

Zur Auswertung lokaler Grauwertstrukturen

Markus Middendorf

Dissertation,
genehmigt von der Fakultät für Informatik
der Universität Fridericiana zu Karlsruhe

Tag der mündlichen Prüfung:	21. Juli 2003
Erster Gutachter:	Prof. Dr. H.-H. Nagel
Zweiter Gutachter:	Prof. Dr. J. Weickert

Zum Geleit

Man kann den Begriff *Optischer Fluss* einführen als die scheinbare Geschwindigkeit der Bewegung einer Grauwertstruktur in der Bildfläche. Bewusst wurde hierbei das Verb ‘einführen’ statt ‘definieren’ verwendet. Dies hängt aber nicht damit zusammen, dass keine Vorschrift zur Schätzung des Optischen Flusses angegeben worden ist.

Die eigentliche Frage muss vielmehr lauten, wie man in der Bildfläche eine ‘Grauwertstruktur’ abgrenzen soll, deren Bewegung den Ausgangspunkt für eine Schätzung des Optischen Flusses darstellen kann. M. Middendorf hat sich dieser Frage zugewendet und ist dabei auf eine überraschende Fülle von Gesichtspunkten gestoßen, die man in diesem Zusammenhang in Betracht zu ziehen hat. Die Schätzung, Segmentation und Interpretation von Optischen-Fluss-Feldern stellt im Rahmen dieser Untersuchung nicht nur ein Ziel an sich dar, sondern dient auch als methodisches Hilfsmittel, um den Einfluss einer Vielzahl von alternativen Vorgehensweisen und Parametern zu studieren. Überrascht wird man bei der Lektüre feststellen, dass manche für selbstverständlich gehaltene Aussagen und Vorgehensweisen unerkannte – oder bisher für irrelevant gehaltene – Probleme aufwerfen. Zugegebenermaßen hängt dies auch damit zusammen, dass die immer noch exponentiell anwachsende Leistungsfähigkeit von Arbeitsplatzrechnern heute Untersuchungen möglich macht, die vor einer Reihe von Jahren schon allein aus Kosten- oder Rechenzeitgründen nur schwer hätten aufgegriffen werden können.

In der Überzeugung, dass sich eine Auseinandersetzung mit den hier dargelegten Untersuchungen lohnt, empfehle ich interessierten Wissenschaftlern, sich die Zeit für eine sorgfältige Lektüre zu nehmen.

Karlsruhe, im Februar 2004

H.-H. Nagel

Dank

Mein Dank gilt an erster Stelle Herrn Prof. Dr. H.-H. Nagel, dessen engagierte Betreuung mit vielseitigen Anregungen meine Arbeit richtungsgebend beeinflusst hat. Unsere Diskussionen waren nicht nur in wissenschaftlicher Hinsicht lehrreich.

Herrn Prof. Dr. J. Weickert danke ich für die freundliche Übernahme des Korreferats trotz des sehr engen zeitlichen Rahmens.

Einen wesentlichen Anteil am erfolgreichen Abschluss meiner Arbeit hatten auch meine Kollegen Michael Arens, Alexander Bachem, Dr. Ralf Gerber, Dr. Ferit Keçeci, Dr. Holger Leuck, Dr. Klaus Mück, Dr. Thomas Müller und Artur Ottlik. Ihnen danke ich für technische Unterstützung, wissenschaftliche Anregungen und praktische Hinweise jeder Art und vor allem für ein hervorragendes Arbeitsklima. Ein besonderer Dank geht auch an Frau Ulla Dietrich, die mit ihrer stets freundlichen Art zur Lösung organisatorischer und verwaltungstechnischer Fragen beitrug.

Herzlich danken möchte ich auch meinen Freunden für ein entspannendes Leben neben der Arbeit, aber auch Vicky Fasen, Matthias Heveling, Sonja Schlauch und Thomas Schmelter für einige mathematische Hinweise sowie Andreas Kunz für das Korrekturlesen, meiner Freundin Dijana Cindrić für ihr Verständnis und ihre Geduld und meinen Eltern für ihren Rückhalt und ein stets offenes Ohr.

Kurzfassung

In den vergangenen Jahren haben Fortschritte der Forschung im Bereich der Bild- und Bildfolgenauswertung dazu beigetragen, dass Anwendungen wie die automatische, videobasierte Erkennung von Staus auf Autobahnen Einsatzreife im Alltagsbetrieb erreicht haben. Komplexere Anwendungen wie die Analyse des Verhaltens von Verkehrsteilnehmern an innerstädtischen Kreuzungen werden noch entwickelt.

Die vorliegende Arbeit beschäftigt sich mit Bewegungsschätzung in Grauwert-Videos und darauf basierender Bildsegmentierung. Beide Aufgaben sind eng miteinander verknüpft und stellen die Grundlage für zahlreiche weiterführende Algorithmen dar. Dabei wird versucht, das Bild durch Auswertung von *lokalen* Grauwertstrukturen, d. h. durch Betrachtung eines im Vergleich zum Gesamtbild kleinen Ausschnitts, in zusammenhängende Bereiche ähnlicher Bewegung zu teilen in der Erwartung, dass jeder dieser Bereiche dem Abbild eines wohldefinierten Körpers in der Szene entspricht.

Dazu baut die Arbeit auf dem „Grauwert-Struktur-Tensor“ (GST) auf, einer gebräuchlichen „Zwischenrepräsentation“, aus der sich Bewegungs- und Segmentierungsinformationen ableiten lassen. Gleichzeitig reduziert der GST durch Mittelung über eine lokale Umgebung rauschbedingte Störungen.

Die Arbeit gliedert sich in drei Teile. Im ersten Teil wird untersucht, inwieweit sich Parametrisierung und Implementierungsdetails der zur GST-Bestimmung nötigen Operationen auf das Ergebnis der GST-Schätzung auswirken. Durch eine präzisere Analyse und Abschätzung der Fehler des bisherigen Vorgehens können dessen Schwächen erkannt und zur Herleitung neuer Verfahren genutzt werden. Die erzielten Verbesserungen werden durch Versuche nachgewiesen. Im zweiten Teil wird ein bisher heuristisch motivierter Ansatz zur GST-Schätzung in einen theoretischen Rahmen eingebettet. Gleichzeitig werden die innerhalb dieses Rahmens noch bestehenden Variationsmöglichkeiten des Algorithmus untersucht und zur gezielten Verbesserung der Ergebnisse im Vergleich zu bestehenden Ansätzen genutzt. Im dritten Teil schließlich wird gezeigt, wie durch eine umfassendere Nutzung der im GST implizit enthaltenen Informationen das Ergebnis von Bewegungsschätzung und Segmentierung weiter verbessert werden kann.

Insgesamt trägt die vorliegende Arbeit dazu bei, das Verständnis der Auswertung *lokaler* Grauwertstrukturen zu verbessern. Die gewonnene Einsicht wird exemplarisch an einigen Fällen zur gezielten Leistungssteigerung bei der Bewegungsschätzung und der Bildsegmentierung genutzt.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Zielsetzung	4
1.3	Aufbau der Arbeit	5
2	Grundlagen der Optischen-Fluss-(OF)-Bestimmung	6
2.1	Grundannahmen der Bildfolgenauswertung	6
2.2	Bemerkung zur Farbe in Bildfolgen	7
2.3	Analyse der Voraussetzungen	9
2.4	Basialgorithmus zur OF-Bestimmung	13
2.4.1	Beschreibung des Algorithmus	13
2.4.2	Ableitungsbestimmung	14
2.4.2.1	Vorbemerkung	14
2.4.2.2	Wahl des Ableitungsoperators	15
2.4.2.3	Umsetzung	18
2.4.3	Vom Gleichungssystem zum Grauwert-Struktur-Tensor (GST)	18
2.4.4	Parameterwahl und Implementierung	21
2.5	Struktur-Tensoren	22
3	Bestimmung der Faltungsmasken	25
3.1	Problembeschreibung	25
3.2	Größe der Faltungsmasken	27
3.2.1	Eindimensionale Masken	27
3.2.2	Dreidimensionale Masken	31
3.2.2.1	Isotrope Masken	31

3.2.2.2	Anisotrope Masken	33
3.3	Bestimmung der Maskenkoeffizienten	35
3.3.1	Bestimmung der vorläufigen Maskenkoeffizienten	36
3.3.2	Korrektur der Maskenkoeffizienten	37
3.3.2.1	Tiefpassmasken	37
3.3.2.2	Ableitungsmasken	39
3.4	Korrespondenzen zw. Tiefpass- und Ableitungsmasken	46
3.4.1	Motivation	46
3.4.2	Ansatz und Umsetzung	46
3.4.3	Bewertung	48
3.4.4	Anmerkungen	50
3.5	Fazit zur Maskenbestimmung	51
4	Versuche zur Verfahrensbewertung	52
4.1	Vergleich bei eindimensionalen Funktionen	52
4.1.1	Tiefpassfilterung	53
4.1.2	Ableitungsbestimmung	56
4.1.3	Versuchsfazit	60
4.2	Fehlerempfindlichkeit der Filterungen	60
4.2.1	Theoretische Analyse der Fehlerfortpflanzung	60
4.2.1.1	Eindimensionale Betrachtung	61
4.2.1.2	Dreidimensionaler Fall	63
4.2.2	Fehlerfortpflanzung bei Faltungsmasken	64
4.2.3	Versuch zur Ermittlung der tatsächlichen Fehler	66
4.2.4	Versuchsfazit	67
4.3	Vergleich der Ergebnisse bei Bildfolgen	67
4.3.1	Korrektheit der „Invarianzebenen“	67
4.3.1.1	Ansatz	67
4.3.1.2	Invarianzebene und Normal-Fluss	69
4.3.1.3	Versuchsdurchführung	69
4.3.1.4	Ergebnisse	71
4.3.2	Korrektheit der GST- und OF-Schätzung	74
4.3.2.1	Ansatz	74

4.3.2.2	Ergebnisse	75
4.3.3	Versuchsfazit	78
5	Skalen-Bestimmung	80
5.1	Motivation	80
5.2	Skalenabhängigkeit der Invarianzebenen	81
5.2.1	Analyse	81
5.2.2	Auswertung 1: Durchschnittsfehler und Median	82
5.2.3	Auswertung 2: Kleinster Fehler der Invarianzebenen	85
5.3	Skalenabhängigkeit des GST	98
5.3.1	Analyse	98
5.3.2	Auswertung 1: Durchschnittsfehler und Median	99
5.3.3	Auswertung 2: Kleinster Fehler des GST	102
5.4	Fazit	104
6	Zu einer adaptiven GST-Bestimmung	105
6.1	Motivation	105
6.2	Exkurs: Referenz-Werte	107
6.3	Verfahrensbeschreibung	109
6.4	Experimente zur iterativen GST-Bestimmung	117
6.4.1	Motivation	117
6.4.2	Qualität der GST-Schätzung	118
6.4.2.1	Qualität der Kategorisierung	118
6.4.2.2	Qualität der OF-Schätzung	122
6.4.3	Konvergenz der OF-Schätzung	125
6.4.3.1	Konvergenz und Oszillation an dominierenden Grau- wertstrukturen	126
6.4.3.2	Konvergenz und Oszillation bei mangelnder Grauwert- struktur	135
6.4.3.3	Fazit zur Konvergenz	139
6.4.4	Robustheit gegenüber Parameterwahl	139
6.4.5	Konvergenzbegriff und Konvergenzkriterium	142
6.4.5.1	Wieder-Überschreiten der Konvergenzschwelle	143
6.4.5.2	Zyklen ohne Überschreiten der Konvergenzschwelle	147

6.4.5.3	Verhalten nach Unterschreiten der Konvergenzschwelle	147
6.4.5.4	Konvergenzkriterium	149
6.4.5.5	Fazit	152
6.5	Weitere Untersuchung zur iterativen GST-Bestimmung	152
6.5.1	Beweis der Fixpunkt-Existenz	152
6.5.2	Praktische Probleme	155
6.5.3	Verbesserung des Konvergenzverhaltens	159
6.5.3.1	Gedämpfte Iteration	160
6.5.3.2	Anpassung der Adaptionfunktion	165
6.5.4	Hinweise auf langsame Konvergenz	166
6.5.5	Anmerkungen zum Konvergenzverhalten	167
6.6	Beurteilung des adaptiv geschätzten OFs	169
6.7	Fazit zur adaptiven GST-Bestimmung	172
7	Vertiefende Untersuchungen	173
7.1	Diskontinuitäten und Diskontinuitätshinweise	173
7.1.1	„Verschobene“ Diskontinuitätsbereiche	174
7.1.2	Unerwartete Diskontinuitäten	180
7.1.3	Löcher in der Diskontinuitätsmauer	182
7.1.4	Integration von Diskontinuitäten in die OF-Schätzung	185
7.1.4.1	Motivation	185
7.1.4.2	Ansatz	187
7.1.5	Ergebnisse	191
7.1.6	Fazit	191
7.2	Aufbau einer Bildpyramide	192
7.2.1	Ansatz	193
7.2.2	Berechnung verkleinerter Bilder	193
7.2.3	Kombination der Ebenen zum Gesamtergebnis	194
7.2.4	Experimentelle Ergebnisse	195
7.2.5	Folgerungen	198
7.3	Anwendungsbeispiele	198
7.4	Fazit	204
8	Zusammenfassung und Ausblick	205

8.1	Zusammenfassung	205
8.2	Ausblick	206
A	Referenzbildfolgen	209
A.1	trans_tree	209
A.2	div_tree	211
A.3	yosemite	211
A.4	tr_bsb	212
A.5	kwbB	214
A.6	stau02	215
A.7	reading	216
A.8	marmor_stat	217
	Literaturverzeichnis	221
	Verzeichnis der Annahmen	233
	Abkürzungsverzeichnis	234
	Symbolverzeichnis	237
	Index	240

Kapitel 1

Einleitung

1.1 Motivation

Betrachten wir als Mensch eine Videofolge, so ist es uns meist ohne Probleme möglich, in den beobachteten Aufnahmen die Abbilder einzelner Objekte zu erkennen, voneinander zu trennen und sowohl deren Bewegung im Bild- als auch die Bewegung der Objekte selbst im Szenenbereich zu beschreiben.¹ Aus Sicht der Informatik findet bei einem solch alltäglichen Vorgang eine Reihe von Auswertungsprozessen statt: Das Auge wandelt das eintreffende Licht punktweise in Nervenimpulse um und führt bereits eine lokale Verarbeitung unter Reduktion der Datenmengen durch. Im Gehirn werden daraus globale visuelle Informationen zusammengesetzt, die als Geschehen in der beobachteten Szene interpretiert werden. Parallel dazu kann – wie auch bei der direkten Betrachtung einer Szene – eine Rückkopplung erfolgen: Die weitere gedankliche Verarbeitung des Geschehens kann beispielsweise dazu führen, dass die Augen auf andere Bereiche des Bildschirms gerichtet werden; eine plötzliche Aufhellung des Bildes kann zu einer Verkleinerung der Pupillen führen.

In der algorithmischen Bildfolgenauswertung wird oft versucht, eine ähnliche Vorgehensweise abzubilden. [Nagel 00] entwirft dafür ein Schichtenmodell (siehe Abbildung 1.1, auch [Nagel 01]), das die Vorgehensschritte in mehrere Ebenen einteilt:

- Die Sensor-Aktuator-Ebene (SAE) umfasst die Gewinnung der Rohdaten mit Hilfe einer Kamera oder – z. B. im Falle von Stereo – eines Systems von mehreren Kameras. Parameter der Kameras (z. B. Fokus oder Blickrichtung) können durch die übergeordneten Ebenen beeinflusst werden.
- Die Roh-Bilddaten aus der SAE werden an die Bild-Signal-Ebene (BSE) weitergereicht. In dieser Ebene werden *lokale* Verarbeitungsschritte – d. h. unter Ver-

¹Unter einem Objekt versteht man in diesem Zusammenhang meist alle wohldefinierten Körper, also z. B. im Diskursbereich „Innerstädtische Straßenkreuzung“ Fahrzeuge, Personen, die Straße selbst, aufgebrachte Markierungen, Straßenschilder, Signalanlagen, Gebäude, etc.

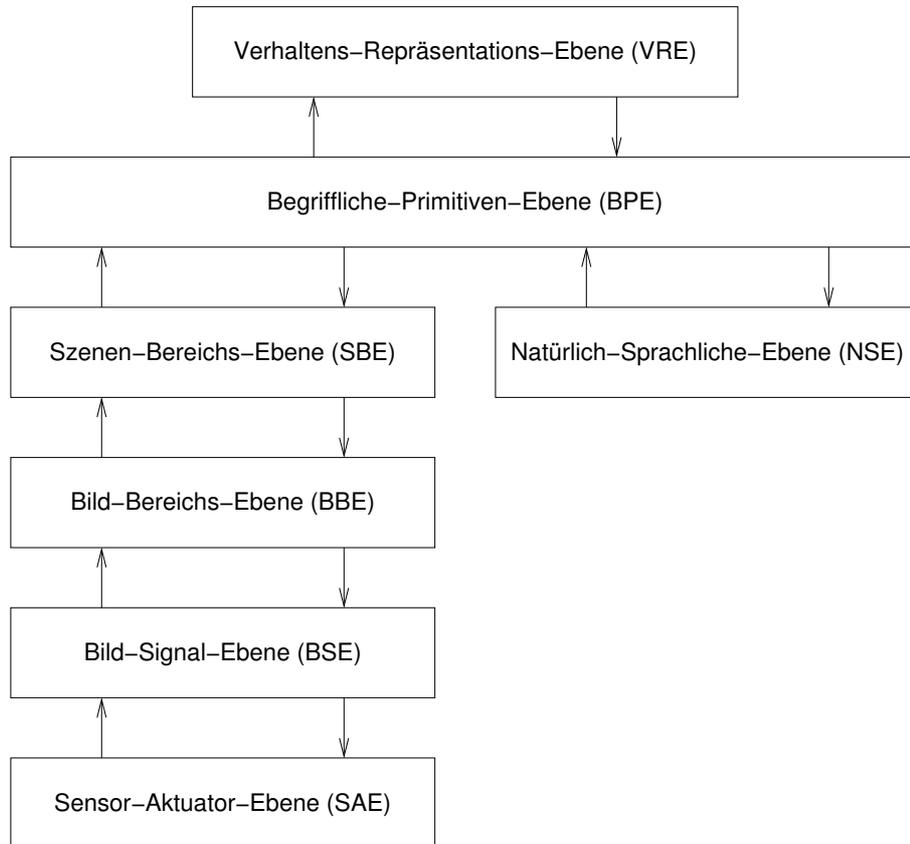


Abbildung 1.1: Skizze der Teil-Struktur eines Kognitiven Systems – hier noch eingeschränkt auf die Bild(folgen)auswertung vom Signal bis zur Erzeugung natürlichsprachlicher Texte (entnommen aus [Nagel 00], deutsche Fassung aus [Nagel 02]). Die Kästen symbolisieren die Schichten, denen die verschiedenen Verarbeitungsprozesse zugeordnet werden, die Pfeile den Informationsaustausch zwischen diesen Schichten.

wendung einer im Vergleich zum Gesamtbild und zur Länge der Bildfolge jeweils nur kleinen räumlich-zeitlichen Umgebung des betrachteten Bildpunkts – durchgeführt wie z.B. Tiefpassfilterung oder Ableitungsbestimmung (vgl. Seite 14). Als Ergebnis erhält man die sogenannten Bildsignalbeschreibungen (BSB). Auch in dieser Ebene können Resultate übergeordneter Ebenen, z. B. Hypothesen über „zusammengehörende“ Bildbereiche, die Einzelheiten des Vorgehens beeinflussen.

- Aus den BSB werden in der Bild-Bereichs-Ebene (BBE) durch Betrachtung eines größeren Bildausschnitts oder des Gesamtbildes komplexere, *globalere* Merkmale, die sogenannten Bild-Bereichs-Hinweise (BBH) abgeleitet. So kann z. B. eine Segmentierung des Bildes vorgenommen werden mit dem Ziel, dass jedes dieser Segmente dem Abbild eines Objektes in der Szene entspricht.
- In höheren Auswertungs- und Abstraktionsebenen werden die so erhaltenen Da-

ten weiterverarbeitet. Es können z. B. aus der Repräsentation im Bild eine Repräsentation in der Szene oder begriffliche Primitiven und daraus natürlichsprachliche Darstellungen abgeleitet werden.

Diese Darstellung eines Kognitiven Systems erhebt nicht den Anspruch auf Vollständigkeit, nicht einmal im Bereich der Bild(folgen)auswertung. Bei der Auswertung können Ebenen wegfallen (z. B. die Szenen-Bereichs-Ebene), und auch bei der Rückkoppelung können Ebenen übersprungen werden. Außerdem können einzelne Ebenen Informationen an Systeme, die nicht dem Bereich Bildauswertung zuzuordnen sind, weitergeben oder von diesen erhalten. In der Praxis können sich Auswertungsprozesse über mehrere Ebenen erstrecken, ohne dass eine genaue Differenzierung möglich ist. Trotzdem ist der in [Nagel 00] gegebene Rahmen hilfreich, um Einzelverfahren im Gesamtkontext „Kognitives Sichtsystem“ einzuordnen.

Im Bereich der Bild- und Bildfolgenauswertung wurden in den vergangenen Jahren erhebliche Fortschritte vor allem bei der „anwendungsnahen“ Forschung erzielt. Einige Anwendungen, z. B. die automatische Erkennung von Staus und Unfällen auf Autobahnen, haben Einsatzreife im Alltagsbetrieb erreicht (vgl. Übersicht in [Martin et al. 01]). Die Europäische Union unterstützt gleich mehrere Konsortien, die basierend auf Videoauswertung Kognitive Sichtsysteme aufbauen wollen ([ECVision]). Diese Arbeiten setzen vor allen in den höheren Schichten (BPE, VRE, NSE) des in Abbildung 1.1 dargestellten Modells an.

Die schon relativ hohen Erfolgsraten der zugrundeliegenden Algorithmen machen aber gleichzeitig eine noch sorgfältigere Analyse notwendig, um die Fälle des Scheiterns solch komplexer Anwendungen erklären zu können. Eine solche Analyse muss schon auf bildnaher Ebene einsetzen, da Fehlerraten, die vor einigen Jahren möglicherweise noch toleriert werden konnten, nun bei anspruchsvolleren Anwendungen zu Problemen auf höheren Auswertungsebenen führen können.

So zeigt beispielsweise [Bachem 03], dass die Auswahl besserer Basisalgorithmen zur Schätzung der Bewegung von Bildstrukturen („Bewegungsschätzung“) das Ergebnis der Kamera-Kalibrierung verbessert, einem Schritt, der bei zahlreichen Anwendungen beim Übergang vom Bild- in den Szenenbereich benötigt wird.

Ein weiteres Beispiel, das die Notwendigkeit solcher Untersuchungen aufzeigt, ist die automatische Generierung von geometrischen Beschreibungen innerstädtischer Kreuzungen ([Mück 00]), die wiederum für die modellgestützte Fahrzeugverfolgung im *Szenenbereich* (vgl. z. B. [Kollnig 95]) benötigt werden. Der Basisschritt besteht dabei in der Erzeugung von Fahrzeugtrajektorien im *Bildbereich* ([Mück et al. 00]) mit dem von [Nagel & Gehrke 98b] vorgeschlagenen und von [Middendorf 99] erweiterten und verbesserten Algorithmus (Abbildung 1.2). Dafür ist die robuste Verfolgung von Fahrzeugabbildern notwendig, möglichst so lange, wie das Fahrzeug im Blickfeld der Kamera ist, also meist über mehrere hundert Bilder. Dazu muss das Fahrzeugabbild in *jedem* oder (bei Verwendung einiger Heuristiken oder einer „stabilisierenden Fortschreibung“ wie z. B. Kalman-Filtern) in fast jedem Bild vom Hintergrund getrennt werden („Bild-

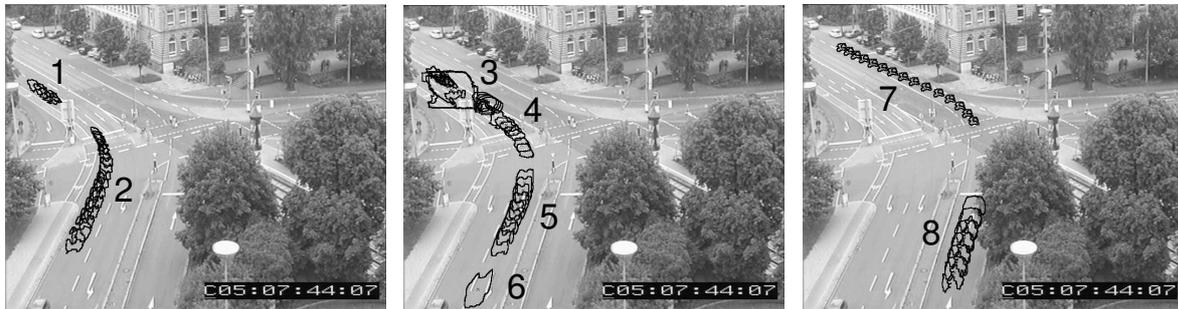


Abbildung 1.2: Trajektorien von acht Fahrzeugabbildern – nach Fahrspuren gruppiert – zur automatischen Fahrspurerkennung. Die detektierten Konturen der Fahrzeugabbilder sind jeweils alle 10 Halbbilder eingezeichnet. Um eine solche Verfolgung des Fahrzeugabbilds im Bildbereich durchzuführen, muss die Abgrenzung zum Hintergrund über 100 oder mehr Bilder zuverlässig erfolgen. Nur bei Fahrzeug Nr. 3 scheitert die Verfolgung des Abbilds aufgrund einer algorithmischen Schwäche – erkennbar an der viel zu weit ausgedehnten Konturlinie –, in den übrigen Fällen sind die Trajektorien wegen Verdeckungen durch einen Baum (Nr. 8) oder ein Verkehrsschild (Nr. 1) unterbrochen, enden mit dem Ende der Bildfolge (Nr. 2, 4, 5, 7) oder wenn das Fahrzeug das Blickfeld der Kamera verlässt (Nr. 6). Entnommen aus [Mück 00], modifiziert.

segmentierung“). Erste Ergebnisse der vorliegenden Arbeit (Bestimmung der Maskengröße, vgl. Kapitel 3) ermöglichten in [Mück 00] im Vergleich zu [Mück et al. 00] eine Ausdehnung der Versuche auf eine weitere Fahrspur (Abbildung 1.2 rechts).

Sowohl die von [Bachem 03] verwendete Bewegungsschätzung als auch die Trennung zwischen Fahrzeugabbildern und Hintergrund erfolgen durch Auswertung einer *lokalen* Grauwertrepräsentation, wie sie in der vorliegenden Arbeit verwendet wird.

1.2 Zielsetzung

In dieser Arbeit soll demonstriert werden, wie durch die Auswertung *lokaler* Grauwertstrukturen eine Repräsentation gewonnen werden kann, die noch möglichst alle für nachgelagerte Auswertungsschritte benötigten Informationen enthält und gleichzeitig störende Einflüsse, z. B. durch Rauschen, reduziert.

Die Qualität einer solchen Repräsentation definiert sich durch die Nutzbarkeit für die nachfolgenden Auswertungsschritte „Bewegungsschätzung“ und „Bildsegmentierung“.

Die in dieser Arbeit dargestellten Untersuchungen erstrecken sich im Wesentlichen auf die Bild-Signal-Ebene, teilweise auch auf die Bild-Bereichs-Ebene. Wie sich zeigen wird, sind für eine korrekte Anwendung aber Annahmen über die Sensor-Aktuator-Ebene nötig. Die dabei zugrundegelegten Annahmen werden herausgearbeitet, die Möglichkeiten und Beschränkungen der eingesetzten Verfahren aufgezeigt.

1.3 Aufbau der Arbeit

Der Hauptteil der vorliegenden Arbeit ist in drei Teile gegliedert. Im ersten Teil – Kapitel 2 bis 5 – wird untersucht, inwieweit sich algorithmische Modifikationen, Implementierungsdetails und Parametrisierung der zur Auswertung von lokalen Grauwertstrukturen nötigen Basisoperationen auf das Ergebnis nachfolgender Auswertungsschritte auswirken. Dazu werden in Kapitel 2 zunächst die zentralen Begriffe „Grauwert-Struktur-Tensor“ (GST), „Optischer Fluss“ (OF) sowie „Skala“ eingeführt und die bei ihrer Bestimmung zugrundegelegten Annahmen herausgearbeitet. In Kapitel 3 werden Möglichkeiten hergeleitet, wie der – oft stillschweigend übergangene – Fehler der dabei benötigten diskreten Faltungen reduziert werden kann. Experimente in Kapitel 4 zeigen, dass diese Überlegungen tatsächlich zu einer Verbesserung der Ergebnisse führen. Kapitel 5 schließlich widmet sich ganz der Untersuchung eines Parameters – der „Skala“ – und zeigt, dass die Wahl der Skala entscheidender für gute Ergebnisse ist als alle zuvor untersuchten Einflussfaktoren.

Im zweiten Teil – Kapitel 6 – wird ein iterativer Ansatz zur Schätzung der „optimalen“ Skala untersucht. Wie sich zeigen wird, ist die iterative Formulierung insofern „wohldefiniert“, als sie in den meisten Fällen ein Ergebnis liefert, das sowohl mathematisch begründbar als auch anschaulich sinnvoll ist.

Im dritten Teil – Kapitel 7 – wird schließlich gezeigt, wie durch eine bessere Nutzung der im GST *implicit* enthaltenen Informationen das Ergebnis weiter verbessert werden kann.

Eine Abgrenzung zu anderen Arbeiten und eine Bewertung der relevanten Literatur findet sich jeweils im Zusammenhang mit der Vorstellung der verwendeten Verfahren, um auch dem Leser, der nur an Einzelaspekten dieser Arbeit interessiert ist, eine zügige Durchsicht der für ihn relevanten Verweise zu ermöglichen. So finden sich in Abschnitt 2.3 Verweise auf OF-Schätzansätze, in Abschnitt 2.4.2.2 Hintergründe zur Skalenraum-Theorie, in Abschnitt 3.1 Überlegungen zur diskreten Faltung und in Abschnitt 7.1.4 Verweise auf Ansätze zur gleichzeitigen OF-Schätzung und Segmentierung.

Es sei an dieser Stelle auch auf den Anhang verwiesen, in dem u. a. die wichtigsten Abkürzungen noch einmal zusammengestellt sind, so dass der Leser diese stets als Referenz zur Verfügung hat.

Kapitel 2

Grundlagen der Optischen-Fluss-Bestimmung

In diesem Kapitel werden die für diese Arbeit zentralen Begriffe „Optischer Fluss“ (OF, Seite 9), „Skala“ (Seite 15) und „Grauwert-Struktur-Tensor“ (GST, Seite 21) eingeführt. Im Rahmen einer ausführlichen Herleitung eines (bekannten) Verfahrens zur OF-Schätzung werden die dabei verwendeten Annahmen herausgearbeitet. In späteren Kapiteln wird sich zeigen, dass viele Fälle, in denen nicht das gewünschte oder erwartete Ergebnis erzielt wurde, auf eine Verletzung dieser Annahmen zurückzuführen sind. Schließlich wird gezeigt, dass der Grauwert-Struktur-Tensor, zunächst nur ein Hilfsmittel zur OF-Schätzung, weitere nutzbare Informationen über die lokale Bildstruktur enthält.

2.1 Grundannahmen der Bildfolgenauswertung

Ziel der Bildauswertung (in Abgrenzung zur reinen Bildverarbeitung) ist es, durch Analyse eines *Bildes* Aussagen über das Abgebildete, die sogenannte *Szene* zu erhalten. Die elementare Grundannahme bei diesem Prozess besteht darin, dass das Bild, also die Verteilung von Helligkeit und Farbe nicht zufällig, sondern durch einen mathematisch beschreibbaren physikalischen Abbildungsprozess in systematischer Art und Weise entstanden ist (oder – bei synthetisch erzeugten Bildern – entstanden sein könnte). Nur diese Annahme rechtfertigt es, aufgrund von Bildprimitiven wie z. B. (Bild-)Kanten auf Szeneneigenschaften wie z. B. Objektgrenzen zurückzuschließen.

Annahme 1: Die verarbeiteten Bilder entstehen in systematischer Art und Weise durch einen mathematisch zumindest annähernd beschreibbaren physikalischen Abbildungsprozess.

Bei der *Bildfolgen*auswertung besteht ein zusätzlicher Informationsgewinn darin, dass von einer systematischen Änderung von Bild zu Bild ausgegangen wird. Diese Annahme darf keinesfalls so missverstanden werden, dass die Abbildung derselben Szene mit derselben Kamera unter denselben Rahmenbedingungen immer zum selben Abbild führt. Vielmehr sind die dabei bekanntermaßen auftretenden Änderungen (Rauschen) Bestandteil des Abbildungsprozesses und müssen (sollten) beim Entwurf von Auswertungsprozessen berücksichtigt werden. Darüber hinaus ist es in der Praxis selbst im Laborumfeld nicht immer einfach, wirklich konstante Rahmenbedingungen zu garantieren.

Im Vergleich z. B. zur Stereo-Bildauswertung, bei der ein systematischer Zusammenhang zwischen gleichzeitig aufgenommenen Bildern vorausgesetzt wird, besteht eine weitere Annahme der *Bildfolgen*auswertung darin, dass durch die schnelle zeitliche Abfolge der Bilder nur eine begrenzte Änderung der Szene (der abgebildeten Objekte, der Beleuchtungsbedingungen, ...) und der Kameraparameter (Kameraposition, Kameraorientierung, ...) von Bild zu Bild möglich ist. Wie groß die Änderung sein darf, so dass noch eine „sinnvolle“ *Bildfolgen*auswertung möglich ist, hängt davon ab, welche Aussagen über die Szene gewünscht sind und mit welchen Algorithmen diese Aussagen gewonnen werden.

Annahme 2: Von Bild zu Bild erfolgen nur kleine, systematische Änderungen der Szene und der Kameraparameter.

Im Rahmen der folgenden Untersuchungen soll zunächst davon ausgegangen werden, dass die vorangehend skizzierten Annahmen erfüllt sind, wobei in Annahme 2 offen bleibt, was eine „kleine“ Änderung ist. Schon [Nagel 81] beschäftigt sich mit dieser Frage, und je nach Kontext mag die Antwort unterschiedlich ausfallen. Probleme, die durch die offensichtliche Verletzung dieser Annahmen entstehen, sollen nicht betrachtet werden. Dazu zählt beispielsweise, wenn sich durch einen Szenenwechsel („Schnitt“) das Bild von einem Aufnahmezeitpunkt zum nächsten in großen Teilen oder völlig ändert (siehe z. B. [Venegas-Andraca 02] für einen Überblick aktueller Literatur).

2.2 Bemerkung zur Farbe in Bildfolgen

Im Rahmen der in dieser Arbeit beschriebenen Versuche wurde auf einer Vielzahl von Bildfolgen gearbeitet, deren wesentliche, als versuchsrelevant betrachtete Merkmale im Anhang A beschrieben werden. Die Aspekte, die die Wahl einer bestimmten Bildfolge bei der Durchführung eines Versuches bedingen, werden bei den einzelnen Versuchen genannt.

Allen Versuchen gemeinsam ist jedoch, dass grundsätzlich nur die Helligkeitsinformation der Bilder genutzt wird. Liegen Bilder in einem Farbformat vor, so werden sie vor der Durchführung von Berechnungen in Grauwertbilder umgewandelt. Ist im Folgenden

von Grauwertbildern die Rede, so sind damit auch stets die *umgewandelten* Farbbilder gemeint.

Alle vorgestellten und eingesetzten Algorithmen lassen sich auch auf andere aus dem Bild punktweise abgeleitete, skalare Merkmale anwenden, solange diese Merkmale den Annahmen 1 und 2 genügen. Ein Beispiel wäre bei Farbbildern, die durch die drei Farbkanäle Rot, Grün und Blau repräsentiert sind (RGB-Bild), die Verwendung nur eines der drei Kanäle oder deren Minimum oder Maximum ([Schäfer 00], vgl. dazu Anmerkungen in [Fermüller et al. 01]). Die Helligkeit hat sich dabei als ein Merkmal erwiesen, das einerseits möglichst viel der im Bild vorhandenen Information nutzt und andererseits die oft relativ unzuverlässige Farbinformation ausschließt. Untersuchungen von [Schäfer 00] haben gezeigt, dass die Berechnung des Optischen Flusses auf RGB-Farbbildern mit paralleler Nutzung aller drei Farbkanäle keine wesentlichen Zusatzinformationen liefert, sondern in ungünstigen Fällen sogar weitere Probleme verursacht.

Für die Umwandlung eines RGB-Bildes in Graustufen gibt es mindestens zwei gebräuchliche Vorgehensweisen:

- Gängige Programme wie z. B. `xv` oder `ppmtopgm` verwenden dazu den Y-Kanal, also die „Luminanz“, des YIQ-Farbsystems:

$$Y = 0,299 r + 0,587 g + 0,114 b \quad , \quad (2.1)$$

wobei r , g und b für den Rot-, Grün- und Blauanteil des Farbbildes stehen. Dieses Farbsystem wurde in den USA zur Kodierung des Farbfernseh-Signals gemäß NTSC („National Television System Committee“) eingeführt (siehe [Pritchard 77; Foley et al. 90]). Eine solches Vorgehen orientiert sich an der menschlichen Wahrnehmung und entspricht der Farbe-nach-Graustufen-Umsetzung bei Schwarz-Weiß-Fernsehern (mit NTSC-Norm). Dabei wird aber unterstellt, dass der Grün-Kanal mehr Informationen enthält als die übrigen Kanäle.

- Ein „moderneres“ Farbsystem, das z. B. der JPEG-Komprimierung zugrunde liegt (vgl. [Russ 95]), ist das HSI-Farbsystem („Hue-Saturation-Intensity“). Die Umrechnung von RGB nach Grauwert erfolgt dabei gemäß der Formel:

$$I = \frac{r + g + b}{3} \quad , \quad (2.2)$$

wobei I („Intensity“) für den Helligkeitswert steht. Dieses Vorgehen entspricht der Annahme, dass alle Farbkanäle gleiches Gewicht haben.

Da einige am IAKS verwendete Bilddaten im JPEG-Format vorliegen, verwenden wir Formel (2.2) für die Umwandlung von RGB- in Grauwert-Bilder. Dafür spricht auch, dass einige als Referenzversuche dienende Experimente (z. B. [Middendorf 99]) ebenfalls diese Formel verwendeten. In der vorliegenden Arbeit werden die Begriffe (Bild-)Helligkeit, (Bild-)Intensität und Grauwert äquivalent verwendet.

2.3 Analyse der Voraussetzungen

Bilder entstehen aus der Szene durch einen Abbildungsprozess, der – bei Vernachlässigung von Diskretisierung¹ und Quantisierung² – eine stetige Funktion darstellt, gegeben durch das Abbildungsmodell³ und die Kameraparameter. Aus den in Annahme 2 geforderten kleinen Änderungen von Szene und Kameraparametern von einem Aufnahmezeitpunkt zum nächsten folgt somit, dass auch die Änderungen von einem Bild zum nächsten klein sind (im Vergleich zu den Änderungen, die sich zwischen beliebigen Bildern ergeben können).

Daraus ergibt sich die Folgerung, dass es für die meisten Bereiche in einem Bild in den vorangehenden und nachfolgenden Bildern korrespondierende Bereiche gibt, d. h. Bildausschnitte, die dasselbe Szenenobjekt abbilden oder die ähnlich aussehen. Eine Charakterisierung der Bildfolge erfolgt nun über die Bewegung zwischen solchen korrespondierenden Bereichen. Dabei werden zwei Fälle unterschieden:

- Die Bewegungsgeschwindigkeit der *Projektion eines Szenenpunktes* im Bild nennt man *Verschiebungsrate (VR)*.
- Die *scheinbare* Bewegungsgeschwindigkeit von *Strukturen im Bild* hingegen bezeichnet man als *Optischen Fluss (OF)*.

In der Definition des OF wird bewusst von *Strukturen* im Bild und nicht von Bildpunkten gesprochen, da es offensichtlich nicht sinnvoll ist, *einzelne Bildpunkte*, die beispielsweise dieselbe Helligkeit besitzen, als korrespondierend zu betrachten, da es i. Allg. in jedem Bild Hunderte von Bildpunkten mit derselben Helligkeit gibt. Ist im Folgenden also zur leichteren Lesbarkeit manchmal trotzdem von der Bewegung von Bildpunkten die Rede, so ist damit stets die sich aus der Einbettung in die lokale Bildstruktur ergebende Bewegung gemeint.

In der vorliegenden Arbeit wird die Verschiebungsrate durchgängig mit dem zweidimensionalen Vektor $\mathbf{v} = (v_1, v_2)^T$ bezeichnet, der Optische Fluss mit dem Vektor $\mathbf{u} = (u_1, u_2)^T$. Die Funktion $\mathbf{u}(\mathbf{x})$ nennt man *OF-Feld*. Die Komponenten der VR- bzw. OF-Vektoren geben die tatsächliche bzw. scheinbare Bewegung in x - bzw. y -Richtung pro Zeitschritt, d. h. von einem Aufnahmezeitpunkt zum nächsten an. Bei zeilenversetzten Bildern soll ein Zeitschritt dem Abstand zwischen zwei *Halbbildern* entsprechen. Ist im Folgenden von der Verschiebung *pro Bild* die Rede, so meint dies bei zeilenversetzten Bildern stets implizit die Verschiebung *pro Halbbild* (HB).

¹Diskretisierung: Abtastung der kontinuierlichen Intensitätsverteilung des auf die Wandlerfläche der Kamera einfallenden Lichts an festen Rasterpunkten oder – präziser – stückweise Integration der Intensitätsverteilung in abgegrenzten Bereichen der Wandlerfläche, die jeweils einem Rasterpunkt zugeordnet werden.

²Quantisierung: Abbildung der einem Rasterpunkt zugeordneten (kontinuierlichen) Energiemenge auf die zur Verfügung stehende Skala der unterschiedlichen Grauwerte; oft Ganzzahlen von 0 bis 255.

³Bei der Modellierung des Abbildungsverhaltens von Standard-Videokameras wird meist idealisiert von einem Lochkameramodell ausgegangen.

OF und VR müssen nicht übereinstimmen. Beispielsweise kann eine kleine Drehung einer spiegelnden Fläche eine große Verschiebung eines Glanzlichtes und somit einen großen OF verursachen, während die VR sehr klein ist. Andererseits kann bei einer rotierenden Billardkugel der OF Null sein, während die VR ungleich Null ist. Mit Mitteln der Bildfolgenauswertung kann stets nur der OF bestimmt werden. Um mit Hilfe des OFs brauchbare Rückschlüsse auf das Geschehen in der Szene ziehen zu können, trifft man

Annahme 3: Bei sorgfältigem Vorgehen ist der geschätzte Optische Fluss eine gute Näherung an die tatsächliche Verschiebungsrate.

Manche Autoren, z. B. [Mester & Hötter 95], verwenden daher auch beide Begriffe gleichwertig. Um bei Experimenten mit bekannter VR diese deutlich vom geschätzten OF abgrenzen zu können, soll in dieser Arbeit die begriffliche Unterscheidung aber beibehalten werden.

Im weiteren Verlauf dieser Arbeit soll herausgearbeitet werden, worin ein sorgfältiges Vorgehen im Sinne dieser Annahme bestehen sollte. Ziel ist es, eine OF-Schätzung zu entwickeln, die an möglichst vielen Bildpunkten mit der VR weitgehend übereinstimmt. Ferner soll auf algorithmischem Wege erkannt werden, wenn an bestimmten Bildpunkten die OF-Schätzung nur unzuverlässig, unvollständig oder gar nicht möglich ist. Außerdem wird eine Repräsentation der lokalen Grauwertstruktur angestrebt, die eine Segmentierung des Bildes aufgrund der Bewegungsinformation unterstützt.

Zur besseren Übersicht werden OF-Schätz-Ansätze oft in Gruppen eingeteilt. Die derzeit wohl bekannteste Übersichtsarbeit stammt von [Barron et al. 94]. Die Autoren geben einen Überblick über die wichtigsten Ansätze zur OF-Schätzung bis etwa 1992 und schlagen eine Unterteilung der Ansätze in gradienten-basierte (differentielle) Ansätze, regionen-basierte Ansätze („Region-matching“, Korrelation), energie-basierte Ansätze und phasen-basierte Ansätze vor. Einige Autoren fassen energie- und phasen-basierte Ansätze zusammen [Fermüller et al. 01], teilweise unter dem Schlagwort frequenz-basierte Ansätze, oder führen neue Gruppen wie z. B. sogenannte modell-basierte Ansätze ein [Beghdadi et al. 03].

Eine andere Klassifizierung teilt die Ansätze in solche, die an jedem Punkt einen OF-Vektor schätzen („dichte“ OF-Felder), und solche, die nicht an jedem Punkt eine OF-Schätzung liefern („spärlich besetzte“ OF-Felder).

Eine solche Einteilung ist hilfreich, um die grobe Vorgehensweise eines (neuen) Ansatzes einfach zu vermitteln. Allerdings ist die Zuordnung eines Ansatzes zu einer Gruppe nicht unbedingt eindeutig. [Beghdadi et al. 03] nennen einen Ansatz modell-basiert, wenn eine systematische Variation des OF-Feldes bei dessen Bestimmung modelliert wird. Allerdings treffen alle Ansätze (zumindest implizit) Annahmen über die Variation der Bewegung und damit des OF-Feldes (siehe [Bergen et al. 92, Abs. 1.2]). [Thompson 98] verwendet einen Ansatz, der durch (implizite) Interpolation zwischen Punkten

mit zuverlässiger OF-Schätzung dichte Vektorfelder erzeugt. Mit einer solchen Interpolation könnte aber aus jedem spärlich besetzten ein dichtes OF-Feld erzeugt werden.

Aus diesem Grund soll der verwendete Ansatz nicht (nur) durch Zuordnung zu der einen oder anderen Gruppe definiert werden, sondern es erfolgt eine Abgrenzung aufgrund der dem Ansatz zugrunde liegenden Annahmen. Für eine neuere Übersicht existierender OF-Schätz-Ansätze sei der interessierte Leser auf [Haußecker & Spies 99] verwiesen.

Im Folgenden sei eine Bildfolge – eine zeitliche Folge von Grauwertbildern – als diskrete Funktion der Grauwerte $g = g(x, y, t)$ in Abhängigkeit von Pixelposition (x, y) und Zeitpunkt t repräsentiert⁴, und es sei $\mathbf{x} = (x, y, t)^T$.

Unter der

Annahme 4: Die Helligkeit des Abbilds eines Szenenpunkts bleibt in aufeinander folgenden Aufnahmen gleich.

ergibt sich für kleine Zeiträume δt folgender Ansatz zur Schätzung des OF:

$$g(x_0, y_0, t_0) = g(x_0 + \delta x(x_0, y_0, t_0, \delta t), y_0 + \delta y(x_0, y_0, t_0, \delta t), t_0 + \delta t), \quad (2.3)$$

wobei $(\delta x, \delta y)$ die unbekannte Bewegung der Grauwertstruktur in Abhängigkeit von räumlich-zeitlicher Ursprungsposition und δt bezeichnet. Diese Annahme ist die Grundlage der meisten OF-Schätz-Ansätze, stellt aber gleichzeitig eine wesentliche Einschränkung der OF-Schätzung dar. Vor allem bei realen Bildfolgen wird diese Annahme oft durch sich ändernde Beleuchtungsverhältnisse oder Schattenwurf verletzt. Oft wird versucht, den dadurch verursachten Fehler mit statistischen Methoden zu kompensieren. So gehen bspw. [Hager & Belhumeur 98] davon aus, dass die Helligkeitsänderung das gesamte Bild – oder zumindest die gesamte zu verfolgende Region – in gleicher Art und Weise betrifft und dass eine hinreichende Anzahl möglicher „Beleuchtungssituationen“, d. h. verschiedener Beleuchtungsrichtungen, in einer Trainings-Phase „gelernt“ werden kann. Trotz der bekannten Probleme werden die Folgen der Verletzung von Annahme 4 in der vorliegenden Arbeit nicht weiter untersucht, da die Berücksichtigung von Beleuchtungsunterschieden schon bei – im Vergleich zur OF-Schätzung – einfacheren Aufgaben wie der Änderungsdetektion bei stationärer Kamera ([Xie et al. 02]) oft nur zu – wie die Autoren selber feststellen – unbefriedigenden Ergebnissen führt.

Trifft man zusätzlich noch die

Annahme 5: Die Bewegung eines Bildpunktes ist in kleinen Zeitintervallen gleichförmig.

⁴Ab und zu werden mit g auch nur ein Bild $g = g(x, y)$ oder ein eindimensionales Grauwertsignal $g = g(x)$ bezeichnet. Durch die Anzahl der Argumente wird aber stets verdeutlicht, welcher Fall gemeint ist.

so lässt sich Gleichung (2.3) unter Verwendung des OF-Vektors \mathbf{u} vereinfachen zu:

$$g(x_0, y_0, t_0) = g(x_0 + u_1(x_0, y_0, t_0)\delta t, y_0 + u_2(x_0, y_0, t_0)\delta t, t_0 + \delta t). \quad (2.4)$$

Die Bestimmung von \mathbf{u} kann durch direkten Vergleich der Grauwerte in unmittelbar aufeinander folgenden Bildern erfolgen (Korrelationsverfahren, z. B. [Anandan 89]). Dieses Vorgehen ist vor allem dann vorteilhaft, wenn die Verschiebung sehr groß ist ([Yacoob & Davis 99]), wenn man Annahme 5 nur für zwei oder wenig mehr Bilder für erfüllt hält oder wenn nur zwei Bilder für die Berechnung zur Verfügung stehen.

Stehen mehr als zwei Bilder zur Verfügung – und das ist bei der Bildfolgenauswertung i. Allg. der Fall – so bietet sich ein *differentielles* Verfahren an, das auch mehrere Bilder einbezieht, um so die Empfindlichkeit gegenüber Rauschen zu vermindern. Durch Taylor-Entwicklung der rechten Seite von Gleichung (2.4) erhält man:

$$\begin{aligned} g(\mathbf{x}_0) &= g(\mathbf{x}_0) + \delta t \left(u_1 \frac{\partial g}{\partial x}(\mathbf{x}_0) + u_2 \frac{\partial g}{\partial y}(\mathbf{x}_0) + \frac{\partial g}{\partial t}(\mathbf{x}_0) \right) \\ &+ \frac{1}{2} \delta t^2 \left(u_1^2 \frac{\partial^2 g}{\partial x^2}(\mathbf{x}_0) + u_2^2 \frac{\partial^2 g}{\partial y^2}(\mathbf{x}_0) + \frac{\partial^2 g}{\partial t^2}(\mathbf{x}_0) \right) \\ &+ \delta t^2 \left(u_1 u_2 \frac{\partial^2 g}{\partial x \partial y}(\mathbf{x}_0) + u_1 \frac{\partial^2 g}{\partial x \partial t}(\mathbf{x}_0) + u_2 \frac{\partial^2 g}{\partial y \partial t}(\mathbf{x}_0) \right) \\ &+ \dots \end{aligned} \quad (2.5)$$

Vernachlässigt man die Terme höherer als erster Ordnung, setzt $\delta t = 1$ und definiert den erweiterten OF-Vektor $\mathbf{u}_+ = (u_1, u_2, 1)^T$ sowie abkürzend

$$g_x = g_x(\mathbf{x}) = \frac{\partial g}{\partial x}(\mathbf{x}), \quad g_y = g_y(\mathbf{x}) = \frac{\partial g}{\partial y}(\mathbf{x}), \quad g_t = g_t(\mathbf{x}) = \frac{\partial g}{\partial t}(\mathbf{x}), \quad (2.6)$$

$$\nabla g = (g_x, g_y, g_t)^T, \quad (2.7)$$

so erhält man die sogenannte OF-Beschränkungsgleichung („Optical Flow Constraint Equation“ – OFCE) [Horn & Schunck 81]:

$$(g_x, g_y, g_t) \mathbf{u}_+ = (\nabla g)^T \mathbf{u}_+ = 0. \quad (2.8)$$

Damit die Taylor-Entwicklung und die Vernachlässigung von Termen höherer als erster Ordnung überhaupt zulässig sind, muss gelten:

Annahme 6: Die Bildfunktion g ist differenzierbar.

Annahme 7: Die Bildfunktion ändert sich in der lokalen Umgebung jedes Punktes näherungsweise linear.

Betrachtet man die Bildfunktion als eine Treppenfunktion mit konstantem Wert in der $1 \times 1 \times 1$ Pixel großen Umgebung eines jeden Abtastpunktes und mit Sprungstellen zwischen benachbarten Abtastpunkten, so ist die Bildfunktion nie differenzierbar (außer in Bereichen konstanter Grauwerte). Annahme 6 fordert demnach, dass die Funktion der Helligkeitsverteilung auf der Wandlerfläche der Kamera *vor* der Abtastung differenzierbar ist.

Gleichung (2.8) ist offensichtlich unterbestimmt. Um eine Schätzung für \mathbf{u}_+ und somit \mathbf{u} zu erhalten, sind also weitere Annahmen erforderlich. Üblicherweise benutzt man

Annahme 8: Alle Punkte in einer hinreichend großen räumlich-zeitlichen Umgebung des jeweils betrachteten Punktes (Aufpunkt) bewegen sich im Wesentlichen genauso wie der Aufpunkt.

Durch Aufstellen von Gleichung (2.8) für mehrere Punkte in der räumlich-zeitlichen Umgebung des jeweils betrachteten Punktes erhält man ein überbestimmtes Gleichungssystem, das sich mit Standardverfahren nach \mathbf{u} lösen lässt. Man kann dabei die lokale Änderung von \mathbf{u} explizit modellieren (z. B. [Campani & Verri 92; Farnebäck 01]), implizit zulassen (z. B. bei der *globalen* Glattheitsforderung von [Horn & Schunck 81]) oder bewusst ausschließen ([Lucas & Kanade 81; Uras et al. 88]). Beim hier gewählten Ansatz wird von lokal konstantem \mathbf{u} ausgegangen.

Offensichtlich besteht ein Zusammenhang zwischen der Erfüllbarkeit der Annahmen 6, 7 und 8. So ist eine nicht-differenzierbare Bildfunktion i. Allg. auch nicht „näherungsweise linear“, und auch wenn sich Bildbereiche unterschiedlich bewegen, wird die Linearitätsannahme meist verletzt. Eine kurze Diskussion weitergehender Annahmen findet sich in [Barron et al. 94].

2.4 Basisalgorithmus zur OF-Bestimmung

2.4.1 Beschreibung des Algorithmus

Aufgrund dieser Vorüberlegungen ergibt sich für jeden Punkt \mathbf{x}_0 folgendes Verfahren zur Bestimmung des OFs:

1. Bestimmung der notwendigen ersten partiellen Ableitungen g_x , g_y und g_t der Grauwertfunktion $g(\mathbf{x})$ in einer lokalen räumlich-zeitlichen Umgebung $\mathcal{U}_M(\mathbf{x}_0)$ von \mathbf{x}_0 (diskutiert in Abschnitt 2.4.2).

2. Aufstellung eines überbestimmten Gleichungssystems auf Basis der OF-Beschränkungsgleichungen aller Punkte $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{U}_M(\mathbf{x}_0)$:

$$\begin{aligned} (\nabla g(\mathbf{x}_1))^T \mathbf{u}_+ &= 0 \\ &\vdots \\ (\nabla g(\mathbf{x}_n))^T \mathbf{u}_+ &= 0 \end{aligned} \tag{2.9}$$

und Lösung nach \mathbf{u}_+ (vgl. Abschnitt 2.4.3).

2.4.2 Ableitungsbestimmung

2.4.2.1 Vorbemerkung

Schon fast vom Beginn der maschinellen Bildauswertung an stellte sich die Aufgabe, Bilder von Rauschen zu befreien [Nathan 66], oder Bildmerkmale wie z. B. Kanten(elemente) zu extrahieren [Roberts 65]. Beide Aufgaben sind miteinander verwandt: Im ersten Fall möchte man vermeiden, dass ein Grauwert *nur zufällig* stark von den Grauwerten der Umgebung abweicht, im zweiten Fall möchte man gerade die Grauwerte finden, die *systematisch* stark von Grauwerten der Umgebung abweichen. Überlegungen aus der Theorie der Signalverarbeitung zeigen, dass auch die zur Lösung dieser Aufgaben geeigneten Operationen ähnlich sind.

Dazu betrachtet man eine Darstellung des Bildes im sogenannten *Orts-Frequenzraum*. In dieser Darstellung wird ein Bild durch die auftretenden Orts-Frequenzen (vereinfacht: den auftretenden Änderungen) repräsentiert. Große Grauwertunterschiede zwischen benachbarten Pixeln führen zu hohen, geringe Grauwertunterschiede zu niedrigen Orts-Frequenzen. Im Gegensatz zum Orts-Frequenzraum spricht man bei der üblichen Bilddarstellung vom *Ortsraum* (und bei Bildfolgen vom *Orts-/Zeitraum*).

Möchte man im Orts-Frequenzraum das Rauschen entfernen, so wendet man einen *Filter* an, der nur niedrige Frequenzen passieren lässt („Tiefpassfilter“) und hohe Frequenzen löscht; das Bild wird geglättet. Möchte man starke Änderungen erkennen, so wendet man einen Filter an, der nur hohe Frequenzen passieren lässt („Hochpassfilter“). Zur Ableitungsbestimmung in Grauwertbildern verknüpft man oft beide Filter miteinander: Man löscht niedrige, aber auch zu hohe Frequenzen („Bandpassfilter“).

Offensichtlich sind solche Filterungen nicht nur im Orts-Frequenzraum möglich. Führt man beispielsweise eine lokale Mittelung der Grauwerte durch, so können keine beliebig großen Ableitungen mehr auftreten; die Mittelung ist also ein Tiefpassfilter.

Im Frequenzraum wird eine solche Filterung durch eine entsprechende *diskrete Faltung* einer endlich ausgedehnten Faltungsmaske – speziell einer *Tiefpass-* bzw. *Glättungsmaske* oder einer *Ableitungsmaske* – mit dem (untransformierten) Bild realisiert. Eine

solche diskrete Faltung berechnet sich im Eindimensionalen nach:

$$(g \otimes m)(x_0) = \sum_{n=-N}^{+N} g(x_0-n) \cdot m(n) , \quad (2.10)$$

analog im Zwei- oder Dreidimensionalen. Dabei bezeichnen g die Grauwert-Funktion, m die im Intervall $[-N, N]$ definierte Maskenfunktion und \otimes den Faltungsoperator. Die Abbildung eines Bildes auf seine geglättete Version wird auch als *Glättungsoperator* bezeichnet; analog definiert man *Ableitungsoperator*.

2.4.2.2 Wahl des Ableitungsoperators

Wie bereits bei Annahme 6 angemerkt, ist die Bildfunktion g im mathematischen Sinne nicht differenzierbar. Man nähert daher die Differentiation durch eine Differenzbildung an. Oft verknüpft man die Ableitungsbestimmung mit einer impliziten Glättung, durch die die Differenzierbarkeit sichergestellt ist.

Frühe Versionen von Faltungsmasken wie beim Roberts-Operator [Roberts 65], Sobel-Operator [Sobel 70] oder Kirsch-Operator [Kirsch 71] waren von den begrenzten technischen Möglichkeiten bestimmt, und wiesen dadurch bedingte Mängel auf. Durch die oft auf 3×3 oder 5×5 Pixel beschränkte Maskengröße waren die mit solchen Operatoren ermittelten Ergebnisse rauschanfällig oder stark von der Orientierung von Kanten im Bild beeinflusst.

Ein weiterer wesentlicher Nachteil dieser frühen Ansätze bestand darin, dass für alle Anwendungsfälle stets dieselbe Maskengröße und dieselbe Gewichtung der von den Masken erfassten Punkte verwendet wurde. [Rosenfeld & Thurston 71] verwenden daher Operatoren verschiedener Größe und versuchen, die optimale Operator-Größe (aus einer vorgegebenen Menge) zu ermitteln. [Klinger 71] schlägt eine *Bildrepräsentation* mit Ebenen verschiedener räumlicher Auflösung vor, d. h. Ebenen, die durch die Anwendung von Operatoren unterschiedlicher Größe und unterschiedlicher Gewichtung der erfassten Punkte – und somit unterschiedlicher „Glättungswirkung“ – erzeugt wurden.

[Marr & Hildreth 80] verwendeten daraufhin für die *Bildauswertung* Masken unterschiedlicher (fest vorgegebener) Größen mit unterschiedlicher Glättungswirkung. [Witkin 83] und [Koenderink 84] formalisieren diese Idee und schlagen die Einbettung eines Bildes in einen ein-parametrischen Raum, den *Skalenraum* vor, der durch Faltung des Bildes mit Gaußfunktionen verschiedener Varianz σ^2 entsteht. Der (kontinuierliche) Parameter $s = \sigma^2$ wird als *Skalierungsparameter* oder kurz *Skala* bezeichnet⁵ und gibt die Auflösung an, die sich in der Glättungswirkung der Faltung niederschlägt. Ausgehend von dieser Definition handelt es sich bei der Arbeit von [Marr & Hildreth 80] um einen sogenannten Mehr-Skalen-Ansatz. Diese Ideen werden daher manchmal auch als

⁵Eine Motivation dieser Bezeichnung findet sich in [Babaud et al. 86].

Marr-Hildreth-Koenderink-Witkin-Theorie bezeichnet. [Weickert et al. 99] verweisen später darauf, dass die Grundideen bereits 20 Jahre zuvor von [Iijima 62] auf Japanisch veröffentlicht wurden.

Losgelöst von der Gaußfunktion definieren [Alvarez et al. 93] einen Skalenraum als eine Menge von Bildern $u(s, x, y)$, die durch Anwendung eines glättenden *Skalenoperators* T_s mit Skala s aus dem Ursprungsbild $g(x, y)$ entsteht – $u(s, x, y) = (T_s \circ g)(x, y)$ – wobei die Glättungswirkung mit wachsendem s zunehmen und der Operator bei Skala 0 der Identitätsabbildung entsprechen muss, also $u(0, x, y) = g(x, y)$.

[Koenderink 84] leitet die Verwendung der Gaußfunktion aus der Tatsache ab, dass bei Faltung des Bildes mit einer isotropen Gaußfunktion der Skalenraum genau der Lösung der Wärmeleitungsgleichung bzw. der linearen homogenen Diffusionsgleichung

$$\frac{\partial u(t, x, y)}{\partial t} = \Delta u(t, x, y), \quad u(0, x, y) = g(x, y) \quad (2.11)$$

entspricht. Dabei bezeichnen g die anfängliche Temperatur bzw. Stoffkonzentration (analog zum ungeglätteten Bild) und u die aktuelle Temperatur bzw. Stoffkonzentration (analog zum geglätteten Bild) zum Zeitpunkt t (analog zur Skala). Δ bezeichnet den Laplace-Operator, d. h. $\Delta u = u_{xx} + u_{yy}$. Man nennt den so erzeugten Skalenraum auch Gauß'schen Skalenraum oder linearen (Diffusions-)Skalenraum. [Alvarez et al. 93] zeigen schließlich, dass sich unter vernünftigen Annahmen alle Skalenräume als Lösung einer partiellen Differentialgleichung („partial differential equation“ – PDE) zweiter Ordnung ausdrücken lassen⁶:

$$\frac{\partial u(t, x, y)}{\partial t} = F(D^2 u, Du, t), \quad u(0, x, y) = g(x, y) . \quad (2.12)$$

Ist diese PDE linear, so spricht man von einem linearen Skalenraum mit einem linearen Skalenoperator, ansonsten von einem nicht-linearen Skalenraum mit einem nicht-linearen Operator. Um Verwechslungen mit dem bei Bildfolgen auftretenden Zeitparameter t zu vermeiden, bezeichnen wir die Skala abweichend von der nach Gleichung (2.11) und (2.12) naheliegenden Bezeichnung mit s oder geben sie direkt durch σ^2 bzw. im Mehrdimensionalen durch die Kovarianzmatrix Σ an⁷.

Die naheliegende Frage, ob es außer der Gaußfunktion auch andere ein-parametrische Funktionen gibt, die als Skalenoperator in Frage kommen, greifen [Babaud et al. 86] auf. Sie analysieren die Anforderungen der Skalenraum-Theorie und zeigen im *eindimensionalen* Fall, dass bei Beschränkung der möglichen Glättungsfunktionen auf die sogenannten schnell abnehmenden Funktionen [Rudin 74] die Gaußfunktion die einzige ist, die einige allgemein als vernünftig anerkannte Grundforderungen erfüllt. Die

⁶ $F(D^2 u, Du, t)$ ist die Kurzschreibweise für eine beliebige Funktion in Abhängigkeit von t sowie ersten und zweiten partiellen Ableitungen von u , so dass noch eine zulässige PDE entsteht.

⁷Man bezeichnet bei Verwendung der Gaußfunktion σ^2 und nicht σ als Skala, weil durch diese Festlegung die Skalenoperatoren eine Halbgruppe bilden, d. h. $T_s(T_t(g)) = T_{s+t}(g)$ für $s, t \geq 0$. Einige Autoren verwenden $\frac{1}{2}\sigma^2$, da dieser Vorfaktor bei der Lösung von Gleichung (2.11) „auftaucht“. Zur besseren Lesbarkeit wird in dieser Arbeit nicht mit diesem Faktor gearbeitet.

wichtigste dieser Forderungen ist, dass bei wachsender Skala und somit zunehmender Glättung keine Extrema an Stellen auftreten dürfen, an denen bei kleinerer Skala kein Extremum vorlag.

[Alvarez et al. 93] stellen einen Katalog solcher Grundforderungen/Axiome auf, charakterisieren die bisherigen Ansätze durch die von ihnen erfüllten Axiome und zeigen, dass es bei anderen Voraussetzungen als von [Koenderink 84] oder [Babaud et al. 86] gefordert – bei Verzicht auf die Forderung nach Linearität des Operators – auch möglich ist, zu einem Skalenraum im oben definierten Sinne zu gelangen, der nicht durch Faltungen mit der Gaußfunktion erzeugt werden kann.

[Felsberg & Sommer 01] stellen schließlich einen linearen Skalenraum zur Diskussion, der nicht auf der Wärmeleitungsgleichung (2.11) beruht, sondern auf der 3D-Laplace-Gleichung

$$\Delta_3 p(x, y, s) = 0, \text{ mit } \Delta_3 = \partial_x^2 + \partial_y^2 + \partial_s^2, \quad (2.13)$$

weisen aber selbst darauf hin, dass sie damit dem Eindeutigkeitsbeweis von [Iijima 62] widersprechen.

Wie die bisherige Darstellung gezeigt hat, gibt es bei der Erzeugung der geglätteten Version eines Bildes – und damit der Ableitung durch anschließende Differenzen-Bildung – zwei grundsätzliche Vorgehensweisen: Faltung mit einer Glättungsfunktion („Faltungsansatz“) oder Lösung einer (partiellen) Differentialgleichung („Variationsansatz“). Es wurde gezeigt, dass die Faltung mit einer *isotropen* Gaußfunktion und das Lösen einer *linearen homogenen* Diffusionsgleichung äquivalent sind.

Verallgemeinert man beide Ansätze, d. h. faltet man mit *punktweise adaptierten* Gaußfunktionen oder verwendet man eine *nicht-lineare* PDE, so ist die Äquivalenz nicht mehr gegeben. Faltungsansätze und Variationsansätze unterscheiden sich also nicht nur in der Formulierung.

Variationsansätze bieten eine knappe und elegante Formulierung der Bildglättung. Durch Modifikationen der Diffusionsgleichung können sich unterschiedlich verhaltende PDEs erzeugt werden (siehe z. B. Übersicht in [Weickert & Brox 02]). Auf Grund der iterativen, adaptiven Berechnung können sehr komplexe Zusammenhänge zwischen der Bildstruktur und dem Verlauf der Glättung modelliert werden, was allerdings das intuitive Verständnis dafür erschwert, wie die Lösung zustande kam. Eine Visualisierung des Diffusionsprozesses wie von [Dam & Nielsen 01] vorgeschlagen kann aber das Verständnis verbessern.

Bei Faltungsansätzen hingegen entsteht das Ergebnis in der Regel durch nur einen einzelnen Faltungsschritt mit meist zum Maskenmittelpunkt symmetrischen Masken. Dies führt im Vergleich zum Variationsansatz zu einer Einschränkung der Möglichkeiten. Vorteile des Faltungsansatzes bestehen in einer stärkeren Lokalität – der in das Ergebnis an einem Punkt eingehende Bereich kann ohne Durchführung der Faltung selbst angegeben werden – und einem damit verbundenen intuitiven Zugang zur „Wirkungsweise“.

Aufgrund der scheinbar überwiegenden Vorteile der Variationsansätze haben sich diese in den vergangenen Jahren zu einem Standardverfahren zur kanten- oder struktur-erhaltenden Glättung und der damit verbundenen Ableitungsbestimmung entwickelt. Dennoch wird in der Folge ein *Faltungsansatz* zur Bestimmung der notwendigen partiellen Ableitungen verwendet. Für diese Entscheidung waren drei Punkte ausschlaggebend:

- Es soll untersucht werden, ob nicht auch mit dem theoretisch eingeschränkteren Ansatz zumindest vergleichbare Ergebnisse erzielt werden können wie mit Variationsansätzen.
- Die stärkere Lokalität des Faltungsansatzes bietet die Möglichkeit für eine einfache Parallelisierung des Ansatzes auf Rechnernetzen, wobei aufgrund der Bandbreite des verfügbaren Netzwerkes der Datenaustausch minimiert werden soll.⁸
- Bekanntermaßen sind Differentiation und Faltung mit der Gaußfunktion vertauschbar. Die partiellen Ableitungen können also in einem Schritt durch Faltung mit den partiellen Ableitungen der Gaußfunktion bestimmt werden.

2.4.2.3 Umsetzung

Die Ableitungsbestimmung auf *Bildern* im Zweidimensionalen durch Faltung mit den partiellen Ableitungen der zweidimensionalen Gaußfunktion kann problemlos für die Ableitungsbestimmung auf *Bildfolgen* verallgemeinert werden: Die – hier verwendete – Methode zur Bestimmung der notwendigen partiellen Ableitungen von $g(x, y, t)$ besteht also in der Faltung von $g(x, y, t)$ mit den ersten partiellen Ableitungen einer dreidimensionalen Gaußfunktion – aufgrund der Form im Ein- und Zweidimensionalen auch Gaußglocke genannt – $G_{3D}(\mathbf{x})$:

$$\nabla g(\mathbf{x}) = (\nabla G_{3D} \otimes g)(\mathbf{x}), \quad (2.14)$$

mit

$$G_{3D}(\mathbf{x}) = \frac{1}{(2\pi)^{3/2} \sqrt{|\Sigma_A|}} \cdot e^{-\frac{1}{2} \mathbf{x}^T \Sigma_A^{-1} \mathbf{x}}. \quad (2.15)$$

Die Kovarianzmatrix Σ_A bestimmt dabei die Form der Gaußglocke, also die Skala, woraus sich die Gewichtung der einzelnen Pixel und somit die Glättungswirkung ergibt.

2.4.3 Vom Gleichungssystem zum GST

Das überbestimmte Gleichungssystem (2.9) besitzt im Allgemeinen keine exakte Lösung, die alle Gleichungen erfüllt. Daher ist ein Vektor \mathbf{u}_+ zu finden, so dass alle

⁸Erst kurz vor Beendigung der hier vorgestellten Untersuchungen wurde auch eine effiziente, auf Rechnernetzen parallelisierte Implementierung eines nicht-linearen Diffusionsansatzes vorgestellt [Bruhn et al. 02].

Gleichungen *möglichst gut* erfüllt werden. Dieses Problem tritt in vielen Bereichen der Ingenieur- und Naturwissenschaften auf, und es sind mehrere Lösungsverfahren bekannt. Um das „optimale“ Verfahren auszuwählen, sind jedoch einige weitere Überlegungen (siehe auch ausführlich in [Weber & Malik 93], mathematische Grundlagen in [Van Huffel & Vandewalle 91]) notwendig:

- Zum einen muss verstanden werden, wie die Koeffizienten des Gleichungssystems zustande kamen und warum die Gleichung nicht exakt erfüllt ist. Beim betrachteten Verfahren sind die Koeffizienten des Gleichungssystems Ableitungswerte, die durch gewichtete Addition und Subtraktion von Grauwerten entstehen (siehe Kapitel 3). Diese Grauwerte sind – gegenüber einer kontinuierlichen Funktion – durch Diskretisierung und Quantisierung verfälscht und mit einem Rauschen versehen, was sich in den Ableitungen niederschlägt. Die GST-Bestimmung in der vorliegenden Form basiert auf

Annahme 9: Alle Grauwerte sind mit einem additiven, normalverteilten Fehler (Rauschen) mit Mittelwert 0 und positions-, helligkeits- sowie zeitunabhängiger Varianz $\hat{\sigma}^2$ behaftet, dessen Verteilungen voneinander unabhängig sind.

und der darauf aufbauenden

Annahme 10: Die Ableitungsschätzung ist mit einem normalverteilten Rauschen behaftet, das für alle partiellen ersten Ableitungen an allen Bildpunkten derselben Verteilung unterliegt und dessen Verteilungen voneinander unabhängig sind.

Annahme 9 stellt eine grobe Vereinfachung der tatsächlichen Verhältnisse dar, da Effekte wie die Quantisierung der Grauwerte oder die Tatsache, dass ein „tatsächlicher Grauwert“ von 0 durch Rauschen nicht kleiner werden kann, dabei unberücksichtigt bleiben. Außerdem gibt es beispielsweise Bildpunkte, die aufgrund von Fertigungsfehlern der Kamerawandlerfläche stets heller oder dunkler erscheinen, als es der eingefallenen Lichtmenge entspricht. [Betke & Makris 01] untersuchen jedoch für eine von ihnen verwendete CCD⁹-Kamera das statistische Verhalten der Grauwerte und stellen fest, dass die Annahme eines additiven, normalverteilten Fehlers im Wesentlichen erfüllt ist, dass die Verteilungen beliebiger Pixel voneinander unabhängig sind und dieselbe Varianz besitzen. Ausführliche Diskussionen der verschiedenen Fehlerfaktoren finden sich in [Tsin et al. 01] oder in [Healey & Kondepudy 94].

Auch die in Annahme 10 postulierte Unabhängigkeit der Verteilungen der Ableitungsschätzungen ist im Allgemeinen *nicht* erfüllt.

⁹„Charge Coupled Device“

- Zum anderen muss festgelegt werden, was unter einer „möglichst guten“ Erfüllung des Gleichungssystems verstanden werden soll. Unter der Annahme, dass es keine Ausreißer gibt, liegt es nahe, eine Lösung zu suchen, die bei möglichst kleiner Korrektur der Koeffizienten der Gleichungen das Gleichungssystem exakt erfüllt.

Nach [Van Huffel & Vandewalle 91] wird in einem solchen Fall die beste Lösung mit dem Verfahren der kleinsten Summe der absoluten Fehlerquadrate („Total Least Squares“ – TLS) erzielt, was der Lösung des Minimierungsproblems:

$$\min_{\mathbf{u}_+} \left(\sum_{\mathbf{x}_i \in \mathcal{U}_M(\mathbf{x}_0)} |(\nabla g(\mathbf{x}_i))^T \mathbf{u}_+|^2 \right) \quad (2.16)$$

unter der Nebenbedingung (u. d. N.) $u_{+,3} = 1$ entspricht¹⁰.

Dieses Standard-TLS-Verfahren führt jedoch zu einem scharfen „Abschneiden“ am Rand des Bereichs $\mathcal{U}_M(\mathbf{x}_0)$. Wünschenswert ist vielmehr ein „sanfter“ Übergang zwischen Punkten in $\mathcal{U}_M(\mathbf{x}_0)$ und Punkten außerhalb von $\mathcal{U}_M(\mathbf{x}_0)$. Dies wird durch Multiplikation aller Gleichungen mit einem Gewichtungsfaktor $\sqrt{w(\mathbf{x}_i)}$ erreicht, mit großem $w(\mathbf{x}_i)$ nahe \mathbf{x}_0 und abfallendem $w(\mathbf{x}_i)$ zum Rand hin. Es bietet sich an, $w(\mathbf{x}_i)$ durch Abtastung einer dreidimensionalen Gaußfunktion mit Kovarianzmatrix Σ_M zu bestimmen, also:

$$w(\mathbf{x}_i) = \frac{1}{(2\pi)^{3/2} \sqrt{|\Sigma_M|}} \cdot e^{-\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_i)^T \Sigma_M^{-1} (\mathbf{x}_0 - \mathbf{x}_i)}. \quad (2.17)$$

Die Beschränkung der Summation auf $\mathcal{U}_M(\mathbf{x}_0)$ kann dabei aufgegeben werden; eine Beschränkung des Summationsbereiches ergibt sich nun in der Praxis „von alleine“ durch den raschen Abfall der Gaußfunktion auf sehr kleine Werte bei wachsendem $\|\mathbf{x}_i - \mathbf{x}_0\|$.

Das neue Minimierungsproblem lautet also:

$$\min_{\mathbf{u}_+} \left(\sum_{\mathbf{x}_i \in \mathbb{Z}^3} w(\mathbf{x}_i) |(\nabla g(\mathbf{x}_i))^T \mathbf{u}_+|^2 \right) \quad \text{u. d. N. } u_{+,3} = 1. \quad (2.18)$$

Durch diese Skalierung wird auf den ersten Blick die Annahme des TLS-Ansatzes verletzt, dass alle Gleichungskoeffizienten dieselbe Fehler-Verteilung besitzen, da der Fehler zum Rand hin gedämpft wird. Wenn man allerdings bedenkt, dass Annahmen 7 und 8 oft nur näherungsweise erfüllt werden, so ist es durchaus gerechtfertigt, bei weiter von \mathbf{x}_0 entfernt liegenden Pixeln implizit einen größeren Fehler der geschätzten partiellen Ableitungen in Hinblick auf die Modellannahmen zuzulassen.

¹⁰Lässt man hingegen Ausreißer zu, d. h. geht man davon aus, dass eine – unbekannte – Teilmenge der Gleichungen fälschlicherweise in das Gleichungssystem aufgenommen wurde, so kann z. B. der Ansatz des kleinsten Median der Fehlerquadrate („Least Median of Squares“ – LMedS) bessere Ergebnisse liefern (siehe [Bab-Hadiashar & Suter 98]).

Nach [Weber & Malik 93] und [Van Huffel & Vandewalle 91] ergibt sich eine robuste Lösung \mathbf{u}_+ von (2.18) durch die Bestimmung des Eigenvektors \mathbf{e}_{\min} zum kleinsten Eigenwert λ_{\min} der Matrix:

$$\overline{(\nabla g)(\nabla g)^T} = \sum_{\mathbf{x}_i \in \mathbb{Z}^3} w(\mathbf{x}_i) \nabla g(\mathbf{x}_i) (\nabla g(\mathbf{x}_i))^T \quad (2.19)$$

$$= \sum_{\mathbf{x}_i \in \mathbb{Z}^3} \frac{\nabla g(\mathbf{x}_i) (\nabla g(\mathbf{x}_i))^T}{(2\pi)^{3/2} \sqrt{|\Sigma_M|}} \cdot e^{-\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_i)^T \Sigma_M^{-1} (\mathbf{x}_0 - \mathbf{x}_i)} \quad (2.20)$$

$$\approx \int_{-\infty}^{+\infty} d\xi \frac{\nabla g(\xi) (\nabla g(\xi))^T}{(2\pi)^{3/2} \sqrt{|\Sigma_M|}} \cdot e^{-\frac{1}{2}(\mathbf{x}_0 - \xi)^T \Sigma_M^{-1} (\mathbf{x}_0 - \xi)}, \quad (2.21)$$

und Skalierung von \mathbf{e}_{\min} :

$$\mathbf{u}_+ = \left(\frac{\mathbf{e}_{\min,1}}{\mathbf{e}_{\min,3}}, \frac{\mathbf{e}_{\min,2}}{\mathbf{e}_{\min,3}}, 1 \right)^T. \quad (2.22)$$

Das Ergebnis der Mittelung in Gleichung (2.20) wird als *Grauwert-Struktur-Tensor* (GST), gelegentlich auch als *Grauwert-Ableitungstensor* bezeichnet.

Es sei noch angemerkt, dass eine eindeutige Lösung des Minimierungsproblems (2.18) nur dann existiert, wenn sich die Bildfunktion *nicht* in der gesamten berücksichtigten Umgebung des Punktes \mathbf{x}_0 *exakt* linear verhält, da sonst jeder Punkt dieselbe Gleichung liefert. In Verbindung mit Annahme 7 ließe sich also fordern: Die Bildfunktion verlaufe in einer kleinen Umgebung näherungsweise linear, in einer etwas größeren Umgebung aber nicht „zu linear“. Es liegt daher nahe zu fordern, dass die Mittelungskovarianzmatrix Σ_M mindestens so groß ist wie die Ableitungskovarianzmatrix Σ_A , in der Hoffnung, dass sich dann die Gradienten an „entgegengesetzten Enden“ des Mittelungsbereichs gerade noch genug voneinander unterscheiden. [Nagel & Gehrke 98a; Nagel & Gehrke 98b] folgen einem heuristischen Vorschlag von [Lindeberg 95], der $\Sigma_M = 2\Sigma_A$ setzt. Aufgrund der vorangehenden Überlegung erscheint dieser Wert plausibel und soll auch in dieser Arbeit stets beibehalten werden.

2.4.4 Parameterwahl und Implementierung

Offensichtlich wird das Ergebnis der GST-Bestimmung sowohl durch Implementierungsdetails der notwendigen diskreten Faltungen als auch durch die Wahl der Kovarianzmatrizen Σ_A und Σ_M der Gaußfunktionen beeinflusst.

In Kapitel 3 und 4 finden sich zunächst ausführliche Untersuchungen zur Implementierung der Faltung und zur Wahl der damit verbundenen – auf den ersten Blick „unsichtbaren“ – Parameter wie der Maskengröße oder der Maskenbestimmungsmethode. Einflüsse auf das Ergebnis durch die Rechnerarithmetik (Rundungsverhalten, Stellig-

keit) und die verwendeten Datentypen sollen dabei aber vernachlässigt werden, da solche Faktoren erfahrungsgemäß nur einen sehr geringen Einfluss haben.¹¹

Zur Beurteilung der Implementierung wird dazu in Kapitel 4 der GST mit fest vorgegebenem Σ_A und Σ_M bestimmt.

Ab Kapitel 5 wird dann der Einfluss der Skalenparameter Σ_A und Σ_M untersucht. [Nagel & Gehrke 98a; Nagel & Gehrke 98b] führen eine „adaptive“ Bestimmung von Σ_A und Σ_M ein, um von der anfänglichen, heuristischen Wahl dieser Parameter unabhängiger zu werden. Dazu geben sie in einem Start-Schritt Σ_A und Σ_M fest vor, unabhängig von dem Punkt, für den der GST bestimmt wird, bestimmen den GST und nutzen das Ergebnis, um Σ_A und Σ_M für einen folgenden Adaptionsschritt festzulegen. Dieser Ansatz wird in mehreren Schritten analysiert, ausgeweitet und verbessert.

2.5 Struktur-Tensoren

Struktur-Tensoren treten bei verschiedenen Aufgaben der Bild- und Bildfolgenauswertung auf. [Bigün & Granlund 87] benutzen als erste einen *zweidimensionalen* Tensor, um die „Orientierung“ einer lokalen Umgebung zu beschreiben. [Nitzberg & Shiota 92] nutzen einen solchen *zweidimensionalen* Tensor, um die Glättung eines Bildes an die lokale Grauwertstruktur zu adaptieren. [Haußecker & Jähne 97] führen die Bestimmung des OFs auf das Finden von Bereichen konstanter Helligkeit in Bildfolgen zurück und leiten so eine allgemeinere Form des in Gleichung (2.20) gegebenen Tensors her. [Nagel & Gehrke 98a; Nagel & Gehrke 98b] verfeinern schließlich die räumlich-zeitliche *dreidimensionale* Tensor-Bestimmung durch ein adaptives Vorgehen, das die Ergebnisse weniger stark von der A-priori-Festlegung der Parameter Σ_A und Σ_M abhängig macht. Beispiele für aktuelle Anwendungen von Struktur-Tensoren liefern [Hladůvka & Gröllner 01] (Interpolation von Volumen-Daten; die *z*-Dimension kann im Wesentlichen wie die Zeit-Dimension behandelt werden), [Bakker 02] (Auswertung seismischer Bilder) oder [Andersson & Knutsson 03] (medizinische Bildauswertung).

Bei der Schätzung des OFs wird nur *ein* Eigenvektor des GST genutzt. Man kann jedoch unter Berücksichtigung von Annahme 7 den GST als (unter den genannten Voraussetzungen) umfassende Beschreibung der lokalen Grauwertstruktur ansehen. Daraus ergeben sich weitergehende Auswertungsmöglichkeiten wie z. B. erst kürzlich durch [Hummel 03] untersucht.

Im Rahmen dieser Arbeit soll aber der GST – abgesehen von der OF-Bestimmung – nur für eine relativ einfache Kategorisierung der OF-Vektoren genutzt werden. Die

¹¹So besitzen Gleitkommazahlen einfacher Genauigkeit gemäß [IEEE 754-1985] eine Mantisse mit 23 Bit, was etwa 7 signifikanten Dezimalstellen entspricht. Bei Gleitkommazahlen doppelter Genauigkeit besitzt die Mantisse sogar 52 Binärstellen, was etwa 15 Dezimalstellen entspricht. In Einzelfällen *kann* sich ein Rundungsfehler überproportional auf das Ergebnis auswirken (vgl. [Golub & van Loan 96, Beispiel 7.2.3]), doch ist dies nur für eine im Kontext dieser Arbeit vernachlässigbar kleine Zahl von Fällen zu erwarten.

Grundidee dazu stammt von [Nagel & Gehrke 98a]. In [Middendorf & Nagel 01] wurde eine Verfeinerung der Kategorisierung vorgestellt und deren Nutzen nachgewiesen. Die folgende Darstellung wurde unter Anpassung der Bezeichner im Wesentlichen aus [Middendorf & Nagel 01] entnommen. Dabei seien $\lambda_1 \geq \lambda_2 \geq \lambda_3$ die sortierten Eigenwerte des GST am Punkt \boldsymbol{x} und \boldsymbol{e}_1 , \boldsymbol{e}_2 und \boldsymbol{e}_3 die entsprechenden Eigenvektoren:

- **Neutral (N)** — Pixel, bei denen die Grauwertstruktur mehr oder weniger homogen ist und deshalb keine verlässliche Bewegungsschätzung möglich ist. Diese Pixel werden daran erkannt, dass alle Eigenwerte des GST fast verschwinden:

$$\text{Spur}(\text{GST}(\boldsymbol{x})) = \lambda_1 + \lambda_2 + \lambda_3 \leq \tau_N . \quad (2.23)$$

- **Dominierende Gradientenrichtung (DGD¹²)** — Pixel, die durch einen ausgeprägten, (nahezu) geradlinigen Grauwertübergang dominiert werden. An solchen Stellen tritt das Aperturproblem auf: Alle Gradientenvektoren sind mehr oder weniger gleich orientiert, im Allgemeinen kann nur eine Komponente des OF (diejenige senkrecht zur Kante) zuverlässig geschätzt werden. Erkennbar sind solche Punkte daran, dass der größte Eigenwert sehr viel größer ist (Faktor 100 oder mehr) als die beiden kleineren Eigenwerte, oder – umformuliert – dass die beiden kleineren Eigenwert klein sind im Vergleich zur Spur:

$$\frac{\lambda_2 + \lambda_3}{\frac{2}{3} \cdot (\lambda_1 + \lambda_2 + \lambda_3)} \leq \tau_{\text{DGD}} . \quad (2.24)$$

Innerhalb dieser Klasse lässt sich eine weitere Unterteilung vornehmen:

- **DGD-OF** — Wenn der kleinste Eigenwert wesentlich kleiner ist als der mittlere ($\lambda_3/\lambda_2 < \tau_{\text{DGD-PP}} \ll 1$) und die dominierende Grauwertkante nicht fast genau senkrecht zur Bewegungsrichtung steht (d. h. wenn der Winkel zwischen der Projektion P der Eigenvektoren \boldsymbol{e}_1 und \boldsymbol{e}_3 in die Bildebene weit genug von 90° abweicht, $|\angle(P(\boldsymbol{e}_1), P(\boldsymbol{e}_3)) - 90^\circ| > \angle_{\text{DGD-PP}}$), so deutet dies darauf hin, dass der Unterschied zwischen mittlerem und kleinstem Eigenwert durch eine kleine, aber systematische Änderung der Grauwertstruktur entlang der dominierenden Kante bestimmt ist und deshalb der Eigenvektor zum kleinsten Eigenwert als (relativ zuverlässige) OF-Schätzung angesehen werden kann.
- **DGD-PP¹³** — Ist eine dieser beiden Bedingungen nicht erfüllt, so deutet dies auf eine unzuverlässige OF-Schätzung hin, bei der tatsächlich nur die Bewegungskomponente senkrecht zur Kante geschätzt werden kann.

