum (AC1C2). Dies lässt sich leicht mit der Verschlechterung der Ergebnisse in den Testdaten 2003 bei Verwendung von Farbinformation erklären.

Tabelle 9.2 zeigt die Ergebnisse unter Aussparung von Methoden ohne Gewichtsverwendung für die Testdaten. Der Luminanz-Chrominanz Farbraum AC1C2 und der I1I2I3 Farbraum erzielen die besten Ergebnisse, gefolgt von RGB und LUV. Alle Farbräume erzielen bessere Ergebnisse als der intensitätsbasierte Ansatz für die 2005 Testdaten.

### 9.1.2 Globaler Ansatz

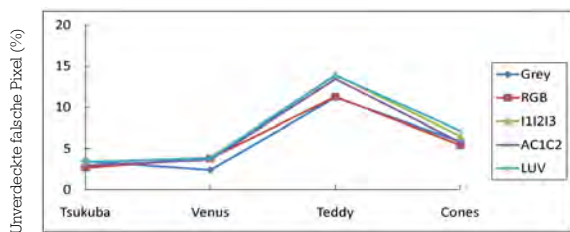
Abbildung 9.3 zeigt die Performance der untersuchten Farbräume auf den Testdaten. Auch im globalen Fall zeigt sich für die Testdaten 2003, dass eine Verwendung von Farbe wenig zur Ergebnisverbesserung beiträgt. Einzig RGB und AC1C2 können in manchen Fällen geringfügig bessere Ergebnisse als der intensitätsbasierte Ansatz erzielen. Die Testdaten 2005 zeigen ein anderes Bild. Hier zeigt sich ein klares Argument für den Einsatz von Farbe. In fast allen Fällen verbessern sich die Ergebnisse. Analog zum lokalen Fall zeigt sich, dass der Luminanz-Chrominanz Farbraum AC1C2 und der statistisch-unabhängige Farbraum I1I2I3 besonders gute Ergebnisse liefern. Interessant ist auch die Tatsache, dass im Vergleich zum lokalen Fall der RGB Farbraum ähnlich gute Ergebnisse liefert wie AC1C2 und I1I2I3. Anzumerken ist, dass auf Grund der schlechten mittleren Ergebnisse der radiometrischen Transformationen LoG und Mittelwert (siehe Abschnitt 9.2), die Ergebnisse dieser beiden Transformationen für die Farbraumvergleiche im globalen Ansatz unberücksichtigt bleiben.

Tabelle 9.3 zeigt wieder eine quantitative Aufschlüsselung der untersuchten Farbräume und Testdaten. Wie schon in Abschnitt 8 erwähnt, erfolgt die Auflistung in Anlehnung an das Ranking der Middlebury Website für Stereo Algorithmen.

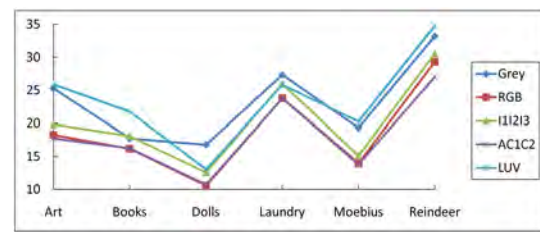
Abbildung 9.4 zeigt die selben Ergebnisse wie Abbildung 9.3, jedoch unter Aussparung von Ergebnissen ohne Gewichtsverwendung. Wie im lokalen Ansatz, zeigt sich

Tabelle 9.2: Farbräume im Vergleich bei lokaler Optimierung unter Aussparung von Ergebnissen ohne Gewichtsverwendung.

Farbraum	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
AC1C2	2.0 <sub>1</sub>	20.3 <sub>1</sub>	2.8 <sub>3</sub>	16.7 <sub>3</sub>	1.5 <sub>1</sub>	22.7 <sub>1</sub>
111213	2.2 <sub>2</sub>	20.5 <sub>2</sub>	3.3 <sub>4</sub>	17.3 <sub>4</sub>	1.5 <sub>1</sub>	22.7 <sub>1</sub>
RGB	2.8 <sub>3</sub>	20.5 <sub>2</sub>	2.3 <sub>2</sub>	14.4 <sub>2</sub>	3.2 <sub>3</sub>	24.7 <sub>3</sub>
Grey	3.7 <sub>4</sub>	21.6 <sub>4</sub>	1.8 <sub>1</sub>	12.8 <sub>1</sub>	5.0 <sub>5</sub>	27.6 <sub>5</sub>
LUV	4.3 <sub>5</sub>	23.2 <sub>5</sub>	5.0 <sub>5</sub>	18.9 <sub>5</sub>	3.8 <sub>4</sub>	26.1 <sub>4</sub>



(a) Testdaten 2003



(b) Testdaten 2005

Abbildung 9.3: Vergleich der Gesamtperformance der Farbräume in den Testsets bei globaler Optimierung.

Tabelle 9.3: Farbräume im Vergleich bei globaler Optimierung.

Farbraum	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
RGB	1.6 <sub>1</sub>	13.5 <sub>1</sub>	1.8 <sub>1</sub>	5.8 <sub>2</sub>	1.5 <sub>1</sub>	18.7 <sub>2</sub>
AC1C2	1.8 <sub>2</sub>	13.5 <sub>1</sub>	2.3 <sub>2</sub>	6.4 <sub>3</sub>	1.5 <sub>1</sub>	18.2 <sub>1</sub>
Grey	3.5 <sub>3</sub>	16.3 <sub>4</sub>	2.5 <sub>3</sub>	5.7 <sub>1</sub>	4.2 <sub>4</sub>	23.3 <sub>4</sub>
111213	3.8 <sub>4</sub>	15.0 <sub>3</sub>	4.5 <sub>5</sub>	6.9 <sub>4</sub>	3.3 <sub>3</sub>	20.3 <sub>3</sub>
LUV	4.3 <sub>5</sub>	17.0 <sub>5</sub>	4.0 <sub>4</sub>	7.1 <sub>5</sub>	4.5 <sub>5</sub>	23.6 <sub>5</sub>

auch hier auf den Testdaten 2005, dass eine Verwendung von Farbe sinnvoll ist. In einer Mehrzahl der Fälle werden bessere Ergebnisse erzielt. Die besten Ergebnisse liefern der Luminanz-Chrominanz Farbraum AC1C2 und der statistisch-unabhängige Farbraum I1I2I3.

Tabelle 9.4 zeigt die Ergebnisse unter Aussparung von Methoden ohne Gewichtsverwendung für die Testdaten. Der Luminanz-Chrominanz Farbraum AC1C2 und der RGB Farbraum erzielen die besten Ergebnisse, gefolgt vom I1I2I3 Farbraum. Einzig LUV erzielt schlechtere Ergebnisse als der intensitätsbasierte Ansatz.

## 9.2 Radiometrische Transformation

Folgender Abschnitt zeigt alle Ergebnisse in Bezug auf radiometrische Transformationen. Für alle folgenden Abbildungen und Tabellen gilt, dass bis auf den Parameter Radiometrische Transformation alle Parameter aus Tabelle 8 variiert werden. Das heißt zum Beispiel, dass in den mittleren Ergebniswerten für die Rank Transformation alle Ergebnisse aller durchgeführten Parametervariationen enthalten sind, in denen die Rank Transformation Anwendung fand.

### 9.2.1 Lokaler Ansatz

Abbildung 9.5 zeigt die Performance der untersuchten radiometrischen Transformationen auf den Testdaten. Es zeigt sich, dass sowohl in den Testdaten 2003 als auch in den Testdaten 2005 Rank und SoftRank die besten Ergebnisse liefern.

Tabelle 9.5 zeigt die quantitative Auflistung der untersuchten radiometrischen Transformationen auf den Testdaten. Wie auch in [HS09, p.17] zeigt sich, dass auch auf Bildern ohne radiometrische Störungen die Anwendung von radiometrischen Transformationen bessere Ergebnisse liefert. Dies erklärt sich aus der Tatsache, dass die Testbilder zwar unter kontrollierten Bedingungen aufgenommen wurden, jedoch einige radiometrische Probleme immer auftreten.

Betrachtet man die radiometrischen Transformationen ohne den Einsatz von gewichteten Transformationen, wie in Abbildung 9.6 dargestellt, erhält man ein ähnliches Ergebnis. Auffällig ist das bessere Abschneiden von bilateraler Subtraktion, welche nun

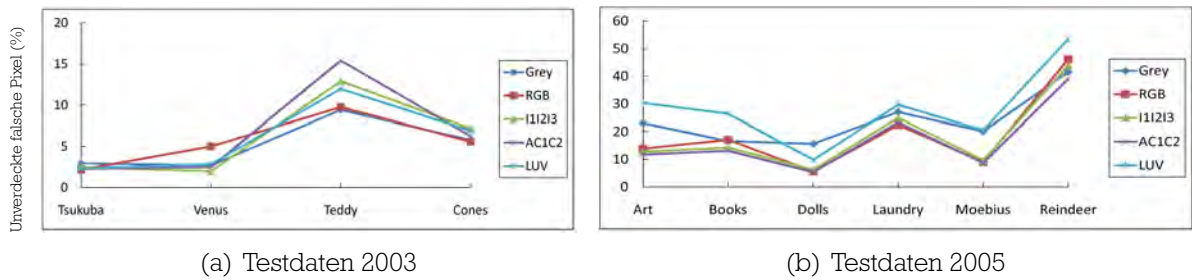


Abbildung 9.4: Gesamtperformance der Farbräume auf den Testdaten, bei Betrachtung nur jener Ergebnisse mit Gewichtsverwendung.

Tabelle 9.4: Farbräume im Vergleich bei globaler Optimierung unter Aussparung von Ergebnissen ohne Gewichtsverwendung.

Farbraum	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
AC1C2	1.9 <sub>1</sub>	12.8 <sub>1</sub>	3.0 <sub>3</sub>	6.6 <sub>5</sub>	1.2 <sub>1</sub>	16.9 <sub>1</sub>
RGB	2.5 <sub>2</sub>	13.7 <sub>3</sub>	2.3 <sub>1</sub>	5.6 <sub>2</sub>	2.7 <sub>2</sub>	19.0 <sub>3</sub>
I1I2I3	3.0 <sub>3</sub>	13.6 <sub>2</sub>	3.5 <sub>4</sub>	6.1 <sub>4</sub>	2.7 <sub>2</sub>	18.7 <sub>2</sub>
Grey	3.3 <sub>4</sub>	16.5 <sub>4</sub>	2.8 <sub>2</sub>	5.2 <sub>1</sub>	3.7 <sub>4</sub>	24.0 <sub>4</sub>
LUV	4.3 <sub>5</sub>	19.4 <sub>5</sub>	3.5 <sub>4</sub>	6.0 <sub>3</sub>	4.8 <sub>5</sub>	28.4 <sub>5</sub>

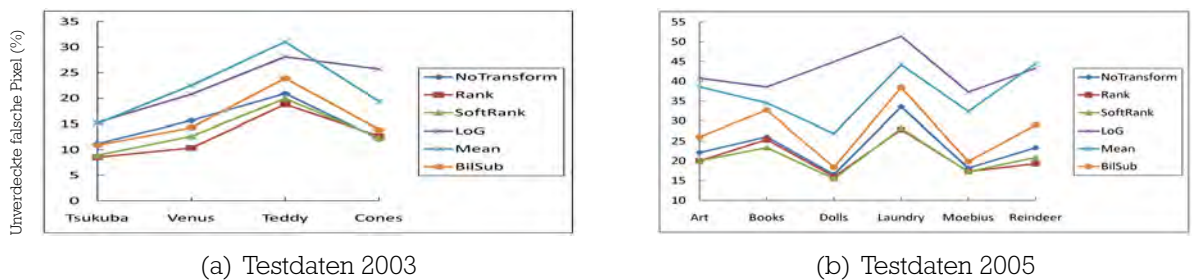


Abbildung 9.5: Vergleich der Gesamtperformance bei radiometrischen Transformationen der Testsets bei lokaler Optimierung.

ebenfalls durchgehend bessere Ergebnisse liefert als nicht transformierte Testdaten. Dies erklärt sich wohl aus der Tatsache, dass bilaterale Subtraktion intern schon eine Gewichtung der Daten vornimmt und eine erneute Gewichtung wenig zu einer Verbesserung der Ergebnisse beiträgt, sondern sogar zu einer Verschlechterung führt.