¹²Die Abkürzung DGD ist auf die englische Bezeichnung „dominant gradient direction“ zurückzuführen.

¹³PP steht hier für „perpendicular“, vgl. Text zur Erklärung.

Parameter	Wert	Bemerkung
τ_N	1,0	
τ_{DGD}	0,050	
$\tau_{\text{DGD-PP}}$	$\frac{1}{50}$	
$\sphericalangle_{\text{DGD-PP}}$	$5,74^\circ$	$= \arcsin(0,1)$
τ_{OFD}	0,0120	

Tabelle 2.1: Kategorisierungsparameter, die für alle in dieser Arbeit beschriebenen Versuche verwendet wurden.

- **Diskontinuitäten (OFD)** — Wir erwarten, dass der kleinste Eigenwert λ_3 nahe bei 0 liegt. Ist dies nicht der Fall, so deutet dies auf eine Diskontinuität hin, eine Grenze zwischen Bereichen, die sich unterschiedlich bewegen. Solche Bereiche können durch einen Vergleich des kleinsten Eigenwerts mit der Spur des GST erkannt werden, wobei berücksichtigt wird, dass bei insgesamt größeren Gradienten auch die nicht zu vermeidende Störung größer wird:

$$\frac{\lambda_3}{\frac{1}{3} \cdot (\lambda_1 + \lambda_2 + \lambda_3)} > \tau_{\text{OFD}} . \quad (2.25)$$

- **Reguläre OF-Vektoren (ROF)** — An den verbliebenen Pixeln kann von einer zuverlässigen Schätzung des Optischen-Flusses ausgegangen werden.

Die Schwellwerte müssen vom Benutzer aufgrund von Erfahrungswerten, theoretischen Überlegungen und einer Betrachtung der Ergebnisse interaktiv eingegeben werden. Ausführliche Überlegungen zur Wahl geeigneter Schwellwerte finden sich in [Nagel & Gehrke 98c]. Bei allen in dieser Arbeit beschriebenen Experimenten wurden – wenn nicht ausdrücklich auf eine Abweichung hingewiesen wird – stets dieselben Schwellwerte verwendet, die auf Erfahrungswerten unter Berücksichtigung der theoretischen Überlegungen und im Hinblick auf die Vergleichbarkeit mit alten Versuchen gewählt wurden (siehe Tabelle 2.1).¹⁴

¹⁴Durch einen Umrechnungsfehler wurde in [Middendorf & Nagel 01] ein um mehr als eine Größenordnung zu kleiner Wert für τ_{DGD} angegeben.

Kapitel 3

Bestimmung der Faltungsmasken

In diesem Kapitel wird zunächst untersucht, wie sich der durch die Beschränkung der Maskengröße vernachlässigte Anteil der Gaußfunktion oder ihrer partiellen Ableitungen quantifizieren lässt. Anschließend wird ein zweistufiges Verfahren zur Bestimmung der eigentlichen Faltungsmasken inklusive Korrektur des durch Diskretisierung und Beschränkung der Maskengröße verursachten Fehlers vorgestellt, analysiert und erweitert. Am Ende des Kapitels wird noch die Frage untersucht, ob sich Zuordnungen zwischen Verfahren zur Korrektur von Tiefpass- und Ableitungsmasken herstellen lassen.

3.1 Problembeschreibung

Die Verwendung von Faltungen mit Masken auf der Basis der Gaußfunktion wurde mit Hilfe von Überlegungen an kontinuierlichen Funktionen motiviert. Die Bildfunktion g ist aber nur an diskreten, festen Rasterpunkten definiert. Das „Standard-Verfahren“ besteht daher meist darin, die kontinuierliche Maskenfunktionen durch Abtastung in eine diskrete Funktion umzuwandeln und dann diskrete Faltungen durchzuführen.

Obwohl zahlreiche Algorithmen im Bereich Bild- und Bildfolgenauswertung auf diskreten Faltungen mit Gaußglocken oder deren Ableitungen beruhen, gibt es nur wenige Arbeiten, die sich mit der praktischen Umsetzung der Faltungen und den dabei auftretenden Problemen beschäftigen.¹ Einige Arbeiten (z. B. [Deriche 87; Vliet et al. 98]) zielen darauf ab, die Faltung durch eine effiziente, rekursive Implementierung zu beschleunigen, was aber zu Lasten der Präzision geht. Im Kontext der vorliegenden Arbeit interessieren aber vielmehr solche Arbeiten, die versuchen, den durch Diskretisierung und Beschränkung der Maskengröße verursachten Fehler abzuschätzen oder zu reduzieren.

Die Beschäftigung mit dieser Frage ist keineswegs nebensächlich. Schon ein wissen-

¹Allerdings ist es nicht auszuschließen, dass einige Arbeiten sich „nebenbei“ mit diesem Thema auseinandersetzen und deshalb bei der Literaturdurchsicht übersehen wurden.

schaftlicher Schlagabtausch zweier amerikanischer Forschungsgruppen vor mehr als 10 Jahren zeigt, dass Implementierungsdetails nicht vernachlässigt werden dürfen: [Sotak & Boyer 89] kritisieren an einer Arbeit von [Chen et al. 87], dass diese zu kleine Masken zur Berechnung von zweiten Ableitungen verwenden würden. [Chen et al. 89] können aber zeigen, dass der Fehler hauptsächlich durch eine inadäquate Implementierung der anderen Autoren zustande kam. Eine ähnliche Kontroverse – ebenfalls um Implementierungsfragen – trugen schon einige Jahre zuvor [Grimson & Hildreth 85] und [Haralick 84; Haralick 85] aus. Untersuchungen von [Middendorf 02] zeigen, dass die OF-Schätzung bei manchen Bildfolgen durchschnittlich mit einem Fehler behaftet ist, dessen Betrag nur 1% der Länge des VR-Vektors entspricht. Um einen so kleinen Fehler weiter zu reduzieren, ist eine möglichst exakte Implementierung der Ableitungsschätzung und der anschließenden Mittelung notwendig.

Als einer der ersten Autoren weist [Lindeberg 93] darauf hin, dass vor der Diskretisierung noch erfüllte Axiome wie beispielsweise die bereits erwähnte Halbgruppen-Eigenschaft der Skalenoperatoren (vgl. Seite 16) nach der Diskretisierung im Allgemeinen verletzt werden. Obwohl er einräumt, dass die einfache Diskretisierung bei sorgfältiger Vorgehensweise und nicht zu kleinen Skalen zu akzeptablen Ergebnissen führen kann, entwickelt er eine neue Theorie mit dem Ziel, den durch die Diskretisierung verursachten Fehler zu minimieren, indem er *diskrete* Funktionen findet, die dieselben elementaren Grundforderungen erfüllen wie bisher nur kontinuierliche Funktionen.

Auch [Otte 94] untersucht den im diskreten Fall auftretenden Fehler, orientiert sich aber stark an der Anwendung. So versucht er beispielsweise die durch die Beschränkung der Maskengröße verursachte systematische Unterschätzung der partiellen Ableitungen zu korrigieren. Sein Hauptaugenmerk bei der Maskenbestimmung richtet sich auf die Abschätzung und Korrektur des durch die Beschränkung der Maskengröße *bei der Anwendung* verursachten Fehlers.

Mit einem Spezialfall desselben Problems – nämlich der Verschiebung des Nulldurchgangs des auf die Gaußfunktion angewandten Laplace-Operators aufgrund der Beschränkung der Maskengröße – beschäftigt sich auch [Gunn 99] und schlägt ebenfalls ein Korrekturverfahren vor.

[Van den Boomgaard & van der Weij 01] regen an, die (unbekannte) kontinuierliche Bildfunktion näherungsweise durch Interpolation der diskreten Bildfunktion wiederherzustellen und dann eine kontinuierliche Faltung durchzuführen. Es stellt sich heraus, dass sich Interpolation und Faltung zusammenfassen lassen, so dass letztendlich doch eine diskrete Faltung durchgeführt wird. Die vorgeschlagene Interpolation 0-ten Grades führt übrigens zu einer Faltungsmaske, die exakt mit dem übereinstimmt, was [Otte 94] als „integrierten Maskenkoeffizienten“ bezeichnet (vgl. auch Seite 37).

Im Gegensatz zu den bisher genannten Arbeiten analysieren einige Autoren den Diskretisierungsfehler im Frequenzraum. So weisen bspw. [Haußecker & Jähne 97] darauf hin, dass bei ihren Untersuchungen allein durch eine sorgfältige, durch eine Minimierung

des Fehlers im Frequenzraum motivierte Implementierung der Maskenbestimmung der Fehler um eine Größenordnung reduziert werden konnte. Ebenfalls auf einer Analyse im Frequenzraum baut [Florack 00] auf.

Dieses Kapitel setzt hauptsächlich auf den Arbeiten von [Otte 94] auf. Es zeigt, wie die Ergebnisse durch ein noch sorgfältigeres Vorgehen weiter verbessert werden können. In Abschnitt 3.4 wird abschließend ein Ansatz vorgestellt, der Parallelen zur Arbeit von [Van den Boomgaard & van der Weij 01] aufweist.

3.2 Größe der Faltungsmasken

Gaußfunktionen und ihre Ableitungen nehmen für beliebig große Entfernungen vom Ursprung Werte an, die noch etwas von Null abweichen. Faltungsmasken können aber nur eine beschränkte Größe haben. Der Wahl der „richtigen“ Maskengröße – bei vorgegebener Skala – kommt in der Anwendung wesentliche Bedeutung zu: Sind die Masken zu klein, so verschlechtern sich die erzielten Ergebnisse, sind die Masken zu groß, steigt die Rechenzeit stark an. Für eine „gute Wahl“ bei der Maskengröße ist eine genaue Kenntnis über die Auswirkungen dieser Entscheidung notwendig. In seiner Dissertation schätzt [Otte 94] den Fehler ab, der durch die endliche Maskenausdehnung verursacht wird, beschränkt sich dabei aber auf den eindimensionalen Fall. In den folgenden Abschnitten werden die Überlegungen von Otte präzisiert und auf den dreidimensionalen Fall ausgedehnt.

3.2.1 Eindimensionale Masken

In Anlehnung an die Begriffsbildung von [Otte 94] soll bei einer eindimensionalen Maske, die sich vom Maskenmittelpunkt um jeweils N Pixel in beide Richtungen erstreckt, von einer *Maskenausdehnung* N und von einer *Maskengröße* oder *Maskenbreite* $2N+1$ gesprochen werden (siehe Abbildung 3.1). N kann nur ganzzahlige Werte annehmen. Da sich die nachfolgenden Überlegungen auf kontinuierliche Gaußfunktionen beziehen, wird darüber hinaus der Begriff des *Maskenbereichs* \mathcal{M} eingeführt. Dabei handele es sich um das zum Ursprung symmetrische Teilintervall der x -Achse, in dem die Maske die Gaußfunktion oder deren Ableitungen „angemessen“ wiedergibt. [Otte 94] unterscheidet nicht zwischen Maskengröße und Maskenbereich, was – wie sich zeigen wird – eine exakte Analyse verhindert.

Die i -te Ableitung der Gaußfunktion in Abhängigkeit vom Skalenparameter σ werde mit $f^{(i)}(x, \sigma)$, $i = 0, 1, 2, \dots$ bezeichnet. Zur Vereinfachung der Darstellung wird die Gaußfunktion selbst im Folgenden auch als „0-te Ableitung“ bezeichnet, und es werden f und $f^{(0)}$ gleichwertig benutzt. Es gilt also:

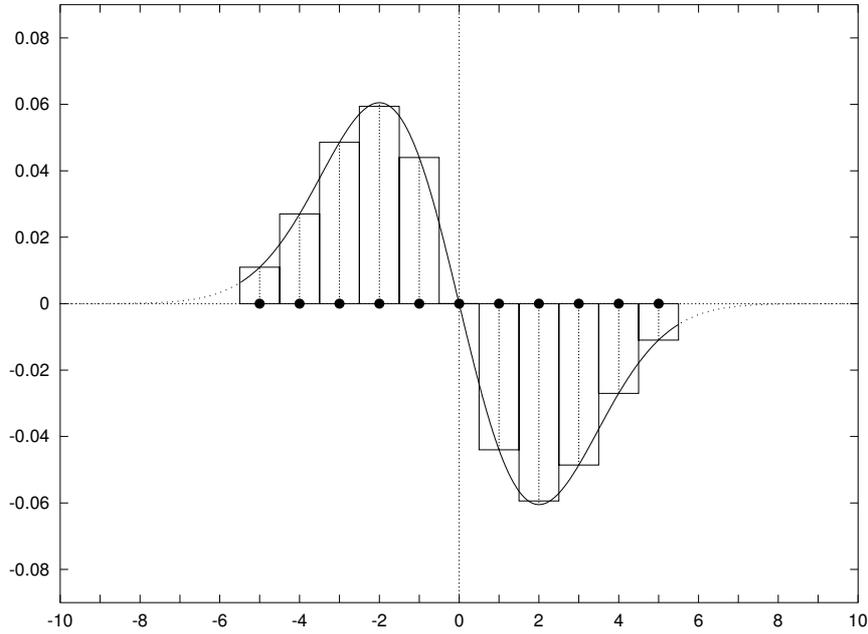


Abbildung 3.1: Erste Ableitung einer Gaußfunktion mit Standardabweichung $\sigma = 2$ (schematisch, modifiziert aus [Otte 94]) und daraus gewonnene Faltungsmaske der Größe 11. Die Punkte auf der x -Achse markieren diejenigen Stellen, an denen Maskenkoeffizienten durch Abtastung oder stückweise Integration bestimmt werden (vgl. Abschnitt 3.3.1). Die Höhe der punktierten Linien in der Mitte der Rechtecke markieren die Größe der abgetasteten Koeffizienten, die Fläche der Rechtecke entspricht der Fläche unter der Ableitung der Gaußglocke und somit der Größe der integrierten Maskenkoeffizienten. Der punktierte Verlauf der Gaußfunktion wurde vernachlässigt. Sowohl abgetastete als auch integrierte Maske repräsentieren den Verlauf der Funktion im Intervall $[-5,5; 5,5]$, obwohl die äußersten Maskenkoeffizienten „nur“ bei -5 bzw. $+5$ liegen.

$$f(x, \sigma) = f^{(0)}(x, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (3.1a)$$

$$f^{(1)}(x, \sigma) = -\frac{x}{\sigma^2} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (3.1b)$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

Außerdem bezeichne $F^{(i)}(x_1, x_2, \sigma)$ die Fläche unter der i -ten Ableitung der Gaußfunktion im Intervall $[x_1, x_2]$:

$$F^{(0)}(x_1, x_2, \sigma) = \frac{1}{2} \left(\operatorname{erf} \left(\frac{x_2}{\sqrt{2}\sigma} \right) - \operatorname{erf} \left(\frac{x_1}{\sqrt{2}\sigma} \right) \right) \quad (3.2a)$$

$$\begin{aligned}
F^{(1)}(x_1, x_2, \sigma) &= f(x_2, \sigma) - f(x_1, \sigma) \\
&\vdots \\
&\vdots
\end{aligned} \tag{3.2b}$$

Die i -te Ableitung $f^{(i)}$, $i > 0$ der Gaußfunktion besitzt genau i Nulldurchgänge, d. h. Nullstellen, davon aufgrund der Achsen- bzw. Punktsymmetrie mindestens eine im nicht-negativen Teil der x -Achse. $z^{(i)}(\sigma)$ sei der vom Ursprung aus gesehen am weitesten entfernt liegende nicht-negative Nulldurchgang der i -ten Ableitung. $z^{(0)}(\sigma)$ wird per Definition auf 0 gesetzt. Es gilt also:

$$z^{(i)}(\sigma) = \begin{cases} 0 & \text{für } i = 0 \\ \max \{ x \geq 0 \mid f^{(i)}(x) = 0 \} & \text{für } i > 0 \end{cases} . \tag{3.3}$$

$F^{(i)}(z^{(i)}(\sigma), \infty, \sigma)$ ist also die Fläche unter der i -ten Ableitung der Gaußfunktion jenseits des vom Ursprung aus gesehen am weitesten entfernt liegenden nicht-negativen Nulldurchgangs. Das Verhältnis der Fläche unter dem durch die Beschränkung des *Maskenbereichs* abgeschnittenen Teil der Gaußfunktion oder ihrer Ableitungen im positiven Teil der x -Achse zur Fläche $F^{(i)}(z^{(i)}(\sigma), \infty, \sigma)$ werde als *erwarteter* oder *theoretischer Maskenfehler* c bezeichnet.² Dabei wird davon ausgegangen, dass die Maske den Verlauf der Gaußglocke perfekt beschreibt. Der *tatsächliche Maskenfehler* kann größer sein – wenn z. B. die einzelnen Maskenkoeffizienten den Verlauf der Gaußglocke nicht ausreichend gut wiedergeben – oder kleiner sein, wenn z. B. aufgrund der Ganzzahligkeit der *Maskengröße* N der von der Maske tatsächlich berücksichtigte Teil der Gaußglocke größer ist als der zum Unterschreiten eines vorgegebenen Maskenfehlers erforderliche Maskenbereich.

Zum besseren Verständnis der späteren Überlegungen bei dreidimensionalen Masken soll die Bestimmung von c im Eindimensionalen kurz dargestellt werden. c ist vom Maskenbereich, gegeben durch die obere Intervallgrenze μ , und von σ abhängig. Die Indizierung des theoretischen Maskenfehlers c deutet dabei an, um welche Ableitung es sich handelt, und ob es um den ein- oder dreidimensionalen Fehler geht.

Bei der Tiefpassfilterung gilt bei Wahl des Maskenbereichs $\mathcal{M} = [-\mu, \mu]$, also bei einem Abschneiden der Maske an der Stelle μ ($\mu > 0$):

$$\int_{\mu}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = c_{1D}^{(0)}(\mu, \sigma) \cdot \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \tag{3.4a}$$

$$\iff c_{1D}^{(0)}(\mu, \sigma) = 2 \cdot \int_{\mu}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \tag{3.4b}$$

²[Otte 94] verwendet in der analogen Definition die Maskenausdehnung anstelle des Maskenbereichs, was unpräzise ist. Außerdem bezeichnet er dieses Verhältnis als Maskengenauigkeit, was aber missverstanden werden könnte, da kleine Werte einer großen Maskengenauigkeit entsprechen.

oder abkürzend

$$c_{1D}^{(0)}(\mu, \sigma) = 1,0 - \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right). \quad (3.4c)$$

Analog gilt bei der ersten Ableitung der Gaußfunktion:

$$\int_{\mu}^{\infty} -\frac{\xi}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = c_{1D}^{(1)}(\mu, \sigma) \cdot \int_0^{\infty} -\frac{\xi}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \quad (3.5a)$$

$$\iff c_{1D}^{(1)}(\mu, \sigma) = e^{-\frac{\mu^2}{2\sigma^2}}, \quad (3.5b)$$

bei der zweiten Ableitung (für $\mu > \sigma$):

$$\int_{\mu}^{\infty} \left(\frac{\xi^2}{\sigma^4} - \frac{1}{\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = c_{1D}^{(2)}(\mu, \sigma) \int_{\sigma}^{\infty} \left(\frac{\xi^2}{\sigma^4} - \frac{1}{\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \quad (3.6a)$$

$$\iff c_{1D}^{(2)}(\mu, \sigma) = \frac{\mu}{\sigma} e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 + \frac{1}{2}}, \quad (3.6b)$$

und bei der dritten Ableitung (für $\mu > \sqrt{3}\sigma$):

$$\int_{\mu}^{\infty} -\left(\frac{\xi^3}{\sigma^6} - \frac{3\xi}{\sigma^4}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = c_{1D}^{(3)}(\mu, \sigma) \int_{\sqrt{3}\sigma}^{\infty} -\left(\frac{\xi^3}{\sigma^6} - \frac{3\xi}{\sigma^4}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \quad (3.7a)$$

$$\iff c_{1D}^{(3)}(\mu, \sigma) = \frac{1}{2} \left(\left(\frac{\mu}{\sigma}\right)^2 - 1 \right) e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 + \frac{3}{2}}. \quad (3.7b)$$

Mit Hilfe dieser Formeln ist es möglich, bei vorgegebenem theoretischen Maskenfehler c den erforderlichen *Maskenbereich* zu bestimmen. Aus dem Maskenbereich kann dann durch Rundung der Intervallgrenzen auf die nächste ganze Zahl die notwendige *Maskenausdehnung* bestimmt werden. Die dabei in der Hälfte der Fälle auftretende Rundung nach unten führt nicht zum Überschreiten der Schwelle für den vorgegebenen Maskenfehler, da ein Maskenkoeffizient an der Stelle $n \in \mathbb{Z}$ die Gaußfunktion (oder deren Ableitung) im Intervall $[n - \frac{1}{2}, n + \frac{1}{2}]$ widerspiegelt (vgl. auch Erläuterungen bei Abbildung 3.1). Somit deckt eine Maske der Ausdehnung N den Maskenbereich $[-(N + \frac{1}{2}), N + \frac{1}{2}]$ ab.

[Otte 94] gibt in Tabelle 2.1, Seite 25, die Werte an, die sich für vorgegebene theoretische Maskenfehler von 1, 5 oder 10 % bei Auflösung der voranstehenden Formeln nach μ ergeben. Genau genommen kann in dieser Tabelle aber nicht der Mindestwert für N , sondern für $N + \frac{1}{2}$ abgelesen werden. In der vorliegenden Implementierung von Otte wird dies aber korrekt gehandhabt.

3.2.2 Dreidimensionale Masken

Bei mehrdimensionalen Masken werden die für den eindimensionalen Fall eingeführten Begriffe analog verwendet. Die Form der Gaußfunktion wird nun durch eine 3×3 -Kovarianzmatrix Σ bestimmt. Der Maskenbereich $\mathcal{M} = [-\mu_x, +\mu_x] \times [-\mu_y, +\mu_y] \times [-\mu_t, +\mu_t]$ wird nun vereinfachend durch einen Vektor $\mathbf{m} = (\mu_x, \mu_y, \mu_t)^T$ repräsentiert.

Bisher wurde bei 3D-Masken die notwendige Maskenausdehnung aus den Formeln für eindimensionale Masken ermittelt. Dies führt aber dazu, dass der tatsächliche Maskenfehler größer ist³ als der bei der Bestimmung des Maskenbereichs vorgegebene Wert c . Um dies zu demonstrieren, wird zunächst angenommen, dass die Kovarianzmatrix Σ Diagonalgestalt hat. In diesem Fall ist bekanntermaßen eine Faktorisierung der dreidimensionalen Faltung in drei eindimensionale Faltungen „entlang der Koordinatensystem-Achsen“ möglich. Wenn dann bei jeder der drei Teilfilterungen 5 % der Fläche unter der eindimensionalen Gaußglocke nicht berücksichtigt werden, so liegt der Gesamtfehler viel höher, nämlich bei rund 14,3 %, bei 10 % 1D-Fehler ergeben sich sogar 27,1 % 3D-Fehler.

Die Vorgehensweise von Otte soll daher erweitert werden, so dass bei Vorgabe des tolerierbaren Gesamtfehlers im Dreidimensionalen der dafür notwendige Maskenbereich bestimmt wird. Zur kompakten Darstellung wird bei Ableitungsmasken die Indizierung von c um die Ableitungsrichtung ergänzt.

3.2.2.1 Isotrope Masken

Erfolgt eine isotrope Filterung, d. h. ist in der Kovarianzmatrix $\Sigma = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_t^2)$ der Gaußparameter σ für alle drei Teilfaltungen gleich, so ist es naheliegend, einen würfelförmigen Maskenbereich $[-\mu, +\mu]^3 \subset \mathbb{R}^3$ zu wählen, also $\mathbf{m} = (\mu, \mu, \mu)^T$.

Bei dreidimensionalen Tiefpassmasken findet in jede der drei Richtungen eine eindimensionale Tiefpassfilterung statt. Es gilt also:

$$c_{3D}^{(0)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu, \sigma)\right)^3 \quad (3.8a)$$

$$= 1,0 - \left(\text{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)^3. \quad (3.8b)$$

Bei der Bestimmung der 1. partiellen Ableitung muss berücksichtigt werden, dass sich diese aus zwei Tiefpassfilterungen und einer Ableitungsberechnung zusammensetzt. Wegen der Isotropie der Masken ist es irrelevant, ob die partielle Ableitung in x -, y -

³wenn nicht andere Faktoren diesem Fehler entgegenwirken.

oder in t -Richtung betrachtet wird. Es gilt also:

$$\begin{aligned} c_{3D_x}^{(1)}(\mathbf{m}, \Sigma) &= c_{3D_y}^{(1)}(\mathbf{m}, \Sigma) = c_{3D_t}^{(1)}(\mathbf{m}, \Sigma) \\ &= 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu, \sigma)\right)^2 \cdot \left(1,0 - c_{1D}^{(1)}(\mu, \sigma)\right) \end{aligned} \quad (3.9a)$$

$$= 1,0 - \left(\operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)^2 \cdot \left(1,0 - e^{-\frac{\mu^2}{2\sigma^2}}\right). \quad (3.9b)$$

Bei Masken zur Bestimmung der 2. partiellen Ableitung muss unterschieden werden, ob zweimal in dieselbe Richtung oder in verschiedene Richtungen abgeleitet wird. Bei Ableitung in dieselbe Richtung – hier o. B. d. A.⁴ in x -Richtung – gilt:

$$c_{3D_{xx}}^{(2)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu, \sigma)\right)^2 \cdot \left(1,0 - c_{1D}^{(2)}(\mu, \sigma)\right) \quad (3.10a)$$

$$= 1,0 - \left(\operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)^2 \cdot \left(1,0 - \frac{\mu}{\sigma} e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 + \frac{1}{2}}\right), \quad (3.10b)$$

bei Ableitung in verschiedene Richtungen – o. B. d. A. x - und y -Richtung – gilt:

$$c_{3D_{xy}}^{(2)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu, \sigma)\right) \cdot \left(1,0 - c_{1D}^{(1)}(\mu, \sigma)\right)^2 \quad (3.11a)$$

$$= 1,0 - \left(\operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right) \cdot \left(1,0 - e^{-\frac{\mu^2}{2\sigma^2}}\right)^2. \quad (3.11b)$$

Bei der 3. partiellen Ableitung müssen sogar drei Fälle unterschieden werden. Wird dreimal in dieselbe Richtung abgeleitet, so gilt:

$$c_{3D_{xxx}}^{(3)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu, \sigma)\right)^2 \cdot \left(1,0 - c_{1D}^{(3)}(\mu, \sigma)\right), \quad (3.12)$$

wird zweimal in dieselbe und einmal in eine andere Richtung abgeleitet, so gilt:

$$c_{3D_{xxy}}^{(3)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu, \sigma)\right) \cdot \left(1,0 - c_{1D}^{(1)}(\mu, \sigma)\right) \cdot \left(1,0 - c_{1D}^{(2)}(\mu, \sigma)\right). \quad (3.13)$$

und wird in jede Richtung genau einmal abgeleitet, so gilt:

$$c_{3D_{xyt}}^{(3)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(1)}(\mu, \sigma)\right)^3. \quad (3.14)$$

Aus diesen Formeln ergeben sich die in Tabelle 3.1 angegebenen Mindest-Maskenbereiche in Abhängigkeit von σ und (vorgegebenem) theoretischen Maskenfehler. Man sieht, dass bei mehrfacher Ableitung *in dieselbe Richtung* im Dreidimensionalen nur

⁴ohne Beschränkung der Allgemeinheit

Maskenfehler	nach [Otte 94] (1D)				auf 3D erweitert, isotrope Masken			
	10 %	5 %	3 %	1 %	10 %	5 %	3 %	1 %
Tiefpass	1,645 σ	1,960 σ	2,170 σ	2,576 σ	2,114 σ	2,388 σ	2,572 σ	2,934 σ
∂x	2,146 σ	2,448 σ	2,648 σ	3,035 σ	2,342 σ	2,614 σ	2,798 σ	3,157 σ
$\partial x \partial x$	2,764 σ	3,035 σ	3,217 σ	3,572 σ	2,803 σ	3,067 σ	3,244 σ	3,592 σ
$\partial x \partial y$					2,489 σ	2,755 σ	2,935 σ	3,287 σ
$\partial x \partial x \partial x$	3,283 σ	3,534 σ	3,703 σ	4,034 σ	3,290 σ	3,539 σ	3,707 σ	4,037 σ
$\partial x \partial x \partial y$					2,853 σ	3,111 σ	3,285 σ	3,627 σ
$\partial x \partial y \partial t$					2,595 σ	2,856 σ	3,032 σ	3,377 σ

Tabelle 3.1: Mindestwert μ für den Maskenbereich $[-\mu, +\mu]$ (1D) bzw. $[-\mu, +\mu]^3$ (3D) für isotrope Masken in Abhängigkeit von Standardabweichung und (vorgegebenem) theoretischen Maskenfehler.

geringfügig größere Masken als im Eindimensionalen notwendig sind, um eine vorgegebene Filtergenauigkeit zu erreichen. Dies ist darauf zurückzuführen, dass die Teilmaske, die für die Ableitung zuständig ist, sowieso schon so groß ist, dass der Fehler der Teilmasken zur Tiefpassfilterung bei dieser Maskengröße kaum mehr ins Gewicht fällt. Als Beispiel sei die dreidimensionale dreifache Ableitung nach x mit einer gewünschten Filtergenauigkeit von 5% herausgegriffen. Schon im Eindimensionalen muss diese Maske einen Bereich von 3,534 σ abdecken. Bei einer solchen Ausdehnung hat eine eindimensionale Tiefpassmaske aber nur noch einen Fehler von 0,04%, der Gesamtfehler läge bei 5,07%. Durch minimale Vergrößerung der Maske kann der Gesamtfehler auf 5% reduziert werden. Bei gemischten Ableitungen ist die erforderliche Maskengröße im Dreidimensionalen sogar geringer als im Eindimensionalen, da sich die Gesamtfaltung aus Ableitungen niedrigerer Ordnung zusammensetzt.

Man sieht auch, dass die von vielen Autoren (z. B. [Fisher et al. 94]) praktizierte Beschränkung der Maskenausdehnung auf drei Standardabweichungen für Tiefpass und erste Ableitung eine gute Wahl ist, wenn man keine genauere Analyse anstellen möchte. Der erwartete 3D-Maskenfehler liegt dann bei 0,8% für die Tiefpassfilterung und bei 1,6% für die Bestimmung der ersten partiellen Ableitung.

3.2.2.2 Anisotrope Masken

Ist Σ keine Diagonalmatrix oder haben σ_x , σ_y und σ_t nicht alle denselben Wert, so muss die Vorgehensweise modifiziert werden.

Diagonalmatrix Besitzt Σ Diagonalgestalt, so ist die Bestimmung der Maskengröße mit einer einfachen Erweiterung des Vorgehens bei isotropen Masken möglich. Dies soll am Beispiel der 1. partiellen Ableitung demonstriert werden. Sei nun $\mathbf{m} = (\mu_x, \mu_y, \mu_t)^T$.

Für die erste partielle Ableitung in x -Richtung gilt dann:

$$c_{3D_x}^{(1)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(1)}(\mu_x, \sigma_x)\right) \cdot \left(1,0 - c_{1D}^{(0)}(\mu_y, \sigma_y)\right) \cdot \left(1,0 - c_{1D}^{(0)}(\mu_t, \sigma_t)\right) , \quad (3.15a)$$

während in t -Richtung gilt:

$$c_{3D_t}^{(1)}(\mathbf{m}, \Sigma) = 1,0 - \left(1,0 - c_{1D}^{(0)}(\mu_x, \sigma_x)\right) \cdot \left(1,0 - c_{1D}^{(0)}(\mu_y, \sigma_y)\right) \cdot \left(1,0 - c_{1D}^{(1)}(\mu_t, \sigma_t)\right) . \quad (3.15b)$$

Hält man in diesem Fall an einer würfelförmigen Maske fest, so ergibt sich für die (oft verwendete) Kombination $\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$ bei einem vorgegebenen theoretischen Maskenfehler von 5% für die x - und die y -Ableitung eine halbe Kantenlänge des Maskenbereiches von 3,589, für die t -Ableitung von 3,376. Durch Rundung auf die nächste ganze Zahl hätten die Masken für die x - und y -Ableitung die Größe $9 \times 9 \times 9$, für die t -Ableitung $7 \times 7 \times 7$.

Sinnvoller ist es in diesem Fall, eine quaderförmige Maskenausdehnung zuzulassen, die das Verhältnis der σ zueinander widerspiegelt und für x -, y - und t -Ableitung gleichermaßen gilt. Dies kann dadurch erreicht werden, dass bei vorgegebenem Maskenfehler zunächst der dafür notwendige würfelförmige Maskenbereich $\mathcal{M}_W = [-\mu, +\mu]^3$ für $\Sigma = I_{3 \times 3}$ bestimmt wird und dieser Bereich anschließend in x -Richtung um den Faktor σ_x , in y -Richtung um σ_y und in t -Richtung um σ_t gestreckt wird. Der quaderförmige Maskenbereich lautet dann $\mathcal{M}_Q = [-\sigma_x \cdot \mu, +\sigma_x \cdot \mu] \times [-\sigma_y \cdot \mu, +\sigma_y \cdot \mu] \times [-\sigma_t \cdot \mu, +\sigma_t \cdot \mu]$. Im vorangehenden Beispiel ergibt sich somit $\mathbf{m} = (1,41 \cdot 2,614; 1,41 \cdot 2,614; 1,00 \cdot 2,614)^T = (3,686; 3,686; 2,614)^T$, woraus eine Maskengröße von $9 \times 9 \times 7$ folgt.

Durch Einsetzen kann verifiziert werden, dass bei einem so bestimmten Maskenbereich \mathcal{M}_Q der zu erwartende Maskenfehler für alle drei partiellen 1. Ableitungen gleich ist. Eine analoge Anwendung des Verfahrens auf höhere Ableitungen ist ohne Probleme möglich.

Tabelle 3.2 gibt die sich aufgrund dieser Überlegungen ergebende Maskenausdehnung für gängige Werte von σ an. Insbesondere lässt sich ablesen, dass die bisher häufig (z. B. bei den Untersuchungen zu [Nagel & Gehrke 98a]) verwendete Kombination einer Maskengröße von $11 \times 11 \times 11$ und einem maximalen σ von 2,83 im Mittelungsschritt der adaptiven GST-Berechnung nicht brauchbar ist. Entweder muss die Maskengröße auf mindestens $13 \times 13 \times 13$, besser $15 \times 15 \times 15$ erhöht werden, oder das maximale σ muss noch weiter reduziert werden (z. B. auf 2,41, was bei $11 \times 11 \times 11$ -Masken dann einem Fehler von 10% entspräche).

Keine Diagonalmatrix Ist Σ keine Diagonalmatrix, so ist der maximale x -Wert des Ellipsoids, das einer Standardabweichung entspricht, bekanntermaßen $\sqrt{\Sigma_{1,1}}$, der

	theor. Maskenfehler	Tiefpass				1. part. Abl.			
		10 %	5 %	3 %	1 %	10 %	5 %	3 %	1 %
$\sigma = 1,0$	nach [Otte 94] (1D)	5	5	5	7	5	5	7	7
	auf 3D erweitert	5	5	7	7	5	7	7	7
$\sigma = 1,41$	nach [Otte 94] (1D)	5	7	7	9	7	7	9	9
	auf 3D erweitert	7	7	9	9	7	9	9	9
$\sigma = 2,0$	nach [Otte 94] (1D)	7	9	9	11	9	11	11	13
	auf 3D erweitert	9	11	11	13	11	11	13	13
$\sigma = 2,83$	nach [Otte 94] (1D)	11	13	13	15	13	15	15	19
	auf 3D erweitert	13	15	15	17	15	15	17	19
$\sigma = 4,0$	nach [Otte 94] (1D)	15	17	19	21	19	21	23	25
	auf 3D erweitert	17	21	21	25	19	21	23	27
$\sigma = 6,0$	nach [Otte 94] (1D)	21	25	27	31	27	31	33	37
	auf 3D erweitert	27	29	31	37	29	33	35	39

Tabelle 3.2: Mindestmaskengröße in Abhängigkeit von Standardabweichung und vorgegebenem theoretischen Maskenfehler.

maximale y -Wert $\sqrt{\Sigma_{2,2}}$ und der maximale t -Wert $\sqrt{\Sigma_{3,3}}$. Man kann also analog zu anisotropen Gaußglocken mit diagonalen Kovarianzmatrizen vorgehen, indem man σ_x durch $\sqrt{\Sigma_{1,1}}$ ersetzt, analog für σ_y und σ_t .

3.3 Bestimmung der Maskenkoeffizienten

Im vorangehenden Abschnitt wurde bei der Fehlerabschätzung davon ausgegangen, dass die Maskenkoeffizienten die Gaußglocke oder ihre Ableitungen exakt wiedergeben. Um die Maskenkoeffizienten einfacher oder schneller bestimmen zu können, werden jedoch oft Näherungsverfahren eingesetzt, die einen zusätzlichen Fehler verursachen.

Die praktische Bestimmung einer Faltungsmaske lässt sich in zwei Schritte unterteilen:

1. Bestimmung von vorläufigen Maskenkoeffizienten, die die Gaußglocke möglichst gut annähern, unter Berücksichtigung von Randbedingungen wie einfache Implementierung oder schnelle Ausführung und
2. „Korrektur“ der Maskenkoeffizienten, um den durch die verwendete Näherung oder durch die Beschränkung der Maskengröße verursachten Fehler auszugleichen.

Beide Aspekte sollen in den folgenden Abschnitten untersucht werden. Dabei werden zunächst nur eindimensionale Faltungen behandelt, da sich zwei- oder dreidimensionale

Faltungen mit einer Gaußglocke mit Kovarianzmatrix in Diagonalgestalt in drei 1D-Faltungen faktorisieren lassen.

3.3.1 Bestimmung der vorläufigen Maskenkoeffizienten

Zur Bestimmung der Maskenkoeffizienten (auf Basis der Gaußfunktion) sind vor allem zwei Verfahren von Bedeutung:

- Direkte Bestimmung durch Auswertung der Gaußfunktion. Dieses Vorgehen lässt sich einfach implementieren und bildet die Gaußfunktion exakt ab. Nachteilig ist ein vergleichsweise hoher Rechenaufwand, insbesondere wenn mehrere Tiefpassfilterungen mit verschiedenen Skalen nacheinander auf dasselbe Bild angewandt werden sollen.
- Annäherung durch einen Binomialfilter. Die Masken weichen vor allem für kleine Skalenparameter merklich von der Gaußglocke ab, können aber bei geeigneter Implementierung sehr schnell für mehrere Skalen berechnet und angewandt werden ([Crowley et al. 02]). Es ist nicht jeder Wert für σ möglich; für $\sigma \neq \frac{n}{4}, n \in \mathbb{N}$ ist eine (ungenaue) Interpolation nötig.

Da im Folgenden jeder beliebige Wert für σ ohne Interpolation möglich sein soll, kommt nur die direkte Bestimmung der Maskenkoeffizienten aus der Gaußglocke in Frage.

Im Folgenden bezeichne $k^{(0)}(n, \sigma)$ bzw. $k^{(1)}(n, \sigma)$ den Koeffizienten der (eindimensionalen) Tiefpassmaske bzw. der ersten Ableitungsmaske für die Standardabweichung σ an der Position $n, n \in \mathbb{Z}$, wobei die Position 0 dem Maskenmittelpunkt entspricht. Die Maskengröße sei $2N + 1$. Außerdem bezeichne – wie in Gleichung (3.1) – $f^{(0)}(x, \sigma)$ die Gaußfunktion mit der Standardabweichung σ und $f^{(1)}(x, \sigma)$ die erste Ableitung von $f^{(0)}(x, \sigma)$.

Für die Nummerierung der Maskenkoeffizienten gibt es zwei Möglichkeiten. Meistens benutzt man die „normale“ Reihenfolge, so dass eine diskrete Faltung \otimes der Grauwertfunktion g mit der Maskenfunktion $k^{(i)}$ geschrieben werden kann als:

$$(g \otimes k)(x_0) = \sum_{n=-N}^{+N} g(x_0 - n) \cdot k(n, \sigma) \quad , \quad (3.16)$$

während einige Autoren Faltungsmasken so angeben, dass Funktion und Maske „übereinander geschoben“ werden können:

$$(g \otimes k)(x_0) = \sum_{n=-N}^{+N} g(x_0 + n) \cdot k(n, \sigma) \quad . \quad (3.17)$$

Im Folgenden werden Masken in Übereinstimmung mit der in Formel (3.16) erforderlichen Nummerierung dargestellt.

Das einfachste Verfahren zur Bestimmung der Maskenkoeffizienten besteht darin, die Gaußfunktion (bzw. ihre Ableitungen) abzutasten:

$$k^{(0)}(n, \sigma) = f^{(0)}(n, \sigma), \quad (3.18)$$

$$k^{(1)}(n, \sigma) = f^{(1)}(n, \sigma). \quad (3.19)$$

Vor allem bei kleinen Standardabweichungen sind aber größere numerische Abweichungen zwischen diesen *abgetasteten Filtermaskenkoeffizienten* und approximierter zugrunde liegender (Gauß-)Filterfunktion festzustellen.

Aus diesem Grund verwendet [Otte 94] *integrierte Koeffizienten*:

$$k^{(0)}(n, \sigma) = \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} f^{(0)}(\xi, \sigma) d\xi = F^{(0)}\left(n - \frac{1}{2}, n + \frac{1}{2}, \sigma\right), \quad (3.20)$$

$$k^{(1)}(n, \sigma) = \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} f^{(1)}(\xi, \sigma) d\xi = F^{(1)}\left(n - \frac{1}{2}, n + \frac{1}{2}, \sigma\right). \quad (3.21)$$

Dabei steht $F^{(i)}(x_1, x_2, \sigma)$ wie schon im vorangehenden Abschnitt für die Fläche unter der i -ten Ableitung der Gaußfunktion im Intervall $[x_1, x_2]$.

3.3.2 Korrektur der Maskenkoeffizienten

Bei den Masken sowohl mit abgetasteten als auch mit integrierten Koeffizienten treten bedingt durch die beschränkte Maskengröße Fehler auf, bei abgetasteten Maskenkoeffizienten entsteht zusätzlich ein Fehler durch die Abtastung. Beide Fehler sollen – wie nachfolgend diskutiert – durch Skalierung aller oder Korrektur einzelner Koeffizienten behoben werden.

Die korrigierten Maskenkoeffizienten werden durch $\tilde{k}^{(0)}(n, \sigma)$ bzw. $\tilde{k}^{(1)}(n, \sigma)$ bezeichnet.

3.3.2.1 Tiefpassmasken

Ein wesentlicher Fehler bei Tiefpassfilterungen tritt dadurch auf, dass die Summe der durch Abtastung oder stückweise Integration der Gaußfunktion gewonnenen Maskenkoeffizienten stets kleiner als eins ist. In diesem Fall entspricht das Ergebnis der Filterung selbst bei über den gesamten Bildbereich konstanten Grauwerten nicht wie erwartet dem Eingabebild, sondern enthält verminderte Grauwerte.

Die Maskenkoeffizienten müssen daher auf jeden Fall so korrigiert werden, dass ihre Summe eins ergibt. In diesem Bericht sollen die folgenden zwei Möglichkeiten näher untersucht werden:

1. Einheitliche *Skalierung* der Koeffizienten: Alle Maskenkoeffizienten werden mit demselben Faktor multipliziert, so dass ihre Summe 1 ergibt:

$$\tilde{k}^{(0)}(n, \sigma) = \frac{k^{(0)}(n, \sigma)}{\sum_{i=-N}^N k^{(0)}(i, \sigma)}. \quad (3.22)$$

Dies ist das allgemein anerkannte „Standardverfahren“, im Folgenden als *Skalierungsverfahren* bezeichnet.

2. *Randkorrektur*: Nur die beiden Maskenkoeffizienten am Rand der Maske werden vergrößert, so dass sich 1 als Summe ergibt. [Otte 94] führte dieses *Randkorrekturverfahren* für Masken mit integrierten Koeffizienten ein:

$$\tilde{k}^{(0)}(n, \sigma) = \begin{cases} F^{(0)}(-\infty, -N + \frac{1}{2}, \sigma) & \text{für } n = -N \\ F^{(0)}(n - \frac{1}{2}, n + \frac{1}{2}, \sigma) & \text{für } |n| < N \\ F^{(0)}(N - \frac{1}{2}, \infty, \sigma) & \text{für } n = +N \end{cases}. \quad (3.23)$$

Eine naheliegende Verallgemeinerung, die auch für abgetastete Koeffizienten gelten soll, könnte lauten:

$$\tilde{k}^{(0)}(n, \sigma) = \begin{cases} k(n, \sigma) & \text{für } |n| < N \\ \sum_{i=N}^{\infty} k^{(0)}(i, \sigma) & \text{für } |n| = N \end{cases}. \quad (3.24)$$

Allerdings kann die unendliche Summe nicht ohne weiteres berechnet werden. Außerdem ist die Summe der Maskenkoeffizienten nach Formel (3.24) bei abgetasteten Koeffizienten immer noch ungleich 1. Eine korrekte Formulierung muss deshalb lauten:

$$\tilde{k}^{(0)}(n, \sigma) = \begin{cases} k^{(0)}(n, \sigma) & \text{für } |n| < N \\ \frac{1}{2} \left(1 - \sum_{i=-(N-1)}^{N-1} k^{(0)}(i, \sigma) \right) & \text{für } |n| = N \end{cases}. \quad (3.25)$$

Bei der Randkorrektur kann es vor allem bei großem σ passieren, dass der Maskenkoeffizient für die äußeren Pixel größer wird als für „innere“ Pixel, bei integrierten Maskenkoeffizienten z.B. bei $\sigma = \sqrt{8}$ und $N \leq 4$ oder bei $\sigma = 4$ und $N \leq 9$. Wie aus Tabelle 3.2 hervorgeht, ist in diesen Fällen die Maskenausdehnung im Verhältnis zu σ aber sowieso viel zu klein gewählt, d. h. es handelt sich mehr um eine theoretische als um eine praktisch relevante Schwäche.

Welches dieser beiden Verfahren man verwendet, hängt wesentlich davon ab, welche der implizit zugrunde gelegten Annahmen man für richtig hält.

Die Randkorrektur basiert auf der Annahme, dass die Randpixel die durch die Beschränkung der Maskengröße abgeschnittenen Pixel besser repräsentieren als das gewichtete Mittel aller Pixel im Maskenbereich. Diese Annahme ist gerechtfertigt, wenn

die Grauwertfunktion im Randbereich der Maske und im abgeschnittenen Bereich monoton oder zumindest ohne sprunghafte Grauwertänderungen verläuft und das Rauschen klein ist. Nachteilig ist, dass sich dadurch verrauschte Randpixel besonders stark auf das Ergebnis auswirken. Bei nicht-monotoner oder sprunghafter Grauwertfunktion sind durch Verwendung der Randkorrektur entstehende Fehler mehr auf störende Eigenschaften der Grauwertfunktion oder eine inadäquate Wahl der Maskenparameter zurückzuführen als auf die Randkorrektur selbst.

[Otte 94] erzielt mit der Randkorrektur gute Ergebnisse, was aber nicht zuletzt daran liegt, dass er für Vergleiche eine rauschfreie Grauwertstruktur wählt, bei der die Grauwerte im abgeschnittenen Bereich und am Rand der Maske absolut homogen sind, d. h. eine Grauwertstruktur, die genau der zugrundeliegenden Annahme entspricht.

3.3.2.2 Ableitungsmasken

Auch bei der Ableitungsbestimmung führt die Beschränkung der Maskengröße zu einer systematischen Fehlschätzung. Ähnlich wie bei der Tiefpassfilterung ist die Summe der positiven (negativen) Maskenkoeffizienten kleiner (größer) als der Wert, der sich bei unendlicher Maskenausdehnung ergeben würde. Bei monotonem Anstieg der Grauwerte kommt es dadurch offensichtlich zu einer Unterschätzung der Ableitungen. Der so verursachte Fehler wird durch Diskretisierungseffekte überlagert, es bleibt jedoch meist bei einer Unterschätzung der Ableitungen.

Im Folgenden werden sieben Verfahren zur Korrektur der Maskenkoeffizienten vorgestellt, die auf unterschiedlichen Anforderungen an das Resultat basieren und somit auch zu unterschiedlichen Ergebnissen führen. Vier dieser Verfahren skalieren alle Maskenkoeffizienten mit demselben Faktor – und werden nachfolgend zusammenfassend als *Skalierungsverfahren* bezeichnet –, die drei anderen Verfahren passen nur bestimmte Koeffizienten an, um den Fehler zu korrigieren. Beide Grundideen wurden bereits bei der Korrektur der Tiefpassmasken angewandt. Es sei schon an dieser Stelle auf Abbildung 4.4 (Seite 56) verwiesen, die alle Korrekturverfahren für Ableitungsmasken gegenüberstellt und dabei auch konzeptionelle Parallelen aufzeigt.

σ -basierte Skalierung In Analogie zum Vorgehen bei der Skalierung der Tiefpass-Maskenkoeffizienten können auch bei der Ableitungsbestimmung die Koeffizienten so skaliert werden, dass ihre Summe dem bei unendlicher Maskenausdehnung zu erwartenden Wert entspricht. Dabei müssen aufgrund der Antisymmetrie der ersten Ableitung der Gaußfunktion positive und negative Maskenkoeffizienten getrennt betrachtet werden.

Ausgehend von

$$\int_0^{\infty} f^{(1)}(\xi, \sigma) d\xi = -\frac{1}{\sigma\sqrt{2\pi}} \quad (3.26)$$

sollte auch nach Diskretisierung und Beschränkung der Maskengröße

$$\sum_{i=0}^N k^{(1)}(i, \sigma) = -\frac{1}{\sigma\sqrt{2\pi}} \quad (3.27)$$

gelten. Um dies zu erreichen, werden die skalierten Maskenkoeffizienten $\tilde{k}^{(1)}(n, \sigma)$ wie folgt berechnet:

$$\tilde{k}^{(1)}(n, \sigma) = k^{(1)}(n, \sigma) \frac{-f^{(0)}(0, \sigma)}{\sum_{i=1}^N k^{(1)}(i, \sigma)}. \quad (3.28)$$

Otte definiert diese Art der Skalierung in [Otte 94], Definition 2.1, Formel (2.9) für integrierte Maskenkoeffizienten. Formel (3.28) verallgemeinert dies für integrierte und abgetastete Koeffizienten.

Dieses Vorgehen soll als σ -basierte Maskenskalierung bezeichnet werden, da bei der Ableitungsskalierung außer den vorläufigen Maskenkoeffizienten auch $f^{(0)}(0, \sigma)$, also $(\sqrt{2\pi} \sigma)^{-1}$, in die Skalierung eingeht.

modifizierte σ -basierte Skalierung Ebenfalls in Definition 2.1 gibt [Otte 94] als Skalierungsformel (2.11) – umformuliert mit den hier verwendeten Bezeichnungen – an:

$$\tilde{k}^{(1)}(n, \sigma) = k^{(1)}(n, \sigma) \frac{-f^{(0)}(\frac{1}{2}, \sigma)}{\sum_{i=1}^N k^{(1)}(i, \sigma)}. \quad (3.29)$$

Dabei wird im Gegensatz zur obigen Formel bei der Skalierung der Anteil der Ableitung der Gaußfunktion zwischen $-1/2$ und $1/2$ vernachlässigt. Diese Skalierung wird im Folgenden als *modifizierte σ -basierte Maskenskalierung* bezeichnet. Otte verwendet bei den vorgestellten Versuchen diese Variante.

Skalierung nach Korn [Korn 88] schlägt eine Skalierung der zur Ableitungsbestimmung verwendeten ersten Ableitung der Gaußfunktion vor, deren Hauptziel nicht die Korrektur des durch die Beschränkung der Maskengröße verursachten Fehlers ist, sondern die „Vereinheitlichung“ der Filterantwort bei verschiedenen Skalen σ . Die Grundforderung von Korn besteht darin, dass bei der Faltung mit Gaußglocken-basierten Ableitungsoperatoren die Filterantwort an Grauwertkanten unabhängig vom Skalenfaktor σ der verwendeten Gaußglocke sein soll. Erreicht wird dies durch Multiplikation der Maskenkoeffizienten mit dem Korrekturfaktor $\mathcal{K}(\sigma) = \sigma\sqrt{2\pi}$.

Dieser Ansatz ist verallgemeinert auch auf den hier untersuchten Fall anwendbar. Die zugrundeliegende Basisoperation, nämlich die „Normierung“ der Fläche unter dem positiven bzw. negativen Teil der Gaußfunktion auf 1 bzw. -1 kann bei diskreten, abgeschnittenen Masken durch Wahl des Korrekturfaktors

$$\tilde{\mathcal{K}}(\sigma) = \sum_{i=-N}^{-1} k^{(1)}(i, \sigma) \quad (3.30)$$

erreicht werden. Die Ableitungsmaskenkoeffizienten $\tilde{k}^{(1)}(n, \sigma)$ ergeben sich dann aus

$$\tilde{k}^{(1)}(n, \sigma) = k^{(1)}(n, \sigma)\tilde{\mathcal{K}}(\sigma). \quad (3.31)$$

Korn weist auf eine weitere interessante Eigenschaft seines Ansatzes hin: Berechnet man die Gradienten für zwei eindimensionale Funktionen $P_1(x)$ und $P_2(x)$ mit $P_2(x) = P_1(x/a)$, $a > 1$ unter Verwendung von Gaußglocken mit Standardabweichung σ_1 und $\sigma_2 = a\sigma_1$, so ergibt sich an Position x der Funktion P_1 derselbe Gradient wie an Position ax der Funktion P_2 . Dieses Ergebnis widerspricht zwar der Intuition, da die Funktion P_2 wegen $a > 1$ offensichtlich „flacher“ verläuft, es zeigt jedoch eine Invarianz des Ansatzes gegenüber einer gleichzeitigen räumlichen Stauchung oder Streckung der zugrundeliegenden Daten und der Faltungsmasken. [Alvarez et al. 93] nennen eine solche Eigenschaft (bei Skalen-Operatoren) *Skalen-Invarianz*.

Otte verwirft diesen Ansatz ohne nähere Untersuchung, da eine gleichzeitige Verwendung von ersten und höheren Ableitungen zu einer unterschiedlichen Gewichtung der ersten und der höheren Ableitungen führen würde. Bei der nachfolgenden Untersuchung soll diese Möglichkeit aber trotzdem berücksichtigt werden, da zur Bestimmung des GST nur die erste Ableitung verwendet wird.

Eine Schwäche des Ansatzes von Korn zeigt sich aber jetzt schon: Die anfangs geforderte Unabhängigkeit des Faltungsergebnisses vom Parameter σ gilt nur an (Sprung-)Kanten oder bei gleichzeitiger räumlicher Skalierung des Bildes. Die von Korn präsentierten Ergebnisse sind zwar gut, die verwendeten Bilder enthalten aber oft nur nahezu homogene Flächen mit dazwischenliegenden scharfen Kanten. Bei andersartigen Bildern sind schlechtere Ergebnisse zu erwarten.

Randkorrektur Alternativ zur σ -basierten Skalierung verwendet Otte die Randkorrektur. Wie bei der Bestimmung der Masken zur Tiefpassfilterung wird den am Maskenrand liegenden Koeffizienten die „Gesamtwichtung bis ins Unendliche“ zugewiesen. Die übrigen Maskenkoeffizienten bleiben unverändert:

$$\tilde{k}^{(1)}(n, \sigma) = \begin{cases} F^{(1)}(-\infty, -N + \frac{1}{2}, \sigma) & \text{für } n = -N \\ F^{(1)}(n - \frac{1}{2}, n + \frac{1}{2}, \sigma) & \text{für } |n| < N \\ F^{(1)}(N - \frac{1}{2}, \infty, \sigma) & \text{für } n = +N \end{cases}. \quad (3.32)$$

Die Randkorrektur ist nur für integrierte Masken definiert, da es nicht sinnvoll erscheint, die „mittleren“ Maskenkoeffizienten durch Abtastung, die „äußeren“ Maskenkoeffizienten aber durch Integration zu gewinnen.

Mittelkorrektur Auch nach der Randkorrektur entspricht die Summe der Beträge der integrierten Maskenkoeffizienten nicht der (jeweils positiven) Gesamtfläche zwischen x -Achse und erster Ableitung der Gaußfunktion.

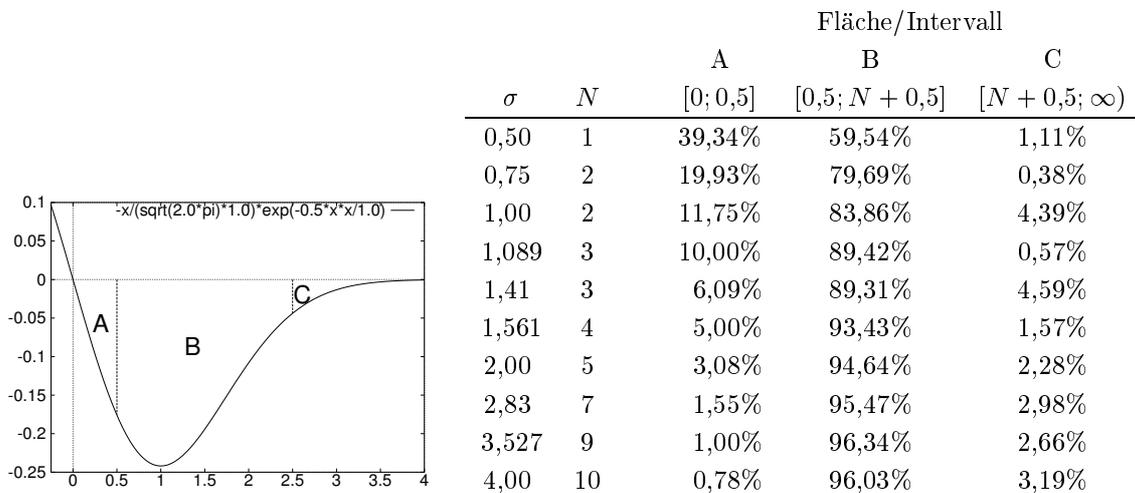


Tabelle 3.3: Prozentuale Verteilung der Fläche zwischen x -Achse und erster Ableitung der Gaußfunktion auf die angegebenen Intervalle in Abhängigkeit vom Skalenparameter σ und der bei vorgegebenem maximalem theoretischen Maskenfehler von 5% automatisch bestimmten Maskenausdehnung N . Untersucht wurden gebräuchliche Werte für σ sowie die Werte von σ , für die der Anteil der Fläche zwischen $x = 0$ und $x = 1/2$ an der Gesamtfläche im positiven Teil der x -Achse 1, 5 und 10% beträgt. Man erkennt, dass für kleine σ bei integrierten Ableitungsmasken der Fehler im Innern der Maske (A) viel größer ist als am Rand (C).

Der Grund dafür ist offensichtlich: Der Teil der Fläche zwischen 0 und $1/2$ bzw. zwischen 0 und $-1/2$ wird nicht berücksichtigt. Dadurch ergibt sich ein Fehler von

$$\left| \int_0^{1/2} f^{(1)}(\xi, \sigma) d\xi \right| = f^{(0)}(0, \sigma) - f^{(0)}(1/2, \sigma) \quad . \quad (3.33)$$

Tabelle 3.3 gibt an, wieviel Prozent der Fläche zwischen x -Achse und erster Ableitung der Gaußfunktion zwischen $-1/2$ und $1/2$ liegen und somit bisher nicht berücksichtigt werden. Bei eindimensionalen Faltungen sind dies für $\sigma < 1,561$ mehr als fünf Prozent und damit mehr als der (in den bisherigen Versuchen) maximal tolerierte Fehler durch die endliche Maskenausdehnung. Insbesondere ist dieser Fehler in der Mitte der Maske für die von Otte untersuchten Fälle ($\sigma = 1,0$ oder $\sigma = 1,5$) größer als am Rand der Maske.

Um den Fehler zu korrigieren wird folgendes Vorgehen gewählt: Die Fläche zwischen 0 und $1/2$ wird dem Maskenkoeffizienten bei $x = 1$, die Fläche zwischen 0 und $-1/2$ dem

Koeffizienten bei $x = -1$ hinzugerechnet. Es gilt also:

$$\tilde{k}^{(1)}(n, \sigma) = \begin{cases} F^{(1)}(-\infty, -N + \frac{1}{2}, \sigma) & \text{für } n = -N \\ F^{(1)}(n - \frac{1}{2}, n + \frac{1}{2}, \sigma) & \text{für } -N < n < -1 \\ F^{(1)}(-\frac{3}{2}, 0, \sigma) & \text{für } n = -1 \\ 0 & \text{für } n = 0 \\ F^{(1)}(0, \frac{3}{2}, \sigma) & \text{für } n = 1 \\ F^{(1)}(n - \frac{1}{2}, n + \frac{1}{2}, \sigma) & \text{für } 1 < n < N \\ F^{(1)}(N - \frac{1}{2}, \infty, \sigma) & \text{für } n = +N \end{cases} . \quad (3.34)$$

Dieses Verfahren wird als *Randkorrektur mit Mittelkorrektur* oder kurz als *Mittelkorrektur* bezeichnet. Aus demselben Grund wie bei der Randkorrektur ist dieses Verfahren nur für integrierte Masken definiert.

Gradienten-orientierte Skalierung Sowohl die beiden Versionen der σ -basierten Skalierung als auch Rand- oder Mittelkorrektur haben einen wesentlichen Fehler: In allen Fällen wird nur das Gesamtgewicht der vernachlässigten *Koeffizienten* auf die verbliebenen Maskenkoeffizienten verteilt.

Bei der modifizierten σ -basierten Skalierung und der Randkorrektur sind die bei der Korrektur angepassten Koeffizienten aber näher am Maskenmittelpunkt als die vernachlässigten Koeffizienten. Dies führt dazu, dass die Unterschätzung der Ableitung auch nach der Korrektur fortbesteht, wie folgender Gedankengang verdeutlicht:

Wir betrachten eine Grauwertfunktion mit der Steigung $m > 0$, d. h. mit wachsendem x nimmt der Grauwert von Pixel zu Pixel um m zu. Ferner nehmen wir an, dass die Faltung mit einer unendlich ausgedehnten Maske als Ableitung genau m berechnet – Diskretisierungseffekte bleiben außer Acht. Durch Beschränkung der Maskenausdehnung auf N werden die Maskenkoeffizienten an den Positionen $N + 1, N + 2, \dots$ sowie $-(N + 1), -(N + 2), \dots$ auf 0 gesetzt. Als Ableitungsschätzung m_{est} ohne Korrektur ergibt sich dann:

$$m_{\text{est}} = m - \underbrace{\sum_{i=-\infty}^{-(N+1)} \underbrace{-m \cdot i \cdot k^{(1)}(i, \sigma)}_{>0}}_{>0} - \underbrace{\sum_{i=N+1}^{\infty} \underbrace{-m \cdot i \cdot k^{(1)}(i, \sigma)}_{>0}}_{>0} < m, \quad (3.35a)$$

was sich bei antisymmetrischen Masken, also $k^{(1)}(-i, \sigma) = -k^{(1)}(i, \sigma)$, vereinfachen lässt zu:

$$m_{\text{est}} = m - 2 \cdot \underbrace{\sum_{i=N+1}^{\infty} \underbrace{-m \cdot i \cdot k^{(1)}(i, \sigma)}_{>0}}_{>0} < m. \quad (3.35b)$$

Bei *Randkorrektur* wird zu diesem Wert noch ein Korrekturterm addiert, der dem Funktionswert beim äußersten Maskenpunkt ($\mp m \cdot N$) multipliziert mit der Summe der vernachlässigten Koeffizienten $\pm \sum_{i=N+1}^{\infty} k^{(1)}(i, \sigma)$ entspricht, also:

$$m_{\text{est,RK}} = m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot (-m \cdot N) \cdot \sum_{i=N+1}^{\infty} k^{(1)}(i, \sigma) < m \quad (3.36a)$$

$$= m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot N \cdot k^{(1)}(i, \sigma) < m \quad (3.36b)$$

$$= m - 2 \cdot \underbrace{\sum_{i=N+1}^{\infty} \underbrace{\overbrace{-m \cdot (i-N) \cdot k^{(1)}(i, \sigma)}^{<0} \cdot \overbrace{k^{(1)}(i, \sigma)}^{>0}}_{>0}}_{>0} < m. \quad (3.36c)$$

Bei der *modifizierten σ -basierten Skalierung* werden (verallgemeinert) den Maskenkoeffizienten Korrekturterme $k_+^{(1)}(i, \sigma, N)$ aufaddiert, deren Summe (jeweils für die negative und positive x -Achse) gleich der Summe der vernachlässigten Koeffizienten ist:

$$m_{\text{est},\sigma\text{-mod}} = m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot \sum_{i=1}^N -m \cdot i \cdot k_+^{(1)}(i, \sigma, N) \quad (3.37a)$$

$$< m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot \sum_{i=1}^N -m \cdot N \cdot k_+^{(1)}(i, \sigma, N) \quad (3.37b)$$

$$= m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot (-m \cdot N) \cdot \sum_{i=1}^N k_+^{(1)}(i, \sigma, N) \quad (3.37c)$$

$$= m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot (-m \cdot N) \cdot \sum_{i=N+1}^{\infty} k^{(1)}(i, \sigma) \quad (3.37d)$$

$$= m - 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot i \cdot k^{(1)}(i, \sigma) + 2 \cdot \sum_{i=N+1}^{\infty} -m \cdot N \cdot k^{(1)}(i, \sigma) \quad (3.37e)$$

$$< m \quad [\text{vgl. Formel (3.36)}]. \quad (3.37f)$$

Man kann sich leicht klarmachen, dass die Unterschätzung sogar schon dann auftritt, wenn für wachsendes x ein streng monotoner Anstieg der Grauwerte vorausgesetzt wird und kein stets gleicher Anstieg um m .

Bei der nicht-modifizierten σ -basierten Skalierung und der Mittelkorrektur wird zusätzlich zu den vorangehend untersuchten Korrekturen die Fläche unter der abgeleiteten Gaußfunktion zwischen $-1/2$ und $1/2$ Koeffizienten zugeschlagen, die weiter vom Maskenmittelpunkt entfernt sind. Dies führt tendentiell zu einer Überschätzung der Ableitung. In seltenen Fällen werden sich beide Effekte gegenseitig aufheben, aber in der

Mehrzahl der Fälle wird bei linearem Grauwertverlauf (also einem Grauwertverlauf, der Annahme 7 – vgl. Seite 13 – entspricht) die Ableitung falsch geschätzt.

Aus diesem Grund wurde ein Ansatz entwickelt, der sicherstellt, dass bei linearen Grauwertübergängen die korrekten Ableitungen geschätzt werden. Die Forderung an die Maskenkoeffizienten muss deshalb lauten:

$$\sum_{i=-N}^{+N} (-i)\tilde{k}^{(1)}(i, \sigma) = 1. \quad (3.38)$$

Dies soll durch Skalierung aller Koeffizienten mit demselben Faktor erreicht werden, also:

$$\tilde{k}^{(1)}(n, \sigma) = \frac{k^{(1)}(n, \sigma)}{\sum_{i=-N}^N (-i)k^{(1)}(i, \sigma)}. \quad (3.39)$$

Da dieser Ansatz versucht, möglichst exakte Gradienten für Grauwertübergänge mit konstanter Steigung zu berechnen, wird er im Folgenden als gradienten-orientierte oder gradienten-basierte Skalierung bezeichnet. Im Gegensatz zur Mittel- oder Randkorrektur ist dieser Ansatz auch für abgetastete Gradienten konsistent einsetzbar, da für die Skalierung selbst keine Integration, sondern nur die Kenntnis der unskalierten Koeffizienten nötig ist. Somit ist zum Zeitpunkt der Skalierung nicht einmal die explizite Kenntnis von σ nötig, was implementierungstechnische Vorteile bietet. Insofern ist dieser Ansatz der flexibelste aller bisher vorgestellten Ansätze, um den Fehler durch die Beschränkung der Maskengröße zu korrigieren.

Gradienten-orientierte Randkorrektur Eine Abwandlung der gradienten-orientierten Skalierung besteht darin, nur die Koeffizienten der Randpixel der Maske zu verändern, so dass Formel (3.38) erfüllt wird. Dies kann erreicht werden durch:

$$\tilde{k}^{(1)}(n, \sigma) = \begin{cases} k^{(1)}(n, \sigma) & \text{für } |n| < N \\ -\frac{1}{2n} \left(1 - \sum_{i=-(N-1)}^{N-1} (-i)k^{(1)}(i, \sigma) \right) & \text{für } |n| = N \end{cases}. \quad (3.40)$$

Im Gegensatz zur von Otte vorgeschlagenen „einfachen“ Randkorrektur ist dies sowohl für abgetastete als auch für integrierte Masken und ohne explizite Kenntnis von σ möglich.

3.4 Korrespondenzen zwischen Tiefpass- und Ableitungsmasken

3.4.1 Motivation

In den vorangehenden Abschnitten wurden zwei Verfahren zur Korrektur von Tiefpassmasken, aber insgesamt sieben Verfahren zur Korrektur von Ableitungsmasken entwickelt. Die naheliegende Frage ist deshalb, welche der Ableitungs-Korrekturverfahren den beiden Verfahren zur Korrektur der Tiefpassmasken entsprechen, wenn es solche Korrespondenzen überhaupt gibt.

3.4.2 Ansatz und Umsetzung

Bisher wurden Ableitungen im Orts-/Zeitraum durch diskrete Faltungen mit Masken bestimmt, die durch Abtastung bzw. stückweise Integration der ersten partiellen Ableitungen der Gaußfunktion ermittelt wurden. Die Grundidee besteht nun darin, durch polynomiale Regression die unbekannte, unverrauschte Grauwertfunktion zu schätzen und dann Funktionswert und Ableitungen unmittelbar aus den Parametern des angepassten Polynoms abzulesen. Das Regressionspolynom sollte dabei denselben Grad wie die höchste gesuchte partielle Ableitung haben, da es kaum sinnvoll ist, z. B. ein Polynom zweiten Grades anzupassen, dann aber bei der OF-Schätzung die Ableitungen zweiten Grades zu vernachlässigen.

Macht man sich also Annahme 7 (Seite 13) zunutze, dass nämlich der Verlauf der Bildfunktion in einer hinreichend großen lokalen Umgebung des jeweils betrachteten Punktes im Wesentlichen linear ist, so ist ein linearer Regressions-Ansatz möglich. O. B. d. A. sollen zur Vereinfachung der Darstellung die Ableitungen für den Punkt im Ursprung des Koordinatensystems geschätzt werden. Seien g die Funktion der verrauschten, gemessenen Grauwerte und \hat{g} die unbekannte, „tatsächliche“ Grauwertfunktion mit ersten partiellen Ableitungen \hat{g}_x , \hat{g}_y und \hat{g}_t . $\epsilon(\mathbf{x})$ sei das (unbekannte) Rauschen der Grauwerte im Punkt \mathbf{x} . Ferner sei $\mathcal{U}(\mathbf{0})$ die Umgebung des Ursprungs $\mathbf{0}$, in der Annahme 7 erfüllt ist. Dann gilt für alle $\mathbf{x} \in \mathcal{U}(\mathbf{0})$:

$$g(\mathbf{x}) = \hat{g}(\mathbf{0}) + \begin{pmatrix} \hat{g}_x(\mathbf{0}) \\ \hat{g}_y(\mathbf{0}) \\ \hat{g}_t(\mathbf{0}) \end{pmatrix}^T \mathbf{x} + \epsilon(\mathbf{x}). \quad (3.41)$$

Um die vier Unbekannten $\hat{g}(\mathbf{0})$, $\hat{g}_x(\mathbf{0})$, $\hat{g}_y(\mathbf{0})$ und $\hat{g}_t(\mathbf{0})$ zu bestimmen, kann man für alle Punkte $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathcal{U}(\mathbf{0})$ Gleichung (3.41) aufstellen und erhält folgendes überbestimmtes Gleichungssystem in Matrixschreibweise:

$$\underbrace{\begin{pmatrix} g(\mathbf{x}_1) \\ g(\mathbf{x}_2) \\ \vdots \\ g(\mathbf{x}_n) \end{pmatrix}}_{=: \mathbf{g}} \approx \underbrace{\begin{pmatrix} 1 & x_1 & y_1 & t_1 \\ 1 & x_2 & y_2 & t_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & t_n \end{pmatrix}}_{=: A} \cdot \underbrace{\begin{pmatrix} \hat{g}(\mathbf{0}) \\ \hat{g}_x(\mathbf{0}) \\ \hat{g}_y(\mathbf{0}) \\ \hat{g}_t(\mathbf{0}) \end{pmatrix}}_{=: \hat{\mathbf{g}}}, \quad (3.42)$$

oder mit den eingeführten Abkürzungen:

$$A\hat{\mathbf{g}} \approx \mathbf{g}. \quad (3.43)$$

Da die Matrix A rauschfrei ist, kann dieses Gleichungssystem mit Hilfe der Pseudo-Inversen-Methode gelöst werden (Kleinste-Fehlerquadrate-Methode, „Least Squares“ – LS, [Van Huffel & Vandewalle 91]). Es gilt also:

$$\hat{\mathbf{g}} = \underbrace{(A^T A)^{-1} A^T}_{=: A^\dagger} \mathbf{g}. \quad (3.44)$$

A^\dagger bezeichnet die „Pseudo-Inverse“ der Matrix A ([Penrose 55]).

Wie schon beim TLS-Problem in Abschnitt 2.4.3 führt dieses Vorgehen zu einem scharfen Abschneiden am Rand des berücksichtigten Bereiches. Daher ist es sinnvoll, die einzelnen Gleichungen des Systems so zu gewichten, dass weiter vom Ursprung entfernt liegende Punkte geringeren Einfluss auf das Ergebnis haben. Es ist naheliegend, dabei wieder auf eine Gaußglocke mit Kovarianzmatrix Σ_A zurückzugreifen. Um zu erreichen, dass die Fehlerquadrate mit einem Gewicht in die Mittelung eingehen, das den abgetasteten Werten der Gaußglocke entspricht, müssen die Gleichungen mit den Quadratwurzeln der Werte der Gaußfunktion multipliziert werden. Mit

$$w(\mathbf{x}_i) = \frac{1}{(2\pi)^{3/2} \sqrt{|\Sigma_A|}} \cdot e^{-\frac{1}{2} \mathbf{x}_i^T \Sigma_A^{-1} \mathbf{x}_i}, \quad \mathbf{x}_i = (x_i, y_i, t_i)^T \quad (3.45)$$

kann man so die Gewichtungsmatrix $W_- = \text{diag}(\dots, \sqrt{w(\mathbf{x}_i)}, \dots)$ definieren und erhält dann die „gewichtete Regressionsgleichung“:

$$W_- A \hat{\mathbf{g}} \approx W_- \mathbf{g}. \quad (3.46)$$

Die Lösung lautet in diesem Fall:

$$\hat{\mathbf{g}} = ((W_- A)^T (W_- A))^{-1} (W_- A)^T W_- \mathbf{g} \quad (3.47)$$

$$= (A^T (W_-^T W_-) A)^{-1} A^T (W_-^T W_-) \mathbf{g}. \quad (3.48)$$

Wie man sieht, tritt W_{\checkmark} nur in Termen der Form $W_{\checkmark}^T W_{\checkmark}$ auf. Mit $W = W_{\checkmark}^T W_{\checkmark}$ gilt also:

$$\hat{\mathbf{g}} = \underbrace{(A^T W A)^{-1} A^T W}_{=: M} \mathbf{g}. \quad (3.49)$$

Einsetzen von W_{\checkmark} ergibt: $W = \text{diag}(\dots, w(\mathbf{x}_i), \dots)$, die Wurzeln von $w(\mathbf{x}_i)$ müssen also nicht berechnet werden.

3.4.3 Bewertung

In Gleichung (3.49) werden \hat{g} , \hat{g}_x , \hat{g}_y und \hat{g}_t als Produkt der Zeilenvektoren von M mit dem Grauwert-(Spalten-)Vektor berechnet. Die Zeilen von M können also als Faltungsmasken betrachtet werden. Der Unterschied zum bisherigen Vorgehen besteht darin, dass an keiner Stelle die Ableitung der Gaußfunktion explizit bestimmt wird.

Eine weitere wichtige Eigenschaft des hier beschriebenen Vorgehens besteht darin, dass die Korrektur der Masken „automatisch“ erfolgt: Betrachtet man in Gleichung (3.42) einen Bereich $\mathcal{U}(\mathbf{0})$, der symmetrisch zum Ursprung ist (also wie bisher bei allen Faltungsmasken) und verwendet man eine symmetrische Gewichtungsfunktion, so hat $\tilde{A} = A^T W A$ Diagonalgestalt:

$$\tilde{A} = A^T W A = \text{diag}(\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4) \quad (3.50)$$

mit

$$\tilde{a}_1 = \sum_{(x_i, y_i, t_i) \in \mathcal{U}(\mathbf{0})} w(x_i, y_i, t_i), \quad (3.51)$$

$$\tilde{a}_2 = \sum_{(x_i, y_i, t_i) \in \mathcal{U}(\mathbf{0})} x_i x_i w(x_i, y_i, t_i), \quad (3.52)$$

$$\tilde{a}_3 = \sum_{(x_i, y_i, t_i) \in \mathcal{U}(\mathbf{0})} y_i y_i w(x_i, y_i, t_i), \quad (3.53)$$

$$\tilde{a}_4 = \sum_{(x_i, y_i, t_i) \in \mathcal{U}(\mathbf{0})} t_i t_i w(x_i, y_i, t_i). \quad (3.54)$$

Für die erste Zeile von $M = (m_{ij})$ gilt also:

$$m_{1j} = \left(\sum_{(x_i, y_i, t_i) \in \mathcal{U}(\mathbf{0})} w(x_i, y_i, t_i) \right)^{-1} \cdot w(x_j, y_j, t_j), \quad (3.55)$$

was der skalierten, abgetasteten Tiefpassmaske (vgl. Gleichung (3.22), Seite 38) entspricht. Analog lässt sich herleiten, dass die zweite bis vierte Zeile von M den abgetasteten Ableitungsmasken nach gradienten-orientierter Skalierung entsprechen.

Somit bildet die *gradienten-orientierte Skalierung* der Ableitungsmasken die „natürliche“ Ergänzung zu den *skalierten* Tiefpassmasken.

Um die Frage zu beantworten, welches Korrekturverfahren für Ableitungsmasken der *Randkorrektur* von Tiefpassmasken entspricht, kann man die Gewichtungsfaktoren $w(\mathbf{x}_i)$ so wählen, dass diese der Tiefpassmaske nach Randkorrektur entsprechen. Durch Nachrechnen erkennt man, dass sich bei den Ableitungsmasken dann ein Mittel zwischen *gradienten-basierter Skalierung* und *gradienten-basierter Randkorrektur* ergibt.

Diese Aussagen gelten so nur für abgetastete Masken, da sich nur in diesem Fall die Koeffizienten der Ableitungsmaske aus den Koeffizienten der Tiefpassmaske durch Multiplikation mit den Koordinaten des Abtastpunktes (und mit dem für alle Punkte konstanten Faktor $-1/\sigma^2$) ergeben. Bei integrierten Masken hingegen zeigen sich geringe Unterschiede zwischen den Ableitungsmasken, die durch stückweise Integration der Ableitung der Gaußfunktion bzw. durch das hier vorgeschlagene Verfahren ermittelt werden.

Die Ermittlung der Ableitungsmasken mit Hilfe von Formel (3.49) hat – neben der Explizierung der Zusammenhänge zwischen Tiefpass- und Ableitungsmasken – noch weitere Vorteile:

- Maskenbestimmung und Maskenkorrektur werden in einer Formel zusammengefasst. Der Teilausdruck $A^T W$ entspricht der Maskenbestimmung, die Multiplikation mit $(A^T W A)^{-1}$ entspricht der Korrektur.
- Es können gezielt Pixel bei der Ableitungsberechnung ausgespart werden, indem der entsprechende Gewichtungsfaktor auf Null gesetzt wird; die Ableitungsmasken bleiben trotzdem ohne weitere Vorkehrungen konsistent. Ein solches Vorgehen kann in mehreren Fällen sinnvoll sein:
 - zur Wahrung der zeitlichen Kausalität: Blendet man alle Pixel aus, die Bildern nach dem aktuellen betrachteten Zeitpunkt entsprechen, so kann trotzdem noch eine Ableitung in Zeitrichtung berechnet werden. Versuche mit diesem Ansatz zeigten jedoch, dass dadurch – wie bei einer Reduzierung der Anzahl der betrachteten Punkte auch nicht anders zu erwarten – der Fehler der darauf basierenden OF-Schätzung im Vergleich zum sonst üblichen Vorgehen zunimmt.
 - zur Ableitungsberechnung nahe am Bildrand: Mit diesem Ansatz können Ableitungen bis an den Bildrand berechnet werden, der Fehler nimmt aber – erwartungsgemäß – bei abnehmender Maskengröße zu.
 - zum Ausblenden defekter Abtastpunkte: Vor allem bei Digital-Kameras für den Massenmarkt oder bei Industrie-Kameras mit hoher Grauwertdynamik treten mehr oder weniger viele Abtastpunkte auf, die aufgrund von Defekten der Wandlerfläche immer denselben oder immer einen stark fehlerhaften Abtastwert liefern. Solche Punkte können nun ausgeblendet werden.

- zur Arbeit auf zeilenversetzten Halbbildern: Handelsübliche Digitalkameras nach NTSC- oder PAL-Standard produzieren zeilenversetzte Halbbilder, d. h. zu jedem Aufnahmezeitpunkt ist nur jede zweite Zeile vorhanden. [Otte 94] bestimmt daher Faltungsmasken für zeilenversetzte Bilder durch Halbierung der zugrunde liegenden Standardabweichung in y -Richtung und anschließende Streckung der Masken in y -Richtung. Der hier vorgeschlagene Ansatz deckt hingegen den Fall zeilenversetzter Bilder ohne Sonderbehandlung mit ab.
- Sollten sich in einem besonderen Fall die Gaußlocken-basierten Faltungsmasken als unangemessen erweisen, so muss nur die (Tiefpass-)Gewichtungsfunktion ausgetauscht werden, die entsprechenden Ableitungsmasken ergeben sich „automatisch“.

Der Haupt-Nachteil des hier skizzierten Verfahrens besteht in der mangelnden Erfahrung bei dessen Anwendung. Versuche auf der zeilenversetzten Bildfolge `kwbB` zeigten, dass dort die Anwendung dieses Verfahrens nur geringe Unterschiede zum Ansatz von [Otte 94] ergibt.

3.4.4 Anmerkungen

Anwendung bei höheren Ableitungen Wie die vorangehenden Ausführungen gezeigt haben, ergibt die „gewichtete lineare Regression“ im Wesentlichen dieselben Tiefpass- und Ableitungsmasken wie die Abtastung der Gaußfunktion und ihrer ersten Ableitung. Ab der zweiten Ableitung ist dies nicht mehr der Fall, wie folgendes einfaches Beispiel zeigt. Gegeben seien folgende Werte einer eindimensionalen Funktion:

x	-3	-2	-1	0	1	2	3
$g(x)$	9	4	1	0	1	4	9

Wendet man auf diese Funktion am Punkt $x = 0$ Tiefpass- und 1. sowie 2. Ableitungsmasken der Breite 7 an („Standardverfahren“), die durch Abtastung einer Gaußglocke mit $\sigma = 1,0$ bzw. deren Ableitungen entstanden sind, so erhält man $\bar{g} = 0,996$, $g_x = 0,000$ und $g_{xx} = 1,934$. Führt man jedoch eine „gewichtete Regression“ mit Anpassung eines Polynoms zweiten Grades durch, so erhält man $\bar{g} = 0,000$, $g_x = 0,000$ und $g_{xx} = 1,000$.

Die Unterschiede lassen sich auf zwei Weisen begründen:

- Beim Standardverfahren werden durch die implizite Tiefpassfilterung die hohen Frequenzen gedämpft, was bei der gewichteten Regression nicht der Fall ist.
- Ab der zweiten Ableitung hat die Matrix A^TWA nicht mehr Diagonalgestalt, d. h. die Masken beeinflussen sich gegenseitig. Beim Standardverfahren sind die Masken unabhängig voneinander.

Da noch keinerlei Erfahrungen mit diesem Verfahren existieren, sollte es nur zur Bestimmung der ersten, aber keiner höheren partiellen Ableitungen verwendet werden.

Abgrenzung zu bekannten Verfahren Das vorgestellte Verfahren weist Parallelen zum Ansatz von [Van den Boomgaard & van der Weij 01] auf: In beiden Fällen wird der Verlauf der kontinuierlichen Bildfunktion geschätzt. Der wesentliche Unterschied besteht darin, dass beim hier vorgeschlagenen Verfahren nicht stückweise interpoliert wird, sondern an jedem Punkt die jeweils „optimale“ den Annahmen entsprechende Funktion angepasst wird. D. h. es wird nicht der „Umweg“ über eine Interpolation der diskreten Bildfunktion gegangen, sondern die Linearitätsannahme wird direkt in der Faltung ausgenutzt.

3.5 Fazit zur Maskenbestimmung

Aufgrund der durchgeführten Analyse ist nun die genaue Abschätzung der zum Unterschreiten einer vorgegebenen Fehler-Schwelle notwendigen Maskengröße im Dreidimensionalen möglich. Zur Korrektur des durch Diskretisierung und Beschränkung der Maskengröße verursachten Fehlers wurden mehrere – bekannte und neue – Korrekturverfahren vorgestellt und theoretisch analysiert. Durch Versuche muss nun geklärt werden, ob die tatsächlichen Ergebnisse auch den Erwartungen entsprechen.

Kapitel 4

Versuche zur Bewertung der Verfahren zur Maskenbestimmung

In diesem Kapitel werden Versuche beschrieben, die die Ergebnisse beim Einsatz der verschiedenen Faltungsmasken vergleichen. In Abschnitt 4.1 werden die Unterschiede bei Glättung und Ableitungsbestimmung von eindimensionalen Funktionen untersucht. Abschnitt 4.2 vergleicht die Empfindlichkeit der verschiedenen Masken gegenüber veräuschten Eingangsdaten. In Abschnitt 4.3 werden schließlich die Ergebnisse bei der Anwendung der Masken zur Ableitungs- und OF-Bestimmung auf einer Bildfolge verglichen.