Hier soll nochmals darauf hingewiesen werden, dass die hier genannten gewichteten Transformationen sich von der gewichteten Aggregation abgrenzen. Gewichtete Transformationen verwenden die berechneten Gewichte in der Berechnung der Transformation, während gewichtete Aggregation nach der Kostenberechnung durchgeführt wird. Um diesen Unterschied nochmals herauszuarbeiten, zeigt Abbildung 9.8 die Performanceunterschiede zwischen dem Einsatz von radiometrischen Transformationen mit und ohne Gewichten, aufgeteilt in Gewichte nach Yoon und geodätische Gewichte. Abbildung 9.20 in Abschnitt 9.4 zeigt die Unterschiede bei gewichteter Aggregation.

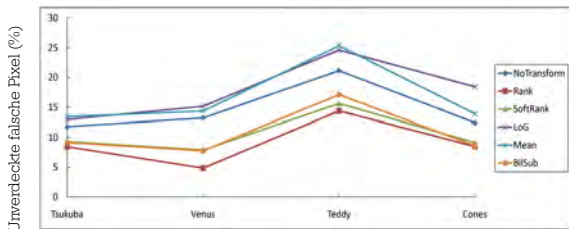
Abbildung 9.7 zeigt die Unterschiede in der Performance von radiometrischen Transformationen in den untersuchten Farbräumen. Für die Testdaten 2003 zeigt sich fast durchgehend eine Verschlechterung der Ergebnisse bei Farbverwendung. In den Testdaten 2005 treten teilweise Verbesserungen, teilweise Verschlechterungen auf. Die besten Ergebnisse aus den Farbräumen erzielen AC1C2 und I1I2I3. Die quantitative Auswertung findet sich in Tabelle 9.6.

Abbildung 9.8 zeigt die Ergebnisse der radiometrischen Transformationen mit und ohne Gewichtsverwendung in der Transformation. Im Falle von Rank und SoftRank verbessert eine Gewichtsverwendung die Ergebnisse nicht. Im Falle von Mittelwert und LoG Transformation verbessern sich die Ergebnisse stark. Die Verschlechterung der Ergebnisse bei Gewichtsverwendung im Falle von bilateraler Subtraktion ist wenig überraschend. Bilaterale Subtraktion arbeitet intern schon mit Gewichten und eine nochmalige Gewichtung der Ergebnisse verschlechtert diese merklich. In fast allen Fällen sind geodätische Gewichte im Vorteil. Tabelle 9.7 zeigt die Ergebnisse numerisch. Der rote Index stellt den Rang innerhalb der Zeile bezogen auf das jeweilige Datenset (Alle, 2003, 2005) dar. Wieder liefern, in Summe gesehen, die geodätischen Gewichte das beste Ergebnis (Rang 1.6 im Set „Alle Daten“).

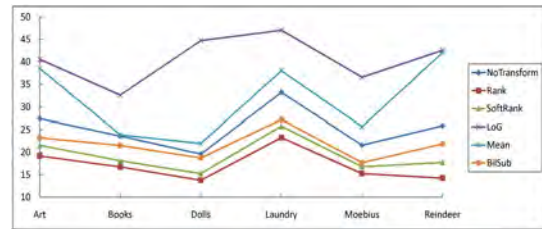
Betrachtet man nur die radiometrischen Transformationen mit Gewichtsunterstützung, werden die Effekte von großen Fenstergrößen etwas abgeschwächt, und es ergibt sich das eindeutige Bild aus Abbildung 9.9. Für die Testdaten 2003 zeigt sich, dass die Verwendung von Farbinformation bei radiometrischen Transformationen keine Verbesse-

Tabelle 9.5: Radiometrische Transformationen im Vergleich bei lokaler Optimierung.

Radio. Transform.	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
Rank	1.5 <sub>1</sub>	17.6 <sub>1</sub>	1.5 <sub>1</sub>	12.6 <sub>1</sub>	1.5 <sub>1</sub>	21.0 <sub>2</sub>
SoftRank	1.7 <sub>2</sub>	17.9 <sub>2</sub>	2.0 <sub>2</sub>	13.5 <sub>2</sub>	1.5 <sub>1</sub>	20.9 <sub>1</sub>
NoTransform	3.0 <sub>3</sub>	20.0 <sub>3</sub>	3.0 <sub>3</sub>	15.0 <sub>3</sub>	3.0 <sub>3</sub>	23.3 <sub>3</sub>
BilSub	3.8 <sub>4</sub>	22.4 <sub>4</sub>	3.5 <sub>4</sub>	15.4 <sub>4</sub>	4.0 <sub>4</sub>	27.4 <sub>4</sub>
Mean	5.3 <sub>5</sub>	30.9 <sub>5</sub>	5.5 <sub>5</sub>	22.0 <sub>5</sub>	5.2 <sub>5</sub>	36.8 <sub>5</sub>
LoG	5.7 <sub>6</sub>	34.6 <sub>6</sub>	5.5 <sub>6</sub>	22.5 <sub>6</sub>	5.8 <sub>6</sub>	42.7 <sub>6</sub>

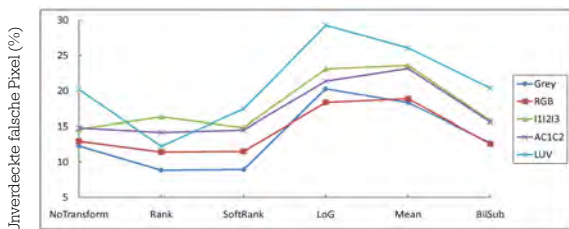


(a) Testdaten 2003

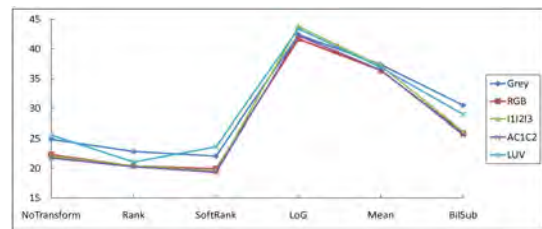


(b) Testdaten 2005

Abbildung 9.6: Vergleich der Gesamtperformance bei radiometrischen Transformationen der Testsets bei lokaler Optimierung, ohne den Einsatz von Gewichten in den Transformationen.



(a) Testdaten 2003



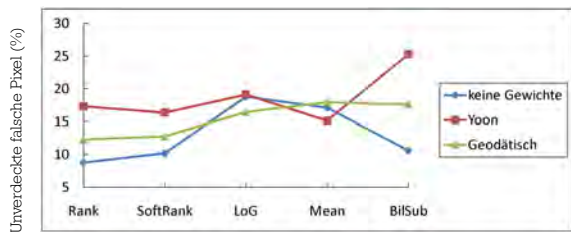
(b) Testdaten 2005

Abbildung 9.7: Vergleich der Gesamtperformance bei radiometrischen Transformationen in den untersuchten Farbräumen.

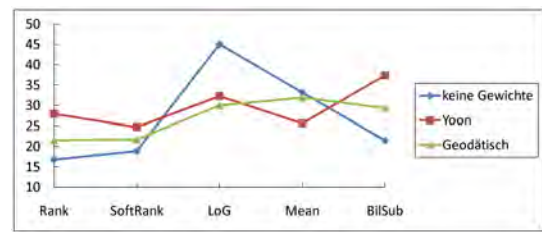


Tabelle 9.6: Vergleich der Gesamtperformance bei radiometrischen Transformationen in den untersuchten Farbräumen.

Radiom. Transf.	Alle Daten					Testdaten 2003					Testdaten 2005				
	Grey	RGB	I1I2I3	AC1C2	LUV	Grey	RGB	I1I2I3	AC1C2	LUV	Grey	RGB	I1I2I3	AC1C2	LUV
NoTransform	19.8 <sub>4</sub>	18.6 <sub>1</sub>	19.1 <sub>3</sub>	19.0 <sub>2</sub>	23.4 <sub>5</sub>	12.3 <sub>1</sub>	13.0 <sub>2</sub>	14.6 <sub>3</sub>	14.8 <sub>4</sub>	20.2 <sub>5</sub>	24.9 <sub>4</sub>	22.3 <sub>3</sub>	22.0 <sub>2</sub>	21.7 <sub>1</sub>	25.6 <sub>5</sub>
Rank	17.2 <sub>2</sub>	16.8 <sub>1</sub>	18.8 <sub>5</sub>	17.8 <sub>4</sub>	17.5 <sub>3</sub>	8.9 <sub>1</sub>	11.4 <sub>2</sub>	16.3 <sub>5</sub>	14.2 <sub>4</sub>	12.2 <sub>3</sub>	22.8 <sub>5</sub>	20.4 <sub>2</sub>	20.4 <sub>3</sub>	20.3 <sub>1</sub>	21.0 <sub>4</sub>
SoftRank	16.8 <sub>2</sub>	16.6 <sub>1</sub>	17.7 <sub>4</sub>	17.4 <sub>3</sub>	21.1 <sub>5</sub>	9.0 <sub>1</sub>	11.5 <sub>2</sub>	14.9 <sub>4</sub>	14.5 <sub>3</sub>	17.5 <sub>5</sub>	22.0 <sub>4</sub>	19.9 <sub>3</sub>	19.6 <sub>2</sub>	19.3 <sub>1</sub>	23.6 <sub>5</sub>
LoG	33.6 <sub>2</sub>	32.4 <sub>1</sub>	35.5 <sub>4</sub>	33.9 <sub>3</sub>	37.7 <sub>5</sub>	20.3 <sub>2</sub>	18.4 <sub>1</sub>	23.1 <sub>4</sub>	21.4 <sub>3</sub>	29.3 <sub>5</sub>	42.4 <sub>3</sub>	41.7 <sub>1</sub>	43.8 <sub>5</sub>	42.3 <sub>2</sub>	43.4 <sub>4</sub>
Mean	29.8 <sub>2</sub>	29.4 <sub>1</sub>	31.7 <sub>4</sub>	31.1 <sub>3</sub>	32.5 <sub>5</sub>	18.3 <sub>1</sub>	18.9 <sub>2</sub>	23.6 <sub>4</sub>	23.2 <sub>3</sub>	26.1 <sub>5</sub>	37.4 <sub>5</sub>	36.4 <sub>1</sub>	37.2 <sub>4</sub>	36.4 <sub>2</sub>	36.9 <sub>3</sub>
BilSub	23.0 <sub>4</sub>	20.3 <sub>1</sub>	21.8 <sub>3</sub>	21.4 <sub>2</sub>	25.4 <sub>5</sub>	12.7 <sub>2</sub>	12.6 <sub>1</sub>	15.8 <sub>4</sub>	15.6 <sub>3</sub>	20.4 <sub>5</sub>	30.5 <sub>5</sub>	25.9 <sub>2</sub>	26.1 <sub>3</sub>	25.6 <sub>1</sub>	29.0 <sub>4</sub>
Mittelwert	23.3 <sub>2,6</sub>	22.3 <sub>1</sub>	24.0 <sub>3,8</sub>	23.4 <sub>2,8</sub>	26.2 <sub>4,6</sub>	13.6 <sub>1,3</sub>	14.3 <sub>1,6</sub>	18.0 <sub>4</sub>	17.2 <sub>3,3</sub>	20.9 <sub>4,6</sub>	29.8 <sub>4,3</sub>	27.7 <sub>2</sub>	28.1 <sub>3,2</sub>	27.5 <sub>1,3</sub>	29.8 <sub>4,2</sub>



(a) Testdaten 2003



(b) Testdaten 2005

Abbildung 9.8: Vergleich der Gesamtperformance bei radiometrischen Transformationen mit und ohne Gewichten bei lokaler Optimierung.

Tabelle 9.7: Vergleich der Gesamtperformance bei radiometrischen Transformationen ohne und mit Gewichtsverwendung bei lokaler Optimierung.

Radiom. Transf.	Alle Daten			Testdaten 2003			Testdaten 2005		
	keine	Yoon	Geodätisch	keine	Yoon	Geodätisch	keine	Yoon	Geodätisch
Rank	13.6 <sub>1</sub>	23.7 <sub>3</sub>	17.8 <sub>2</sub>	8.8 <sub>1</sub>	17.3 <sub>3</sub>	12.3 <sub>2</sub>	16.8 <sub>1</sub>	28.0 <sub>3</sub>	21.5 <sub>2</sub>
SoftRank	15.4 <sub>1</sub>	21.4 <sub>3</sub>	18.1 <sub>2</sub>	10.2 <sub>1</sub>	16.4 <sub>3</sub>	12.8 <sub>2</sub>	18.9 <sub>1</sub>	24.7 <sub>3</sub>	21.7 <sub>2</sub>
LoG	34.6 <sub>3</sub>	27.0 <sub>2</sub>	24.6 <sub>1</sub>	18.8 <sub>2</sub>	19.1 <sub>3</sub>	16.5 <sub>1</sub>	45.1 <sub>3</sub>	32.3 <sub>2</sub>	30.0 <sub>1</sub>
Mean	26.8 <sub>3</sub>	21.5 <sub>1</sub>	26.3 <sub>2</sub>	17.1 <sub>2</sub>	15.1 <sub>1</sub>	17.9 <sub>3</sub>	33.2 <sub>3</sub>	25.7 <sub>1</sub>	31.9 <sub>2</sub>
BilSub	17.1 <sub>1</sub>	31.3 <sub>3</sub>	23.5 <sub>2</sub>	10.6 <sub>1</sub>	25.3 <sub>3</sub>	17.6 <sub>2</sub>	21.5 <sub>1</sub>	37.4 <sub>3</sub>	29.4 <sub>2</sub>
Mittelwert	21.5 <sub>1,8</sub>	24.1 <sub>2,2</sub>	21.9 <sub>1,6</sub>	13.1 <sub>1,3</sub>	18.6 <sub>2,6</sub>	15.4 <sub>2</sub>	27.1 <sub>1,8</sub>	29.6 <sub>2,4</sub>	26.9 <sub>1,6</sub>

rung bringt. Einzig bei der bilateralen Subtraktion stellt sich eine durchgehende Verbesserung durch Farbeinsatz ein. Bei den Testdaten 2005 zeigt sich ein gegenteiliges Bild. Bis auf einen Ausreißer bei der Mittelwerttransformation erzielen alle Farbräume bessere Ergebnisse als reine intensitätsbasierte Transformationen. Wieder zeigt sich ein sehr gutes Abschneiden des AC1C2 Farbraumes und des I1I2I3 Farbraumes.

## 9.2.2 Globaler Ansatz

Abbildung 9.10 zeigt die Performance der untersuchten radiometrischen Transformationen auf den Testdaten. Auffällig ist das ungewöhnlich schlechte Abschneiden von LoG und der Mittelwert Transformation. Dies könnte sich aus der verwendeten, sehr großen Fenstergröße (29x29 Pixel) erklären. Besonders LoG und Mittelwert liefern bei Anwendung ohne Gewichtung sehr schlechte Ergebnisse ab einer Fenstergröße von ungefähr 15x15 Pixel. Dies erklärt wohl auch die in [HS09] gewählten Fenstergrößen von 5x5 Pixel für LoG und 15x15 Pixel für die Mittelwerttransformation. Im Training schwächten die guten Ergebnisse der restlichen Transformationen die LoG und Mittelwert Transformationsfehler jedoch genügend ab, um eine Wahl der großen Fenstergröße zu rechtfertigen. Wie im lokalen Fall liefert Rank sehr gute Ergebnisse. Im Unterschied zum lokalen Fall liefert auch die bilaterale Subtraktion sehr gute Ergebnisse, jedoch zeigt sich, dass die Anwendung keiner Transformation durchgehend die besten Ergebnisse liefert.

Tabelle 9.8 zeigt die quantitative Auflistung der untersuchten radiometrischen Transformationen auf den Testdaten. Im globalen Fall scheint eine Anwendung von radiometrischen Transformationen nur beschränkt sinnvoll. Jedoch liefern besonders Rank und bilaterale Subtraktion gute Ergebnisse.

Abbildung 9.11 zeigt die Unterschiede in der Performance von radiometrischen Transformationen in den untersuchten Farbräumen. Es zeigt sich in fast allen Farbräumen und Testdaten eine Verbesserung der Ergebnisse gegenüber rein intensitätsbasierten Transformationen. Die quantitative Auswertung findet sich in Tabelle 9.9. Es ist zu beachten, dass der rote Index hier den Rang innerhalb der jeweiligen Testdaten bezeichnet. Zum Beispiel belegt der Farbraum RGB für die Rank Transformation in den Testdaten „Alle Daten“ den Rang 1. Weiters wäre somit der Spitzenreiter im Datenset „Alle Daten“ der AC1C2 Farbraum mit einer mittleren Fehlerrate von 24.2 Prozent und einem mittleren Rang von 1.6.

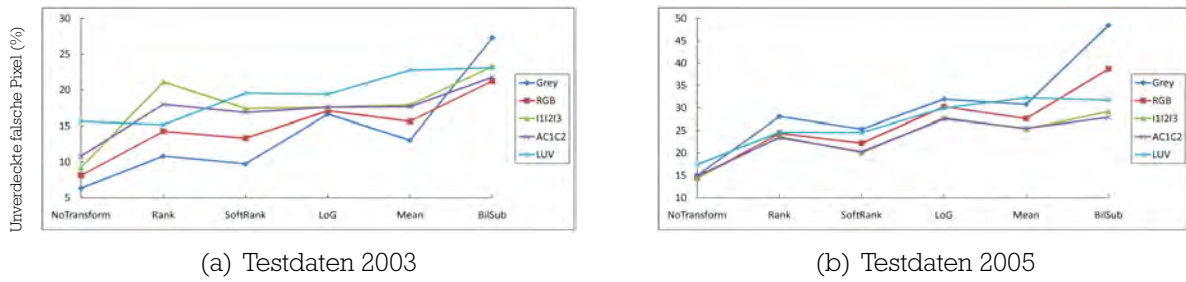


Abbildung 9.9: Vergleich der Gesamtperformance bei radiometrischen Transformationen mit Gewichtsverwendung in den untersuchten Farbräumen.

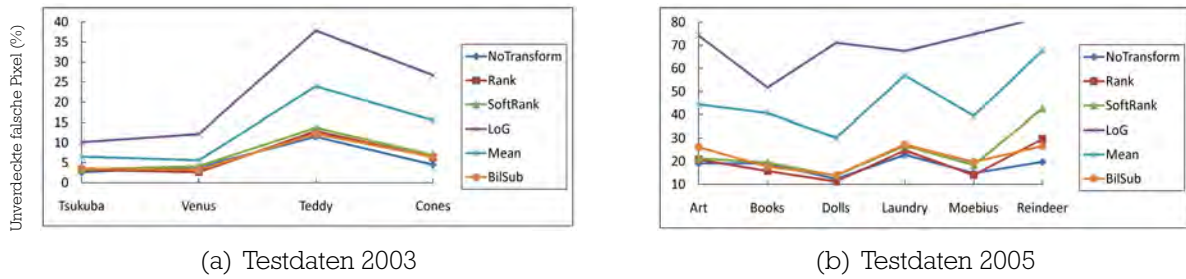


Abbildung 9.10: Vergleich der Gesamtperformance bei radiometrischen Transformationen der Testsets bei globaler Optimierung.

Tabelle 9.8: Radiometrische Transformationen im Vergleich bei globaler Optimierung.

Radio. Transform.	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
NoTransform	1.6 <sub>1</sub>	13.1 <sub>1</sub>	1.5 <sub>1</sub>	5.7 <sub>1</sub>	1.7 <sub>1</sub>	18.0 <sub>1</sub>
Rank	1.8 <sub>2</sub>	14.2 <sub>2</sub>	2.0 <sub>2</sub>	6.3 <sub>2</sub>	1.7 <sub>1</sub>	19.4 <sub>2</sub>
BilSub	3.1 <sub>3</sub>	15.7 <sub>3</sub>	2.8 <sub>3</sub>	6.4 <sub>3</sub>	3.3 <sub>2</sub>	21.9 <sub>3</sub>
SoftRank	3.5 <sub>4</sub>	17.1 <sub>4</sub>	3.8 <sub>4</sub>	7.0 <sub>4</sub>	3.3 <sub>2</sub>	23.7 <sub>4</sub>
Mean	5.0 <sub>5</sub>	33.1 <sub>5</sub>	5.0 <sub>5</sub>	12.9 <sub>5</sub>	5.0 <sub>5</sub>	46.6 <sub>5</sub>
LoG	6.0 <sub>6</sub>	50.9 <sub>6</sub>	6.0 <sub>6</sub>	21.7 <sub>6</sub>	6.0 <sub>6</sub>	70.3 <sub>6</sub>

## Kapitel 9 Ergebnisse

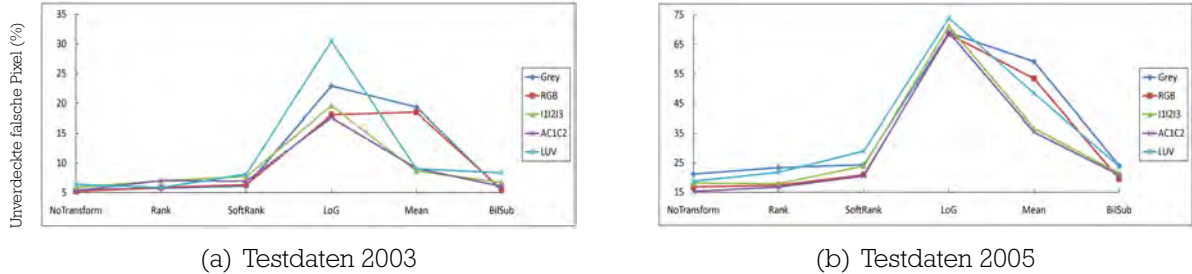


Abbildung 9.11: Vergleich der Gesamtperformance bei radiometrischen Transformationen in den untersuchten Farbräumen.

Tabelle 9.9: Vergleich der Gesamtperformance bei radiometrischen Transformationen in den untersuchten Farbräumen.

Radiom. Transf.	Alle Daten					Testdaten 2003					Testdaten 2005				
	Grey	RGB	I1I2I3	AC1C2	LUV	Grey	RGB	I1I2I3	AC1C2	LUV	Grey	RGB	I1I2I3	AC1C2	LUV
NoTransform	14.9 <sub>5</sub>	12.2 <sub>2</sub>	13.2 <sub>3</sub>	11.3 <sub>1</sub>	13.8 <sub>4</sub>	5.5 <sub>3</sub>	5.2 <sub>1</sub>	5.9 <sub>4</sub>	5.3 <sub>2</sub>	6.4 <sub>5</sub>	21.2 <sub>5</sub>	16.8 <sub>2</sub>	18.1 <sub>3</sub>	15.3 <sub>1</sub>	18.7 <sub>4</sub>
Rank	16.3 <sub>5</sub>	12.8 <sub>1</sub>	13.5 <sub>3</sub>	12.9 <sub>2</sub>	15.4 <sub>4</sub>	5.8 <sub>1</sub>	5.8 <sub>2</sub>	7.0 <sub>4</sub>	7.0 <sub>5</sub>	5.9 <sub>3</sub>	23.4 <sub>5</sub>	17.4 <sub>2</sub>	17.9 <sub>3</sub>	16.74 <sub>1</sub>	21.76 <sub>4</sub>
SoftRank	17.1 <sub>3</sub>	15.1 <sub>2</sub>	17.4 <sub>4</sub>	15.1 <sub>1</sub>	20.6 <sub>5</sub>	6.1 <sub>1</sub>	6.3 <sub>2</sub>	7.7 <sub>4</sub>	6.9 <sub>3</sub>	8.1 <sub>5</sub>	24.4 <sub>4</sub>	20.9 <sub>2</sub>	23.9 <sub>3</sub>	20.5 <sub>1</sub>	29.0 <sub>5</sub>
LoG	50.5 <sub>4</sub>	48.3 <sub>1</sub>	50.5 <sub>3</sub>	48.5 <sub>2</sub>	56.5 <sub>5</sub>	22.9 <sub>4</sub>	18.0 <sub>2</sub>	19.6 <sub>3</sub>	17.6 <sub>1</sub>	30.5 <sub>5</sub>	69.0 <sub>2</sub>	68.5 <sub>1</sub>	71.1 <sub>4</sub>	69.02 <sub>3</sub>	73.9 <sub>5</sub>
Mean	43.2 <sub>5</sub>	39.4 <sub>4</sub>	25.4 <sub>2</sub>	24.9 <sub>1</sub>	32.6 <sub>3</sub>	19.3 <sub>5</sub>	18.5 <sub>4</sub>	8.6 <sub>1</sub>	9.1 <sub>3</sub>	9.0 <sub>2</sub>	59.1 <sub>5</sub>	53.4 <sub>4</sub>	36.7 <sub>2</sub>	35.4 <sub>1</sub>	48.4 <sub>3</sub>
BilSub	16.5 <sub>4</sub>	13.9 <sub>1</sub>	15.6 <sub>3</sub>	14.9 <sub>2</sub>	17.6 <sub>5</sub>	5.3 <sub>1</sub>	5.5 <sub>2</sub>	6.7 <sub>4</sub>	6.0 <sub>3</sub>	8.4 <sub>5</sub>	24.0 <sub>5</sub>	19.5 <sub>1</sub>	21.5 <sub>3</sub>	20.9 <sub>2</sub>	23.8 <sub>4</sub>
Mittelwert	27.9 <sub>4,3</sub>	25.1 <sub>1,8</sub>	23.7 <sub>3</sub>	22.3 <sub>1,5</sub>	27.5 <sub>4,3</sub>	11.6 <sub>2,5</sub>	10.5 <sub>2,2</sub>	9.7 <sub>3,3</sub>	9.1 <sub>2,8</sub>	11.9 <sub>4,2</sub>	38.8 <sub>4,3</sub>	34.8 <sub>2</sub>	33.1 <sub>3</sub>	31.2 <sub>1,5</sub>	37.9 <sub>4,2</sub>