4.1 Vergleich der Ergebnisse bei eindimensionalen Funktionen

[Otte 94] vergleicht an einer Stufenkante, d. h. an einem sprunghaften Grauwertübergang, einem Grauwertsprung von 0 auf 255 die theoretisch erwarteten und die mit verschiedenen Verfahren tatsächlich erzielten Ergebnisse für eine Tiefpassfilterung sowie die erste bis vierte Ableitung jeweils mit Gauß-basierten Faltungsmasken. Dabei werden die besten Ergebnisse bei integrierten Maskenkoeffizienten mit Randkorrektur erzielt. Dies ist aber nicht weiter verwunderlich, weil bedingt durch die abschnittsweise konstanten Grauwerte die Faltung mit einer *unendlich ausgedehnten* diskretisierten, durch stückweise Integration bestimmten Maske *exakt* mit einer Faltung mit der kontinuierlichen Gaußfunktion übereinstimmt und weil darüber hinaus bei einer Stufenkante die implizite Annahme der Randkorrektur „automatisch“ erfüllt ist, der äußerste berücksichtigte Punkt repräsentiere die Grauwertstruktur weiter außerhalb.

In diesem Abschnitt sollen deshalb alle im vorangehenden Kapitel vorgestellten Verfahren zur Maskenbestimmung unter Beschränkung auf die Tiefpassfilterung und die erste Ableitung erneut miteinander verglichen werden. Ziel ist es herauszufinden, ob

integrierte oder abgetastete Maskenkoeffizienten bessere Resultate liefern, und ob mit kleineren, aber „korrigierten“ Masken ähnliche Ergebnisse erzielt werden können wie mit größeren unkorrigierten Masken.

Die Vergleiche werden an drei verschiedenen Grauwertstrukturen durchgeführt:

- an einem Grauwertsprung: In realen Bildfolgen treten sprunghafte Grauwertänderungen z. B. an den Grenzen von Objektabbildern regelmäßig auf. Trotz der Verletzung von Annahme 7 – vgl. Seite 13 – sollten die Ergebnisse möglichst gut sein, damit die Faltungsergebnisse „universell“ verwendbar sind. Ein direkter Vergleich der Ergebnisse mit [Otte 94] ist möglich.
- an einer Sinus-Funktion: Die Sinus-Funktion wurde beispielhaft als eine Struktur betrachtet, die weder linear verläuft, noch abrupte Sprünge enthält. Bei der Ermittlung der Grauwerte wurden zwei Fälle unterschieden: Die Grauwerte können durch Abtastung oder durch stückweise Integration der kontinuierlichen Sinus-Funktion gewonnen werden.
- an einem linearen Grauwertübergang: Dies ist der „Idealfall“, wenn Annahme 7 erfüllt ist. Hier sollte jedes korrekt entworfene Verfahren zur Glättung oder Ableitungsbestimmung gute Ergebnisse liefern.

Als Parameter wurden zwecks Vergleichbarkeit der Resultate die von [Otte 94] verwendeten Werte übernommen: Standardabweichung $\sigma = 1,0$ und Maskenbreite von 7 Pixeln. Durch die Korrekturverfahren sind nur geringe Änderungen der Ergebnisse im Vergleich zur Verwendung der nicht-korrigierten Masken zu erwarten, da der Maskenfehler bei der betrachteten Kombination von $\sigma = 1,0$ und Maskenbreite 7 sowohl beim Tiefpass als auch bei der Ableitung sowieso schon unter 1% liegt. Daher wurde beim linearen Grauwertverlauf bei gleicher Maskenbreite auch mit größeren Werten für σ experimentiert.

4.1.1 Tiefpassfilterung

Bei der Tiefpassfilterung wurden sechs verschiedene Methoden zur Maskenbestimmung untersucht: Durch Abtastung (A) oder Integration (I) gewonnene vorläufige Koeffizienten können unverändert übernommen (A-, I-) oder einheitlich skaliert werden (AS, IS), oder es kann eine Randkorrektur (AR, IR) vorgenommen werden. Zum Vergleich wird auch das Ergebnis bei Verbreiterung der Masken um ein Pixel in jede Richtung (A+1, I+1), also auf 9 Pixel Maskenbreite angegeben.

Die Ergebnisse dieser Untersuchung sind in Tabellen 4.1, 4.2 und 4.3 dargestellt:

- Beim Grauwertsprung (Tabelle 4.1) liefert bei *integrierten* Maskenkoeffizienten die skalierte Maske in allen Fällen bessere Ergebnisse als die unskalierte und die Randkorrektur bessere Ergebnisse als die Skalierung. Beides war aufgrund

$\sigma = 1,00$		abgetastet				integriert			
x	kont.*	A-*	AS	AR	A+1	I-	IS*	IR*	I+1
-4	0,0009	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3	0,0593	0,0000	0,0000	0,0000	0,0341	0,0000	0,0000	0,0000	0,0585
-2	1,5835	1,1301 [†]	1,1304	1,1646	1,1643	1,5241	1,5249	1,5835	1,5826
-1	17,0358	14,8978	14,9019	14,9323	14,9319	16,9765	16,9844	17,0358	17,0350
0	78,6771	76,6004	76,6211	76,6349	76,6345	78,6178	78,6543	78,6771	78,6762
1	176,3229	178,3306	178,3789	178,3651	178,3648	176,2636	176,3457	176,3229	176,3221
2	237,9642	240,0332	240,0981	240,0677	240,0673	237,9048	238,0156	237,9642	237,9633
3	253,4165	253,8009	253,8696	253,8354	253,8350	253,3572	253,4751	253,4165	253,4157
4	254,9407	254,9310	255,0000	255,0000	254,9651	254,8814	255,0000	255,0000	254,9398
5	254,9991	254,9310	255,0000	255,0000	254,9992	254,8814	255,0000	255,0000	254,9983

Tabelle 4.1: Vergleich der Faltungsergebnisse bei Tiefpassfilterung an einer *Stufenkante* mit einem Übergang von 0 auf 255 an der Stelle $x = 0,5$. Die erste Spalte zeigt das Ergebnis bei Faltung mit einer kontinuierlichen, unendlich ausgedehnten Gaußfunktion, die übrigen Spalten zeigen die Ergebnisse bei mit verschiedenen Verfahren bestimmten endlichen Masken: Koeffizienten unverändert (A-, I-), Koeffizienten skaliert (AS, IS), Randkorrektur (AR, IR), Maske beiderseits um 1 Pixel vergrößert (A+1, I+1); Details siehe Text. Maskenbreite 7 (bei A+1 und I+1: 9), Standardabweichung $\sigma = 1,0$. Alle mit * markierten Spalten aus [Otte 94], der mit † markierte Wert wurde korrigiert.

der Vorüberlegungen zu erwarten. Die Verbesserung durch Skalierung und Randkorrektur ist aber jeweils nur sehr gering. Bei der Randkorrektur ergeben sich im durch die Maskenbreite abgedeckten Bereich sogar „perfekte“ Ergebnisse, da ein Grauwertsprung exakt den impliziten Annahmen dieses Verfahrens entspricht und dadurch die diskrete Faltung mit der kontinuierlichen Faltung übereinstimmt. Die *abgetasteten* Masken führen hingegen in fast allen Fällen zu Ergebnissen, die deutlich stärker von den Ergebnissen einer kontinuierlichen Faltung abweichen als bei integrierten Masken. Skalierung und Randkorrektur führen manchmal zu keiner Verbesserung, sondern zu einer Verschlechterung der Ergebnisse.

- Ein gegensätzliches Ergebnis zeigt sich bei der Sinus-Funktion (Tabelle 4.2). In diesem Fall sind die Ergebnisse mit der abgetasteten Maske in allen Fällen besser als mit der integrierten Maske. Etwas überraschend nimmt der Fehler insgesamt leicht zu, wenn die Grauwerte nicht durch Abtastung, sondern durch stückweise Integration der Sinus-Funktion gewonnen werden. Eine weitere Besonderheit dieses Testfalls liegt darin, dass Skalierung und Randkorrektur bei den abgetasteten Masken zu einer (geringfügigen) Verschlechterung der Ergebnisse führen. Ein möglicher Grund für diese Beobachtung liegt darin, dass die Sinus-Funktion im von der Maske erfassten Bereich nicht monoton ist und somit eigentlich eine kleinere Skala gewählt werden müsste.
- Beim linearen Grauwertverlauf (Tabelle 4.3) führen Skalierung und Randkorrek-

$\sigma = 1,00$		abgetastet				integriert			
x	kont.	A-	AS	AR	A+1	I-	IS	IR	I+1
Sinus-Funktion abgetastet									
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	35,4285	35,4286	35,4381	35,4325	35,4286	35,2014	35,2178	35,2082	35,2014
2	65,4633	65,4634	65,4811	65,4707	65,4634	65,0437	65,0740	65,0563	65,0437
3	85,5320	85,5321	85,5552	85,5417	85,5321	84,9837	85,0232	85,0001	84,9837
4	92,5791	92,5793	92,6043	92,5896	92,5793	91,9857	92,0285	92,0035	91,9857
Sinus-Funktion stückweise integriert									
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	35,4285	35,2013	35,2109	35,2053	35,2013	34,9756	34,9919	34,9824	34,9756
2	65,4633	65,0436	65,0612	65,0509	65,0436	64,6266	64,6567	64,6391	64,6266
3	85,5320	84,9836	85,0066	84,9931	84,9836	84,4387	84,4780	84,4550	84,4387
4	92,5791	91,9856	92,0105	91,9959	91,9856	91,3958	91,4383	91,4135	91,3958

Tabelle 4.2: Wie Tabelle 4.1, aber Vergleich der Faltungsergebnisse bei Tiefpassfilterung einer *Sinusschwingung* mit Amplitude 100 und Periodenlänge 16. Es ergeben sich unterschiedliche Werte je nachdem, ob die Sinus-Funktion an den Rasterpunkten x abgetastet wird (oben) oder ob der Grauwert am Rasterpunkt x durch Integration der zugrundeliegenden Sinus-Funktion im Intervall $[x - 1/2, x + 1/2]$ gewonnen wird (unten). Die erste Spalte zeigt das Ergebnis bei Faltung einer *kontinuierlichen* Sinus-Funktion mit einer kontinuierlichen, unendlich ausgedehnten Gaußfunktion.

$x = 10$		abgetastet				integriert			
σ	kont.	A-	AS	AR	A+1	I-	IS	IR	I+1
1,00	10,000	9,99729	10,000	10,000	9,99997	9,99535	10,000	10,000	9,99993
1,12	10,000	9,98756	10,000	10,000	9,99966	9,98222	10,000	10,000	9,99941
1,26	10,000	9,95648	10,000	10,000	9,99751	9,94527	10,000	10,000	9,99645
1,41	10,000	9,88760	10,000	10,000	9,98879	9,86945	10,000	10,000	9,98585
1,50	10,000	9,82560	10,000	10,000	9,97755	9,80369	10,000	10,000	9,97300

Tabelle 4.3: Wie Tabelle 4.1, aber Vergleich der Faltungsergebnisse bei Tiefpassfilterung eines *linearen Grauwertübergangs* mit $g(x) = x$. Im Gegensatz zu den anderen Versuchen werden nur ein Punkt ($x = 10$), aber unterschiedliche Skalen σ betrachtet.

tur sowohl für abgetastete als auch für integrierte Masken zum korrekten Ergebnis. Auch dies ist nicht überraschend, da bei einer solchen Grauwertstruktur *jede* symmetrische Maske, deren Koeffizienten sich zu 1 aufsummieren, zu diesem Ergebnis führen würde. Ohne Korrektur wird, wie bereits wiederholt dargestellt, ein zu kleiner Wert berechnet. Der Fehler nimmt bei wachsendem σ zu, da bei konstanter Maskengröße dadurch der vernachlässigte Teil der Gaußlocke wächst. In allen Fällen ohne Korrektur liefern die abgetasteten Maske etwas bessere Ergebnisse als die integrierten Masken.

keine Korrektur	Korrektur des Maskenfehlers						
	... durch Betrachtung der Maske					... durch Betrachtung des Ergebnisses der Anwendung der Maske	
	... und Skalierung			... und Randkorr.		... und Skalierung	... und Randkorr.
	... mit $[-\frac{1}{2}, \frac{1}{2}]$... ohne $[-\frac{1}{2}, \frac{1}{2}]$	sonst.	... mit $[-\frac{1}{2}, \frac{1}{2}]$... ohne $[-\frac{1}{2}, \frac{1}{2}]$		
A-, I-	AS _σ , IS _σ	AS _{mod} , IS _{mod}	AK, IK	IM	IR	AG, IG	AGR, IGR

Tabelle 4.4: Überblick über die Methoden zur Bestimmung der Ableitungsmaske. Erklärung der Abkürzungen im Text.

Als Fazit lässt sich festhalten, dass es bei der Tiefpassfilterung von der betrachteten Grauwertstruktur abhängt, ob abgetastete oder integrierte Masken bessere Ergebnisse liefern. Wegen der sonst vor allem bei größerem σ auftretenden systematischen Unterschätzung des Glättungsergebnisses (Tabelle 4.3) sollte die Summe der Maskenkoeffizienten stets auf 1 korrigiert werden, wobei Skalierung und Randkorrektur in etwa gleichwertig sind.

4.1.2 Ableitungsbestimmung

Bei der Ableitungsbestimmung gibt es eine viel größere Zahl an alternativen Vorgehensweise als bei der Tiefpassfilterung, insgesamt vierzehn Varianten (siehe auch Tabelle 4.4):

- durch Abtastung oder Integration der zugrundeliegenden ersten partiellen Ableitung der Gaußfunktion gewonnene Maskenkoeffizienten ohne Korrekturen (A-, I-), das Referenzverfahren;
- abgetastete oder integrierte Masken mit σ -basierter Skalierung (AS_σ, IS_σ);
- abgetastete oder integrierte Masken mit modifizierter, σ -basierter Skalierung (AS_{mod}, IS_{mod});
- abgetastete oder integrierte Masken, nach der Idee von Korn skaliert (AK, IK);
- integrierte Masken mit Randkorrektur (IR);
- integrierte Masken mit Rand- und Mittelkorrektur (IM);
- abgetastete oder integrierte, gradienten-basiert skalierte Masken (AG, IG);
- abgetastete oder integrierte, gradienten-basiert randkorrigierte Masken (AGR, IGR).

$\sigma = 1,00$		abgetastet								
x	kont.*	A-*	AS $_{\sigma}$	AS $_{\text{mod}}$	AK	–	–	AG	AGR	A+1
–4	0,0041	0,0000	0,0000	0,0000	0,0000	–	–	0,0000	0,0000	0,0000
–3	0,2225	0,0000	0,0000	0,0000	0,0000	–	–	0,0000	0,0000	0,1365
–2	4,4697	3,3904	3,7235	3,2860	9,3335	–	–	3,4052	3,5756	3,5269
–1	33,0270	30,9258	33,9646	29,9737	85,1367	–	–	31,0611	31,1110	31,0623
0	89,7767	92,6283	101,7303	89,7767	255,0000	–	–	93,0337	92,8135	92,7648
1	89,7767	92,6283	101,7303	89,7767	255,0000	–	–	93,0337	92,8135	92,7648
2	33,0270	30,9258	33,9646	29,9737	85,1367	–	–	31,0611	31,1110	31,0623
3	4,4697	3,3904	3,7235	3,2860	9,3335	–	–	3,4052	3,5756	3,5269
4	0,2225	0,0000	0,0000	0,0000	0,0000	–	–	0,0000	0,0000	0,1365
5	0,0041	0,0000	0,0000	0,0000	0,0000	–	–	0,0000	0,0000	0,0000
\emptyset -%-F.	0,0	11,2	10,9	11,9	150,2	–	–	11,1	9,73	10,1
$\sigma = 1,00$		integriert								
x	kont.*	I–	IS $_{\sigma}$	IS $_{\text{mod}}$ *	IK	IR*	IM	IG	IGR	I+1
–4	0,0041	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
–3	0,2225	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,2185
–2	4,4697	4,2472	4,8246	4,2577	12,0933	4,4697	4,4697	4,2772	4,5453	4,4656
–1	33,0270	32,8045	37,2647	32,8860	93,4067	33,0270	33,0270	33,0361	33,1025	33,0229
0	89,7767	89,5541	101,7303	89,7767	254,9946	89,7767	101,7303	90,1865	89,8522	89,7726
1	89,7767	89,5541	101,7303	89,7767	254,9946	89,7767	101,7303	90,1865	89,8522	89,7726
2	33,0270	32,8045	37,2647	32,8860	93,4067	33,0270	33,0270	33,0361	33,1025	33,0229
3	4,4697	4,2472	4,8246	4,2577	12,0933	4,4697	4,4697	4,2772	4,5453	4,4656
4	0,2225	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,2185
5	0,0041	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
\emptyset -%-F.	0,0	1,97	11,4	1,72	179,1	0,00	4,44	1,60	0,67	0,04

Tabelle 4.5: Vergleich der Faltungsergebnisse bei Bestimmung der 1. Ableitung an einer *Stufenkante* mit einem Übergang von 0 auf 255 an der Stelle $x = 0,5$. Die zweite Spalte zeigt das Ergebnis bei Faltung mit einer kontinuierlichen, unendlich ausgedehnten Gaußfunktion, die übrigen Spalten zeigen die Ergebnisse bei mit verschiedenen Verfahren (siehe Text) bestimmten endlichen Masken. In der Zeile „ \emptyset -%-F.“ ist der durchschnittliche prozentuale Fehler der geschätzten Werte im Bereich -2 bis 3 angegeben. Maskenbreite 7, Standardabweichung $\sigma = 1,0$. Alle mit * markierten Spalten aus [Otte 94].

Wie schon beim vorangehenden Vergleich werden als Referenzwerte die theoretisch zu erwartenden Ergebnisse sowie die Resultate bei vergrößerten Masken (A+1, I+1) angegeben.

- Beim Grauwertsprung (Tabelle 4.5) liefern die integrierten Masken in fast allen Fällen bessere Ergebnisse als die abgetasteten Masken. Die besten Ergebnisse – gemessen an der durchschnittlichen prozentualen Abweichung der berechneten Werte von den bei kontinuierlicher Faltung auftretenden Werten – bei integrierten

Sinus-Funktion abgetastet											
$\sigma = 1,00$			abgetastet								
x	kont.	kont.TP	A-	AS_σ	AS_{mod}	AK	-	-	AG	AGR	A+1
0	39,270	36,356	36,247	39,809	35,131	99,787	-	-	36,406	36,382	36,354
1	36,281	33,588	33,488	36,779	32,457	92,191	-	-	33,635	33,612	33,587
2	27,768	25,707	25,631	28,149	24,842	70,560	-	-	25,743	25,726	25,706
3	15,028	13,913	13,871	15,234	13,444	38,187	-	-	13,931	13,923	13,912
4	0,000	0,000	0,000	0,000	0,000	0,000	-	-	0,000	0,000	0,000
\emptyset -%-F.	-	0,00	0,298	9,50	3,37	174,5	-	-	0,138	0,071	0,004
$\sigma = 1,00$			integriert								
x	kont.	kont.TP	I-	IS_σ	IS_{mod}	IK	IR	IM	IG	IGR	I+1
0	39,270	36,356	35,948	40,836	36,038	102,361	36,110	39,697	36,202	36,164	36,120
1	36,281	33,588	33,212	37,728	33,294	94,569	33,361	36,676	33,447	33,412	33,370
2	27,768	25,707	25,419	28,875	25,483	72,380	25,533	28,070	25,599	25,572	25,540
3	15,028	13,913	13,757	15,627	13,791	39,172	13,819	15,192	13,854	13,840	13,822
4	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
\emptyset -%-F.	-	0,00	1,12	12,3	0,875	181,6	0,677	9,19	0,422	0,526	0,649
Sinus-Funktion stückweise integriert											
$\sigma = 1,00$			abgetastet								
x	kont.	kont.TP	A-	AS_σ	AS_{mod}	AK	-	-	AG	AGR	A+1
0	39,270	36,356	36,015	39,554	34,906	99,147	-	-	36,172	36,148	36,121
1	36,281	33,588	33,273	36,543	32,249	91,600	-	-	33,419	33,397	33,372
2	27,768	25,707	25,466	27,969	24,682	70,107	-	-	25,578	25,561	25,542
3	15,028	13,913	13,782	15,137	13,358	37,942	-	-	13,843	13,833	13,823
4	0,000	0,000	0,000	0,000	0,000	0,000	-	-	0,000	0,000	0,000
\emptyset -%-F.	-	0,00	0,938	8,80	3,99	172,7	-	-	0,504	0,571	0,645
$\sigma = 1,00$			integriert								
x	kont.	kont.TP	I-	IS_σ	IS_{mod}	IK	IR	IM	IG	IGR	I+1
0	39,270	36,356	35,718	40,574	35,807	101,704	35,878	39,443	35,970	35,932	35,888
1	36,281	33,588	32,999	37,486	33,081	93,962	33,147	36,440	33,232	33,197	33,156
2	27,768	25,707	25,256	28,690	25,319	71,916	25,370	27,890	25,435	25,408	25,377
3	15,028	13,913	13,669	15,527	13,703	38,921	13,730	15,094	13,765	13,751	13,734
4	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
\emptyset -%-F.	-	0,00	1,75	11,6	1,51	179,7	1,31	8,49	1,06	1,16	1,29

Tabelle 4.6: Wie Tabelle 4.5, aber mit *Sinusschwingung* mit Amplitude 100 und Periodenlänge 16. Es ergeben sich unterschiedliche Werte je nachdem, ob die Sinus-Funktion an den Rasterpunkten x abgetastet wird (oben) oder ob der Grauwert am Rasterpunkt x durch Integration der zugrundeliegenden Sinus-Funktion im Intervall $[x - 0,5; x + 0,5]$ gewonnen wird (unten). Die erste Spalte („kont.“) gibt die tatsächliche Ableitung der *kontinuierlichen* Sinus-Funktion, die zweite Spalte die Ableitung der mit einer kontinuierlichen, unendlich ausgedehnten Gaußglocke tiefpassgefilterten („kont. TP“) Funktion an.

$x = 10$		abgetastet								
σ	kont.	A-	AS_σ	AS_{mod}	AK	-	-	AG	AGR	A+1
1,00	1,00	0,9956	1,0935	0,9650	2,7409	-	-	1,0000	1,0000	0,99993
1,12	1,00	0,9839	1,0636	0,9628	2,9861	-	-	1,0000	1,0000	0,99932
1,26	1,00	0,9547	1,0286	0,9507	3,2486	-	-	1,0000	1,0000	0,99603
1,41	1,00	0,9041	0,9870	0,9269	3,4885	-	-	1,0000	1,0000	0,98551
1,50	1,00	0,8660	0,9604	0,9085	3,6109	-	-	1,0000	1,0000	0,97407
$x = 10$		integriert								
σ	kont.	I-	IS_σ	IS_{mod}	IK	IR	IM	IG	IGR	I+1
1,00	1,00	0,9930	1,1280	0,9955	2,8275	0,9982	1,0920	1,0000	1,0000	0,99984
1,12	1,00	0,9782	1,0898	0,9864	3,0595	0,9944	1,0619	1,0000	1,0000	0,99888
1,26	1,00	0,9454	1,0468	0,9675	3,3060	0,9855	1,0335	1,0000	1,0000	0,99457
1,41	1,00	0,8923	0,9990	0,9382	3,5310	0,9702	1,0047	1,0000	1,0000	0,98233
1,50	1,00	0,8536	0,9697	0,9173	3,6460	0,9584	0,9872	1,0000	1,0000	0,96977

Tabelle 4.7: Vergleich der Faltungsergebnisse bei Bestimmung der 1. Ableitung eines *linearen Grauwertübergangs*. Maskenbreite 7, verschiedene Standardabweichungen.

Masken liefert erwartungsgemäß die Randkorrektur nach Otte. Auf dem zweiten Platz folgen die vergrößerten Masken, dahinter die beiden gradienten-basierten Verfahren. Diese führen auch bei den abgetasteten Masken zum zweit- und viertbesten Ergebnis, mit den vergrößerten Masken auf Platz 1 und der σ -basierten Skalierung auf dem dritten Platz. Die Verfahren, die das Intervall $[-\frac{1}{2}, +\frac{1}{2}]$ berücksichtigen (AS_σ , IS_σ , IM), führen zu viel zu großen Werten in unmittelbarer Nähe des Grauwertsprungs. Dasselbe gilt für die Skalierung nach Korn (AK, IK), doch war in diesem Fall auch nichts Anderes erwartet worden.

- Bei der Sinus-Funktion (Tabelle 4.6) kommen zwei verschiedene Referenzwerte in Frage: Die Ableitung der zugrundeliegenden kontinuierlichen Funktion („kont.“) und die Ableitung der mit der Gaußfunktion gefalteten kontinuierlichen Funktion („kont. TP“). Aufgrund des Maskenentwurfs (Ableitung der Gaußfunktion) wird der zweite dieser Werte als Referenzwert herangezogen.

Davon ausgehend sind im unskalierten Fall die Ergebnisse mit abgetasteten Masken geringfügig besser als mit integrierten Masken, wobei die Ergebnisse in beiden Fällen nahe am zu erwartenden Wert liegen. Die gradienten-basierten Korrekturverfahren (AG, IG, AGR, IGR) und die Vergrößerung der Masken führen in allen Fällen zu einer Verbesserung der Ergebnisse. Nur die Randkorrektur nach Otte führt zu einem ähnlich guten Resultat. Hingegen führen die Verfahren, die das Intervall $[x - \frac{1}{2}, x + \frac{1}{2}]$ mitberücksichtigen (AS_σ , IS_σ , AM, IM) stets zu einer deutlichen Überschätzung, die modifizierte σ -basierte Skalierung zu einer Unterschätzung der Ableitung. Bei der Skalierung nach Korn sind die Werte erwartungsgemäß vielfach größer als die korrekten Werte.

- Bei Betrachtung eines linearen Grauwertüberganges (Tabelle 4.7) zeigt sich deutlich, dass die Gradienten-basierten Korrekturverfahren (AG, IG, AGR, IGR) die einzigen unter den untersuchten sind, die eine zuverlässige Korrektur des Fehlers unabhängig von der Wahl von σ gewährleisten. Ähnlich gute Ergebnisse ergeben sich nur bei Vergrößerung der Masken.

4.1.3 Versuchsfazit

Zusammenfassend kann man feststellen, dass wie schon bei der Tiefpassfilterung die abgetasteten Masken nicht grundsätzlich schlechtere Ergebnisse produzieren als die integrierten Masken, manchmal sogar bessere. Von den untersuchten Korrekturverfahren waren die beiden gradienten-basierten Verfahren und die Vergrößerung der Faltungsmasken am erfolgreichsten, während alle anderen Verfahren beim ein oder anderen Testfall wesentliche Fehler zeigten.

4.2 Fehlerempfindlichkeit der Filterungen

Für die Auswahl der bestgeeigneten Methode zur Korrektur des durch die Beschränkung der Maskengröße verursachten Fehlers kann nicht nur die Qualität der Ergebnisse bei „idealen“, also rauschfreien Eingabedaten ausschlaggebend sein. Vielmehr muss auch berücksichtigt werden, wie sich das Verfahren bei verrauschten Eingabedaten verhält. Entsprechende Untersuchungen sollen in diesem Abschnitt durchgeführt werden.

4.2.1 Theoretische Analyse der Fehlerfortpflanzung

Die nachfolgenden Untersuchungen basieren auf der Grundannahme, dass – wie bereits als Annahme 9 (Seite 19) formuliert und dort diskutiert – in einem mit einer digitalen Kamera aufgenommenen Grauwertbild die aufgezeichneten Grauwerte $\hat{g} = \hat{g}(\mathbf{x})$ mit einem additiven normalverteilten Fehler mit Mittelwert 0 behaftet sind, d. h. $\hat{g}(\mathbf{x}) \sim N(g(\mathbf{x}), \hat{\sigma}(\mathbf{x})^2)$, wobei $g(\mathbf{x})$ den „fehlerfreien“ Grauwert bezeichne. Außerdem wird angenommen, dass die Fehler an den einzelnen Pixelpositionen voneinander unabhängig sind und dass $\hat{\sigma}(\mathbf{x})$ für alle Pixel – unabhängig von deren Position oder dem Betrag des (unbekannten) fehlerfreien Grauwerts – denselben Wert hat, also $\hat{g}(\mathbf{x}) \sim N(g(\mathbf{x}), \hat{\sigma}^2)$.

Unter denselben Grundannahmen untersuchen [Nagel 95] und [Nagel & Haag 98] die Auswirkungen der Fortpflanzung des Fehlers der Grauwerte auf die Schätzung des OFs mit dem Verfahren der Nachbarschaftsabtastung (vgl. [Otte & Nagel 95]), ohne den Zwischenschritt der Ableitungsschätzung explizit durchzuführen.

Alle Berechnungen beruhen auf zwei bekannten Formeln (siehe z. B. [Henze 95]):

1. Lineare Transformation von Zufallsvariablen:

$$X \sim N(\mu, \sigma^2) \Rightarrow \alpha X \sim N(\alpha\mu, \alpha^2\sigma^2) \text{ für } \alpha \neq 0. \quad (4.1)$$

2. Summe mehrerer unabhängiger Zufallsvariablen:

$$X_i \sim N(\mu_i, \sigma_i^2), i = 1, \dots, n \Rightarrow X_1 + \dots + X_n \sim N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2) \quad (4.2)$$

4.2.1.1 Eindimensionale Betrachtung

Zunächst sollen der Fall einer eindimensionalen Tiefpassfilterung und einer eindimensionalen Ableitungsfilterung betrachtet werden.

Bei der Tiefpassfilterung mithilfe einer Maske mit Koeffizienten α_i gilt:

$$\bar{g}(x_0) = \sum_{x_i \in \mathcal{U}_M(x_0)} \alpha_i \hat{g}(x_i) \sim N\left(\sum \alpha_i g(x_i), \left(\sum \alpha_i^2\right) \hat{\sigma}^2\right) \quad (4.3)$$

$$= N\left(\bar{g}(x_0), \left(\sum \alpha_i^2\right) \hat{\sigma}^2\right). \quad (4.4)$$

Dabei sei $\mathcal{U}_M(x_0)$ die bei der Mittelung berücksichtigte lokale Umgebung von x_0 . Die Maskenkoeffizienten α_i werden dabei aus einer Gauß-Funktion mit Standardabweichung σ gewonnen. Es gilt also bei abgetasteten Maskenkoeffizienten:

$$\sum \alpha_i^2 \approx \sum_{k=-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{k^2}{2\sigma^2}}\right)^2 \quad (4.5a)$$

$$= \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{2k^2}{2\sigma^2}} \quad (4.5b)$$

$$= \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \frac{\frac{1}{\sqrt{2}}}{\sqrt{2\pi}\left(\frac{\sigma}{\sqrt{2}}\right)} \exp\left(-\frac{k^2}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2}\right) \quad (4.5c)$$

$$= \frac{1}{2\sqrt{\pi}\sigma} \underbrace{\sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\left(\frac{\sigma}{\sqrt{2}}\right)} \exp\left(-\frac{k^2}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2}\right)}_{\approx 1} \quad (4.5d)$$

$$\approx \frac{1}{2\sqrt{\pi}\sigma}. \quad (4.5e)$$

Somit ergibt sich näherungsweise:

$$\overline{\hat{g}(x_0)} \sim N\left(\overline{g(x_0)}, \frac{1}{2\sqrt{\pi}\sigma}\hat{\sigma}^2\right). \quad (4.6)$$

Analog gilt bei der Ableitungsfilterung bei berücksichtigter Umgebung $\mathcal{U}_A(x_0)$:

$$\frac{\partial \hat{g}}{\partial x}(x_0) = \sum_{x_i \in \mathcal{U}_A(x_0)} \alpha_i \hat{g}(x_i) \sim N\left(\frac{\partial g}{\partial x}(x_0), \left(\sum \alpha_i^2\right) \hat{\sigma}^2\right), \quad (4.7)$$

mit:

$$\sum \alpha_i^2 \approx \sum_{k=-\infty}^{\infty} \left(-\frac{k}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{k^2}{2\sigma^2}}\right)^2 \quad (4.8a)$$

$$= \sum_{k=-\infty}^{\infty} \frac{k^2}{\sigma^4} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^2 \exp\left(-\frac{2k^2}{2\sigma^2}\right) \quad (4.8b)$$

$$\begin{aligned} \text{Subst.} \\ \sigma' = \frac{\sigma}{\sqrt{2}} &= \frac{1}{8} \left(\frac{1}{\sqrt{2\pi}(\sigma')}\right) \sum_{k=-\infty}^{\infty} \frac{k^2}{(\sigma')^4} \underbrace{\left(\frac{1}{\sqrt{2\pi}(\sigma')}\right) \exp\left(-\frac{k^2}{2(\sigma')^2}\right)}_{=: f(k, \sigma')} \end{aligned} \quad (4.8c)$$

$$\begin{aligned} &= \frac{1}{8} \left(\frac{1}{\sqrt{2\pi}(\sigma')}\right) \\ &\quad \left(\sum_{k=-\infty}^{\infty} \frac{1}{(\sigma')^2} f(k, \sigma') + \sum_{k=-\infty}^{\infty} \left(\frac{k^2}{(\sigma')^4} - \frac{1}{(\sigma')^2}\right) f(k, \sigma')\right) \end{aligned} \quad (4.8d)$$

$$\begin{aligned} &\approx \frac{1}{8} \left(\frac{1}{\sqrt{2\pi}(\sigma')}\right) \\ &\quad \left(\frac{1}{(\sigma')^2} \int_{-\infty}^{\infty} f(x, \sigma') dx + \int_{-\infty}^{\infty} \left(\frac{x^2}{(\sigma')^4} - \frac{1}{(\sigma')^2}\right) f(x, \sigma') dx\right) \end{aligned} \quad (4.8e)$$

$$= \frac{1}{8} \left(\frac{1}{\sqrt{2\pi}(\sigma')}\right) \left(\frac{1}{(\sigma')^2} 1 + 0\right) \quad (4.8f)$$

$$\text{Rück-} \\ \text{subst.} = \frac{1}{4\sqrt{\pi}\sigma^3}, \quad (4.8g)$$

zusammenfassend also:

$$\frac{\partial \hat{g}}{\partial x}(x_0) \sim N\left(\frac{\partial g}{\partial x}(x_0), \left(\frac{1}{4\sqrt{\pi}\sigma^3}\right) \hat{\sigma}^2\right) \quad (4.9)$$

Es ist zu beachten, dass $\hat{\sigma}$ in den Gleichungen (4.6) und (4.9) und σ in den Gleichungen (4.5a) und (4.8a) im Allgemeinen nicht denselben Wert haben. $\hat{\sigma}$ beschreibt die Standardabweichung der verrauschten Grauwerte, σ die Ausdehnung der Faltungsmaske.

4.2.1.2 Dreidimensionaler Fall

Dreidimensionale Faltungen mit abgeleiteten Gaußmasken lassen sich bei Kovarianzmatrix in Diagonalgestalt in eindimensionale Faltungen faktorisieren. Somit kann der dreidimensionale Fall einfach auf den eindimensionalen Fall zurückgeführt werden. Beispielfolgt soll die Bestimmung der partiellen Ableitung in x -Richtung im Punkt (x_0, y_0, t_0) betrachtet werden, die durch Glättung in t -Richtung, anschließende Glättung in y -Richtung und letztendlich eine Ableitungsfilterung in x -Richtung realisiert ist. Die Standardabweichung der Grauwerte sei $\hat{\sigma}$.

Die Tiefpassfilterung in t -Richtung muss an allen Punkten (x, y, t) berechnet werden, an denen $t = t_0$ gilt. Dadurch geht der Grauwert am Punkt (x, y, t) nur in das Ergebnis einer einzigen Tiefpassfilterung, nämlich am Punkt (x, y, t_0) ein. Die Ergebnisse der Tiefpassfilterung sind somit unter den getroffenen Annahmen stochastisch voneinander unabhängig und unterliegen nach Gleichung (4.6) und (4.5e) jeweils einer Normalverteilung mit Standardabweichung

$$\sqrt{\frac{1}{2\sqrt{\pi}\sigma_t}} \hat{\sigma} \quad . \quad (4.10)$$

Auch die Ergebnisse der anschließenden Tiefpassfilterung in y -Richtung (Maske gegeben durch σ_y) sind voneinander unabhängig und normalverteilt mit Standardabweichung

$$\sqrt{\frac{1}{2\sqrt{\pi}\sigma_y}} \sqrt{\frac{1}{2\sqrt{\pi}\sigma_t}} \hat{\sigma} \quad . \quad (4.11)$$

Das Endergebnis erhält man schließlich nach einer durch σ_x gegebenen Ableitungsfilterung, mit einer Standardabweichung von

$$\sqrt{\frac{1}{4\sqrt{\pi}\sigma_x^3}} \sqrt{\frac{1}{2\sqrt{\pi}\sigma_y}} \sqrt{\frac{1}{2\sqrt{\pi}\sigma_t}} \hat{\sigma} \quad . \quad (4.12)$$

Für $\sigma_x = \sigma_y = \sigma_t = \sigma$ kann dies vereinfacht werden zu:

$$\frac{1}{4} \sqrt{\frac{1}{\sqrt{\pi^3}\sigma^5}} \hat{\sigma} \quad . \quad (4.13)$$

	kont.	abgetastet				integriert			
σ	lt. Gl. (4.6)	A-	AS	AR	A+1	I-	IS	IR	I+1
Theoretische Werte									
1,00	0,2820948	0,28212	0,28228	0,28213	0,28212	0,27091	0,27117	0,27092	0,27091
1,12	0,2518704	0,25187	0,25250	0,25190	0,25187	0,24383	0,24469	0,24387	0,24383
1,26	0,2238848	0,22388	0,22584	0,22405	0,22388	0,21818	0,22058	0,21842	0,21819
1,41	0,2000672	0,20002	0,20459	0,20074	0,20007	0,19591	0,20113	0,19682	0,19598
1,50	0,1880632	0,18795	0,19468	0,18935	0,18806	0,18451	0,19197	0,18619	0,18465
Tatsächliche Werte									
1,00	–	0,28215	0,28230	0,28215	0,28215	0,27094	0,27119	0,27094	0,27094
1,12	–	0,25189	0,25252	0,25192	0,25189	0,24385	0,24471	0,24389	0,24385
1,26	–	0,22389	0,22586	0,22407	0,22389	0,21819	0,22060	0,21844	0,21819
1,41	–	0,20003	0,20460	0,20075	0,20003	0,19592	0,20114	0,19683	0,19592
1,50	–	0,18796	0,19469	0,18937	0,18796	0,18452	0,19199	0,18621	0,18452

Tabelle 4.8: Vergleich der Varianz des durch Rauschen verursachten Fehlers bei der Tiefpassfilterung. Maskenbreite 7, verschiedene Standardabweichungen.

Das bedeutet, dass für $\sigma = 1,0$ die Standardabweichung der Ergebnisse etwa ein Zehntel der Standardabweichung der Eingangsdaten beträgt, für $\sigma = 1,41$ sogar nur rund 4,5 %.

4.2.2 Fehlerfortpflanzung bei Faltungsmasken

Bei den vorangehenden Berechnungen wurden mehrere Approximationen vorgenommen, die nur für große Werte von σ bzw. σ' mathematisch gerechtfertigt sind. Darüber hinaus wurden nur abgetastete Masken betrachtet, und es wurde von einer unendlichen Maskenausdehnung ausgegangen. Bei den tatsächlich von uns verwendeten Masken sind durchaus Abweichungen zu erwarten von den Werten, die sich aus den Formeln (4.6) und (4.9) ergeben.

Für fünf ausgewählte Werte von σ wurde deshalb bei einer festen Maskengröße von sieben Pixeln ermittelt, wie groß die zu erwartende Varianz der Ergebnisse bei abgetasteten oder integrierten, mit verschiedenen Verfahren korrigierten Masken ist. Da die Varianz der Ergebnisse proportional zur Varianz der Eingangsdaten ist, wurde von standard-normalverteilten Eingangsdaten ausgegangen.

Die Ergebnisse dieser Betrachtung sind jeweils in der oberen Hälfte der Tabellen 4.8 und 4.9 dargestellt. Es fällt auf, dass die zu erwartende Varianz des Ergebnisses bei integrierten Masken etwas geringer ist als bei abgetasteten Masken, wobei die Ergebnisse bei abgetasteten Masken nur wenig von den hergeleiteten Werten abweichen.

Außerdem kann festgestellt werden, dass die Varianz durch alle Korrekturverfahren im Vergleich zum nicht-korrigierten Fall zunimmt. Die Ursache dafür liegt darin, dass bei allen Korrekturverfahren der Betrag mindestens eines Maskenkoeffizienten zunimmt,

Theoretische Werte										
kont.		abgetastet								
σ	lt. Gl. (4.9)	A-	AS $_{\sigma}$	AS $_{\text{mod}}$	AK	-	-	AG	AGR	A+1
1,00	0,1410474	0,14077	0,16980	0,13224	1,06688	-	-	0,14201	0,14081	0,14077
1,12	0,1003948	0,10037	0,11730	0,09610	0,92451	-	-	0,10368	0,10064	0,10037
1,26	0,0705104	0,07046	0,08179	0,06987	0,81584	-	-	0,07730	0,07163	0,07051
1,41	0,0503162	0,05011	0,05972	0,05267	0,74606	-	-	0,06130	0,05346	0,05031
1,50	0,0417918	0,04142	0,05093	0,04558	0,72003	-	-	0,05522	0,04670	0,04178
kont.		integriert								
σ	lt. Gl. (4.9)	I-	IS $_{\sigma}$	IS $_{\text{mod}}$	IK	IR	IM	IG	IGR	I+1
1,00	0,1410474	0,12469	0,16091	0,12531	1,01100	0,12475	0,17088	0,12646	0,12477	0,12469
1,12	0,1003948	0,09100	0,11296	0,09254	0,89027	0,09131	0,11753	0,09511	0,09142	0,09102
1,26	0,0705104	0,06516	0,07988	0,06824	0,79685	0,06626	0,08052	0,07291	0,06669	0,06524
1,41	0,0503162	0,04702	0,05895	0,05198	0,73635	0,04974	0,05758	0,05906	0,05095	0,04728
1,50	0,0417918	0,03912	0,05049	0,04518	0,71374	0,04314	0,04875	0,05369	0,04508	0,03954
Tatsächliche Werte										
		abgetastet								
σ		A-	AS $_{\sigma}$	AS $_{\text{mod}}$	AK	-	-	AG	AGR	A+1
1,00	-	0,14078	0,16981	0,13225	1,06695	-	-	0,14202	0,14082	0,14078
1,12	-	0,10038	0,11731	0,09611	0,92461	-	-	0,10369	0,10065	0,10039
1,26	-	0,07047	0,08180	0,06988	0,81596	-	-	0,07732	0,07164	0,07052
1,41	-	0,05011	0,05974	0,05268	0,74619	-	-	0,06131	0,05347	0,05032
1,50	-	0,04142	0,05094	0,04558	0,72016	-	-	0,05523	0,04671	0,04179
		integriert								
σ		I-	IS $_{\sigma}$	IS $_{\text{mod}}$	IK	IR	IM	IG	IGR	I+1
1,00	-	0,12470	0,16092	0,12532	1,01109	0,12476	0,17089	0,12647	0,12478	0,12471
1,12	-	0,09101	0,11297	0,09256	0,89038	0,09132	0,11754	0,09512	0,09143	0,09103
1,26	-	0,06517	0,07990	0,06825	0,79697	0,06627	0,08053	0,07292	0,06671	0,06525
1,41	-	0,04703	0,05896	0,05199	0,73648	0,04975	0,05759	0,05907	0,05096	0,04728
1,50	-	0,03912	0,05050	0,04519	0,71387	0,04315	0,04876	0,05370	0,04509	0,03955

Tabelle 4.9: Vergleich der Varianz des durch Rauschen verursachten Fehlers bei Bestimmung der 1. Ableitung. Maskenbreite 7, verschiedene Standardabweichungen.

aber keiner abnimmt. Ein Fehler des entsprechenden Pixels im Bild wirkt sich dadurch stärker auf das Ergebnis aus, die Varianz nimmt zu.

Darüber hinaus erkennt man, dass bei vergleichbaren Paaren von Skalierungs- und Randkorrekturverfahren (AS vs. AR, IS vs. IR beim Tiefpass, AG vs. AGR, IG vs. IGR bei der Ableitung) die Randkorrektur stets zu einer kleineren Varianz führt als die Skalierung. Dies ist auf den ersten Blick überraschend, weil bei der Randkorrektur die Gewichtung nur der Randpixel zunimmt und damit wenige (bei 1D-Masken

Art der Maske	Maske	Varianz
nicht-korrigiert	(0,2 0,5 0,2)	0,33
skaliert	($0,\bar{2}$ $0,\bar{5}$ $0,\bar{2}$)	$0,\overline{407}$
Rand-korrigiert	(0,25 0,5 0,25)	0,375
nicht-korrigiert	(0,3 0,3 0,3)	0,27
skaliert	($0,\bar{3}$ $0,\bar{3}$ $0,\bar{3}$)	$0,\bar{3}$
Rand-korrigiert	(0,35 0,3 0,35)	0,335

Tabelle 4.10: Auswirkung von Skalierung und Randkorrektur auf die Varianz des Faltungsergebnisses: Dargestellt ist die Varianz des Faltungsergebnisses, wenn alle drei Eingabewerte unabhängig voneinander normalverteilt mit Mittelwert 0 und Varianz 1 sind.

nur zwei) Pixel für die Veränderung der Varianz verantwortlich sind. Allerdings ergibt sich aus den Formeln (4.1) und (4.2), dass die Varianz des Faltungsergebnisses kleiner wird, wenn die Summe der Quadrate der Maskenkoeffizienten abnimmt. Da die von einer Randkorrektur betroffenen Maskenkoeffizienten ein insgesamt geringes Gewicht bei der Faltung haben, führt ihre Vergrößerung zu einer geringeren Zunahme der Quadratsumme – und somit der Varianz – als eine gleichmäßige Vergrößerung aller Maskenkoeffizienten. Tabelle 4.10 verdeutlicht dies an zwei einfachen Beispielen.

4.2.3 Versuch zur Ermittlung der tatsächlichen Fehler

Zur Prüfung der bisherigen Überlegungen vergleicht nun ein abschließender Versuch noch die theoretisch ermittelten Werte für die Varianz der Faltungsergebnisse mit den bei der Durchführung der Faltung tatsächlich auftretenden Werten.

Wie sich aus den Formeln (4.1), (4.2), (4.6) und (4.9) ablesen lässt, gehen die (unbekannten) tatsächlichen Grauwerte selbst in keiner Weise in die Berechnung der *Varianz* des Faltungsergebnisses ein, wenn man annimmt, dass der Fehler der Grauwerte additiv und normalverteilt mit Mittelwert 0 ist. Der Fehler der Ergebnisse ist also unabhängig von der (unbekannten) zugrundeliegenden fehlerfreien Grauwertstruktur. Zur Ermittlung des bei verschiedenen Masken zur Tiefpass- und Ableitungsbestimmung tatsächlich auftretenden Fehlers wurde deshalb ein Bild mit anfangs konstanten Grauwerten verwendet ($g(\mathbf{x}) = c \forall \mathbf{x}$), das durch punktweise Addition eines normalverteilten Fehlers mit Mittelwert 0 und Standardabweichung 1 verrauscht wurde. Die sich so ergebenden verrauschten Grauwerte wurden weder auf den sonst üblichen Wertebereich 0–255 eingeschränkt noch auf die nächstgelegene ganze Zahl gerundet, um das Ergebnis nicht dadurch zu verfälschen. Die normalverteilten Zufallszahlen wurden mithilfe des Zufallszahlengenerators der C++-Standardbibliothek und der Box-Müller-Transformation ([Press et al. 90]) erzeugt. Die Periodizität dieses Zufallszahlengenerators liegt nach Autorenangaben bei etwa $16 \cdot (2^{31} - 1) \approx 34 \cdot 10^9$, was für unsere Zwecke

ausreichend sein sollte. Bei 60 Millionen verrauschten Grauwerten, aus denen 50 Millionen Tiefpass- und Ableitungswerte gewonnen wurden, lag der tatsächliche Mittelwert des aufaddierten Fehlers bei $19,0 \cdot 10^{-6}$, die tatsächliche Varianz bei $1 + 19,31 \cdot 10^{-6}$.

Die Ergebnisse dieser Simulation sind jeweils in der unteren Hälfte der Tabellen 4.8 und 4.9 dargestellt. Man erkennt, dass Unterschiede zu den theoretischen Werten erst ab der fünften Nachkommastelle auftreten.

4.2.4 Versuchsfazit

Insgesamt lässt sich feststellen, dass die Varianz der Ergebnisse im Wesentlichen vom Skalenparameter σ und nur zu einem kleinen Teil von der Implementierung der Faltung (mit Ausnahme der Skalierung nach Korn) abhängt.

4.3 Vergleich der Ergebnisse bei Bildfolgen

Die Versuchsergebnisse aus Abschnitt 4.1 zeigen, dass es unter anderem von der zum Vergleich der Masken herangezogenen Grauwert-Struktur abhängt, welches Korrekturverfahren die besten Ergebnisse liefert. Die drei dort untersuchten Grauwertstrukturen geben zwar einen ersten Einblick in die Qualität der Korrekturverfahren, liefern aber nicht genug Hinweise für eine abschließende Beurteilung. Aus diesem Grund sollen die Vergleiche nun auf eine Bildfolge ausgedehnt werden. In Abschnitt 4.3.1 werden die mit den verschiedenen Masken ermittelten Ableitungen, in Abschnitt 4.3.2 die geschätzten OF-Vektoren miteinander und mit Referenzwerten verglichen.

4.3.1 Korrektheit der „Invarianzebenen“

4.3.1.1 Ansatz

Bei der Ableitungsbestimmung auf Bildern oder Bildfolgen ist es im Gegensatz zu den bisherigen Versuchen nicht mehr ohne weiteres möglich, die „korrekte“ Ableitung anzugeben. Dies gilt selbst bei synthetischen Bildern, die nicht durch Auswertung einer sehr einfachen Funktion erzeugt wurden.

Bei bekannter VR ergibt sich aus folgender Überlegung dennoch ein Bewertungskriterium für die Qualität der Ableitungsschätzung: An jedem Bildpunkt gibt der Gradientenvektor ∇g die Richtung der maximalen Grauwertänderung an. Senkrecht dazu ändert sich der Grauwert (lokal) nicht. Der Gradientenvektor ist also der Normalenvektor einer Ebene, in der sich die Grauwerte nicht ändern, der „Invarianzebene“ \mathbf{E}_0 . Als Eigenwert-Problem formuliert bedeutet das: Die Invarianzebene wird durch die Eigenvektoren zum doppelten Eigenwert 0 des lokalen Tensors $(\nabla g)(\nabla g)^T$ aufgespannt. Eine perfekte Übereinstimmung zwischen VR und OF ergibt sich, wenn der erweiterte

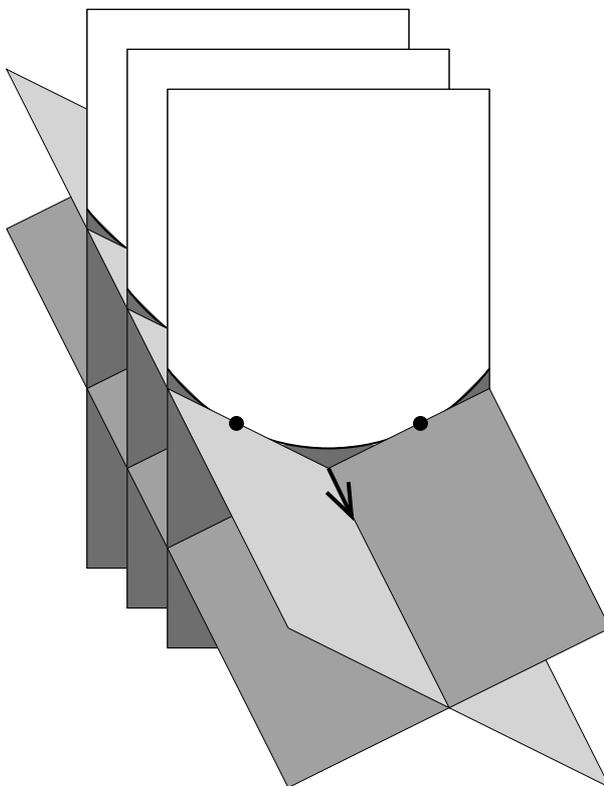


Abbildung 4.1: Zusammenhang zwischen Invarianzebenen und OF-Schätzung (schematisch): Durch mindestens zwei nicht-koplanare Invarianzebenen wird der Optische Fluss eindeutig bestimmt.

VR-Vektor $\mathbf{v}_+ = (v_1, v_2, 1)$ Eigenvektor zum Eigenwert 0 der GST-Matrix $\overline{(\nabla g)(\nabla g)^T}$ ist. Da jeder der lokalen Tensoren positiv definit ist, kann der VR-Vektor diese Eigenschaft nur erhalten, wenn er auch für jeden der lokalen Tensoren in der berücksichtigten Umgebung Eigenvektor zum Eigenwert 0 ist, also in der Invarianzebene liegt¹. Jede Abweichung aus dieser Ebene führt zumindest zu einer Zunahme des kleinsten Eigenwertes des GST, im ungünstigen Fall – wenn sich nicht die fehlerhaften Orientierungen an mehreren Punkten gegenseitig aufheben – aber auch zu einer Verfälschung der Richtung des Eigenvektors und somit zu einer schlechten OF-Schätzung. Ein Anhaltspunkt für die Qualität der Ableitungsbestimmung ist deshalb der Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ zwischen dem Vektor \mathbf{v}_+ und dieser Ebene, der im Idealfall 0 sein sollte:

$$\alpha_{\mathbf{v}_+, \mathbf{E}_0} := \angle(\mathbf{v}_+, \mathbf{E}_0) = \min\{\angle(\mathbf{v}_+, \mathbf{e}) \mid \mathbf{e} \in \mathbf{E}_0\} .$$

¹[Haußecker & Jähne 97] definieren sogar den OF (sinngemäß) als den Vektor der „so senkrecht wie möglich“ auf allen Gradientenvektoren steht.

4.3.1.2 Invarianzebene und Normal-Fluss

Um der Tatsache abzuweichen, dass der OF nicht aus der Gradienteninformation eines einzelnen Punktes bestimmt werden kann, definiert man üblicherweise den Normal-Fluss (NF) als die Komponente der Bewegung eines Bildpunktes in Gradientenrichtung im Bild, dargestellt durch den NF-Vektor $\mathbf{u}_N = (u_{N,1}, u_{N,2})^T$. Der NF kann für jeden Bildpunkt aus dem Gradientenvektor eindeutig bestimmt werden. Offensichtlich besteht ein Zusammenhang zwischen NF und Invarianzebene:

- Der NF-Vektor muss im Bild in Gradientenrichtung verlaufen:

$$(u_{N,1}, u_{N,2})^T \parallel (g_x, g_y)^T . \quad (4.14)$$

- Der erweiterte NF-Vektor $\mathbf{u}_{N+} = (n_1, n_2, 1)^T$ muss in der Invarianzebene liegen (vgl. auch OFCE, Gleichung (2.8), Seite 12):

$$\mathbf{u}_{N+} \perp \nabla g . \quad (4.15)$$

Der erweiterte NF-Vektor ist also derjenige Vektor aus der Invarianzebene mit t -Komponente 1, dessen Projektion in die Bildebene (bis auf das Vorzeichen) mit der Projektion des Gradientenvektors in die Bildebene übereinstimmt.

Da der NF aber „nur“ ein zweidimensionales Merkmal in der Bildebene ist und die Invarianzebene anschaulicher ist, wird im Folgenden mit der Invarianzebene gearbeitet.

4.3.1.3 Versuchsdurchführung

Mit Hilfe dieses Kriteriums soll ein weiterer Vergleich aller Verfahren zur Bestimmung der Ableitungsmasken durchgeführt werden. Alle Versuche wurden auf Bild 150 der Bildfolge `marmor_stat` durchgeführt. Dabei wurde das gesamte Bild bis auf einen 20 Pixel breiten Streifen am Rand untersucht. Bei diesem Versuch wäre zwar ein so breiter Rand nicht nötig, ist aber zwecks späterer Vergleichbarkeit bei anderer Parametrisierung, v. a. bei größeren Masken, sinnvoll. Um verschiedene Effekte, die Einfluss auf das Ergebnis nehmen, besser voneinander unterscheiden zu können, wird das Bild bei der Auswertung in drei Bereiche geteilt:

1. Punkte, die so nahe am Abbild von Objektkanten liegen, dass bei der Ableitungsbestimmung Pixel einen wesentlichen Einfluss haben, zwischen denen eine Bewegungsdiskontinuität besteht („Rand“). Zu dieser Kategorie wurden Punkte gezählt, deren räumlicher euklidischer Abstand zur nächstgelegenen verdeckenden Objektkante im betrachteten Bild kleiner ist als die Maskenausdehnung N . An diesen Stellen ist ein relativ großer Winkelfehler zu erwarten, wobei bei Randkorrektur ein schlechteres Ergebnis zu erwarten ist als bei einer Skalierung aller Koeffizienten.

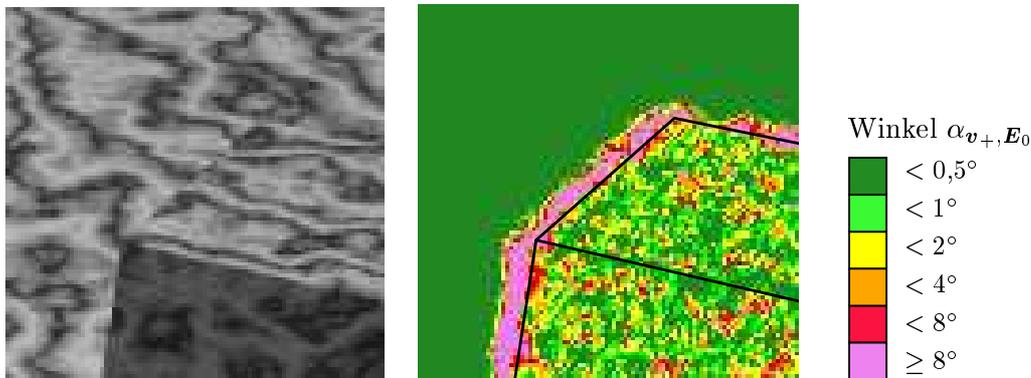


Abbildung 4.2: [marmor_stat] Winkel α_{v+, E_0} für einen Ausschnitt aus Bild 150, für $\sigma_x = \sigma_y = 1,41, \sigma_t = 1,0$, abgetastete Koeffizienten ohne Korrektur, Maskengröße $13 \times 13 \times 11$. Der im unteren rechten Bildbereich zu sehende Block bewegt sich nach links, während der Hintergrund stationär ist. Man erkennt, dass die größten Winkel an der Grenze zwischen bewegtem Vordergrund und stationärem Hintergrund auftreten. Eine Erklärung, warum große Fehler nur an der im Hintergrund liegenden Seite der Grenze des Objektbildes auftreten, findet sich in Abschnitt 7.1.1.

2. Punkte, die zum Abbild eines sich bewegenden Vordergrundobjektes gehören, aber nicht zu nahe an der Objektkante liegen (ohne Rand – „VG o. R.“). Hier sollte der Winkelfehler relativ niedrig sein, da die Daten rauschfrei sind.
3. Punkte, die zum Abbild des stationären Bildhintergrundes gehören und nicht zu nahe an den Kanten eines Vordergrundobjektes liegen. Da die synthetischen Bilder rauschfrei sind, ist hier der Fehler nur dann ungleich 0, wenn im Maskenbereich eine Schattenkante liegt. Von einer Untersuchung dieser Punkte wird abgesehen, da die Schattenkante eine scheinbare Bewegungsdiskontinuität darstellt und somit dieser Fall durch die Testfälle des erstgenannten Falles bereits abgedeckt wird.

Die Zuordnung eines Punktes zu einer dieser drei Gruppen kann sich ändern, wenn sich die Maskenausdehnung ändert. Um Versuche, die mit unterschiedlichen Maskenausdehnungen durchgeführt wurden, miteinander vergleichen zu können, wird daher als vierte Gruppe der zu Vordergrundobjekten gehörende Bereich *inklusive* Randanteil (mit Rand – „VG m. R.“) betrachtet. Abbildung 4.2 zeigt an einem Bildausschnitt, dass der Fehler tatsächlich im Bereich der Bewegungsdiskontinuität am größten und im Hintergrund am geringsten ist. Die Einteilung des Bildes in diese Bereiche ist also sinnvoll.

Zur Beurteilung des Fehlers werden zwei Kenngrößen herangezogen: der Median der Fehlers und der durchschnittliche Fehler. Eine Betrachtung von sowohl Median als auch Mittelwert erscheint sinnvoll, da zu erwarten ist, dass die Schätzung der Invarianzebene für einen kleinen Prozentsatz der betrachteten Punkte einen weit überdurchschnittli-

	abgetastet							
	A-	AS $_{\sigma}$	AS $_{\text{mod}}$	AK	-	-	AG	AGR
Median Rand	7,535	7,748	7,555	9,347	-	-	7,560	7,727
Mittelwert Rand	19,496	19,997	19,338	18,042	-	-	19,392	19,630
Median VG o. R.	1,060	1,476	1,025	5,226	-	-	1,024	0,914
Mittelwert VG o. R.	2,013	2,276	1,998	5,431	-	-	1,997	1,792
Median VG m. R.	1,144	1,549	1,107	5,344	-	-	1,111	1,000
Mittelwert VG m. R.	2,360	2,613	2,347	5,695	-	-	2,345	2,179
	integriert							
	I-	IS $_{\sigma}$	IS $_{\text{mod}}$	IK	IR	IM	IG	IGR
Median Rand	7,676	7,938	7,671	9,293	7,891	8,056	7,682	7,917
Mittelwert Rand	19,531	20,235	19,546	18,161	19,667	20,222	19,529	19,791
Median VG o. R.	1,048	1,621	1,050	5,048	0,949	1,593	1,048	0,920
Mittelwert VG o. R.	1,984	2,351	1,986	5,255	1,841	2,389	1,984	1,782
Median VG m. R.	1,128	1,693	1,133	5,157	1,041	1,669	1,128	0,998
Mittelwert VG m. R.	2,345	2,699	2,347	5,531	2,231	2,734	2,345	2,183

Tabelle 4.11: Durchschnittlicher Winkel α_{v^+, E_0} (in Grad) für Bild 150 der Bildfolge [marmor_stat] bei verschiedenen Verfahren zur Maskenbestimmung: Aufgelistet sind für die Bereiche Rand, Vordergrund ohne Randanteil (VG o. R.) und Vordergrund mit Randanteil (VG m. R.) jeweils Median und Mittelwert der Winkel. Parameter: $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$, Maskengröße: $7 \times 7 \times 5$. Der Rand umfasst 3 Pixel beiderseits der verdeckenden Objektkanten, insgesamt 3890 Punkte, der Vordergrund ohne Randanteil 17241 Punkte, der Vordergrund mit Randanteil 19143 Punkte.

chen Fehler aufweist. Die Ergebnisse werden zeigen, dass tatsächlich der Mittelwert des Fehlers stets deutlich größer ist (etwa um einen Faktor 2 bis 3) als der Median.

Beim Versuch wurden $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$ gewählt, da bei isotropen Masken alle *Skalierungsverfahren* offensichtlich dieselben Ergebnisse liefern; es fände lediglich eine Skalierung des Normalenvektors der Invarianzebene statt. Bei Vorgabe eines theoretischen Maskenfehlers von drei Prozent ergibt sich eine Maskengröße von $9 \times 9 \times 7$ Pixeln (vgl. Tabelle 3.2). Der tatsächliche Maskenfehler liegt bei dieser Ausdehnung bei rund 0,8 Prozent bei der x - und y -Ableitung und bei 0,5 Prozent bei der t -Ableitung. Zum Vergleich wurden auch Masken betrachtet, die eine jeweils um einen Pixel kleinere oder größere Ausdehnung aufweisen, also $7 \times 7 \times 5$ Pixel (tatsächlicher Maskenfehler von rund 7,0 % bei x - und y - sowie 6,9 % bei t -Ableitung) und $11 \times 11 \times 9$ Pixel (tatsächlicher Maskenfehler von etwa 0,06 % bzw. 0,02 %).

4.3.1.4 Ergebnisse

Tabelle 4.11 zeigt die Ergebnisse für eine Maskengröße von $7 \times 7 \times 5$ Pixeln, Tabelle 4.12 für eine Maskengröße von $9 \times 9 \times 7$ und Tabelle 4.13 für eine Maskengröße von $11 \times 11 \times 9$.

	abgetastet							
	A-	AS _σ	AS _{mod}	AK	–	–	AG	AGR
Median Rand	6,267	6,483	6,217	8,812	–	–	6,269	6,309
Mittelwert Rand	17,811	18,275	17,667	16,682	–	–	17,729	17,783
Median VG o. R.	0,815	1,240	0,809	5,299	–	–	0,797	0,789
Mittelwert VG o. R.	1,654	1,943	1,659	5,312	–	–	1,647	1,628
Median VG m. R.	0,920	1,327	0,912	5,442	–	–	0,898	0,889
Mittelwert VG m. R.	2,083	2,353	2,088	5,644	–	–	2,077	2,065
	integriert							
	I-	IS _σ	IS _{mod}	IK	IR	IM	IG	IGR
Median Rand	6,461	6,737	6,471	8,766	6,531	6,636	6,460	6,521
Mittelwert Rand	18,010	18,640	17,982	16,857	17,993	18,468	17,923	17,988
Median VG o. R.	0,772	1,431	0,760	5,023	0,744	1,436	0,752	0,739
Mittelwert VG o. R.	1,563	2,016	1,557	5,042	1,538	2,099	1,553	1,536
Median VG m. R.	0,877	1,512	0,866	5,172	0,848	1,525	0,851	0,843
Mittelwert VG m. R.	2,021	2,444	2,015	5,394	2,003	2,504	2,012	2,002

Tabelle 4.12: Wie Tabelle 4.11, aber mit Maskengröße $9 \times 9 \times 7$. Randbreite: 4 Pixel, insgesamt 5190 Punkte, Vordergrund ohne Randanteil: 16629 Punkte, Vordergrund mit Randanteil wie oben 19143 Punkte.

Maskengröße Betrachtet man den Vordergrund inklusive Randanteil, so nimmt – abgesehen von der Skalierung nach Korn, wo sowohl Median als auch Mittelwert des Fehlers bei zunehmender Maskengröße wachsen – erwartungsgemäß der Fehler in fast allen Fällen² mit wachsender Maskengröße ab. Dasselbe gilt für die Bereiche Rand und Vordergrund ohne Randanteil, wobei hier ein Vergleich der Ergebnisse nur eingeschränkt möglich ist, weil bei kleineren Masken ein kleinerer Rand- und ein größerer Vordergrundbereich (ohne Randanteil) betrachtet werden.

Abgetastete vs. integrierte Maskenkoeffizienten Bei 108 Paaren vergleichbarer Kennzahlen (Mediane und Mittelwerte) liefern 51 mal die abgetasteten und 55 mal die integrierten Maskenkoeffizienten die besseren Ergebnisse. In zwei Fällen stimmen die Zahlen in den drei betrachteten Nachkommastellen überein. Somit sind die theoretisch erwarteten Schwächen der abgetasteten Maskenkoeffizienten (schlechtere Modellierung der Gaußglocke) in der Praxis kaum nachweisbar.

Skalierungen und Randkorrekturen Die σ -basierte Skalierung (AS_σ , IS_σ) führt – im Vergleich zu den nicht-korrigierten Masken – in allen Fällen zu einer Zunahme von Median und Mittelwert des Fehlers, wobei die Zunahme im Randbereich gering

²Die einzige Ausnahme wird bei der Bewertung der gradienten-basierten Randkorrektur auf Seite 74 diskutiert.

	abgetastet							
	A-	AS _σ	AS _{mod}	AK	–	–	AG	AGR
Median Rand	4,834	5,052	4,854	7,985	–	–	4,843	4,837
Mittelwert Rand	15,398	15,838	15,294	14,658	–	–	15,389	15,395
Median VG o. R.	0,774	1,175	0,797	5,335	–	–	0,772	0,771
Mittelwert VG o. R.	1,576	1,851	1,598	5,306	–	–	1,575	1,574
Median VG m. R.	0,895	1,272	0,919	5,512	–	–	0,893	0,892
Mittelwert VG m. R.	2,064	2,314	2,085	5,695	–	–	2,064	2,063
	integriert							
	I-	IS _σ	IS _{mod}	IK	IR	IM	IG	IGR
Median Rand	4,949	5,290	4,960	7,876	4,979	5,192	4,985	4,979
Mittelwert Rand	15,654	16,251	15,651	14,867	15,652	16,057	15,643	15,651
Median VG o. R.	0,725	1,350	0,724	5,074	0,722	1,421	0,723	0,722
Mittelwert VG o. R.	1,477	1,913	1,477	5,043	1,476	2,054	1,477	1,475
Median VG m. R.	0,846	1,455	0,845	5,245	0,846	1,526	0,847	0,845
Mittelwert VG m. R.	2,001	2,399	2,000	5,453	2,000	2,505	2,000	2,000

Tabelle 4.13: Wie Tabelle 4.11, aber mit Maskengröße $11 \times 11 \times 9$. Randbreite: 5 Pixel, insgesamt 6850 Punkte, Vordergrund ohne Randanteil: 15864 Punkte, Vordergrund mit Randanteil wie oben 19143 Punkte.

ausfällt (ca. 2,6 bis 6,9 Prozent), während es im Vordergrund einmal sogar fast zu einer Verdoppelung des Medians kommt. Eine ähnliche Verschlechterung der Ergebnisse zeigt sich bei der *Mittelkorrektur (IM)*. Aufgrund dieser Beobachtungen erscheinen beide Verfahren ungeeignet zur Korrektur des Maskenfehlers.

Alle anderen Korrekturverfahren führen in mindestens einem Testfall zu einer Ergebnisverbesserung. Bei der *Skalierung nach Korn (AK,IK)* tritt diese Verbesserung jedoch nur beim durchschnittlichen Fehler im Randbereich auf. In allen anderen Fällen kommt es zu einer zum Teil erheblichen Verschlechterung. Die Verbesserung des Mittelwerts im „schwierigen Fall“ um fünf bis sieben Prozent rechtfertigt keine Vervielfachung des Fehlers im „unproblematischen Fall“ des bewegten Vordergrundes. Aus diesem Grund ist auch dieses Verfahren ungeeignet zur Korrektur des Maskenfehlers.

Die beiden noch verbliebenen Skalierungsverfahren (AS_{mod}, IS_{mod}, AG, IG) verändern die Ergebnisse in den meisten Fällen kaum (maximal 1 Prozent besser oder schlechter). Auffällig ist aber, dass die *modifizierte σ -basierte Skalierung* bei den größten Masken – in dem Fall also, in dem die Skalierung eigentlich das Ergebnis kaum mehr verändern sollte – bei abgetasteten Koeffizienten im „unproblematischen Fall“ des bewegten Vordergrundes zu einer Zunahme des Fehlers führt.³ Dies führt zur zunächst vorläufigen Annahme, dass die *gradienten-basierte Skalierung* besser zur Korrektur der Masken

³Die Ergebnisse wurden durch eine Wiederholung des Versuchs verifiziert, um einen Fehlübertrag der Ergebnisse auszuschließen.

geeignet ist als die modifizierte σ -basierte Skalierung.

Im Vergleich zu den Skalierungsverfahren führen die beiden Randkorrektur-Verfahren (IR, AGR, IGR) zu einer Vergrößerung des Fehlers in Randbereichen und zu einer Abnahme des Fehlers im Vordergrundbereich. Die Zunahme des Fehlers in Randbereichen ist eigentlich keine Schwäche der Randkorrektur, sondern ist auf die Verletzung der Annahmen 7 und 8 (vgl. Seite 13) zurückzuführen, d. h. die Skala wurde falsch (zu groß) gewählt. Trotzdem soll dieser Aspekt bei der Beurteilung der Korrekturverfahren berücksichtigt werden, da es sich selbst bei besserer Berücksichtigung der Diskontinuitäten nicht vermeiden lassen wird, dass von den Faltungsmasken Punkte erfasst werden, die Annahmen 7 und 8 verletzen.

Die *gradienten-basierte Randkorrektur* angewandt auf $9 \times 9 \times 7$ -Masken führt im Vordergrund-Bereich mit Rand zu einem geringfügig niedrigeren Median des Fehlers als die nicht-korrigierten (und als die gradienten-basiert randkorrigierten) $11 \times 11 \times 9$ -Masken; d. h. eine passend gewählte Maskenkorrektur ist eine Rechenzeit-sparende Alternative zu größeren Masken, wenn die Korrektur selbst effizient durchgeführt werden kann.

4.3.2 Korrektheit der GST- und OF-Schätzung

Die Schätzung des OFs erfolgt – wie in Abschnitt 2.4.1 dargestellt – mit Hilfe des GST, der aus den lokalen Tensoren durch eine gewichtete Mittelung entsteht.

Einerseits kann dadurch der Fehler einzelner lokaler Tensoren ausgeglichen werden, andererseits wirken sich fehlerhafte lokale Tensoren dadurch an allen Pixelpositionen aus, in deren Mittelungsumgebung sie liegen, wobei sich an einem Pixel auch beide Effekte überlagern können.

Wie stark sich eine fehlerhafte Orientierung der Invarianzebenen auf die OF-Schätzung auswirkt, kann letztendlich nur durch Betrachtung des GST oder des OF selbst beurteilt werden.

4.3.2.1 Ansatz

Als Maß für die Zuverlässigkeit der Schätzung des GST greifen wir auf das von [Barron et al. 94] eingeführte und von zahlreichen Autoren verwendete Winkelmaß zurück. Seien – wie bisher – $\mathbf{u}_+ = (u_x, u_y, 1)^T$ der erweiterte OF-Vektor und $\mathbf{v}_+ = (v_x, v_y, 1)^T$ der erweiterte VR-Vektor, so definieren [Barron et al. 94] (unter Verwendung anderer Bezeichnungen):

$$\psi_{\mathbf{v}_+, \mathbf{u}_+} := \sphericalangle(\mathbf{v}_+, \mathbf{u}_+). \quad (4.16)$$

Im hier verwendeten Verfahren ist \mathbf{u}_+ ein – allerdings nicht normierter – Eigenvektor zum kleinsten Eigenwert des Grauwertstrukturensors. ψ gibt also den Winkel an, um den der GST mindestens rotiert werden muss, damit der OF mit der tatsächlichen VR übereinstimmt.

Dieses Winkelmaß besitzt jedoch, wie die Autoren z. T. selbst feststellen, mehrere Nachteile, die folgende Tabelle an einem Zahlenbeispiel illustriert:

Bsp. Nr.	\mathbf{v}_+^T	\mathbf{u}_+^T	$\psi_{\mathbf{v}_+, \mathbf{u}_+}$	$\ \mathbf{v} - \mathbf{u}\ _2$
1	(5,0; 0,0; 1,0)	(4,0; 0,0; 1,0)	2,72°	1,0
2	(2,5; 0,0; 1,0)	(2,0; 0,0; 1,0)	4,76°	0,5
3	(3,0; 0,0; 1,0)	(2,0; 0,0; 1,0)	8,13°	1,0
4	(5,0; 0,0; 1,0)	(5,0; 0,2428; 1,0)	2,72°	0,2428

Bei großer VR führen vergleichsweise große Fehler der OF-Schätzung zu einem relativ kleinen Winkelfehler. Dies gilt sowohl bei Betrachtung der relativen (Bsp. 1 und 2) als auch der absoluten (Bsp. 1 und 3) Unterschätzung der Verschiebung. Außerdem werden Fehler senkrecht zur Richtung der Verschiebung stärker berücksichtigt als Fehler in Richtung der Verschiebung (Bsp. 1 und 4).

Aus diesem Grund soll zusätzlich auch die Euklid-Norm $\|\mathbf{v} - \mathbf{u}\|_2$ des Differenzvektors aus tatsächlicher VR \mathbf{v} und geschätztem OF \mathbf{u} betrachtet werden.⁴

Wie schon bei der Untersuchung der durch die lokalen Tensoren gegebenen Invarianzebenen werden die Ableitungen mit $\Sigma_A = \text{diag}(1,41^2; 1,41^2; 1,00^2)$ berechnet, bei Maskengrößen von $7 \times 7 \times 5$ bis $11 \times 11 \times 9$. Für die anschließende Mittelung über die Tensoren gilt $\Sigma_M = 2\Sigma_A$. Dadurch sind die Mittelungsmasken größer als die entsprechenden Ableitungsmasken: Gibt man – wie schon bei den Ableitungsmasken – einen Fehler von 3% vor, so ergeben sich Maskengrößen von $11 \times 11 \times 9$ (Standardgröße⁵, tatsächlicher Maskenfehler von rund 1,3 Prozent), $9 \times 9 \times 7$ (um ein Pixel verkleinerte Ausdehnung, tatsächlicher Maskenfehler von rund sechs Prozent) bzw. $13 \times 13 \times 11$ (um ein Pixel vergrößerte Ausdehnung, tatsächlicher Maskenfehler von etwa 0,2 Prozent).

4.3.2.2 Ergebnisse

Tabelle 4.14 zeigt die Ergebnisse für eine Ableitungsmaskengröße von $7 \times 7 \times 5$ Pixeln und eine Mittelungsmaske mit $9 \times 9 \times 7$ Pixeln, Tabelle 4.15 für Maskengrößen von $9 \times 9 \times 7$ bzw. $11 \times 11 \times 9$ und Tabelle 4.16 für Maskengrößen von $11 \times 11 \times 9$ und $13 \times 13 \times 11$.

Bei zunehmender **Maskengröße** nimmt der Fehler in fast allen Fällen ab. Die Ausnahmen treten bei der Skalierung nach Korn und bei der modifizierten σ -basierten Skalierung auf, sowie in drei Einzelfällen bei den gradienten-basierten Verfahren, wo die durchschnittliche Länge des Differenzvektors um 0,002 bis 0,02 Pixel/Bild zunimmt.

⁴Schon [Otte 94] weist auf die Probleme des Winkelmaßes nach [Barron et al. 94] hin und verwendet daher ausschließlich die Euklid-Norm der Differenz.

⁵nach Tabelle 3.1: $\mu_x = \mu_y = \sqrt{2} \cdot 1,41 \cdot 2,572 = 5,13$, d. h. Maskenbreite $2 \cdot 5 + 1 = 11$; $\mu_t = \sqrt{2} \cdot 1,00 \cdot 2,572 = 3,64$, d. h. Maskenbreite $2 \cdot 4 + 1 = 9$

		abgetastet							
		A-	AS _{σ}	AS _{mod}	AK	-	-	AG	AGR
Winkel $\psi_{\mathbf{v}_+, \mathbf{u}_+}$	Median Rand	13,627	13,980	13,511	14,152	-	-	13,570	14,118
	Mittelwert Rand	25,719	26,446	25,486	23,443	-	-	25,568	25,890
	Median VG o. R.	0,742	1,628	0,639	7,423	-	-	0,647	0,448
	Mittelwert VG o. R.	0,890	1,703	0,801	7,559	-	-	0,808	0,619
	Median VG m. R.	0,851	1,702	0,765	7,608	-	-	0,767	0,533
	Mittelwert VG m. R.	3,373	4,051	3,307	9,417	-	-	3,310	3,217
$\ \mathbf{v} - \mathbf{u}\ _2$	Median Rand	0,4093	0,4341	0,4023	0,4901	-	-	0,4048	0,4181
	Mittelwert Rand	1,0742	1,1379	1,0468	0,8406	-	-	1,0562	1,0961
	Median VG o. R.	0,0389	0,1141	0,0279	0,4281	-	-	0,0294	0,0229
	Mittelwert VG o. R.	0,0456	0,1144	0,0347	0,4173	-	-	0,0366	0,0300
	Median VG m. R.	0,0453	0,1196	0,0332	0,4310	-	-	0,0352	0,0276
	Mittelwert VG m. R.	0,1721	0,2331	0,1615	0,4833	-	-	0,1636	0,1663
		integriert							
		I-	IS _{σ}	IS _{mod}	IK	IR	IM	IG	IGR
Winkel $\psi_{\mathbf{v}_+, \mathbf{u}_+}$	Median Rand	13,771	14,286	13,776	14,264	14,257	14,462	13,768	14,392
	Mittelwert Rand	25,743	26,742	25,757	23,606	25,931	26,857	25,740	26,050
	Median VG o. R.	0,737	1,911	0,750	7,107	0,579	1,846	0,735	0,542
	Mittelwert VG o. R.	0,904	1,989	0,914	7,249	0,757	1,953	0,903	0,710
	Median VG m. R.	0,862	1,987	0,872	7,304	0,692	1,948	0,861	0,635
	Mittelwert VG m. R.	3,429	4,333	3,437	9,175	3,373	4,347	3,428	3,340
$\ \mathbf{v} - \mathbf{u}\ _2$	Median Rand	0,4134	0,4466	0,4134	0,4844	0,4197	0,4471	0,4132	0,4248
	Mittelwert Rand	1,0743	1,1836	1,0769	0,8426	1,0890	1,1740	1,0739	1,0977
	Median VG o. R.	0,0344	0,1359	0,0356	0,4112	0,0273	0,1298	0,0343	0,0284
	Mittelwert VG o. R.	0,0426	0,1362	0,0438	0,4022	0,0350	0,1324	0,0424	0,0360
	Median VG m. R.	0,0411	0,1416	0,0426	0,4141	0,0329	0,1366	0,0410	0,0343
	Mittelwert VG m. R.	0,1750	0,2598	0,1762	0,4672	0,1682	0,2572	0,1749	0,1712

Tabelle 4.14: Mittelwert von Winkel $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ (in Grad) und Euklidnorm $\|\mathbf{v} - \mathbf{u}\|_2$ (in Pixel/Bild) für Bild 150 der Bildfolge [marmor_stat] bei verschiedenen Verfahren zur Maskenbestimmung: Aufgelistet sind für die Bereiche Rand und Vordergrund (VG) Median und Mittelwert der Winkel. Parameter: $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$; Maskengröße Ableitung: $7 \times 7 \times 5$, Mittelung: $9 \times 9 \times 7$. Der Rand umfasst 7 Pixel beiderseits der verdeckenden Objektkanten, insgesamt 9437 Punkte, der Vordergrund ohne Randanteil 14709 Punkte, der Vordergrund mit Randanteil 19143 Punkte.

Beim Vergleich **abgetastete vs. integrierte Maskenkoeffizienten** sind in 149 von 216 Fällen die Ergebnisse bei den abgetasteten Masken besser, 56 mal bei den integrierten Masken, und in 11 Fällen stimmen die Kennzahlen überein. Die abgetasteten Masken können somit mit hoher Sicherheit als zumindest gleichwertig zu den integrierten Masken betrachtet werden.

		abgetastet							
		A-	AS_σ	AS_{mod}	AK	–	–	AG	AGR
Winkel ψ_{v_+, u_+}	Median Rand	9,616	9,832	9,611	11,143	–	–	9,609	9,719
	Mittelwert Rand	22,264	23,029	22,068	20,563	–	–	22,141	22,245
	Median VG o. R.	0,302	1,289	0,354	7,666	–	–	0,261	0,241
	Mittelwert VG o. R.	0,384	1,329	0,439	7,725	–	–	0,356	0,335
	Median VG m. R.	0,387	1,365	0,444	7,793	–	–	0,349	0,319
	Mittelwert VG m. R.	3,117	3,879	3,165	9,663	–	–	3,097	3,095
$\ v - u\ _2$	Median Rand	0,2675	0,2893	0,2644	0,4458	–	–	0,2640	0,2668
	Mittelwert Rand	0,9062	0,9703	0,8992	0,7237	–	–	0,8996	0,9022
	Median VG o. R.	0,0168	0,0934	0,0206	0,4462	–	–	0,0134	0,0123
	Mittelwert VG o. R.	0,0197	0,0921	0,0233	0,4324	–	–	0,0169	0,0159
	Median VG m. R.	0,0212	0,0973	0,0256	0,4474	–	–	0,0178	0,0163
	Mittelwert VG m. R.	0,1599	0,2242	0,1624	0,4955	–	–	0,1565	0,1586
		integriert							
		I-	IS_σ	IS_{mod}	IK	IR	IM	IG	IGR
Winkel ψ_{v_+, u_+}	Median Rand	9,871	10,248	9,851	11,067	10,036	10,177	9,845	10,055
	Mittelwert Rand	22,490	23,501	22,445	20,748	22,481	23,377	22,349	22,473
	Median VG o. R.	0,319	1,629	0,280	7,265	0,242	1,671	0,263	0,241
	Mittelwert VG o. R.	0,401	1,672	0,372	7,328	0,339	1,691	0,363	0,339
	Median VG m. R.	0,406	1,704	0,371	7,390	0,325	1,746	0,356	0,321
	Mittelwert VG m. R.	3,202	4,225	3,180	9,340	3,169	4,220	3,175	3,171
$\ v - u\ _2$	Median Rand	0,2760	0,3071	0,2754	0,4340	0,2777	0,3043	0,2727	0,2777
	Mittelwert Rand	0,9139	0,9998	0,9092	0,7391	0,9102	0,9904	0,8998	0,9088
	Median VG o. R.	0,0179	0,1204	0,0149	0,4268	0,0124	0,1219	0,0134	0,0123
	Mittelwert VG o. R.	0,0208	0,1181	0,0183	0,4138	0,0161	0,1200	0,0171	0,0160
	Median VG m. R.	0,0225	0,1240	0,0193	0,4284	0,0164	0,1264	0,0179	0,0164
	Mittelwert VG m. R.	0,1689	0,2521	0,1665	0,4796	0,1665	0,2519	0,1647	0,1663

Tabelle 4.15: Wie Tabelle 4.14, aber mit Maskengröße Ableitung: $9 \times 9 \times 7$, Mittelung: $11 \times 11 \times 9$. Der Rand umfasst 9 Pixel beiderseits der verdeckenden Objektkanten, insgesamt 12391 Punkte, der Vordergrund ohne Randanteil 13449 Punkte und mit Randanteil 19143 Punkte.

Bei der Betrachtung der Korrekturverfahren gleichen die Ergebnisse sehr stark den Ergebnissen bei der Untersuchung der Invarianzebenen. σ -basierte Skalierung (AS_σ, IS_σ) und Mittelkorrektur (IM) führen stets, Skalierung nach Korn (AK, IK) in fast allen Fällen zu einer Verschlechterung der Ergebnisse und sind daher ungeeignet zur Korrektur des Maskenfehlers.