Abbildung 9.12 zeigt die Ergebnisse der radiometrischen Transformationen mit und ohne Gewichtsverwendung in der Transformation. Im Falle von Rank und SoftRank verbessert eine Gewichtsverwendung die Ergebnisse. Im Gegensatz zum lokalen Fall verschlechtern sich die Ergebnisse jedoch für Mittelwert und LoG Transformation. Geodätische Gewichte sind gegenüber Gewichten nach Yoon im Vorteil. Tabelle 9.10 zeigt die Ergebnisse numerisch. Der rote Index stellt den Rang innerhalb der Zeile bezogen auf das jeweilige Datenset (Alle, 2003, 2005) dar. Eine Gewichtsverwendung in radiometrischen Transformationen scheint somit wenig sinnvoll (mittlerer Rang von 1.5 auf dem Datenset „Alle Daten“ bei Verwendung keiner Gewichte). Anzumerken bleibt, dass für die bilaterale Subtraktion keine Tests mit Gewichten im globalen Fall durchgeführt wurden, da diese Transformation zu einer der rechenintensivsten Transformationen zählt und die Verarbeitungszeit gerade im globalen Fall vervielfacht hätte.

Betrachtet man nur die radiometrischen Transformationen mit Gewichtsunterstützung in Abbildung 9.13, ergibt sich ein ähnliches Bild wie im lokalen Fall, jedoch fallen die erzielten Verbesserungen durch Farbeinsatz geringer aus.

### 9.3 Kostenberechnungsmethode

Folgender Abschnitt zeigt alle Ergebnisse in Bezug auf die untersuchten Kostenberechnungsmethoden. Für alle folgenden Abbildungen und Tabellen gilt, dass bis auf den Parameter Kostenberechnungsmethode alle Parameter aus Tabelle 8 variiert werden. Das heißt zum Beispiel, dass in den mittleren Ergebniswerten für HMI alle Ergebnisse aller durchgeführten Parametervariationen enthalten sind, in denen HMI Anwendung fand.

#### 9.3.1 Lokaler Ansatz

In Abbildung 9.14 ist ein Vergleich der Gesamtperformance der untersuchten Kostenberechnungsmethoden dargestellt. Als Erstes fällt das überraschend gute Abschneiden von NCC auf. In fast allen Fällen liefert es bessere Ergebnisse als der erwartete Spitzenreiter HMI.

Besonders die NCC Version, die Gewichte benutzt, scheint zu diesem guten Ergebnis beizutragen. Zweitens erkennt man, dass BT immer schlechter als SAD abschneidet.

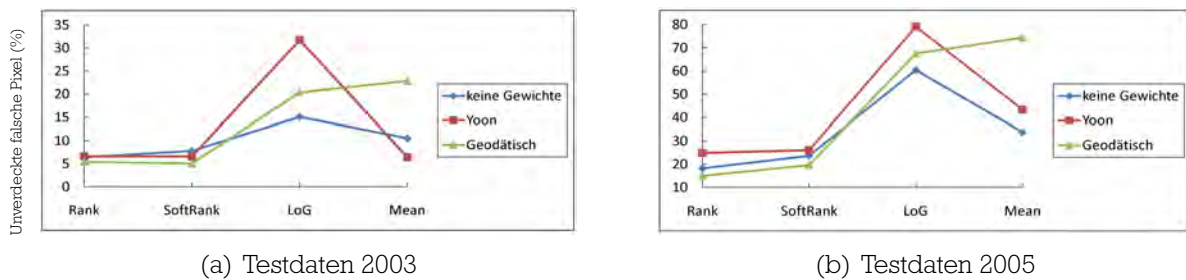


Abbildung 9.12: Vergleich der Gesamtperformance bei radiometrischen Transformationen mit und ohne Gewichten bei globaler Optimierung.

Tabelle 9.10: Vergleich der Gesamtperformance bei radiometrischen Transformationen ohne und mit Gewichtsverwendung bei globaler Optimierung.

Radiom. Transf.	Alle Daten			Testdaten 2003			Testdaten 2005		
	keine	Yoon	Geodätisch	keine	Yoon	Geodätisch	keine	Yoon	Geodätisch
Rank	13.5 <sub>2</sub>	17.6 <sub>3</sub>	11.2 <sub>1</sub>	6.4 <sub>2</sub>	6.6 <sub>3</sub>	5.4 <sub>1</sub>	18.2 <sub>2</sub>	24.8 <sub>3</sub>	15.0 <sub>1</sub>
SoftRank	17.3 <sub>2</sub>	18.3 <sub>3</sub>	13.8 <sub>1</sub>	7.8 <sub>3</sub>	6.6 <sub>2</sub>	5.1 <sub>1</sub>	23.7 <sub>2</sub>	26.1 <sub>3</sub>	19.6 <sub>1</sub>
LoG	42.3 <sub>1</sub>	60.2 <sub>2</sub>	48.7 <sub>3</sub>	15.1 <sub>1</sub>	31.7 <sub>3</sub>	20.4 <sub>2</sub>	60.4 <sub>1</sub>	79.1 <sub>3</sub>	67.6 <sub>2</sub>
Mean	24.4 <sub>1</sub>	28.6 <sub>2</sub>	53.8 <sub>3</sub>	10.5 <sub>2</sub>	6.4 <sub>1</sub>	22.9 <sub>3</sub>	33.7 <sub>1</sub>	43.4 <sub>2</sub>	74.4 <sub>3</sub>
Mittelwert	24.4 <sub>1.5</sub>	31.1 <sub>2.5</sub>	31.9 <sub>2</sub>	10.0 <sub>1.3</sub>	12.8 <sub>2.6</sub>	13.4 <sub>2</sub>	34.0 <sub>1.5</sub>	43.4 <sub>2.8</sub>	44.2 <sub>1.8</sub>

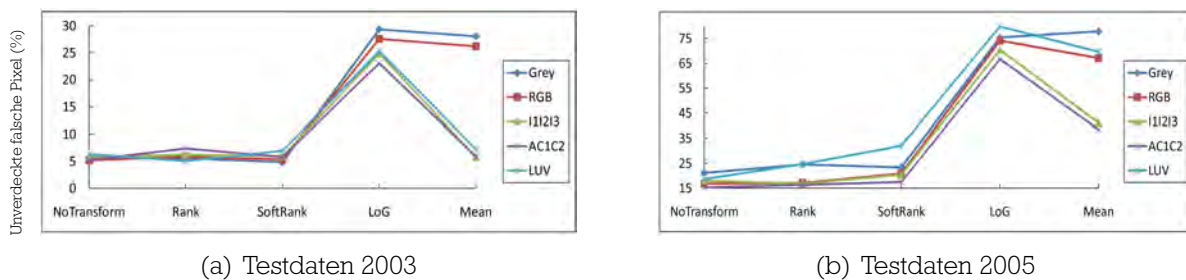


Abbildung 9.13: Vergleich der Gesamtperformance bei radiometrischen Transformationen mit Gewichtsverwendung in den untersuchten Farbräumen.

Hier scheint ein Zusammenhang mit der gewählten Aggregationsmethode zu bestehen. Verwendet man reine quadratische Aggregation, wie in Abbildungen 9.15(a) und 9.15(b) dargestellt, zeigt sich zumindest teilweise ein Vorteil für BT. Kommt hingegen gewichtete Aggregation zur Anwendung, wie in Abbildungen 9.15(c) und 9.15(d) dargestellt, ist BT immer klar im Nachteil.

Tabelle 9.11 zeigt die quantitative Auswertung des Vergleiches der Kostenberechnungsmethoden. Betrachtet man die Ränge der Kostenberechnungsmethoden, zeigt sich, dass NCC (1.3 zu 1.0) und HMI (2.8 zu 2.3) für die weniger anspruchsvollen Testdaten 2003 im Mittel schlechtere Ergebnisse erzielen als für die anspruchsvolleren Testdaten 2005. Bei SAD (2.5 zu 2.7) und BT (3.5 zu 4.0) verhält es sich genau umgekehrt. Dies lässt die Vermutung zu, dass NCC und HMI besonders bei anspruchsvolleren Bilddaten gute Ergebnisse erzielen, während die Ergebnisqualität von SAD und BT mit der Zunahme der Bilddatenkomplexität abnimmt.

### 9.3.2 Globaler Ansatz

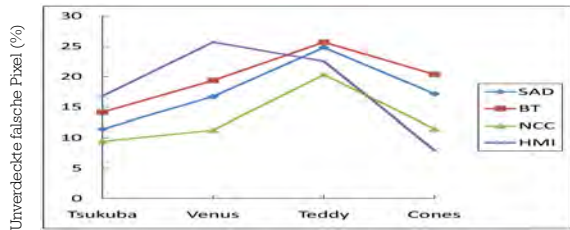
In Abbildung 9.16 ist ein Vergleich der Gesamtperformance der untersuchten Kostenberechnungsmethoden dargestellt. Hier zeigt sich ein eindeutiges, den Erwartungen entsprechendes Ergebnis. HMI erzielt in fast allen Fällen die besten Ergebnisse, gefolgt von BT und SAD. Besonders im Fall der Testdaten 2005 kann HMI wesentlich bessere Ergebnisse erzielen als die beiden anderen Methoden. Dies bestätigt die Vermutung aus dem lokalen Ansatz, dass HMI besonders bei anspruchsvolleren Bilddaten gute Ergebnisse erzielt.

Tabelle 9.12 zeigt die quantitative Auswertung des Vergleiches der Kostenberechnungsmethoden. Wie auch schon aus Abbildung 9.16 ersichtlich, liefert HMI sehr gute Ergebnisse.

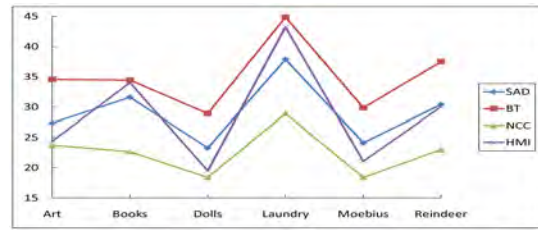
## 9.4 Fenstergröße, Gewichtsverwendung und Gewichtstyp

Folgender Abschnitt zeigt alle Ergebnisse in Bezug auf variierte Fenstergröße, Gewichtsverwendung und Gewichtstyp. Für alle folgenden Abbildungen und Tabellen

## Kapitel 9 Ergebnisse



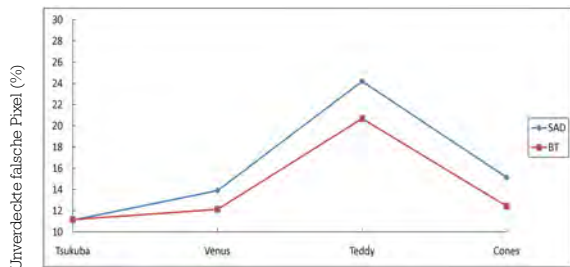
(a) Testdaten 2003



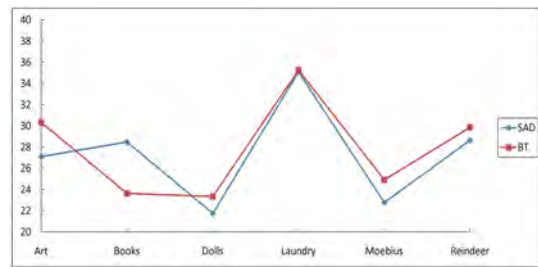
(b) Testdaten 2005

Abbildung 9.14: Vergleich der Gesamtperformance der Kostenberechnungsmethoden auf den Testdaten bei lokaler Optimierung.