Beim Vergleich der verbliebenen beiden Skalierungsverfahren ($AS_{\text{mod}}, IS_{\text{mod}}, AG, IG$) zeigt die modifizierte σ -basierte Skalierung Schwächen bei abgetasteten Koeffizienten ausgerechnet bei schon vorher geringem Maskenfehler und kann daher auch nicht als zuverlässig betrachtet werden.

		abgetastet							
		A-	AS $_{\sigma}$	AS $_{\text{mod}}$	AK	–	–	AG	AGR
Winkel ψ_{v_+, u_+}	Median Rand	6,298	6,449	6,268	9,148	–	–	6,302	6,340
	Mittelwert Rand	19,205	19,959	19,057	18,114	–	–	19,191	19,213
	Median VG o. R.	0,231	1,201	0,396	7,737	–	–	0,231	0,230
	Mittelwert VG o. R.	0,294	1,233	0,444	7,787	–	–	0,294	0,293
	Median VG m. R.	0,316	1,287	0,486	7,869	–	–	0,315	0,315
	Mittelwert VG m. R.	3,086	3,823	3,205	9,745	–	–	3,086	3,088
$\ v - u\ _2$	Median Rand	0,1763	0,1888	0,1765	0,4099	–	–	0,1761	0,1769
	Mittelwert Rand	0,7671	0,8187	0,7573	0,6356	–	–	0,7658	0,7672
	Median VG o. R.	0,0118	0,0880	0,0251	0,4525	–	–	0,0118	0,0118
	Mittelwert VG o. R.	0,0144	0,0863	0,0261	0,4395	–	–	0,0143	0,0143
	Median VG m. R.	0,0162	0,0912	0,0299	0,4521	–	–	0,0161	0,0161
	Mittelwert VG m. R.	0,1583	0,2170	0,1689	0,5001	–	–	0,1581	0,1599
		integriert							
		I-	IS $_{\sigma}$	IS $_{\text{mod}}$	IK	IR	IM	IG	IGR
Winkel ψ_{v_+, u_+}	Median Rand	6,636	6,872	6,637	9,043	6,666	6,788	6,633	6,664
	Mittelwert Rand	19,447	20,436	19,442	18,290	19,438	20,310	19,429	19,437
	Median VG o. R.	0,229	1,526	0,230	7,353	0,228	1,655	0,230	0,228
	Mittelwert VG o. R.	0,294	1,561	0,293	7,403	0,292	1,661	0,294	0,292
	Median VG m. R.	0,318	1,611	0,317	7,482	0,315	1,735	0,316	0,314
	Mittelwert VG m. R.	3,159	4,164	3,159	9,436	3,154	4,204	3,159	3,154
$\ v - u\ _2$	Median Rand	0,1847	0,2054	0,1844	0,4017	0,1847	0,2054	0,1842	0,1848
	Mittelwert Rand	0,7788	0,8450	0,7783	0,6522	0,7723	0,8468	0,7770	0,7722
	Median VG o. R.	0,0118	0,1139	0,0118	0,4339	0,0117	0,1215	0,0117	0,0117
	Mittelwert VG o. R.	0,0144	0,1113	0,0143	0,4216	0,0143	0,1196	0,0143	0,0143
	Median VG m. R.	0,0162	0,1167	0,0162	0,4341	0,0160	0,1249	0,0161	0,0160
	Mittelwert VG m. R.	0,1652	0,2466	0,1651	0,4854	0,1628	0,2479	0,1649	0,1628

Tabelle 4.16: Wie Tabelle 4.14, aber mit Maskengröße Ableitung: $11 \times 11 \times 9$, Mittelung: $13 \times 13 \times 11$. Der Rand umfasst 11 Pixel beiderseits der verdeckenden Objektkanten, insgesamt 15108 Punkte, der Vordergrund ohne Randanteil 12347 Punkte und mit Randanteil 19143 Punkte.

Die beiden Randkorrektur-Verfahren (IR, AGR, IGR) erweisen sich als etwa gleichwertig, wobei der Vorteil der *gradienten-basierten Randkorrektur* darin besteht, auch für abgetastete Masken anwendbar zu sein.

4.3.3 Versuchsfazit

Aufgrund dieser Untersuchungen zur Korrektheit der Invarianzebenen und der OF-Schätzung kommen noch vier Verfahren zur Korrektur des Maskenfehlers in Frage:

die gradienten-basierte Skalierung, die einfache Randkorrektur, die gradienten-basierte Randkorrektur sowie mit Abstrichen die modifizierte σ -basierte Skalierung.

Eine *Vergrößerung* der Masken bringt aber *stets mindestens genauso gute Ergebnisse* wie jede der Korrekturen. Hinzu kommt, dass die Korrekturverfahren Probleme bei der Anwendung mit sich bringen:

- Sind Σ_A oder Σ_M keine Diagonalmatrizen, so können die Faltungsmasken bisher nur durch Abtastung der dreidimensionalen Gaussfunktion oder ihrer partiellen Ableitungen gewonnen werden. Die einfache Randkorrektur ist in diesem Fall nicht anwendbar. Auch die modifizierte σ -basierte Skalierung erscheint für abgetastete Masken ungeeignet, da sie in den hier beschriebenen Versuchen oft zu einer Zunahme des Fehlers führte.
- Wenn an jedem Bildpunkt eine andere Maske verwendet wird, so führen die beiden gradienten-basierten Verfahren meist zu einem höheren Rechenaufwand als die Vergrößerung der Masken, da eine „Testfaltung“ mit einer linearen Grauwertstruktur durchgeführt werden muss.

Aus diesen Gründen wird im Folgenden ausschließlich mit nicht-korrigierten, aber möglichst großen Masken gerechnet.

Sinnvoll ist der Einsatz der in diesem Kapitel untersuchten Korrekturverfahren allerdings, wenn z. B. die Größe der Faltungsmasken aufgrund von Echtzeit-Anforderungen oder wegen Realisierung der Faltung als Spezielschaltung beschränkt ist, Σ_A und Σ_M Diagonalmatrizen sind und nur wenige verschiedene Masken verwendet werden. In solchen Fällen können die Korrekturverfahren zu einer signifikanten Verbesserung der Ergebnisse führen.

Kapitel 5

Skalen-Bestimmung

In diesem Kapitel wird untersucht, inwiefern sich die Qualität der Gradientenschätzung – gemessen am Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ – und der OF-Schätzung – gemessen an der Differenz zur VR – bei Variation der Skala, gegeben durch Σ , ändern. Diese Untersuchung ist für spätere Versuche von Bedeutung, um dort die Bestimmung von Σ auf einen vernünftigen Wertebereich einzuschränken und um zu verstehen, was eine „optimale Skala“ ausmacht.

5.1 Motivation

Wie die vorangegangenen Versuche gezeigt haben, liegt der durchschnittliche Winkel zwischen Invarianzebene und VR-Vektor im Bereich der Abbilder der bewegten Marmorblöcke *ohne* Rand im besten Fall immer noch bei knapp 1,5 Grad, *mit* Rand sogar bei 2,0 Grad (vgl. Tabelle 4.13, Spalte „IGR“).

Bei Betrachtung des OFs liegt im selben Bereich (ohne Rand) der kleinste durchschnittliche Winkel $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ bei $0,292^\circ$, der durchschnittliche Fehler $\|\mathbf{v} - \mathbf{u}\|_2$ des OF bei 0,0143 Pixel/Bild (vgl. Tabelle 4.16, Spalte „IGR“). [Schar et al. 97] schätzen hingegen, dass für einen OF-Fehler von höchstens 0,01 Pixel/Bild der Fehler in der Orientierung des Eigenvektors zum kleinsten Eigenwert des GST bis zu $0,57^\circ$ betragen darf.

Diese Abschätzung gilt aber nur, wenn die tatsächliche VR Null ist. Beträgt die tatsächliche VR 1,0 Pixel/Bild, so darf – um einen OF-Fehler von 0,01 Pixel/Bild zu unterschreiten – der Orientierungsfehler höchstens $0,288^\circ$ in Bewegungsrichtung oder höchstens $0,405^\circ$ senkrecht zur Bewegungsrichtung betragen. Bei 1,6 Pixeln/Bild – wie durchschnittlich in Bild 150 der Bildfolge `marmor_stat` – entspricht ein Fehler von 0,01 Pixel/Bild einer Fehlorientierung von $0,162^\circ$ bzw. $0,304^\circ$, was in Einklang mit den tatsächlich beobachteten Werten steht.

Obwohl die für diesen Versuch verwendeten Daten rauschfrei sind, tritt also selbst in Bereichen, die nicht durch Bewegungsdiskontinuitäten beeinflusst sind, ein nicht un-

erheblicher Fehler bei der Schätzung des lokalen Tensors und des GST auf. Aufgrund der sorgfältigen Generierung der Bilder und der Referenzverschiebungsrates (siehe Diskussion in Anhang A.8) kann ein in der Bildfolge selbst liegender Fehler mit hoher Sicherheit ausgeschlossen werden.

Der wahrscheinlichste Grund für den Restfehler ist somit die Wahl eines ungeeigneten Algorithmus, eine inadäquate Parametrisierung des Algorithmus oder eine Kombination beider Ursachen. In der Folge soll daher untersucht werden, welche Parameterkombination die besten Resultate liefert. Es ist zwar nicht zu erwarten, dass die in diesem untersuchten Fall optimalen Parameter auch in allen anderen Fällen gute Ergebnisse liefern, aber evtl. sichtbar werdende bzw. vermutete mögliche Gesetzmäßigkeiten könnten zur automatischen Bestimmung der optimalen Parameter eingesetzt werden.

5.2 Skalenabhängigkeit der Invarianzebenen

5.2.1 Analyse

Die Bestimmung der Invarianzebene wird durch drei Parameter wesentlich beeinflusst:

- Die Methode zur Bestimmung der vorläufigen Maske und zur Korrektur des Maskenfehlers: Dies wurde bereits ausführlich untersucht, soll aber auch in dieser Analyse berücksichtigt werden.
- Die Größe des Maskenbereichs: Tabellen 4.11, 4.12 und 4.13 zeigen, dass der Fehler bei Vergrößerung der Masken abnimmt. Eine weitere Vergrößerung der Faltungsmasken (auf bis zu $15 \times 15 \times 15$ Pixel) brachte aber nur eine minimale Verbesserung der Ergebnisse (um maximal 0,002 Grad).
- Die Wahl der Ableitungs-Kovarianzmatrix Σ_A . Der Einfluss dieses Wertes soll im Folgenden untersucht werden.

Für die Analyse des Einflusses von Σ_A auf das Ergebnis wurde eine Versuchsreihe unter folgenden Rahmenbedingungen durchgeführt:

- Die Ableitungen wurden auf dem gesamten Bild 150 der Bildfolge `marmor_stat` berechnet, unter Auslassung eines 20 Pixel breiten Randes.
- Als Kovarianzmatrix wurde stets eine Diagonalmatrix $\Sigma_A = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_t^2)$ gewählt mit $\sigma_x = \sigma_y$.
- Es galt entweder $\sigma_x = \sigma_t$ oder $\sigma_x = \sqrt{2}\sigma_t$.
- σ_x wurde im Intervall $[0,50; 6,00]$ mit der Schrittlänge 0,01 variiert. Nach Überlegungen von [Weber & Malik 93] sollte bei der Verwendung der ersten Ableitung

der Gaußfunktion σ_x mindestens so groß sein wie der Betrag der Verschiebung. Bei einer VR zwischen 1 und 2 Pixeln pro Bild ist deshalb eine Betrachtung für $\sigma_x < 0,50$ nicht sinnvoll.

- Die Maskengröße wurde automatisch gewählt unter Vorgabe eines Maskenfehlers von 1 %, bei einer maximalen Maskengröße von $31 \times 31 \times 31$ Pixel. Es wurde bewusst in Kauf genommen, dass der Fehler dadurch etwas höher ist als bei größeren Masken; dies erleichtert aber die Versuchsdurchführung erheblich.
- Die Masken wurden sowohl mit abgetasteten als auch mit integrierten Koeffizienten bestimmt und die Ergebnisse ohne Maskenkorrektur (A-, I-) und unter Verwendung der drei (AS_{mod}, AG, AGR) bzw. vier (IS_{mod}, IR, IG, IGR) noch verbliebenen Korrekturverfahren bestimmt.
- Zum Vergleich wurden auch unkorrigierte abgetastete oder integrierte Masken verwendet, die in jede Richtung einen Pixel weiter ausgedehnt waren als bei automatisch bestimmter Maskengröße, bei gleichzeitiger Anhebung der maximalen Maskengröße auf $33 \times 33 \times 33$ Pixel. Diese Masken werden nachfolgend als „vergrößerte Masken“ bezeichnet.

Insgesamt wurden also 12122 verschiedene¹ Ableitungsfelder berechnet, die zur Auswertung zur Verfügung stehen.

5.2.2 Auswertung 1: Durchschnittsfehler und Median

In einem ersten Vergleich werden die Werte von $\alpha_{v_+, \mathbf{E}_0}$ im gesamten Bereich der Abbilder der bewegten Marmorblöcke inklusive Randanteil betrachtet.

Abbildung 5.1 zeigt, wie sich Mittelwert $\overline{\alpha_{v_+, \mathbf{E}_0}}$ und Median $\text{med}(\alpha_{v_+, \mathbf{E}_0})$ des Winkels $\alpha_{v_+, \mathbf{E}_0}$ bei veränderlichem σ_x verhalten. Der Unterschied zwischen abgetasteten und integrierten Maskenkoeffizienten ist dabei sehr gering, nur für $\sigma_x < 1,5$ treten nennenswerte Unterschiede zugunsten der integrierten Maskenkoeffizienten auf. In allen vier Grafiken tritt ein globales Minimum des Mittelwerts und des Medians des Fehlers im Bereich zwischen $\sigma_x = 1,5$ und $\sigma_x = 3,5$ auf. Zum Rand hin nimmt der Fehler deutlich zu, der untersuchte Wertebereich für σ_x wurde also offensichtlich vernünftig gewählt. Die Minima des Mittelwerts $\arg \min_{\sigma_x} \overline{\alpha_{v_+, \mathbf{E}_0}}(\sigma_x)$ und des Medians $\arg \min_{\sigma_x} \text{med}(\alpha_{v_+, \mathbf{E}_0})$ werden für unterschiedliche σ_x angenommen, mit

$$\arg \min_{\sigma_x} \overline{\alpha_{v_+, \mathbf{E}_0}}(\sigma_x) < \arg \min_{\sigma_x} \text{med}(\alpha_{v_+, \mathbf{E}_0}).$$

Dies deutet darauf hin, dass im Bereich zwischen den beiden Minima der Fehler für die meisten Punkte noch abnimmt, für wenige Punkte aber bereits deutlich zunimmt. Diese

¹551 Schritte für σ_x mit je zwei Alternativen für σ_t , 2× unkorrigierte, 2× vergrößerte, 7× korrigierte Masken, also $551 \cdot 2 \cdot (2 + 2 + 7) = 12122$ Kombinationen.

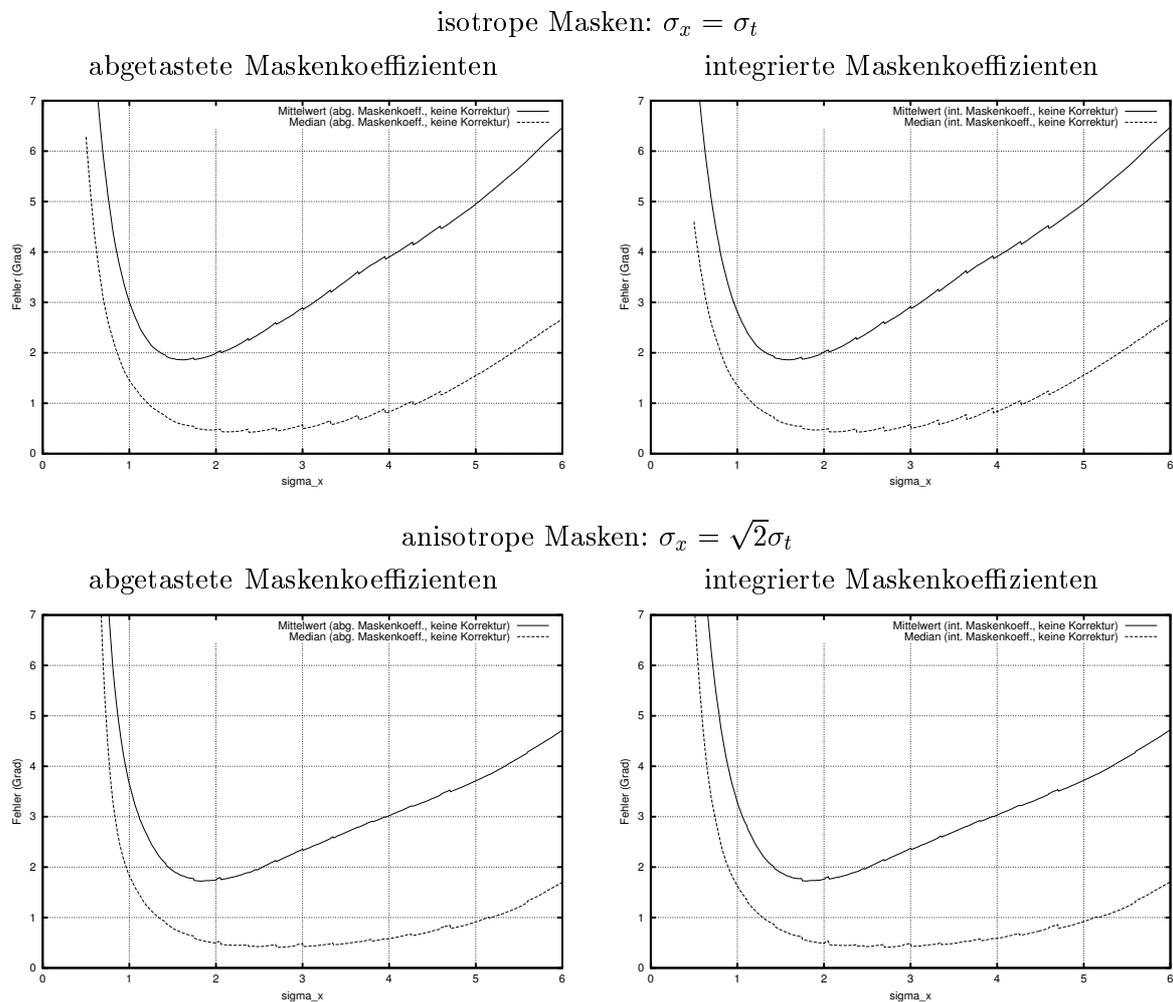


Abbildung 5.1: Mittelwert (—) und Median (— —) des Winkels $\alpha_{v_+, \mathbf{E}_0}$ im gesamten Bereich der Abbilder der bewegten Marmorblöcke inklusive Randanteil für $\sigma_x \in [0, 5..6, 0]$ bei unkorrigierten Masken. Mittelwert und Median haben jeweils ein Minimum zwischen $\sigma_x = 1,5$ und $\sigma_x = 4,5$ und steigen für $\sigma_x \rightarrow 0,5$ und $\sigma_x \rightarrow 6,0$ deutlich an. Das untersuchte Intervall für σ_x wurde also offensichtlich hinreichend groß gewählt. Weitere Kommentare im Text.

Hypothese wird durch Abbildung 5.2 bestätigt: Während der Fehler für die meisten Punkte im Inneren der Objektabbilder abnimmt, nimmt er an deren Rand zu.

Bei genauer Betrachtung der Graphen in Abbildung 5.1 fallen mehrere Sprünge der Funktionen auf.² Diese treten genau an den Stellen auf, an denen es zu einer Vergrößerung der Masken kommt. Abbildung 5.3 zeigt, dass diese „Sprünge“ durch die Maskenkorrekturen erheblich reduziert werden können: Im isotropen Fall führen bei abgetasteten Masken nur die gradienten-basierte Randkorrektur und die Vergrößerung

²Der Begriff „Unstetigkeit“ anstelle von „Sprung“ wird bewusst vermieden, da es bei diskreten Funktionen problematisch ist, von Stetigkeit zu sprechen.

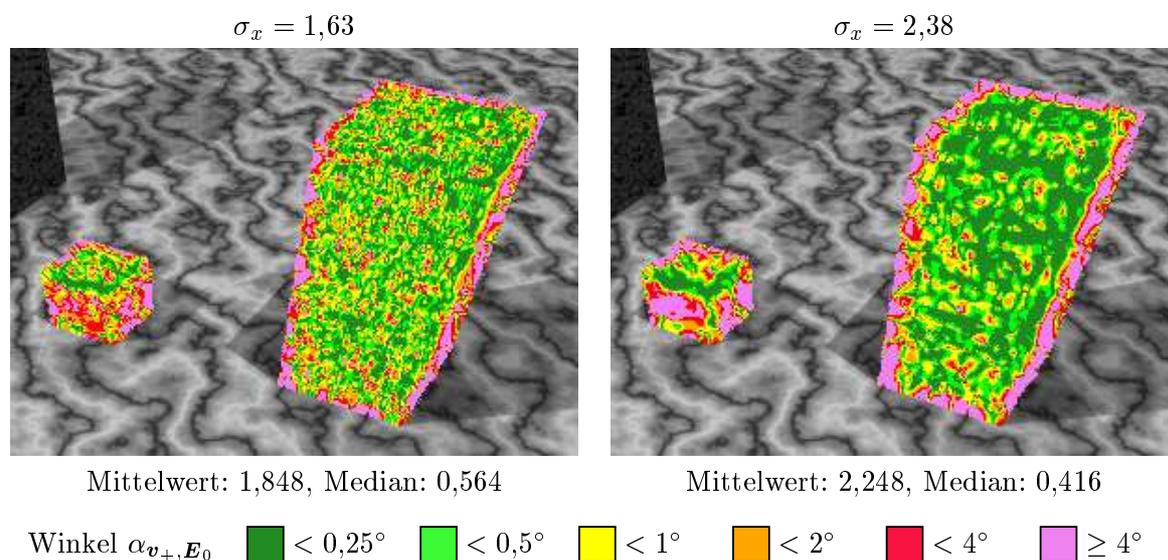


Abbildung 5.2: Vergleich von $\alpha_{v_+, \mathbf{E}_0}$ für minimalen Mittelwert (links) und minimalen Median des Winkels (rechts) im Bereich der bewegten Objektabbilder bei abgetasteten, unkorrigierten, isotropen, vergrößerten Masken: $\sigma_x = \sigma_y = \sigma_t$. Es sind nur solche Punkte eingefärbt, die zu den Abbildern der bewegten Marmorblöcke gehören und somit zu diesen Werten beitragen. Man erkennt, dass bei wachsendem σ_x der Fehler im Innern des Abbilds der bewegten Marmorblöcke abnimmt, am Rand der Blöcke aber deutlich zunimmt. Bildausschnitt: $(162,68)+[304,220]$.

der Masken zu einer Verringerung der „Sprunghöhe“, da die (nicht eingezeichneten) Skalierungsverfahren alle drei Ableitungen mit demselben Faktor skalieren und sich somit die Orientierung der geschätzten Invarianzebenen nicht ändert. Auch bei den integrierten Maskenkoeffizienten führen die gradienten-basierte Randkorrektur und die Vergrößerung der Masken zu den größten Verbesserungen, mit geringen Vorteilen gegenüber der einfachen Randkorrektur. Auch im – hier nicht dargestellten – anisotropen Fall führen die beiden Randkorrektur-Verfahren zur stärksten Reduzierung der „Sprünge“, während Skalierungen nur zu einer geringfügigeren Verbesserung oder sogar zu einer Verschlechterung führen.

Auch wenn die Sprünge in Abbildung 5.3 groß erscheinen, so ist zu bedenken, dass die Unterschiede zwischen den verschiedenen Korrekturverfahren im ungünstigsten Fall bei nur 0,06 Grad liegen. In einer ähnlichen Größenordnung liegen für $\sigma_x > 1,5$ die Unterschiede zwischen integrierten und abgetasteten Maskenkoeffizienten. Wesentlich größer ist hingegen der Unterschied zwischen verschiedenen Werten für σ_x . Die folgenden Untersuchungen sollen daher auf die Wahl von Σ konzentriert werden.

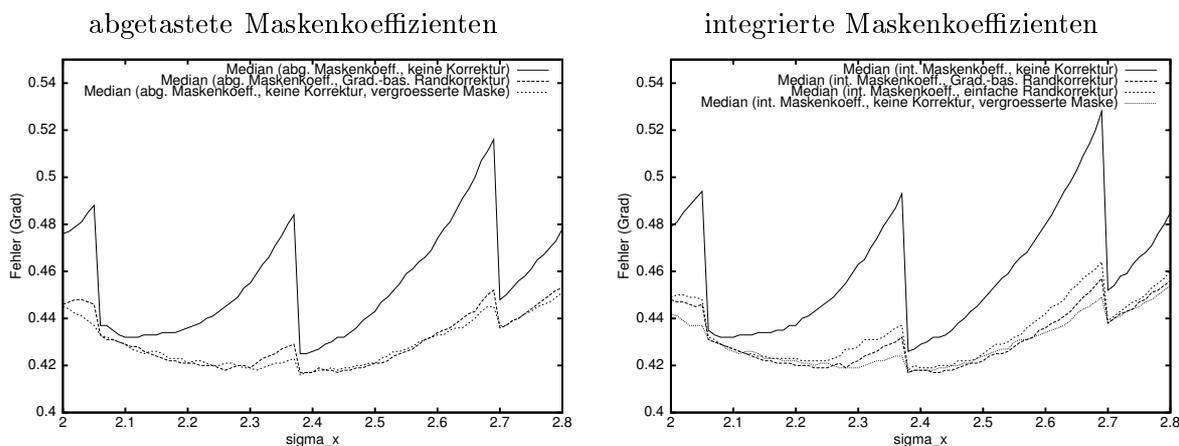


Abbildung 5.3: Median des Winkels $\alpha_{v_+, \mathbf{E}_0}$ für $\sigma_x \in [2,0..2,8]$ bei isotropen Masken und verschiedenen Korrekturverfahren. Durch die Korrekturen werden die „Sprünge“ z. T. deutlich reduziert.

5.2.3 Auswertung 2: Kleinster Fehler der Invarianzebenen

Offensichtlich gibt es kein isotropes Σ , das für alle Punkte optimal ist, vielmehr ist die optimale „Skala“ eine lokale Eigenschaft. Diese Erkenntnis ist weder neu noch überraschend. Eine ausführliche Diskussion findet sich z. B. in [Lindeberg 98a], ein Anwendungsbeispiel bei [Crowley 01]. Auch von den getesteten anisotropen Masken ist keine für alle Punkte optimal.

Daher sollen folgende Fragen geklärt werden:

1. Gibt es für jeden Punkt ein isotropes $\Sigma_{A, \text{opt}}$, für das der Winkel $\alpha_{v_+, \mathbf{E}_0}$ (nahezu) 0 ist?
2. Wenn nicht, gibt es ein anisotropes $\Sigma_{A, \text{opt}}$, für das der Winkel $\alpha_{v_+, \mathbf{E}_0}$ (nahezu) 0 ist?
3. Ist die Wahl von $\Sigma_{A, \text{opt}}$ eindeutig?
4. Wie kann $\Sigma_{A, \text{opt}}$ gefunden werden, ohne auf die bekannte VR zurückzugreifen?
5. Wie lassen sich die Punkte kennzeichnen, an denen es kein solches $\Sigma_{A, \text{opt}}$ gibt?

Die erste und dritte Frage lassen sich durch Ausprobieren empirisch beantworten. Betrachtet man σ_x wie bisher auch schon im Intervall $[0,50; 6,00]$ und inkrementiert man in Schritten der Größe 0,01, so müssen nur 551 Ableitungsfelder berechnet werden. Bei der zweiten Frage scheitert die empirische Beantwortung selbst bei Beschränkung auf diagonale Kovarianzmasken, da nun schon $551^3 \approx 167 \cdot 10^6$ Kombinationen betrachtet werden müssten. Dies ist höchstens für eine stark eingeschränkte Anzahl von Pixeln

möglich. Lässt man beliebige Kovarianzmatrizen zu, so nimmt die Anzahl der zu untersuchenden Fälle noch einmal um mindestens drei Größenordnungen zu, was nur für vereinzelte Punkte möglich ist. Die nachfolgende Betrachtung beschränkt sich daher auf Masken, die auf isotropen Gaußlocken basieren.

Für alle Berechnungen von Vergleich 2 wurden – soweit nicht ausdrücklich anders erwähnt – abgetastete Masken verwendet, bei denen die Maskengröße zunächst auf Basis eines vorgegebenen Fehlers von 1 % bestimmt und um ein Pixel erweitert wurde, um die zuvor beobachteten Sprünge bei Veränderung der Maskengröße zu reduzieren. Wie schon im ersten Vergleich werden die Werte von $\alpha_{v_+, \mathbf{E}_0}$ nur im Bereich der Abbilder der bewegten Marmorblöcke inklusive Randanteil betrachtet.

Winkel $\alpha_{v_+, \mathbf{E}_0}$ für verschiedene σ_x Abbildung 5.4 illustriert für verschiedene Werte von σ_x für jeden Punkt im Bereich des Abbilds der bewegten Marmorblöcke den Winkel $\alpha_{v_+, \mathbf{E}_0}$. Für den kleinsten betrachteten Wert $\sigma_x = 0,50$ ist der Winkel für den Großteil der Punkte größer als 4 Grad, was darauf hindeutet, dass dieser Wert für σ_x zu klein ist. Wenn man bedenkt, dass in diesem Fall ein erheblicher Fehler durch die Diskretisierung der Gaußlocke auftritt, so ist dieses Ergebnis aber auch nicht überraschend. Für $\sigma = 1,00$ und $\sigma = 1,50$ nimmt der durchschnittliche Fehler deutlich ab, bei der Verteilung der Winkel macht sich eine gewisse Struktur bemerkbar, es gibt aber immer noch Paare unmittelbar benachbarter Punkte, von denen der eine einen Winkel kleiner als $0,25^\circ$ und der andere einen Winkel von mehr als $4,00^\circ$ aufweist. Bei weiter wachsendem σ_x nimmt die „Strukturierung“ der Ergebnisse zu: Zwischen Punkten mit kleinem und großem Winkel liegen nun fast immer Punkte mit einem mittelgroßen Winkel, was auf die zunehmende Überlappung der Masken zurückzuführen ist. Gleichzeitig wächst aber auch der Randbereich an, in dem der Großteil der Punkte einen größeren Winkel aufweist.

Es fällt auf, dass es auch bei großem σ_x immer noch Punkte im Innern der Marmorblock-Abbilder gibt, an denen der Winkel sehr groß ist. Dies ist insofern überraschend, weil bei bisherigen Untersuchungen mehr oder weniger stillschweigend davon ausgegangen wurde, dass bei wachsendem Sigma, also zunehmender impliziter Glättung der Bild-daten, der Fehler des lokalen Tensors immer mehr abnimmt, solange ein Punkt durch die wachsenden Masken nicht in den Bereich von Bewegungsdiskontinuitäten gerät.

Minimum des Winkels Dies führt unmittelbar zur Frage, ob es für die betroffenen Punkte im Innern der Abbilder der bewegten Marmorblöcke überhaupt ein σ_x gibt, bei dem der Winkelfehler klein ist. Eine Antwort liefert Abbildung 5.5. Diese Abbildung zeigt für jeden Punkt den kleinsten Winkel, der bei Variation von σ_x , σ_y und σ_t im Intervall $[0,5; 6,0]$ in den angegebenen Schrittweiten auftritt. Wie man sieht, gibt es schon bei Betrachtung der Schrittweite 0,1 für die meisten Punkte (mindestens) ein σ_x , für das der Winkel kleiner ist als 0,25 Grad. Verkleinert man die Schrittweite auf 0,01, so nimmt die Anzahl der Punkte, für die der Winkel von 0,25 Grad nie unterschritten

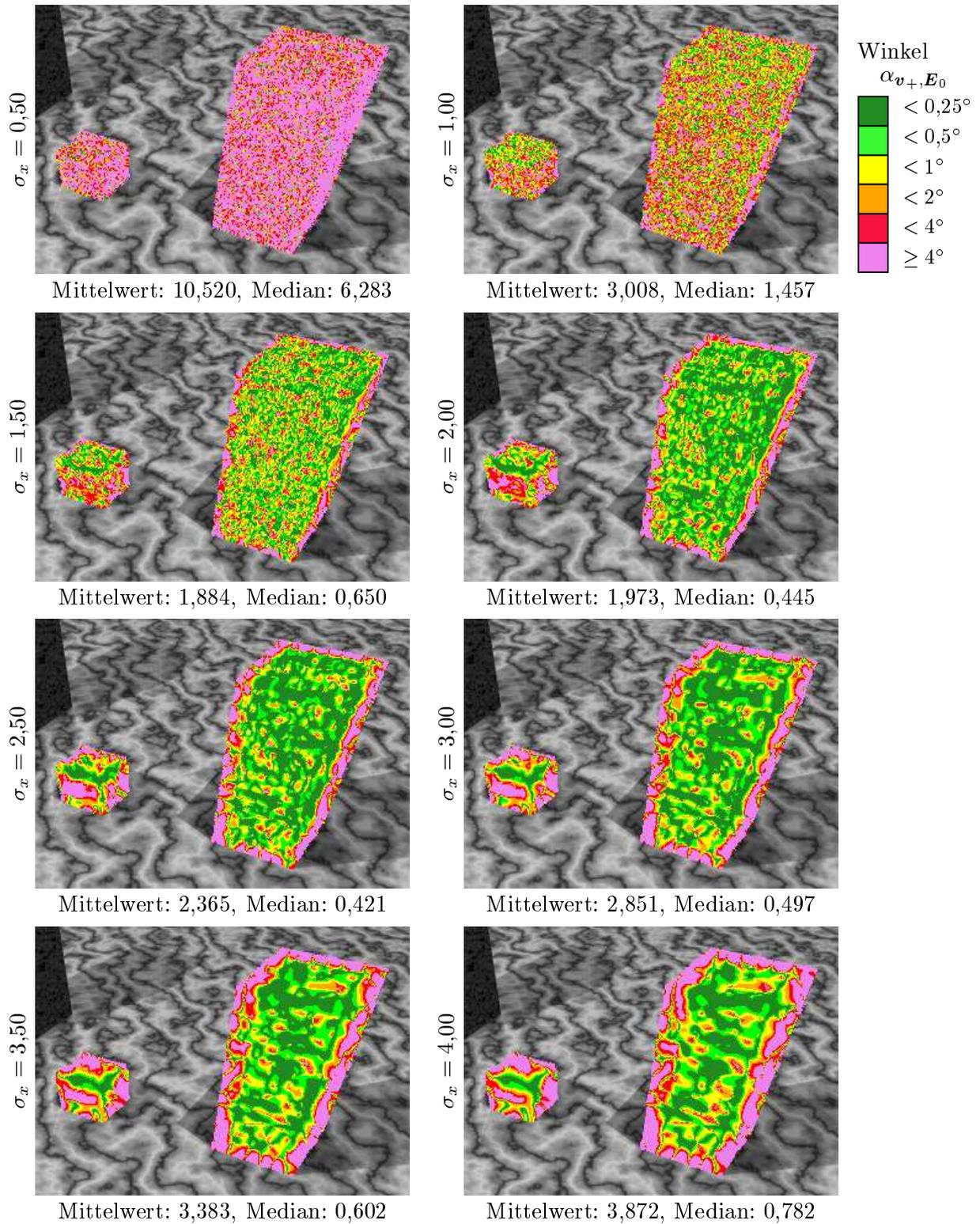


Abbildung 5.4: Vergleich von $\alpha_{v+, \mathbf{E}_0}$ für verschiedene σ_x bei abgetasteten, unkorrigierten, isotropen, vergrößerten Masken: $\sigma_x = \sigma_y = \sigma_t$. Bei wachsendem σ nimmt der Fehler im Innern des bewegten Objektabbildes ab und am Rande zu.

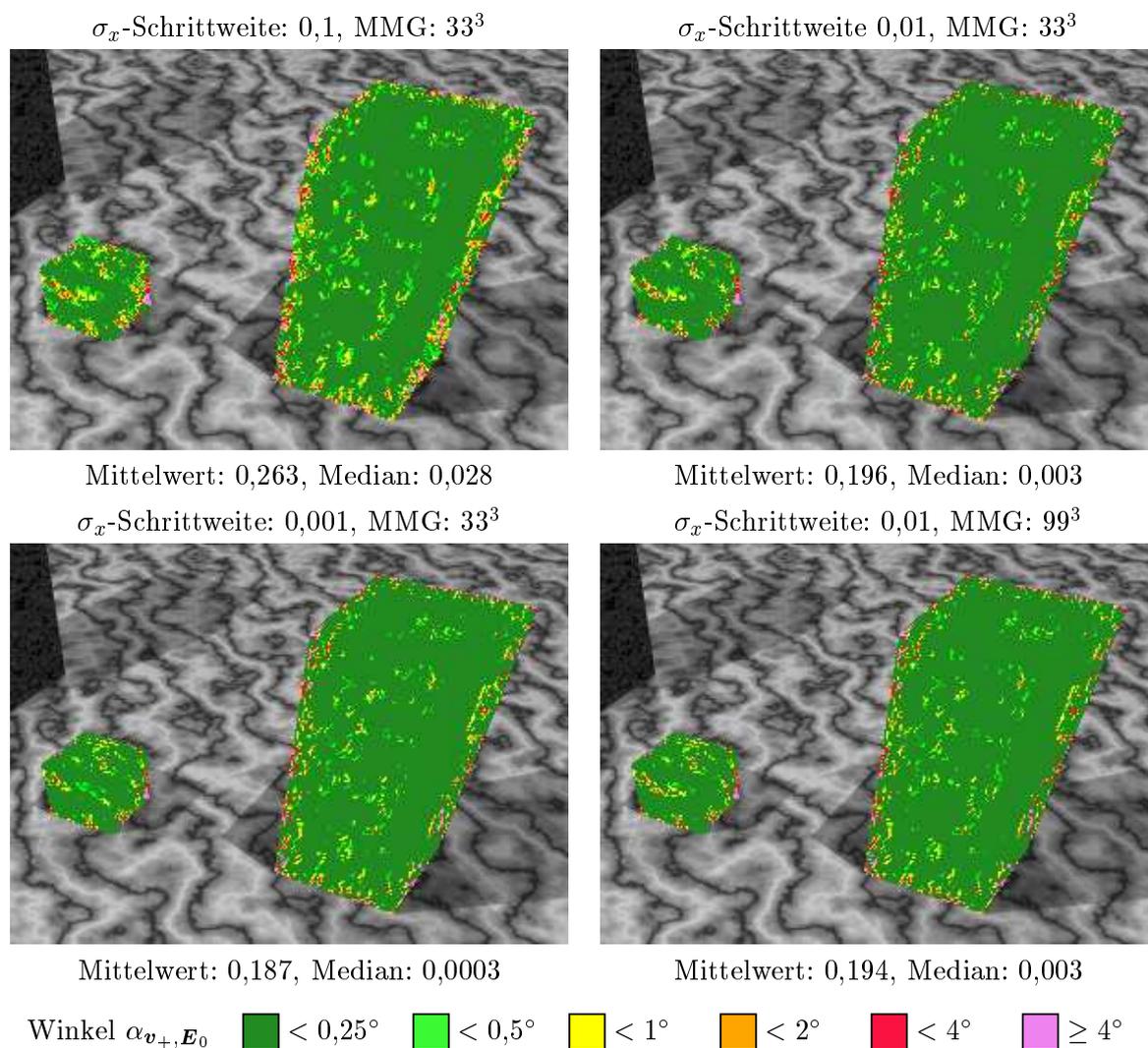


Abbildung 5.5: Kleinster auftretender Winkel $\alpha_{v_+, \mathbf{E}_0}$ bei Variation von σ_x im Bereich $[0,50; 6,00]$ in unterschiedlichen Schrittweiten und bei maximalen Maskengrößen (MMG) von 33^3 und 99^3 ; abgetastete, isotrope, vergrößerte Masken: $\sigma_x = \sigma_y = \sigma_t$. Für die meisten Punkte (dunkelgrün) gibt es mindestens ein σ , so dass der Winkel $\alpha_{v_+, \mathbf{E}_0}$ unter $0,25^\circ$ sinkt.

wird, erkennbar ab. Bei erneuter Verkleinerung der Schrittweite auf 0,001 nimmt zwar der durchschnittliche Winkel erneut ab, aber es gibt nur vereinzelt Punkte, bei denen dadurch der Winkel noch unter die Schwelle von 0,25 Grad sinkt. Eine Schrittweite von 0,01 erscheint daher als guter Kompromiss um verwertbare Ergebnisse bei akzeptabler Rechenzeit zu erzielen.

Die Punkte, die die Winkel-Schwelle von 0,25 Grad noch überschreiten, liegen zum Großteil am Rand der bewegten Marmorblöcke, es gibt aber auch solche Punkte im Innern der Blöcke. Um auszuschließen, dass das Überschreiten der Schwelle durch die Beschränkung der Maskengröße auf 33^3 verursacht wird, wurde diese Begrenzung auf

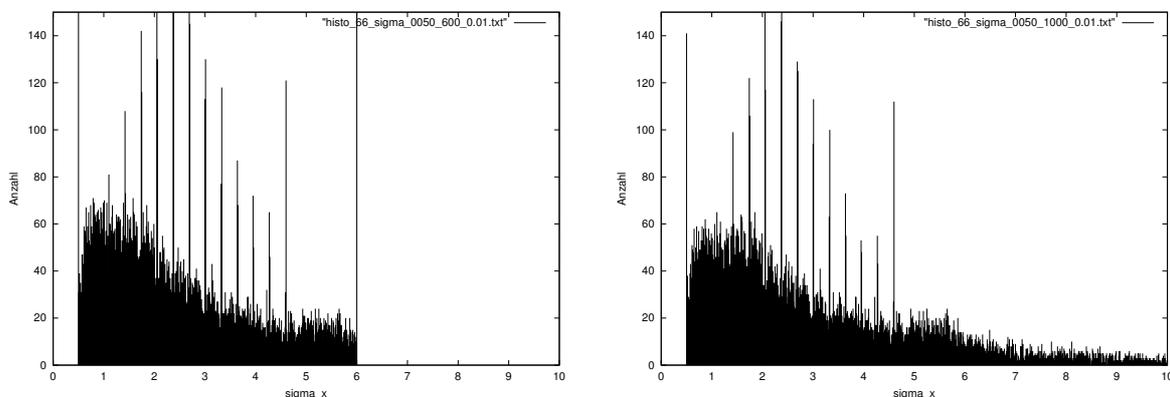


Abbildung 5.6: σ_x , bei dem der kleinste Winkel $\alpha_{v_+, \mathbf{E}_0}$ auftritt bei Variation von σ_x in Schritten von 0,01 im Bereich $[0,50; 6,00]$ (links) oder im Bereich $[0,50; 10,00]$ (rechts); abgetastete, isotrope, vergrößerte Masken: $\sigma_x = \sigma_y = \sigma_t$, max. Maskengröße: 33^3 .

99^3 angehoben – eine Grenze, die bei diesem Versuch nie erreicht wird – und der Vergleich wiederholt. Das Ergebnis unterscheidet sich aber kaum vom Ergebnis mit kleinerer maximaler Maskengröße.

Die nun naheliegende Frage lautet, für welche Werte von σ_x diese kleinen Werte für die Winkel erreicht werden.

Die Histogramme in Abbildung 5.6 geben einen ersten Überblick. Im linken Histogramm wurde σ_x wie bisher in Schritten der Länge 0,01 im Bereich $[0,50; 6,00]$ variiert. Deutlich fallen einige Werte von σ_x auf, bei denen signifikant mehr Punkte den kleinsten Winkel aufweisen als bei benachbarten Werten von σ_x . Abgesehen von diesen „Spitzen“ hat die Histogramm-Kurve ein Maximum etwa zwischen $\sigma_x = 0,8$ und $\sigma_x = 1,2$ und fällt zu kleineren Werten hin steil, zu größeren Werten hin langsam ab. Der Durchschnitt der optimalen Werte von σ_x liegt etwa bei 2,5. Dies spricht dafür, dass der Wertebereich von σ_x vernünftig gewählt wurde.

Die „Spitzen“ treten zum einen am unteren und am oberen Ende des untersuchten Wertebereichs von σ_x auf. In einer Erweiterung der bisherigen Versuche wurden daher für σ_x Werte von bis zu 10,0 untersucht. Das Ergebnis zeigt das rechte Histogramm in Abbildung 5.6. Wie man sieht, wurden nicht nur die Punkte, bei denen bisher $\sigma_x = 6,0$ optimal war, auf größere Werte von σ_x verteilt, sondern es nahmen fast alle Häufigkeiten für $\sigma_x < 6,0$ ab. Allerdings handelt es sich dabei um insgesamt eher wenige Punkte, so dass die bisherige Beschränkung des Wertebereichs aufrecht erhalten werden soll.

Abbildung 5.7 zeigt, wo im Bild sich die Punkte befinden, an denen das kleinste (links) bzw. größte (rechts) zulässige σ_x optimal ist. Die meisten dieser Punkte finden sich am Rand des bewegten Marmorblocks, aber auch im Innern der Marmorblock-Abbilder gibt es solche Punkte, wobei deren Verteilung augenscheinlich nicht völlig zufällig ist, sich die zugrundeliegende Struktur aber auch nicht sofort erschließen lässt.

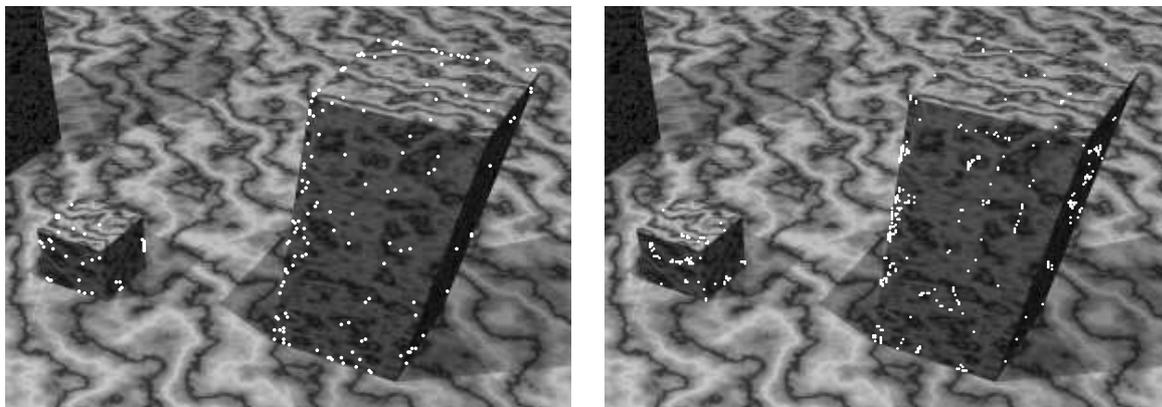
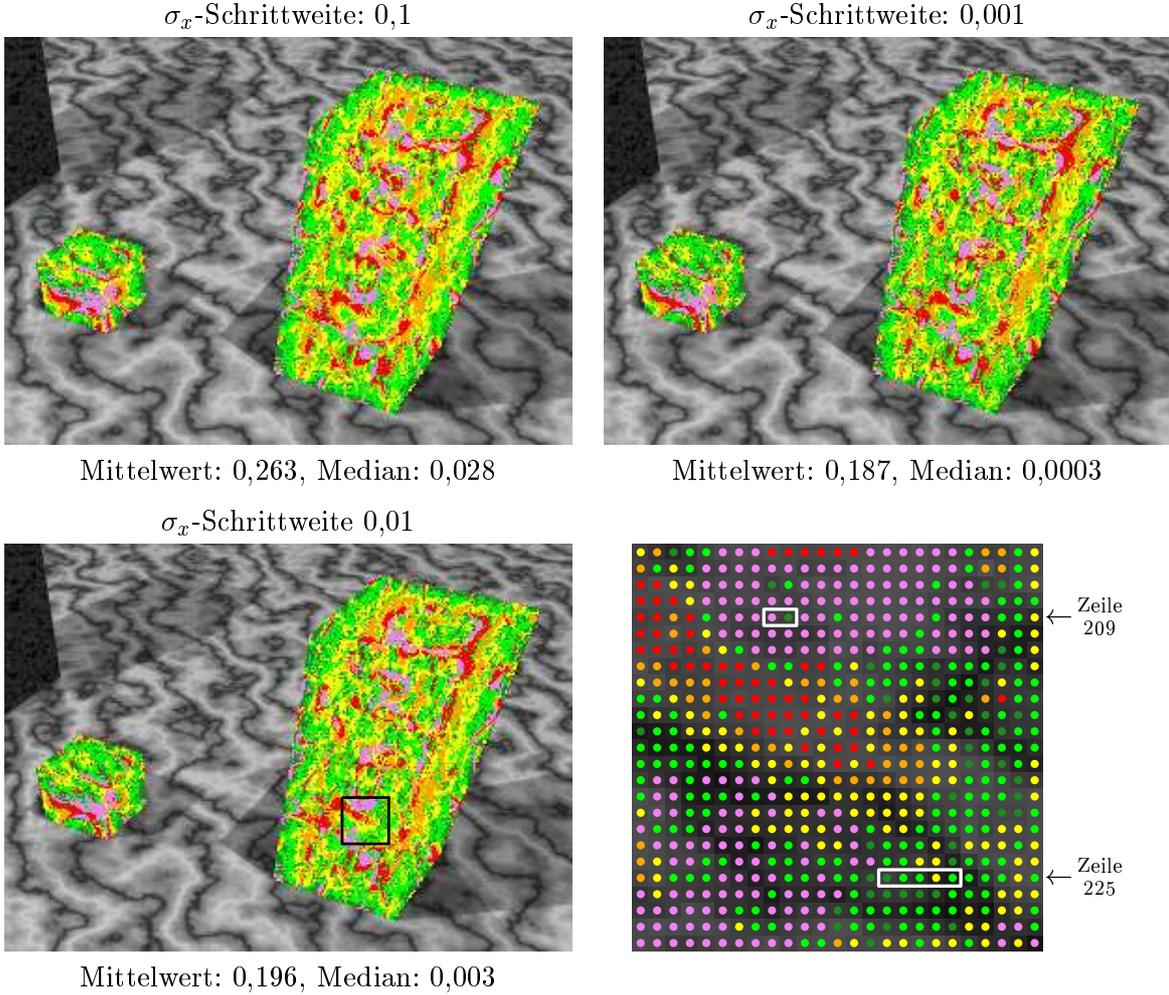


Abbildung 5.7: Lage der Punkte im Bild, an denen das kleinste (links) bzw. größte (rechts) zulässige σ_x optimal ist; abgetastete, isotrope, vergrößerte Masken: $\sigma_x = \sigma_y = \sigma_t$, max. Maskengröße: 33^3 .

Die übrigen Spitzen der Histogrammfunktion treten bei 1,42, 1,74, 2,05, 2,38, 2,69, 3,01, 3,33 usw. auf. Alle diese Werte liegen unmittelbar über oder unter einer Grenze, bei der es zu einer Veränderung der Maskengröße kommt. An diesen Stellen ist die Ableitungsfunktion bedingt durch die Beschränkung der Maskengröße unstetig in σ_x , was schon in Abbildung 5.3 augenscheinlich wurde. Die letzte auffällige Spitze im Histogramm tritt bei $\sigma_x = 4,60$ auf. Bei diesem Wert wird die maximale Maskengröße von 33^3 erreicht. Etwas überraschend ist, dass diese Effekte so deutlich auftreten, obwohl bereits ein maximaler Maskenfehler von 1% zugrunde gelegt wurde und diese Masken aufgrund vorangehender Überlegungen noch einmal um einen Pixel in jede Richtung vergrößert wurden. Als Konsequenz daraus sollte *keinesfalls* mit Masken gearbeitet werden, die einen Fehler von mehr als 1% aufweisen.

Untersucht man nun, an welchen Stellen im Bild welches σ_x optimal ist (Abbildung 5.8), so erkennt man, dass bei benachbarten Punkten in den meisten Fällen ähnliche Werte für σ_x optimal sind. Eine mögliche Erklärung für diesen Effekt besteht in der Überlappung des von den Ableitungsmasken an benachbarten Punkten erfaßten Bereichs.

Um diesen Effekt quantisieren zu können, wird der „Überlappungsfaktor“ η_A eingeführt. Er soll große Werte annehmen, wenn die Ableitungsmasken bei der Bestimmung zweier Gradientenvektoren im wesentlichen dieselben Bildpunkte mit ähnlicher Gewichtung berücksichtigen, und kleine Werte, wenn dieselben Bildpunkte mit stark unterschiedlicher Gewichtung oder gar unterschiedliche Bildbereiche in das Ergebnis der Faltungen einfließen. Sei $w_x(\Sigma_A, \mathbf{x}, \mathbf{x}_0)$ der Wert einer am Punkt \mathbf{x}_0 zentrierten dreidimensionalen x -Ableitungsmaske am Punkt \mathbf{x} , basierend auf einer Gaußfunktion mit Kovarianzmatrix Σ_A , und seien w_y und w_t für die partielle Ableitung nach y bzw. t analog definiert. Außerdem sei $\mathcal{U}_A(\mathbf{x}_0)$ die bei der Berechnung der partiellen



σ_x , bei dem der kleinste Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ auftritt

$< 1,0$
 $< 2,0$
 $< 3,0$
 $< 4,0$
 $< 5,0$
 $\geq 5,0$

Abbildung 5.8: σ_x , bei dem der kleinste Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ auftritt, bei Variation von σ_x im Bereich $[0,50; 6,00]$ in verschiedenen Schrittweiten. In der unteren Zeile wird der links markierte und rechts vergrößert dargestellte Ausschnitt für die Schrittweite 0,01 näher untersucht. Man erkennt deutlich, dass meist für größere Bereiche ähnliche Werte von σ_x optimal sind, dass es aber auch benachbarte Punkte gibt, an denen große Unterschiede zwischen den optimalen Werten von σ_x bestehen. Markiert sind hier die Punkte (351, 209) und (352, 209) (im oberen Teil des Bildes) sowie (358, 225) bis (362, 225) (im unteren Teil des Bildes), die näher untersucht werden. Einzelheiten siehe Text; abgetastete, isotrope, vergrößerte Masken: $\sigma_x = \sigma_y = \sigma_t$, max. Maskengröße: 33^3 .

Ableitungen berücksichtigte Umgebung von \mathbf{x}_0 , d. h.

$$\mathcal{U}_A(\mathbf{x}_0) = \{\mathbf{x} \mid (w_x(\Sigma_A, \mathbf{x}, \mathbf{x}_0) \neq 0) \vee (w_y(\Sigma_A, \mathbf{x}, \mathbf{x}_0) \neq 0) \vee (w_t(\Sigma_A, \mathbf{x}, \mathbf{x}_0) \neq 0)\} . \quad (5.1)$$

σ_x	$\Sigma_A = \text{diag}(\sigma_x^2, \sigma_x^2, \sigma_x^2)$								
	$\mathbf{x}_1 - \mathbf{x}_2 = (1; 0; 0)^T$			$\mathbf{x}_1 - \mathbf{x}_2 = (1; 1; 0)^T$			$\mathbf{x}_1 - \mathbf{x}_2 = (2; 0; 0)^T$		
	η_x	$\eta_y = \eta_t$	η_A	$\eta_x = \eta_y$	η_t	η_A	η_x	$\eta_y = \eta_t$	η_A
0,5	0,005	0,213	0,144	0,005	0,191	0,067	0,000	0,107	0,071
0,6	0,030	0,336	0,234	0,028	0,284	0,113	0,000	0,171	0,114
0,8	0,165	0,501	0,389	0,144	0,408	0,232	0,005	0,273	0,184
1,0	0,334	0,601	0,512	0,283	0,498	0,354	0,037	0,359	0,251
1,41	0,589	0,717	0,674	0,478	0,623	0,527	0,203	0,496	0,399
2,0	0,689	0,800	0,763	0,631	0,726	0,663	0,439	0,624	0,562
3,0	0,795	0,867	0,843	0,753	0,814	0,773	0,609	0,741	0,697
4,0	0,847	0,900	0,882	0,815	0,860	0,830	0,703	0,803	0,770
5,0	0,877	0,920	0,906	0,851	0,887	0,863	0,759	0,842	0,814
6,0	0,895	0,933	0,920	0,871	0,904	0,882	0,793	0,867	0,842
7,0	0,905	0,941	0,929	0,882	0,915	0,893	0,812	0,883	0,860
8,0	0,910	0,947	0,935	0,887	0,922	0,899	0,822	0,895	0,871
9,0	0,913	0,952	0,939	0,889	0,926	0,902	0,826	0,903	0,878
10,0	0,913	0,955	0,941	0,889	0,929	0,902	0,827	0,909	0,882

Tabelle 5.1: Richtungsüberlappungsfaktoren η_x , η_y und η_t sowie Gesamt-Überlappungsfaktor η_A für gebräuchliche Werte der Ableitungskovarianzmatrix Σ_A bei unmittelbar benachbarten, diagonal benachbarten und nicht benachbarten Punkten: Der Gesamt-Überlappungsfaktor unmittelbar benachbarter Bildpunkte liegt – für $\sigma_x \in [1/2, 6]$ – zwischen 0,144 und 0,920. Abgetastete Maskenkoeffizienten, maximale Maskengröße 33^3 .

Dann sei der Richtungsüberlappungsfaktor η_x zweier Masken zur Berechnung der partiellen Ableitung in x -Richtung mit Aufpunkten \mathbf{x}_1 und \mathbf{x}_2 und Kovarianzmatrizen Σ_1 , Σ_2 definiert durch:

$$\eta_x(\mathbf{x}_1, \mathbf{x}_2, \Sigma_1, \Sigma_2) = 1,0 - \frac{\sum_{\mathbf{x} \in \mathcal{U}_A(\mathbf{x}_1) \cup \mathcal{U}_A(\mathbf{x}_2)} |w_x(\Sigma_1, \mathbf{x}, \mathbf{x}_1) - w_x(\Sigma_2, \mathbf{x}, \mathbf{x}_2)|}{\sum_{\mathbf{x} \in \mathcal{U}_A(\mathbf{x}_1)} |w_x(\Sigma_1, \mathbf{x}, \mathbf{x}_1)| + \sum_{\mathbf{x} \in \mathcal{U}_A(\mathbf{x}_2)} |w_x(\Sigma_2, \mathbf{x}, \mathbf{x}_2)|}. \quad (5.2)$$

Mögliche Werte von η_x liegen – durch die Skalierung im Nenner – zwischen 0 und 1 mit z. B. $\eta_x = 1,0$ für $\mathbf{x}_1 = \mathbf{x}_2$, $\Sigma_1 = \Sigma_2$ und $\mathcal{U}(\mathbf{x}_1) = \mathcal{U}(\mathbf{x}_2)$, sowie $\eta_x = 0,0$ für $\mathcal{U}(\mathbf{x}_1) \cap \mathcal{U}(\mathbf{x}_2) = \emptyset$. Der Gesamt-Überlappungsfaktor η_A , der die Richtungs-Überlappungsfaktoren zusammenfasst, ergibt sich dann durch Mittelung:

$$\eta_A = \frac{\eta_x + \eta_y + \eta_t}{3}. \quad (5.3)$$

Wie man aus Tabelle 5.1 entnehmen kann, beträgt der Gesamt-Überlappungsfaktor der Ableitungsmasken unmittelbar benachbarter Punkte bei diesem Versuch zwischen

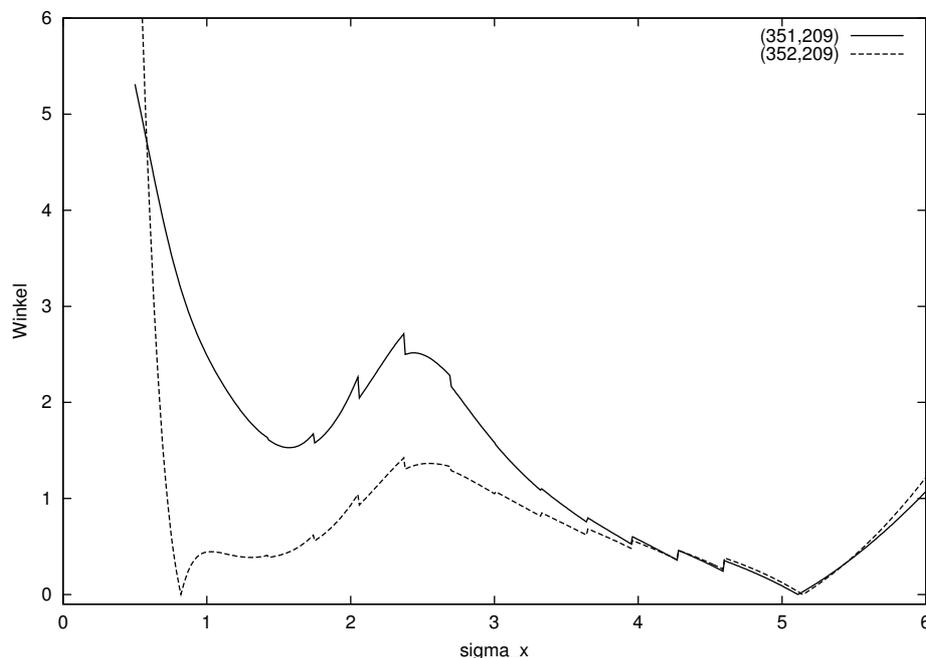


Abbildung 5.9: Vergleich des Winkels α_{v_+, E_0} in Abhängigkeit von σ_x für die Punkte (351, 209) und (352, 209).

0,144 (für $\sigma_x = 0,5$) und 0,920 (für $\sigma_x = 6,0$). Aufgrund dieser Werte ist es also nicht überraschend, wenn zumindest für große Werte von σ_x auch das optimale σ_x benachbarter Punkte oft ähnliche Werte annimmt.

Hinzu kommt, dass man auch bei geringer Überlappung das optimale σ_x als strukturelle Eigenschaft der lokalen Grauwertumgebung ansehen kann: Benachbarte Punkte haben oft eine ähnliche lokale Grauwertumgebung, und deshalb haben sie meist ähnliche Werte für das Optimum von σ_x .

Untersuchung benachbarter Punkte Durch Vergleich benachbarter Punkte soll nun versucht werden, einen genaueren Einblick in die Ergebnisse zu gewinnen.

Betrachtet man beispielsweise das benachbarte Punktepaar (351, 209) und (352, 209) (im unteren rechten Teilbild von Abbildung 5.8 durch weißen Kasten markiert), so ist für Punkt (351, 209) $\sigma_x = 5,11$ optimal, während bei Punkt (352, 209) $\sigma_x = 0,82$ die besten Ergebnisse liefert.

Abbildung 5.9 zeigt die Veränderung des Winkels α_{v_+, E_0} in Abhängigkeit von σ_x . Für große Werte von $\sigma_x (> 4)$ verhält sich der Winkel für beide Punkte in etwa gleich. Dies ist nicht weiter überraschend, da sich die Masken in diesem Fall stark überlappen (vgl. Tabelle 5.1). Für beide Punkte gibt es ein lokales Minimum des Winkel bei $\sigma_x = 5,11$ bzw. bei $\sigma_x = 5,14$ (Überlappung $\eta_A \approx 0,908$), wobei der Winkel in beiden Fällen

kleiner als 0,005 Grad ist. Für kleinere Werte von σ_x unterscheiden sich die Winkel zunehmend. Am Punkt (351, 209) besitzt die Winkelfunktion noch ein weiteres lokales Minimum bei $\sigma_x = 1,57$, wenn man von den durch die Veränderung der Maskengröße hervorgerufenen Sprüngen absieht. Dieses lokale Minimum ist aber deutlich größer als das globale Minimum. Am Punkte (352, 209) besitzt die Winkelfunktion zwei weitere Minima bei $\sigma_x = 1,31$ und $\sigma_x = 0,82$, wobei der Winkel bei letzterem das globale Minimum darstellt. Mit einem Winkel von 0,0033 Grad ist das globale Minimum aber nur geringfügig kleiner als das lokale Minimum von 0,0039 Grad für $\sigma_x = 5,14$. Inkrementiert man σ_x in kleineren Schritten, so erhält man bei $\sigma_x = 0,8205$ ein lokales Minimum von 0,00015 Grad und bei $\sigma_x = 5,1442$ ein lokales Minimum von 0,00002 Grad.

Aus dieser Untersuchung lassen sich zwei Schlussfolgerungen ziehen:

- An benachbarter Punkten ergeben sich bei großem σ_x ähnliche Gradienten ∇g und somit meist ähnliche Winkelfehler $\alpha_{v_+, \mathbf{E}_0}$, während sich bei kleinem σ_x völlig unterschiedliche Gradienten ergeben können. Ist also für einen Punkt A ein großes σ_x optimal, so kann man am benachbarten Punkt B ein zumindest lokales Minimum bei ähnlichem σ_x erwarten. Aufgrund der großen Unterschiede bei kleinem σ_x kann aber Punkt B sein global optimales σ_x auch bei kleinem σ_x aufweisen. Außerdem reagieren die Gradienten bei kleinem σ_x stärker auf Änderungen von σ_x als bei großem σ_x .
- Die Frage nach der Eindeutigkeit des optimalen $\Sigma_{A, \text{opt}}$ (vgl. Seite 85) muss aufgrund der Beobachtungen für Punkt (352, 209) offensichtlich verneint werden, da es zwei unterschiedliche Skalen gibt, für die $\alpha_{v_+, \mathbf{E}_0}$ jeweils kleiner ist als 0,0002 Grad.

In einer zweiten Detail-Untersuchung wurden die fünf benachbarten Punkte (358, 225) bis (362, 225) (im unteren rechten Teilbild von Abbildung 5.8 durch weißen Kasten markiert) betrachtet, die sich im kleinsten auftretenden Winkel $\alpha_{v_+, \mathbf{E}_0}$ deutlich unterscheiden (Abbildung 5.10). Wie schon beim vorangehend untersuchten Punktepaar zeigt sich für große σ_x ein ähnlicher Kurvenverlauf. Bemerkenswert ist aber das Verhalten des Winkels beim Punkte (359, 225). An diesem Punkt wächst der Winkel $\alpha_{v_+, \mathbf{E}_0}$ für $\sigma_x \in [2,09; 2,35]$ auf Werte von über 10 Grad an und erreicht ein Maximum bei 30 Grad. Eine genaue Betrachtung der Ergebnisse zeigt aber, dass der Betrag des Gradienten an diesem Punkt für diese Werte von σ kleiner als 0,1 ist, was etwa $\frac{1}{10}$ bis $\frac{1}{100}$ der bei dieser Bildfolge sonst üblichen Werte entspricht. Das bedeutet aber, dass diese ungenaue Invarianzebene bei einer nachfolgenden Mittelung kaum eine Rolle spielt.

Der Winkel $\alpha_{v_+, \mathbf{E}_0}$ alleine reicht also zur Beurteilung der „Qualität“ der Invarianzebenen nicht aus. Vielmehr müssen auch die Auswirkungen auf die GST-Schätzung betrachtet werden. Es lassen sich vier Fälle unterscheiden:

- kleiner Winkel $\alpha_{v_+, \mathbf{E}_0}$, großer Gradient: Dies ist der wünschenswerte Fall. Ein solcher lokaler Tensor geht mit großem Gewicht in die Glättung (Gl. 2.19) ein und „stützt“ die korrekte Lösung.

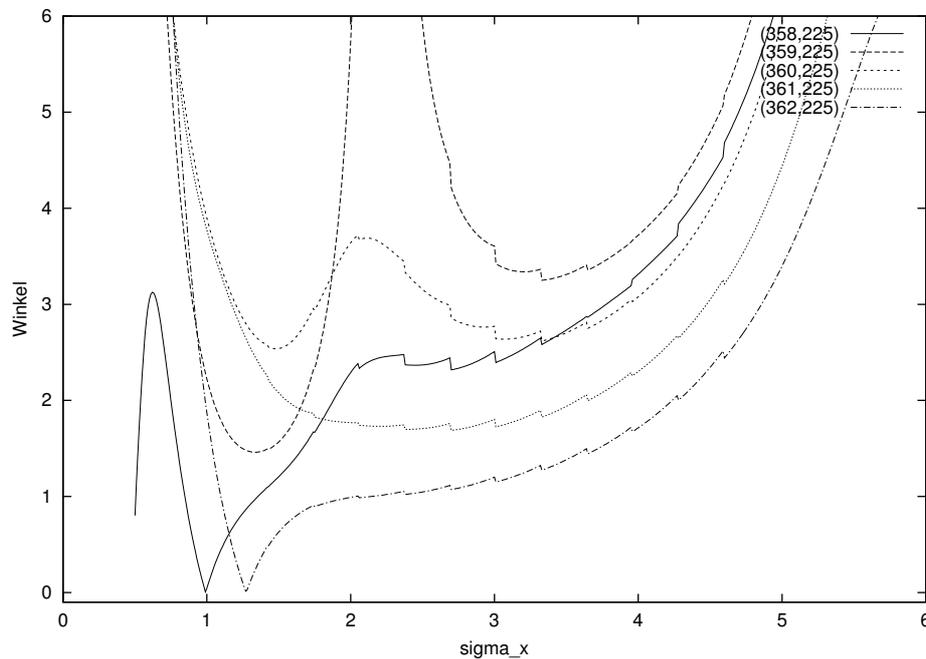


Abbildung 5.10: Wie Abbildung 5.9, aber für die Punkte (358, 225) bis (362, 225).

- kleiner Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$, kleiner Gradient: Dies ist der zweitbeste Fall. Der Tensor trägt zur korrekten Lösung bei, hat aber kein allzu großes Gewicht.
- großer Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$, kleiner Gradient: Obwohl die Invarianzebene fehlerhaft ist, richtet ein solcher lokaler Tensor kaum Schaden an, kann also noch toleriert werden.
- großer Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$, großer Gradient: Dieser Fall ist der ungünstigste Fall, da die fehlerhafte Invarianzebene auch noch großen Einfluss auf das Ergebnis hat.

Projektion $p(\nabla g, \mathbf{v}_+)$ des Gradienten auf den VR-Vektor Als Konsequenz dieser Beobachtung soll nun nicht mehr der Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ selbst, sondern der letztendlich dadurch verursachte „Schaden“ betrachtet werden. Im Idealfall stehen der Gradientenvektor ∇g und der erweiterte VR-Vektor \mathbf{v}_+ senkrecht aufeinander. Die Länge der Projektion des Gradientenvektors auf den erweiterten VR-Vektor (im Folgenden kurz Fehlerprojektion genannt) ist daher ein geeignetes Maß für den Schätzfehler des Gradienten:

$$p(\nabla g, \mathbf{v}_+) = \frac{(\nabla g)^T \mathbf{v}_+}{\|\mathbf{v}_+\|}.$$

Wie Abbildung 5.11 zeigt, nimmt die Fehlerprojektion im Innern des Marmorblock-Abbildes bei wachsendem σ ab, während ein zunehmend breiterer Rand große Werte der Fehlerprojektion aufweist. Ähnliches Verhalten war bereits bei der Betrachtung

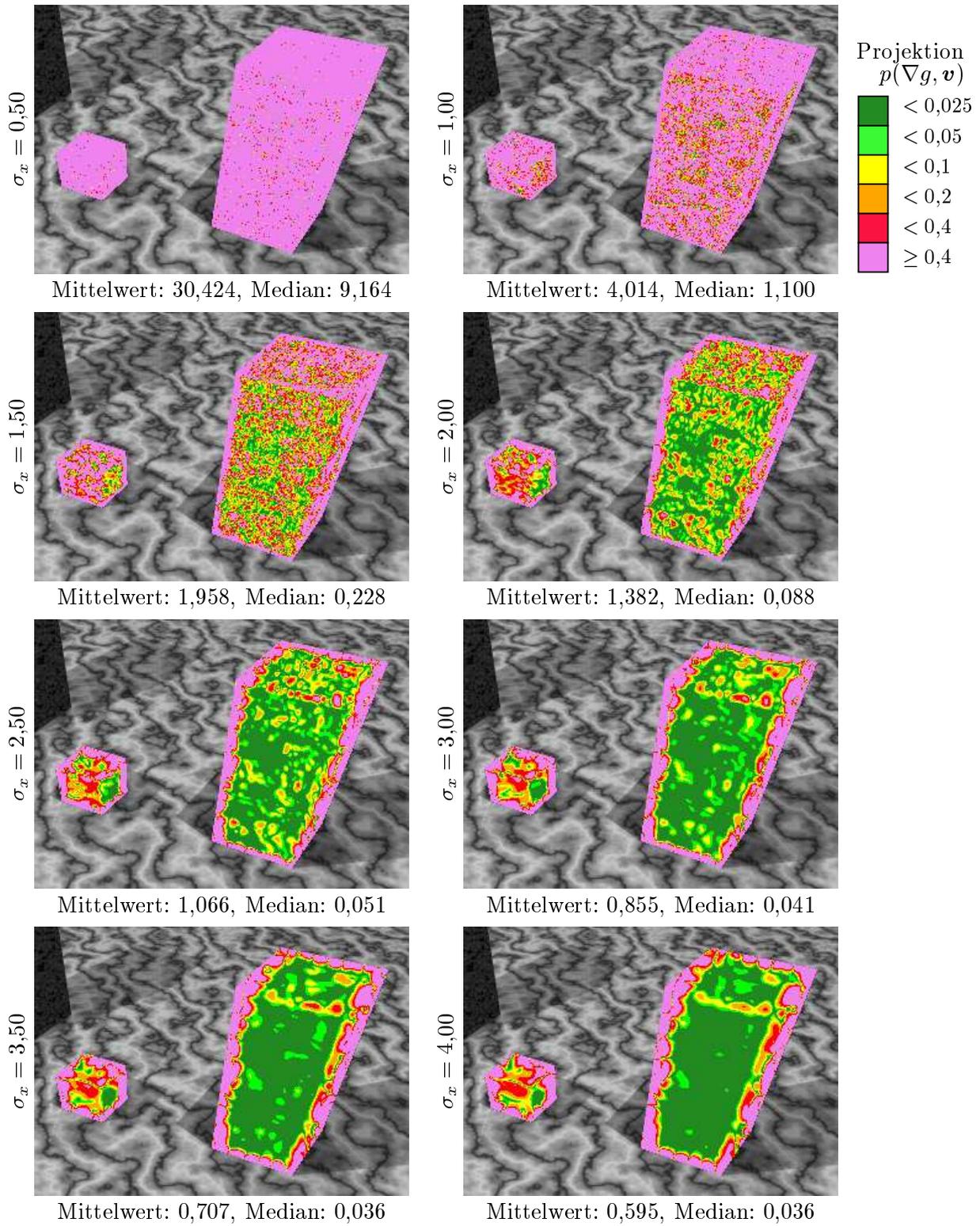


Abbildung 5.11: Vergleich von $p(\nabla g, \mathbf{v}_+) = \frac{(\nabla g)^T \mathbf{v}_+}{\|\mathbf{v}_+\|}$ für verschiedene σ_x bei abgetasteten, unkorrigierten, isotropen, vergrößerten Masken: $\sigma_x = \sigma_y = \sigma_t$.

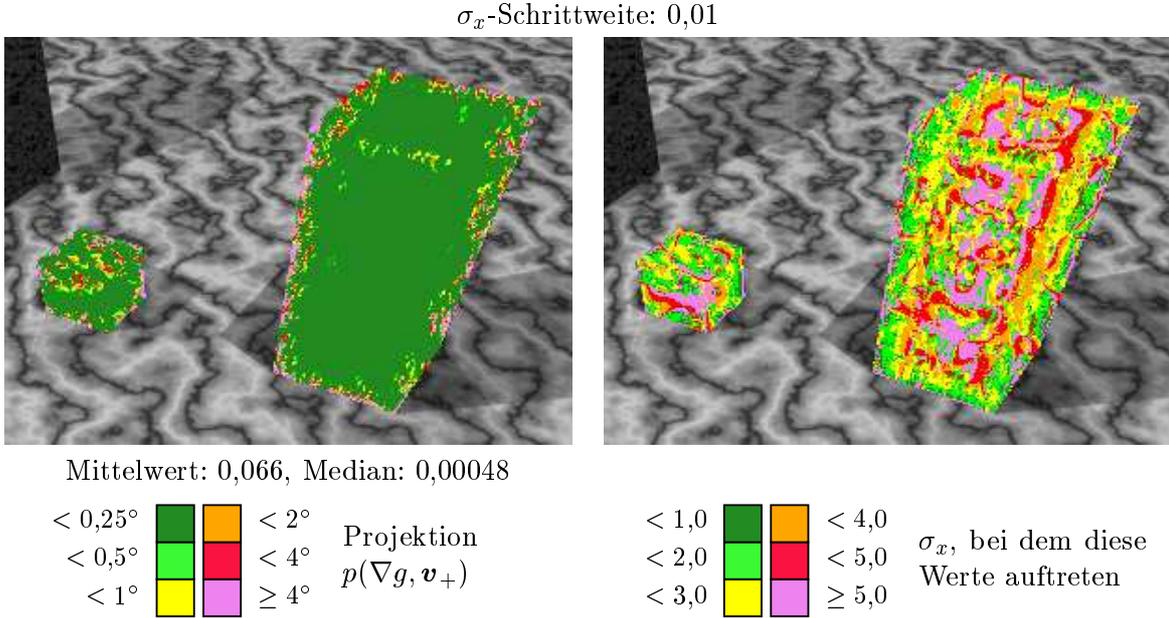


Abbildung 5.12: Kleinste Werte für $p(\nabla g, \mathbf{v}_+)$ (links) und σ_x , bei dem diese Werte auftreten, (rechts) bei Variation von σ_x im Bereich $[0,50; 6,00]$ in Schritten von 0,01; abgetastete, isotrope, vergrößerte Masken: $\sigma_x = \sigma_y = \sigma_t$, max. Maskengröße: 33^3 . Im Vergleich mit Abbildung 5.5 und Abbildung 5.8 erkennt man, dass an fast allen Punkten *im Innern* des Marmorblocks, bei denen der kleinste auftretende Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ groß ist, $p(\nabla g, \mathbf{v}_+)$ trotzdem einen kleinen Wert annimmt, und dass der ‘optimale’ Wert für σ_x oft zunimmt.

des Winkels $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ festgestellt worden. Für großes σ_x ist sogar zu erkennen, wie sich die Kanten im Innern der Marmorblöcke auswirken. Im Gegensatz zur Betrachtung des Winkels $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ liegt das Minimum des Mittelwerts der Projektion (0,595) erst am Rand des betrachteten Bereichs, nämlich bei $\sigma_x = 6,00$.

Auch die Darstellung der kleinsten auftretenden Werte von $p(\nabla g, \mathbf{v}_+)$ und der zugehörigen σ_x (Abbildung 5.12) ähnelt stark der entsprechenden Abbildung für $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$.

Beim Histogramm der σ_x , für die die optimalen Werte erreicht werden (Abbildung 5.13), ist im Vergleich zu dem entsprechenden Histogramm beim Winkel $\alpha_{\mathbf{v}_+, \mathbf{E}_0}$ eine Verschiebung zu größeren σ_x hin festzustellen. Diese Verschiebung ist – ebenso wie die Verschiebung des Medians – darauf zurückzuführen, dass bei wachsendem σ_x die implizite Glättung der Bilder zunimmt und dadurch die Gradienten und somit auch die Fehlerprojektion kleinere Werte annehmen.

Zusammenfassend kann festgestellt werden, dass für fast alle rand-fernen Punkte des Abbilds des bewegten Marmorblocks die *Fehlerprojektion* mindestens für ein σ_x sehr kleine Werte annimmt, auch an den Punkten, bei denen der kleinste auftretende *Winkel* noch groß ist.

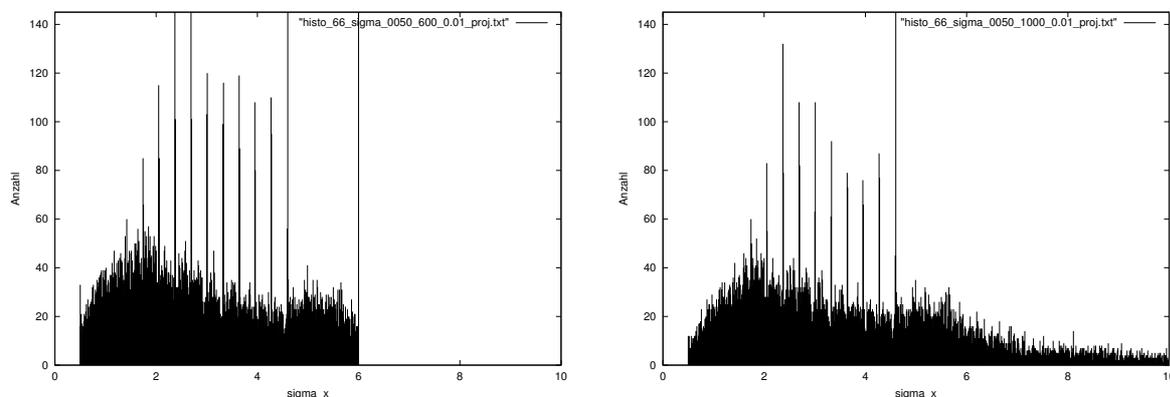


Abbildung 5.13: σ_x , bei dem die kleinste Fehlerprojektion $p(\nabla g, \mathbf{v}_+)$ auftritt bei Variation von σ_x in Schritten von 0,01 im Bereich $[0,50; 6,00]$ (links) oder im Bereich $[0,50; 10,00]$ (rechts); abgetastete, isotrope, vergrößerte Masken: $\sigma_x = \sigma_y = \sigma_t$, max. Maskengröße: 33^3 .

5.3 Skalenabhängigkeit des GST

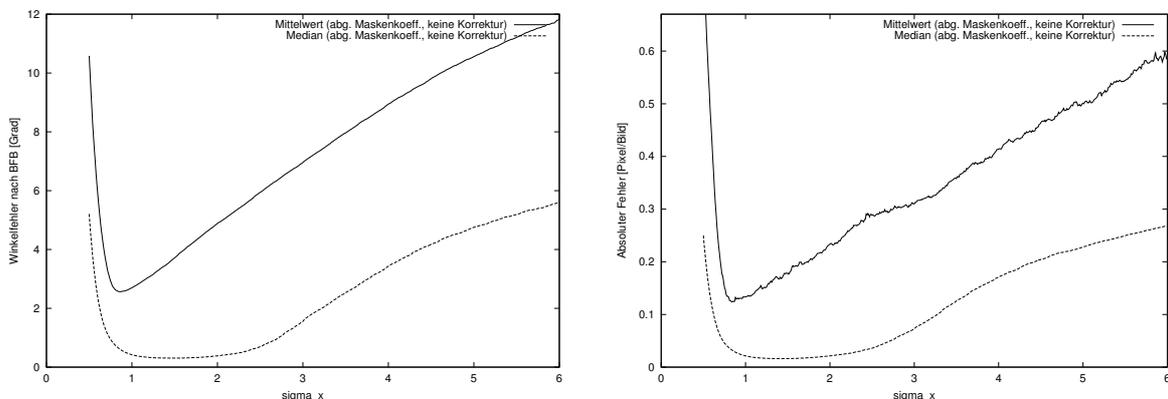
5.3.1 Analyse

In einer weiteren Versuchsreihe sollen die Untersuchungen nun auf die Bestimmung des GST und des OF ausgedehnt werden.

Die Bestimmung des GST wird im Wesentlichen von denselben Parametern beeinflusst wie die Bestimmung der Invarianzebenen (siehe Abschnitt 5.2.1). Es besteht allerdings die Möglichkeit, bei der Ableitungsbestimmung und der anschließenden Mittelung unterschiedliche Kovarianzmatrizen Σ_A bzw. Σ_M zu verwenden. In Anlehnung an [Lindenberg & Gårding 94; Nagel & Gehrke 98a] fixieren wir das Verhältnis $\Sigma_M = 2\Sigma_A$.

Für die nachfolgende Versuchsreihe gelten als Rahmenbedingungen:

- Der GST wird in einem Ausschnitt aus Bild 150 der Bildfolge `marmor_stat` berechnet, der alle bewegten Objektbilder umfasst.
- Als Kovarianzmatrizen werden stets Diagonalmatrizen $\Sigma_A = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_t^2)$, $\Sigma_M = 2\Sigma_A$ gewählt mit $\sigma_x = \sigma_y = \sigma_t$. Die Identifizierung der Skala erfolgt vereinfacht durch Angabe von σ_x .
- σ_x wurde im Intervall $[0,50; 6,00]$ mit der Schrittlänge 0,01 variiert.
- Die Maskengröße wurde in jede Richtung um einen Pixel größer gewählt als die unter Vorgabe eines Maskenfehlers von 1% automatisch ermittelte Größe, bei maximaler Maskengröße von $33 \times 33 \times 33$ Pixel.
- Die Maskenkoeffizienten wurden durch Abtastung gewonnen und keinerlei Korrekturen unterworfen. Nach der Sprechweise der vorangehenden Versuche wurden



Winkelfehler $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ nach [Barron et al. 94]

absoluter Fehler $\|\mathbf{v} - \mathbf{u}\|_2$

Abbildung 5.14: Mittelwert (—) und Median (---) des Winkels $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ (links) und der Euklidnorm $\|\mathbf{v} - \mathbf{u}\|_2$ (rechts) im gesamten Bereich der Abbilder der bewegten Marmorblöcke inklusive Randanteil für $\sigma_x \in [0,5..6,0]$ bei unkorrigierten, vergrößerten, abgetasteten, anisotropen Masken. Beide „Qualitätsmaße“ liefern in etwa dieselben Ergebnisse. Der Mittelwert hat jeweils ein Minimum etwa bei $\sigma_x = 0,8 - 0,9$, der Median zwischen $\sigma_x = 1,0$ und $\sigma_x = 2,0$. Weitere Kommentare im Text.

also abgetastete, isotrope, vergrößerte, unkorrigierte Masken verwendet.

Es wurden in diesem Versuch 551 verschiedene GST-Felder berechnet und verglichen.

5.3.2 Auswertung 1: Durchschnittsfehler und Median

Abbildung 5.14 zeigt Mittelwert und Median des Winkels $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ und der Euklidnorm $\|\mathbf{v} - \mathbf{u}\|_2$ in Abhängigkeit von σ_x bei unkorrigierten, vergrößerten, abgetasteten Masken. Dabei werden nur Punkte betrachtet, die im Bereich der Abbilder der bewegten Marmorblöcke liegen. Für ausgewählte Werte von σ_x zeigen Abbildungen 5.15 und 5.16 den Fehler an den einzelnen Punkten bei abgetasteten, unkorrigierten, isotropen, vergrößerten Masken.

- Beide „Qualitätsmaße“ liefern in etwa dieselben Ergebnisse.
- Der Mittelwert hat jeweils ein Minimum etwa bei $\sigma = 0,8$ bis $0,9$, bei größeren und kleineren Werten steigt der durchschnittliche Fehler schnell an.
- Der Median nimmt im gesamten Intervall zwischen $\sigma = 1,0$ und $\sigma = 2,0$ etwa dieselben niedrigen Werte an, nur bei noch größeren und kleineren Werten nimmt der Median merklich zu.
- Die Beobachtung, dass der Median über ein größeres Intervall von Skalen nahezu konstant bleibt, der Mittelwert aber im selben Intervall um rund 80 Prozent

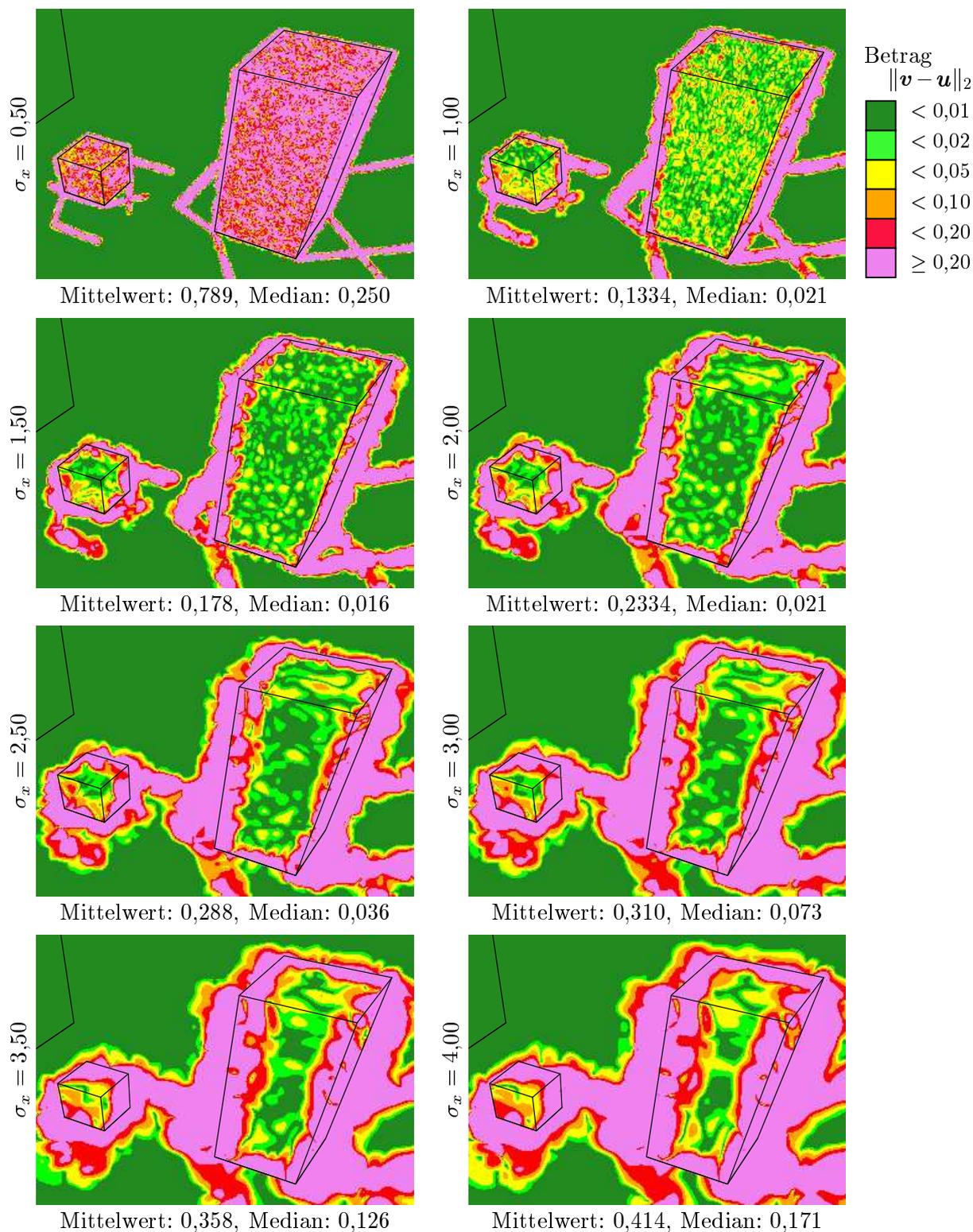


Abbildung 5.15: Vergleich des absoluten Fehlers der OF-Schätzung $\|\mathbf{v} - \mathbf{u}\|_2$ für verschiedene σ_x ($= \sigma_y = \sigma_t$). Bei wachsendem σ_x treten in einem zunehmend breiten Streifen (bedingt durch die Verbreiterung der Faltungsmasken) große Fehler auf, während der Fehler im Innern des bewegten Objektabbildes bis etwa $\sigma = 3,0$ abnimmt und dann auch anwächst.

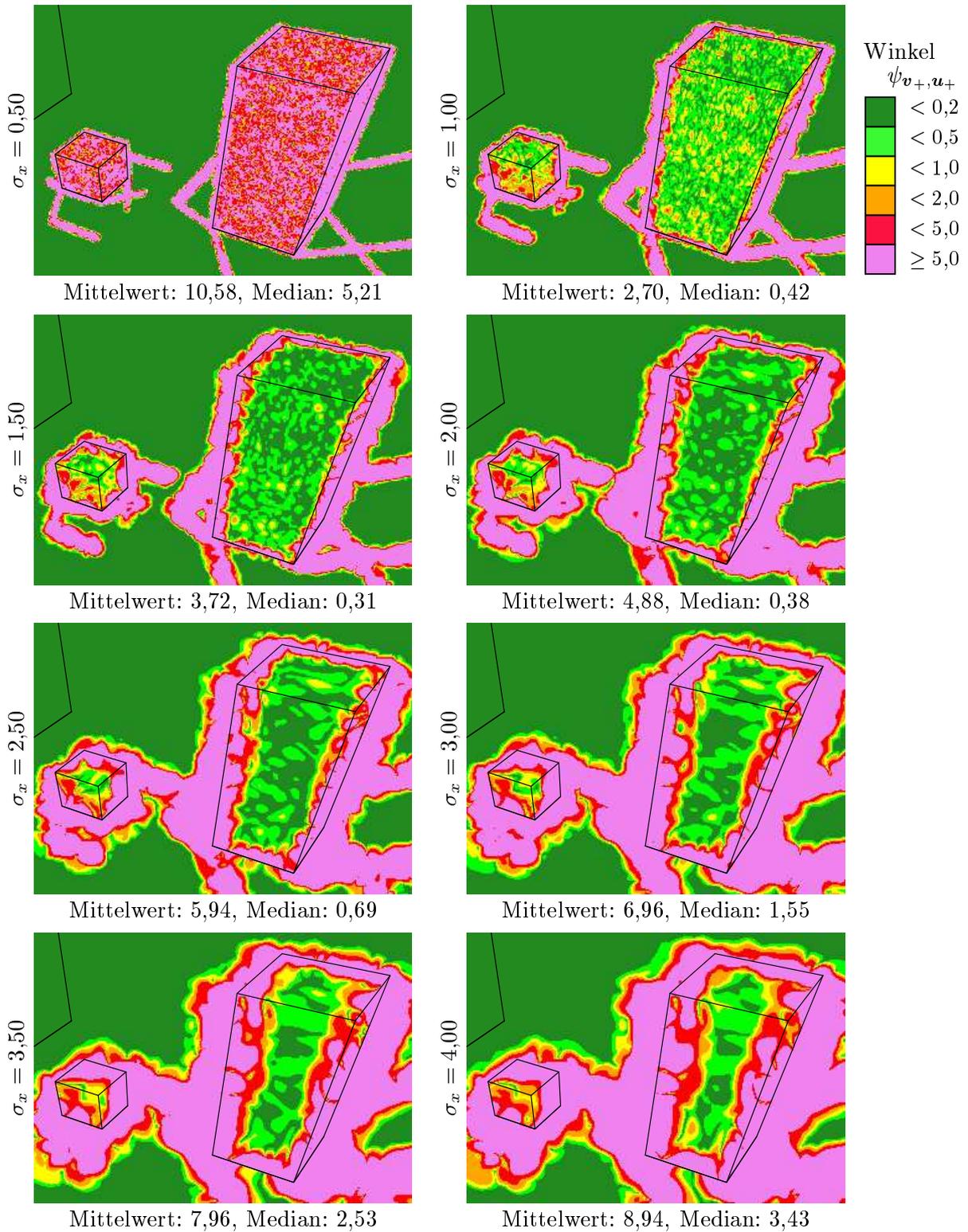


Abbildung 5.16: Vergleich des Fehlers der GST-Orientierung ψ_{v_+,u_+} für verschiedene σ_x bei abgetasteten, unkorrigierten, isotropen, vergrößerten Masken: $\sigma_x = \sigma_y = \sigma_t$. Es lassen sich die gleichen Beobachtungen wie bei Abbildung 5.15 machen, nur dass der Fehler im Innern des bewegten Objektbildes schon ab etwa $\sigma_x = 2,5$ mehrheitlich zunimmt.

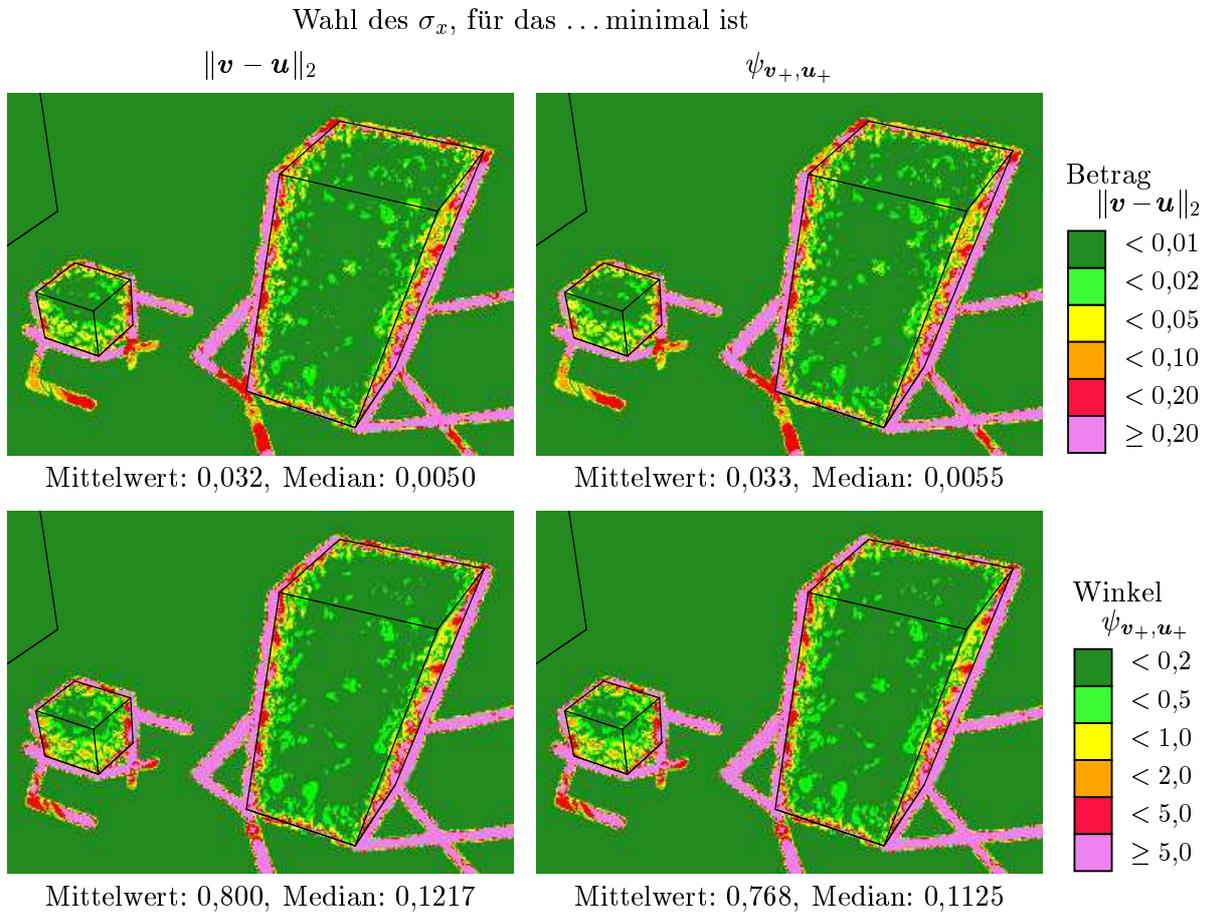


Abbildung 5.17: Kleinste auftretende Fehler $\|\mathbf{v} - \mathbf{u}\|_2$ bzw. $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ bei Variation von σ_x bei abgetasteten, unkorrigierten, isotropen, vergrößerten Masken: $\sigma_x = \sigma_y = \sigma_t$, $\Sigma_M = 2\Sigma_A$.

wächst, ist darauf zurückzuführen, dass der Fehler an relativ wenigen Punkten (im Randbereich) stark zunimmt, während der Fehler an den übrigen Punkten (im Innern der Blöcke) sogar abnimmt.

5.3.3 Auswertung 2: Kleinster Fehler des GST

Abbildung 5.17 zeigt, dass es für fast alle Punkte ein σ_x gibt, für das sowohl $\|\mathbf{v} - \mathbf{u}\|_2$ als auch $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ sehr kleine Werte annehmen. Nur entlang der bewegten Objektkanten, d. h. bei Bewegungsdiskontinuitäten, treten selbst im besten Fall größere Werte auf.

In Analogie zu Abbildung 5.9 untersucht Abbildung 5.18 die Veränderung des Fehlers des GST bei Variation von σ_x an den Punkten (351, 209) und (352, 209). Im Gegensatz zur Betrachtung der Invarianzebenen verhält sich der Fehler an beiden Punkten schon für Werte von $\sigma_x > 1,5$ sehr ähnlich. Dies ist aufgrund der Überlappung der Mittelungsmasken auch nicht weiter erstaunlich. Zur Quantisierung der Überlappung

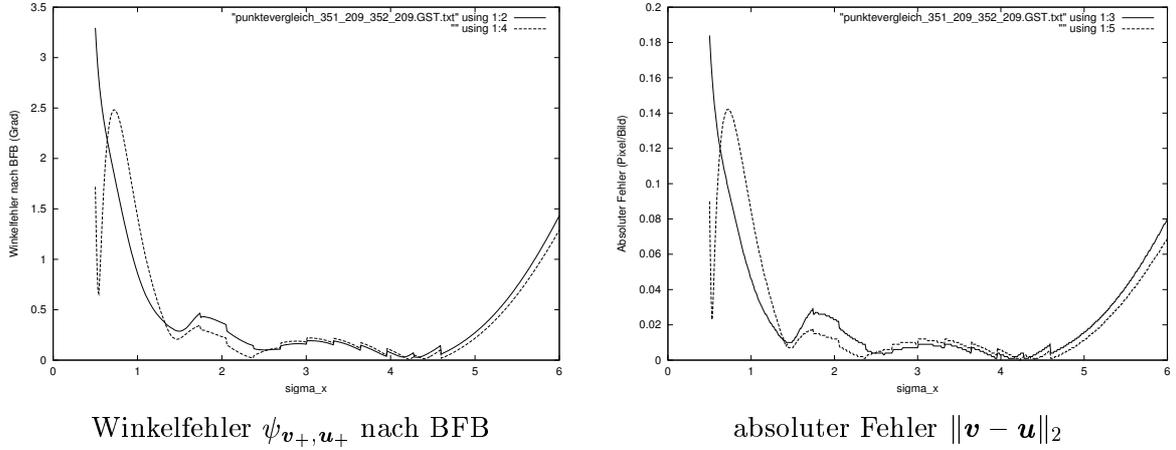


Abbildung 5.18: Vergleich des Winkelfehlers $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ und des absoluten Fehlers $\|\mathbf{v} - \mathbf{u}\|_2$ in Abhängigkeit von σ_x für die Punkte (351, 209) und (352, 209).

kann auch für Mittelungsmasken ein „Überlappungsfaktor“ η_M eingeführt werden. Sei $w(\Sigma_M, \mathbf{x}, \mathbf{x}_0)$ der Wert einer am Punkt \mathbf{x}_0 zentrierten dreidimensionalen Mittelungsmaske am Punkt \mathbf{x} mit Mittelungskovarianzmatrix Σ_M , und $\mathcal{U}_M(\mathbf{x}_0)$ die bei der Mittelung berücksichtigte Umgebung von \mathbf{x}_0 , d. h. $\mathcal{U}_M(\mathbf{x}_0) = \{\mathbf{x} \mid w(\Sigma_M, \mathbf{x}, \mathbf{x}_0) \neq 0\}$. Dann sei der Überlappungsfaktor zweier Mittelungsmasken mit Aufpunkten \mathbf{x}_1 und \mathbf{x}_2 sowie Kovarianzmatrizen Σ_1, Σ_2 definiert durch:

$$\eta_M(\mathbf{x}_1, \mathbf{x}_2, \Sigma_1, \Sigma_2) = 1,0 - \frac{\sum_{\mathbf{x} \in \mathcal{U}_M(\mathbf{x}_1) \cup \mathcal{U}_M(\mathbf{x}_2)} |w(\Sigma_1, \mathbf{x}, \mathbf{x}_1) - w(\Sigma_2, \mathbf{x}, \mathbf{x}_2)|}{\sum_{\mathbf{x} \in \mathcal{U}_M(\mathbf{x}_1)} |w(\Sigma_1, \mathbf{x}, \mathbf{x}_1)| + \sum_{\mathbf{x} \in \mathcal{U}_M(\mathbf{x}_2)} |w(\Sigma_2, \mathbf{x}, \mathbf{x}_2)|}. \quad (5.4)$$

Durch die gewählte Skalierung ist sichergestellt, dass die möglichen Werte von η_M zwischen 0 und 1 liegen.

Wie man Tabelle 5.2 entnehmen kann, beträgt der Überlappungsfaktor für $\sigma_x = 1,41$ bereits 0,8, für $\sigma = 2,0$ sogar 0,860. Für größere σ_x ist daher eine näherungsweise Übereinstimmung des Verlaufs von Winkelfehler und absolutem Fehler an benachbarten Punkten zu erwarten.

Das „optimale“ σ_x ist auch bei der GST-Bestimmung nicht eindeutig. So besitzt z. B. der absolute Fehler $\|\mathbf{v} - \mathbf{u}\|_2$ am Punkt (351, 209) neben dem globalen Minimum von 0,00106 Pixel/Bild für $\sigma_x = 4,28$ noch zahlreiche weitere lokale Minima, von denen zwei – nämlich 0,00115 Pixel/Bild für $\sigma_x = 4,13$ und 0,00127 Pixel/Bild für $\sigma_x = 3,95$ – nahe am Wert des globalen Minimum liegen. Auch für den Punkt (352, 209) gibt es vier lokale Minima – 0,00126 Pixel/Bild für $\sigma_x = 2,37$, 0,00034 Pixel/Bild für $\sigma_x = 4,22$, 0,00049 Pixel/Bild für $\sigma_x = 4,41$, 0,00066 Pixel/Bild für $\sigma_x = 4,60$ – die zumindest in der praktischen Anwendung gleichwertig sind.

σ_x	$\Sigma_A = \text{diag}(\sigma_x^2, \sigma_x^2, \sigma_x^2)$					
	$\mathbf{x}_1 - \mathbf{x}_2 = (1; 0; 0)^T$		$\mathbf{x}_1 - \mathbf{x}_2 = (1; 1; 0)^T$		$\mathbf{x}_1 - \mathbf{x}_2 = (2; 0; 0)^T$	
	$\Sigma_M = \Sigma_A$	$\Sigma_M = 2\Sigma_A$	$\Sigma_M = \Sigma_A$	$\Sigma_M = 2\Sigma_A$	$\Sigma_M = \Sigma_A$	$\Sigma_M = 2\Sigma_A$
	η_M	η_M	η_M	η_M	η_M	η_M
0,5	0,179	0,436	0,156	0,357	0,068	0,228
0,75	0,468	0,624	0,382	0,521	0,250	0,383
1,0	0,601	0,719	0,498	0,625	0,359	0,499
2,0	0,801	0,860	0,727	0,804	0,625	0,728
3,0	0,867	0,907	0,815	0,868	0,742	0,816
4,0	0,901	0,930	0,860	0,901	0,804	0,862
5,0	0,921	0,947	0,888	0,923	0,843	0,894
6,0	0,935	0,959	0,907	0,939	0,870	0,919
7,0	0,946	0,969	0,922	0,951	0,893	0,937
8,0	0,955	0,976	0,934	0,961	0,911	0,952
9,0	0,963	0,981	0,944	0,969	0,926	0,962
10,0	0,969	0,985	0,952	0,975	0,939	0,970
1,41		0,800				
2,81		0,900				

Tabelle 5.2: Überlappungsfaktor η_M für gebräuchliche Werte der Mittelungskovarianzmatrix Σ_M bei unmittelbar benachbarten, diagonal benachbarten und nicht benachbarten Punkten: Für $\Sigma_M = 2\Sigma_A$ – wie bisher stets verwendet – und $\sigma_x \in [1/2, 6]$ liegt der Überlappungsfaktor unmittelbar benachbarter Mittelungsmasken zwischen 0,436 und 0,959. Abgetastete Maskenkoeffizienten, maximale Maskengröße 33^3 .

5.4 Fazit

- Selbst bei den hier untersuchten rauschfreien Bilddaten ist die Wahl von Σ_A und Σ_M keineswegs unbedeutend, auch bei Punkten, die nicht in der Nähe von Bewegungsdiskontinuitäten liegen.
- Abgesehen von Punkten in der Nähe von Bewegungsdiskontinuitäten gibt es aber für fast alle Punkte ein *isotropes* Σ , für das die korrekte Invarianzebene geschätzt wird bzw. der OF mit der VR übereinstimmt. Dieses Σ ist aber nicht unbedingt eindeutig, wie einige der untersuchten Beispiele zeigen.
- Die Invarianzebenen sind u. a. fehlerbehaftet, wenn der geschätzte Gradient aufgrund der impliziten Mittelung klein wird. Solche Punkte wirken sich aber bei einer anschließenden GST-Bestimmung durch Mittelung kaum aus.
- Es gibt keine offensichtliche Vorgehensweise, wie das optimale Σ bei unbekannter VR gefunden werden kann.

Kapitel 6

Zu einer adaptiven GST-Bestimmung

In diesem Kapitel wird ein iterativer Ansatz zur Schätzung der „optimalen“ Skala untersucht. Wie sich zeigen wird, ist die iterative Formulierung insofern „wohldefiniert“, als sie in den meisten Fällen ein Ergebnis liefert, das sowohl mathematisch begründbar als auch anschaulich sinnvoll ist.

6.1 Motivation

Wie die bisherigen Versuche gezeigt haben, gibt es keine Diagonalmatrizen Σ_A und Σ_M (in einer vernünftig gewählten Teilmenge aller Diagonalmatrizen), so dass die Bestimmung der Invarianzebene, des GST oder des daraus abgeleiteten OFs für *alle* Punkte gute bis sehr gute Ergebnisse liefert. Ebenso sind keine Anzeichen dafür zu erkennen, dass es fest vorgebbare Diagonalmatrizen Σ_A und Σ_M gibt, so dass die Ergebnisse zumindest zufriedenstellend für alle Punkte sind. Wie bereits mehrfach erwähnt, ist also die Wahl der optimalen Skala (und somit der optimalen Matrizen Σ_A und Σ_M) eine *lokale* Eigenschaft und darüber hinaus abhängig von der gewünschten Anwendung.

Im Folgenden soll davon ausgegangen werden, dass die Skala sowohl die Größe als auch die Orientierung des betrachteten Bereichs umfasst. Jede Skala entspricht also umkehrbar eindeutig einer Kovarianzmatrix der verwendeten Gaußglocken bzw. ihrer Ableitungen.

Folgende Forderungen werden vorgegeben:

- Bei Erfüllung aller in Kapitel 2 getroffenen Annahmen (siehe auch Überblick im Anhang) sollten die Ergebnisse möglichst gut sein, d. h. die Invarianzebenen sollten korrekt geschätzt werden, der OF mit der VR übereinstimmen.
- Eine Verletzung der Annahmen sollte sich nur in einer möglichst kleinen Umge-

bung der Punkte auswirken, an denen die Annahme verletzt wird.

- Aufgrund der geschätzten OF-Felder oder der GS-Tensoren sollte eine Segmentierung des Bildes in sich unabhängig voneinander bewegende Bereiche möglich sein.

Ziel ist es, für jeden Punkt die Kovarianzmatrizen Σ_A und Σ_M zu finden, die diese drei Forderungen möglichst gut erfüllen. Dafür gibt es mindestens zwei grundsätzlich verschiedene Ansätze:

1. Auswahl: Die erste Möglichkeit besteht darin, die Ableitungen, die lokalen Tensoren, den GST (kurz: die „Merkmale“) für eine größere Menge von Skalen zu bestimmen und dann aufgrund von Kenngrößen, die entweder direkt aus den zu bestimmenden Merkmalen abgeleitet oder gesondert berechnet werden, die optimale Skala durch Auswahl aus der untersuchten Menge von Skalen zu bestimmen (siehe z. B. [Lindeberg 98a; Lindeberg 98b]). Der große Nachteil dieser Vorgehensweise besteht darin, dass aufgrund von praktischen Beschränkungen (Rechenzeit, Speicher) nur eine kleine Teilmenge der möglichen Skalen untersucht werden kann¹. Durch Interpolation können zwar einige Berechnungen eingespart werden, durch geschickte Implementierung ([Crowley et al. 02]) die Berechnungen selbst beschleunigt werden, dennoch muss stets eine A-priori-Beschränkung der betrachteten Kovarianzmatrizen (z. B. auf Diagonalmatrizen) erfolgen.
2. Adaption: Die zweite Möglichkeit besteht darin, zunächst die Berechnung der Merkmale auf *einer* ausgewählten Skala durchzuführen und auf Basis des Ergebnisses die optimale Skala zu schätzen, also die Skala aufgrund des Ergebnisses zu *adaptieren*. Eine solche Adaption wird im Zweidimensionalen schon von [Nitzberg & Shiota 92] vorgeschlagen. [Nagel & Gehrke 98a] wenden diese Idee als erste auf Bildfolgen an. Der Vorteil dieses Vorgehens besteht darin, dass bei geeignet gewähltem Adaptionsverfahren die Menge der zulässigen Skalen weit weniger stark eingeschränkt wird als bei der Skalen-Auswahl.

Während die Skalenbestimmung durch Auswahl bereits von den oben genannten und weiteren Autoren untersucht wurde, gibt es bisher erst recht wenig Erfahrungen mit dem Adaptionsverfahren. In der vorliegenden Arbeit soll das existierende Adaptionsverfahren analysiert, erweitert und ausgiebig getestet werden.

¹Eine Kovarianzmatrix hat sechs freie Parameter: drei Winkel für die Hauptachsen-Orientierung, drei Werte für die Ausdehnung der Hauptachsen. Lässt man für jeden dieser Parameter nur 100 Werte zu, so ergeben sich schon 10^{12} Kombinationen. Könnte man auf einem heutigen Rechner (mit 2,5 GHz Prozessor) eine Kombination in nur 10 Taktzyklen berechnen und auswerten, also $2,5 \cdot 10^8$ Kombinationen pro Sekunden, so ergäbe sich eine Rechenzeit von $10^{12}/2,5 \cdot 10^8 = 4000$ Sekunden pro Pixel.

6.2 Exkurs: Referenz-Werte

Meßlatte für den Erfolg oder Mißerfolg der Vorgehensweisen zur GST- und OF-Bestimmung soll die Qualität der OF-Schätzung sein, die jeweils mit Hilfe von Mittelwert und Median des Winkelmaßes $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ nach [Barron et al. 94] und des euklidischen Abstands $\|\mathbf{v} - \mathbf{u}\|_2$ zwischen OF-Schätzung und VR beurteilt werden soll.

Dazu wird auf mehrere Bildfolgen zurückgegriffen, bei denen die VR zuverlässig bekannt ist:

- die Yosemite-Bildfolge (`yosemite`),
- die Translating-Tree-Bildfolge (`trans_tree`),
- die Diverging-Tree-Bildfolge (`div_tree`),
- die Marmor-Block-Bildfolge von Otte (`tr_bsb`) und
- die neue Marmor-Block-Bildfolge (`marmor_stat`).

Eine ausführliche Darstellung und Diskussion der Eigenschaften dieser Bildfolgen finden sich in Anhang A.

Tabellen 6.1 bis 6.5 vergleichen die Ergebnisse des nicht-adaptiven Verfahrens zur GST- und OF-Schätzung mit den publizierten Ergebnissen anderer Verfahren. Man erkennt, dass das nicht-adaptive Verfahren ähnlich gute oder sogar bessere Ergebnisse als die meisten anderen Verfahren produziert.

Zur Berechnung des OFs wurden jeweils abgetastete, nicht-korrigierte Masken der angegebenen Ausdehnung verwendet, die durch Abtastung der Gaußfunktion bzw. ihrer Ableitungen mit $\Sigma_A = \text{diag}(1,41; 1,41; 1,41)$, $\Sigma_M = 2\Sigma_A$ ermittelt wurden. Eine zeitliche Ausdehnung der Mittelungsmasken von 1 bedeutet, dass in diesen Fällen keine zeitliche Mittelung erfolgte.

Alle eigenen Werte wurden bei einem Bildrand von 20 Pixeln berechnet. Bei der Yosemite-Bildfolge wurden keine Punkte berücksichtigt, die zum „Himmel“ gehören. Bei der neuen Marmor-Block-Bildfolge wurden nur Punkte berücksichtigt, die zum bewegten Bereich des Bildes gehören. Bei den selbst berechneten Werten wird auch angegeben, welches Ergebnis sich bei Beschränkung auf Punkte ergibt, die der Kategorie ROF (vgl. Abschnitt 2.5, Seite 23, Schwellwerte: $\tau_{\text{OFD}} = 0,0120$, $\tau_{\text{DGD}} = 0,050$) zugeordnet wurden.

Die Werte anderer Autoren wurden ungeprüft aus den jeweils angegebenen Quellen entnommen. Oft ist nicht bekannt, welcher Bildausschnitt zur Ermittlung dieser Werte herangezogen wurden.

Verfahren	$\psi_{\mathbf{v}_+, \mathbf{u}_+}$		$\ \mathbf{v} - \mathbf{u}\ _2$		Dichte	Quelle (vgl. Legende)
	\emptyset	Med.	\emptyset	Med.		
Sun (2002)	9,21	-	-	-	> 100	[Sun 02]
Pedersen & Nielsen (2001)	8,51	-	-	-	(>) 100	[P&N 2001]
Alvarez et al. (2000)	5,53	-	-	-	> 100	[Alvarez et al. 00]
Mémin & Pérez (1998)	5,38	-	-	-	> 100	[M&P 98]
Black & Anandan (1996)	4,46	-	-	-	100	[BHS 98]
Black (1994)	3,52	-	-	-	100	[BHS 98]
Mémin & Pérez (1998)	2,34	-	-	-	100	[M&P 98]
Black & Jepson (1994)	2,29	-	-	-	100	[BHS 98]
Ju et al. (1996)	2,16	-	-	-	100	[BHS 98]
BHS WLS6 (1998)	2,02	-	-	-	100	[BHS 98]
Lai & Vemuri (1997)	1,99	-	-	-	100	[Lai & Vemuri 97]
BHS WTLS6 (1998)	1,97	-	-	-	100	[BHS 98]
Farnebäck (2001)	1,14	-	-	-	100	[Farnebäck 01]
nicht-adaptiv, alle (1)	2,62	1,93	0,1216	0,0761	100	
nicht-adaptiv, nur ROF (1)	2,20	1,82	0,0976	0,0635	80,4	
nicht-adaptiv, alle (2)	2,27	1,54	0,0940	0,0581	100	
nicht-adaptiv, nur ROF (2)	1,88	1,46	0,0741	0,0506	80,1	
nicht-adaptiv, alle (3)	2,35	1,59	0,0991	0,0613	100	
nicht-adaptiv, nur ROF (3)	1,95	1,51	0,0773	0,0528	79,0	
nicht-adaptiv, alle (4)	2,81	1,90	0,1311	0,0740	100	
nicht-adaptiv, nur ROF (4)	2,30	1,78	0,0968	0,0604	76,4	

Tabelle 6.1: Referenzwerte für Bild Nr. 7 (von 0–14) der Yosemite-Bildfolge. $\psi_{\mathbf{v}_+, \mathbf{u}_+}$ und $\|\mathbf{v} - \mathbf{u}\|_2$ geben Aufschluss über die Qualität der OF-Schätzung. Der Eintrag in der Spalte „Dichte“ gibt an, wieviel Prozent der Punkte im betrachteten Bereich in das angegebene Ergebnis eingingen. Eine Dichte > 100 bei aus fremden Quellen stammenden Daten bedeutet, dass die vorliegenden Zahlen die Wolkenregion berücksichtigen, was meist zu deutlich schlechteren Ergebnissen führt, wie die Ergebnisse von [Mémin & Pérez 98] zeigen, die für beide Fälle vorliegen. [Pedersen & Nielsen 01] geben nicht an, ob sie die Wolkenregion berücksichtigen oder nicht.

Abgekürzte Quellenangaben: BHS 98 = [Bab-Hadiashar & Suter 98], P&N 2001 = [Pedersen & Nielsen 01], M&P 98 = [Mémin & Pérez 98]

Wolkenregion nicht berücksichtigt. Standardparameter; Maskengröße: (1) $9 \times 9 \times 7$ (Abl.) bzw. $13 \times 13 \times 7$ (Mitt.), (2) $9 \times 9 \times 9$ (Abl.) bzw. $13 \times 13 \times 7$ (Mitt.), (3) $11 \times 11 \times 11$ (Abl.) bzw. $13 \times 13 \times 5$ (Mitt.), (4) $11 \times 11 \times 11$ (Abl.) bzw. $13 \times 13 \times 1$ (Mitt.).

Verfahren	ψ_{v_+, u_+}		$\ v - u\ _2$		Dichte	Quelle (vgl. Legende)
	\emptyset	Med.	\emptyset	Med.		
Horn & Schunk (orig.)	38,72	-	-	-	100	[Barron et al. 94]
Anandan	4,54	-	-	-	100	[Barron et al. 94]
Horn & Schunk (mod.)	2,02	-	-	-	100	[Barron et al. 94]
Singh (2. Schritt)	1,25	-	-	-	100	[Barron et al. 94]
Uras et al.	0,62	-	-	-	100	[Barron et al. 94]
Lucas & Kanade	0,56	-	-	-	13,1	[Barron et al. 94]
Lai & Vemuri (1997)	0,40	-	-	-	100	[Lai & Vemuri 97]
Fleet & Jepson	0,32	-	-	-	74,5	[Barron et al. 94]
Fleet & Jepson	0,23	-	-	-	49,7	[Barron et al. 94]
Pedersen & Nielsen	0,15	-	-	-	100	[P&N 2001]
nicht-adaptiv, alle (1)	0,43	0,38	0,0292	0,0270	100	
nicht-adaptiv, ROF (1)	0,39	0,35	0,0271	0,0256	82,0	
nicht-adaptiv, alle (2)	0,36	0,29	0,0217	0,0187	100	
nicht-adaptiv, ROF (2)	0,31	0,26	0,0197	0,0167	82,0	
nicht-adaptiv, alle (3)	0,42	0,32	0,0259	0,0203	100	
nicht-adaptiv, ROF (3)	0,36	0,28	0,0229	0,0183	68,6	

Tabelle 6.2: Referenzwerte für Bild Nr. 20 (von 1–40) der Translating-Tree-Bildfolge.

Abgekürzte Quellenangabe: P&N 2001 = [Pedersen & Nielsen 01]

Ausschnitt (20, 20) + [110, 110]. Standardparameter; Maskengröße: (1) $9 \times 9 \times 9$ (Abl.) bzw. $13 \times 13 \times 13$ (Mitt.), (2) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 15$ (Mitt.), (3) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 1$ (Mitt.).

6.3 Verfahrensbeschreibung

Die Grundidee des adaptiven Verfahrens zur GST-Bestimmung, wie es von [Nagel & Gehrke 98a] eingeführt wurde, besteht in einem iterativen Vorgehen, bei dem die Kovarianzmatrizen Σ_A und Σ_M der bei der Bestimmung der lokalen Tensoren bzw. der anschließenden Mittelungen verwendeten Gaußlocken auf Basis des GSTs der vorangehenden Iteration geschätzt werden. Durch dieses Vorgehen werden die Faltungsmasken an die lokale Grauwertstruktur, insbesondere an Orientierung und Betrag eventueller Grauwertübergänge angepasst. [Nagel & Gehrke 98a] geben im Startschritt als „Start-Kovarianz“ Diagonalmatrizen Σ_A und Σ_M vor – mit $\Sigma_M = 2\Sigma_A$ – und erzielen bereits durch eine einzige Iteration bemerkenswerte Verbesserungen der Ergebnisse.

Seit der erstmaligen Anwendung des adaptiven Verfahrens sind mehrere Untervarianten entstanden, die nachfolgend vorgestellt werden. Dabei bezeichne $(x, y, t)^T$ die räumlich-zeitliche Position eines Punkte in der betrachteten Bildfolge und $(x_0, y_0, t_0)^T$ den Punkt, für den der OF berechnet wird. Der Begriff „betrachtetes Bild“ beziehe sich stets auf das

Verfahren	$\psi_{\mathbf{v}_+, \mathbf{u}_+}$		$\ \mathbf{v} - \mathbf{u}\ _2$		Dichte	Quelle (vgl. Legende)
	\emptyset	Med.	\emptyset	Med.		
Horn & Schunk (orig.)	12,02	-	-	-	100	[Barron et al. 94]
Singh (2. Schritt)	8,60	-	-	-	100	[Barron et al. 94]
Anandan	7,64	-	-	-	100	[Barron et al. 94]
Uras et al.	4,64	-	-	-	100	[Barron et al. 94]
Horn & Schunk (mod.)	2,55	-	-	-	100	[Barron et al. 94]
Lucas & Kanade	1,65	-	-	-	24,3	[Barron et al. 94]
Lai & Vemuri (1997)	1,34	-	-	-	100	[Lai & Vemuri 97]
Fleet & Jepson	0,99	-	-	-	61,0	[Barron et al. 94]
Pedersen & Nielsen	0,80	-	-	-	100	[P&N 2001]
Fleet & Jepson	0,73	-	-	-	28,2	[Barron et al. 94]
nicht-adaptiv, alle (1)	2,08	1,66	0,0464	0,0386	100	
nicht-adaptiv, ROF (1)	1,88	1,52	0,0423	0,0358	79,1	
nicht-adaptiv, alle (2)	2,08	1,65	0,0465	0,0385	100	
nicht-adaptiv, ROF (2)	1,88	1,52	0,0424	0,0359	79,4	
nicht-adaptiv, alle (3)	2,58	2,09	0,0588	0,0485	100	
nicht-adaptiv, ROF (3)	2,39	1,95	0,0547	0,0455	75,4	

Tabelle 6.3: Referenzwerte für Bild Nr. 20 (von 1–40) der Diverging-Tree-Bildfolge.

Abgekürzte Quellenangabe: P&N 2001 = [Pedersen & Nielsen 01]

Ausschnitt (20, 20) + [110, 110]. Standardparameter; Maskengröße: (1) $9 \times 9 \times 9$ (Abl.) bzw. $13 \times 13 \times 13$ (Mitt.), (2) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 15$ (Mitt.), (3) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 1$ (Mitt.).

Bild zum Zeitpunkt t_0 . Ferner seien GST_i der Grauwertstruktur-Tensor und $\Sigma_{A,i}$ und $\Sigma_{M,i}$ die Kovarianzmatrizen der i -ten Iteration, wobei der Startschritt als 0-te Iteration bezeichnet wird. Schließlich bezeichne f_A die Funktion, die eine Kovarianzmatrix aus einem GST berechnet, – im Folgenden *Adaptionsfunktion* f_A genannt – und auf die anschließend näher eingegangen wird.

- In der von [Nagel & Gehrke 98a] vorgeschlagenen Version des Adaptionsverfahrens hängen sowohl Σ_A bei der Ableitungsschätzung als auch Σ_M bei der Mittelung nur vom GST an dem Punkt ab, für den der GST berechnet werden soll, also vom Aufpunkt. Wie beim nicht-adaptiven GST gilt dabei $\Sigma_{M,i} = 2\Sigma_{A,i}$, also:

$$\Sigma_{A,i}(x, y, t) = f_A(\text{GST}_{i-1}(x_0, y_0, t_0)), \quad (6.1)$$

$$\Sigma_{M,i}(x_0, y_0, t_0) = 2 \cdot f_A(\text{GST}_{i-1}(x_0, y_0, t_0)). \quad (6.2)$$

Zur Berechnung des GST_i an einem Punkt inklusive Ableitungsbestimmung ist also nur die Kenntnis des GST_{i-1} am selben Punkt nötig. Die GST-Berechnungen

Verfahren	$\psi_{\mathbf{v}_+, \mathbf{u}_+}$		$\ \mathbf{v} - \mathbf{u}\ _2$		Dichte	Quelle (vgl. Legende)
	\emptyset	Med.	\emptyset	Med.		
Otte SB _{2,1}	-	-	0,139	-	60,8	[Otte 94]
Otte NA	-	-	0,107	-	59,1	[Otte 94]
Otte NA	-	-	0,087	-	25,6	[Otte 94]
BHS WTLS6	3,67	-	-	-	100	[BHS 98]
BHS WLS2	3,39	-	-	-	100	[BHS 98]
Fleet & Jepson	2,05	-	-	-	55,8	[BHS 98]
BHS WTLS6	1,55	-	-	-	51,6	[BHS 98]
BHS WLS6	1,44	-	-	-	55,9	[BHS 98]
nicht-adaptiv, alle (1)	4,05	1,26	0,2304	0,0530	100	
nicht-adaptiv, ROF (1)	1,51	1,01	0,0714	0,0394	72,0	
nicht-adaptiv, alle (2)	4,00	1,22	0,2237	0,0506	100	
nicht-adaptiv, ROF (2)	1,48	0,98	0,0690	0,0378	72,0	
nicht-adaptiv, alle (3)	4,52	2,07	0,2045	0,0714	100	
nicht-adaptiv, ROF (3)	2,40	1,85	0,0969	0,0585	73,9	

Tabelle 6.4: Referenzwerte für Bild Nr. 35 (von 20–50) der alten Marmorblock-Bildfolge `tr_bsb`.

Abgekürzte Quellenangabe: BHS 98 = [Bab-Hadiashar & Suter 98]

Ausschnitt (20, 20) + [472, 472]. Standardparameter; Maskengröße: (1) $9 \times 9 \times 9$ (Abl.) bzw. $13 \times 13 \times 13$ (Mitt.), (2) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 15$ (Mitt.), (3) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 1$ (Mitt.).

benachbarter Punkte beeinflussen sich nicht gegenseitig. Dieses Vorgehen soll im Folgenden als *einfach-adaptives Verfahren* bezeichnet werden.

Nachteil dieses Vorfahrens ist, dass an ein- und demselben Punkt die Ableitungsberechnung möglicherweise hundertfach ausgeführt wird, jedesmal – da sich der Aufpunkt ändert – mit anderen Kovarianzmatrizen, was zu einer schlechten Performanz dieser Methode führt.

- Eine Variation des Verfahrens wenden [Nagel & Gehrke 98a] an, wenn der GST nicht für einen Zeitpunkt, sondern für zeitlich aufeinanderfolgende Bilder einer Bildfolge berechnet wird. In diesem Fall wird zur Bestimmung der ersten Iteration der adaptiven GST-Bestimmung nicht der Startschritt mit festen Kovarianzmatrizen, sondern der ebenfalls adaptiv bestimmte GST des vorangehenden Bildes herangezogen:

$$\Sigma_{A,1}(x, y, t) = f_A(\text{GST}_1(x_0, y_0, t_0 - 1)), \quad (6.3)$$

$$\Sigma_{M,1}(x_0, y_0, t_0) = 2 \cdot f_A(\text{GST}_1(x_0, y_0, t_0 - 1)) . \quad (6.4)$$

Verfahren	$\psi_{\mathbf{v}_+, \mathbf{u}_+}$		$\ \mathbf{v} - \mathbf{u}\ _2$		Dichte
	\emptyset	Med.	\emptyset	Med.	
nicht-adaptiv, alle (1)	3,5073	0,3561	0,1698	0,0191	100
nicht-adaptiv, ROF (1)	0,6129	0,2810	0,0272	0,0151	77,1
nicht-adaptiv, alle (2)	3,5003	0,3086	0,1706	0,0162	100
nicht-adaptiv, ROF (2)	0,5874	0,2333	0,0254	0,0122	76,9
nicht-adaptiv, alle (3)	3,5030	0,5261	0,1913	0,0288	100
nicht-adaptiv, ROF (3)	0,8409	0,4332	0,0389	0,0234	79,1

Tabelle 6.5: Referenzwerte für Bild Nr. 150 (von 100–300) der neuen Marmorblock-Bildfolge `marmor_stat`. Berücksichtigt wurden alle 19143 Punkte, die zu den bewegten Blöcken gehören. Standardparameter; Maskengröße: (1) $9 \times 9 \times 9$ (Abl.) bzw. $13 \times 13 \times 13$ (Mitt.), (2) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 15$ (Mitt.), (3) $11 \times 11 \times 11$ (Abl.) bzw. $15 \times 15 \times 1$ (Mitt.).

Dieses *fortschreibend einfach-adaptive Verfahren* hat allerdings den Nachteil, dass die Punkte $(x_0, y_0, t_0)^T$ und $(x_0, y_0, t_0 - 1)$ bei nicht-stationären Szenen – also im eigentlich interessanten Fall – möglicherweise nicht demselben Szenenpunkt entsprechen und somit der GST an der Stelle $(x_0, y_0, t_0 - 1)$ völlig ungeeignet sein kann, um die Kovarianzmatrix an der Stelle $(x_0, y_0, t_0)^T$ zu schätzen. Unter Berufung auf Annahme 2 kann zwar argumentiert werden, dass die Verschiebung sowieso nicht zu groß sein darf, aber gerade in der Nähe von Bewegungsdiskontinuitäten ist mit Schwierigkeiten zu rechnen. Die vorliegende Arbeit konzentriert sich auf die Frage, wie der OF möglichst gut „aus dem Nichts“ geschätzt werden kann, wenn also noch kein GST aus zeitlich vorangehenden Bildern vorliegt. Diese Variante wird daher nicht weiter betrachtet.

- Unter dem Aspekt der Performanzverbesserung² stellt [Middendorf 99] das *doppelt-adaptive Verfahren* zur GST-Berechnung vor. Dabei werden an jedem Punkt, der von irgendeiner Mittelungsoperation erfasst wird, die Ableitungen nur noch ein einziges Mal berechnet, wobei Σ_A stets aus dem GST des Punktes des betrachteten Bildes an derselben räumlichen Position berechnet wird.

Es gilt also für alle x_0, y_0, t_0 und für alle $x, y, t \in \mathcal{U}(x_0, y_0, t_0)$:

$$\Sigma_{A,i}(x, y, t) = f_A(\text{GST}_{i-1}(x, y, t_0)), \quad (6.5)$$

$$\Sigma_{M,i}(x_0, y_0, t_0) = 2 \cdot f_A(\text{GST}_{i-1}(x_0, y_0, t_0)). \quad (6.6)$$

Ein solches Vorgehen ist aber nur zu rechtfertigen, wenn die Verschiebung von Bild zu Bild relativ klein ist, da sonst ähnliche Probleme zu befürchten sind wie beim fortschreibend einfach-adaptiven Verfahren.

²Die Berechnung nur einer Iteration des einfach-adaptiven Verfahren auf einem 3000 Pixel umfassenden Bildausschnitt benötigte auf den damals verfügbaren Rechnern etwa drei Minuten, was ein erhebliches Hindernis für die Weiterentwicklung der Verfolgungsalgorithmen darstellte.

Die Bezeichnung *doppelt-adaptives Verfahren* wurde gewählt, weil Mittelungs- und Ableitungsmasken getrennt voneinander adaptiert werden. Die Ergebnisse benachbarter Punkte sind voneinander abhängig, da der GST_{*i*-1} eines Punktes über die Kovarianzen der Ableitungsmasken den GST_{*i*} all jener Punkte beeinflusst, in deren Mittelungsbereich er liegt.

- In einer weiteren Version des Algorithmus – der im Rahmen der vorliegenden Arbeit entworfen und umgesetzt wurde – erfolgt die Ableitungsberechnung für jeden Punkt $(x, y, t)^T$ in Abhängigkeit vom vorangehend berechneten GST am Punkt $(x, y, t)^T$ selbst, unabhängig von räumlicher oder zeitlicher Position des Aufpunktes. Es gilt also:

$$\Sigma_{A,i}(x, y, t) = f_A(\text{GST}_{i-1}(x, y, t)), \quad (6.7)$$

$$\Sigma_{M,i}(x_0, y_0, t_0) = 2 \cdot f_A(\text{GST}_{i-1}(x_0, y_0, t_0)) . \quad (6.8)$$

Dieses Verfahren wird im Folgenden als *mehrfach-adaptiv* bezeichnet, da für jede adaptive GST-Berechnung eine häufigere Adaption stattfindet als bei allen anderen Vorgehensweisen.

Der Nachteil dieses Verfahrens besteht darin, dass zur Berechnung des GSTs für ein Bild der GST der vorangehenden Iteration für alle Bilder vorliegen muss, die in den zeitlichen Mittelungsbereich des betrachteten Bildes fallen. Bei einer zeitlichen Mittelungs-Maskengröße von $2N + 1$ muss also zur ersten Iteration der GST-Berechnung eines Bildes der nicht-adaptive GST für $2N + 1$ Bilder vorliegen. In der zweiten Iteration müsste dann der GST der ersten Iteration für $2N + 1$ Bilder und somit der nicht-adaptive GST für $4N + 1$ Bilder vorliegen. Bei diesem Ansatz ist also ein iteratives Vorgehen (über eine Iteration hinaus) praktisch ausgeschlossen.

[Middendorf 02] vergleicht einfach-adaptives, doppelt-adaptives und mehrfach-adaptives Verfahren hinsichtlich mehrerer Kriterien und zeigt, dass signifikante Unterschiede im Wesentlichen nur an Stellen auftreten, an denen aufgrund nahezu homogener Grauwerte sowieso keine zuverlässige Schätzung des OF mit Hilfe des GST zu erwarten war. Um weitere Versuche in vertretbarer Zeit durchführen zu können, wird daher bis auf weiteres das doppelt-adaptive Verfahren verwendet.

Entscheidend für die Adaption ist die Adaptionsfunktion f_A . Die Grundidee besteht darin, dass Ableitungs- und Mittelungsmasken groß ausgedehnt sind in Richtungen, in denen sich die Grauwertstruktur wenig ändert, und nur geringe Ausdehnung in Richtungen besitzen, in denen sich die Grauwertstruktur stark ändert. Wie stark sich die Grauwertstruktur in einer bestimmten Richtung ändert, lässt sich an den Eigenwerten des GST ablesen: je größer die Eigenwerte, desto größer die Änderung in Richtung des zugehörigen Eigenvektors. Die Funktion f_A berechnet daher im Wesentlichen das Inverse des GST. Da die Eigenwerte des GST erfahrungsgemäß zwischen 0,01 und 1000

schwanken, würde aber die „reine“ Invertierung des GST manchmal zu extremen Ausdehnungen der Faltungsmasken führen. Um dies zu verhindern, wird der Wertebereich von f_A eingeschränkt. Dazu wird zunächst eine symmetrische Schur-Zerlegung des GST bestimmt:

$$\text{GST} = U \cdot \text{diag}(\lambda_1, \lambda_2, \lambda_3) \cdot U^T \quad \text{mit} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0 \quad , \quad (6.9)$$

was einer Drehung des Koordinatensystems entspricht, so dass die Achsen des transformierten Koordinatensystems den Eigenvektoren des GST entsprechen. λ_i ($i = 1, 2, 3$) bezeichnen hierbei die Eigenwerte des GST.

Aus den Eigenwerten λ_i sollen nun *Ausdehnungsfaktoren* α_i bestimmt werden, aus denen sich durch Rückrotation des Koordinatensystems die Kovarianzmatrix Σ_A ergibt, also:

$$\Sigma_A = f_A(\text{GST}) = U^{-T} \cdot \text{diag}(\alpha_1, \alpha_2, \alpha_3) \cdot U^{-1} . \quad (6.10)$$

Für die Bestimmung von α_i aus λ_i gibt es mehrere Möglichkeiten:

- Mit $\alpha_i = (\lambda_i)^{-1}$ wäre $f_A(\text{GST}) = \text{GST}^{-1}$, was – wie bereits dargestellt – wegen der daraus resultierenden extremen Ausdehnungen der Gaußglocken nicht in Frage kommt.
- [Nagel & Gehrke 98a] schlagen zunächst eine Normierung der Eigenwerte vor, um Invarianz gegenüber einer globalen Skalierung der Grauwerte zu erreichen:

$$\tilde{\lambda}_i = \frac{\lambda_i}{\text{Spur}(\text{GST})} \in [0; 1], \quad i = 1, 2, 3 \quad , \quad (6.11)$$

und bestimmen die Ausdehnungsfaktoren aus diesen *normierten Eigenwerten* durch Linearkombination von minimal und maximal zugelassener Varianz σ_{\min}^2 bzw. σ_{\max}^2 der Gaußglocke:

$$\alpha_i = \tilde{\lambda}_i \cdot \sigma_{\min}^2 + (1 - \tilde{\lambda}_i) \cdot \sigma_{\max}^2, \quad i = 1, 2, 3 . \quad (6.12)$$

Die Standardabweichung der adaptierten Gaußglocke liegt somit immer im „Skalierungsbereich“ $[\sigma_{\min}, \sigma_{\max}]$, wobei ein normierter Eigenwert nahe bei 1 zu einem Ausdehnungsfaktor von $\approx \sigma_{\min}^2$ und ein Eigenwert nahe 0 zu einem Ausdehnungsfaktor von $\approx \sigma_{\max}^2$ führt. Aufgrund dieses linearen Zusammenhangs soll dieses Verfahren als *lineare Adaption* bezeichnet werden.

Die lineare Adaption führt fast immer zu $\alpha_1 \approx \sigma_{\min}^2$, $\alpha_2 \approx \sigma_{\max}^2$ und $\alpha_3 \approx \sigma_{\max}^2$, da die normierten Eigenwerte linear in die Ausdehnungsfaktoren eingehen und auch der mittlere (normierte) Eigenwert des GST meist klein ist im Vergleich zum größten Eigenwert (Abbildung 6.1).

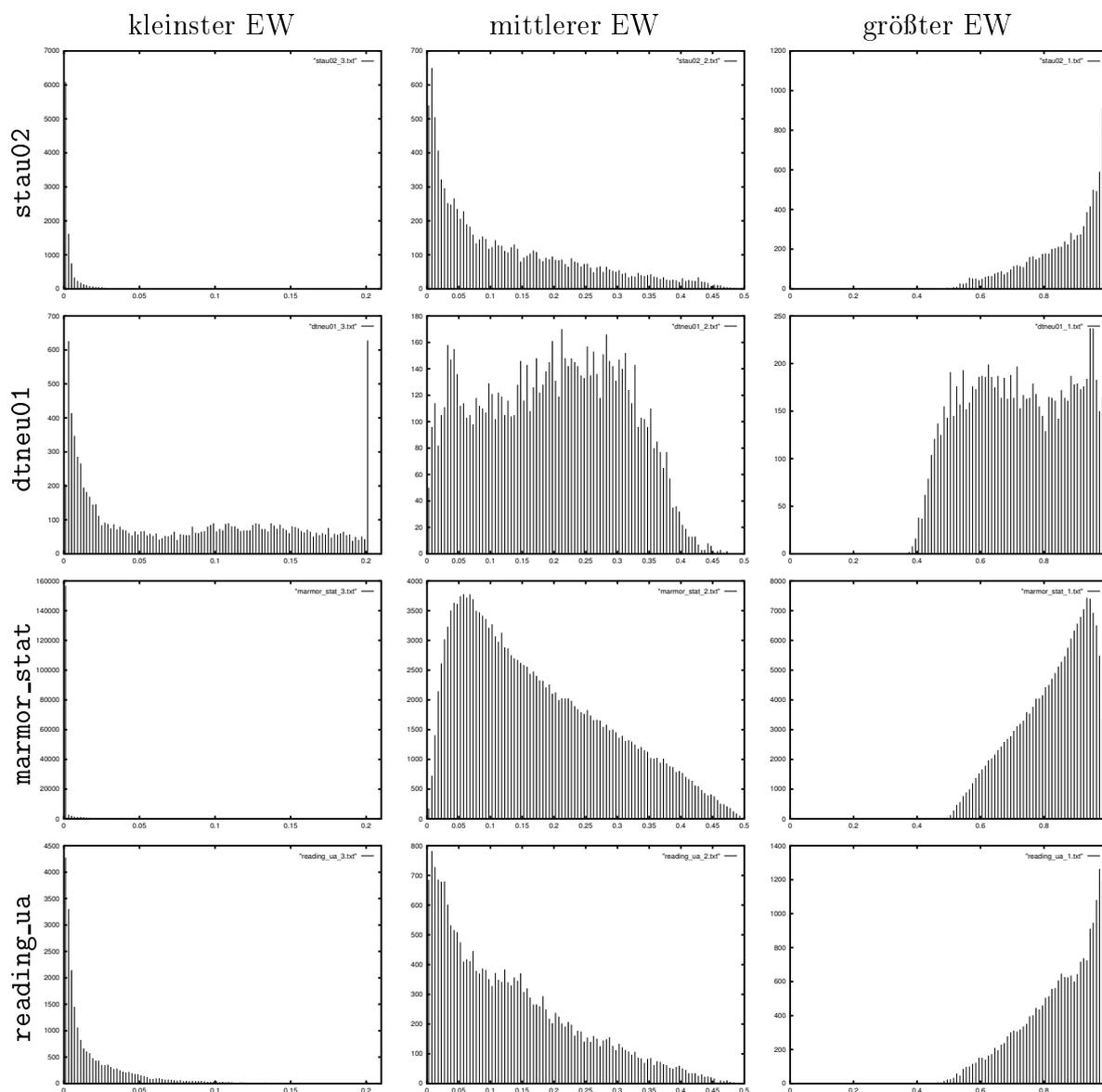


Abbildung 6.1: Verteilung der normierten Eigenwerte in ausgewählten Bildern der Bildfolgen `stau02`, `dtneu01`, `marmor_stat` und `reading_ua`. Die größte Streuung der Eigenwerte tritt bei der Bildfolge `dtneu01` auf, die kleinste bei der Bildfolge `stau02`. Bei allen Histogrammen stellen die beiden äußeren Striche alle Werte dar, die nicht im dargestellten Wertebereich liegen. Dies ist jedoch nur bei der Darstellung des kleinsten Eigenwerts bei `dtneu01` von Bedeutung.

- Um den zwar kleinen, aber keineswegs zufälligen Unterschieden zwischen den beiden kleinsten Eigenwerten besser Rechnung zu tragen, wurde ein neues Verfahren zur Bestimmung der Ausdehnungsfaktoren eingeführt. Nach der Skalierung der

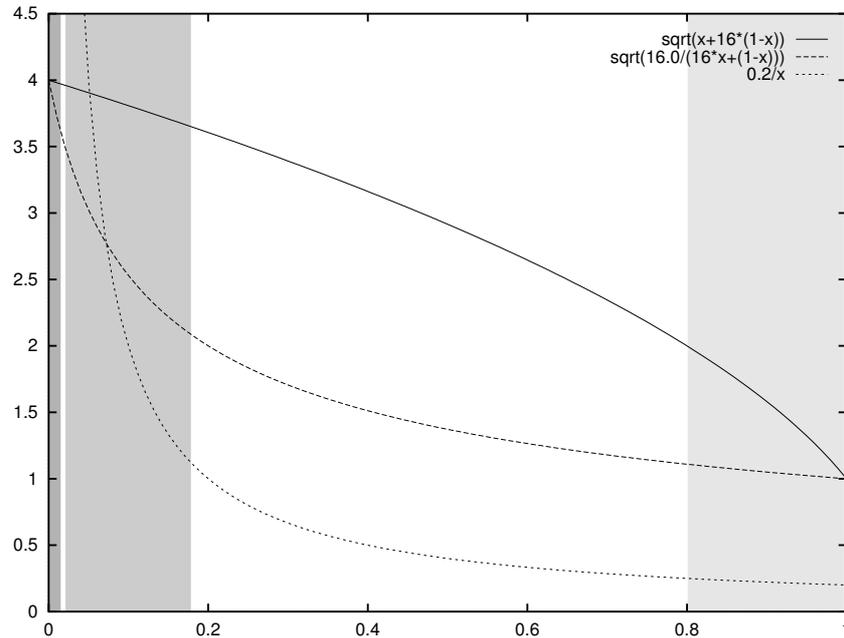


Abbildung 6.2: Vergleich der linearen und der quasi-inversen Adaption: Die Grafik zeigt die Standardabweichung in Richtung des Eigenvektors e_i in Abhängigkeit vom normierten Eigenwert $\tilde{\lambda}_i$ für $\sigma_{\min} = 1,0$ und $\sigma_{\max} = 4,0$ bei Verwendung der linearen Adaption (—) und der quasi-inversen Adaption (- -). Zum Vergleich (- - -) das Ergebnis, wenn der GST nur invertiert würde ohne Einschränkung durch σ_{\min} und σ_{\max} und ohne Skalierung der Eigenwerte (Annahme: $\text{spur}(\text{GST})=5$). Grau unterlegt sind die Bereiche, in denen typischerweise die normierten Eigenwerte liegen (schematisch).

Eigenwerte gemäß Formel (6.11) berechnet man α_i wie folgt:

$$\alpha_i = \frac{\sigma_{\max}^2 \cdot \sigma_{\min}^2}{(1 - \tilde{\lambda}_i) \cdot \sigma_{\min}^2 + \tilde{\lambda}_i \cdot \sigma_{\max}^2}, \quad i = 1, 2, 3. \quad (6.13)$$

Für $\tilde{\lambda}_i = 0$ und $\tilde{\lambda}_i = 1$ liefert Formel 6.13 dieselben Ergebnisse wie Formel 6.12, für $\tilde{\lambda}_i \in (0; 1)$ entspricht Formel 6.13 aber eher einer Invertierung der Diagonalmatrix $\text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3)$. Deshalb soll dieses Verfahren als *quasi-inverse Adaption* bezeichnet werden.

Abbildung 6.2 verdeutlicht grafisch die Unterschiede zwischen linearer und quasi-inverser Adaption. Man erkennt, dass die quasi-inverse Adaption zu einer deutlichen Differenzierung im Bereich des mittleren Eigenwertes führt. [Middendorf 02] und [Middendorf & Nagel 01] zeigen darüber hinaus, dass die quasi-inverse Adaption auch in der Anwendung bei realen Bildfolgen tatsächlich Vorteile gegenüber der linearen Adaption bringt. Soweit nicht ausdrücklich anders erwähnt, wird daher die quasi-inverse Adaption als Referenzverfahren verwendet.

6.4 Experimentelle Evaluation der iterativen GST-Bestimmung

6.4.1 Motivation

Die vorangehend beschriebene Adaption ist grundsätzlich beliebig oft wiederholbar. Aus einem adaptiv-bestimmten GST können neue Masken und damit ein neuer adaptiver GST bestimmt werden. Bisher ([Nagel & Gehrke 98a], [Middendorf 99]) wurde die Anzahl der Iterationen v. a. aus Rechenzeitgründen auf eins begrenzt.

Bevor wir uns in Abschnitt 6.5 einer theoretischen Betrachtung zuwenden, sollen zunächst einige Fragen experimentell betrachtet werden, um ein besseres Verständnis der Theorie zu ermöglichen. Im Folgenden sollen vor allem vier Fragen untersucht werden:

- **Qualität der GST-Schätzung:** Ergibt die adaptive GST-Schätzung mit einer oder mehr Iterationen bessere GST als das nicht-adaptive Verfahren, und zwar sowohl in Hinblick auf Länge und Orientierung der geschätzten OF-Vektoren als auch auf die Kategorisierung?
- **Konvergenz der OF-Schätzung:** Unter welchen Bedingungen konvergiert das adaptive Verfahren? Liefert irgendwann eine weitere Iteration dasselbe Ergebnis wie die vorangehende?
- **Robustheit gegenüber Parameterwahl:** In wie weit hängen das Ergebnis der Iteration (OF-Qualität, Kategorisierung) und das Konvergenzverhalten von den gewählten Parametern ab? Untersucht werden sollen der Einfluss der Start-Kovarianzen σ_x , σ_y und σ_t in der initialen, nicht-adaptiven Iteration, der Größe des Skalierungsbereiches und des Adaptionsverfahrens (linear vs. quasi-invers, vgl. vorangehender Abschnitt).
- **Konvergenzbegriff und Konvergenzkriterium:** Sind der gewählte Konvergenzbegriff und das Konvergenzkriterium in Hinblick auf die durchgeführten Experimente sinnvoll?

Zur Beantwortung dieser Fragen wurden Teilbereiche repräsentativer Bilder der Bildfolgen `marmor_stat`, `yosemite` und `kwbB` betrachtet (vgl. Abbildung 6.3). Eine Diskussion der Eigenschaften dieser Bildfolgen findet sich in Anhang A. Die `yosemite`- und die `kwbB`-Bildfolge werden nur betrachtet, wenn sich dadurch zusätzliche Aufschlüsse über die jeweils untersuchten Aspekte erwarten lassen.

Für jeden Bildausschnitt wurden jeweils 100 Iterationen des doppelt-adaptiven Verfahrens unter Verwendung der in Tabelle 6.6 aufgeführten Parameter berechnet, mit jeweils zwei unterschiedlichen Start-Kovarianzen, zwei Skalierungsbereichen und den beiden Adaptionsverfahren, insgesamt also acht Versuche pro Bildfolge.

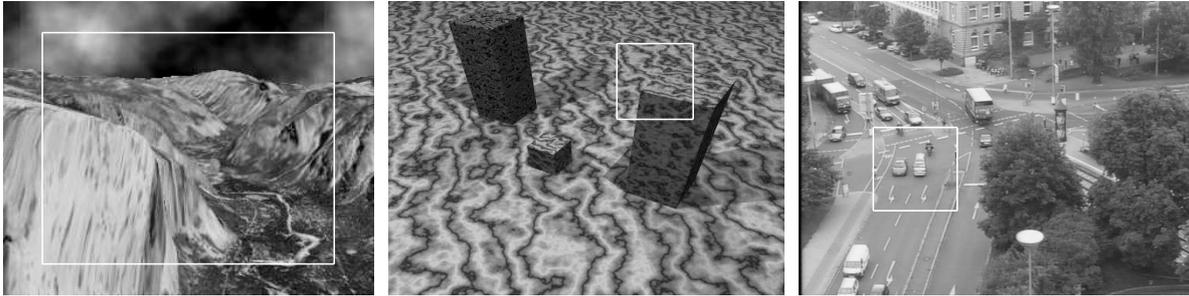


Abbildung 6.3: Ausschnitte, für die mehrere Iterationen der GST-Bestimmung durchgeführt wurden: `yosemite`, Bild 7 (links), `marmor_stat`, Bild 150 (Mitte), `kwbB`, Bild 2540 (rechts).

Parameter	Versuchsnummer			
	40	41	42	43
σ_x (nicht-it.)	1,00 oder 1,41 (vgl. Text)			
σ_t (nicht-it.)	1,00			
Max. Maskengröße (nicht-it.)	$19 \times 19 \times 19$			
Adaptionsverfahren	doppelt-adaptiv			
Invertierung	linear	quasi-inverse	linear	quasi-inverse
Schwelle auf Spur (τ_N)	1,00			
Diskontinuitäts-Schwelle (τ_{OFD})	0,0120			
σ_{\min}	1,00			
σ_{\max}	2,85		4,00	
Korrektur der Ableitungsmasken	nein			
DGD-Schwelle (τ_{DGD})	0,050			
Max. Maskengröße	$19 \times 19 \times 19$		$25 \times 25 \times 25$	
$\Sigma_M = 2 \cdot \Sigma_A$	ja			

Tabelle 6.6: Parameter für Versuch 40–43

Zur vereinfachten Darstellung wird in diesem Abschnitt die initiale, nicht-adaptive GST-Schätzung als 0-te Iteration bezeichnet.

6.4.2 Qualität der GST-Schätzung

6.4.2.1 Qualität der Kategorisierung

Zunächst soll die Kategorisierung im betrachteten Ausschnitt ohne Adaption und nach 100 Iterationen verglichen und beurteilt werden. Eine Kategorisierung wird als gut betrachtet, wenn folgende Kriterien erfüllt sind:

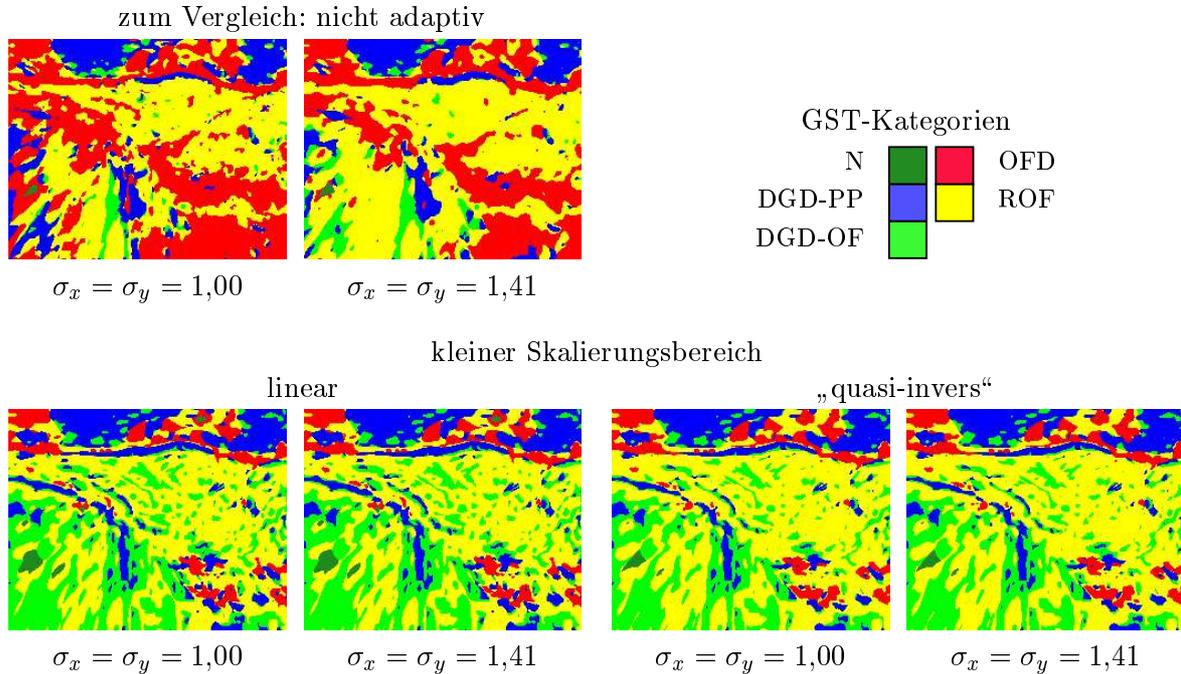


Abbildung 6.4: [yosemite, Bild 7] Vergleich der Kategorisierung der GSTs nach 100 Iterationen (unten) mit der Kategorisierung ohne Adaption (oben). Durch die Adaption geht die Anzahl der fälschlicherweise erkannten Diskontinuitäten deutlich zurück.

- Im Bereich des Abbilds *eines* Objekts sollten möglichst nur ROF-Vektoren liegen.
- Im Bereich zwischen *zwei* sich unterschiedlich bewegendem Objektabbildern sollte ein möglichst schmaler, aber zusammenhängender Bereich von OFD- und DGD-PP-Vektoren liegen.
- DGD-Bereiche im Innern eines Objektbildes sollten möglichst klein sein.

Bei der *yosemite*-Bildfolge (Abbildung 6.4) geht die Anzahl der fälschlicherweise erkannten Diskontinuitäten durch die Adaption deutlich zurück. Im Bereich der Bewegungsdiskontinuität zwischen dem „Felsen“ vorne links und dem Hintergrund wird ein schmaler Bereich der Kategorie DGD-PP erkannt, was darauf zurückzuführen ist, dass die Bewegungsdiskontinuität mit einem signifikanten Grauwertunterschied zusammenfällt. Nur im unteren Bereich dieser Bewegungsdiskontinuität werden Punkte als DGD-OF (statt wie gewünscht als DGD-PP) klassifiziert, da dort die Bewegungen beiderseits der Kante im Wesentlichen in dieselbe Richtung verlaufen. Keine der getesteten Adaptionvarianten zeigt signifikante Vorteile gegenüber einer der anderen Methoden.

Bei Bildfolge *marmor_stat* (Abbildung 6.5) wird das Ergebnis durch die Adaption deutlich schlechter: die Breite der Diskontinuitätsbereiche nimmt zu, es treten mehr OFD innerhalb der Objektabbilder auf. Beim Vergleich der Adaptionansätze untereinander fällt die „quasi-inverse“ Adaption durch schmalere Diskontinuitätsbereiche als bei

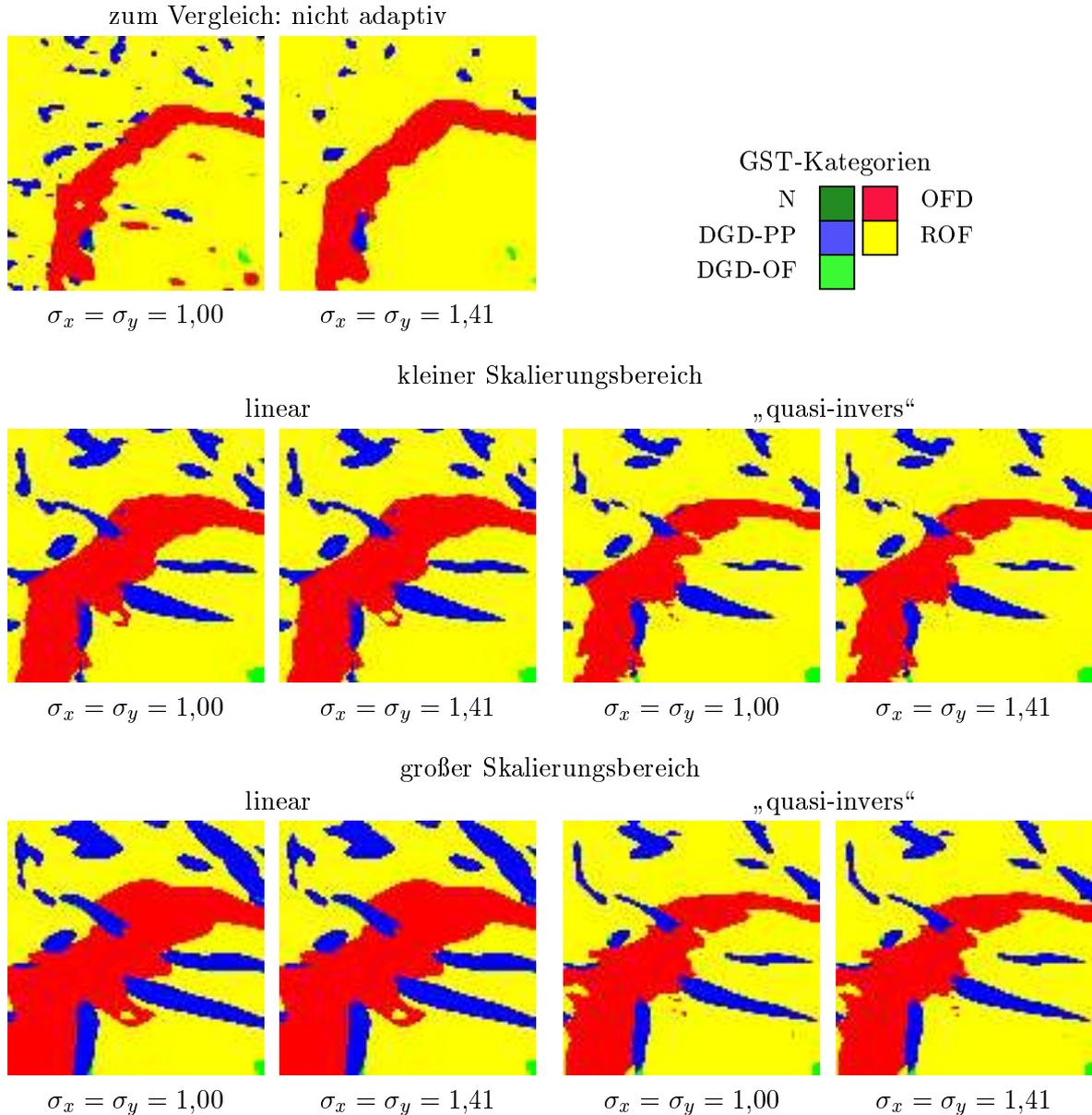


Abbildung 6.5: [marmor_stat, Bild 150] Vergleich der Kategorisierung der GSTs nach 100 Iterationen (Mitte, unten) mit der Kategorisierung ohne Adaption (oben). Man erkennt, dass durch die Adaption die Diskontinuitätsbereiche deutlich breiter werden und im Übergangsbereich zwischen der dunklen Vorderseite des bewegten Marmorblocks und dessen hellerer Oberseite sowie dem helleren Hintergrund mehr Punkte als DGD kategorisiert werden. Zum Vergleich der verschiedenen Adaptionsverfahren siehe Text.

linearer Adaption auf. Die unterschiedlichen Start-Kovarianzen wirken sich nur bei einzelnen Pixeln aus. Die Größe des Skalierungsbereiches beeinflusst zwar das Ergebnis, jedoch ist keine der beiden Varianten der anderen überlegen.

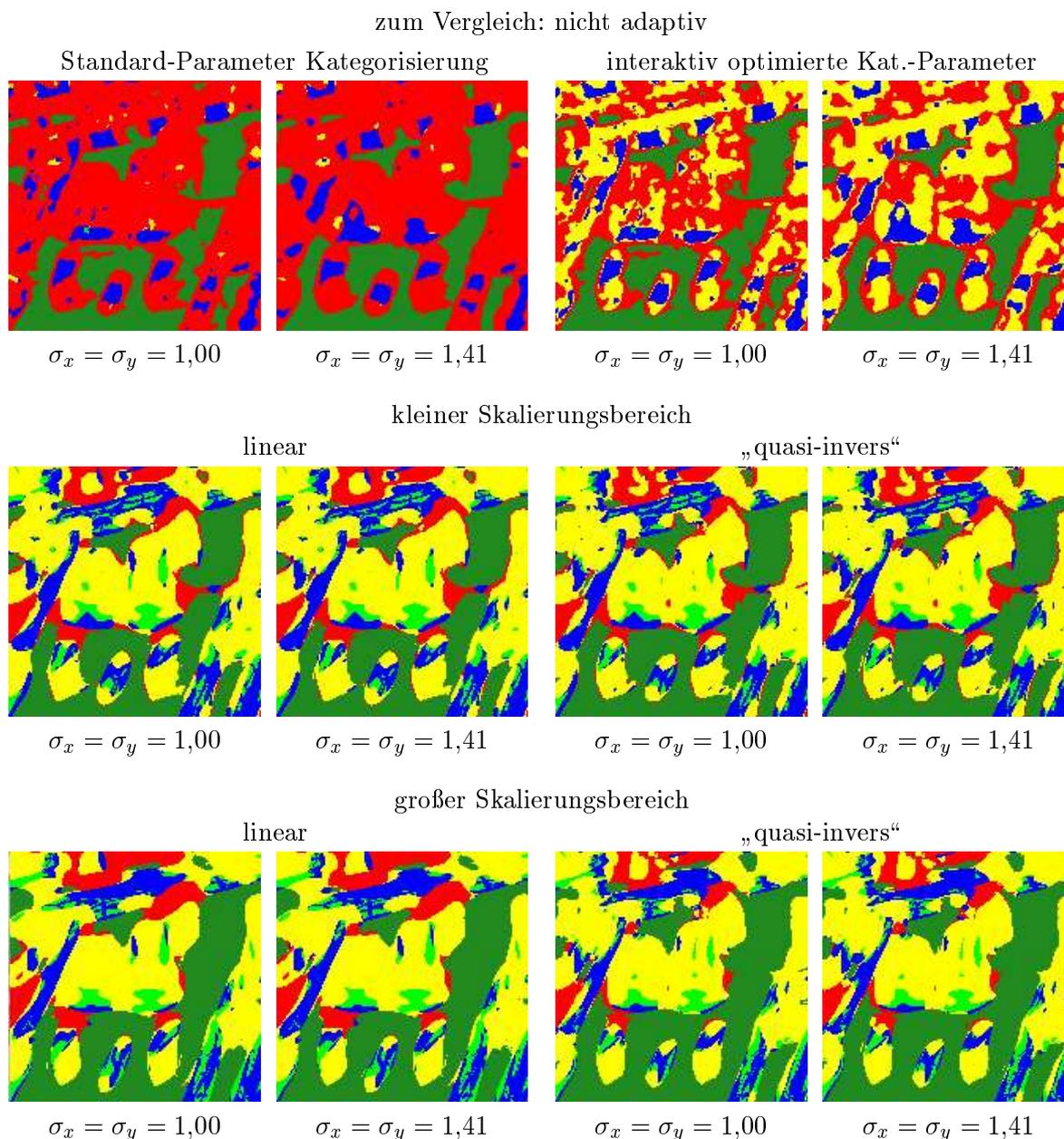


Abbildung 6.6: [kwbB, Halbbild 2540] Vergleich der Kategorisierung der GSTs nach 100 Iterationen (Mitte, unten) mit der Kategorisierung ohne Adaption (oben). Ohne Adaption ist das Kategorisierungsergebnis sowohl mit den Standardparametern (oben links) als auch mit interaktiv durch deutliche Erhöhung der OFD-Schwelle (0,0120 \rightarrow 0,0500) optimierten Parametern mangelhaft.

Bei der Bildfolge kwbB hingegen (Abbildung 6.6) führt die Adaption zu einer Ergebnisverbesserung: Die Objektabbilder zeichnen sich als zusammenhängende Bereiche von ROF-Vektoren besser ab als ohne Adaption. Allerdings geht die Diskontinuität zwi-

schen den Abbildern der beiden im Ausschnitt sichtbaren Autos verloren, was auf die nahezu identische Bewegungsrichtung und -geschwindigkeit zurückzuführen ist. Eine kategorienbasierte Trennung zwischen dem (in Blickrichtung) rechtem Auto und dem in Blickrichtung dahinter rechts befindlichen Motorrad ist weder mit noch ohne Adaption gegeben.

Bei allen drei Bildfolgen ist augenscheinlich kein Unterschied zwischen den Ergebnissen bei den verschiedenen Start-Kovarianzen feststellbar.

Zusammenfassend kann man feststellen, dass die Adaption die Kategorisierungsergebnisse verbessern, aber auch verschlechtern kann. Besonders deutlich treten die Vorteile bei einer natürlichen, rauschbehafteten Bildfolge auf, während die Nachteile bei einer künstlichen, besonders kontraststarken Bildfolge hervortreten. Durch geschickte Wahl des Adaptionsverfahrens und der Parameter kann der Hauptnachteil (breite Diskontinuitäten) aber reduziert werden.

6.4.2.2 Qualität der OF-Schätzung

Die Qualität der OF-Schätzung wird nur bei der Bildfolge `marmor_stat` beurteilt, da bei der Bildfolge `kwbB` gar keine und bei der Bildfolge `yosemite` nur durch Quantisierung verfälschte Referenzwerte vorliegen.

Tabellen 6.7 und 6.8 stellen die Abweichung der OF-Schätzung von der tatsächlichen VR dar. Zunächst fällt auf, dass der durchschnittliche Fehler bei der OF-Schätzung bei Betrachtung aller Vektoren durch die Adaption und mehrfache Iteration nicht ab-, sondern zunimmt. Dies ist darauf zurückzuführen, dass durch die größeren Masken bei der Adaption mehr Punkte in den Einflussbereich der Bewegungsdiskontinuität geraten. Erst wenn man die Betrachtung auf diejenigen ROF-Vektoren beschränkt, die im Bereich des Abbilds des bewegten Klotzes liegen³, so ist der durchschnittliche Fehler in der ersten Iteration stets geringer als ohne Adaption. Dieser Vergleich ist aber dadurch verfälscht, dass der Durchschnitt auf unterschiedlichen Punkt-Mengen berechnet wurden. Aus diesem Grund wurde zum besseren Vergleich als Referenzwert der durchschnittliche Fehler (Referenzfehler) in der 0. Iteration für *dieselben* Punkte berechnet, die bei der Durchschnittsberechnung in den einzelnen Iterationen berücksichtigt wurden (vgl. Spalte „Ref. \emptyset Fehler“). Man erkennt, dass bei quasi-inverser Adaption der Durchschnittsfehler der ROF-Vektoren im Vergleich zum Referenzwert durch die Adap-

³Anmerkung: In einer früheren Version dieser Tabelle in [Middendorf 02] wurden anstelle nur der ROF-Vektoren, die zum bewegten Marmor-Block gehören, alle ROF-Vektoren betrachtet, mit dem Ergebnis, dass der durchschnittliche Fehler aller ROF-Vektoren durch die Adaption stets zunahm. Es wäre jedoch eine verfrühte Interpretation, dass die Adaption zu einer Verschlechterung der OF-Schätzung ausgerechnet bei den als besonders zuverlässig erachteten ROF-Vektoren führt. Bei genauer Betrachtung stellte sich heraus, dass dieses Ergebnis durch die Überlagerung zweier Effekte zustande kam: Aufgrund der besonderen Art der Bildfolge sind die ROF-Vektoren im stationären Hintergrund nahezu fehlerfrei. Durch die Adaption und die damit verbundene Zunahme der Diskontinuitäten nahm der Anteil dieser Vektoren an der Gesamtheit der ROF-Vektoren ab, der durchschnittliche Fehler nahm insgesamt stets zu, obwohl der Fehler vieler ROF-Vektoren abnahm.

It.	Anz. Vekt.	alle Vektoren				nur ROF-Vektoren aus bewegtem Klotz					
		Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kummulativ)			\emptyset Fehler	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kummulativ)			\emptyset Fehler	Ref. \emptyset Fehler
0	10000	4534	5526	8713	0,1235	3634	8	750	3508	0,0302	—

lineare Adaption, $\sigma_{\max} = 2,85$ (V 40)

1	10000	3029	5070	7832	0,1876	2913	13	1227	2808	0,0222	0,0239
2	10000	3034	4917	7827	0,1881	2850	15	1052	2716	0,0252	0,0236
3	10000	3019	4905	7827	0,1920	2786	19	1035	2668	0,0234	0,0236
50	10000	3017	4915	7823	0,1915	2782	17	1044	2648	0,0251	0,0237
100	10000	3017	4915	7823	0,1915	2779	17	1044	2646	0,0250	0,0237

„quasi-inverse“ Adaption, $\sigma_{\max} = 2,85$ (V 41)

1	10000	3271	5850	8231	0,1615	3375	48	1898	3239	0,0208	0,0276
2	10000	3255	5694	8197	0,1664	3318	36	1717	3159	0,0219	0,0288
3	10000	3240	5653	8173	0,1699	3287	27	1665	3116	0,0235	0,0287
50	10000	3244	5661	8182	0,1689	3278	30	1671	3121	0,0223	0,0287
100	10000	3243	5660	8181	0,1689	3278	29	1670	3120	0,0223	0,0287

lineare Adaption, $\sigma_{\max} = 4,00$ (V 42)

1	10000	2576	4357	7350	0,2064	2504	11	932	2449	0,0217	0,0233
2	10000	2527	4195	7331	0,2202	2432	3	793	2262	0,0309	0,0233
3	10000	2496	4152	7362	0,2300	2347	3	763	2256	0,0274	0,0234
50	10000	2523	4074	7272	0,2275	2290	33	711	2159	0,0302	0,0229
100	10000	2525	4073	7275	0,2268	2289	33	709	2159	0,0301	0,0229

„quasi-inverse“ Adaption, $\sigma_{\max} = 4,00$ (V 43)

1	10000	2895	5383	8111	0,1641	3373	31	1856	3260	0,0194	0,0269
2	10000	2902	5350	8081	0,1778	3291	51	1762	3177	0,0195	0,0289
3	10000	2874	5279	8022	0,1825	3213	42	1686	3054	0,0214	0,0277
50	10000	2875	5318	8018	0,1803	3175	33	1702	3033	0,0206	0,0282
100	10000	2875	5318	8016	0,1803	3175	33	1702	3031	0,0206	0,0282

Tabelle 6.7: [marmor_stat, Bild 150] Auswirkungen wiederholter Iterationen bei der adaptiven GST-Bestimmung auf den Fehler der OF-Schätzung: Die Tabelle stellt den Fehler der OF-Schätzung ohne Adaption (= 0. Iteration) und für unterschiedlich viele Iterationen des doppelt-adaptiven Verfahrens zur GST-Bestimmung dar. In der Spalte „Ref. \emptyset Fehler“ ist der durchschnittliche Fehler *derselben* Pixel, die in die Berechnung des links davon stehenden Wertes eingehen, in der 0. Iteration angegeben. Start-Kovarianz: $\sigma_x = \sigma_y = \sigma_t = 1,00$. Min. t -Komponente: 0,25.

tion stets abnimmt, während bei linearer Adaption der Fehler – wenn überhaupt – nur in den ersten Iterationen kleiner, nach großer Iterationszahl aber stets größer ist als der Referenzfehler ohne Adaption.