### Quadratische Aggregation

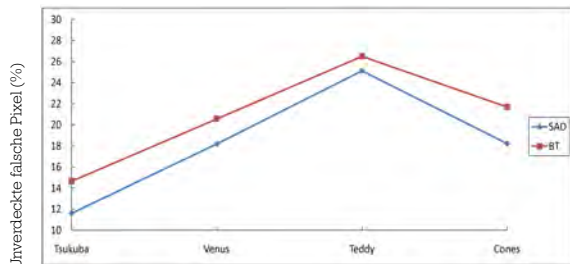


(a) Testdaten 2003

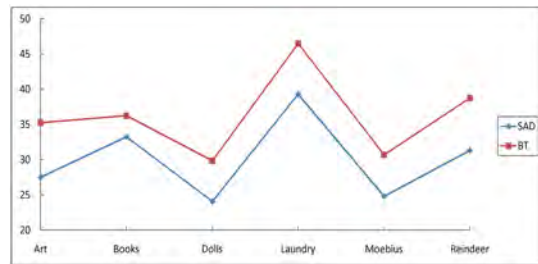


(b) Testdaten 2005

### Aggregation mit Gewichten

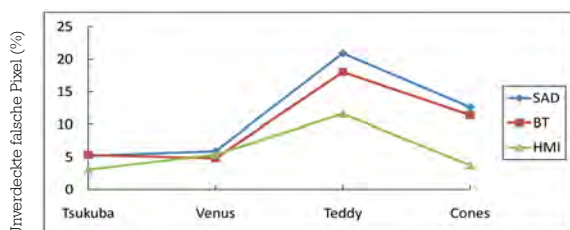


(c) Testdaten 2003

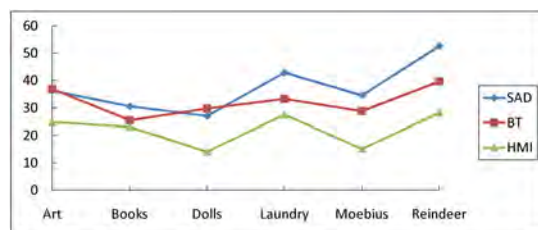


(d) Testdaten 2005

Abbildung 9.15: Vergleich der Gesamtperformance von SAD und BT unter Berücksichtigung der Aggregationsmethode.



(a) Testdaten 2003



(b) Testdaten 2005

Abbildung 9.16: Vergleich der Gesamtperformance der Kostenberechnungsmethoden auf den Testdaten bei globaler Optimierung.



gilt, dass bis auf die betrachteten Parameter alle Parameter aus Tabelle 8 variiert werden und die mittleren Ergebnisse aller durchgeführten Kombinationen angezeigt werden. Hier wurde keine Aufteilung in lokalen und globalen Ansatz vorgenommen, da ein Großteil der Betrachtungen nur für den lokalen Fall durchgeführt wurde. Weiters wurde die Betrachtung von Gewichtsverwendung bei radiometrischen Transformationen schon im Abschnitt 9.2 behandelt.

Abbildung 9.17 zeigt die Performance unterschiedlicher Fenstergrößen. Die mittlere Fenstergröße erzielt die besten Ergebnisse. Weiters ist festzustellen, dass im Bezug auf die Fenstergröße die Wahl der Gewichtsrechnungsmethode (in dieser Arbeit Gewichte nach Yoon oder geodätische Gewichte) keinen Einfluss hat. Tabelle 9.13 zeigt die zugehörigen quantitativen Ergebnisse.

Abbildung 9.18 visualisiert die Performance mit und ohne Gewichtsverwendung bei Aggregation, radiometrischen Transformationen und NCC. Hier ist anzumerken, dass die Wahl der Gewichtsrechnungsmethode einen Unterschied darstellt. In fast allen Fällen führt die Verwendung von geodätischen Gewichten durchgehend zu einer Verbesserung der Ergebnisse. In der Mehrzahl der Fälle liefern geodätische Gewichte bessere Ergebnisse als Gewichte nach Yoon. Tabelle 9.14 zeigt die Ergebnisse wieder numerisch.

Bei quadratischer Aggregation liefert die kleinste Fenstergröße die besten Ergebnisse, während bei gewichteter Aggregation die größte Fenstergröße empfehlenswert ist. Abbildung 9.19 visualisiert diesen Sachverhalt.

Betrachtet man die Verwendung von Gewichten nur in der Aggregation, zeigt sich das Bild aus Abbildung 9.20. Es ist erkennbar, dass in fast allen Fällen der Einsatz von Gewichten bei der Aggregation sinnvoll ist. Weiters sind geodätische Gewichte klar im Vorteil. Tabelle 9.15 zeigt wieder die numerischen Ergebnisse. Die Betrachtung von Gewichtsverwendung in den radiometrischen Transformationen erfolgte schon in Abschnitt 9.2.1.

## 9.5 Beispielhafte Ergebnisse

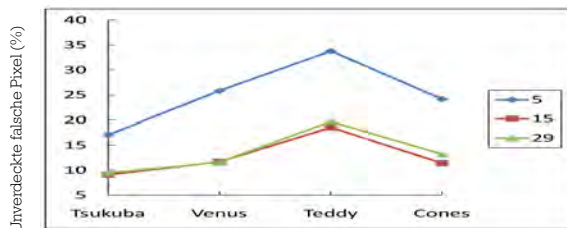
Im folgenden Abschnitt finden sich beispielhafte Ergebnisse, die mit dem erstellten Stereoframework erzielt wurden.

Tabelle 9.11: Kostenberechnungsmethoden im Vergleich bei lokaler Optimierung.

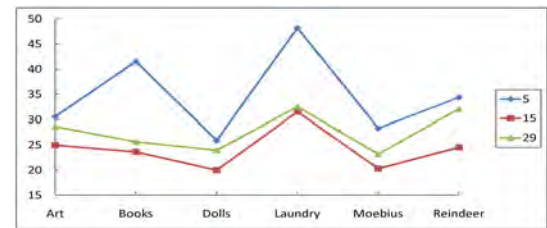
Methode	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
NCC	1.1 <sub>1</sub>	18.8 <sub>1</sub>	1.3 <sub>1</sub>	13.1 <sub>1</sub>	1.0 <sub>1</sub>	22.5 <sub>1</sub>
HMI	2.5 <sub>2</sub>	24.5 <sub>3</sub>	2.8 <sub>3</sub>	18.2 <sub>3</sub>	2.3 <sub>2</sub>	28.7 <sub>2</sub>
SAD	2.6 <sub>3</sub>	24.4 <sub>2</sub>	2.5 <sub>2</sub>	17.4 <sub>2</sub>	2.7 <sub>3</sub>	29.1 <sub>3</sub>
BT	3.8 <sub>4</sub>	29.0 <sub>4</sub>	3.5 <sub>4</sub>	19.9 <sub>4</sub>	4.0 <sub>4</sub>	35.0 <sub>4</sub>

Tabelle 9.12: Kostenberechnungsmethoden im Vergleich bei globaler Optimierung.

Methode	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
HMI	1.1 <sub>1</sub>	15.7 <sub>1</sub>	1.3 <sub>1</sub>	6.0 <sub>1</sub>	1.0 <sub>1</sub>	22.1 <sub>1</sub>
BT	2.2 <sub>2</sub>	23.3 <sub>2</sub>	2.0 <sub>2</sub>	9.9 <sub>2</sub>	2.3 <sub>2</sub>	32.3 <sub>2</sub>
SAD	2.7 <sub>3</sub>	26.8 <sub>3</sub>	2.8 <sub>3</sub>	11.2 <sub>3</sub>	2.7 <sub>3</sub>	37.3 <sub>3</sub>



(a) Testdaten 2003



(b) Testdaten 2005

Abbildung 9.17: Vergleich verschiedener Fenstergrößen auf den Testdaten bei lokaler Optimierung.

Tabelle 9.13: Fenstergröße im Vergleich bei lokaler Optimierung.

Fenstergröße	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>	Rang <sub>∅</sub>	Fehler <sub>∅</sub>
15	1.3 <sub>1</sub>	19.8 <sub>1</sub>	1.0 <sub>1</sub>	12.9 <sub>1</sub>	1.1 <sub>1</sub>	24.2 <sub>1</sub>
29	1.8 <sub>2</sub>	21.5 <sub>2</sub>	2.0 <sub>2</sub>	13.1 <sub>2</sub>	1.9 <sub>2</sub>	27.7 <sub>2</sub>
5	3.0 <sub>3</sub>	31.0 <sub>3</sub>	3.0 <sub>3</sub>	25.2 <sub>3</sub>	3.0 <sub>3</sub>	34.8 <sub>3</sub>

## Kapitel 9 Ergebnisse

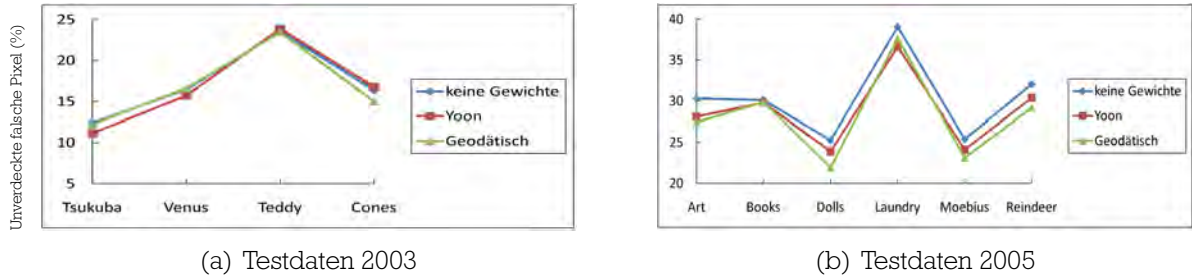
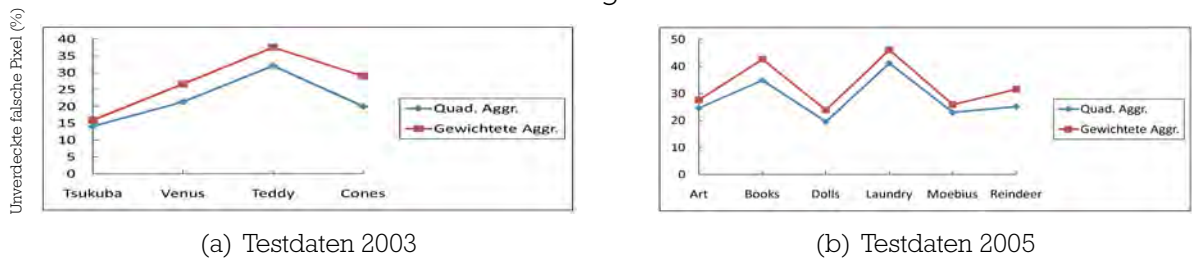


Abbildung 9.18: Vergleich der Verwendung von Gewichten in Aggregation und radiometrischen Transformationen sowie NCC auf den Testdaten bei lokaler Optimierung.

Tabelle 9.14: Gewichtsverwendung im Vergleich bei lokaler Optimierung.

Gewichte	Alle Daten		Testdaten 2003		Testdaten 2005	
	↓Rang $\emptyset$	Fehler $\emptyset$	Rang $\emptyset$	Fehler $\emptyset$	Rang $\emptyset$	Fehler $\emptyset$
Geodätisch	1.5 <sub>1</sub>	23.7 <sub>1</sub>	1.8 <sub>1</sub>	16.8 <sub>1</sub>	1.3 <sub>1</sub>	28.2 <sub>1</sub>
Yoon	1.8 <sub>2</sub>	24.1 <sub>2</sub>	2.0 <sub>2</sub>	16.9 <sub>2</sub>	1.7 <sub>2</sub>	28.8 <sub>2</sub>
ohne	2.7 <sub>3</sub>	25.1 <sub>3</sub>	2.3 <sub>3</sub>	17.2 <sub>3</sub>	3.0 <sub>3</sub>	30.4 <sub>3</sub>

### Fenstergröße 5



### Fenstergröße 29

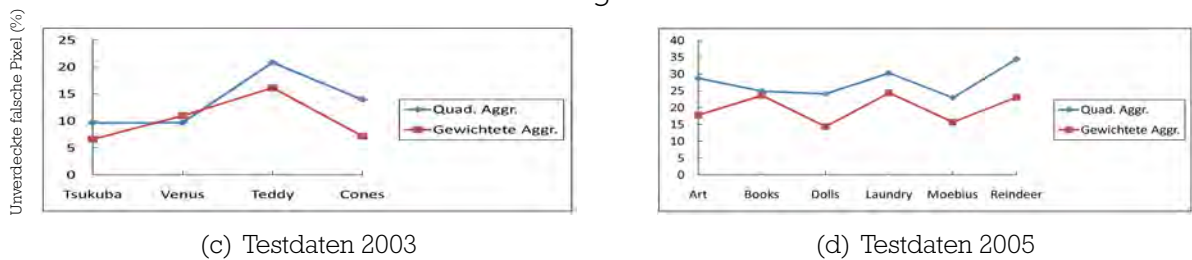


Abbildung 9.19: Vergleich der Gesamtperformance der Fenstergröße bei Einsatz von Gewichten in Aggregation und radiometrischen Transformationen sowie NCC.

### 9.5.1 Lokaler Ansatz

Abbildung 9.21 zeigt das jeweils beste Ergebnis (Disparitätskarte) je Testbild und das zugehörige Fehlerbild. Das Fehlerbild wird erzeugt, indem Pixel mit einer Abweichung in der Disparität vom Vergleichswert (der Groundtruth) um mehr als einen Pixel gesetzt werden. Schwarze Pixel stellen Fehler in unverdeckten Bereichen und graue Pixel Fehler in verdeckten Bereichen dar. In Tabelle 9.16 sind die verwendeten (relevanten) Parameter je Ergebnisbild aufgeschlüsselt. Abschließend sind in Tabelle 9.17 die Ergebnisse numerisch dargestellt.

### 9.5.2 Globaler Ansatz

Wie im lokalen Ansatz auch, zeigt Abbildung 9.21 das jeweils beste Ergebnis (Disparitätskarte) je Testbild und das zugehörige Fehlerbild. Das Fehlerbild wird erzeugt, indem Pixel mit einer Abweichung in der Disparität vom Vergleichswert (der Groundtruth) um mehr als einen Pixel gesetzt werden. Wobei schwarze Pixel Fehler in unverdeckten Bereichen und graue Pixel Fehler in verdeckten Bereichen darstellen. In Tabelle 9.18 sind die verwendeten Parameter je Ergebnisbild aufgeschlüsselt. Tabelle 9.19 zeigt die quantitativen Ergebnisse.

## Kapitel 9 Ergebnisse

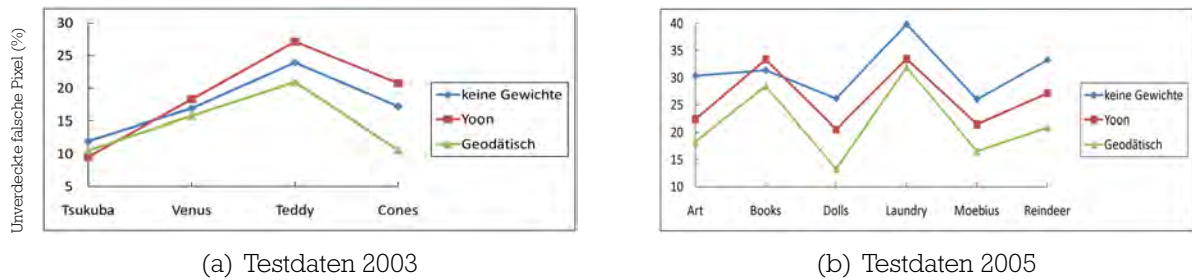


Abbildung 9.20: Vergleich der Verwendung von Gewichten im Aggregationsschritt bei lokaler Optimierung.

Tabelle 9.15: Gewichtsverwendung im Aggregationsschritt im Vergleich bei lokaler Optimierung.

Gewichte	Alle Daten		Testdaten 2003		Testdaten 2005	
	$\downarrow$ Rang $_{\emptyset}$	Fehler $_{\emptyset}$	Rang $_{\emptyset}$	Fehler $_{\emptyset}$	Rang $_{\emptyset}$	Fehler $_{\emptyset}$
Geodätisch	1.1 <sub>1</sub>	18.7 <sub>1</sub>	1.3 <sub>1</sub>	14.5 <sub>1</sub>	1.0 <sub>1</sub>	21.6 <sub>1</sub>
Yoon	2.3 <sub>2</sub>	23.4 <sub>2</sub>	2.5 <sub>3</sub>	18.9 <sub>3</sub>	2.2 <sub>2</sub>	26.4 <sub>2</sub>
ohne	2.6 <sub>3</sub>	25.7 <sub>3</sub>	2.3 <sub>2</sub>	17.5 <sub>2</sub>	2.8 <sub>3</sub>	31.2 <sub>3</sub>

Tabelle 9.16: Übersicht der verwendeten Parametereinstellungen bei lokaler Optimierung für die Erzeugung der Ergebnisbilder in Abbildung 9.21.

Testbild	Parameterset						
	Fenstergröße	Supporttype	Radio. Transf.	Farbraum	Kostenf.	Aggregation	Gewichte
Tsukuba	29	Geodesic	NoTransform	RGB	NCC	Gewichtet	Benutzt
Venus	15	Geodesic	Rank	Grey	SAD	Gewichtet	Benutzt
Teddy	15	Geodesic	Rank	Grey	BT	Gewichtet	Unbenutzt
Cones	15	Geodesic	NoTransform	Grey	NCC	Gewichtet	Benutzt
Art	5	Yoon	NoTransform	RGB	NCC	Gewichtet	Benutzt
Books	15	Yoon	NoTransform	RGB	NCC	Gewichtet	Benutzt
Dolls	15	Geodesic	SoftRank	RGB	SAD	Gewichtet	Benutzt
Laundry	29	Geodesic	Rank	RGB	SAD	Gewichtet	Benutzt
Moebius	29	Geodesic	Rank	111213	SAD	Gewichtet	Benutzt
Reindeer	15	Geodesic	Rank	AC1C2	SAD	Gewichtet	Benutzt

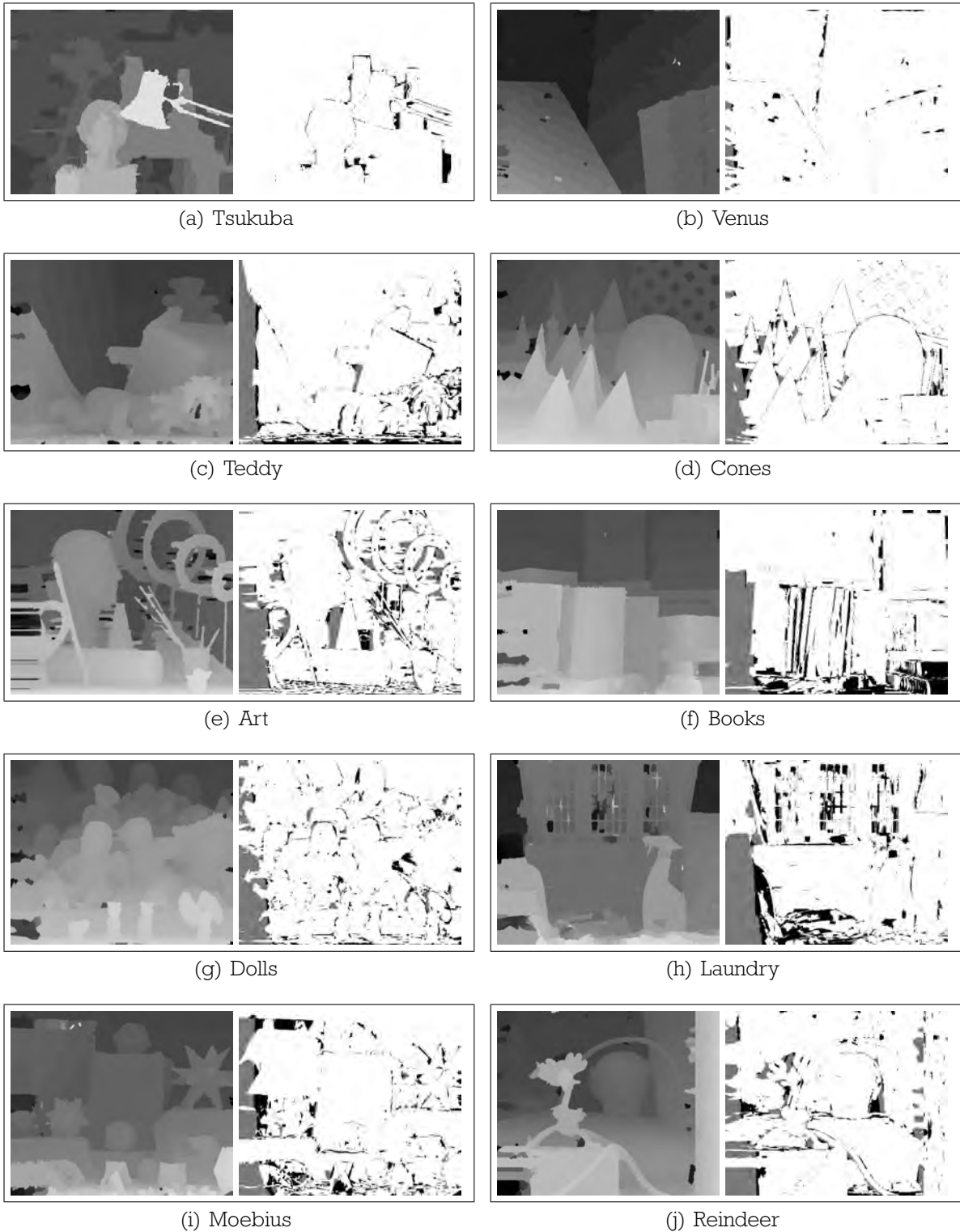


Abbildung 9.21: Bestes Ergebnis je Testbild (links) und Fehlerbild (rechts) bei lokaler Optimierung.

Tabelle 9.17: Bestes Ergebnis der lokalen Optimierung je Testbild. Aufgeteilt in Fehler in den unverdeckten Bereichen in Prozent (Nicht Verdeckt), Fehler in den verdeckten Bereichen in Prozent (Verdeckt) und Gesamtfehler in Prozent (Gesamt).

Testdaten 2003	Nicht Verdeckt (%)	Verdeckt (%)	Gesamt (%)
Tsukuba	2.6	20.7	3.2
Venus	1.6	30.0	2.7
Teddy	7.4	77.2	14.8
Cones	3.0	55.6	9.3
Testdaten 2005			
Art	7.4	58.0	18.9
Books	12.4	73.3	19.6
Dolls	5.1	57.9	12.6
Laundry	13.5	60.3	21.5
Moebius	8.1	56.4	14.5
Reindeer	5.5	48.5	13.2

Tabelle 9.18: Übersicht der verwendeten Parametereinstellungen bei globaler Optimierung für die Erzeugung der Ergebnisbilder in Abbildung 9.22.