It.	alle Vektoren				nur ROF-Vektoren aus bewegtem Klotz						
	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kummulativ)			\emptyset Fehler	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kummulativ)			\emptyset Fehler	Ref. \emptyset Fehler
0	10000	4286	5484	8596	0,1429	3664	12	901	3597	0,0247	—
lineare Adaption, $\sigma_{\max} = 2,85$ (V 140)											
1	10000	3011	4924	7847	0,1905	2898	18	1161	2831	0,0198	0,0216
2	10000	3022	4900	7849	0,1890	2890	14	1075	2759	0,0244	0,0214
3	10000	3013	4864	7829	0,1921	2809	19	1011	2710	0,0220	0,0219
50	10000	3017	4915	7824	0,1914	2780	17	1044	2645	0,0251	0,0217
100	10000	3017	4914	7824	0,1914	2779	17	1043	2645	0,0250	0,0217
„quasi-inverse“ Adaption, $\sigma_{\max} = 2,85$ (V 141)											
1	10000	3260	5841	8291	0,1639	3417	37	1872	3288	0,0201	0,0261
2	10000	3266	5709	8218	0,1672	3358	42	1733	3216	0,0212	0,0279
3	10000	3246	5654	8188	0,1692	3288	30	1664	3132	0,0229	0,0278
50	10000	3243	5660	8165	0,1690	3281	29	1670	3108	0,0228	0,0280
100	10000	3243	5660	8166	0,1690	3280	29	1670	3109	0,0228	0,0280
lineare Adaption, $\sigma_{\max} = 4,00$ (V 142)											
1	10000	2574	4214	7378	0,2168	2539	13	856	2471	0,0223	0,0201
2	10000	2513	4042	7369	0,2246	2438	7	738	2285	0,0295	0,0199
3	10000	2505	4052	7383	0,2302	2363	4	736	2282	0,0268	0,0202
50	10000	2524	4075	7275	0,2274	2291	32	713	2160	0,0302	0,0199
100	10000	2527	4075	7279	0,2267	2291	34	713	2162	0,0302	0,0199
„quasi-inverse“ Adaption, $\sigma_{\max} = 4,00$ (V 143)											
1	10000	2883	5406	8167	0,1695	3431	38	1875	3326	0,0182	0,0255
2	10000	2872	5345	8123	0,1787	3320	28	1759	3229	0,0188	0,0280
3	10000	2865	5269	8040	0,1827	3221	30	1672	3076	0,0205	0,0274
50	10000	2875	5320	8004	0,1805	3175	33	1705	3021	0,0210	0,0278
100	10000	2875	5321	8004	0,1805	3175	33	1706	3021	0,0210	0,0278

Tabelle 6.8: [marmor_stat, Bild 150] Wie in Tabelle 6.7, aber mit anderer Start-Kovarianz: $\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$.

Bei linearer Adaption sind die Ergebnisse beim kleinen Skalierungsbereich signifikant besser als beim großen Skalierungsbereich, während bei quasi-inverser Adaption kleiner und großer Skalierungsbereich zu etwa gleichwertigen Ergebnissen führen. Außerdem lässt sich ablesen, dass das Ergebnis nach 100 Iterationen kaum von der Start-Kovarianz abhängt. Bemerkenswert ist noch, dass der Fehler in fast allen Fällen durch die wiederholte Iteration im Vergleich zur ersten zunimmt, was aber beim quasi-inversen Verfahren mit einer gleichzeitigen Zunahme des Referenzfehlers einhergeht.

Wie schon bei der Kategorisierung, so kann auch bei der Qualität der OF-Schätzung

das Ergebnis durch die Adaption besser oder schlechter werden. Es zeichnet sich ein vorteilhaftes Verhalten der neuen, quasi-inversen Adaption im Vergleich zur linearen Adaption ab.

6.4.3 Konvergenz der OF-Schätzung

Die bisherigen Beobachtungen bezogen sich auf das *durchschnittliche* Verhalten der iterativen GST-Schätzung. In diesem Abschnitt soll nun eine Betrachtung der GST-Schätzungen an einzelnen Punkten erfolgen. Es stellt sich die Frage, ob nach mehreren Iterationen nur der Durchschnittsfehler nahezu konstant bleibt, oder ob sich die OF-Schätzung für jeden *einzelnen* Vektor kaum mehr ändert.

Zur Beantwortung dieser Frage wird das Konvergenzverhalten der GST- und damit der OF-Schätzung betrachtet. Konvergenz von GST- und OF-Schätzung sind offensichtlich voneinander abhängig, aber keineswegs zueinander äquivalent:

- Aus Konvergenz der OF-Schätzung kann nicht auf Konvergenz der GST-Schätzung geschlossen werden, da der OF nur auf *einem* der Eigenvektoren der GST beruht. Die beiden übrigen Eigenvektoren können sich beliebig verhalten.
- Umgekehrt ist es denkbar, dass eine Folge von GST-Schätzungen mit paarweise unterschiedlichen Eigenwerten gegen einen GST mit doppeltem Eigenwert konvergiert. Während also die OF-Schätzung für jedes Folgenglied möglich ist (und auch selbst eine konvergente Folge bilden *kann*), ist in diesem Fall für den GST-Grenzwert keine OF-Schätzung mehr möglich⁴.

Der zweite Fall wird bei realen Bildfolgen (und auch bei nicht speziell konstruierten künstlichen Bildfolgen) kaum auftreten. Im Allgemeinen kann daher aus der Konvergenz der GST- auf Konvergenz der OF-Schätzungen geschlossen werden. Aufgrund von Rechenungenauigkeiten und Quantisierungseffekten ist nicht zu erwarten, dass die GST- oder OF-Schätzungen nach mehreren Iterationen *exakt* übereinstimmen. Es soll daher schon dann von einer Konvergenz der GST- oder OF-Schätzungen gesprochen werden, wenn der Unterschied der Schätzungen in aufeinander folgenden Iterationen hinreichend klein wird⁵. Welche Norm zur Quantisierung des Unterschieds herangezogen wird, ist dabei (für die mathematische Konvergenzaussage) gleichgültig. Im Folgenden soll – soweit nicht ausdrücklich anders erwähnt – bei Vektoren die Euklid-Norm und bei Matrizen (also auch Tensoren) die Frobenius-Norm verwendet werden, also:

$$d(\mathbf{u}_i(\mathbf{x}_1), \mathbf{u}_j(\mathbf{x}_2)) = \|\mathbf{u}_i(\mathbf{x}_1) - \mathbf{u}_j(\mathbf{x}_2)\|_2, \quad (6.14)$$

⁴Sofern die OF-Schätzung selbst eine konvergente Folge bildet, so liegt der erweiterte OF-Vektor des OF-Grenzwerts immerhin im zum doppelten Eigenwert des GST-Grenzwerts gehörenden Eigenraum.

⁵Einer strengen mathematischen Betrachtung hält dieses Vorgehen nicht stand: Bei der Reihe $\alpha_N = \sum_{i=1}^N 1/i$ sinkt die Differenz aufeinanderfolgender Reihenglieder irgendwann unter jede noch so kleine Schwelle, trotzdem divergiert die Reihe. Ob dieses Kriterium also geeignet ist, muss sich in der Anwendung zeigen.

$$d(\text{GST}_i(\mathbf{x}_1), \text{GST}_j(\mathbf{x}_2)) = \|\text{GST}_i(\mathbf{x}_1) - \text{GST}_j(\mathbf{x}_2)\|_F . \quad (6.15)$$

Da der OF-Vektor durch Skalierung nur des Eigenvektors zum kleinsten Eigenwert des GST ermittelt wird, kann beim Vergleich zweier GST und der daraus ermittelten OF-Vektoren die Euklid-Norm der OF-Differenz groß sein, obwohl die Frobenius-Norm der GST-Differenz klein ist oder umgekehrt. Für eine umfassende Konvergenzbetrachtung ist es also unumgänglich, sowohl GST- als auch OF-Konvergenz zu betrachten.

Im Rahmen der hier durchgeführten Versuche soll aber zunächst trotzdem nur die Konvergenz der OF-Schätzung betrachtet werden. Dafür sprechen vor allem zwei Gründe:

- Für die Beurteilung der Auswirkungen der iterativen Adaption auf die OF-Schätzung ist es wichtig zu wissen, ob die OF-Schätzung konvergiert.
- Der noch zu bestimmende OF-Konvergenz-Schwellwert hat eine intuitiv einleuchtende Bedeutung.

Als OF-Konvergenz-Schwellwert wird zunächst heuristisch 0,001 Pixel pro Bild bzw. Halbbild gewählt. Dieser Wert liegt deutlich unter der mit heutigen Verfahren durchschnittlich erreichten Genauigkeit der OF-Schätzung (vgl. Abschnitt 6.2). Kleine Änderungen können daher ohne wesentliche Verfälschung der Ergebnisse vernachlässigt werden. Es wird auch dann von Konvergenz in Iteration i gesprochen, wenn die Konvergenzschwelle in nachfolgenden Iterationen wieder überschritten wird. Eine Bewertung der Brauchbarkeit des hier gewählten Konvergenzbegriffs findet sich in Abschnitt 6.4.5.

6.4.3.1 Konvergenz und Oszillation an dominierenden Grauwertstrukturen

Für die Bildfolge `marmor_stat` geben die Tabellen 6.9 und 6.10 Aufschluss über das Konvergenzverhalten. Während der Unterschied in der OF-Schätzung zwischen 0. (nicht-adaptiver) und 1. (adaptiver) Iteration je nach Adaptionsverfahren und Parameterwahl noch bei rund 52 bis 63 Prozent der Vektoren größer als $\frac{1}{100}$ Pixel pro Bild ist, liegt der Unterschied zwischen der 3. und 4. Iteration bei nur noch 8 bis 22 Prozent der Vektoren über $\frac{1}{100}$ Pixel pro Bild. Die meisten Punkte mit großen Unterschieden treten jeweils bei linearer Adaption und großem Skalierungsbereich (Tabelle 6.10) auf. Zwischen 9. und 10. Iteration liegt bei rund 86 bis 95 Prozent der Vektoren der Unterschied sogar unter $\frac{1}{1000}$ Pixel pro Bild. Nur bei linearer Adaption und großem Skalierungsbereich treten noch wesentlich häufiger größere Abweichungen auf.

Die auftretenden Unterschiede zwischen OF-Schätzungen in aufeinander folgenden Iterationen führen aber nicht zu einer systematischen Verbesserung der OF-Schätzung. Wie die Tabellen 6.11 und 6.12 zeigen, nimmt der Fehler für etwa genauso viele Vektoren zu wie ab. In den ersten Iterationen überwiegt sogar meist die Anzahl der Vektoren mit zunehmendem Fehler. Dieses Verhalten stimmt mit den Beobachtungen in Abschnitt 6.4.2.2 überein und kann auch hier auf das Wachsen des Bereichs zurückgeführt werden, in dem sich die Bewegungsdiskontinuität auswirkt.

Iteration	lineare Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 2,85$ (VID 40)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	3020	1221	3621	2138	3002	1511	3353	2134
1 → 2	3573	3337	2146	944	3654	3374	2249	723
2 → 3	5057	3005	1614	324	5280	2823	1631	266
3 → 4	6655	2193	999	153	6512	2245	1102	141
4 → 5	7361	1816	702	121	7233	1830	815	122
5 → 6	7867	1545	488	100	7771	1518	599	112
6 → 7	8221	1263	423	93	8020	1375	495	110
7 → 8	8532	981	394	93	8271	1227	399	103
8 → 9	8683	845	385	87	8431	1108	363	98
9 → 10	8846	695	371	88	8636	920	344	100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	9069	495	346	90	9049	518	342	91
29 → 30	9069	495	346	90	9056	510	341	93
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	9091	479	340	90	9042	518	340	100
99 → 100	9091	479	340	90	9047	511	343	99

Iteration	„quasi-inverse“ Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 2,85$ (VID 41)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	3225	1341	3754	1680	3183	1609	3593	1615
1 → 2	4553	3060	1705	682	4986	2688	1735	591
2 → 3	6195	2332	1265	208	6376	2068	1413	143
3 → 4	7531	1657	745	67	7539	1636	764	61
4 → 5	8211	1319	428	42	8161	1417	381	41
5 → 6	8684	1067	221	28	8688	1062	220	30
6 → 7	9066	747	163	24	9048	770	158	24
7 → 8	9318	523	136	23	9308	545	123	24
8 → 9	9449	403	126	22	9445	426	108	21
9 → 10	9531	327	121	21	9535	345	100	20
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	9663	205	101	31	9667	222	80	31
29 → 30	9663	204	101	32	9666	223	80	31
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	9674	194	97	35	9679	207	77	37
99 → 100	9674	195	96	35	9679	207	77	37

Tabelle 6.9: [marmor_stat, Bild 150] Auswirkungen wiederholter Iterationen bei der adaptiven GST-Bestimmung: Anzahl der Bildpunkte, an denen der Betrag der Differenz zwischen OF-Schätzungen in aufeinander folgenden Iterationen (in Pixel pro Bild) unterhalb des im Spaltenkopf angegebenen Grenzwertes liegt. Kleiner Skalierungsbereich.

Iteration	lineare Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 4,00$ (VID 42)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	2566	1175	3563	2696	2566	1225	3467	2742
1 → 2	2642	2401	3499	1458	2656	2547	3502	1295
2 → 3	3118	3551	2669	662	3121	3698	2606	575
3 → 4	3995	3774	1979	252	3966	3809	2020	205
4 → 5	4913	3395	1495	197	4823	3457	1555	165
5 → 6	5872	2969	1026	133	5783	2965	1118	134
6 → 7	6558	2743	596	103	6346	2507	995	152
7 → 8	7133	2183	583	101	6661	2363	872	104
8 → 9	7533	1892	474	101	6884	2233	783	100
9 → 10	7832	1442	612	114	7020	2165	684	131
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	8477	937	457	129	8282	1092	524	102
29 → 30	8432	990	476	102	8208	1101	568	123
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	8429	936	516	119	8390	985	491	134
99 → 100	8431	998	470	101	8465	981	451	103

Iteration	„quasi-inverse“ Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 4,00$ (VID 43)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	2877	1246	4023	1854	2809	1364	4008	1819
1 → 2	3801	3136	2098	965	4100	3024	1991	885
2 → 3	5172	2677	1695	456	5361	2514	1754	371
3 → 4	6305	2290	1254	151	6443	2170	1257	130
4 → 5	7185	1787	948	80	7235	1761	930	74
5 → 6	7740	1560	645	55	7759	1612	579	50
6 → 7	8189	1305	459	47	8159	1363	434	44
7 → 8	8518	1085	355	42	8526	1093	340	41
8 → 9	8759	904	294	43	8786	894	278	42
9 → 10	8889	814	257	40	8921	807	229	43
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	9433	339	175	53	9418	369	156	57
29 → 30	9430	341	176	53	9418	369	156	57
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	9421	359	165	55	9430	354	156	60
99 → 100	9422	358	165	55	9428	356	156	60

Tabelle 6.10: [marmor_stat, Bild 150] Wie in Tabelle 6.9, aber mit größerem Skalierungsbereich.

Iteration	lineare Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 2,85$ (VID 40)													
	$\sigma_x = \sigma_y = \sigma_t = 1,00$							$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$						
	---	--	-	=	+	++	+++	---	--	-	=	+	++	+++
0 → 1	1494	1599	1216	3183	748	1383	377	1403	1511	1256	3172	879	1286	493
1 → 2	415	1065	1593	4191	1485	875	376	285	1028	1512	4379	1524	964	308
2 → 3	164	814	1380	5667	1256	620	99	125	814	1245	5822	1265	651	78
3 → 4	53	495	836	7070	1077	409	60	27	549	953	6973	977	452	69
4 → 5	46	406	918	7634	707	257	32	55	471	850	7526	780	295	23
5 → 6	26	236	661	8086	717	238	36	24	301	643	7961	755	265	51
6 → 7	31	238	676	8395	463	173	24	47	294	644	8202	613	177	23
7 → 8	19	181	441	8675	445	207	32	21	179	559	8433	563	200	45
8 → 9	29	214	494	8794	285	166	18	40	200	546	8574	465	157	18
9 → 10	18	164	290	8949	347	201	31	18	157	410	8759	433	185	38
10 → 11	31	196	356	9005	233	162	17	37	183	428	8864	315	154	19
11 → 12	16	163	235	9046	315	195	30	17	158	304	8968	335	183	35
12 → 13	29	190	322	9064	216	162	17	36	185	338	9017	250	156	18
13 → 14	18	161	213	9100	288	191	29	18	156	235	9072	296	186	37
14 → 15	30	187	292	9118	196	159	18	39	185	287	9098	219	154	18
15 → 16	17	161	193	9125	289	185	30	18	156	215	9114	275	185	37
16 → 17	30	182	290	9130	190	161	17	37	180	276	9126	207	156	18
17 → 18	16	161	188	9135	285	185	30	18	155	200	9131	275	184	37
18 → 19	30	182	287	9137	188	160	16	37	182	279	9125	205	154	18
19 → 20	16	160	192	9134	286	180	32	18	155	211	9132	270	178	36
20 → 21	33	178	284	9144	185	161	15	35	167	275	9145	204	156	18
21 → 22	15	162	182	9145	284	178	34	18	155	198	9149	269	176	35
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	34	171	286	9155	177	161	16	35	174	261	9158	195	159	18
29 → 30	16	161	178	9157	283	171	34	18	158	196	9154	261	178	35
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	34	166	248	9193	176	166	17	38	176	262	9149	196	158	21
99 → 100	17	166	175	9192	250	167	33	21	160	200	9152	256	173	38

Tabelle 6.11: [marmor_stat, Bild 150] Auswirkungen wiederholter Iterationen bei der adaptiven GST-Bestimmung: Anzahl der Bildpunkte, an denen der Fehler der OF-Schätzung (im Vergleich zur VR) in aufeinander folgenden Iterationen (0. Iteration $\hat{=}$ nicht-adaptiv) abnimmt (+, ...), zunimmt (-, ...) oder nahezu gleich bleibt (=). Bei weniger als $1/1000$ Pixel/Bild Unterschied werden OF-Schätzungen als gleich gut angesehen. Ansonsten wird unterschieden zwischen *leichter* Verbesserung (+) bzw. Verschlechterung (-) bei $1/1000$ bis $1/100$ Pixel/Bild Unterschied, *mittlerer* Verbesserung (++) bzw. Verschlechterung (--) bei $1/100$ bis $1/10$ Pixel/Bild Unterschied und *starker* Verbesserung (+++) bzw. Verschlechterung (---) bei mehr als $1/10$ Pixel/Bild Unterschied.

Im Vergleich zu Tabelle 6.9 fällt auf, dass es mehr Punkte gibt, an denen der Fehler in der 100. Iteration sich kaum mehr ändert (9192 bzw. 9152), als es Punkte gibt, an denen Konvergenz festgestellt wurde (9091 bzw. 9047). Diese anscheinende Diskrepanz kann dadurch erklärt werden, dass sich die OF-Schätzung innerhalb einer Iteration stark ändern kann, der Fehler im Vergleich mit der VR aber in beiden Iterationen gleich (groß) sein kann.

Iteration	„quasi-inverse“ Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 2,85$ (VID 41)													
	$\sigma_x = \sigma_y = \sigma_t = 1,00$							$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$						
	---	--	-	=	+	++	+++	---	--	-	=	+	++	+++
0 → 1	1046	1289	1109	3426	901	1821	408	853	1273	1062	3397	1114	1790	511
1 → 2	388	861	1470	5153	1235	690	203	313	882	1291	5556	1094	671	193
2 → 3	130	695	1237	6590	851	474	23	105	730	1062	6712	838	527	26
3 → 4	18	345	748	7769	788	287	45	12	366	776	7727	761	310	48
4 → 5	36	205	588	8371	658	139	3	38	203	670	8311	637	139	2
5 → 6	2	72	469	8832	478	122	25	1	79	467	8806	492	127	28
6 → 7	22	105	318	9171	338	45	1	24	85	359	9126	348	58	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14 → 15	22	74	140	9653	91	18	2	17	41	118	9664	126	33	1
15 → 16	3	19	83	9657	143	73	22	3	34	120	9665	120	41	17
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	25	61	126	9705	58	21	4	16	37	110	9702	89	32	14
99 → 100	4	21	58	9705	126	61	25	14	32	89	9702	110	37	16

Iteration	lineare Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 4,00$ (VID 42)													
	$\sigma_x = \sigma_y = \sigma_t = 1,00$							$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$						
	---	--	-	=	+	++	+++	---	--	-	=	+	++	+++
0 → 1	1939	1858	1232	2678	624	1176	493	1838	1883	1222	2730	686	1051	590
1 → 2	805	1698	1454	3042	1247	1321	433	645	1781	1538	3071	1277	1244	444
2 → 3	366	1250	1901	3667	1620	1027	169	270	1212	1887	3749	1680	1042	160
3 → 4	88	884	1641	4678	1723	863	123	55	895	1571	4682	1819	852	126
4 → 5	104	822	1789	5486	1276	481	42	94	848	1636	5408	1466	509	39
5 → 6	43	416	1087	6353	1512	517	72	37	391	947	6288	1661	599	77
6 → 7	58	388	1451	6959	993	118	33	73	605	1282	6679	1060	261	40
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	63	282	601	8553	296	166	39	64	287	704	8531	220	153	41
99 → 100	37	137	319	8557	638	259	53	34	140	244	8601	681	248	52

Iteration	„quasi-inverse“ Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 4,00$ (VID 43)													
	$\sigma_x = \sigma_y = \sigma_t = 1,00$							$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$						
	---	--	-	=	+	++	+++	---	--	-	=	+	++	+++
0 → 1	1139	1615	1027	3072	919	1760	468	1015	1538	1018	3031	1040	1815	543
1 → 2	583	967	1536	4454	1310	905	245	517	937	1397	4729	1340	832	248
2 → 3	272	914	1359	5635	1082	639	99	232	900	1245	5818	1034	682	89
3 → 4	55	611	1054	6698	990	515	77	33	604	1008	6815	946	505	89
4 → 5	61	433	933	7436	739	387	11	59	427	942	7510	683	367	12
5 → 6	9	252	709	7946	763	280	41	3	245	698	7995	757	262	40
6 → 7	41	200	597	8366	597	197	2	37	211	605	8332	642	171	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	43	66	240	9474	114	57	6	31	57	167	9502	159	69	15
99 → 100	6	57	114	9474	240	66	43	15	69	160	9502	166	57	31

Tabelle 6.12: [marmor_stat, Bild 150] Wie in Tabelle 6.11, aber mit „quasi-inverser“ Adaption (oben), mit größerem Skalierungsbereich (Mitte) oder mit „quasi-inverser“ Adaption und größerem Skalierungsbereich (unten).

Auffällig ist, dass etwa ab der 10. Iteration für eine etwa gleichbleibende, in den kommenden 20 Iterationen noch langsam und danach gar nicht mehr abnehmende Zahl von Vektoren der Fehler abwechselnd deutlich zu- und abnimmt (siehe Tabelle 6.11) – unabhängig von der untersuchten Kombination von Parametern und Adaptionsvarianze. Eine genauere Analyse zeigt, dass es sich dabei um Vektoren an immer denselben Positionen handelt. Beispielhaft wird im Folgenden der Punkt (339, 129) herausgegriffen, der bei quasi-inverser Adaption, kleinem Skalierungsbereich und Start-Kovarianzen $\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$ zu den 17 Punkten gehört (vgl. Tabelle 6.12, oben rechts), bei denen ab der 14. Iteration der Fehler abwechselnd deutlich zu- und abnimmt. Ab der 17. Iteration wiederholen sich an diesem Punkt auch die wesentlichen Größen des Algorithmus alle zwei Iterationen (für die hier dargestellten Nachkommastellen):

Iteration 17, 19, 21, ...	Iteration 18, 20, 22, ...
$\sigma_1 = 2,8481; \sigma_2 = 1,9322; \sigma_3 = 1,0815$	$\sigma_1 = 2,799; \sigma_2 = 1,529; \sigma_3 = 1,203$
$\Sigma_M = \begin{pmatrix} 13,834 & 1,643 & -3,542 \\ 1,643 & 3,396 & -2,607 \\ -3,542 & -2,607 & 8,799 \end{pmatrix}$	$\Sigma_M = \begin{pmatrix} 11,325 & 2,002 & -5,488 \\ 2,002 & 5,254 & -1,898 \\ -5,488 & -1,898 & 6,656 \end{pmatrix}$
↓	↓
$\overline{\nabla g(\nabla g)^T} = \begin{pmatrix} 1,639 & -0,261 & 2,241 \\ -0,261 & 2,863 & 0,931 \\ 2,241 & 0,931 & 3,797 \end{pmatrix}$	$\overline{\nabla g(\nabla g)^T} = \begin{pmatrix} 0,321 & -0,297 & 0,440 \\ -0,297 & 5,215 & 1,719 \\ 0,440 & 1,719 & 1,525 \end{pmatrix}$
↓	↓
$\lambda_1 = 5,3800; \lambda_2 = 2,8758; \lambda_3 = 0,0432$	$\lambda_1 = 5,8947; \lambda_2 = 1,1657; \lambda_3 = 0,0013$
$\mathbf{e}_1 = (-0,4830; -0,2592; -0,8364)^T$	$\mathbf{e}_1 = (-0,0208; 0,9310; 0,3643)^T$
$\mathbf{e}_2 = (0,3519; -0,9321; 0,0856)^T$	$\mathbf{e}_2 = (0,5218; -0,3007; 0,7983)^T$
$\mathbf{e}_3 = (-0,8018; -0,2530; 0,5415)^T$	$\mathbf{e}_3 = (0,8528; 0,2067; -0,4795)^T$
↓	↓
OF = $(-1,4808; -0,4672)^T$	OF = $(-1,7785; -0,4311)^T$

Während das Ergebnis der OF-Schätzung in der 18. (20., 22., ...) Iteration um weniger als $1/100$ Pixel von der tatsächlichen VR von $(-1,785; -0,430)^T$ abweicht, ist der Fehler in der 17. (19., 21., ...) Iteration deutlich größer. Gleichzeitig führt der kleinste normierte Eigenwert bei der 17. Iteration zu einer Kategorisierung als Diskontinuität, während der Vektor der 18. Iteration als ROF eingestuft wird. Abbildung 6.7 verdeutlicht die unterschiedliche Maskenausdehnung in den einzelnen Iterationen.

Dies legt die Vermutung nahe, dass bei einigen der Punkte, an denen keine Konvergenz auftritt, eine Oszillation zwischen zwei GST- (und somit OF- und Skalen-) Schätzungen vorliegt.



Abbildung 6.7: [marmor_stat, Bild 150] Ausdehnung der dreidimensionalen Mittelungsmasken am Punkt (339,129) für Iterationen 17 (links), 18 (Mitte) und 19 (rechts). Jede Ellipse entspricht einer Standardabweichung. Parameter: „quasi-inverse“ Adaption, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$, $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$.

Tabelle 6.13 zeigt im oberen Teil das Ergebnis einer Untersuchung, welche Regelmäßigkeiten bei Schätzung der OF-Vektoren bei Verwendung der „Standardparameter“ (vgl. Tabelle 6.6) in der 100. Iteration vorliegen. Dazu wird der Begriff des „OF-Schätz-Zyklus“ eingeführt:

Ein Punkt gehört in der k -ten Iteration zu einem OF-Schätz-Zyklus der Länge i genau dann, wenn in *jeder* der Iterationen $k - i + 1, \dots, k - 1, k$ die Differenz des aktuellen OF zum jeweils i Iterationen zuvor berechneten OF unter der Konvergenz-Schwelle liegt, und der Punkt zu keinem OF-Schätz-Zyklus mit einer Länge $< i$ gehört. Eine Oszillation stellt also einen OF-Schätz-Zyklus der Länge zwei dar.

Wie man der Tabelle entnimmt, liegt an den meisten Bildpunkten die Differenz der OF-Schätzung zur vorangehenden Iteration unter der Konvergenz-Schwelle. An den meisten übrigen Punkten liegt tatsächlich eine Oszillation vor. Es treten aber auch längere Zyklen von OF-Schätzungen auf, und vor allem bei großer Start-Kovarianz, großem Skalierungsbereich und linearer Adaption ist für eine nicht unerhebliche Zahl von Punkten gar keine Regelmäßigkeit der OF-Schätzung festzustellen.

Eine vertiefende Untersuchung zeigte, dass die Wahl der Maskengröße das Konvergenzverhalten nicht unwesentlich beeinflusst. Bisher wurde mit Masken gearbeitet, deren theoretischer Maskenfehler (vgl. Abschnitt 3.2.1, Seite 29) bei 1% liegt, bei Beschränkung der maximalen Maskengröße auf $19 \times 19 \times 19$ Pixel. Vergrößert man nun die Maskenausdehnung um jeweils einen Pixel bei Beibehaltung der bisherigen maximalen Maskengröße, so ergibt sich ein geringfügig, aber insgesamt signifikant verbessertes Konvergenzverhalten (unterer Teil von Tabelle 6.13).

Die naheliegende Frage ist nun, wo im Bild sich die Punkte befinden, an denen keine Konvergenz eintritt. Wie man Abbildung 6.8, obere Hälfte, entnimmt, liegen fast alle betroffenen Punkte *in der Nähe* einer besonders starken Grauwertkante oder -ecke, we-

V-ID	40		41		42		43		
Adapt.-Verf.	linear		quasi-invers		linear		quasi-invers		
$\sigma_{\min}, \sigma_{\max}$	1,00; 2,85				1,00; 4,00				
σ_x, σ_y	1,00	1,41	1,00	1,41	1,00	1,41	1,00	1,41	ges.
σ_t	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	

	Bildfolge marmor_stat, „Standardmasken“, Iteration 100								
Anz. Pixel gesamt	10000								80000
Diff. unter Konv.-Schwelle	9091	9047	9674	9679	8431	8465	9422	9428	73237
Oszillation (2er-Zyklus)	853	842	325	321	898	916	576	571	5302
3er Zyklus	3	0	0	0	0	2	1	1	7
4er Zyklus	51	89	0	0	49	10	0	0	199
5er Zyklus	0	0	0	0	15	0	0	0	15
6er Zyklus	1	0	0	0	0	0	0	0	1
7er Zyklus	0	0	0	0	0	0	0	0	0
8er Zyklus	0	0	0	0	3	0	0	0	3
mehr als 8er oder kein Zyklus	1	22	1	0	604	607	1	0	1236

	Bildfolge marmor_stat, um 1 Pixel vergrößerte Masken, It. 100								
Anz. Pixel gesamt	10000								80000
Diff. unter Konv.-Schwelle	9138	9085	9696	9713	8555	8604	9459	9463	73713
Oszillation (2er-Zyklus)	774	802	302	287	828	835	539	535	4902
3er Zyklus	0	0	0	0	53	3	1	1	58
4er Zyklus	85	42	0	0	5	2	0	0	134
5er Zyklus	0	0	0	0	0	0	0	0	0
6er Zyklus	0	0	0	0	174	99	0	0	273
7er Zyklus	0	0	0	0	0	0	0	0	0
8er Zyklus	0	48	0	0	0	1	0	0	49
mehr als 8er oder kein Zyklus	3	23	2	0	385	456	1	1	871

Tabelle 6.13: Oszillationen und Zyklen der OF-Schätzung: Diese Tabelle untersucht das Konvergenzverhalten der OF-Schätzung in einem Ausschnitt von Bild 150 der Bildfolge `marmor_stat` bei Verwendung der „Standardmasken“ (automatisch bestimmte Maskengröße mit 1 % theoretischen Fehler, maximale Größe 19^3 (oben) sowie bei Vergrößerung der so automatisch ermittelten Maskenausdehnung um ein Pixel unter Beibehaltung der maximalen Größe (unten). Man erkennt, dass der Unterschied der OF-Schätzung zur vorangehenden Iteration für über 90 % der Punkte unter der Konvergenz-Schwelle liegt. An den meisten übrigen Punkten liegt eine Oszillation zwischen zwei OF-Schätzungen vor, es treten aber auch Zyklen unterschiedlicher Länge und OF-Schätzungen ohne direkt erkennbare Regelmäßigkeiten auf.

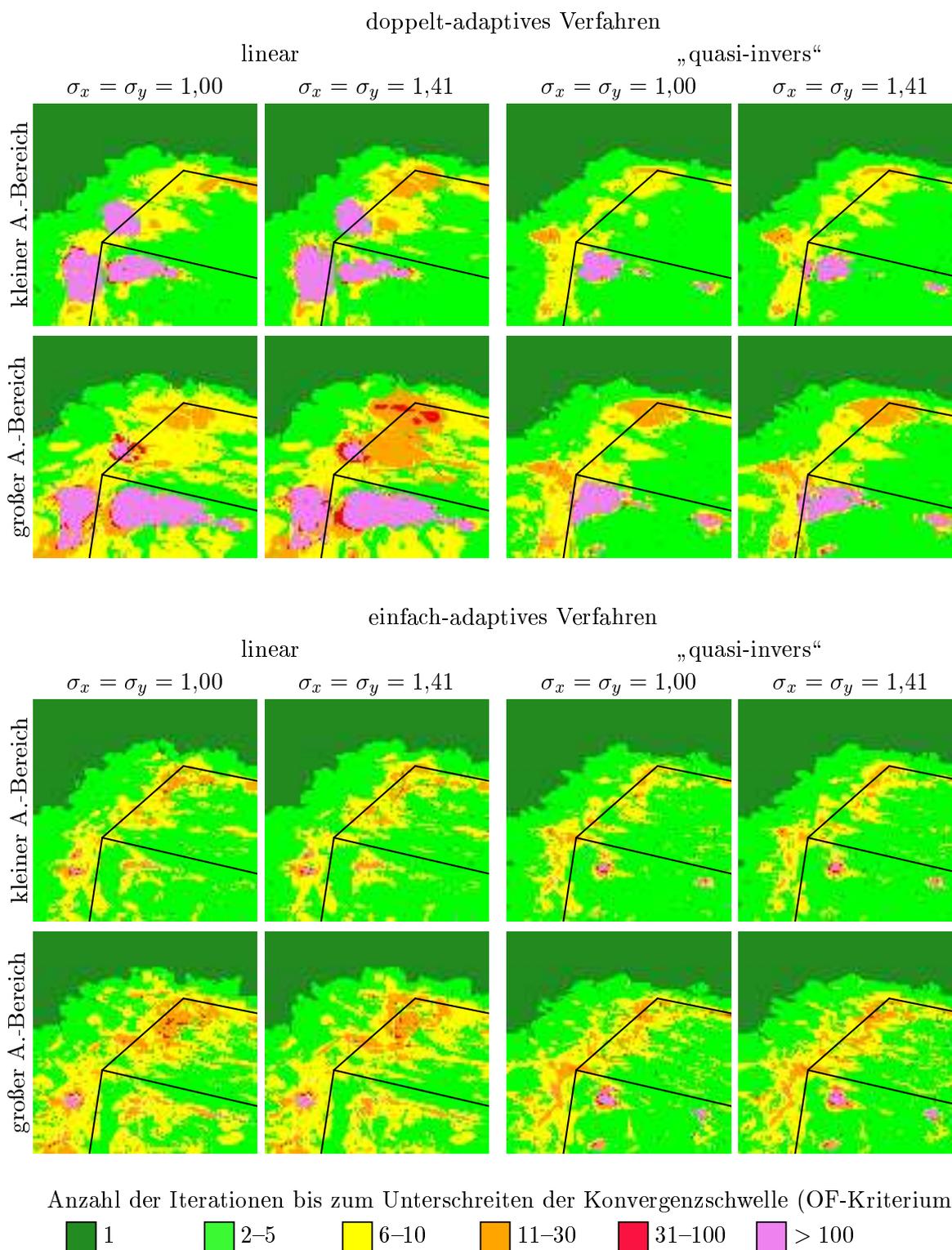


Abbildung 6.8: [marmor_stat, Bild 150] Anzahl der Iterationen bis zur Konvergenz, doppelt-adaptives Verfahren (oben) vs. einfach-adaptives Verfahren (unten).

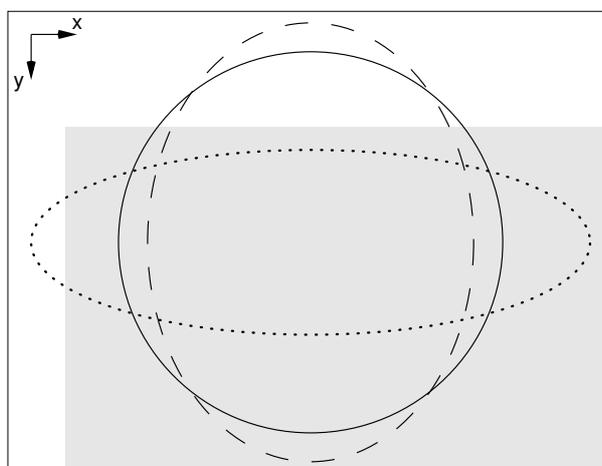


Abbildung 6.9: Schematische Darstellung der Adaption „über die Sichtweite hinaus“: Betrachtet werden die ersten Iterationen der GST-Adaption. Im Startschritt liegt eine waagrecht verlaufende Kante in dem Bereich, der bei der GST-Bestimmung berücksichtigt wird (angedeutet durch den durchgezogenen Kreis). Durch Streckung der Maske entlang der Kante und Stauchung senkrecht dazu gerät eine senkrechte Kante in den Einflussbereich der GST-Schätzung (gepunktete Ellipse), die waagerechte Kante liegt nun außerhalb des berücksichtigten Bereiches. Im nächsten Iterationsschritt wird daher der gestrichelt markierte Bereich berücksichtigt, dann wieder der gepunktete Bereich usw.

nige Punkte genau auf einer Kante. Eine mögliche Ursache für die Oszillationen oder längere Zyklen liegt darin, dass die maximale Maskenausdehnung mehr als zweieinhalb mal so groß ist wie die minimale Maskenausdehnung. Durch den Adaptionsschritt können somit Punkte signifikanten Einfluss auf die GST-Schätzung erlangen, deren Gewicht zuvor vernachlässigbar war, und umgekehrt. Abbildung 6.9 demonstriert an einem Beispiel, wie es dann *in der Nähe von Kanten* zu Oszillationen kommen kann. Durch die wechselseitige Beeinflussung der GST-Schätzung an benachbarten Punkten wegen der Verwendung des doppelt-adaptiven Verfahrens kann es dann auch *auf Kanten* zu Oszillationen kommen. Ein Vergleich mit dem Konvergenzverhalten bei Verwendung des einfach-adaptiven Verfahrens (Abbildung 6.8, untere Hälfte) zeigt, dass ohne wechselseitige Beeinflussung der GST-Schätzung tatsächlich alle Fälle von Nicht-Konvergenz in der Nähe von Kanten, aber nie auf Kanten auftreten. Bemerkenswert ist, dass bei einfach-adaptiven Verfahren zwar insgesamt weniger Fälle von Nicht-Konvergenz auftreten, es sich aber in vielen Fällen um isolierte Punkte handelt, während solche Punkte beim doppelt-adaptiven Verfahren meist geballt auftreten.

6.4.3.2 Konvergenz und Oszillation bei mangelnder Grauwertstruktur

Betrachtet man das Konvergenzverhalten des OF auf der Bildfolge `kwbB` (Tabellen 6.14 und 6.15), so erkennt man ein ähnliches Verhalten wie bei `marmor_stat`, nur dass bei

Iteration	lineare Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 2,85$ (VID 40)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	3	193	9378	16026	3	244	12169	13184
1 → 2	1991	10400	9730	3479	2694	11189	9038	2679
2 → 3	7513	10464	6379	1244	8537	10215	5781	1067
3 → 4	12442	8749	3631	778	13362	8276	3271	691
4 → 5	15954	6552	2437	657	16515	6185	2283	617
5 → 6	18023	4969	1989	619	18506	4576	1918	600
6 → 7	19293	3902	1783	622	19514	3723	1762	601
7 → 8	19964	3346	1694	596	20060	3255	1685	600
8 → 9	20419	2935	1649	597	20364	2993	1658	585
9 → 10	20512	2890	1611	587	20520	2886	1605	589
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	20978	2517	1525	580	20943	2507	1545	605
29 → 30	20984	2513	1522	581	20949	2502	1544	605
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	20982	2522	1506	590	20958	2494	1530	618
99 → 100	20979	2528	1503	590	20958	2495	1529	618

Iteration	„quasi-inverse“ Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 2,85$ (VID 41)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	0	208	9779	15613	2	227	12592	12779
1 → 2	2587	10110	8242	4661	3738	10543	7161	4158
2 → 3	8616	8852	6060	2072	10206	8107	5303	1984
3 → 4	12525	7738	4221	1116	13653	6861	3986	1100
4 → 5	15173	6709	2872	846	15944	6068	2776	812
5 → 6	17229	5353	2309	709	17587	5101	2238	674
6 → 7	18421	4555	1962	662	18769	4279	1923	629
7 → 8	19219	3963	1784	634	19344	3912	1723	621
8 → 9	19622	3663	1688	627	19791	3541	1660	608
9 → 10	19933	3441	1610	616	20014	3408	1577	601
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	20609	2922	1453	616	20549	3015	1404	632
29 → 30	20611	2919	1453	617	20549	3015	1403	633
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	20646	2902	1431	621	20545	3036	1388	631
99 → 100	20655	2894	1432	619	20546	3034	1388	632

Tabelle 6.14: [kwB, Bild 2540] Auswirkungen wiederholter Iterationen bei der adaptiven GST-Bestimmung: Anzahl der Bildpunkte, an denen der Betrag der Differenz zwischen OF-Schätzungen in aufeinander folgenden Iterationen unterhalb des im Spaltenkopf angegebenen Grenzwertes liegt (in Pixel pro Halbbild) – kleiner Skalierungsbereich.

Iteration	lineare Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 4,00$ (VID 42)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	0	169	8871	16560	2	279	11187	14132
1 → 2	986	9014	11330	4270	1389	9462	11036	3713
2 → 3	4306	10913	8216	2165	4706	11099	7779	2016
3 → 4	7926	10710	5350	1614	8530	10557	5052	1461
4 → 5	11119	9280	3700	1501	11453	9156	3617	1374
5 → 6	13505	7445	3212	1438	13737	7385	3151	1327
6 → 7	15057	6319	2807	1417	15285	6203	2817	1295
7 → 8	15978	5624	2601	1397	16247	5414	2634	1305
8 → 9	16508	5194	2509	1389	16849	4922	2507	1322
9 → 10	16978	4842	2402	1378	17252	4590	2439	1319
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	18810	3265	2161	1364	18841	3195	2237	1327
29 → 30	18808	3268	2162	1362	18863	3172	2237	1328
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	18834	3222	2170	1374	18826	3147	2258	1369
99 → 100	18831	3226	2170	1373	18829	3144	2255	1372

Iteration	„quasi-inverse“ Adaption, $\sigma_{\min} = 1,0, \sigma_{\max} = 4,00$ (VID 43)							
	$\sigma_x = \sigma_y = \sigma_t = 1,00$				$\sigma_x = \sigma_y = 1,41, \sigma_t = 1,00$			
	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$	$<^1/1000$	$<^1/100$	$<^1/10$	$\geq^1/10$
0 → 1	0	167	9634	15799	1	236	12197	13166
1 → 2	1893	9225	9020	5462	2585	9881	8232	4902
2 → 3	5938	9127	7720	2815	7154	8832	6799	2815
3 → 4	9293	8457	5831	2019	10189	7903	5570	1938
4 → 5	11528	7946	4296	1830	12230	7385	4205	1780
5 → 6	13277	6897	3694	1732	13763	6522	3614	1701
6 → 7	14408	6070	3455	1667	14787	5766	3393	1654
7 → 8	15167	5502	3280	1651	15504	5263	3192	1641
8 → 9	15707	5080	3176	1637	16033	4799	3146	1622
9 → 10	16021	4852	3082	1645	16297	4612	3079	1612
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28 → 29	17091	3986	2877	1646	17098	3973	2900	1629
29 → 30	17084	3993	2874	1649	17104	3967	2898	1631
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98 → 99	17108	3938	2896	1658	17072	3976	2898	1654
99 → 100	17104	3941	2897	1658	17080	3969	2898	1653

Tabelle 6.15: [kwbb, Bild 2540] Wie in Tabelle 6.14, aber mit größerem Skalierungsbereich.

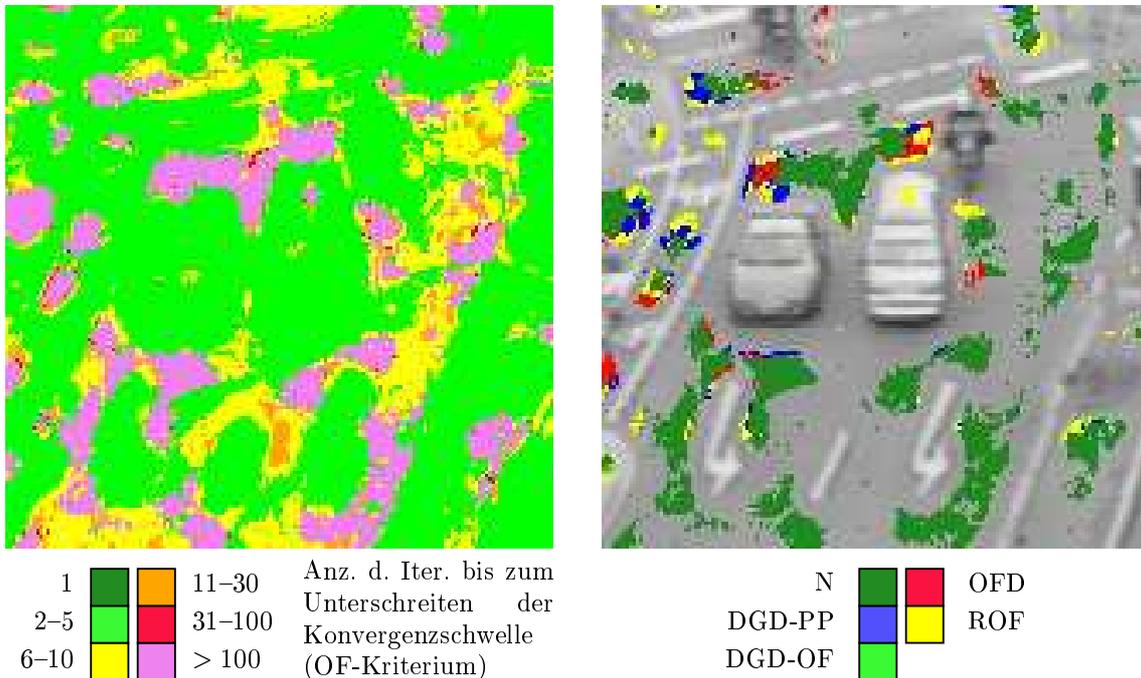


Abbildung 6.10: [kwbB, Bild 2540] Anzahl der Iterationen bis zur Konvergenz. Wie schon in Abbildung 6.8 erkennt man im linken Teil der Abbildung, dass in den meisten Fällen, in denen nach 30 Iterationen keine Konvergenz auftrat, auch nach 100 Iterationen keine Konvergenz auftritt. Rechts ist die Kategorisierung der Punkte ohne Konvergenz dargestellt. Man sieht, dass es sich meist um Punkte der Kategorie „Neutral“ handelt, an denen also keine ausreichende Grauwertstruktur vorliegt. Parameter: „quasi-inverse“ Adaption, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$, $\sigma_x = 1,41$, $\sigma_t = 1,00$. Spur-Schwelle (Kat.): 2,50.

dieser Bildfolge bei einem wesentlich größeren Anteil der Bildpunkte auch nach 100 Iterationen noch keine Konvergenz eingetreten ist. Wie Abbildung 6.10 zeigt, liegen in diesem Fall aber die meisten Punkte ohne Konvergenz in Bereichen, in denen die Spur des GST klein ist, in denen also kaum Grauwertstruktur vorliegt. Kleine (absolute) Unterschiede der GSTs führen an diesen Stellen zu relativ großen OF-Unterschieden. Als Beispiel werde der Punkt (190, 346) betrachtet. Bei quasi-inverser Adaption, großer Start-Kovarianz und kleinem Skalierungsbereich ergeben sich folgende Schätzungen:

	Iteration 50, 52, 54, ...	Iteration 51, 53, 55, ...
GST =	$\begin{pmatrix} 0,0575 & -0,0027 & 0,0305 \\ -0,0027 & 0,0965 & 0,1217 \\ 0,0305 & 0,1217 & 0,3458 \end{pmatrix}$	$\begin{pmatrix} 0,0601 & -0,0059 & 0,0299 \\ -0,0059 & 0,0809 & 0,1023 \\ 0,0299 & 0,1023 & 0,3610 \end{pmatrix}$
\Rightarrow OF =	$(-2,6833; 0,4749)^T$	$(-3,2286; 0,3959)^T$

An diesem Punkt liegt wie schon bei den meisten nicht-konvergenten OF-Schätzungen in der Bildfolge `marmor_stat` eine Oszillation zwischen zwei stabilen Schätzungen vor.

Dies trifft aber auch bei dieser Bildfolge nicht für alle Punkte ohne OF-Konvergenz zu. Bei acht untersuchten Parameter-Algorithmus-Variationen auf einem je 25600 Pixel großen Ausschnitt, also insgesamt $8 \times 25600 = 204800$ iterativen GST-Schätzungen, überschreitet die OF-Differenz zwischen 99. und 100. Iteration bei $204800 - 154982 = 49818$ Punkten den Konvergenz-Schwellwert von 0,001 Pixel pro Halbbild⁶.

Tabelle 6.16 zeigt – analog zu Tabelle 6.13 – das Konvergenzverhalten im untersuchten Ausschnitt der Bildfolge `kwbb`. Bei Verwendung der „Standardmasken“ liegt an 48110 Punkten eine Oszillation vor, an 1166 Punkten durchläuft die OF-Schätzung einen Zyklus von drei bis acht Iterationen. An den verbliebenen 542 Punkte liegt entweder ein Zyklus mit mehr als acht Iterationen oder gar keine regelmäßige Wiederholung der OF-Schätzungen vor.

Vergrößert man nun wieder die Maskenausdehnung um jeweils einen Pixel bei Beibehaltung der bisherigen maximalen Maskengröße, so gibt es nur noch $204800 - 158395 = 46405$ (statt 49818) Pixel, an denen zwischen 99. und 100. Iteration der Konvergenz-Schwellwert überschritten wird (vgl. mittlerer Teil von Tabelle 6.16). Wie schon bei den kleineren Masken liegt in den meisten Fällen eine Oszillation zwischen zwei OF-Schätzungen vor, aber es gibt auch wieder größere Schätz-Zyklen oder Punkte, an denen keine regelmäßige Änderung festzustellen ist.

Eine weitere Betrachtung eben dieser Pixel, an denen keine regelmäßige Änderung festzustellen war, ergab, dass sich die OF-Schätzung dort in vielen Fällen sehr langsam einem regelmäßigen Verhalten annähern. Betrachtet man daher weitere 150 Iterationen (vgl. unterer Teil von Tabelle 6.16), so verbleiben in der 250. Iteration etwas weniger Pixel, an denen die OF-Schätzung keinem Zyklus folgt. Dabei nimmt aber die Anzahl dieser Punkte ab der 200. Iteration kaum mehr ab. Versuche mit bis zu 1000 Iterationen ergaben ebenfalls keine weitere Reduzierung dieser Punkte mehr.

6.4.3.3 Fazit zur Konvergenz

Für einen Großteil der Pixel konvergiert die OF-Schätzung innerhalb weniger Iterationen. An den meisten Punkten ohne Konvergenz liegt Oszillation vor, es verbleiben aber einige wenige Punkte, bei denen gar kein regelmäßiges Verhalten festgestellt werden kann. Allerdings führen an einigen Punkten angesichts der mangelnden Grauwertstruktur sehr kleine Änderungen des GST zu relativ großen Änderungen der OF-Schätzung.

6.4.4 Robustheit gegenüber Parameterwahl

In diesem Abschnitt wird untersucht, wie sich die Wahl der Start-Kovarianzmatrix, des Skalierungsbereiches und des Adaptionsverfahrens auf das Ergebnis der iterativen

⁶Dies sind mehr Punkte als die, an denen Nicht-Konvergenz festgestellt wurde, da bereits dann von Konvergenz ausgegangen wird, wenn die OF-Differenz im Verlauf der Iterationen *einmal* die Konvergenz-Schwelle unterschreitet.

V-ID	40		41		42		43		
Adapt.-Verf.	linear		quasi-invers		linear		quasi-invers		
$\sigma_{\min}, \sigma_{\max}$	1,00; 2,85				1,00; 4,00				
σ_x, σ_y	1,00	1,41	1,00	1,41	1,00	1,41	1,00	1,41	ges.
σ_t	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	

	Bildfolge kwB, „Standardmasken“, Iteration 100								
Anz. Pixel gesamt	25600								204800
Diff. unter Konv.-Schwelle	20979	20958	20655	20546	18831	18829	17104	17080	154982
Oszillation (2er-Zyklus)	4495	4545	4827	4971	6489	6572	8037	8174	48110
3er Zyklus	3	5	1	2	4	5	0	0	20
4er Zyklus	22	15	84	57	202	83	334	193	990
5er Zyklus	1	1	0	0	0	0	0	0	2
6er Zyklus	8	5	6	8	18	24	13	16	98
7er Zyklus	0	2	0	0	0	0	0	0	2
8er Zyklus	4	3	9	0	7	7	15	9	54
> 8er oder kein Zyklus	88	66	18	16	49	80	97	128	542

	Bildfolge kwB, um 1 Pixel vergrößerte Masken, Iteration 100								
Anz. Pixel gesamt	25600								204800
Diff. unter Konv.-Schwelle	21349	21353	21180	21038	19104	19055	17677	17639	158395
Oszillation (2er-Zyklus)	4092	4131	4390	4549	6311	6339	7682	7734	45228
3er Zyklus	13	4	1	1	2	3	0	0	24
4er Zyklus	7	6	15	1	89	149	107	151	525
5er Zyklus	0	0	0	0	0	0	0	0	0
6er Zyklus	8	33	0	5	20	10	27	35	138
7er Zyklus	0	0	0	0	0	0	0	0	0
8er Zyklus	17	1	0	0	2	3	18	3	44
> 8er oder kein Zyklus	114	72	14	6	72	41	89	38	446

	Bildfolge kwB, um 1 Pixel vergrößerte Masken, Iteration 250								
Anz. Pixel gesamt	25600								204800
Diff. unter Konv.-Schwelle	21345	21343	21176	21049	19114	19059	17652	17597	158335
Oszillation (2er-Zyklus)	4126	4160	4405	4544	6324	6339	7757	7775	45430
3er Zyklus	12	5	1	1	3	3	0	0	25
4er Zyklus	20	7	15	1	116	148	103	151	561
5er Zyklus	0	0	0	0	0	0	0	0	0
6er Zyklus	5	29	0	5	20	16	22	41	138
7er Zyklus	0	0	0	0	0	0	0	0	0
8er Zyklus	36	4	1	0	3	3	17	12	76
> 8er oder kein Zyklus	56	52	2	0	20	32	49	24	235

Tabelle 6.16: Oszillationen und Zyklen der OF-Schätzung: Wie Tabelle 6.13, aber für einen Ausschnitt von Bild 2540 der Bildfolge kwB, im unteren Teil ergänzt um eine Untersuchung der 250. Iteration. Etwa 75 bis 80 % der Punkte unterschreiten die Konvergenz-Schwelle.

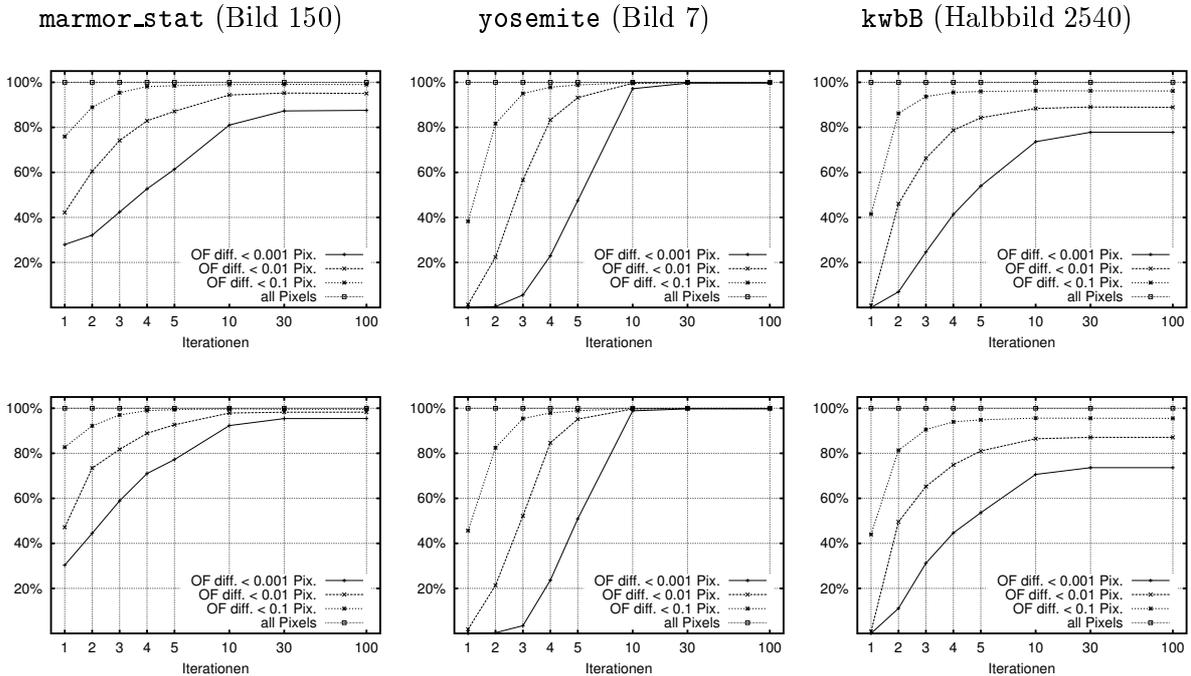


Abbildung 6.11: Anteil der Pixel (in Prozent), an denen nach 1 bis 5, 10, 30 oder 100 Iterationen verschiedene Konvergenz-Schwellwerte unterschritten wurden: Der lineare Adaptionsalgorithmus (oben) konvergiert ein wenig schneller bei einer 10 mal höheren Konvergenzschwelle von $1/100$ Pixel pro Bild (bzw. Halbbild), wohingegen der „quasi-inverse“ Ansatz bei der üblichen Schwelle von $1/1000$ schneller konvergiert (untere Reihe).

GST-Schätzung auswirken.

Bei der Beurteilung der *Qualität der GST-Schätzung* wurden bereits die acht untersuchten Verfahrensvarianten miteinander verglichen. Es zeigte sich, dass die Start-Kovarianz nach mehreren Iterationen nur noch sehr geringen Einfluss auf die GST-Kategorisierung und die OF-Schätzung hat. Skalierungsbereich und Adaptionsverfahren hingegen haben insbesondere bei der Bildfolge *marmor_stat* erkennbaren Einfluss auf das Ergebnis: Nur das quasi-inverse Verfahren mit kleinem Skalierungsbereich führt zu einer Verbesserung der OF-Schätzung und liefert darüber hinaus die besten Kategorisierungsergebnisse. Bei den beiden anderen untersuchten Bildfolgen sind die Ergebnisse nicht so offensichtlich, widersprechen aber auch nicht dieser Beobachtung.

Das *Konvergenzverhalten* der untersuchten Verfahrensvarianten ist in den Abbildungen 6.11, 6.12 und 6.13 dargestellt. Die Grafiken zeigen jeweils in Abhängigkeit von der Iterationsanzahl, an wieviel Prozent der untersuchten Bildpunkte die Differenz der OF-Schätzung unter den Konvergenzschwellwert 0,001 Pixel pro (Halb-)bild bzw. unter eine 10- bzw. 100-mal so große Schwelle fällt. Auch hier zeigen sich ein nur geringer Einfluss der Wahl der Start-Kovarianz (Abbildung 6.12) und leichte Vorteile für den kleinen Skalierungsbereich (Abbildung 6.13). Bei der Untersuchung des Einflusses des

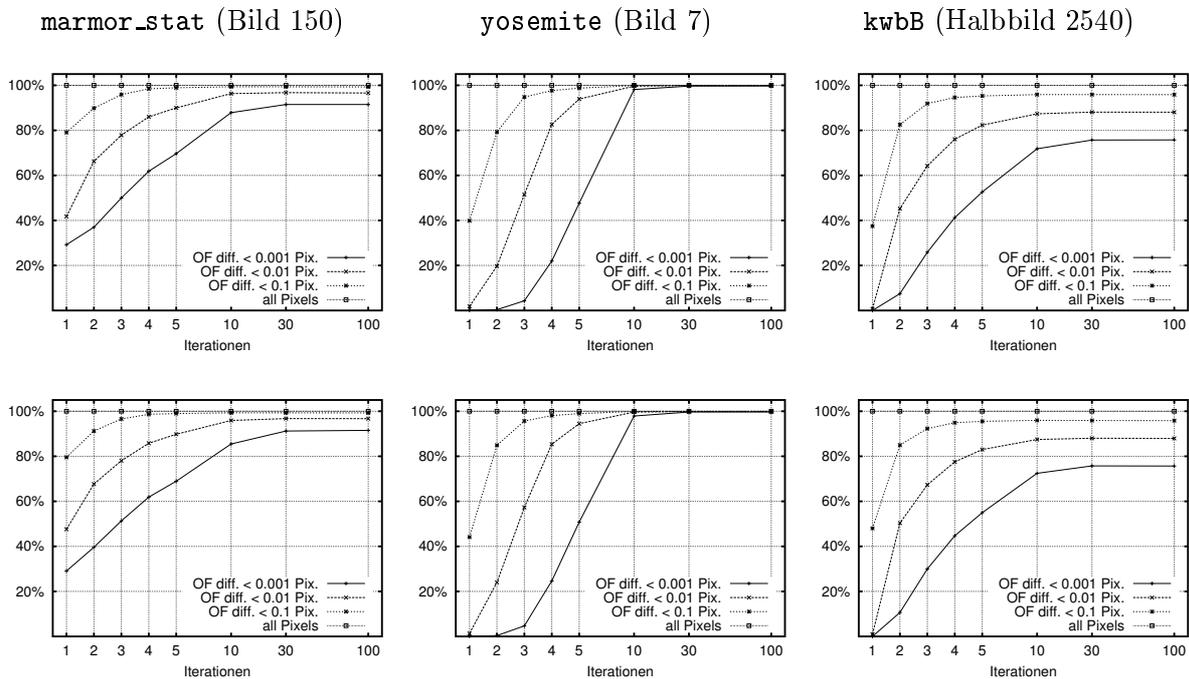


Abbildung 6.12: Vergleich unterschiedlicher Start-Kovarianzen: $\Sigma_A = \text{diag}(1,0; 1,0; 1,0)$ (oben) vs. $\Sigma_A = \text{diag}(2,0; 2,0; 1,0)$ (unten).

Adaptionsverfahrens auf die Konvergenz (Abbildung 6.11) zeigen beide Verfahren für je eine Bildfolge leichte Vorteile. Insgesamt ist unter dem Aspekt des Konvergenzverhaltens keine der untersuchten Verfahrensvarianten den anderen eindeutig überlegen.

6.4.5 Konvergenzbegriff und Konvergenzkriterium

Bisher wurde von Konvergenz der OF-Schätzung an einem Punkt ausgegangen, wenn die Differenz zweier aufeinander folgender Iterationen unter eine heuristisch festgelegte Konvergenzschwelle fiel. Es kann aber passieren, dass trotzdem keine Konvergenz der OF-Schätzung im mathematischen Sinne vorliegt, wenn beispielsweise

1. in nachfolgenden Iterationen diese Konvergenzschwelle wieder überschritten wird,
2. es zu Oszillationen oder Zyklen der OF-Schätzung kommt, ohne dass die Schwelle überschritten wird, oder
3. sich die OF-Schätzung ohne Überschreitung der Konvergenzschwelle systematisch „in eine Richtung“ weiterentwickelt.

Alle drei Fälle werden im Folgenden näher untersucht. Darüber hinaus soll geprüft werden, inwiefern die gewonnenen Ergebnisse davon abhängen, welches Konvergenzkriterium gewählt wird.

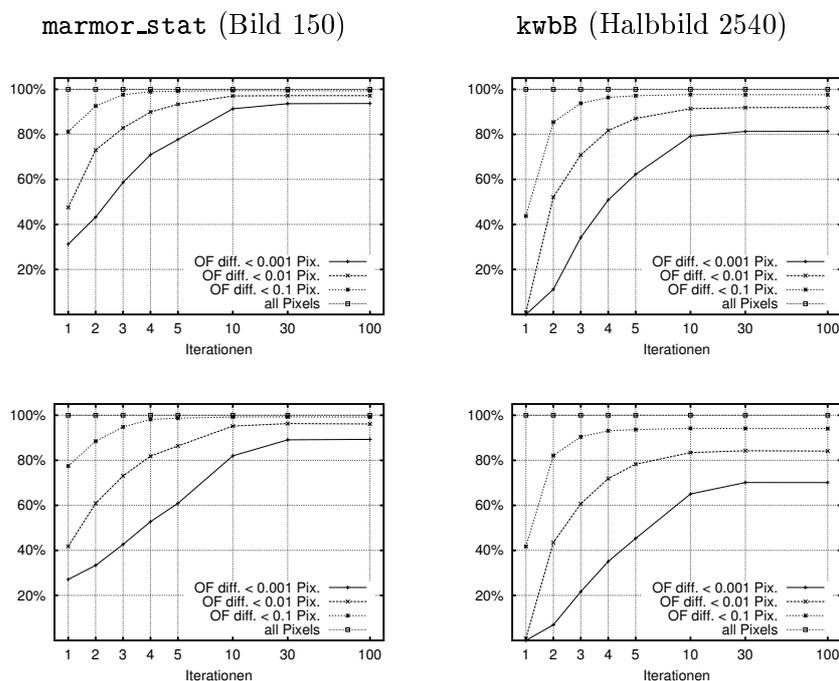


Abbildung 6.13: Vergleich des kleinen Skalierungsbereiches $\sigma_{\max}^2 = 8,0$ Pixel (oben) mit dem großen Skalierungsbereich $\sigma_{\max}^2 = 16,0$ Pixel (unten). Der Vergleich wurde für die **yosemite** Bildfolge nicht durchgeführt, da die vorhandenen 15 Bilder für den großen Skalierungsbereich nicht ausreichen.

6.4.5.1 Wieder-Überschreiten der Konvergenzschwelle

Bisher wurde davon ausgegangen, dass der Unterschied der OF-Schätzung an einem Punkt in zwei aufeinander folgenden Iterationen unterhalb einer Schwelle von $\frac{1}{1000}$ Pixel pro Bild (bzw. Halbbild) bleibt, sobald er einmal unter diese Schwelle gefallen ist. Nur unter dieser Annahme war es gerechtfertigt, in einem solchen Fall von Konvergenz zu sprechen und den Punkt in weiteren Iterationen gar nicht mehr zu betrachten. Im Laufe der Versuche zeigte sich jedoch, dass das Verhalten an einigen Punkten nicht dieser Annahme entspricht.

Tabelle 6.17 stellt das Konvergenzverhalten an Stellen dar, an denen die ursprüngliche Annahme verletzt wird. Dabei sind die Punkte von Interesse, an denen der OF-Unterschied zwischenzeitlich unter die Konvergenzschwelle von $\frac{1}{1000}$ Pixel pro Bild bzw. Halbbild sinkt, später aber wieder über diese Schwelle ansteigt (im Folgenden „Ausreißer“ genannt). Dieses Problem tritt bei allen Versuchen auf, je nach Bildfolge und Versuchsparametern bei 3,8 bis 19,7 Prozent aller Punkte, und meist in der Iteration, die unmittelbar auf die Iteration folgt, in der die Konvergenzschwelle erstmals unterschritten wurde. Bei vereinzelten Punkten wird allerdings die Konvergenzschwelle in 20 oder mehr aufeinander folgenden Iterationen unterschritten, um dann doch wieder überschritten zu werden. Bei vielen dieser Punkte sinkt die OF-Differenz bereits

Bildfolge	marmor_stat									
V-ID	40		41			42		43		
σ_x, σ_y	1,00	1,41	1,00	1,41	2,68	1,00	1,41	1,00	1,41	
σ_t	1,00	1,00	1,00	1,00	2,00	1,00	1,00	1,00	1,00	
Anz. Pixel gesamt	10000									
davon mind. 1× unter K-Schwelle	9188	9206	9748	9786	9787	8972	8936	9597	9635	
davon wieder über Schwelle	609	614	370	430	476	1665	1660	728	712	
sofort in nächster Iteration	537	494	305	346	398	1333	1288	584	572	
in späterer Iteration	72	120	65	84	78	332	372	144	140	
... und sofort wieder unter Schwelle	379	276	217	232	268	528	458	391	339	
... und später wieder unter Schwelle	136	190	65	83	96	585	711	155	172	
... und nie wieder unter Schwelle	94	148	88	115	112	552	491	182	201	

Bildfolge	kwbB									
V-ID	40		41			42		43		
σ_x, σ_y	1,00	1,41	1,00	1,41	2,68	1,00	1,41	1,00	1,41	
σ_t	1,00	1,00	1,00	1,00	2,00	1,00	1,00	1,00	1,00	
Anz. Pixel gesamt	25600									
davon mind. 1× unter K-Schwelle	21846	21768	21599	21503	21411	19931	20123	18370	18373	
davon wieder über Schwelle	2083	2092	2181	1967	2056	3783	3963	3097	2932	
sofort in nächster Iteration	1560	1577	1667	1507	1558	2828	2737	2288	2147	
in späterer Iteration	523	515	514	460	498	955	1226	809	785	
... und sofort wieder unter Schwelle	797	834	829	752	818	1254	1185	1110	944	
... und später wieder unter Schwelle	443	485	425	315	367	1436	1512	685	646	
... und nie wieder unter Schwelle	843	773	927	900	871	1093	1266	1302	1342	

Tabelle 6.17: Konvergenzverhalten einzelner Pixel. Für alle Versuche, die im Zusammenhang mit der Konvergenz durchgeführt wurden, ist dargestellt, an wievielen Punkten die OF-Differenz im Verlauf von 100 Iterationen mindestens einmal unter die Konvergenzschwelle von $\frac{1}{1000}$ Pixel pro (Halb-)Bild sinkt, bei wievielen Punkten und wie schnell diese Schwelle in darauf folgenden Iterationen wieder überschritten wird sowie ob und wie schnell die Schwelle dann wieder (endgültig) unterschritten wird. Für den Fall eines kleinen Skalierungsbereiches und quasi-inverser Adaption wurde zusätzlich zu den üblichen Start-Kovarianzen auch die Kombination $\sigma_x = \sigma_y = 2,68, \sigma_t = 2,00$ getestet, da diese Kombination theoretisch auch als *Ergebnis* einer Adaption auftreten kann. Trotzdem zeigt diese Kombination keine Vorteile gegenüber den beiden üblichen Parameterpaaren.

in der nächsten Iteration wieder unter den Schwellwert und bleibt auch unter dem Schwellwert. Bei anderen Punkten wird die Schwelle nie (innerhalb der betrachteten 100 Iterationen) wieder unterschritten. Darunter sind auch solche Punkte, die zuvor den Schwellwert in mehr als 10 aufeinander folgenden Iterationen unterschritten haben. Abbildung 6.14 zeigt am Beispiel von sechs ausgewählten Punkten, wie sich die OF-

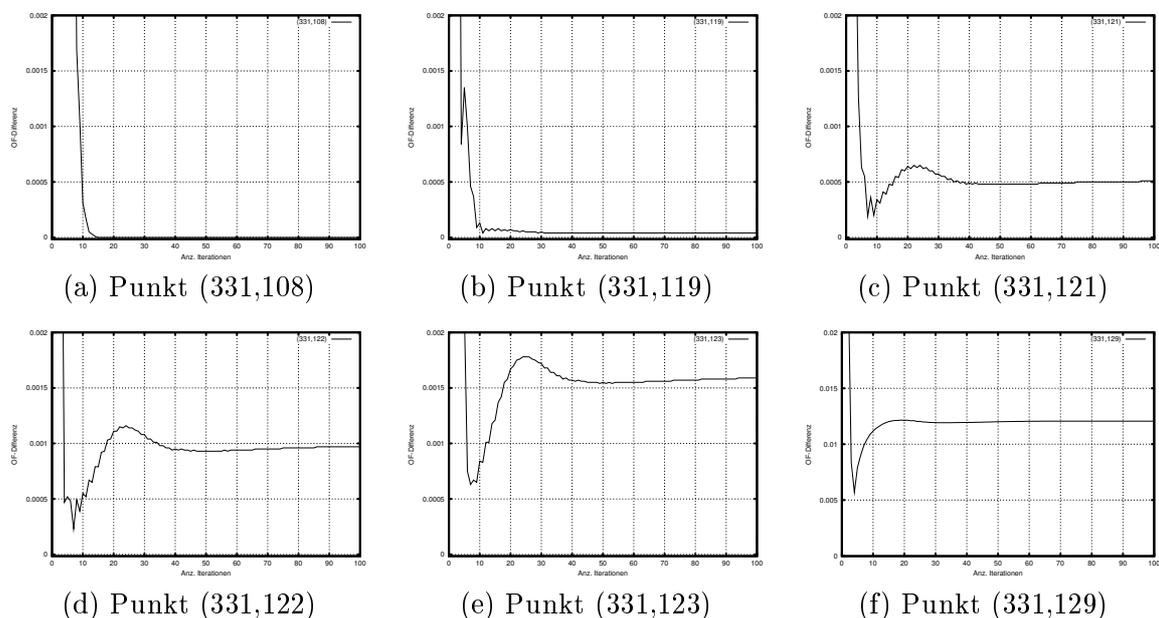


Abbildung 6.14: Mögliche Veränderung der OF-Differenzen im Verlauf der Iteration am Beispiel ausgewählter Punkte aus Bild 150 der Bildfolge `marmor_stat`: (a) Konvergenzschwelle wird unterschritten und nie wieder überschritten, OF-Differenz konvergiert gegen 0. (b) Konvergenzschwelle wird unterschritten, sofort wieder über- und dann endgültig unterschritten, OF-Differenz konvergiert gegen 0,00004. (c) Konvergenzschwelle wird unterschritten und nie wieder überschritten, aber OF-Differenz konvergiert nicht gegen 0. (d) Konvergenzschwelle wird (deutlich) unterschritten, nach mehreren Iterationen wieder über- und nach mehreren weiteren Iterationen endgültig wieder (knapp) unterschritten. (e) Konvergenzschwelle wird unterschritten, aber nach mehreren Iterationen (endgültig) wieder überschritten. (f) Konvergenzschwelle wird nie unterschritten. Die y -Achse hat im Vergleich zu den anderen Grafiken eine 10-mal größere Skalierung!

Differenz aufeinanderfolgender Iterationen entwickeln kann. Grafik (a) zeigt den Idealfall: Die Konvergenzschwelle wird nach wenigen Iterationen unterschritten, die OF-Differenz geht sehr schnell gegen 0. Dieses Verhalten wurde an den meisten Punkten festgestellt, an denen die Konvergenzschwelle unterschritten wird. Der Fall (b) ähnelt Fall (a), nur dass die Konvergenzschwelle kurzfristig wieder überschritten wird. Wie man in Grafik (c) sieht, kann es aber auch passieren, dass die Konvergenzschwelle zwar unterschritten wird, die OF-Differenz aber nicht gegen 0 geht. Dabei ist eine Art „Einschwingen“ der OF-Differenz feststellbar. Dieser Fall wird in Abschnitt 6.4.5.2 näher untersucht. Fall (d) verhält sich ähnlich wie (c), mit dem Unterschied, dass das „Einschwingen“ zu einem zwischenzeitlichen Wieder-Überschreiten der Schwelle führt und die OF-Differenz langfristig nur knapp unter der Konvergenzschwelle bleibt. Grafik (e) gleicht ebenfalls (c) und (d), nur dass die OF-Unterschiede insgesamt größer sind, was dazu führt, dass die Konvergenzschwelle nach dem Wieder-Überschreiten nicht wieder unterschritten wird. In Fall (f) schließlich treten noch einmal größere OF-Differenzen

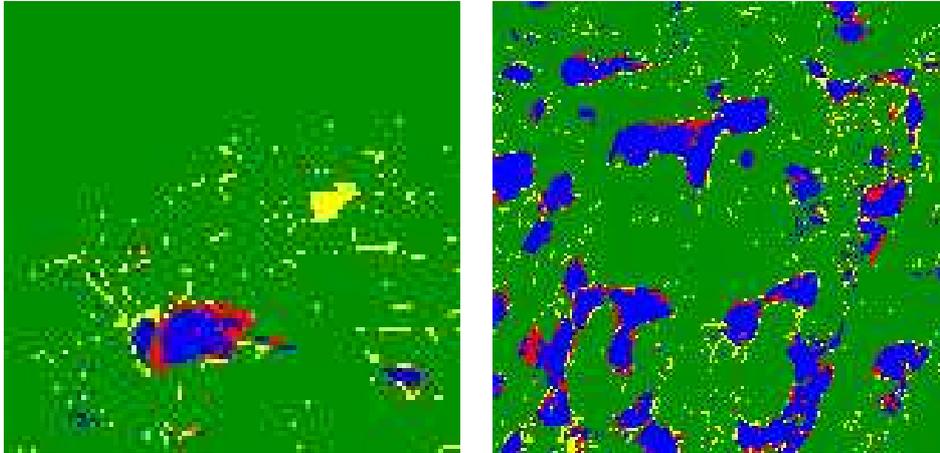


Abbildung 6.15: Lage der Pixel, die die Konvergenzschwelle nach zwischenzeitlichem Unterschreiten wieder überspringen, für Bild 150 der Bildfolge `marmor_stat` (links) und Halbbild 2540 der Bildfolge `kwB` (rechts). Farbcodierung: grün $\hat{=}$ Konvergenzschwelle wird nach einmaligem Unterschreiten nicht mehr überschritten; gelb $\hat{=}$ Konvergenzschwelle wird unterschritten, wieder überschritten, aber letztendlich unterschritten; rot $\hat{=}$ Konvergenzschwelle wird unterschritten, überschritten und nicht mehr unterschritten; blau $\hat{=}$ Konvergenzschwelle wird nie unterschritten.

auf (man beachte die veränderte Skalierung der Ordinate), so dass die Konvergenzschwelle nie unterschritten wird.

Interessant ist nun vor allem die Frage, wodurch sich die Punkte auszeichnen, bei denen trotz eines zwischenzeitlichen Unterschreitens der Schwelle diese wieder überschritten wird. Abbildung 6.15 illustriert die Lage dieser Punkte in jeweils einem Bild der Sequenzen `marmor_stat` und `kwB`. Für beide Bildfolgen gilt, dass die Ausreißer (genauso wie die Punkte, an denen die Konvergenzschwelle nie unterschritten wird) stets geballt auftreten. Es handelt sich also nicht um Zufallseffekte, sondern anscheinend um eine besondere Grauwert-Konstellation, die in Kombination mit einem noch nicht optimalen Adaptionsalgorithmus zu diesem unerwünschten Konvergenz-Verhalten führt. Man erkennt auch, dass diejenigen Punkte, an denen die Konvergenzschwelle unterschritten, dann wieder überschritten, letztendlich aber dauerhaft unterschritten wird (in gelb eingezeichnet) oft in keinem räumlichen Zusammenhang mit Nicht-Konvergenz (blau) stehen. Eine visuelle, stichprobenartige Inspektion der Daten ergab, dass an solchen Punkten meist die in Abbildung 6.14 (b) – seltener (d) – illustrierte Entwicklung vorliegt. Die Punkte, bei denen die Konvergenzschwelle kurzzeitig unter- aber letztendlich überschritten wird, liegen oft im Übergangsbereich zwischen Regionen mit gutem Konvergenzverhalten und Regionen ohne Konvergenz. Ein Beispiel dafür bilden die Punkte (331, 119) bis (331, 123) aus Bild 150 der Bildfolge `marmor_stat`, deren Konvergenzverhalten in Abbildung 6.14 (b) bis (e) illustriert ist: Während im Punkt (331, 119) (b) noch fast der Idealfall vorliegt, konvergiert die OF-Differenz in Punkt (331, 121) (c)

schon nicht mehr gegen 0. Bei Punkt (331, 122) (d) ist es eher zufällig, dass bei großer Iterationenzahl die Konvergenzschwelle knapp unterschritten wird, während am Punkt (331, 123) (e) die Konvergenzschwelle langfristig nicht erreicht wird.

Zusammenfassend kann man feststellen, dass „Ausreisser“ bisher entweder durch (für die Konvergenzaussage nicht störendes) Einschwingen verursacht wurden, oder im örtlichen Zusammenhang mit Punkten ohne Konvergenz auftreten (für die bereits eine Erklärung gegeben wurde).

Eine theoretische Erklärung, wie es dazu kommen kann, dass die Konvergenzschwelle z. T. in mehreren Iterationen unterschritten, dann aber wieder überschritten und vor allem (im Rahmen der untersuchten Iterationen) nie wieder unterschritten wird, findet sich in Abschnitt 6.5.5 (abstoßende Fixpunkte).

6.4.5.2 Zyklen ohne Überschreiten der Konvergenzschwelle

Als weiterer interessanter Fall wird untersucht, an wie vielen Punkten die OF-Schätzung nach Unterschreiten der Konvergenzschwelle wirklich konvergiert, bzw. es zu Zyklen oder Oszillationen ohne Überschreiten der Schwelle kommt (Abbildung 6.14 (b),(c)). Dazu wurde der Konvergenzschwellwert um eine Größenordnung verkleinert und das Konvergenzverhalten bei den um ein Pixel vergrößerten Masken⁷ untersucht (Tabelle 6.18). An mehr als 80 % der Punkte, an denen der ursprüngliche Schwellwert unterschritten wurde, wird auch der verschärfte Schwellwert von 0,0001 Pixel pro Bild (bzw. Halbbild) unterschritten, an den meisten übrigen Punkten kommt es zu Oszillationen. Es treten auch vermehrt größere Zyklen oder unregelmäßiges Verhalten auf.

Als Schlussfolgerung lässt sich ableiten, dass am Großteil der Punkte die OF-Schätzung tatsächlich (im mathematischen Sinne) konvergiert, oszilliert, oder die verbliebenen Fehler in einer Größenordnung liegen, die sich in der Anwendung nicht mehr auswirkt.

6.4.5.3 Verhalten nach Unterschreiten der Konvergenzschwelle

Wie die vorangehenden Auswertungen gezeigt haben, treten nach hinreichend großer Iterationszahl an den meisten Pixeln nur noch kleine Unterschiede der OF-Schätzung auf. Dabei ist es aber offensichtlich ein Unterschied bezüglich der Qualität der Ergebnisse, ob sich diese kleinen Änderungen in aufeinander folgenden Iterationen aufsummieren oder eher gegenseitig aufheben.

Es wurde daher untersucht, wie groß die durchschnittlichen Unterschiede in der OF-Schätzung zwischen der 100. Iteration und der 50., 80. sowie 96. bis 99. Iteration sind

⁷Bei einer tatsächlichen VR von etwa 1 bis 2 Pixel pro Bild (bzw. Halbbild) im Bereich der bewegten Marmorblöcke der Bildfolge `marmor_stat` und der Fahrzeugabbilder der Bildfolge `kwbb` entspricht die verkleinerte Konvergenzschwelle etwa 0,1‰ dieses Wertes. Es erschien nicht sinnvoll, an den meisten Punkten eine Schwankung des OF von weniger als 0,1‰ bei einem Maskenfehler von 1 % zu fordern, weshalb dieser Versuch nur mit den vergrößerten Masken durchgeführt wurde.

V-ID	40		41		42		43		
Adapt.-Verf.	linear		quasi-invers		linear		quasi-invers		
$\sigma_{\min}, \sigma_{\max}$	1,00; 2,85				1,00; 4,00				
σ_x, σ_y	1,00	1,41	1,00	1,41	1,00	1,41	1,00	1,41	ges.
σ_t	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	
marmor_stat, um 1 Pixel vergr. Masken, It. 100, Schwelle 0,0001									
Anz. Pixel gesamt	10000								80000
Diff. unter Konv.-Schwelle	8541	8487	9432	9424	7501	7482	8781	8773	68421
Oszillation (2er-Zyklus)	1222	1246	563	575	1402	1422	1214	1223	8867
3er Zyklus	2	1	0	0	24	4	1	1	33
4er Zyklus	215	67	1	0	2	3	0	0	288
5er Zyklus	0	0	0	0	1	1	0	0	2
6er Zyklus	0	0	0	0	165	119	0	1	285
7er Zyklus	0	0	0	0	0	5	0	0	5
8er Zyklus	0	51	0	0	0	0	0	0	51
> 8er oder kein Zyklus	20	148	4	1	905	964	4	2	2048
zum Vergleich: Schwelle 0,001									
Diff. unter Konv.-Schwelle	9138	9085	9696	9713	8555	8604	9459	9463	73713
Oszillation (2er-Zyklus)	774	802	302	287	828	835	539	535	4902
3er bis 8er Zyklus	85	90	0	0	232	105	1	1	514
> 8er oder kein Zyklus	3	23	2	0	385	456	1	1	871
kwbB, um 1 Pixel vergr. Masken, It. 100, Schwelle 0,0001									
Anz. Pixel gesamt	25600								204800
Diff. unter Konv.-Schwelle	18444	18446	17858	17745	15065	15264	12865	12928	128615
Oszillation (2er-Zyklus)	6557	6642	7610	7734	9603	9473	11766	11724	71109
3er Zyklus	27	16	0	0	24	31	1	1	100
4er Zyklus	18	23	46	10	202	295	377	492	1463
5er Zyklus	0	0	0	0	0	0	0	0	0
6er Zyklus	133	128	1	62	166	83	65	68	706
7er Zyklus	0	0	0	0	0	0	0	0	0
8er Zyklus	22	2	0	0	10	4	75	18	131
> 8er oder kein Zyklus	399	343	85	49	530	450	451	369	2676
zum Vergleich: Schwelle 0,001									
Diff. unter Konv.-Schwelle	21349	21353	21180	21038	19104	19055	17677	17639	158395
Oszillation (2er-Zyklus)	4092	4131	4390	4549	6311	6339	7682	7734	45228
3er bis 8er Zyklus	45	44	16	7	113	165	152	189	731
> 8er oder kein Zyklus	114	72	14	6	72	41	89	38	446

Tabelle 6.18: Oszillationen und Zyklen der OF-Schätzung bei verkleinertem Schwellwert, zum Vergleich jeweils die (verkürzten) Ergebnisse aus Tabellen 6.13 und 6.16: An den meisten Punkten wird auch der verschärfte Schwellwert unterschritten, es treten aber mehr Oszillationen, längere Zyklen oder unregelmäßiges Verhalten auf. Das beste Konvergenzverhalten zeigt sich bei quasi-inversem Adaptionsverfahren und kleinem Skalierungsbereich.