Testbild	Parameterset				
	Supporttype	Radio. Transf.	Farbraum	Kostenf.	Gewichte
Tsukuba	Geodesic	SoftRank	RGB	SAD	Benutzt
Venus	Geodesic	SoftRank	Grey	SAD	Benutzt
Teddy	Geodesic	SoftRank	Grey	SAD	Benutzt
Cones	-	NoTransform	Grey	HMI	Unbenutzt
Art	Geodesic	SoftRank	AC1C2	SAD	Benutzt
Books	Geodesic	SoftRank	Grey	SAD	Benutzt
Dolls	Geodesic	Rank	RGB	SAD	Benutzt
Laundry	-	NoTransform	AC1C2	BT	Unbenutzt
Moebius	Geodesic	SoftRank	RGB	SAD	Benutzt
Reindeer	-	NoTransform	AC1C2	BT	Unbenutzt

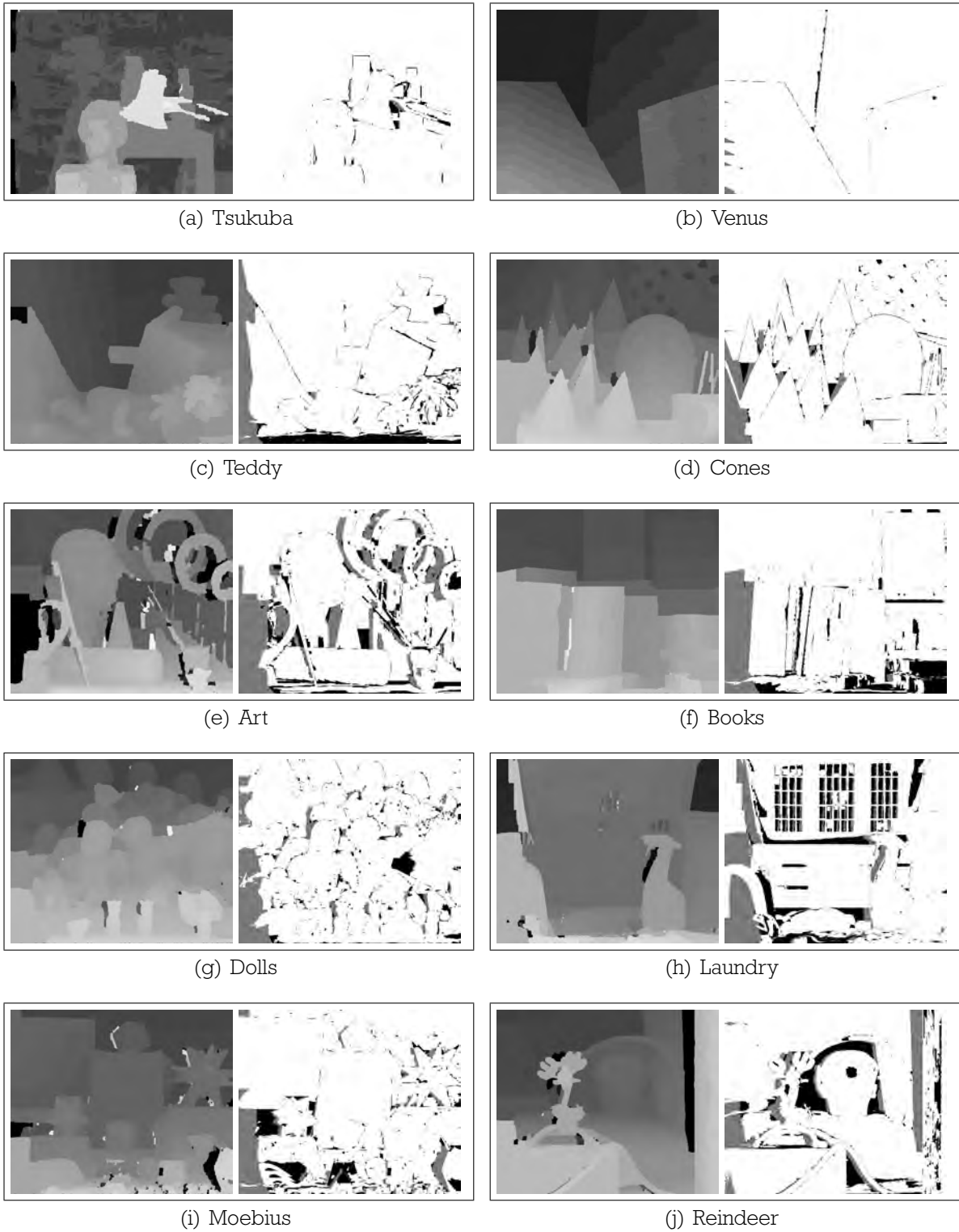


Abbildung 9.22: Bestes Ergebnis je Testbild (links) und Fehlerbild (rechts) bei globaler Optimierung.



Tabelle 9.19: Bestes Ergebnis der globalen Optimierung je Testbild. Aufgeteilt in Fehler in den unverdeckten Bereichen in Prozent (Nicht Verdeckt), Fehler in den verdeckten Bereichen in Prozent (Verdeckt) und Gesamtfehler in Prozent (Gesamt).

Testdaten 2003	Nicht Verdeckt (%)	Verdeckt (%)	Gesamt (%)
Tsukuba	1.6	26.7	2.4
Venus	0.4	18.0	1.1
Teddy	7.3	63.4	13.3
Cones	3.0	54.0	9.2
Testdaten 2005			
Art	10.9	70.8	24.5
Books	9.6	72.6	17.1
Dolls	4.8	62.8	13.1
Laundry	16.8	73.8	26.5
Moebius	7.6	59.2	14.5
Reindeer	10.8	51.4	28.1

# Kapitel 10

## Zusammenfassung

In dieser Arbeit wurde zuerst eine kurze Einführung in Stereo Vision gegeben. Danach wurden, aufbauend auf der Architektur der verwendeten Stereosoftware, die wesentlichen Schritte zur Erzeugung einer Disparitätskarte (Disparity Map) vorgestellt. Im Fall dieser Arbeit sind das:

- Farbraumtransformation
- Radiometrische Transformation
- Kostenberechnung
- Aggregation
- Disparitätsberechnung

Anschließend wurde eine Evaluierung durchgeführt. Die Evaluierung gliederte sich in zwei Schritte. Zuerst wurde ein Trainingsschritt durchgeführt, danach folgte die eigentliche Evaluierung. Im Training wurden mit einer Auswahl von Testbildern die festzuhaltenden Parameter der verschiedenen Ansätze bestimmt. Diese Parameter wurden später in der Evaluierung nicht mehr verändert um eine faire Betrachtung der untersuchten Parameter zu gewährleisten. Die Evaluierung wurde auf zehn ausgewählten Testbildern durchgeführt. Die Testdatensätze stammen aus der Middlebury Groundtruth Datenbank<sup>1</sup>, welche eine Reihe von Testbildern für den wissenschaftlichen Gebrauch zur Verfügung stellt und in vielen Publikationen genutzt wird.

---

<sup>1</sup><http://vision.middlebury.edu/stereo/data/> - Letzter Zugriff: 01.07.2009

Die in der Evaluierung variierten Parameter waren:

- Farbraum
- Radiometrische Transformation
- Kostenberechnungsmethode
- Aggregationsart
- Optimierungsverfahren
- Art der Gewichtsverwendung
- Gewichtstyp
- Fenstergröße (im lokalen Fall)

In Summe wurden knapp 14000 Kombinationen evaluiert. Auf Grund der Verarbeitungsgeschwindigkeit lag der Fokus klar auf den lokalen Methoden. Aus diesem Grund sind auch die Ergebnisse für die lokalen Methoden von geringerer Varianz und höherer Robustheit. In der Evaluierung wurde je Testbildpaar das Ergebnis der Berechnungen - das Disparitätsbild - mit den vorhandenen Groundtruth Daten verglichen und die Fehlerraten in den unverdeckten Bereichen, den verdeckten Bereichen und die Gesamtfehlerrate berechnet. Zum Einen wurde der mittlere Prozentsatz jener Pixel, die einen Fehlergrenzwert von einem Pixel (bei der Klassifizierung) in den nicht verdeckten Regionen der Testdaten übersteigen, berechnet. Zum Anderen wurde für jedes Testbild die jeweilige Position (Rang) innerhalb der Testdaten in Bezug auf die aktuell betrachtete Größe ermittelt.

Bei den lokalen Methoden zeigte sich in den Testfällen, den Farbraum betreffend, ein ähnliches Ergebnis wie in [BCPG08]. Farbverwendung anstelle von Intensitäten verbessert die Ergebnisse. Die besten Ergebnisse wurden mit dem AC1C2 und dem I1I2I3 Farbraum erzielt. Besonders bei den neueren, anspruchsvolleren Testdaten wurde in allen Fällen eine Verbesserung gegenüber dem reinen intensitätsbasierten Ansatz erzielt. Die mittlere Verbesserung der Ergebnisse im Vergleich zwischen dem Farbraum mit den besten Ergebnissen und dem intensitätsbasierten Ansatz beträgt bis zu 17.9 Prozent.

Bei den radiometrischen Transformationen zeigt sich klar, dass im lokalen Fall eine Verwendung zu besseren Ergebnissen führt. Im Test erzielten Rank und SoftRank sowohl

für die Testdaten 2003 als auch für die anspruchsvolleren Testdaten 2005 die besten Ergebnisse. Wie in [HS09, p.17] beschrieben, zeigt sich, dass auch auf Bilddaten, die unter kontrollierten Bedingungen aufgenommen wurden, minimale, nicht offensichtliche, radiometrische Störungen vorhanden sind und der Einsatz von geeigneten radiometrischen Transformationen in jedem Fall sinnvoll erscheint. Im globalen Fall ist ein differenziertes Vorgehen notwendig, da radiometrische Transformationen nicht immer die besten Ergebnisse liefern. In jedem Fall zeigte sich jedoch, dass eine Verwendung von Farbe bei radiometrischen Transformationen bessere Ergebnisse erzielt als rein intensitätsbasiertes Vorgehen.

Bei den untersuchten Kostenberechnungsmethoden ergab sich ein durchwegs gutes Abschneiden von gewichteter normalisierter Kreuzkorrelation (NCC). HMI konnte, im lokalen Fall, die guten Ergebnisse von NCC nicht erreichen. Es zeigte sich jedoch, dass mit einem Ansteigen der Komplexität der Testdaten der Unterschied in der Ergebnisqualität zurückging. Dies deutet darauf hin, dass HMI besonders bei komplexen Bilddaten sehr gute Ergebnisse erzielen kann.

Im lokalen Fall wurde auch der Einfluss der Fenstergröße untersucht. Die mittlere Fenstergröße im Test (15x15 Pixel) erzielte im Durchschnitt die besten Ergebnisse. Weiters beeinflusst die Wahl der Fenstergröße die Performance der Aggregationsmethode. Während quadratische Aggregation mit kleiner Fenstergröße die besten Ergebnisse liefert, erzielt die größte Fenstergröße im Test (29x29 Pixel) für gewichtete Aggregation die besten Ergebnisse. Gewichtete, radiometrische Transformationen und NCC profitieren von einer großen Fenstergröße.

Im Bezug auf den Gewichtstyp konnte festgestellt werden, dass in allen Fällen geodätische Gewichte bessere Ergebnisse als Gewichte nach Yoon liefern. Die Verwendung von Gewichten in Aggregation, radiometrischen Transformation und besonders bei NCC erscheint auf Grund der erzielten Verbesserungen sinnvoll. Ein Problem stellt allerdings die Rechenzeit dar. Anwendungen, die einen Echtzeitanpruch erfüllen müssen, sollten die zusätzliche Rechenzeit in Bezug zur möglichen Verbesserung der Ergebnisse setzen.

Hiermit sind die Betrachtungen den lokalen Ansatz betreffend abgeschlossen, und nachfolgend finden sich Ausführungen zum globalen Ansatz.

Wie im lokalen Fall konnte auch im globalen Fall keine signifikante Verbesserung der Ergebnisse bei Verwendung von Farbe auf den Testdaten 2003 erzielt werden. Der

Analogie folgend konnte jedoch auch im globalen Fall eine Verbesserung der Ergebnisse für die Testdaten 2005 erzielt werden. Wieder konnten der Luminanz-Chrominanz Farbraum AC1C2 und der statistisch-unabhängige Farbraum I1I2I3 besonders gute Ergebnisse liefern. Im Unterschied zum lokalen Fall lieferte RGB auch sehr gute Ergebnisse.

Auch im globalen Fall lieferten die beiden radiometrischen Transformationen Rank und SoftRank sehr gute Ergebnisse. Überraschend war auch das gute Abschneiden von bilateraler Subtraktion, welche ähnlich gute Ergebnisse lieferte wie die anderen beiden Transformationen. Es gilt jedoch zu beachten, dass nicht in allen Fällen eine Verbesserung erzielt werden konnte.

Im globalen Fall konnte sich bei den Kostenberechnungsmethoden HMI klar durchsetzen, gefolgt von BT und AD. HMI als globale Kostenberechnungsmethode scheint besonders in Kombination mit einem globalen Optimierungsansatz sehr gute Ergebnisse zu liefern. Im lokalen Fall liefert HMI zwar ebenfalls gute Ergebnisse, an die sehr guten Ergebnisse der lokalen Methode NCC kann sie jedoch nicht herankommen.

Wie auch schon in [BCPG08] festgestellt, beeinflusst die Wahl des anwendungsspezifischen Farbraumes die Ergebnisqualität. Auf Grundlage dieser Arbeit kann Gleiches für die Wahl der radiometrischen Transformation und der Kostenberechnungsmethode gesagt werden. Bei der Wahl der Gewichtungsmethode sieht diese Arbeit geodätische Gewichte klar im Vorteil.