V-ID	40		41		42		43	
Adapt.-Verf.	linear		quasi-invers		linear		quasi-invers	
$\sigma_{\min}, \sigma_{\max}$	1,00; 2,85				1,00; 4,00			
σ_x, σ_y	1,00	1,41	1,00	1,41	1,00	1,41	1,00	1,41
σ_t	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

	Bildfolge marmor_stat, um 1 Pixel vergrößerte Masken							
99 → 100	7,236	7,443	1,012	1,021	5,583	7,282	1,669	1,630
98 → 100	0,069	0,061	0,001	0,000	2,937	1,348	0,001	0,002
97 → 100	7,206	7,307	1,011	1,021	6,349	6,374	1,668	1,629
96 → 100	0,003	0,032	0,001	0,000	7,973	3,807	0,002	0,002
80 → 100	0,017	0,033	0,005	0,001	3,559	2,481	0,009	0,009
50 → 100	0,098	0,086	0,026	0,012	1,079	1,663	0,029	0,020

	Bildfolge kwB, um 1 Pixel vergrößerte Masken							
99 → 100	9,399	9,520	12,484	12,186	24,102	23,442	34,097	35,087
98 → 100	0,063	0,057	0,004	0,003	0,035	0,044	0,188	0,097
97 → 100	9,424	9,504	12,482	12,186	23,947	23,427	34,150	35,104
96 → 100	0,080	0,058	0,004	0,003	0,030	0,026	0,028	0,022
80 → 100	0,144	0,120	0,040	0,018	0,077	0,238	0,091	0,063
50 → 100	0,254	0,237	0,246	0,116	0,228	0,527	0,761	0,532

Tabelle 6.19: Durchschnittliche Unterschiede in der OF-Schätzung zwischen der 100. Iteration und der 50., 80. sowie 96. bis 99. Iteration in 10^{-3} Pixeln pro Bild (bzw. Halbbild).

(Tabelle 6.19). Man erkennt, dass die Unterschiede zwischen 100. und 99. bzw. 97. Iteration vergleichsweise groß sind, was darauf zurückzuführen ist, dass bei Iterationen mit ungerader Differenz die beiden Zustände von Oszillationen miteinander verglichen werden. In den übrigen Fällen sind die Unterschiede durchweg gering, mit den besten Resultaten bei quasi-inverser Adaption mit kleinem Skalierungsbereich.

Diese Ergebnisse deuten darauf hin, dass auch in den Fällen, in denen die OF-Schätzung im mathematischen Sinne nicht konvergiert, dies nicht zu einer langfristigen Veränderung der OF-Schätzung führt.

6.4.5.4 Konvergenzkriterium

In den vorangehenden Versuchen wurde stets die Konvergenz der OF-Schätzung betrachtet. Wie bereits festgestellt wurde, ist dies *nicht* äquivalent zur Konvergenz der GST-Schätzung. Abbildung 6.16 vergleicht daher das Konvergenzverhalten hinsichtlich der beiden Konvergenzkriterien:

- Euklid-Norm der Differenz der OF-Schätzungen mit Schwellwert 0,001 Pixel pro Bild (bzw. Halbbild), und

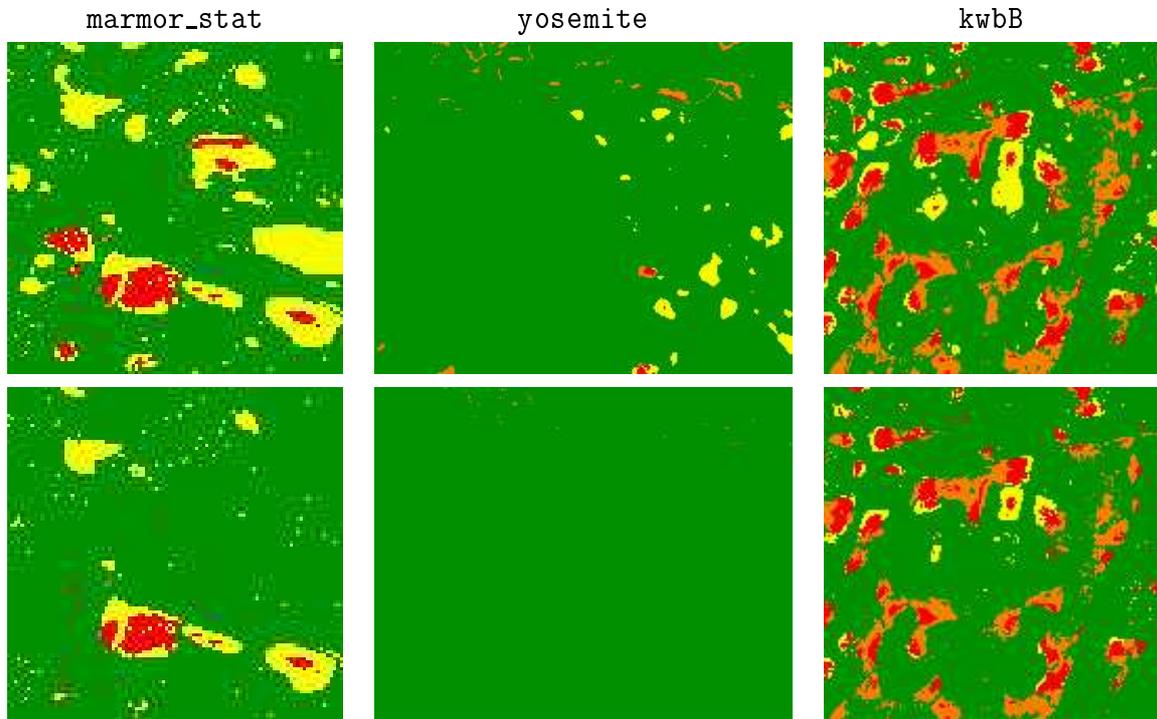


Abbildung 6.16: Frobenius-Norm der GST-Differenz vs. Euklid-Norm der OF-Differenz: Vergleich der Konvergenzkriterien in Iteration 10 (oben) und 30 (unten). An den meisten Pixel-Positionen konvergiert die GST-Schätzung entweder nach beiden Kriterien (grün) oder nach keinem der beiden (rot). In einer beschränkten Zahl von Fällen konvergiert die iterative GST-Schätzung nur gemäß des GST-Kriteriums (orange) bzw. des OF-Kriteriums (gelb). Quasi-inverse Adaption, große Start-Kovarianz, kleiner Skalierungsbereich.

- Frobenius-Norm der Differenz der GST-Schätzungen mit Schwellwert 0,03, was z. B. einem Unterschied von 0,01 in jeder GST-Komponente entspricht.

An den meisten Pixel-Positionen konvergiert die GST-Schätzung entweder nach beiden Kriterien oder nach keinem der beiden. Es gibt aber auch zahlreiche Punkte, die von den beiden Kriterien unterschiedlich bewertet werden:

- Bei der Bildfolge `kwbb` sind dies vor allem Punkte, an denen der GST-Unterschied unter die Konvergenzschwelle fällt, während der OF-Unterschied über der Konvergenzschwelle liegt. Diese Punkte liegen alle in Bereichen mit sehr wenig Grauwertstruktur. Die GSTs selbst haben dort eine kleine Frobenius-Norm, wodurch kleine Unterschiede (unterhalb der Schwelle) zu großen Unterschieden (oberhalb der Schwelle) bei der Eigenwert-/Eigenvektor- und damit der OF-Bestimmung führen können.
- Bei der Bildfolge `marmor_stat` treten hingegen mehrheitlich Fälle auf, bei denen die OF-Schätzung konvergiert, der GST aber nicht. An allen daraufhin visuell

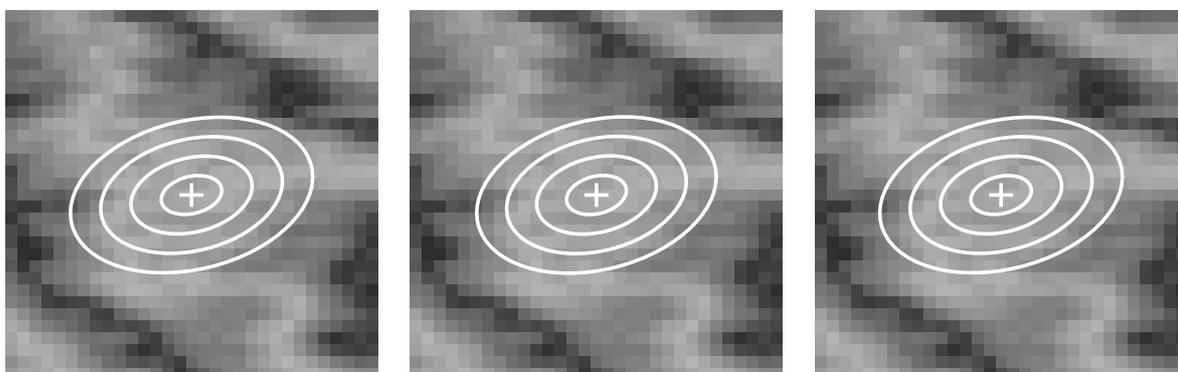


Abbildung 6.17: [marmor_stat, Bild 150] Ausdehnung der dreidimensionalen Mittelungsmasken am Punkt (319, 73) für Iterationen 17 (links), 18 (Mitte) und 19 (rechts). Jede Elipse entspricht einer Standardabweichung. Eine kaum zu erkennende Veränderung der Masken führt zu einer so großen Änderung des GST, dass die Unterschiede den Konvergenzschwellwert überschreiten. Parameter: „quasi-inverse“ Adaption, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$, $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$.

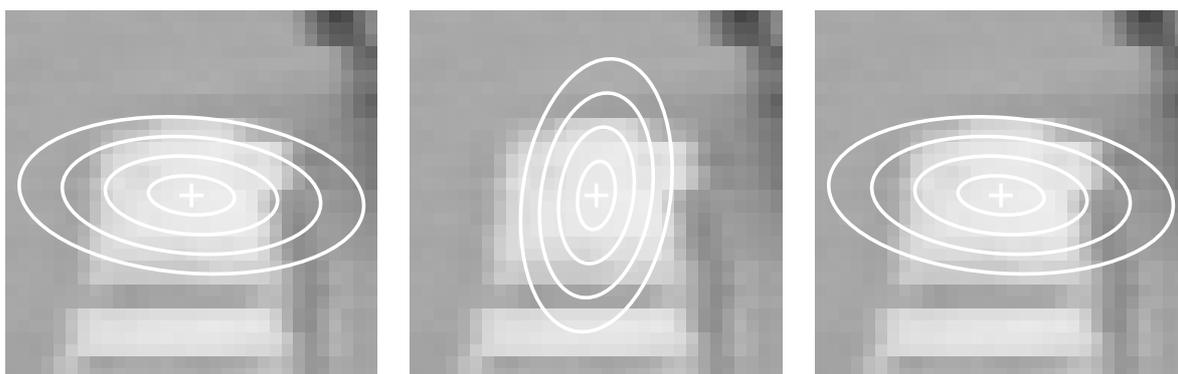


Abbildung 6.18: [kwbB, Bild 2540] Ausdehnung der dreidimensionalen Mittelungsmasken am Punkt (229, 295) für Iterationen 17 (links), 18 (Mitte) und 19 (rechts). Parameter: „quasi-inverse“ Adaption, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$, $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$.

inspizierten Fällen lag der Fall vor, dass sich die Eigenvektoren zum mittleren und größten Eigenwert stark änderten, während der Eigenvektor zum kleinsten Eigenwert (und damit der OF) nahezu konstant blieben. Der GST drehte sich hier also um den Eigenvektor zum kleinsten Eigenwert, und zwar entweder in einem nahezu homogenen Bereich (vgl. Abbildung 6.17) oder in der Nähe einer starken Grauwertkante (vgl. Abbildung 6.18).

Als Fazit lässt sich festhalten, dass verschiedene Konvergenzkriterien zwar im Detail zu unterschiedlichen Konvergenzaussagen führen können, aber in beiden Fälle die Nicht-Konvergenz auf ähnliche Grauwertstrukturen zurückgeführt werden kann.

6.4.5.5 Fazit

Es hat sich in der Praxis bewährt, von Konvergenz auszugehen, wenn die Konvergenzschwelle einmal unterschritten wird. In den meisten Fällen nehmen die Unterschiede der OF- und GST-Schätzung von Iteration zu Iteration dann auch tatsächlich weiter ab. Welches Konvergenz-Kriterium (GST-Differenz oder OF-Differenz) verwendet wird, sollte in der Praxis von der Zielstellung der jeweils durchgeführten Untersuchung abhängen, mit der Kenntnis, dass gegenseitige Rückschlüsse nur möglich sind, wenn die klar identifizierten Problemfälle ausgeschlossen wurden.

6.5 Vertiefende Untersuchungen zur iterativen GST-Bestimmung

Nach dieser experimentellen Voruntersuchung soll nun die Theorie der adaptiven GST-Bestimmung näher untersucht werden, um darauf aufbauend ein im Sinne der Konvergenz robusteres Verfahren zu entwickeln. Die iterative Bestimmung des GST ist ein Prozess, bei dem ausgehend von einer Start-Kovarianzmatrix abwechselnd ein Grauwertstrukturtensor GST und daraus eine Kovarianzmatrix für die darauffolgende Iteration bestimmt werden. Das bisherige Vorgehen setzt implizit voraus, dass die iterative GST-Bestimmung gegen einen Fixpunkt konvergiert. In diesem Abschnitt wird

1. bewiesen, dass ein solcher Fixpunkt bei minimalen Voraussetzungen an die Adaptionfunktion existiert,
2. demonstriert, dass ein solcher Fixpunkt in manchen Fällen *nicht* durch wiederholte Adaption gefunden werden kann,
3. an zwei Beispielen gezeigt, wie Konvergenz der GST-Iteration erreicht werden kann, und
4. untersucht, ob sich vor Beginn der Iterationen Anzeichen für im Sinne der Konvergenz kritische Punkte erkennen lassen,
5. sowie anhand einiger Bemerkungen aufgezeigt, wie sich einige auf den ersten Blick unerwartete Beobachtungen der vorausgegangenen Versuche mit Hilfe der Theorie erklären lassen.

6.5.1 Beweis der Fixpunkt-Existenz

Die Existenz des Fixpunktes wird zunächst für das *einfach-adaptive* Verfahren (vgl. Seite 110) bewiesen, da sich bei diesem Verfahren die GST-Schätzungen benachbarter Punkte nicht gegenseitig beeinflussen, und somit jeder Punkt für sich betrachtet werden

kann. Wie sich abschließend zeigen wird, ist eine Verallgemeinerung auf das doppelt-adaptive Verfahren problemlos möglich.

Zur Vereinfachung der Darstellung des nachfolgenden mathematischen Beweises sollen zunächst einige Bezeichnungen eingeführt werden: $g = g(\mathbf{x})$ sei die Grauwert-Funktion in Abhängigkeit von der räumlich-zeitlichen Bild-Position $\mathbf{x} = (x, y, t)^T$. Ferner seien \mathbf{x}_0 der Aufpunkt, für den die iterative GST-Bestimmung durchgeführt wird, $\mathcal{U}_M(\mathbf{x}_0)$ die räumlich-zeitliche Umgebung dieses Punktes, an denen lokale Tensoren $(\nabla g)(\nabla g)^T$ vorliegen müssen, $\mathcal{U}_A(\mathbf{x})$ die Umgebung des Punktes \mathbf{x} , die bei der Ableitungsbestimmung berücksichtigt wird, und $\mathcal{U}_{\text{GST}}(\mathbf{x}_0)$ die Umgebung des Aufpunktes, die bei der GST-Bestimmung insgesamt – inklusive Ableitungsbestimmung – berücksichtigt wird. Dabei gilt:

$$\mathcal{U}_{\text{GST}}(\mathbf{x}_0) = \bigcup_{\mathbf{x} \in \mathcal{U}_M(\mathbf{x}_0)} \mathcal{U}_A(\mathbf{x}). \quad (6.16)$$

$\Sigma_{A,i}(\mathbf{x})$ und $\Sigma_{M,i}(\mathbf{x})$ bezeichnen die bei der Ableitungsbestimmung bzw. Tensor-Mittlung in der i -ten Iteration am Punkt \mathbf{x} verwendete Kovarianzmatrix der zugrundeliegenden Gaußlocke.

Beim einfach-adaptiven Verfahren gilt für alle Iterationen $\Sigma_{M,i}(\mathbf{x}_0) = c \cdot \Sigma_{A,i}(\mathbf{x}_0)$ mit fest gewähltem $c > 0$ – meist sogar $c \geq 1$ – sowie $\Sigma_{A,i}(\mathbf{x}) = \Sigma_{A,i}(\mathbf{x}_0)$ für alle $\mathbf{x} \in \mathcal{U}_M(\mathbf{x}_0)$. Wir können also vereinfachend Σ_i anstelle von $\Sigma_{A,i}(\mathbf{x})$ schreiben.

Die Standardabweichungen der adaptierten Gaußlocke in Hauptachsenrichtung liegen im „Skalierungsbereich“ $S = [\sigma_{\min}, \sigma_{\max}]$. Die Menge \mathcal{M}_Σ der möglichen Kovarianzmatrizen ist eine Teilmenge des $\mathbb{R}^{3 \times 3}$ und ergibt sich durch Rotation von Diagonalmatrizen, deren Koeffizienten durch den Skalierungsbereich beschränkt sind:

$$\mathcal{M}_\Sigma = \left\{ R \cdot \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2) \cdot R^T \mid RR^T = I, \det(R) = 1, \sigma_1, \sigma_2, \sigma_3 \in [\sigma_{\min}, \sigma_{\max}] \right\}. \quad (6.17)$$

Ferner sei \mathcal{M}_{GST} die Menge der möglichen GST. \mathcal{M}_{GST} hängt vom Wertebereich der Grauwertfunktion, von \mathcal{M}_Σ sowie von Implementierungsdetails ab und ist eine Teilmenge des $\mathbb{R}^{3 \times 3}$.

Schließlich seien $n = |\mathcal{U}_{\text{GST}}(\mathbf{x}_0)|$ die Anzahl der Punkte in $\mathcal{U}_{\text{GST}}(\mathbf{x}_0)$ und $g|_{\mathcal{U}_{\text{GST}}(\mathbf{x}_0)}$ die Einschränkung von g auf $\mathcal{U}_{\text{GST}}(\mathbf{x}_0)$, also ein n -dimensionaler Vektor.

Das Adaptionverfahren lässt sich dann wie folgt darstellen:

1. **Startschritt:** Wähle $\Sigma_0 = \Sigma_{\text{init}} \in \mathcal{M}_\Sigma$.
2. **Bestimmung des GST:** Der GST der i -ten Iteration wird in Abhängigkeit vom Aufpunkt \mathbf{x}_0 , von den Grauwerten in der Umgebung des Aufpunktes $g|_{\mathcal{U}_{\text{GST}}(\mathbf{x}_0)}$ sowie von der Kovarianzmatrix Σ_i mit Hilfe der *GST-Schätzfunktion* f_S bestimmt:

$$f_S : (\mathbb{R}^3, \mathcal{M}_\Sigma, \mathbb{R}^n) \rightarrow \mathcal{M}_{\text{GST}}, \quad \text{GST}_i = f_S(\mathbf{x}_0, \Sigma_i, g|_{\mathcal{U}_{\text{GST}}(\mathbf{x}_0)}). \quad (6.18)$$

\mathbf{x}_0 und g sind für alle Iterationen konstant. Vereinfachend kann also geschrieben werden:

$$f_S : \mathcal{M}_\Sigma \rightarrow \mathcal{M}_{\text{GST}}, \quad \text{GST}_i = f_S(\Sigma_i). \quad (6.19)$$

Die Schätzfunktion f_S ist offensichtlich stetig in Σ_i , wenn $\mathcal{U}_{\text{GST}}(\mathbf{x}_0)$ und $\mathcal{U}_A(\mathbf{x})$ jeweils über alle Iterationen gleich bleiben.⁸

3. **Bestimmung der Kovarianzmatrizen:** Die Kovarianzmatrix Σ_{i+1} wird mit Hilfe der *Adaptionsfunktion* f_A aus GST_i bestimmt:

$$f_A : \mathcal{M}_{\text{GST}} \rightarrow \mathcal{M}_\Sigma, \quad \Sigma_{i+1} = f_A(\text{GST}_i). \quad (6.20)$$

4. **Abbruch bei Konvergenz,** sonst zurück zu Schritt 2.

Ein vollständiger Iterationsschritt besteht also in der Hintereinanderausführung von GST-Schätzfunktion und Adaptionfunktion $f_A \circ f_S$ oder umgekehrt $f_S \circ f_A$. Die Hintereinanderausführung $f_A \circ f_S$ wird im Folgenden als *Skalen-Iterationsfunktion* f_{SI} bezeichnet, die Hintereinanderausführung $f_S \circ f_A$ als *Tensor-Iterationsfunktion* f_{TI} . Fordert man nun die *Stetigkeit* der Adaptionfunktion f_A , so sind sowohl $f_{SI} = f_A \circ f_S$ als auch $f_{TI} = f_S \circ f_A$ stetige Abbildungen von \mathcal{M}_Σ bzw. \mathcal{M}_{GST} in sich selbst. Die jeweiligen Definitionsbereiche \mathcal{M}_Σ bzw. \mathcal{M}_{GST} sind als abgeschlossene^{9,10}, einfach zusammenhängende Teilmengen des $\mathbb{R}^{3 \times 3}$ homöomorph zur abgeschlossenen Einheitskugel des $\mathbb{R}^{3 \times 3}$. Nach einer Verallgemeinerung des Fixpunktsatzes von Brouwer (vgl. z. B. [Heuser 00], Sätze 228.1 und 229.2) besitzen somit beide Abbildungen mindestens einen Fixpunkt \hat{f}_{SI} bzw. \hat{f}_{TI} . Da sich beide Fixpunkte auseinander berechnen lassen ($\hat{f}_{SI} = f_A(\hat{f}_{TI})$, $\hat{f}_{TI} = f_S(\hat{f}_{SI})$), ist im Folgenden oft nur von *dem* Fixpunkt der Iteration die Rede.

Beim *doppelt-adaptiven* Verfahren erfolgt der Konvergenzbeweis ähnlich. Da sich bei diesem Verfahren die GST-Schätzungen benachbarter Punkte gegenseitig beeinflussen,

⁸Würde man hingegen $\mathcal{U}_{\text{GST}}(\mathbf{x}_0)$ und $\mathcal{U}_A(\mathbf{x})$ in jeder Iteration in Abhängigkeit von Σ_i neu festlegen, so würde bei neu hinzukommenden Punkten deren Gewichtung *sprunghaft* von 0 auf einen Wert $w > \varepsilon > 0$ zunehmen und bei nicht mehr berücksichtigten Punkten ebenso *sprunghaft* auf 0 zurückgehen. Die Gewichtungsfunktionen bei Ableitung und Mittelung und somit auch f_S insgesamt wären nicht mehr stetig.

⁹Beweis der Abgeschlossenheit für \mathcal{M}_Σ : Sei $(M_i)_{i \in \mathbb{N}}$ eine Folge von Matrizen $M_i = (m_{ijk})_{j,k=1,2,3} \in \mathcal{M}_\Sigma$, die gegen $M = (m_{jk})_{j,k=1,2,3}$ konvergiert. D. h. $\forall j, k : m_{ijk} \rightarrow m_{jk}$ für $i \rightarrow \infty$. Dann ist M symmetrisch, da alle M_i symmetrisch sind. Nach dem Satz über die Symmetrische Schur-Zerlegung (vgl. [Golub & van Loan 96, Seite 393]) lässt sich M also schreiben als $M = R \cdot \text{diag}(\mu_1, \mu_2, \mu_3) \cdot R^T$ mit Rotationsmatrix R . Falls $\det(R) = -1$, so verwendet man $-R$ statt R . Zu zeigen ist noch, dass $\mu_1, \mu_2, \mu_3 \in [\sigma_{\min}^2, \sigma_{\max}^2]$. Es wird o. B. d. A. gezeigt, dass $\mu_1 \leq \sigma_{\max}^2$: Angenommen, $\mu_1 > \sigma_{\max}^2$. D. h. $\exists \varepsilon > 0 : \mu_1 > \sigma_{\max}^2 + 2\varepsilon$. Wegen Konvergenz gilt aber auch: $\exists n \in \mathbb{N} : \forall i > n : \|M - M_i\|_2 < \varepsilon$. Für ein beliebiges $i > n$ lässt sich M_i schreiben als $M_i = R_i \cdot \text{diag}(\lambda_1, \lambda_2, \lambda_3) \cdot R_i^T$ mit Rotationsmatrix R_i . Für die Konditionszahl κ dieser Rotationsmatrix R_i gilt: $\kappa_2(R_i) = \|R_i\|_2 \cdot \|R_i^{-1}\|_2 = 1$. Dann gilt nach dem Theorem von Bauer-Fike (vgl. [Golub & van Loan 96, Seite 321]): $\min_{\lambda \in \{\lambda_1, \lambda_2, \lambda_3\}} |\lambda - \mu_1| \leq \kappa_2(R_i) \|M - M_i\|_2 < \varepsilon$. D. h. $\lambda_1 > \sigma_{\max}^2 \vee \lambda_2 > \sigma_{\max}^2 \vee \lambda_3 > \sigma_{\max}^2$, was im Widerspruch zur Konstruktion von \mathcal{M}_Σ steht. Daraus folgt, dass $M \in \mathcal{M}_\Sigma$, also ist \mathcal{M}_Σ abgeschlossen.

¹⁰Beweis der Abgeschlossenheit für \mathcal{M}_{GST} : analog zu \mathcal{M}_Σ .

müssen aber auch alle Punkte gleichzeitig betrachtet werden. Wird die doppelt-adaptive Iteration auf einem n Pixel umfassenden Bildausschnitt durchgeführt, so definiert man \mathcal{M}_Σ als Menge aller möglichen n -Tupel von Kovarianzmatrizen, \mathcal{M}_{GST} als Menge aller möglichen n -Tupel von GST und f_S und f_A entsprechend erweitert. Die Adaption auf einem höher-dimensionalen Raum ändert aber nichts an der Anwendbarkeit des Satzes von Brouwer.

6.5.2 Praktische Probleme

Wie die Untersuchungen in Abschnitt 6.4 gezeigt haben, konvergiert die iterative Adaption nicht an allen Punkten. Als eine mögliche Ursache wurde – wie in Abbildung 6.9 illustriert – die durch Formel 6.11 bedingte Koppelung der Maskenausdehnungen in den *verschiedenen* Hauptachsenrichtungen identifiziert. Abgesehen von diesen praktischen Problemen ist es darüber hinaus auch theoretisch anzuzweifeln, ob es – als Effekt der Koppelung – sinnvoll ist, wenn eine sich in *einer* Richtung verändernde Grauwertstruktur *senkrecht dazu* zu einer Änderung der Skala führt. Durch die Koppelung wird zwar eine Invarianz gegenüber einer globalen Skalierung der Grauwerte erreicht, völlig unterschiedliche lokale Grauwertstrukturen (z. B. sanfte und steile Übergänge) können aber so zur selben Kovarianzmatrix führen.

Es liegt daher nahe, den Adaptionsschritt so anzupassen, dass die Maskenausdehnung in Richtung einer der Hauptachsen nur von dem zu dieser Hauptachse gehörenden Eigenwert des GST abhängt. Die beiden bisher verwendeten Adaptionsverfahren „lineare Adaption“ und „quasi-inverse Adaption“ setzen aber normierte Eigenwerte im Intervall $[0, 1]$ voraus, es müsste also ein neues Adaptionsverfahren entwickelt werden. In diesem Abschnitt werden – noch ohne ein konkretes Vorgehen zu benennen – allgemeine Eigenschaften einer solch „entkoppelten“ Adaption untersucht. Die Entkopplung der Maskenskalierung ermöglicht dabei eine Analyse des Problems im Eindimensionalen. Wie sich zeigen wird, kann es aber auch in diesem Fall bei naheliegenden Annahmen über die Adaptionsfunktion zu Nicht-Konvergenz kommen.

O. B. d. A. wird dazu angenommen, dass sich die Grauwertstruktur in der Umgebung des betrachteten Punktes $\mathbf{x}_0 = (x_0, y_0, t_0)^T$ nur in x -Richtung ändert, in y - und t -Richtung aber konstant bleibt. Unter dieser Annahme sind alle partiellen Ableitungen in y - und t -Richtung 0. Der GST am Punkt \mathbf{x}_0 hat also die Form:

$$\text{GST}(\mathbf{x}_0) = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (6.21)$$

Zunächst wird untersucht, welche Werte λ überhaupt annehmen kann:

Der *kleinstmögliche Wert* von λ tritt auf, wenn die Grauwerte im gesamten in die GST-Bestimmung einbezogenen Bereich konstant sind. Es gilt dann $\lambda = 0$.

Der *größtmögliche Wert* λ_{\max} von λ tritt auf, wenn \mathbf{x}_0 genau an einem Grauwertsprung vom kleinsten auf den größten Wert des Definitionsbereichs der Grauwerte liegt. Der in einem solchen Fall auftretenden Wert von λ soll nun abgeschätzt werden.

Unter den obigen Voraussetzungen ergibt sich die einzige von 0 verschiedene Komponente des GST durch eine gewichtete Mittelung über die Quadrate der x -Ableitungen. σ_A bezeichne die Standardabweichung der zur Ermittlung der x -Ableitungsmaske herangezogenen eindimensionalen Gaußglocke, analog σ_M für die Mittelungsmaske.

Das Mittelungsergebnis wird durch mehrere – zumindest teilweise gegeneinander wirkende – Effekte beeinflusst. Bei wachsendem σ_A

- werden die Ableitungswerte in unmittelbarer Nähe des Grauwertsprunges aufgrund der impliziten Glättung kleiner,
- nimmt hingegen in einiger Entfernung zum Grauwertsprung der Einfluss des Grauwertsprunges – ebenfalls aufgrund der impliziten Glättungswirkung – zu, die Ableitungswerte werden größer,
- wird im Mittelungsschritt über einen größeren Bereich gemittelt, wenn – wie bisher immer – σ_M linear von σ_A abhängt, mit sinkendem Gewicht in der Nähe des Grauwertsprunges und mit zunehmenden Gewicht in einiger Entfernung dazu.

Folgende (eindimensionale) Betrachtung soll die genauen Zusammenhänge aufdecken. Dabei wird die gewichtete Mittelung der Ableitungsquadrate in Abhängigkeit von σ_A und σ_M an der Stelle 0 bestimmt, wenn an der Stelle 0 ein Grauwertsprung von 0 auf 1 vorliegt, d. h. alle Grauwerte $g(x) = g(x, y, t)$ sind 0 für $x < 0$ und 1 für $x \geq 0$. Vereinfachend werden dabei die aufgrund der Rasterung auftretenden Summen durch die entsprechenden Integrale angenähert. Zur besseren Lesbarkeit gelte außerdem $\sigma_A = \sigma$ und $\sigma_M = k\sigma$.

Für die mit Hilfe der abgeleiteten Gaußfunktion bestimmten Ableitungen an der Stelle x gilt:

$$g_x(x, \sigma) = \int_{-\infty}^{+x} -\frac{\xi}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \quad (6.22)$$

Die Mittelung ergibt dann:

$$\overline{g_x^2}(0, \sigma, k) = \int_{-\infty}^{\infty} \frac{(g_x(x, \sigma))^2}{\sqrt{2\pi}k\sigma} e^{-\frac{x^2}{2(k\sigma)^2}} dx \quad (6.23a)$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \right)^2 \frac{1}{\sqrt{2\pi}k\sigma} e^{-\frac{x^2}{2(k\sigma)^2}} dx \quad (6.23b)$$

$$= \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{2\pi}k\sigma} \int_{-\infty}^{\infty} e^{-\frac{2x^2}{2\sigma^2}} e^{-\frac{x^2}{2(k\sigma)^2}} dx \quad (6.23c)$$

$$= \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{2\pi}k\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{2k^2+1}{k^2\sigma^2}\right)x^2} dx \quad (6.23d)$$

$$= \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{2\pi}k\sigma} \frac{\sqrt{2\pi}}{\sqrt{\frac{2k^2+1}{k^2\sigma^2}}} \quad (6.23e)$$

$$= \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{2\pi}k\sigma} \frac{\sqrt{2\pi}k\sigma}{\sqrt{2k^2+1}} \quad (6.23f)$$

$$= \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{2k^2+1}}. \quad (6.23g)$$

Ist g_{\min} der kleinste und g_{\max} der größte Grauwert, so beträgt der größtmögliche Wert von λ also näherungsweise:

$$\lambda \approx (g_{\max} - g_{\min})^2 \cdot \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{2k^2+1}}. \quad (6.24)$$

Für bisher häufig benutzte Kombinationen von σ_A und σ_M ergeben sich folgende theoretische Maximalwerte von λ :

$g_{\max} - g_{\min}$	σ_A	σ_M	σ	k	$\lambda_{\max} = \max(\lambda)$
255	1,0	1,0	1,0	1,0	5975
255	1,0	$\sqrt{2}$	1,0	$\sqrt{2}$	4628
255	$\sqrt{2}$	2	$\sqrt{2}$	$\sqrt{2}$	2314

Ein Vergleich der tatsächlichen Werte mit den theoretisch erwarteten Werten in [Middendorf 02] zeigt eine gute Übereinstimmung. Die Abweichungen sind auf Fehler durch die Diskretisierung, Beschränkung der Maskengröße und Rechenungenauigkeiten zurückzuführen.

Mit wachsendem σ oder k , also mit wachsendem σ_A oder σ_M nimmt der größtmögliche Wert von λ ab. Setzt man wie bisher $\sigma_{\min} = 1,0$, so können also Eigenwerte des GST zwischen 0 und fast 6000 auftreten. Tabelle 6.20 und Abbildung 6.19 zeigen aber, dass die tatsächlich auftretenden Werte für einen Großteil der Bildpunkte nicht einmal ein Zehntel des Maximalwertes betragen.

Gesucht ist also eine Adaptionfunktion f_A , die Eigenwerte zwischen 0 und 6000 auf den zugelassenen *Skalierungsbereich* $S := [\sigma_{\min}, \sigma_{\max}]$ abbildet, wobei – wie bisher – kleine Eigenwerte auf eine große Skala und große Eigenwerte auf eine kleine Skala abgebildet werden sollen. Dabei ist zu fordern, dass die Adaptionfunktion stetig und monoton fallend ist.

Bildfolge (Bildnummer)	Quantile				
	5 %	25 %	50 %	75 %	95 %
<code>marmor_stat</code> (150)	7,5	31,5	66,4	112,3	199,9
<code>stau02</code> (1750)	0,8	7,0	19,3	53,6	217,0
<code>reading</code> (140)	0,4	1,2	4,2	38,8	299,4
<code>kwbB</code> (2660)	0,3	3,9	11,4	30,1	112,2
<code>tr_bsb</code> (35)	0,04	5,9	52,7	130,4	288,8

Tabelle 6.20: Tatsächlich auftretende Werte für den größten Eigenwert des GST bei verschiedenen Bildfolgen. Der zugrunde liegende GST wurde jeweils auf dem gesamten Bild mit $\Sigma_A = \text{diag}(2,0; 2,0; 2,0)$, $\Sigma_M = 2\Sigma_A$, abgetasteten Maskenkoeffizienten und vorgegebenem Maskenfehler von 1 % ermittelt. Bei allen untersuchten Bildern liegt ein Großteil der Werte deutlich unter dem theoretischen Maximum von rund 2314.

Wie gezeigt, besitzen Skalen- und Tensor-Iterationsfunktion jeweils mindestens einen Fixpunkt. Würde man nun zeigen, dass eine dieser Funktionen kontrahierend ist, so wäre nach dem bekannten Banach'schen Fixpunktsatz bewiesen, dass die Adaption für jeden Startwert σ_0 gegen diesen Fixpunkt iteriert (z. B. [Heuser 00]).

Wie aber folgendes Gedankenexperiment zeigt, kann es bei manchen Grauwertkonfigurationen und vernünftigen Annahmen über die Adaptionfunktion dazu kommen, dass die iterierte Adaption nicht konvergiert, sondern zur Oszillation führt. Angenommen, man möchte die „optimale“ Skala für einen Punkt bestimmen, der in jeweils sechs Pixeln Entfernung zwischen zwei Grauwertsprüngen $255 \rightarrow 0$ bzw. $0 \rightarrow 255$ liegt. Abbildung 6.20 skizziert eine solche Grauwertstruktur und zeigt, welche Werte die Mittelung λ über die Ableitungsquadrate an einem solchen Punkte in Abhängigkeit von σ im angenommenen Skalierungsbereich $[1; 4]$ annimmt. Startet man die Adaption mit σ_{init} mit dem Wert 1,0, so erhält man für λ den sehr kleinen Wert von 0,02, der die Wahl der maximalen Skala in der ersten Iteration rechtfertigt. Wie Abbildung 6.20 zeigt, nimmt λ für die maximale Skala ($\sigma = 4,0$) einen Wert von etwa 240 an. Im Vergleich zu den tatsächlich auftretenden größten Eigenwerten (vgl. Tabelle 6.20) ist dieser Wert so groß, dass er im zweiten Iterationsschritt wiederum die Wahl der minimalen Skala rechtfertigt. In diesem Fall käme es also zur Oszillation, egal, wie die Adaptionfunktion im Detail aussieht.

Dieses Gedankenexperiment verwendet eine sehr spezielle Grauwertstruktur, die in realen Bildfolgen wohl kaum auftreten wird, und einen ungünstigen Startwert σ_{init} am Rand des Skalierungsbereichs. Außerdem könnte die Oszillation durch eine (deutliche) Einschränkung des Skalierungsbereiches möglicherweise vermieden werden. Das Beispiel zeigt aber deutlich, dass nicht gleichzeitig erreicht werden kann, dass die Skaleniterationsfunktion in allen Fällen kontrahierend ist *und* dass die Adaptionfunktion einigen intuitiven Annahmen entspricht.

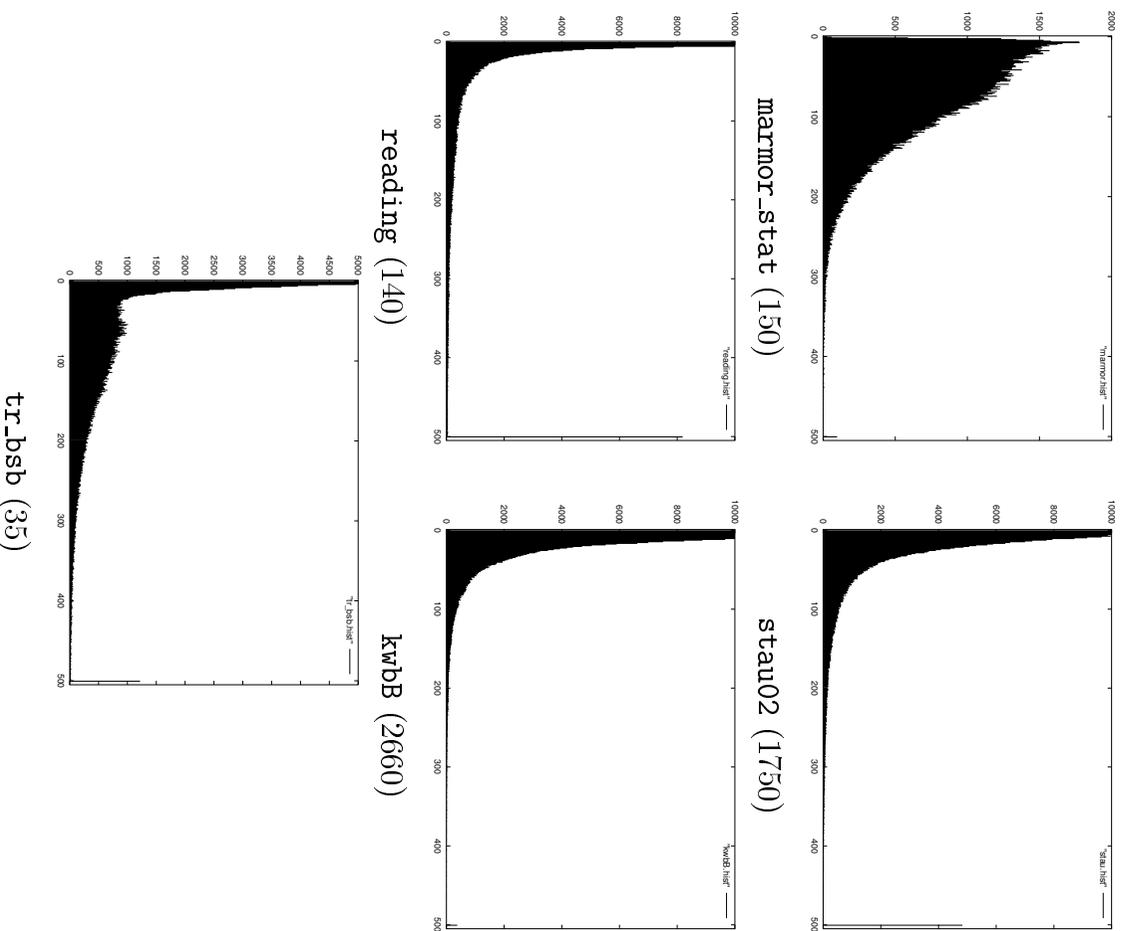


Abbildung 6.19: Wie in Tabelle 6.20, Eigenwerte im Bereich 0–500 als Histogramm mit Schrittweite 1 dargestellt. Der Histogramm-Wert für 501 entspricht der Summe aller auftretenden größeren Werte.

6.5.3 Verbesserung des Konvergenzverhaltens

Angesichts dieser Erkenntnis ergibt sich die Frage nach dem weiteren Vorgehen. Im Rahmen der hier vorgestellten Arbeit wurden drei Alternativen untersucht:

- Es wird ein Verfahren vorgestellt, das durch „gedämpfte Iteration“ für deutlich mehr Punkte als bisher durch iteratives Vorgehen *näherungsweise* einen Fixpunkt bestimmt, und zwar unabhängig von der gewählten Adaptionfunktion, wenn diese gewissen Grundanforderungen entspricht. Die Aussagekraft der so gefundenen

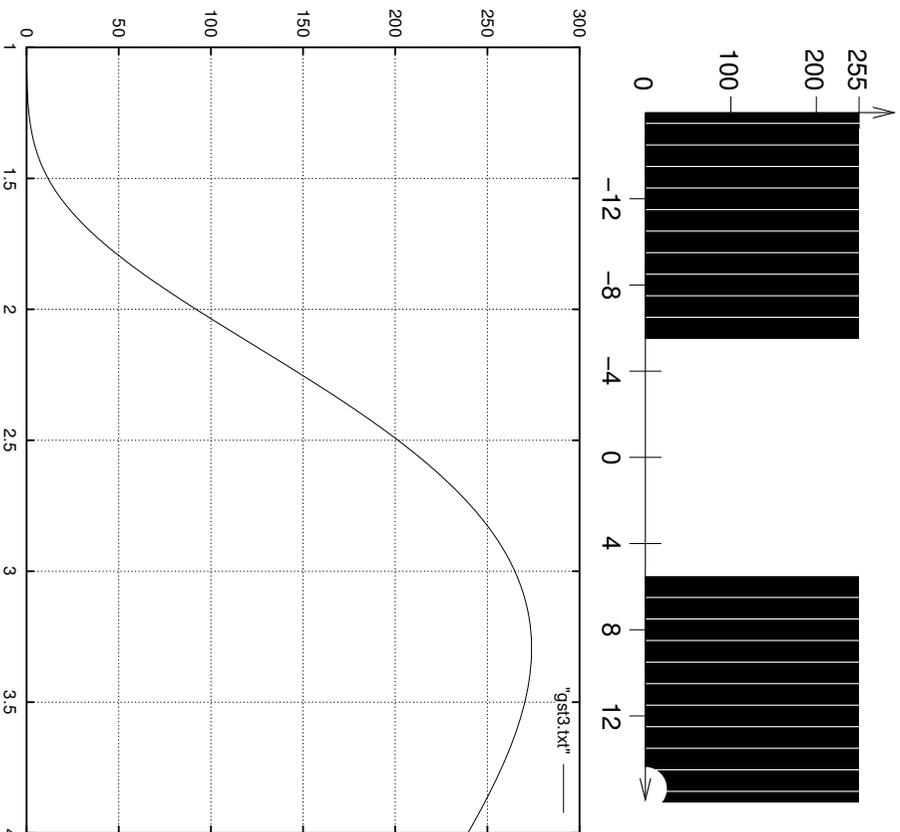


Abbildung 6.20: Ergebnis (oben) der gewichteten Mittelung über die Ableitungsquadrate in Abhängigkeit von σ in der Mitte zwischen zwei Grauwertsprüngen $255 \rightarrow 0$ bzw. $0 \rightarrow 255$ in x -Richtung zwischen Pixel -6 und -5 bzw. zwischen Pixel 5 und 6 (Skizze: oben) bei in y - und t -Richtung konstanten Grauwerten.

Fixpunkte wird diskutiert.

- Es wird eine neue (ungedämpfte) Adaptionfunktion vorgestellt, die deutlich besseres Konvergenzverhalten zeigt als die lineare oder quasi-inverse Adaption.
- Es wird versucht, die im Sinne der Konvergenz kritischen Punkte schon vor Beginn der Adaption zu finden. Bei Erfolg könnten solche Punkte von vornherein ausgesondert und einer speziellen Behandlung unterzogen werden.

6.5.3.1 Gedämpfte Iteration

In der Literatur (vgl. z. B. [Ortega & Rheinboldt 70], [Schwettlick 79], [Press et al. 90], [Acton 90]) sind mehrere Verfahren bekannt, wie ein Fixpunkt bei mehrdimensionalen

len, nicht-linearen Funktionen gefunden werden kann. Die meisten Standard-Verfahren setzen aber voraus, dass die untersuchte Funktion kontrahierend ist – was im hier vorliegenden Fall wie soeben gezeigt nicht zutrifft – oder dass alle partiellen Ableitungen der Funktion bestimmt oder zumindest angenähert werden können – was sehr aufwändig ist.

Betrachtet man die Fälle, in denen das bisherige Vorgehen zu Oszillationen führt (Abbildungen 6.7, 6.9, 6.18), so zeigt sich als eine Ursache der mangelnden Konvergenz, dass oft „über die Sichtweite hinaus“ adaptiert wird: Bei einer Verdoppelung oder gar Vervierfachung der Standardabweichung im Rahmen der Adaption erlangen Punkte wesentlichen Einfluss auf das Ergebnis, deren Gewicht zuvor nahe 0 war. Ein naheliegender Ansatz, der weder die Kontraktivität voraussetzt noch eine Ableitungsberechnung erfordert, ist eine „gedämpfte“ Iteration. Anstelle der Standarditeration:

$$x_{i+1} := f(x_i) \quad (6.25)$$

gelte dann:

$$x_{i+1} := (1 - \mu)f(x_i) + \mu x_i, \quad (6.26)$$

mit Dämpfungsfaktor μ , $0 < \mu < 1$. Das Problem besteht darin, ein geeignetes μ zu finden. Wird es zu klein gewählt, so tritt möglicherweise keine Konvergenz ein, wird es zu groß gewählt, so erfolgt die Konvergenz sehr langsam.

Die Fragen, welcher Wert von μ hinsichtlich möglichst schneller Konvergenz an möglichst allen Punkten optimal ist, ist dem Forschungsbereich der Numerischen Mathematik zuzuordnen. Anhand einiger Beispiele soll hier nur gezeigt werden, dass es grundsätzlich möglich ist, für *jeden* Punkt näherungsweise einen Fixpunkt der Tensor- oder Skaleniteration zu bestimmen.

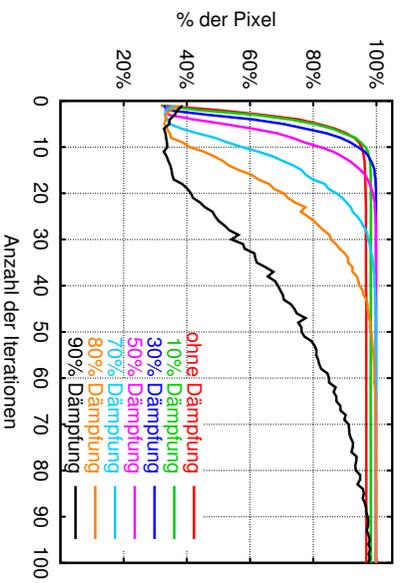
Dazu wurde eine Versuchsreihe durchgeführt, in der das Konvergenzverhalten sowohl der OF-Vektoren als auch der GST bei ungedämpfter Iteration sowie bei Dämpfungsfaktoren μ von 10 bis 90 Prozent (in 10%-Schritten) untersucht wird.

Beim Vergleich der Ergebnisse bei verschiedenen Werten für den Dämpfungsfaktor μ muss berücksichtigt werden, dass durch die Dämpfung die Unterschiede zwischen aufeinander folgenden Iterationen nicht so groß werden können wie bei ungedämpfter Iteration. So ist bei 50 % Dämpfung im Mittel nur mit etwa halb so großen Unterschieden von einer Iteration zur nächsten zu rechnen wie ohne Dämpfung, bei einem Dämpfungsfaktor μ nur mit $(1 - \mu)$ -mal so großen Unterschieden. Durch Skalierung mit dem Kehrwert des Verhältnisses des gedämpft zu erwartenden Unterschieds zum ungedämpften Fall muss die Vergleichbarkeit der mit verschiedenen Dämpfungsfaktoren erzielten Werte wieder hergestellt werden. Der „dämpfungsbereinigte“ Euklid-Abstand:

$$\hat{d}(\mathbf{u}_i(\mathbf{x}), \mathbf{u}_j(\mathbf{x})) := \frac{1}{1 - \mu} \|\mathbf{u}_i(\mathbf{x}) - \mathbf{u}_j(\mathbf{x})\|_2. \quad (6.27)$$

erfüllt diese Vorgabe und verhindert so eine irreleitende visuell bedingte Bevorzugung der Ergebnisse bei gedämpfter Iteration.

marmor_stat



krwb

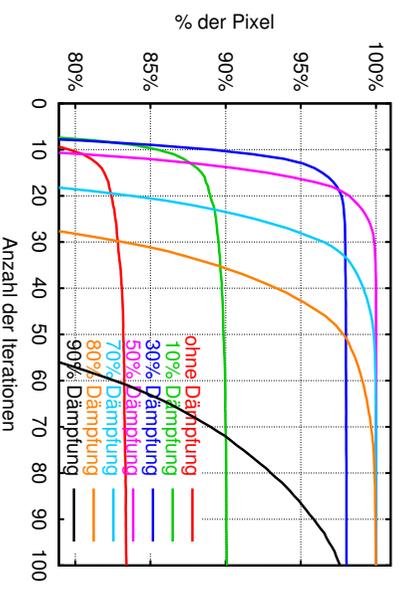
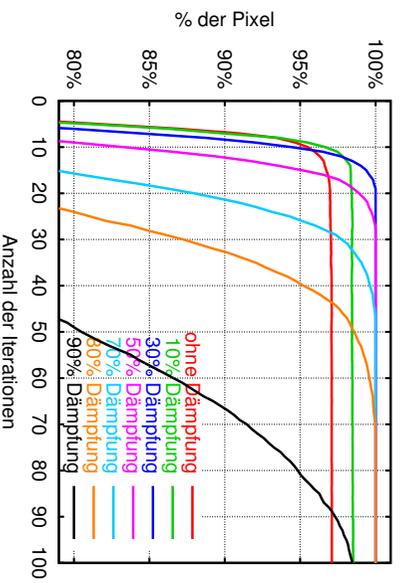
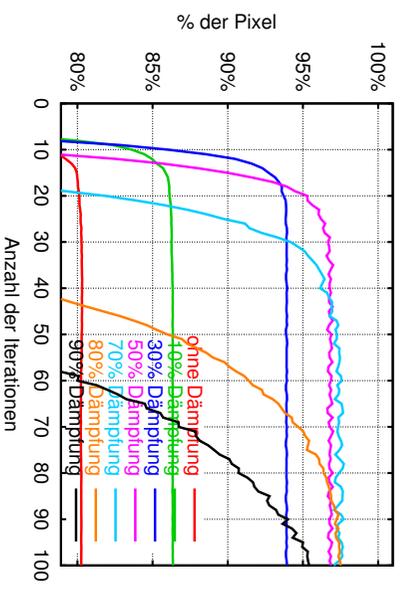
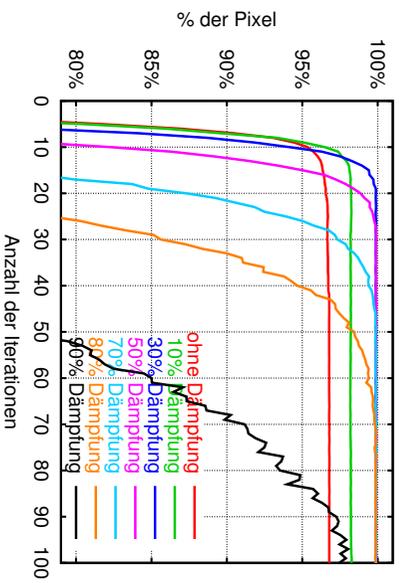
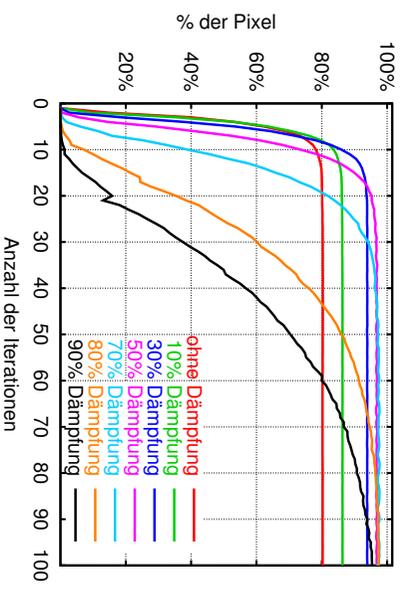


Abbildung 6.21: Konvergenz der OF-Schätzung bei gedämpfter Iteration: Die Grafiken zeigen, am wieviel Prozent der Punkte in Abhängigkeit von der Iterationszahl die Konvergenzschwelle von 0,001 Pixel pro Bild (bzw. Halbbild) unterschritten wird. Bei 1% Maskenfehler (1. Zeile: Überblick; 2. Zeile: Details) ergibt sich der beste Kompromiss zwischen hoher Konvergenzrate und schneller Konvergenz bei 30 bis 70 Prozent Dämpfung. Bei fester Maskengröße von 21^3 (3. Zeile) hingegen kommt es bei mehr als 50 Prozent Dämpfung an allen Punkten zu Konvergenz. Außerdem sind die Kurven deutlich glatter. Parameter: quasi-inverse Adaption, kleiner Skalierungsbereich und große Start-Kovarianz.

Die oberen beiden Zeilen von Abbildung 6.21 zeigen die Auswirkungen der gedämpften Iteration am Beispiel der quasi-inversen Adaption mit kleinem Skalierungsbereich und großer Start-Kovarianz bei 1 % Maskenfehler. Man erkennt, dass bereits eine Dämpfung von 10 Prozent zu einer signifikanten Steigerung der Konvergenzrate (=prozentualer Anteil der Pixel, an denen es zu Konvergenz kommt) führt, ohne dass die Konvergenz selber wesentlich langsamer erfolgt. Der beste Kompromiss zwischen hoher Konvergenzrate und schneller Konvergenz ergibt sich für 30 bis 70 Prozent Dämpfung. Höhere Dämpfungsraten von 80 oder 90 Prozent führen hingegen zu einer wesentlichen Verlangsamung der Konvergenz, ohne dass sich dies in einer erkennbar höheren Konvergenzrate niederschlägt.

Trotz der höheren Konvergenzrate führt die gedämpfte Iteration beim untersuchten Parameteratz auch bei 70 % Dämpfung oder mehr *nicht* für alle Pixel zur Konvergenz. Bei der Bildfolge `kwbb` ist dies augenscheinlich, bei der Bildfolge `marmor_stat` sind es aber – in der Grafik nicht zu erkennen – auch ca. 50 Pixel. Eine Erhöhung der Iterationszahl für 80 % und 90 % Dämpfung – in diesen beiden Fällen ist die Einschwingphase nach 100 Iterationen zumindest bei der Bildfolge `kwbb` noch nicht abgeschlossen – führt letztendlich auch zu keinem besseren Ergebnis als bei 70 % Dämpfung.

Die Erklärung für dieses Problem liefert eine Rückbesinnung auf die theoretischen Überlegungen in Abschnitt 6.5.1. Dort wurde gezeigt, dass ein Fixpunkt existiert, wenn die Iterationsfunktion (unter anderem) stetig ist. Bei den hier durchgeführten Versuchen wird die Stetigkeit aber dadurch verletzt, dass in jeder Iteration die Maskengröße unter Verwendung des vorgegebenen Maskenfehlers von 1 % neu bestimmt wird (vgl. auch Fußnote⁸ auf Seite 154). Verwendet man daher eine feste Maskengröße, die für alle möglichen Kovarianzmatrizen ausreichend ist, also z. B. $21 \times 21 \times 21$ beim kleinen Skalierungsbereich¹¹ und $25 \times 25 \times 25$ beim großen Skalierungsbereich, so erreicht man bei kleiner Start-Kovarianz, kleinem Skalierungsbereich und quasi-inverser Adaption sowohl bei `marmor_stat` als auch bei `kwbb` bei 50 % Dämpfung oder mehr für *alle* Punkte Konvergenz der OF-Schätzung (Abbildung 6.21, unterste Zeile). Auch bei den anderen sieben Kombinationen von Start-Kovarianz, Adaptionsverfahren und Skalierungsbereich werden bei dieser festen Maskengröße bei 50 % Dämpfung sehr gute Konvergenzergebnisse erzielt (Tabelle 6.21). Durch die Größe der Masken vervielfacht sich aber die Rechenzeit. Als Kompromiss wurde mit den bereits erprobten, gegenüber der automatisch bestimmten Größe um 1 Pixel erweiterten Masken gearbeitet. Wie man Tabelle 6.21 entnimmt, führt dies immer noch an nahezu allen Pixeln zur Konvergenz und in jedem Fall zu deutlich besseren Ergebnissen als ohne Dämpfung.

Abbildung 6.22 visualisiert die durch die Dämpfung erhaltenen Mittelungsmasken. Wie man sieht, liegt das Ergebnis in der Mitte der beiden sonst verwendeten Masken. Insbesondere in dem Ausschnitt aus der Bildfolge `kwbb` (rechts) erkennt man, dass sich die Maske nun sehr gut an die Objektkontur anpasst.

¹¹Es wurde übrigens mit $21 \times 21 \times 21$ statt der bisher als maximale Größe verwendeten und daher nahelegenden $19 \times 19 \times 19$ Pixel gearbeitet, da nach dem Verfahren von [Orte 94] bestimmte Masken für zeilenverschränkte Bilder in y -Richtung eine Ausdehnung von $4n + 1$, $n \in \mathbb{N}$ besitzen müssen.

V-ID	40	41	42	43	
Adapt.-Verf.	linear	quasi-invers	linear	quasi-invers	
$\sigma_{\min}, \sigma_{\max}$	1,00; 2,85				
σ_x, σ_y	1,00	1,41	1,00	1,41	1,00
σ_t	1,00	1,00	1,00	1,00	1,00
					ges.

marmor_stat

	feste Maskengröße 21 ³ bzw. 25 ³ , It. 100, 50% Dämpfung								
Anz. Pixel gesamt	10000								
Diff. unter Konv.-Schwelle	9996	9996	10000	10000	9726	9703	10000	10000	80000
Oszillation (2er-Zyklus)	4	3	0	0	250	259	0	0	79421
> 2er oder kein Zyklus	0	1	0	0	24	38	0	0	516
									63

	um 1 Pixel vergrößerte Masken, It. 100, 50% Dämpfung								
Diff. unter Konv.-Schwelle	9996	9995	9997	9997	9672	9662	9994	9994	79307
Oszillation (2er-Zyklus)	1	1	3	3	285	304	5	5	607
> 2er oder kein Zyklus	3	4	0	0	43	34	1	1	86

	zum Vergleich: um 1 Pixel vergr. Masken, It. 100, ohne Dämpfung								
Diff. unter Konv.-Schwelle	9138	9085	9696	9713	8555	8604	9459	9463	73713
Oszillation (2er-Zyklus)	774	802	302	287	828	835	539	535	4902
3er bis 8er Zyklus	85	90	0	0	232	105	1	1	514
> 8er oder kein Zyklus	3	23	2	0	385	456	1	1	871

kwbb

	feste Maskengröße 21 ³ bzw. 25 ³ , It. 100, 50% Dämpfung								
Anz. Pixel gesamt	25600								
Diff. unter Konv.-Schwelle	25554	25494	25600	25600	25553	25541	25564	25559	204800
Oszillation (2er-Zyklus)	11	10	0	0	22	19	36	41	204465
> 2er oder kein Zyklus	35	96	0	0	25	40	0	0	139
									196

	um 1 Pixel vergrößerte Masken, It. 100, 50% Dämpfung								
Diff. unter Konv.-Schwelle	25427	25456	25209	25174	25089	25133	24770	24814	201072
Oszillation (2er-Zyklus)	71	54	321	348	297	282	734	698	2805
> 2er oder kein Zyklus	102	90	70	78	214	185	96	88	923

	zum Vergleich: um 1 Pixel vergr. Masken, It. 100, ohne Dämpfung								
Diff. unter Konv.-Schwelle	21349	21353	21180	21038	19104	19055	17677	17639	158395
Oszillation (2er-Zyklus)	4092	4131	4390	4549	6311	6339	7682	7734	45228
3er bis 8er Zyklus	45	44	16	7	113	165	152	189	731
> 8er oder kein Zyklus	114	72	14	6	72	41	89	38	446

Tabelle 6.21: Oszillationen und Zyklen der OF-Schätzung bei Dämpfung, zum Vergleich jeweils die (verkürzten) Ergebnisse aus Tabellen 6.13 und 6.16:

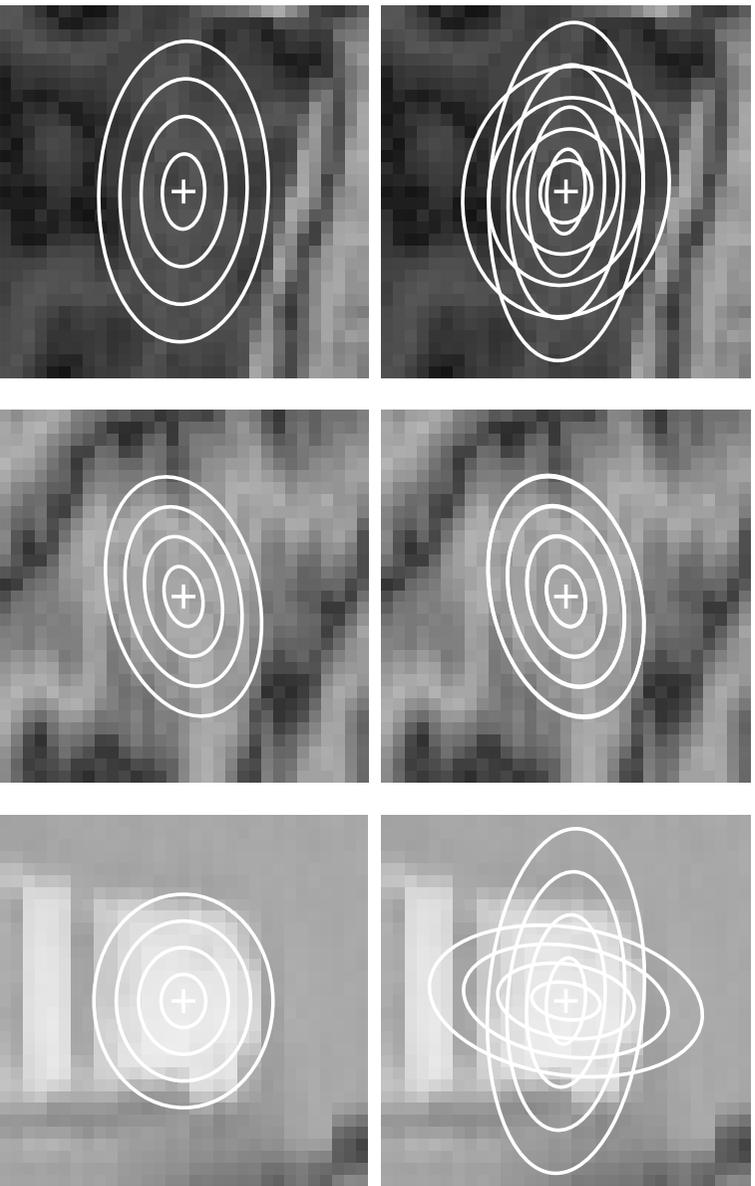


Abbildung 6.22: Ausdehnung der dreidimensionalen Mittelungsmasken in Iteration 19 bei 50% Dämpfung (unten) im Vergleich zu den oszillierenden Masken in Iteration 18 und 19 ohne Dämpfung (oben) an den Punkten (339, 129) (links) und (319, 73) (Mitte) in Bild 150 der Bildfolge `marmor_stat` sowie an Punkt (229, 295) (rechts) in Bild 2540 der Bildfolge `krwbB`. Jede Ellipse entspricht einer Standardabweichung. Parameter: „quasi-inverse“ Adaption, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$, $\sigma_x = \sigma_y = 1,41$, $\sigma_t = 1,00$.

6.5.3.2 Anpassung der Adaptionfunktion

Als Alternative zur Einföhrung der gedämpften Iteration wird nun die Einföhrung einer neuen Adaptionfunktion untersucht. Aufgrund der Überlegungen zur Oszillation (vgl. Abbildung 6.9, Seite 135, auch Abschnitt 6.5.2) soll eine Adaptionfunktion entworfen werden, die *ohne* Skalierung der Eigenwerte mit der Spur arbeitet, da diese Skalierung bisher als Hauptgrund für mangelnde Konvergenz angesehen wird. Dadurch wird der Entwurf dieser Funktion im Eindimensionalen möglich. Folgende Eigenschaften sollen für die eindimensionale Adaptionfunktion $f_{A,1D}$ gefordert werden:

- Ist ein Eigenwert $\lambda = 0$, so wird die maximale Skala, gegeben durch σ_{\max} , verwendet.
- Ist ein Eigenwert $\lambda = 0$, so wird die maximale Skala, gegeben durch σ_{\max} , verwendet.
- Ist ein Eigenwert λ so groß wie der maximal mögliche Eigenwert λ_{\max} , so wird die minimale Skala, gegeben durch σ_{\min} , verwendet.

- Bei λ zwischen 0 und λ_{\max} nimmt die Skala mit wachsendem λ monoton ab.

Da die tatsächlich auftretenden Eigenwerte wesentlich kleiner sind als λ_{\max} (vgl. Tabelle 6.20), erscheint es sinnvoll, schon bei Eigenwerten, die eine Schwelle $\lambda_S < \lambda_{\max}$ überschreiten, die minimale Skala zu verwenden.

Es wurde eine Adaptionfunktionen getestet, die sich an der quasi-inversen Adaption orientiert und die Skalierung der Eigenwerte mit der Spur durch eine (positionsumabhängige) Skalierung mit λ_S ersetzt:

$$f_{A,1D}(\lambda) = \begin{cases} \sigma_{\min}^2 & \text{für } \lambda > \lambda_S \\ \frac{\sigma_{\max}^2 \cdot \sigma_{\min}^2}{\left(1 - \frac{\lambda}{\lambda_S}\right) \cdot \sigma_{\min}^2 + \frac{\lambda}{\lambda_S} \cdot \sigma_{\max}^2} & \text{sonst} \end{cases} \quad (6.28)$$

Dieses Vorgehen soll als ungekoppelte quasi-inverse Adaption (UQI-Adaption) bezeichnet werden.

Bei allen Versuchen wurde λ_S auf 120 gesetzt. Das Ergebnis lässt sich wie folgt zusammenfassen:

- Es treten immer noch Fälle von Nicht-Konvergenz – stets als Oszillation – auf. Bei kleinem Skalierungsbereich handelt es sich aber nur um vereinzelte Punkte, an denen die OF-Schwelle jeweils nur knapp überschritten wird. Vermutlich wird eine Vergrößerung der Masken um 1 Pixel zu Konvergenz an allen Punkten führen. Bei großem Skalierungsbereich treten die Fälle geballt entlang ausgeprägter Kanten auf, und zwar genau so, wie es im Gedankenexperiment (Seite 158) auch war.
- Die Qualität der OF-Schätzung ist schlechter als bei (gekoppelter) quasi-inverser Adaption.
- Die Kategorisierungsergebnisse sind ungefähr gleichwertig zu denen bei (gekoppelter) quasi-inverser Adaption.
- Wie stark das Ergebnis vom Parameter λ_S abhängt, ist (noch) nicht bekannt.

6.5.4 Hinweise auf langsame Konvergenz

In [Middendorf 02] wurde ausführlich untersucht, ob es „Merkmale“ gibt, die vor Beginn der Iteration bestimmt werden können und die einen Rückschluss darauf zulassen, an welchen Punkte Konvergenz schnell oder langsam eintritt. Die Untersuchung soll hier nur zusammengefasst wiedergegeben werden:

Aufgrund der beobachteten Lage der Pixel mit mangelnder Konvergenz in der Nähe von dominierenden Grauwertstrukturen wurde zunächst vermutet, eine starke Änderung

der lokalen Tensoren $\nabla g \nabla g^T$ innerhalb des betrachteten Mittelungsbereiches oder in Richtung des Eigenvektors zum größten Eigenwert führe zu langsamer oder scheinbarer Konvergenz. Diese Änderung wurde durch die „Gesamtschiefe“ S_g – Änderung aller Komponenten des Tensors – bzw. die „Richtungsschiefe“ S_R – Änderung des Gradienten in Richtung des Eigenvektors zum größten Eigenwert – quantitativ erfasst, mit:

$$S_g(\mathbf{x}) = \left| \int d\boldsymbol{\xi} \frac{\nabla g(\boldsymbol{\xi})(\nabla g(\boldsymbol{\xi}))^T}{(2\pi)^{3/2} \sqrt{|\Sigma_M|}} \cdot (\mathbf{e}_1^T(\mathbf{x} - \boldsymbol{\xi}))^3 \cdot e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\xi})^T (\Sigma_M)^{-1} (\mathbf{x} - \boldsymbol{\xi})} \right| \cdot (6.29)$$

$$S_R(\mathbf{x}) = \left| \int d\boldsymbol{\xi} \frac{\mathbf{e}_1^T \nabla g(\boldsymbol{\xi})}{(2\pi)^{3/2} \sqrt{|\Sigma_M|}} \cdot (\mathbf{e}_1^T(\mathbf{x} - \boldsymbol{\xi}))^3 \cdot e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\xi})^T (\Sigma_M)^{-1} (\mathbf{x} - \boldsymbol{\xi})} \right| \cdot (6.30)$$

Dabei konnte kein Zusammenhang zwischen Schiefe in der Anfangsiteration, Schiefe zum Konvergenzzeitpunkt und Anzahl der Iterationen bis zur Konvergenz festgestellt werden. Allerdings wurde beobachtet, dass die Schiefe durch die Adaption in den meisten Fällen geringfügig abnimmt, die Adaption führt also zu in sich homogenen Tensoren.

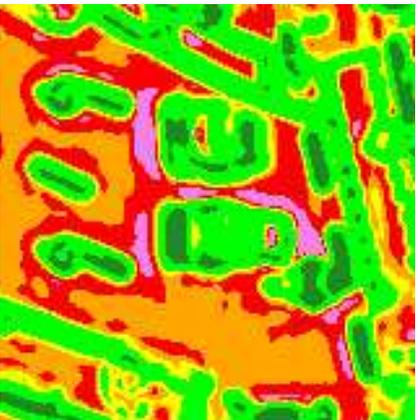
In einem weiteren Versuch wurden die GSTs verglichen, die sich bei der Verwendung der – im Rahmen des vorgegebenen Skalierungsbereichs – größt- bzw. kleinstmöglichen isotropen Kovarianzmatrizen ergaben. Sei also GST_{\min} der mit $\Sigma_A = \Sigma_{\min} = \text{diag}(\sigma_{\min}^2, \sigma_{\min}^2, \sigma_{\min}^2)$ und GST_{\max} der mit $\Sigma_A = \Sigma_{\max} = \text{diag}(\sigma_{\max}^2, \sigma_{\max}^2, \sigma_{\max}^2)$ berechnete GST. Für die Quantifizierung des Unterschieds zwischen zwei GSTs wurden mehrere Alternativen getestet. Die besten Ergebnisse brachte der Vergleich der Wurzel der Spur von GST_{\min} bzw. GST_{\max} , also der gewichtet gemittelten Gradientenbeiträge. Es zeigte sich (Abbildung 6.23), dass die Nicht-Konvergenz vor allem an den Stellen auftrat, an der die Spur des GST mit wachsendem Σ zunahm – während im „Normalfall“ die Vergrößerung der Kovarianz durch die implizite Glättung zu einer Abnahme der Spur des GST führt. Dies sind aber genau die Stellen, an denen mit dem in Abbildung 6.9 illustrierten Effekt gerechnet werden muss. Durch die gegenseitige Koppelung der GST-Schätzungen beim doppelt-adaptiven Verfahren kommt es in einem größeren Bereich als dem durch das Kriterium „Differenz der Spur bei unterschiedlichem Σ “ identifizierten zu Nicht-Konvergenz.

6.5.5 Anmerkungen zum Konvergenzverhalten

Nach dem Fixpunktsatz von Banach konvergiert eine Iteration für jeden beliebigen Startwert gegen den dann eindeutigen Fixpunkt einer Funktion, falls die Funktion global, d. h. auf dem gesamten Definitionsbereich kontrahierend ist. Bei der hier betrachteten Tensor-Iteration bzw. der Skalen-Iteration ist aber nicht bekannt, ob die Funktion an einigen oder den meisten Punkten kontrahierend ist. Das bedeutet aber:

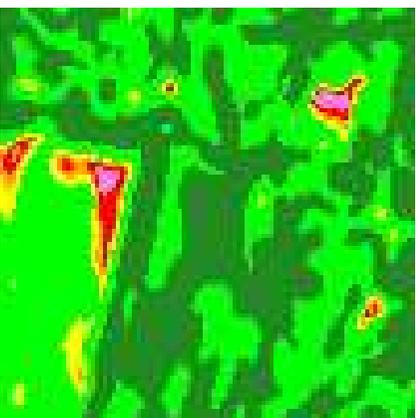
krbb

Bild 2540, Ausschnitt



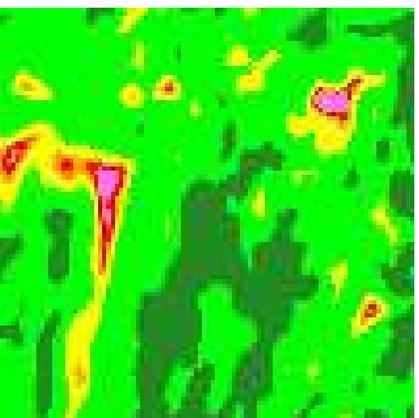
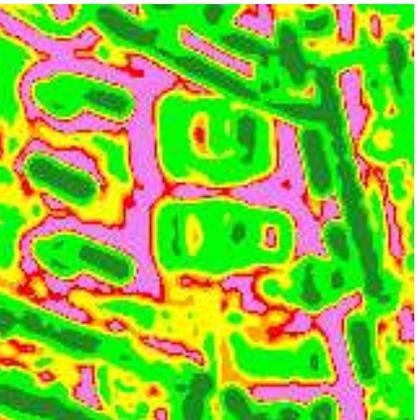
marmor_stat

Bild 150, Ausschnitt



Differenz der Wurzel
der Spur der GSTs

< -10,0
-10,0 - -2,5
-2,5 - -1,0
-1,0 - 0,0
0,0 - 1,0
≥ 1,0



Differenz der Wurzel
der Spur der GSTs,
skaliert mit der Sum-
me der Wurzeln

< -0,6
-0,6 - -0,3
-0,3 - -0,1
-0,1 - 0,0
0,0 - 0,1
≥ 0,1

Abbildung 6.23: Differenz der Wurzel aus der Spur der mit $\sigma_x = \sigma_y = \sigma_t = \sigma_{\max}$ bzw. mit $\sigma_x = \sigma_y = \sigma_t = \sigma_{\min}$ berechneten Tensoren: $\text{GST}(\sigma_x = \sigma_y = \sigma_t = 2,85)$ – $\text{GST}(\sigma_x = \sigma_y = \sigma_t = 1,00)$, in der unteren Zeile skaliert mit Summe der Wurzeln der Spur des GST.

- Die Iteration kann scheitern, obwohl ein Fixpunkt existiert (dieser Fall wurde oft genug beobachtet).
- Konvergiert die Iteration, so ist der Grenzwert ein Fixpunkt. Dieser muss aber nicht eindeutig sein.
- Aus der Konvergenz kann auch nicht geschlossen werden, dass die Funktion global kontrahierend ist. Vielmehr kann bei einer nur lokal kontrahierenden Funktion die Iteration je nach Startwert konvergieren oder nicht.
- Durch eine ungeschickte Implementierung oder Parametrierung (vgl. Wahl der Maskengröße) kann die Konvergenz scheitern, weil dann kein Fixpunkt mehr existieren muss.

- Aus den beiden letztgenannten Punkten folgt, dass Konvergenz der GST-Schätzung keine inhärente Eigenschaft der lokalen Grauerwertung in Kombination mit der Adaptionfunktion ist, sondern auch noch von Implementierung, Parametrierung, Startwert und möglicherweise anderen Faktoren abhängt.
- Bei abstoßenden Fixpunkten (wenn der Gradient in der Umgebung des Fixpunktes größer als 1 ist), kann sich die Iterationsfolge wieder vom Fixpunkt entfernen, wenn dieser nicht exakt getroffen wurde. Bei einem Gradienten knapp über 1 und einer anfangs nur geringen Entfernung zum Fixpunkt kann dies auch erst nach zahlreichen Iterationen geschehen. Dies ist eine mögliche Erklärung für die z. T. erst nach mehreren Iterationen unter der Konvergenzschwelle auftretenden Ausreißer. Eine Verbesserungsmöglichkeit des Algorithmus bestände also darin, einen zunehmenden Abstand der Iterationsergebnisse zu detektieren und die Iteration ggfs. abzubrechen.
- Theoretisch betrachtet, kommt es im Verlauf der Iteration *immer* zu einem Zyklus (möglicherweise der Länge 1, also Konvergenz), wie man sich mit folgender Überlegung verdeutlichen kann: Die Menge der möglichen Kovarianzmatrizen bzw. der GST und somit auch jede Folge innerhalb dieser Mengen sind beschränkt. Also besitzt jede durch Iteration gewonnene Folge eine konvergente Teilfolge. Da die Iterationsfunktion stetig ist, bilden auch die Abbilder der Elemente dieser konvergenten Teilfolge eine konvergente Folge, und somit deren Abbilder usw. Nach einer endlichen Iterationszahl wird wieder ein Element der anfangs ausgesonderten konvergenten Teilfolge erreicht. Somit entsteht immer ein Zyklus durch Elemente jeweils konvergenter Teilfolgen. Die Fälle, bei denen kein Zyklus beobachtet wurde, sind also darauf zurückzuführen, dass die tatsächlichen Zyklen länger waren als die betrachtete Anzahl an Iterationen, dass nach der berechneten Iterationszahl die Konvergenz gegen diese Teilfolgen noch nicht weit genug fortgeschritten ist oder dass Implementierungs- bzw. Parametrierungsdetails die Bildung dieser Teilfolgen verhindern.

6.6 Beurteilung des adaptiv geschätzten OFS

Zur Abrundung der vorangehenden Untersuchungen sollen nun noch die Ergebnisse der auf dem adaptiv-bestimmten GST beruhenden OF-Schätzung mit den in Abschnitt 6.2 genannten Referenzwerten verglichen werden.¹²

Tabelle 6.22 zeigt, dass schon bei Verwendung der von [Nagel & Gehrke 98a] vorgeschlagenen linearen Adaptionfunktion die Ergebnisse in den meisten Fällen besser sind als nicht-adaptiv.

¹²Mit Ausnahme der Bildfolge `marmor_stat`, die im Verlauf der vorangehenden Untersuchungen bereits ausführlich betrachtet wurde.

Verfahren	Maskengröße		ψ_{v^+, u^+} \emptyset Med.	$\ v - u\ _2$ \emptyset Med.	Dichte		
	Ableitung	Mittelung					
yosemite							
(4) alle	15 × 15 × 15	15 × 15 × 1	2,68	1,84	0,1100	0,0720	100
(4) nur ROF	15 × 15 × 15	15 × 15 × 1	2,27	1,81	0,0833	0,0594	66,2
trans_tree							
(1) alle	19 × 19 × 19	19 × 19 × 19	0,35	0,31	0,0227	0,0207	100
(1) nur ROF	19 × 19 × 19	19 × 19 × 19	0,32	0,29	0,0217	0,0198	76,8
(3) alle	19 × 19 × 19	19 × 19 × 1	0,36	0,31	0,0231	0,0202	100
(3) nur ROF	19 × 19 × 19	19 × 19 × 1	0,32	0,29	0,0215	0,0191	67,8
div_tree							
(1) alle	19 × 19 × 19	19 × 19 × 19	3,02	2,39	0,0869	0,0551	100
(1) nur ROF	19 × 19 × 19	19 × 19 × 19	2,63	2,28	0,0592	0,0530	80,6
(3) alle	19 × 19 × 19	19 × 19 × 1	3,24	2,72	0,0741	0,0634	100
(3) nur ROF	19 × 19 × 19	19 × 19 × 1	3,11	2,65	0,0714	0,0620	76,7
tr_bsb							
(1) alle	19 × 19 × 15	19 × 19 × 15	3,62	1,15	0,1473	0,0478	100
(1) nur ROF	19 × 19 × 15	19 × 19 × 15	1,54	0,97	0,0682	0,0361	70,6
(3) alle	19 × 19 × 19	19 × 19 × 1	3,59	1,17	0,1553	0,0475	100
(3) nur ROF	19 × 19 × 19	19 × 19 × 1	1,67	0,97	0,0733	0,0381	70,4

Tabelle 6.22: Ergebnisse der adaptiven OF-Schätzung. Wie bei Tabellen 6.1 bis 6.5 geben ψ_{v^+, u^+} und $\|v - u\|_2$ Aufschluss über die Qualität der OF-Schätzung. In der Spalte „Verfahren“ wird durch die Zahl in Klammern angegeben, auf welcher nicht-adaptiven GST-Schätzung (vgl. Tabelle 6.1 ff, Seite 108 ff) die Iteration aufsetzt, und ob die Werte alle Punkte oder nur die ROF-Vektoren umfassen. Doppelt-adaptives Verfahren, **lineare Adaption**, Start-Kovarianz: $\Sigma_A = \text{diag}(1,41^2; 1,41^2; 1,41^2)$, $\Sigma_M = 2\Sigma_A$, 10 Iterationen, 50 % Dämpfung, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$.

Bei Verwendung der im Rahmen der vorliegenden Arbeit eingeführten *quasi-inversen Adaptionfunktion* (Tabelle 6.23) kommt es im Vergleich zur *linearen Adaptionfunktion* (Tabelle 6.22) in *allen* untersuchten Fällen zu einer Verbesserung der OF-Schätzung bei meist gleichzeitiger Zunahme des Anteils an ROF-Vektoren.

Gegenüber den *nicht-adaptiv* erzielten Werten zeigt sich bei den Bildfolgen `trans_tree` und `tr_bsb` eine klare Verbesserung der Ergebnisse. Bemerkenswert ist, dass dabei in einigen Fällen *ohne zeitliche* Mittelung ebenso gute oder sogar bessere Ergebnisse erzielt werden als *mit räumlich-zeitlicher* Mittelung. Bei der Bildfolge `tr_bsb` sind die Ergebnisse bei *dichten* OF-Feldern besser als alle verfügbaren Vergleichswerte (durchschnittlicher Winkelfehler von 3,29 statt 3,39, vgl. Tabelle 6.4), bei *spärlich besetzten* OF-Feldern geringfügig schlechter (durchschnittlicher Winkelfehler von 1,45 statt 1,44),

Verfahren	Maskengröße Ableitung	Mittelung	ψ_{σ^+, u^+} \emptyset Med.	$\ w - u\ _2$ \emptyset Med.	Dichte		
yosemite							
(4) alle	15 × 15 × 15	15 × 15 × 1	2,35	1,64	0,0980	0,0643	100
(4) nur ROF	15 × 15 × 15	15 × 15 × 1	2,01	1,58	0,0744	0,0534	71,1
trans_tree							
(1) alle	19 × 19 × 19	19 × 19 × 19	0,32	0,27	0,0205	0,0180	100
(1) nur ROF	19 × 19 × 19	19 × 19 × 19	0,28	0,25	0,0190	0,0165	77,6
(3) alle	19 × 19 × 19	19 × 19 × 1	0,33	0,27	0,0209	0,0180	100
(3) nur ROF	19 × 19 × 19	19 × 19 × 1	0,28	0,25	0,0187	0,0163	70,9
div_tree							
(1) alle	19 × 19 × 19	19 × 19 × 19	2,63	2,06	0,0798	0,0472	100
(1) nur ROF	19 × 19 × 19	19 × 19 × 19	2,22	1,91	0,0499	0,0442	79,9
(3) alle	19 × 19 × 19	19 × 19 × 1	2,81	2,39	0,0644	0,0556	100
(3) nur ROF	19 × 19 × 19	19 × 19 × 1	2,62	2,25	0,0600	0,0521	76,8
tr_bsb							
(1) alle	19 × 19 × 15	19 × 19 × 15	3,43	1,02	0,1305	0,0418	100
(1) nur ROF	19 × 19 × 15	19 × 19 × 15	1,45	0,91	0,0635	0,0343	76,1
(3) alle	19 × 19 × 19	19 × 19 × 1	3,29	1,01	0,1551	0,0412	100
(3) nur ROF	19 × 19 × 19	19 × 19 × 1	1,56	0,87	0,0664	0,0345	76,9

Tabelle 6.23: Ergebnisse der adaptiven OF-Schätzung. Wie Tabelle 6.22, aber mit quasi-inverser Adaption.

aber bei wesentlich höherer Dichte (76,1 % statt 55,9%).

Bei der Bildfolge yosemite kann aufgrund der Kürze der Bildfolge bei der Adaption nur eine zweidimensionale Mittelung durchgeführt werden. Dies führt zwar zu besseren Ergebnissen als nicht-adaptiv mit *zwei*dimensionaler Mittelung, aber nicht zu besseren Ergebnissen als nicht-adaptiv mit *drei*dimensionaler Mittelung. Das Problem liegt aber in den zur Verfügung stehenden Daten, nicht im Algorithmus. Bei der *div_tree*-Bildfolge kommt es zu einer Verschlechterung des Ergebnisses, was auf zwei Probleme zurückgeführt werden kann. Zum einen führt die Kamerabewegung in dieser Bildfolge zu einer Verletzung von Annahme 8, was sich bei durch die Adaption größeren Masken stärker auswirkt. Zum anderen wirkt sich die bei der Adaption stärkere implizite Glättung in den relativ großen kontrastschwachen Flächen negativ aus.

6.7 Fazit zur adaptiven GST-Bestimmung

Die in diesem Kapitel vorgestellten Arbeiten liefern zwei wesentliche Beiträge zur Verbesserung der OF-Schätzung auf Basis eines adaptiv-bestimmten Grauwert-Struktur-Tensors. Zum einen konnte durch die sorgfältige Wahl einer neuen Adaptionfunktion („quasi-inverse Adaption“) eine Verbesserung sowohl der Qualität der OF-Schätzung als auch der GST-Kategorisierung erzielt werden. Zum anderen konnte das bisher heuristisch geprägte Adaptionsverfahren in einen theoretischen Rahmen eingebettet werden. Dies führte wiederum zu einer Reduzierung der Abhängigkeit der Ergebnisse von der gewählten Start-Kovarianz. Die erzielten Verbesserungen wurden durch umfangreiche Versuche – auch auf allgemein bekannten „Standardbildfolgen“ – nachgewiesen.