In dieser Arbeit wurde im lokalen Fall die gewählte Fenstergröße sowohl für die radiometrischen Transformationen und NCC als auch für die anschließende Aggregation verwendet. In einer fortführenden Arbeit sollte untersucht werden, wie sich eine Variation der Fenstergröße auf die Ergebnisse auswirkt. Es scheint logisch, dass zum Beispiel eine kleine Fenstergröße bei der LoG Transformation gefolgt von einer gewichteten Aggregation mit einer großen Fenstergröße bessere Ergebnisse erzielt als die Verwendung der selben Fenstergröße in beiden Fällen. Dies sollte verifiziert werden. Weiters untersucht diese Arbeit als globalen Ansatz nur Graph Cuts. Eine Untersuchung mit einem anderen globalen Ansatz, wie zum Beispiel Dynamischer Programmierung, würde mehr Vergleichsdaten liefern. Ebenfalls wäre eine Untersuchung von anderen Farbräumen (zum Beispiel Lab), sowie die Verwendung weiterer Testbilder, ein Thema für weiterführende Arbeiten.

# Literaturverzeichnis

- [ACM04] A. Ansar, A. Castano, and L. Matthies. Enhanced real-time stereo using bilateral filtering. In *3DPVT04*, pages 455–462, 2004.
- [BBH03] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(8):993–1008, August 2003.
- [BCPG08] M. Bleyer, S. Chambon, U. Poppe, and M. Gelautz. Evaluation of different methods for using colour information in global stereo matching approaches. In *ISPRS Congress*, pages B3a: 415–422, 2008.
- [BF82] S. T. Barnard and M. A. Fischler. Computational stereo. *ACM Comput. Surv.*, 14(4):553–572, 1982.
- [BI99] A. F. Bobick and S. S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, September 1999.
- [Bor86] G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics and Image Processing*, 34:344–371, 1986.
- [Bro92] L.G. Brown. A survey of image registration techniques. *CSURV: Computing Surveys*, 24, 1992.
- [BT98] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:401–406, 1998.
- [BT99] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *ICCV*, pages 489–495, 1999.
- [BVZ98] Y. Boykov, O. Veksler, and R. Zabih. A variable window approach to early vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(12):1283–1294, 1998.

- [CC04] S. Chambon and A. Crouzil. Color stereo matching using correlation measures. In *Complex Systems Intelligence and Modern Technological Applications - CSIMTA 2004*, Cherbourg, 19/09/2004-22/09/2004, pages 520–525. LUSAC, September 2004.
- [CIE86] CIE (Commission Internationale de l’Eclairage). Colorimetry: Technical Report. 1986.
- [DA89] U. R. Dhond and J. K. Aggarwal. Structure from stereo: A review. *IEEE Trans. Systems, Man and Cybernetics*, 19(6):1489–1510, November 1989.
- [Din70] E. A. Dinitz. Algorithm for solution of a problem of maximum flow in networks with power estimation. *Sov. Math. Dokl.*, 11:1277–1280, 1970.
- [FF62] L. R. Ford, jr. and D. R. Fulkerson. *Flows in networks*. Princeton Univ. Press, 1962.
- [FL95] P. V. Fua and Y. G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1):35–56, September 1995.
- [FRT97] A. Fusiello, V. Roberto, and E. Trucco. Efficient stereo with multiple windowing. In *CVPR*, pages 858–863. IEEE Computer Society, 1997.
- [GPS89] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal Royal Statistical Society*, B: 51(2):271–279, 1989.
- [GT88] A.V. Goldberg and R.E. Tarjan. A new approach to the maximum-flow problem. *Journal of the Association for Computing Machinery*. Vol, 35(4):921–940, 1988.
- [HBGR09] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann. Local stereo matching using geodesic support weights. In *IEEE International Conference on Image Processing*, 2009.
- [Hir08] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(2):328–341, February 2008.
- [HS07] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *CVPR*, pages 1–8, 2007.

- [HS09] H. Hirschmuller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(9):1582–1599, September 2009.
- [IG98] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *ECCV*, pages 232–248, 1998.
- [KB01] V. Kolmogorov and Y. Y. Boykov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *EMMVCVPR*, pages 359–374, 2001.
- [KKZ03] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Proceedings of the Ninth IEEE International Conference on Computer Vision, ICCV 2003*, volume 2, pages 1033–1040, Nice, France, October 2003.
- [KZ01] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. In *ICCV*, pages II: 508–515, 2001.
- [LK81] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. DARPA Image Understanding Workshop*, pages 121–130, 1981.
- [MH80] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 207(1167):187–217, 1980.
- [MMHM02] K. Muhlmann, D. Maier, J. Hesser, and R. Manner. Calculating dense disparity maps from color stereo images, an efficient implementation. *International Journal of Computer Vision*, 47(1-3):79–88, April 2002.
- [NVP00] L. Macaire N. Vandembroucke and J.-G. Postaire. Color systems coding for color image processing. In *International Conference on Color in Graphics and Image Processing*, volume 1, pages 180–185, Saint-Etienne, France, 2000.
- [OK91] M. Okutomi and T. Kanade. A stereo matching algorithm with an adaptive window: Theory and experiment. In *CRA*, pages 1088–1095, 1991.
- [OKS80] Y. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer Graphics Image Processing*, 13(3):222–241, July 1980.



- [RC98] S. Roy and I. J. Cox. A maximum-flow formulation of the N-camera stereo correspondence problem. In *ICCV*, pages 492–502, 1998.
- [Sha93] J. Shah. A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. In *CVPR*, pages 34–40, 1993.
- [Shi95] T. Y. Shih. The reversibility of 6 geometric color spaces. *Photogrammetric Engineering and Remote Sensing*, 61(10):1223–1232, October 1995.
- [SK99] S. M. Seitz and K. N. Kutulakos. A theory of shape by space carving. In *ICCV*, pages 307–314, 1999.
- [SP07] D. Scharstein and C. J. Pal. Learning conditional random fields for stereo. In *CVPR*, pages 1–8, 2007.
- [SS96] D. Scharstein and R. S. Szeliski. Stereo matching with non-linear diffusion. In *CVPR*, pages 343–350. IEEE Computer Society, 1996.
- [SS02] D. Scharstein and R. S. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, April 2002.
- [SS03] D. Scharstein and R. S. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, pages I: 195–202, 2003.
- [SSZ02] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. In *ECCV*, page II: 510 ff., 2002.
- [ST97] G. Sharma and H. J. Trussell. Digital color imaging. *IEEE Trans. Image Processing*, 6(7):901–932, July 1997.
- [SZS<sup>+</sup>08] R. S. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, June 2008.
- [Ter86] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:413–424, 1986.

- [TM96] C. Tomasi and R. Manduchi. Stereo without search. In *ECCV*, pages I:452–465, 1996.
- [TM98a] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.
- [TM98b] C. Tomasi and R. Manduchi. Stereo matching as a nearest-neighbor problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(3):333–340, 1998.
- [Vek99] O. Veksler. *Efficient Graph-Based Energy Minimization Methods In Computer Vision*. PhD thesis, April 16 1999.
- [Vek02] O. Veksler. Stereo correspondence with compact windows via minimum ratio cycle. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(12):1654–1660, 2002.
- [Vek03] O. Veksler. Fast variable window for stereo correspondence using integral images. In *CVPR*, pages 556–564. IEEE Computer Society, 2003.
- [VWI95] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. In *ICCV*, pages 16–23, 1995.
- [WKSX04] L. Wang, SB Kang, HY Shum, and G. Xu. Cooperative segmentation and stereo using perspective space search. In *Proc. Asian Conf. Computer Vision*, volume 1, pages 366–371, 2004.
- [YK06] K. J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(4):650–656, April 2006.
- [YYL93] Y. B. Yang, A. Yuille, and J. Lu. Local, global, and multilevel stereo matching. In *CVPR*, pages 274–279, 1993.
- [ZVB99] R. Zabih, O. Veksler, and Y. Y. Boykov. Fast approximate energy minimization via graph cuts. In *ICCV*, pages 377–384, 1999.
- [ZW94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *ECCV*, pages B:151–158, 1994.

# Anhang A

## Bilder

# Anhang A Bilder

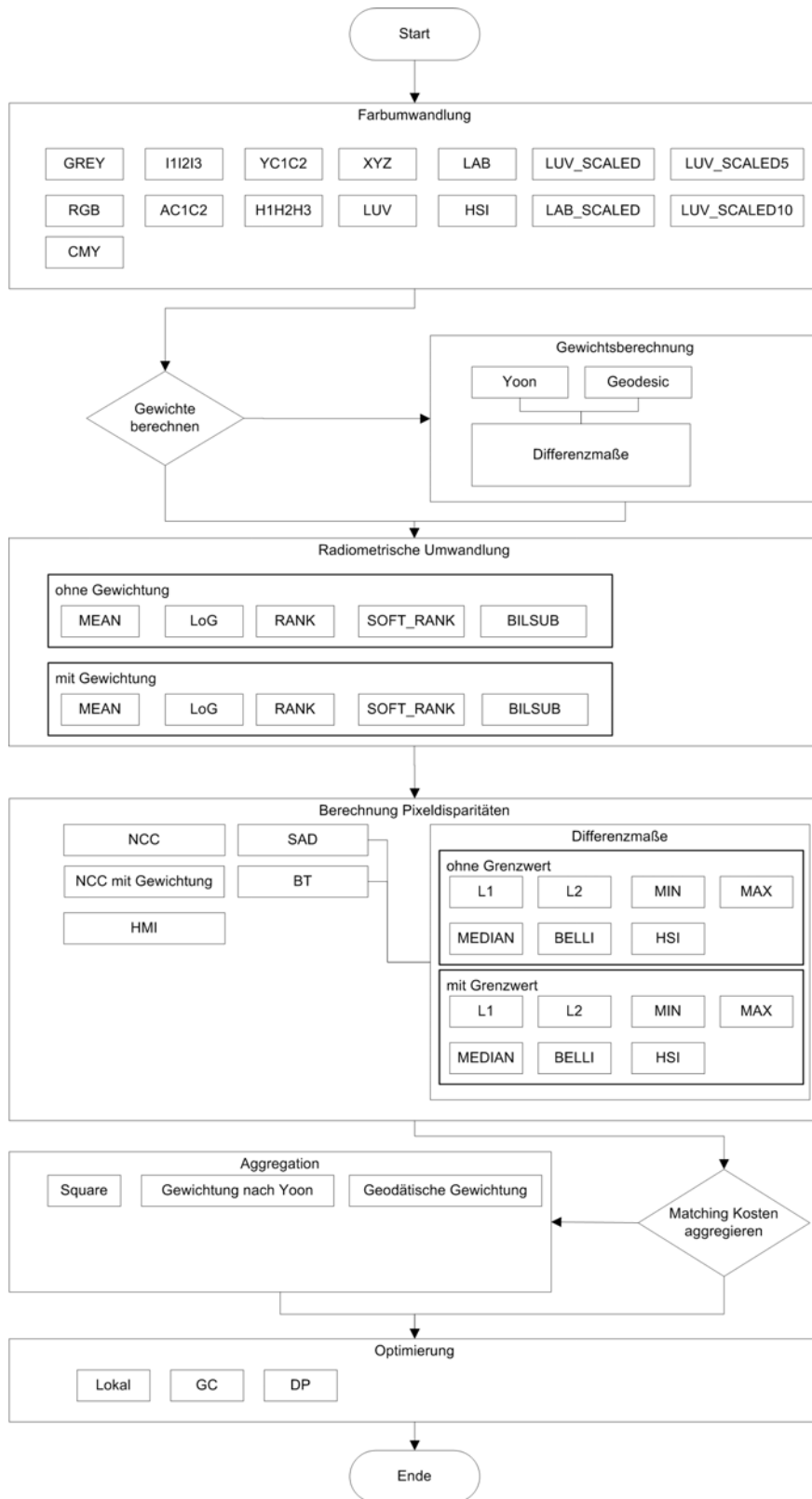


Abbildung A.1: Übersicht der verwendeten Architektur zur Stereoberechnung

# Anhang B

## Abkürzungen und Übersetzungen

**3D** Dreidimensional

**RGB** Rot Grün Blau

**Mean** Mittelwert

**LoG** Laplacian of Gaussian

**Rank** Ranking Transform

**SoftRank** Soft Ranking Transform

**BilSub** Bilateral Subtraction Transform

**SAD** Sum of Absolute Differences

**BT** Birchfield and Tomasi

**NCC** Normalized Cross Correlation

**HMI** Hierarchical Mutual Information

**DP** Dynamic Programming

**GC** Graph Cuts

**DSI** Disparity Space Image

**Mean** Mittelwert

**Bilateral Subtraction Transform** Zweiseitige Subtraktionstransformation

**Disparity Map** Dissparitätskarte

**Ground Truth** Die optimal erreichbare Ergebnisdisparitätskarte

**Support Weight** Gewichtung

**Geodesic** Geodätisch

**Stereo Vision** Stereoskopisches Sehen