Kapitel 7

Vertiefende Untersuchungen

Dieses Kapitel beschäftigt sich mit dem Verhalten des OF-Schätzansatzes in zwei speziellen Fällen: beim Zusammentreffen zweier (oder mehr) unterschiedlicher Bewegungen innerhalb des Mittelungsbereichs, und bei sehr schnellen Bewegungen. Für beide Fälle wird eine Verbesserung des bisherigen Ansatzes vorgeschlagen und experimentell untersucht. Schließllich wird gezeigt, wie sich verbesserte OF-Schätzung und Segmentierung im Anwendungsbeispiel „Fahrzeugverfolgung im Bildbereich“ auswirken.

7.1 Diskontinuitäten und Diskontinuitätshinweise

Das Ergebnis der GST-Kategorisierung wurde neben der Qualität der OF-Schätzung als eines der beiden Qualitätskriterien für die GST-Schätzung definiert. Allerdings treten im Zusammenhang mit als „Diskontinuität“ (OFD) eingeordneten Punkten mehrere Probleme auf, von denen die wichtigsten in diesem Abschnitt behandelt werden sollen. In diesem Zusammenhang soll von einer *Diskontinuität* im Bild an den Stellen geredet werden, an denen eine Unstetigkeit der *Verschiebungsrate* vorliegt. Ein *Diskontinuitätsmaß* sei jede Funktion, die umso größere Werte annimmt, je mehr Anzeichen für das Vorliegen einer solchen Diskontinuität sich an einem Pixel finden lassen. Überschreitet das Diskontinuitätsmaß eine Schwelle, so liegt ein *Diskontinuitätshinweis* vor. Zusammenhängende Bereiche von Diskontinuitätshinweisen bilden *Diskontinuitätsbereiche*. Die Begrenzungslinie eines Objektbildes soll auch als *Objektbildskante* (OAKante) bezeichnet werden, unabhängig davon, ob sie im Bild tatsächlich erkennbar ist oder nicht. Im Gegensatz dazu bezeichne *Grauwertkante* den Ort des Maximums eines signifikanten Helligkeitsunterschiedes im Bild.

Mit diesen Bezeichnungen lassen sich folgende Beobachtungen formulieren:

- Die Mitte des Diskontinuitätsbereichs und das Maximum des Diskontinuitätsmaßes fallen oft nicht mit der Kante zwischen Objektbildern zusammen. Diese Beobachtung wird in Abschnitt 7.1.1 präzisiert und erklärt.

- Es treten Diskontinuitätslinien an Stellen auf, an denen sie nicht erwartet wurden (Abschnitt 7.1.2).
- Es treten keine Diskontinuitätslinien an Stellen auf, an denen Diskontinuitäten vorliegen (Abschnitt 7.1.3).
- Es ergeben sich oft viel zu breite Diskontinuitätsbereiche (Abschnitt 7.1.4).

7.1.1 „Verschobene“ Diskontinuitätsbereiche

Betrachtet man Abbildung 6.5 oder die vergrößerte Darstellung in Abbildung 7.1, so fällt auf, dass die Mitte des Diskontinuitätsbereiches nicht mit der Kante zwischen Hintergrund und Abbild des bewegten Objektes übereinstimmt, sondern in dem Bereich, in dem die Diskontinuität nahezu senkrecht zur Verschiebungsrichtung verläuft (Kanten 1 und 2), um mehrere Pixel nach links in den Bildhintergrund hinein verschoben ist¹. Bei Kante 3 hingegen, die in Bewegungsrichtung verläuft, erkennt man eine symmetrische Anordnung des Diskontinuitätsbereichs beiderseits der Objektabbildskante.

Ein ähnliches – und offensichtlich verwandtes – Phänomen war bereits bei vorangehenden Versuchen (z. B. [Middendorf & Nagel 00], [Middendorf 02]) beobachtet worden: Der Bereich, in dem OF-Vektoren geschätzt werden, die auf eine Bewegung hindeuten, war oft viel weiter ausgedehnt als das bewegte Objektabbild selbst. Allerdings wurde dies bisher nur bei Bildfolgen beobachtet, die einen kontrastarmen Hintergrund besitzen.

Vom Problem der genauen Lokalisierung der verdeckenden Objektabbildkante mit Hilfe von OF-Diskontinuitäts Hinweisen berichtet auch [Thompson 98]. Die in diesem Journalbeitrag gegebenen Beispiele sind aber nicht mit den hier verwendeten Bildfolgen vergleichbar: Das erste Beispiel besteht aus dunklen Balken vor weißem Hintergrund (jeweils ohne Textur), beim zweiten sind die Objekte zwar texturiert, die Helligkeitsunterschiede zwischen den beiden Objekten aber größer als die innerhalb der Objekte. Durch die (implizite) Interpolation des verwendeten OF-Schätzverfahrens – einer modifizierte Version des Verfahrens von [Horn & Schunck 81], die nach Autorenangaben Bewegungshinweise an örtlichen Grauwertkanten besonders stark berücksichtigt – kommt es in den homogenen Bereichen zwangsläufig zu einem sanften OF-Übergang.

Dennoch besteht auch bei dem hier betrachteten Fall zumindest bei Kante 1 (vgl. Abbildung 7.1) eine mögliche Erklärung für die Verschiebung des Diskontinuitätsbereiches im erheblichen Helligkeitsunterschied zwischen dem sich bewegenden dunklen Block und dem hellen Hintergrund. Dies kann man sich auf verschiedenen Wegen deutlich machen.

¹Es handelt sich hierbei wirklich hauptsächlich um eine Verschiebung in der Bildhintergrund und nur zufällig auch um eine Verschiebung in Bewegungsrichtung. Eine systematische Verschiebung in oder gegen die Bewegungsrichtung wäre sowieso nur mit einem Implementierungsfehler zu erklären, da das verwendete Schätzverfahren in Zeit-Richtung absolut symmetrisch ist. Durch mehrere Tests, u. a. Berechnung des GST auf derselben Bildfolge rückwärts abgespielt, konnte aber ein Implementierungs- oder Darstellungsfehler so gut wie ausgeschlossen werden.

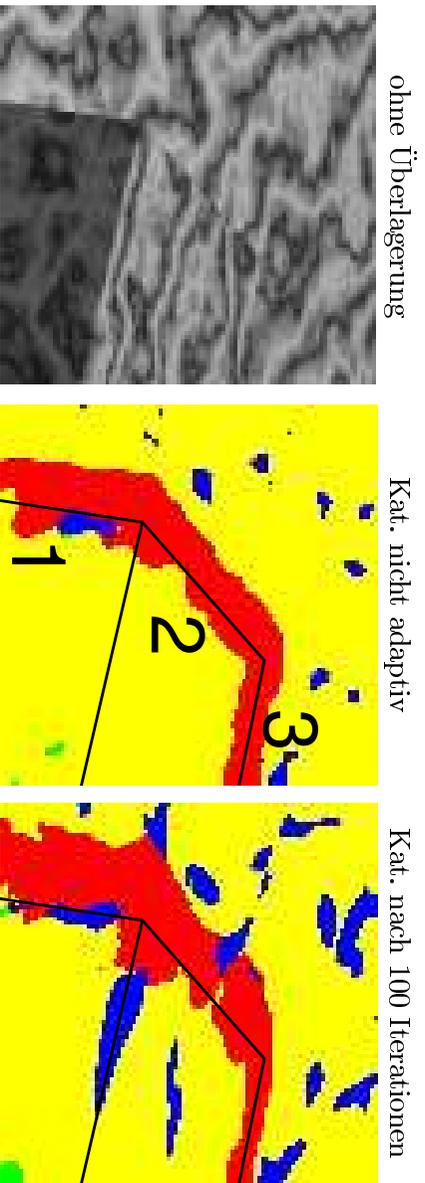


Abbildung 7.1: `larmor_stat`, Bild 1501 Verschiebung des Diskontinuitätsbereichs in den Bildhintergrund gegenüber der schwarz eingezeichneten Grenze des Objektbilds, nicht-adaptiv (Mitte) und nach 100 Iterationen (rechts). Zur späteren Bezugnahme sind die drei Objektabschnittskanten, an denen eine Bewegungsdiskontinuität vorliegt, durchnummeriert. Parameter: große Start-Kovarianz, kleiner Skalierungsbereich, quasi-inverse Adaption.

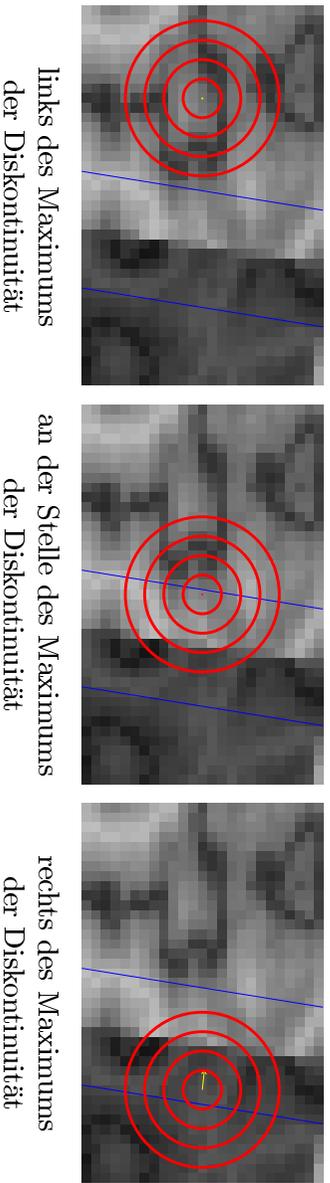


Abbildung 7.2: Erklärung der „verschobenen“ Diskontinuitätsbereiche durch „dominierende Tensoren“ am Beispiel eines Ausschnitts aus Bild 150 der Bildfolge `marmor_stat`. Im rechten Teil des Bildes erkennt man die dunkle Vorderseite des bewegten Marmorblocks, links den helleren Hintergrund. Die roten Kreise deuten die Ausdehnung der Mittelungsmaske für den bei großer Start-Kovarianzmatrix verwendeten Wert $\Sigma_M = \text{diag}(2 \cdot 1, 41^2; 2 \cdot 1, 41^2; 2 \cdot 1, 00^2)$ an. Jeder Kreis entspricht einer Standardabweichung. Im Übergangsbereich, markiert durch die beiden blauen Linien, werden die lokalen Tensoren $(\nabla g)(\nabla g)^T$ durch die Grauwertkante dominiert, die sich wie der rechte Bildteil bewegt. Die größte Änderung der lokalen Tensoren tritt daher im Bereich der linken blauen Linie auf, dort ist die Diskontinuität am größten.

Erklärung durch dominierende Tensoren In der Nähe der Kante dominiert dieser Helligkeitsunterschied über den durch die Textur bedingten; die lokalen Tensoren $(\nabla g)(\nabla g)^T$ werden durch den von der Helligkeitskante verursachten Gradienten dominiert. Abbildung 7.2 verdeutlicht und erklärt dies an einem realen Beispiel:

Im rechten Teil des Bildes erkennt man die dunkle Vorderseite des bewegten Marmor-

blocks, links den helleren Hintergrund. Der dargestellte Ausschnitt lässt sich anhand der Charakteristiken der lokalen Tensoren $(\nabla g)(\nabla g)^T$ in drei Bereiche teilen, angedeutet durch die blauen Linien. Im linken Bereich ergeben sich die lokalen Tensoren durch die Texturierung des Hintergrund, im rechten Bereich durch die Texturierung des bewegten Blocks. Im mittleren Teil – dem Übergangsbereich zwischen den Objekten abbildern – hingegen spielt die Textur nur eine untergeordnete Rolle, die Richtung des stärksten Gradienten wird durch den Helligkeitsunterschied zwischen bewegtem Block und Hintergrund dominiert. Betrachtet man nun einen Punkt, der weit genug von der Kante entfernt im Hintergrund liegt (linkes Bild), so mittelt dort die Tiefpaßmaske (rot) zur Berechnung von $\overline{(\nabla g)}(\overline{\nabla g})^T$ nur über lokale Tensoren $(\nabla g)(\nabla g)^T$, die die Bewegung des linken Teilbereichs wiedergeben. Es wird also ein regulärer OF geschätzt, der der Bewegung des linken Teilbildes entspricht. Im mittleren Bild liegt der Mittelpunkt der Maske immer noch links der Grauwertschwelle, die Maske erfasst aber sowohl Tensoren aus dem linken Teilbereich als auch aus dem Übergangsbereich. Die Tensoren im Übergangsbereich widerspiegeln allerdings die Bewegung des rechten Bildbereichs, da sich die Grauwertkante mit dem Vordergrundobjekt bewegt. Hinzu kommt, dass in diesem Bereich der Gradientenbetrag größer ist und dadurch diese Tensoren überproportional in den Gesamttensor $\overline{(\nabla g)}(\overline{\nabla g})^T$ einfließen. Dadurch erhalten beide Teilbereiche im Gesamttensor ungefähr gleichen Anteil, es kommt zu einem Diskontinuitätshinweis. Im rechten Bild liegt die Mittelungsmaske genau in derselben Entfernung zur Grauwertkante wie im mittleren Bild. In den Gesamttensor gehen die Gradienten des Übergangsbereiches am stärksten ein, gefolgt von lokalen Tensoren der rechten Teilbereichs. Deshalb kommt es in diesem Bereich oft zur Schätzung von ROF-Vektoren (entsprechend der Bewegung der Kante und somit des rechten Teilbereichs) oder zu DGD-Vektoren.

Erklärung über Tensor-Orientierung Auch wenn man die Orientierung des Eigensystems des GST betrachtet, wird die Verschiebung der Diskontinuität deutlich. Man rufe sich in Erinnerung, dass die lokalen Tensoren stets einen doppelten Eigenwert 0 haben, und dass sich der OF im Idealfall aus dem Eigenvektor zum (einzigem) Eigenwert 0 des gemittelten Tensors ergibt. Bei einem Punkt, der weit genug von der Kante entfernt im stationären Hintergrund liegt, ist die t -Ableitung stets 0, d. h. für alle lokalen Tensoren ist der Einheitsvektor in t -Richtung ein Eigenvektor zum Eigenwert 0. Die beiden anderen Eigenvektoren liegen dann beide in der x - y -Ebene. Ihre genaue Lage wird durch die maximale Gradientenrichtung bestimmt, in deren Richtung der Eigenvektor zum größten Eigenwert verläuft. Mittelt man über Tensoren, bei denen jeweils der Einheitsvektor in t -Richtung ein Eigenvektor zum Eigenwert 0 ist, so ist dieser Vektor auch beim GST ein – nun meist eindeutiger – Eigenvektor zum Eigenwert 0. Betrachtet man hingegen einen Punkt genau auf der bewegten Kante, so ergibt sich eine völlig andere lokale Grauwertstruktur. Der Eigenvektor zum größten Eigenwert ist hier durch die dominierende Kante bestimmt, und liegt (bei bewegter Kante) nicht mehr in der x - y -Ebene. Demzufolge kann der Einheitsvektor in t -Richtung kein

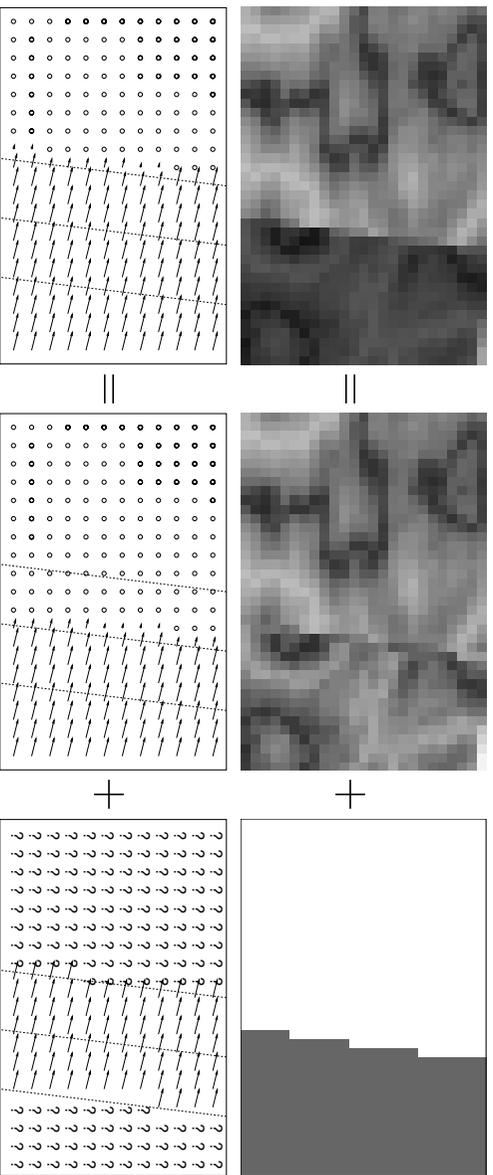


Abbildung 7.3: Erklärung der „verschobenen“ Diskontuitätsbereiche als Überlagerung am Beispiel eines Ausschnitts aus Bild 150 der Bildfolge `marmor_stat`. Zerlegt man die Grauwertstruktur (links oben) gedanklich in eine Überlagerung zweier ähnlich texturierter, aber gleichheller Flächen (oben Mitte) mit einer untexturierten Grauwertkante (oben rechts), so kann der OF durch Zusammensetzen der in den Teilbildern enthaltenen OF-Information erklärt werden.

Eigenvektor zum Eigenwert 0 mehr sein. Somit kommt es also noch im Bereich des Hintergrunds zu einer Drehung des Eigensystems unter Veränderung des Eigenvektors zum kleinsten Eigenwert, und somit zu einem großen Diskontuitätsmaß. Auf der anderen Seite der Kante kann sich die Drehung des Eigensystems fortsetzen, wenn der Einfluss der dominierenden Kante abnimmt. Da die Bewegung des Vordergrunds aber mit der der Kante übereinstimmt, verändert sich dabei der Eigenvektor zum kleinsten Eigenwert nicht, der kleinste Eigenwert bleibt nahe bei 0.

Erklärung als Überlagerung Schließlich kann die Grauwertstruktur an den betroffenen Stellen gedanklich auch als Überlagerung einer Grauwertkante zwischen homogenen Flächen und zweier texturerter, sich unterschiedlich bewegender Flächen aufgefasst werden (Abbildung 7.3). Der tatsächlich auftretende OF kann dann durch Zusammensetzen der in den Teilbildern enthaltenen OF-Information erklärt werden:

- Bei einer Grauwertkante zwischen zwei *homogenen* Flächen (Mitte) *unterschiedlicher Helligkeit* kann nämlich nur in einer durch die Skala und die Maskengröße bestimmten lokalen Umgebung der Kante überhaupt ein OF geschätzt werden. Diese Schätzung ist zu beiden Seiten der Kante symmetrisch, da es nicht möglich ist, lokal und bei Betrachtung nur eines Zeitpunkts² zwischen verdeckter und ver-

²Gibt man diese Einschränkung auf, so liegt genau der Fall vor, für den [Thompson 98] eine Lösung vorschlägt.

deckender Fläche zu unterscheiden. Der so geschätzte OF stimmt aber nur für die Hälfte der Punkte mit der VR überein.

- Beim Aneinandertreffen zweier unterschiedlich bewegter Flächen *gleicher Helligkeit* und *ähnlicher Textur* liegt die größte Änderung des OF und somit der maximale Diskontinuitäts Hinweis genau an der Grenze zwischen den beiden Flächen.

Bei einer Überlagerung der Grauwertstrukturen erfolgt näherungsweise auch eine Überlagerung der OF-Schätzungen. Die auf realen Daten basierende Tabelle 7.1 bestätigt im Wesentlichen diese vereinfachende Darstellung: Bei gleichhellen Texturen werden die größten Werte des Diskontinuitätsmaßes in der betrachteten Zeile 12 bei Pixel 24 und 25 erreicht, also beidseits der Objektabbildskante. Bei einer Aufhellung der rechten Bereichs um 20 Graustufen kommt es zu einer Zunahme des Diskontinuitätsmaßes bei Pixel 19 bis 23, zu einer Abnahme bei Pixel 24 bis 26, das Maximum liegt aber immer noch bei Pixel 24 und 25. Bei einer Aufhellung um 50 Grauwerte verschiebt sich das Maximum des Diskontinuitätsmaßes zu Pixel 22, also in Richtung des Hintergrunds, der verdeckten Fläche.

Anmerkungen Die beschriebene Verschiebung des Diskontinuitätsbereichs tritt offensichtlich nicht auf, wenn die Kante genau in Bewegungsrichtung verläuft (Kante 3 in Abbildung 7.1), da in einem solchen Fall die Grauwertkante selbst gar keine Bewegungsinformation liefert. Bei Kante 2 sind die Helligkeitsunterschiede zum Hintergrund deutlich geringer ausgeprägt als bei Kante 1, aber dennoch vorhanden, da – zufällig – am Rand des bewegten Marmorblock vor allem dunkle Bereiche liegen, die auf einen hellen Hintergrund treffen. Die Diskontinuitäten sind folgerichtig auch schmaler (senkrecht zur Kante gemessen) und nicht ganz so stark in den Hintergrund verschoben.

Diese Erklärungen lassen sich auch auf den Fall der zu großen Ausdehnung des von einem bewegten Objekt verursachten OF-Feldes bei homogenen Hintergrund anwenden. In diesen Fall liegen im Hintergrund nur lokale Tensoren mit sehr kleiner Spur vor, die bei der Gewichtung von den oft deutlich ausgeprägten, vom bewegten Objekt verursachten Tensoren „überstimmt“ werden. Dies führt dazu, dass auch in einigen Pixel Entfernung vom bewegten Objektabbild Punkte auftreten können, bei denen ein OF geschätzt wird, der der Objektbewegung entspricht. Erschwerend kommt hinzu, dass in diesem Fall meist auch keine Diskontinuitäts Hinweise erkannt werden können, da es keine Tensoren mit wesentlichen Einfluss auf das Ergebnis gibt, die mit der Objektbewegung inkompatibel sind.

Fazit Zwei häufig auftretende Probleme konnten auf dasselbe Grundproblem zurückgeführt werden, für das mehrere Erklärungen gegeben wurden. Aufgrund der Beobachtungen kann nicht erwartet werden, dass eine *exakte* bewegungs-basierte Bildsegmentierung ausschließlich auf Basis des bisher verwendeten GT-Schätzverfahrens möglich ist.

	gleichhelle Texturen	20 Grauwerte Unterschied	50 Grauwerte Unterschied
Pixel (..., 12)	$ \nabla g $ OF $\tilde{\lambda}_3$	$ \nabla g $ OF $\tilde{\lambda}_3$	$ \nabla g $ OF $\tilde{\lambda}_3$
15	8,886 $\begin{pmatrix} -0,000 \\ 0,000 \end{pmatrix}$ 0,0000	8,886 $\begin{pmatrix} -0,000 \\ 0,000 \end{pmatrix}$ 0,0000	8,887 $\begin{pmatrix} -0,000 \\ 0,000 \end{pmatrix}$ 0,0000
16	11,034 $\begin{pmatrix} -0,001 \\ 0,001 \end{pmatrix}$ 0,0000	11,034 $\begin{pmatrix} -0,002 \\ 0,002 \end{pmatrix}$ 0,0000	11,036 $\begin{pmatrix} -0,002 \\ 0,002 \end{pmatrix}$ 0,0001
17	12,825 $\begin{pmatrix} -0,003 \\ 0,002 \end{pmatrix}$ 0,0001	12,828 $\begin{pmatrix} -0,006 \\ 0,005 \end{pmatrix}$ 0,0002	12,836 $\begin{pmatrix} -0,007 \\ 0,008 \end{pmatrix}$ 0,0006
18	14,070 $\begin{pmatrix} 0,006 \\ 0,003 \end{pmatrix}$ 0,0005	14,089 $\begin{pmatrix} 0,012 \\ 0,007 \end{pmatrix}$ 0,0011	14,133 $\begin{pmatrix} 0,050 \\ 0,006 \end{pmatrix}$ 0,0032
19	14,717 $\begin{pmatrix} 0,064 \\ -0,002 \end{pmatrix}$ 0,0013	14,796 $\begin{pmatrix} 0,134 \\ -0,003 \end{pmatrix}$ 0,0033	14,991 $\begin{pmatrix} 0,324 \\ -0,026 \end{pmatrix}$ 0,0079
20	14,757 $\begin{pmatrix} 0,116 \\ -0,006 \end{pmatrix}$ 0,0029	14,993 $\begin{pmatrix} 0,239 \\ -0,011 \end{pmatrix}$ 0,0071	15,610 $\begin{pmatrix} 0,490 \\ -0,050 \end{pmatrix}$ 0,0142
21	14,221 $\begin{pmatrix} 0,137 \\ -0,009 \end{pmatrix}$ 0,0062	14,754 $\begin{pmatrix} 0,299 \\ -0,030 \end{pmatrix}$ 0,0150	16,210 $\begin{pmatrix} 0,600 \\ -0,116 \end{pmatrix}$ 0,0244
22	13,258 $\begin{pmatrix} 0,166 \\ -0,021 \end{pmatrix}$ 0,0130	14,175 $\begin{pmatrix} 0,414 \\ -0,121 \end{pmatrix}$ 0,0256	16,826 $\begin{pmatrix} 0,741 \\ -0,295 \end{pmatrix}$ 0,0285
23	12,211 $\begin{pmatrix} 0,232 \\ -0,049 \end{pmatrix}$ 0,0251	13,405 $\begin{pmatrix} 0,583 \\ -0,303 \end{pmatrix}$ 0,0291	17,212 $\begin{pmatrix} 0,815 \\ -0,474 \end{pmatrix}$ 0,0211
24	11,579 $\begin{pmatrix} 0,330 \\ -0,047 \end{pmatrix}$ 0,0399	12,697 $\begin{pmatrix} 0,659 \\ -0,329 \end{pmatrix}$ 0,0313	17,055 $\begin{pmatrix} 0,813 \\ -0,458 \end{pmatrix}$ 0,0173
25	11,639 $\begin{pmatrix} 0,597 \\ -0,082 \end{pmatrix}$ 0,0409	12,301 $\begin{pmatrix} 0,751 \\ -0,225 \end{pmatrix}$ 0,0330	16,248 $\begin{pmatrix} 0,849 \\ -0,311 \end{pmatrix}$ 0,0180
26	12,133 $\begin{pmatrix} 1,098 \\ -0,209 \end{pmatrix}$ 0,0223	12,249 $\begin{pmatrix} 0,929 \\ -0,141 \end{pmatrix}$ 0,0215	14,998 $\begin{pmatrix} 0,943 \\ -0,164 \end{pmatrix}$ 0,0140
27	12,535 $\begin{pmatrix} 1,259 \\ -0,196 \end{pmatrix}$ 0,0077	12,309 $\begin{pmatrix} 1,086 \\ -0,107 \end{pmatrix}$ 0,0083	13,695 $\begin{pmatrix} 1,033 \\ -0,086 \end{pmatrix}$ 0,0068
28	12,470 $\begin{pmatrix} 1,149 \\ -0,104 \end{pmatrix}$ 0,0024	12,163 $\begin{pmatrix} 1,125 \\ -0,083 \end{pmatrix}$ 0,0024	12,590 $\begin{pmatrix} 1,075 \\ -0,057 \end{pmatrix}$ 0,0022
29	11,868 $\begin{pmatrix} 1,043 \\ -0,036 \end{pmatrix}$ 0,0007	11,634 $\begin{pmatrix} 1,070 \\ -0,044 \end{pmatrix}$ 0,0007	11,637 $\begin{pmatrix} 1,068 \\ -0,040 \end{pmatrix}$ 0,0006
30	10,907 $\begin{pmatrix} 1,005 \\ -0,008 \end{pmatrix}$ 0,0002	10,781 $\begin{pmatrix} 1,018 \\ -0,014 \end{pmatrix}$ 0,0002	10,701 $\begin{pmatrix} 1,031 \\ -0,018 \end{pmatrix}$ 0,0002
31	9,898 $\begin{pmatrix} 0,997 \\ 0,000 \end{pmatrix}$ 0,0000	9,850 $\begin{pmatrix} 1,000 \\ -0,001 \end{pmatrix}$ 0,0000	9,802 $\begin{pmatrix} 1,005 \\ -0,003 \end{pmatrix}$ 0,0000
32	9,154 $\begin{pmatrix} 0,998 \\ 0,002 \end{pmatrix}$ 0,0000	9,142 $\begin{pmatrix} 0,998 \\ 0,002 \end{pmatrix}$ 0,0000	9,127 $\begin{pmatrix} 0,999 \\ 0,002 \end{pmatrix}$ 0,0000
33	8,833 $\begin{pmatrix} 0,999 \\ 0,003 \end{pmatrix}$ 0,0000	8,831 $\begin{pmatrix} 0,999 \\ 0,002 \end{pmatrix}$ 0,0000	8,829 $\begin{pmatrix} 0,999 \\ 0,002 \end{pmatrix}$ 0,0000
34	8,877 $\begin{pmatrix} 1,000 \\ 0,002 \end{pmatrix}$ 0,0000	8,877 $\begin{pmatrix} 1,000 \\ 0,002 \end{pmatrix}$ 0,0000	8,877 $\begin{pmatrix} 1,000 \\ 0,002 \end{pmatrix}$ 0,0000
35	9,156 $\begin{pmatrix} 1,001 \\ 0,002 \end{pmatrix}$ 0,0000	9,156 $\begin{pmatrix} 1,001 \\ 0,002 \end{pmatrix}$ 0,0000	9,156 $\begin{pmatrix} 1,001 \\ 0,002 \end{pmatrix}$ 0,0000

Tabelle 7.1: Verschiebung des Maximums des Diskontinuitätsmaßes bei zunehmendem Hel-
 ligkeitsunterschied zwischen den sich unterschiedlich bewegenden Bildbereichen. Im darge-
 stellten Bild verläuft die Objektbildkante senkrecht zwischen den (der Einfachheit halber
 nur durch ihren Spaltenindex bezeichneten) Pixeln 24 und 25. Der linke Bereich ist stationär,
 der rechte Bereich bewegt sich mit 1 Pixel pro Bild nach rechts. Die Pfeile markieren die
 untersuchten Bereich der 12. Bildzeile.

7.1.2 Unerwartete Diskontinuitäten

Immer wieder treten bei der Berechnung von OF-Vektor-Feldern unerwartete Diskontinuitätsflecken auf, d. h. Ansammlungen von OF-Diskontinuitätshinweisen, die von Vektoren der Kategorie „ROF“ umschlossen sind. Diese Beobachtung wurde vor allem bei der Fahrzeugverfolgung auf der Bildfolge `reading` (oder der verkleinerten Version `reading_ua`) gemacht, vereinzelt aber auch bei den `dtneu_` und `stau`-Bildfolgen. Obwohl bei den `reading`-Bildfolgen ein Einfluss durch die JPEG-Codierung, durch uns nicht bekannte Störfaktoren (z. B. Vermreinigungen der Scheibe, durch die die Aufnahme gemacht wurde) oder durch die Verkleinerung des Bildes nicht ausgeschlossen werden kann, so konnten doch drei mögliche Ursachen für diese Probleme ausgemacht werden:

- Eine mögliche Ursache für Diskontinuitätshinweise inmitten des Fahrzeugabbildes sind **Verdeckungen** durch mit bloßem Auge nur schwer erkennbare Objekte (z. B. Straßenbahnoberleitungen, Befestigungsdrähte für Straßenbelichtung, schlanke Äste).

Ein Beispiel dafür liefert Fahrzeug 21 in der Bildfolge `reading_ua` zwischen den Bildern 400 und 440. Bei Betrachtung der OF-Schätzung fallen Diskontinuitätshinweise auf (siehe Abbildung 7.4), die zunächst nicht erklärbar erscheinen. Die Tatsache, dass die Diskontinuitätshinweise immer an der gleichen Position im Bildbereich auftreten und nicht mit dem Fahrzeugabbild „mitwandern“, deutet aber auf eine stationäre Verdeckung hin. Erst bei Betrachtung des nicht-verkleinerten Originalbildes (Abbildung 7.5) zeigt sich, dass Teile eines Baumes das Fahrzeug verdecken. Es liegt also eine Verletzung von Annahme 8 vor. Die Grauwerte des Baumes unterscheiden sich – auch bedingt durch die starke JPEG-Komprimierung der Bilder – nur so wenig von den Grauwerten des Fahrzeugabbildes, dass die Verdeckung mit bloßem Auge kaum zu erkennen ist. Die Diskontinuitätshinweise werden also durch tatsächlich vorhandene Diskontinuitäten verursacht und sind somit berechtigt. Hieran zeigt sich die hohe Sensitivität des Verfahrens.

- Als zweite Ursache für unerwartete Diskontinuitätshinweise kommen **Veränderungen des Fahrzeugabbildes** in Betracht. Besonders deutlich zeigt sich dieses Problem bei Fahrzeug 21 der Bildfolgen `reading` und `reading_ua` (siehe Abbildung 7.6). Dieses Fahrzeug biegt mit hoher Geschwindigkeit ab. Innerhalb weniger Bilder wird das Abbild der Seitenscheiben durch Reflexionen deutlich heller, um dann wieder dunkler zu werden. Dadurch wird Annahme 4 verletzt. Gleichzeitig verschnindet durch die Fahrzeugdrehung die Windschutzscheibe, während durch die Seitenscheiben helle Objekte aus dem Fahrzeuginnern (vermutlich Teile des Armaturenbretts) durchscheinen.
- Es verbleibt eine dritte Gruppe von Diskontinuitätsflecken im Innern des Fahrzeugabbildes, die mit den bisherigen Hypothesen nicht erklärt werden können.

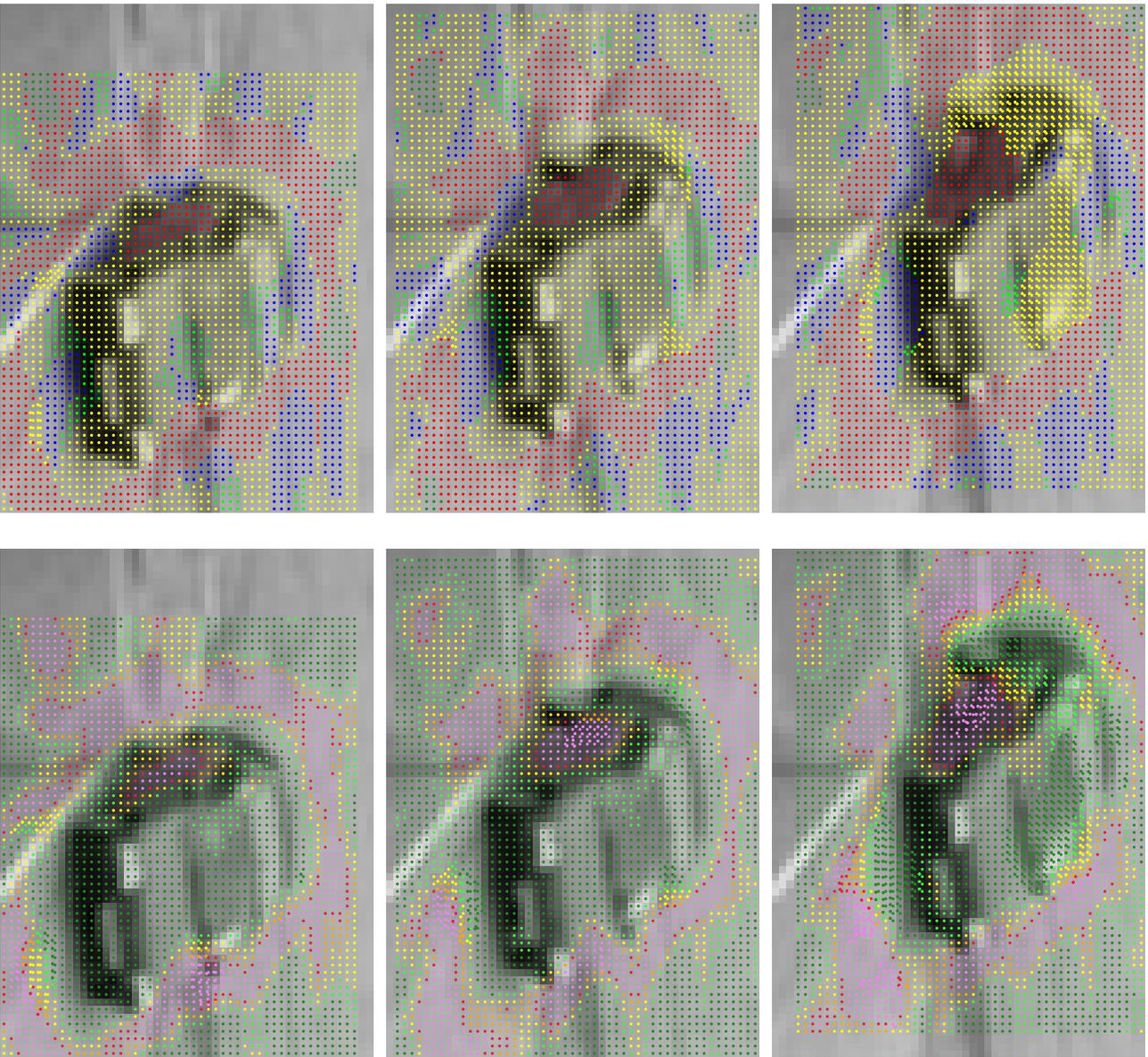


Abbildung 7.4: OF-Diskontinuitäten im Innern des Abbildes von Fahrzeug 21 in den Bildern 410, 420 und 430 der Bildfolge *reading.ya*. Die Diskontinuität ist relativ zum Bildkoordinatensystem nahezu stationär, was auf eine Verdeckung durch ein ortsfestes Objekt (stationäre Kamera) hindeutet.

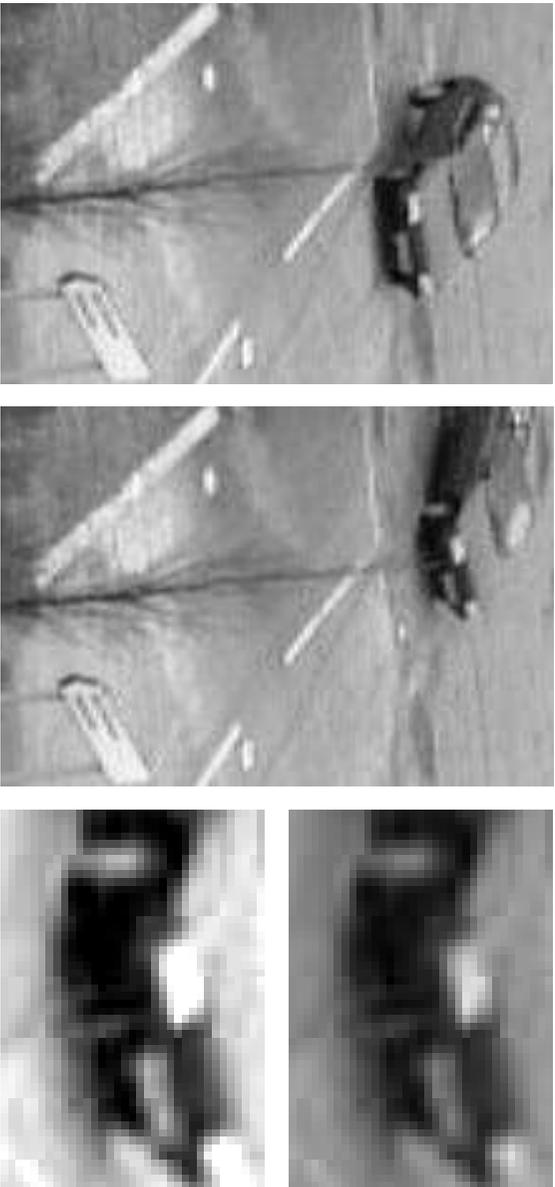


Abbildung 7.5: Diese Ausschnitte aus Aufnahmen der Bildfolge *reading* zeigen den Baum, von dem einige Äste vor das Fahrzeug 21 ragen und so die Diskontinuitätsflecken verursacht. In Bild 410 (links) – zu diesem Zeitpunkt ist der Diskontinuitätsflecken am deutlichsten ausgeprägt – sind die Äste mit bloßem Auge nicht zu erkennen. In Bild 385 (Mitte) dagegen hebt sich der Baum hell gegen die dunkle Front des Fahrzeuges ab, wie die Ausschnitte rechts – der untere mit verstärktem Kontrast – verdeutlichen. Die Spitze des Baumes befindet sich unter dem rechten Rand des Scheinwerfers in Höhe des Kennzeichens. Die Diskontinuitätsflecke werden also durch tatsächlich vorhandene Diskontinuitäten verursacht und sind somit berechtigt und ein Zeichen für die hohe Sensitivität des verwendeten Algorithmus.

Dabei handelt es sich um meist nur kleine Diskontinuitätsflecken von weniger als 20 Pixeln, die zudem stark von der Wahl der Parameter abhängen, wie Abbildung 7.7 illustriert. Eine Betrachtung der zugrundeliegenden Bilder zeigt, dass die betroffenen Bereiche eine im Vergleich zum Rauschen nur sehr **schwache** oder gar keine offensichtliche **Änderung der Grauwertstruktur** aufweisen. Das Auftreten solcher unerwünschter Diskontinuitätsflecken konnte durch die quasi-inverse Adaption im Vergleich zur linearen Adaption leicht reduziert werden.

7.1.3 Löcher in der Diskontinuitätsmauer

Im Idealfall werden zwei sich unterschiedlich bewegende Objektabbilder durch einen Diskontinuitätsbereich vollständig voneinander getrennt. In diesem Fall bilden die Diskontinuitätslinien eine „Diskontinuitätsmauer“. Abbildung 7.8 zeigt, warum es aber an manchen Stellen trotz unterschiedlicher Bewegung nicht zu Diskontinuitätslinien kommen kann. Ein Punkt wird als Diskontinuitätsliniensweise kategorisiert, wenn der kleinste, genormte (d. h. durch die Spur geteilte) Eigenwert von $\nabla g(\nabla g)^T$ einen Schwellwert überschreitet. Jeder lokale Tensor $\nabla g(\nabla g)^T$ des Mittelungsbereichs trägt zu dieser

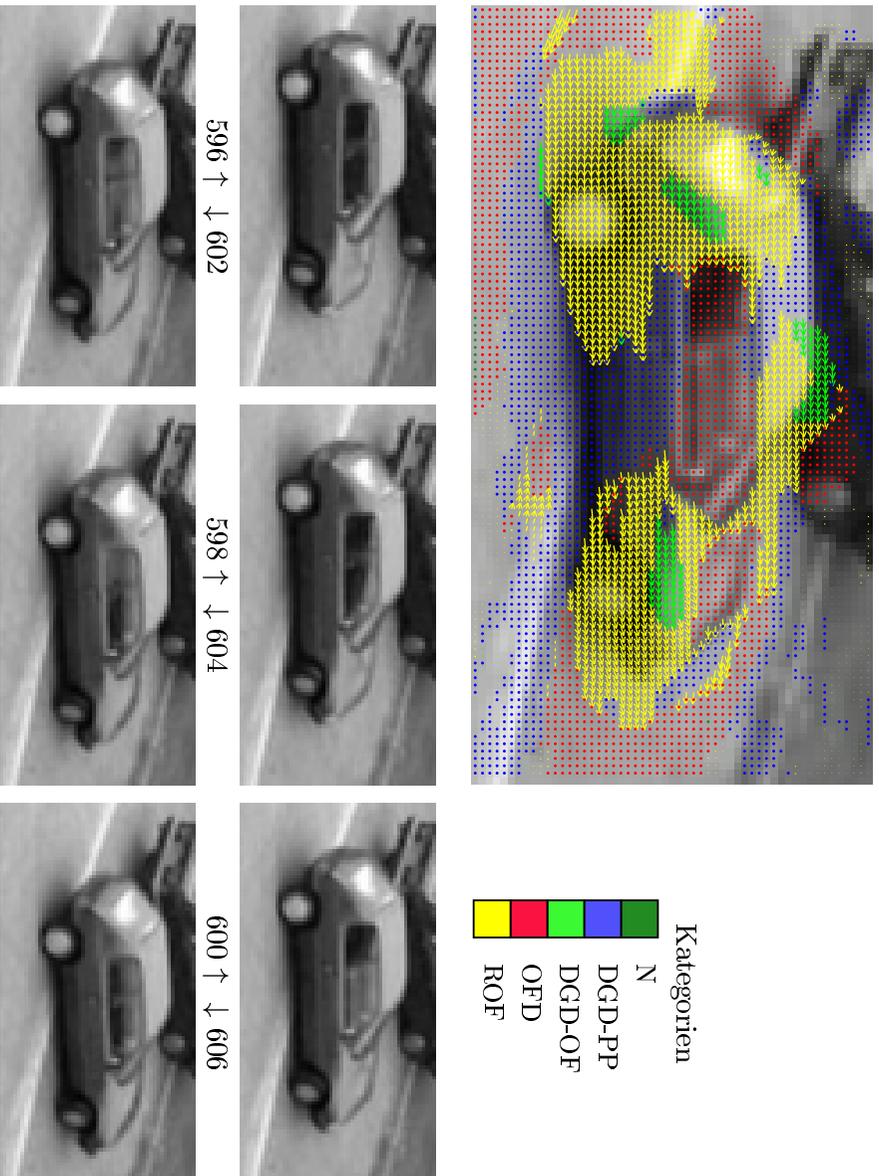


Abbildung 7.6: Das obere Bild zeigt den geschätzten OF für Fahrzeug 21 in Bild 600 der Bildfolge `reading.ra`. Auffällig sind die Diskontinuitätshinweise im Bereich der Seitenscheiben und der Windschutzscheibe. Die Ausschnitte aus dem Bildern 596 bis 606 (im Abstand von zwei Bildern) verdeutlichen, wie es zu diesen Diskontinuitätshinweisen kommt. Innerhalb weniger Bilder wird das Abbild der Seitenscheiben durch Reflexionen deutlich heller, um dann wieder dunkler zu werden. Gleichzeitig verschwindet durch die Fahrzeugdrehung die Windschutzscheibe, während durch die Seitenscheiben helle Objekte aus dem Fahrzeuginnern (vermutlich Teile des Armaturenbretts) durchscheinen.

Schwellwertüber- oder -unterschreitung mehr oder weniger bei. Man erkennt aber, dass im Hintergrund eine Kante nahezu in Bewegungsrichtung des Vordergrundobjektes verläuft. Der maximale Gradient zeigt also in der Bildebene in eine Richtung senkrecht zur Bewegungsrichtung, was dazu führt, dass die Invarianzebene mit der durch die VR-Vektoren von Vordergrund- und Hintergrundobjekt aufgespannten Ebene fast übereinstimmt. Der kleinste Eigenwert bleibt sehr klein, es kommt nicht zum Diskontinuitätshinweisen.

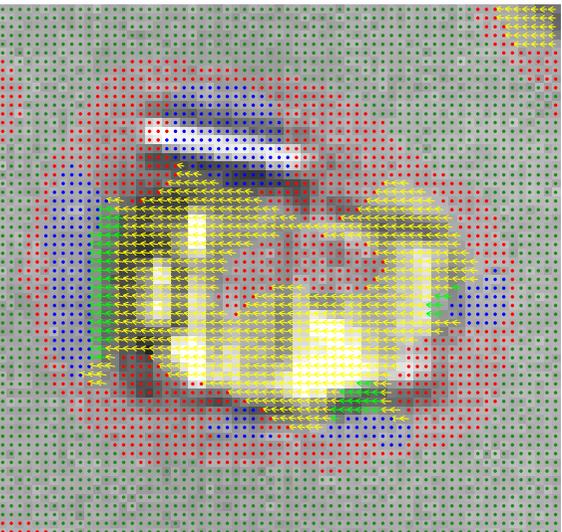
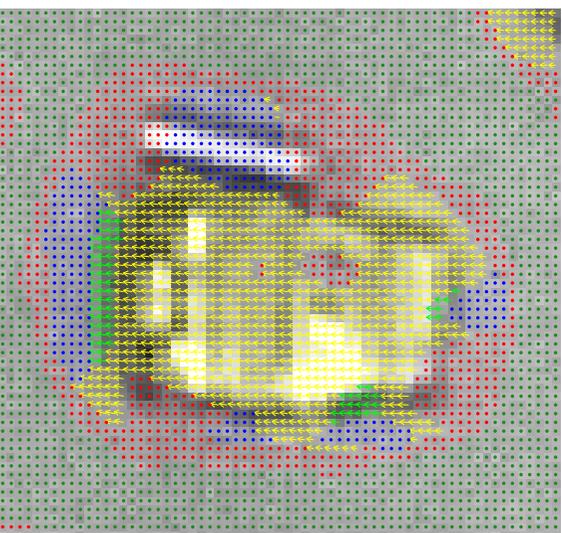
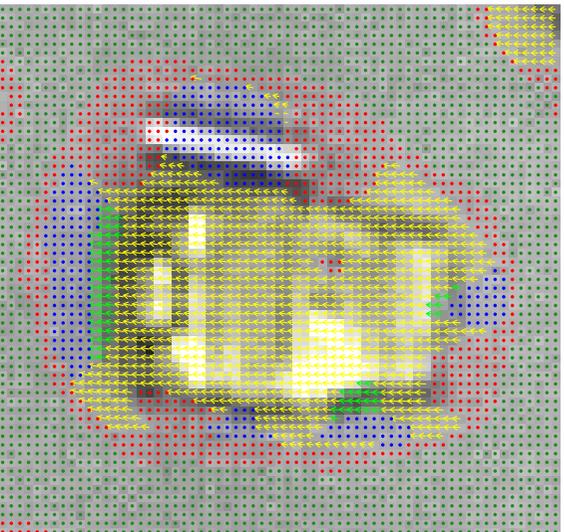
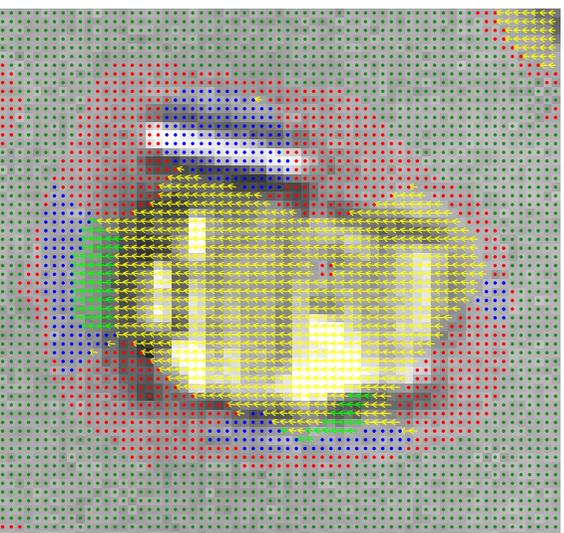
lineare Adaption, $\tau_{\text{OFD}} = 0,0060$ lineare Adaption, $\tau_{\text{OFD}} = 0,0120$ lineare Adaption, $\tau_{\text{OFD}} = 0,0180$ quasi-inverse Adaption, $\tau_{\text{OFD}} = 0,0060$

Abbildung 7.7: Parameterabhängigkeit der Diskontinuitätsflecken im Innern des Fahrzeugabbildes 7 in Halbfolge dtneu01: Bei linearer Adaption – dem „Standard-Verfahren“ vor Beginn der in dieser Arbeit beschriebenen Untersuchungen – überschreitet das Diskontinuitätsmaß an zahlreichen Punkten im Innern des Fahrzeugabbilds den Diskontinuitäts-Schwellwert. Beim „Standard-Wert“ von $\tau_{\text{OFD}} = 0,0120$ tritt dies bei 20 Punkten ein, selbst bei einem um 50% erhöhten Schwellwert von $\tau_{\text{OFD}} = 0,0180$ sind noch fünf Pixel betroffen. Beim neu eingeführten quasi-inversen Verfahren hingegen überschreiten selbst für $\tau_{\text{OFD}} = 0,0060$ nur vier Pixel den Schwellwert, während bei einem so niedrigen Schwellwert beim linearen Verfahren 110 Punkte den Schwellwert überschreiten. Insgesamt ist der Bereich der ROF-Vektoren beim quasi-inversen Verfahren weniger „ausgefranst“ als bei linearen Adaptionsverfahren. Sonstige Parameter: Maskengröße $19 \times 19 \times 19$, eine doppelt-adaptive Iteration, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$, $\tau_N = 2,5$, $\tau_{\text{DGD}} = 0,050$.

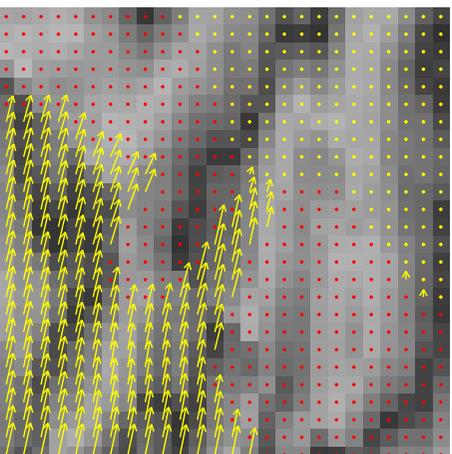


Abbildung 7.8: Dieser Ausschnitt aus Bild 150 der Bildfolge `marmor-stat` illustriert, warum die Diskontinuitätsmaner zwischen Hintergrund (im Ausschnitt oben links) und bewegtem Vordergrund (im Ausschnitt unten rechts) an manchen Stellen nicht geschlossen ist. Zu Zwecken dieser Demonstration wurde der Schwellwert τ_{OFD} auf das vierfache des üblichen Wertes angehoben. Man erkennt, dass im Hintergrund eine Kante nahezu in Bewegungsrichtung des Vordergrundobjektes verläuft. Der maximale Gradient zeigt also in der Bildebene in eine Richtung senkrecht zur Bewegungsrichtung, was dazu führt, dass die Invarianzebene mit der durch die VR-Vektoren von Vordergrund- und Hintergrundobjekt aufgespannten Ebene fast übereinstimmt. Der kleinste Eigenwert bleibt sehr klein, es kommt nicht zum Diskontinuitätshinweis.

7.1.4 Integration von Diskontinuitäten in die OF-Schätzung

7.1.4.1 Motivation

Wie Tabellen 6.7 und 6.8 gezeigt haben, führt die adaptive GST-Schätzung zu einer Verbesserung der OF-Schätzung an den als ROF klassifizierten Pixeln. Der Preis dafür sind allerdings wesentlich breitere Diskontinuitätsbereiche, die eine präzise Segmentierung auf Basis der vorliegenden GSTs an den Grenzen der Objektbilder nahezu unmöglich machen.

Eine Verbesserung der Ergebnisse ist auf drei Arten denkbar:

1. durch Anwendung einer Verfahrens, für das die zuvor extrahierte Bewegungsinformation dennoch ausreichend für eine zuverlässige Segmentierung ist,
2. durch Segmentierung *vor* Durchführung der Bewegungsschätzung oder mit wesentlichem Anteil der nicht aus der Bewegung gewonnenen Information, oder
3. durch gleichzeitige (oder ineinander verschränkte) Bewegungsschätzung und Segmentierung.

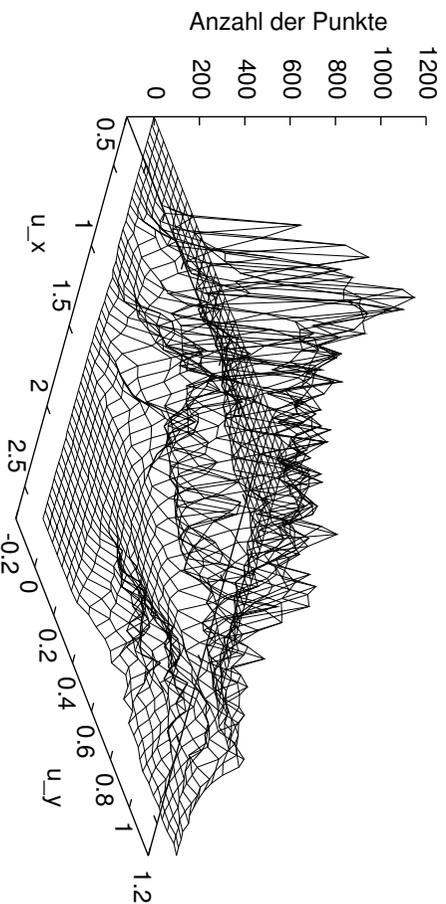


Abbildung 7.9: Histogramm der Komponenten der OF-Vektoren für Bild 35 der Bildfolge `tr_bstb`. Eine klare Trennung zwischen den einzelnen Spitzen des Histogramms ist nicht möglich.

In die erste Gruppe fällt die Arbeit von [Wiskott 99]. Der Autor legt ein Histogramm aller im Bild geschätzten Bewegungen an und nutzt die Spitzen, um die Hauptbewegungsrichtungen im Bild zu bestimmen. Abbildung 7.9 demonstriert das Hauptproblem solcher Ansätze, die erst nach der Bewegungsschätzung aufsetzen: Unterscheiden sich die Bewegungen der verschiedenen Objektabbilder nicht hinreichend voneinander, ist die Bewegungsschätzung durch Rauschen stark gestört oder wird ein fließender Übergang der Bewegung zwischen zwei Objektabbildern geschätzt, so kann ein Scheitern solcher Ansätze kaum verhindert werden.

Aus der zweiten Gruppe werden exemplarisch zwei Ansätze betrachtet: [Kim et al. 01] segmentieren Bilder einer Videosequenz durch Minimierung einer Energiefunktion, die jedes Pixel einer Ballung zuordnet. Die zur Demonstration dieser Vorgehensweise verwendeten Bilder sind allerdings sehr kontraststark, so dass die Wirksamkeit des Algorithmus auf verrauschten oder kontrastschwachen Bildern nicht beurteilt werden kann. [Thompson 98] schlägt als Lösung zur exakteren Lokalisierung der Objektgrenze vor, davon auszugehen, dass Objektkanten immer mit deutlichen Grauwertübergängen zusammenfallen. Erfahrungen aus der Fahrzeugverfolgung zeigen aber, dass oft der Kontrast zwischen bewegten Objekt und Hintergrund sehr gering ist. Selbst beim vorangehend betrachteten Beispiel ist der Helligkeitsunterschied an Kante 2 (vgl. Abbildung 7.1) gering im Vergleich zu den sonst im Bild auftretenden Unterschieden. Ein Vorgehen wie von [Thompson 98] vorgeschlagen würde daher also entweder eine tatsächlich vorliegende Bewegungsdiskontinuität nicht als solche erkennen, weil die Grauwertkante fehlt, oder wegen einer Vielzahl von Grauwertkanten im Bild zu einer Vielzahl von Segmentierungshypothesen führen, auf deren Behandlung der Autor nicht weiter eingeht.

Von den verbliebenen Ansätzen, die Segmentierung und Bewegungsschätzung zusammenfassen, modellieren jedoch die meisten die Diskontinuität in irgendeiner Art und Weise, z. B. durch Niveau-Schnitte ([Brox et al. 03]). Bei dem hier vorgeschlagenen Ansatz werden keine Annahmen über Form oder Lage der Objektgrenzen oder über Anzahl oder Größe der Objekte getroffen. Es wird auch nicht angenommen, dass die Grenze zwischen den Objektabbildern vollständig im Bildausschnitt zu sehen ist. Stattdessen werden nur die im GST schon vorhandenen, aber bisher nicht genutzten Informationen verwendet, um die Ausdehnung der Diskontinuitäten auf möglichst kleine Bereiche zu beschränken. Es handelt sich also um einen streng lokalen Ansatz. Der Nachteil des vorgeschlagenen Ansatzes besteht darin, dass er keine „fertige Segmentierung“ liefert, sondern nur soviel Diskontinuitätshinweise erzeugt, wie es für die Verbesserung der OF-Schätzung unbedingt nötig ist.

7.1.4.2 Ansatz

Seien $\lambda_1 > \lambda_2 > \lambda_3$ die Eigenwerte des GST und $\tilde{\lambda}_i = \lambda_i/\text{spur}(\text{GST})$ die normierten Eigenwerte, dann liefert $\delta := 3,0 \cdot \tilde{\lambda}_3$ ein Maß für die Diskontinuität an einem Pixel, mit möglichen Werten von 0,0 bis 1,0. Ein kleiner Wert deutet darauf hin, dass alle Pixel im Mittelungsbereich des GST im Wesentlichen derselben Bewegung unterliegen; ein großer Wert weist auf unterschiedliche Bewegungen hin. Als Diskontinuitätsschwelle θ_d wird bisher meist 0,0120 angenommen, d. h. überschreitet d diesen Wert, so wird von einer Diskontinuität ausgegangen. Der durchschnittliche Wert von d liegt meist bei unter 0,01, selbst in Diskontinuitätsbereichen überschreitet er selten 0,1. Bei ROF-Vektoren liegt er meist bei weniger als 0,001, der maximale bisher beobachtete Wert beträgt 0,78.

Der auf diesen Beobachtungen basierende Algorithmus wird in drei zunehmend komplexeren Varianten entwickelt, von der bereits die erste zu bemerkenswerten Ergebnissen führt.

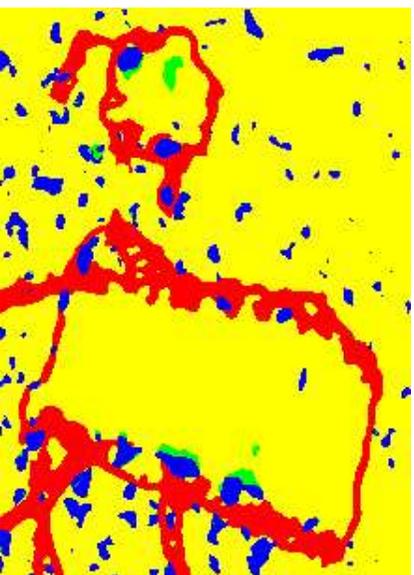
Variante 1 Die Grundidee besteht darin, die aus vorangehenden Iterationen verfügbare Diskontinuitätsinformation in den Folgeiterationen sofort auszunutzen, anstatt – wie bisher – erst das Ergebnis der mehrfachen Iteration zu kategorisieren. Dazu wird im Mittelungsschritt der sich aus der Gaußlocke ergebende Gewichtungsfaktor ω für jeden Tensor mit einem Faktor zwischen 0 und 1 multipliziert, der aus dem Diskontinuitätsmaß abgeleitet ist, mit einem Wert nahe 1 für Punkte mit sehr kleinem $\tilde{\lambda}_3$ und kleineren Werten bei größerem Diskontinuitätsmaß. Die in diesem Fall gewählte Funktion lautet:

$$\omega = e^{-d \cdot \zeta / \tau_{\text{OPD}}} . \quad (7.1)$$

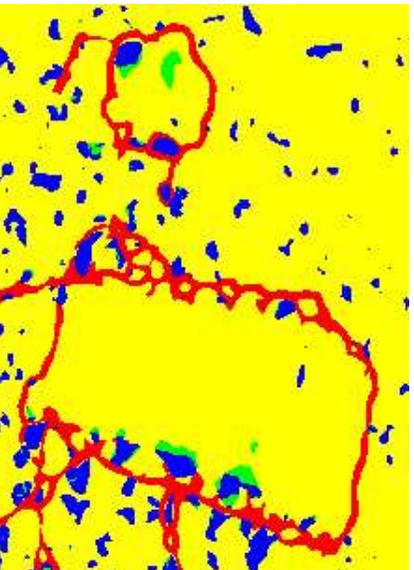
τ_{OPD} bezeichnet den OFD-Schwellwert, der Faktor ζ ermöglicht, den Einfluss der Diskontinuitäten auf die Mittelung zu steuern. Bisher wurde ζ so gewählt, dass ω den Wert 0,9 für $d = \tau_{\text{OPD}}$ annimmt.



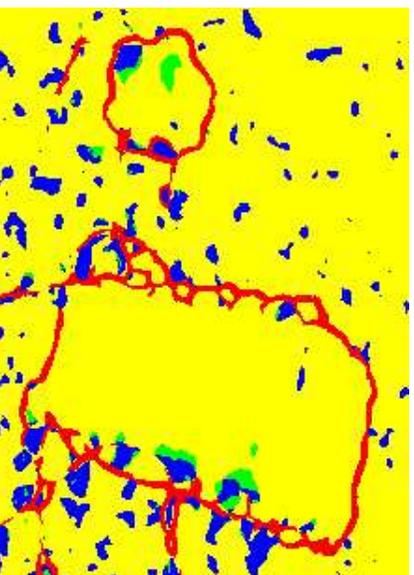
Vergleichsbild



Iteration 10



Iteration 50



Iteration 100

Abbildung 7.10: Integration der Diskontinuitätshinweise in die OF-Schätzung (Variante 1) am Beispiel der Bildfolge *marmor-stat*: Im Vergleich zum bisher erzielten Ergebnis (oben links) ergibt sich nach 10 (oben rechts), 50 (unten links) oder 100 Iterationen (unten rechts) ein deutlich schmalerer Diskontinuitätsbereich. Obwohl der Diskontinuitätsbereich durchschnittlich nur noch drei bis vier Pixel breit ist, ist er an keiner Stelle unterbrochen. Feste Kovarianzmatrizen.

In der bisher beschriebenen Form führt das Verfahren aber dazu, dass das Diskontinuitätsmaß insgesamt abnimmt, was im folgenden Iterationsschritt zu einer stärkeren Berücksichtigung von Punkten führt, die zuvor noch als Diskontinuität erkannt worden waren. Vor allem bei großen Masken kann es daher zu Oszillationen zwischen Iterationen mit sehr schmalen und sehr breiten Diskontinuitätsbereichen kommen. Dies lässt sich dadurch vermeiden, dass an allen Punkten die Diskontinuitätsmaße aller bisheriger Iterationen aufsummiert werden.

Variante 2 Vor allem bei größeren Mittelungsmasken, wie sie bei der Adaption auftreten, kann es vorkommen, dass „über Diskontinuitäten hinweg“ gemittelt wird. In

alle Vektoren				nur ROF-Vektoren aus bewegtem Klotz			
It.	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kumulativ)	\emptyset Fehler	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kumulativ)	\emptyset Fehler	Ref. Fehler
0	66880	30289 37535 52285	0,1739	14766	70 4355 14182	0,0272	—

keine Adaption, $\sigma_x = \sigma_y = \sigma_z = 1,41$ (V 199)

1	66880	30556 38077 52999	0,1668	15083	68 4397 14414	0,0288	0,0301
2	66880	30752 38456 53458	0,1608	15247	71 4441 14551	0,0292	0,0315
3	66880	30887 38708 53801	0,1568	15412	73 4458 14655	0,0302	0,0339
10	66880	31387 39750 55285	0,1425	16044	68 4549 15112	0,0356	0,0482
50	66880	32252 41700 58137	0,1211	17022	67 4699 15900	0,0392	0,0757

Tabelle 7.2: [Marmor_stat, Bild 150] Auswirkungen der Integration der Diskontinuitätshinweise in die OF-Schätzung (Variante 1): Der durchschnittliche OF-Fehler aller Vektoren nimmt deutlich ab, der durchschnittliche Fehler der ROF-Vektoren nimmt zwar zu, aber weit weniger stark als der Referenzfehler (= Fehler derselben Punkte im 0. Schritt). Von 19143 Punkte, die zu einem bewegten Marmor-Block gehören, werden letztendlich 17022 (das entspricht knapp 89%) der Kategorie ROF zugewiesen.

0.7	0.2	0.3	0.8	1.0	0.2	0.2	0.3	0.8	1.0
0.6	0.2	0.5	0.9	1.0	0.2	0.2	0.5	0.9	1.0
0.5	0.2	1.0	1.0	1.0	0.2	0.2	1.0	1.0	1.0
0.4	0.3	1.0	1.0	1.0	0.3	0.3	1.0	1.0	1.0
0.4	0.3	1.0	1.0	1.0	0.3	0.3	1.0	1.0	1.0

Abbildung 7.11: Vergleich Variante 1 vs. Variante 2 (schematisch): Links dargestellt sind die Faktoren ω , die sich für die 5×5 im Mittelungsbereich enthaltenen Punkte gemäß Formel (7.1) ergeben. Die kleinsten Werte (entsprechend dem Maximum des Diskontinuitätsmaßes) finden sich in der 2. Spalte. Rechts die angepasste Diskontinuitätsmaske, die berücksichtigt, dass nicht über Diskontinuitäten hinweg gemittelt werden sollte.

einer verbesserten Version des Algorithmus wird sichergestellt, dass ω mit wachsender Entfernung vom Maskenmittelpunkt nicht zunehmen kann. Dies geschieht durch Verwendung einer zweidimensionalen „Diskontinuitätsmaske“, die vom Maskenmittelpunkt her mit den Werten für ω aufgefüllt wird, wobei aber nie ein neu eingetragener Wert für ω größer sein darf als einer der für die (näher am Maskenmittelpunkt liegenden) Nachbarpunkte schon eingetragenen Werte (Abbildung 7.11). Diese Diskontinuitätsmaske wird im Mittelungsschritt punktweise mit der Tiefpassmaske multipliziert. Durch diese Erweiterung steigt der Rechenaufwand stark an, was sich aber nicht in einer signifikanten Verbesserung der Ergebnisse auszahlt.

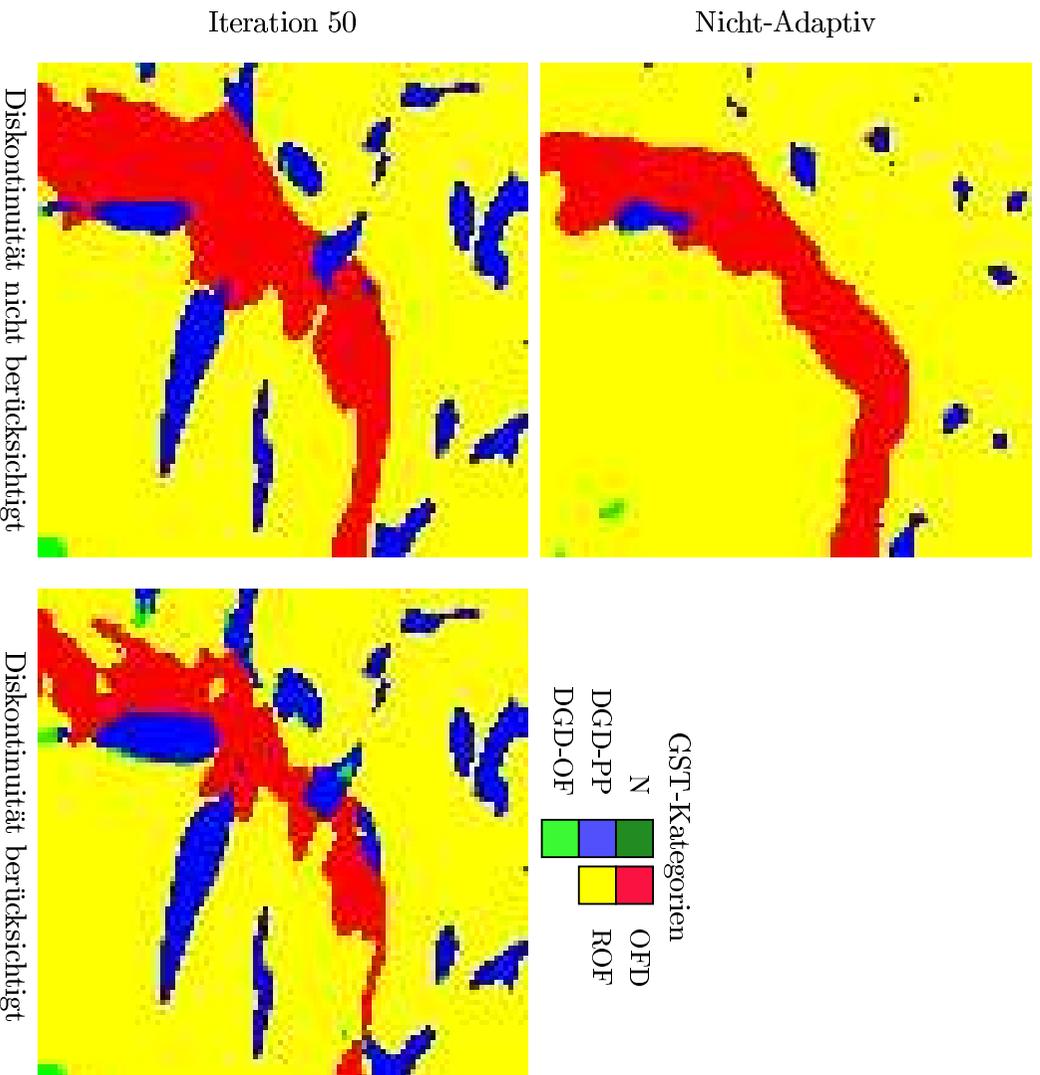


Abbildung 7.12: Integration der Diskontinuitätshinweise in die OF-Schätzung (Variante 2): Bei adaptiver Anpassung der Masken und gleichzeitiger Berücksichtigung der Diskontinuitätshinweise (unten rechts) ergeben sich schmalere Diskontinuitätsbereiche als ohne Berücksichtigung der Diskontinuitätshinweise nicht-adaptiv (oben links) oder adaptiv (unten links). Jeweils große Start-Kovarianzmatrix, kleiner Skalierungsbereich, quasi-inverse Adaption.

Variante 3 Es wird nun zusätzlich berücksichtigt, dass sich die Diskontinuität bewegt, indem die bei Variante 2 ermittelte Diskontinuitätsmaske für Zeitpunkte, die nicht dem aktuellen Bild entsprechen, entsprechend dem am Aufpunkt geschätzten OF verschoben wird. Die Ergebnisse sind aber in diesem Fall nicht besser, sondern z. T. sogar schlechter als bei Variante 2. Der Grund für diese Verschlechterung liegt vermutlich darin, dass die Bewegungsinformation im Bereich der Diskontinuitäten sehr unzuverlässig ist. Es erscheint daher sinnvoller, auf die Berücksichtigung der Bewegung der Diskontinuität zunächst zu verzichten.

	alle Vektoren			nur ROF-Vektoren aus bewegtem Klotz			
It.	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kumulativ)	\emptyset Fehler	Anz. Vekt.	Fehler $\leq \dots$ Pixel/Bild $1/1000$ $1/100$ $1/10$ (kumulativ)	\emptyset Fehler	Ref. Fehler
0	10000	4286 5484 8596	0,1429	3664	12 901 3597	0,0247	—
„quasi-inverse“ Adaption <i>ohne</i> Berücksichtigung der Disk., $\sigma_x = \sigma_y = \sigma_t = 1,41$							
50	10000	3243 5665 8165	0,1690	3283	29 1670 3109	0,0329	0,0357
„quasi-inverse“ Adaption <i>mit</i> Berücksichtigung der Disk., $\sigma_x = \sigma_y = \sigma_t = 1,41$							
50	10000	3370 5908 8502	0,1502	3469	37 1709 3279	0,0241	0,0416

Tabelle 7.3: [marmor_stat, Bild 150] Auswirkungen der Integration der Diskontinuitäts-hinweise in die adaptive OF-Schätzung: Der durchschnittliche OF-Fehler aller Vektoren nimmt nach 50 Iterationen (doppelt-adaptives Verfahren) im Vergleich zur Adaption ohne Berücksichtigung der Diskontinuitätshinweise ab, der durchschnittliche Fehler der ROF-Vektoren nimmt zwar zu, aber weit weniger stark als der Referenzfehler (= Fehler derselben Punkte im 0. Schritt).

7.1.5 Ergebnisse

Abbildung 7.10 und Tabelle 7.2 zeigen die mit diesem Verfahren bei Variante 1 erzielten Ergebnisse. Im Vergleich zu den bisherigen Ergebnissen erkennt man deutlich schmalere, aber nicht unterbrochene Diskontinuitätsbereiche. Bei Anwendung von Variante 2 ergeben sich die in Abbildung 7.12 und Tabelle 7.3 dargestellten Resultate. Auch dieses Beispiel zeigt eine Verbesserung der Segmentierung.

Tabelle 7.4 stellt das Ergebnis der OF-Schätzung bei adaptiver GST-Bestimmung mit gleichzeitiger Berücksichtigung der Diskontinuitäten gemäß Variante 1 dar. Im Vergleich zu Tabelle 6.23 zeigt sich eine weitere Verbesserung der Ergebnisse. Insbesondere die Durchschnittswerte der Fehler nehmen ab, während die Mediane (fast) unverändert bleiben. Dies ist darauf zurückzuführen, dass vor allem OF-Schätzungen, die zuvor aufgrund der Nähe zu Bewegungsdiskontinuitäten besonders schlecht waren, verbessert werden. Angesichts der geringen Anzahl der Punkte, an denen sich die Berücksichtigung der Diskontinuitäten überhaupt auswirkt, ist die Verbesserung der Durchschnittswerte bemerkenswert. Beim Vergleich mit den Ergebnissen anderer Autoren ist zu bedenken, dass die Verfahren mit besseren Ergebnissen (z. B. [Farnebäck 01]) oft von wesentlich weitreichenderen Annahmen wie affiner Bewegung in relativ großen Bildbereichen ausgehen. Insofern sind die im Rahmen dieser Arbeit erzielten Ergebnisse durchaus „konkurrenzfähig“.

7.1.6 Fazit

Durch Analyse der zugrunde liegenden Grauwertstruktur konnten mehrere Problemfälle im Zusammenhang mit Diskontinuitäten erklärt werden. Es wurde ein Verfahren vor-

Verfahren	Maskengröße		$\psi_{v+,u+}$	$\ v - u\ _2$	Dichte		
	Ableitung	Mittelung	\emptyset Med.	\emptyset Med.			
yosemite							
(4) alle	$15 \times 15 \times 15$	$15 \times 15 \times 1$	2,28	1,64	0,0973	0,0639	100
(4) nur ROF	$15 \times 15 \times 15$	$15 \times 15 \times 1$	2,01	1,58	0,0741	0,0532	71,2
(4) alle (τ_{OFD} verkl.)	$15 \times 15 \times 15$	$15 \times 15 \times 1$	2,28	1,64	0,0964	0,0639	100
(4) nur ROF (τ_{OFD} verkl.)	$15 \times 15 \times 15$	$15 \times 15 \times 1$	1,91	1,56	0,0718	0,0523	69,3
tr_bsb							
(1) alle	$19 \times 19 \times 15$	$19 \times 19 \times 15$	3,20	1,02	0,1378	0,0415	100
(1) nur ROF	$19 \times 19 \times 15$	$19 \times 19 \times 15$	1,56	0,91	0,0677	0,0346	77,5
(3) alle	$19 \times 19 \times 19$	$19 \times 19 \times 1$	3,01	1,01	0,1271	0,0407	100
(3) nur ROF	$19 \times 19 \times 19$	$19 \times 19 \times 1$	1,71	0,89	0,0713	0,0350	79,0

Tabelle 7.4: Wie Tabelle 6.23, bei gleichzeitiger Berücksichtigung der Diskontinuitäten. Doppelt-adaptives Verfahren, quasi-inverse Adaption, Start-Kovarianz: $\Sigma_A = \text{diag}(1,41^2; 1,41^2; 1,41^2)$, $\Sigma_M = 2\Sigma_A$, 10 Iter., 50% Dämpfung, $\sigma_{\min} = 1,0$, $\sigma_{\max} = 2,85$.

gestellt, dass durch Ausnutzung der Diskontinuitätshinweise bei der OF-Schätzung die Qualität der Kategorisierung erheblich verbessert, bei gleichzeitiger Verbesserung der OF-Schätzung in den Bereichen, in denen sich durch diese Integration die Kategorie von OFD nach ROF ändert.

7.2 Aufbau einer Bildpyramide

Das Verfahren zur OF-Schätzung in der bisher beschriebenen Form ist noch von zahlreichen Parametern abhängig, die vom Benutzer vorgegeben werden müssen. Besondere Bedeutung kommt dabei den Parametern σ_x , σ_y , σ_t sowie bei adaptiver Vorgehensweise den Parametern σ_{\min} und σ_{\max} zu. Werden sie zu klein gewählt, kommt es bei zu schneller Bewegung des Objektbildes zu schlechten Ergebnissen. Werden Sie zu groß gewählt, so nimmt die Rechenzeit in nicht tolerierbarem Maße zu.

Als Faustregel galt bisher, dass die maximale Objektbewegung in Pixeln pro Halbbild (bei zeilenverschränkten Bildfolgen) bzw. pro Vollbild (sonst) nicht größer sein sollte als die einfache Standardabweichung in räumlicher Richtung und nicht größer sein darf als die zwei- bis dreifache Standardabweichung.

Bei der Bildfolge reading tritt eine maximale VR von 12 bis 16 Pixel pro Vollbild auf. Infolgedessen wurde $\sigma_{\max} = 6,0$ (bei einer maximalen Maskengröße von 31^3 Pixeln) gewählt. Dies führt jedoch bei der adaptiven GST-Bestimmung zu einer Rechenzeit von über zwei Stunden bei einem Speicherbedarf von 670 MByte. Beide Werte sind inakzeptabel.



Abbildung 7.13: Vervielfachung des Fahrzeugabbildes durch 3D-Tiefpassfilterung

Aus diesem Grund wurde bei den Versuchen zum PETS 2000-Workshop mit 2:1-verkleinerten Bildern gearbeitet. Doch selbst in diesen Bildern liegt die VFR noch bei bis zu 8 Pixeln pro Bild.

7.2.1 Ansatz

In Anlehnung an bekannte Ansätze (z. B. [Rosenfeld 84]) nutzen wir eine Pyramide verkleinerter Bilder. Nach Aufbau der Bild-Pyramide berechnen wir auf jeder Ebene den OF, um anschließend diese Schätzungen zu einem Gesamtergebnis zu kombinieren. Im Folgenden werden die einzelnen Schritte genauer vorgestellt.

7.2.2 Berechnung verkleinerter Bilder

Die Bild-Pyramide besteht aus mehreren Ebenen i , in denen die Bilder jeweils um den Faktor 2^i verkleinert wurden. Die verkleinerten Bilder werden dem ‘Stand-der-Kunst’ folgend durch Tiefpassfilterung mit anschließender Unterabtastung erzeugt. Zu Beginn der Versuche stellte sich die Frage, ob eine nicht-adaptive oder eine adaptive Glättung des Bildes vorgenommen werden soll, ob nur in räumlicher Richtung oder auch in zeitlicher Richtung.

Bei bisherigen Versuchen wurde „nur“ eine nicht-adaptive Filterung durchgeführt, mit brauchbaren Ergebnissen. Eigene Erfahrungen bei der GST-Bestimmung und Ergebnisse von [Almansa & Lindeberg 00], die sich ausführlich mit adaptiven Methoden zur Auswertung einzelner Bilder beschäftigen, legen dennoch ein adaptives Vorgehen nahe. Daher wurden die adaptiven 2D- und 3D-Tiefpassfilterungen implementiert und untereinander und mit den 2D-nicht-adaptiven Ergebnissen verglichen.

Verwendet man eine dreidimensionale, nicht-adaptive Tiefpassfilterung, so werden die Ergebnisse in stationären Bereichen besser als im zweidimensionalen Fall, in bewegten Bildbereichen hingegen kommt es zu einer „Vervielfachung“ der Abbilder der bewegten Objekte (siehe Abbildung 7.13).



Abbildung 7.14: Bild 160 der `reading` Bildfolge: Originalbild (links) und 8:1 verkleinertes Bild (rechts). Zum besseren Vergleich wurde das verkleinerte Bild wieder auf dieselbe Größe wie das Originalbild vergrößert.

Verwendet man eine 3-dimensionale, adaptive Tiefpassfilterung und wählt einen großen Skalierungsbereich ($\sigma_{\max} = 4,0$), so kommt es zwar nicht mehr zu einer Vereinfachung der Objektabbilder, aber zu einem für unsere Zwecke zu starken Glättungseffekt.

Die besten Ergebnisse ergeben sich bei Verwendung einer adaptiven 2D Tiefpassfilterung (Abbildung 7.14): In Richtung der größten Gradienten wird möglichst wenig geglättet, während entlang von Kanten oder in mehr oder weniger homogenen Bereichen stärker geglättet wird.

7.2.3 Kombination der Ebenen zum Gesamtergebnis

In Hinblick auf die OF-Schätzung liefern Vektoren der Kategorien ‘DGD-OF’, ‘OFD’ und ‘ROF’ mehr Informationen als die übrigen: Während bei Punkten der Kategorie ‘Neutral’ kaum eine Aussage über den OF möglich ist, und bei DGD-PP-Vektoren nur *eine* Komponente zuverlässig geschätzt werden kann, liefern ‘DGD-OF’ und ‘OFD’ Vektoren einigermäßen verlässliche Aussagen über beide Komponenten des OFs. Eine hohe Zuverlässigkeit gibt es nur bei ROF-Vektoren. Deshalb sollte versucht werden, möglichst viele ROF-Vektoren zu erhalten, möglichst wenige ‘DGD-PP’ und ‘N’, sowie ‘DGD-OF’ und ‘OFD’ nur, wo es unvermeidbar ist.

Dennoch liefern auch die Vektoren, die nicht als ROF kategorisiert werden, wichtige Informationen, insbesondere in Hinblick auf eine Segmentierung des Bildes.

Zur Vereinfachung der Darstellung wird das Originalbild im Folgenden als Verkleinerungsebene 0 bezeichnet, die $2^i : 1$ verkleinerten Bilder als Ebene i mit $i = 1, \dots, i_{\max}$. Bisher arbeiten wir mit $i_{\max} = 3$, d.h. die maximale Verkleinerung liegt bei $8 : 1$, was z. B. bei der Bildfolge `reading` zu einer Verkleinerung der PAL Bilder von 768×576 Pixel auf 96×72 Pixel führt. Eine weitere Verkleinerung erscheint nicht sinnvoll, weil

die Größe der Faltungsmasken einen Rand erzwingt, in dem kein OF ausgerechnet werden kann. Auf Bildern der Größe 48×36 Pixel wäre eine Berechnung nur auf einem Bereich von etwa 30×18 Pixeln³ möglich, weniger als ein Drittel der Gesamtgröße.

Die Abarbeitung der Bildpyramide startet mit dem kleinsten Bild. Für jedes Ebenenpaar $i+1/i$ (beginnend mit $i+1 = i_{\max}$) testen wir die Konsistenz der „Kategorisierungsübergänge“ von Ebene $i+1$ nach Ebene i („Kompatibel“ oder „Inkompatibel“?) und beurteilen die Zuverlässigkeit der OF-Schätzung sowie ihren Beitrag zur Bildsegmentation („Informationsgehalt“).

Diese Tests führen zu einer der folgenden drei möglichen Aktionen:

- Wenn Ebene i mindestens so viel Information enthält wie Ebene $i+1$, so behalten wir die Kategorisierung aus Ebene i als Zwischenergebnis bei, setzen den Vergleich aber trotzdem mit der nächstfeineren Auflösung, also zwischen Ebene i und $i-1$ fort. In Tabelle 7.5 wird dieser Fall mit „W“ abgekürzt (für „Weitermachen“).
- Wenn Ebene i weniger Informationen enthält als Ebene $i+1$, beenden wir die Auswertung und übernehmen das Ergebnis von Ebene $i+1$ („A+1“ für Akzeptiere Ebene $i+1$).
- Wenn beide Ebenen Informationen enthalten, die in der jeweils anderen Ebene nicht vorhanden sind, so beenden wir die Auswertung mit einer Kombination der Kategorisierung aus Ebene i und der OF-Schätzung aus Ebene $i+1$ („Komb“ für Kombination).

Tabelle 7.5 fasst die möglichen Übergänge zusammen. Auf den ersten Blick ist es überraschend, dass fast alle Übergänge als kompatibel betrachtet werden und die Auswertung fortgesetzt wird. Dies ist aber darauf zurückzuführen, dass so viel Informationen wie möglich bei höchster Bildauflösung gewonnen werden sollen.

Zusätzlich zu den oben angegebenen Bedingungen brechen wir die Auswertung für einen Punkt auch dann auf Ebene i ab, wenn er einer der Kategorien ‘ROF’ oder ‘DGD-OF’ zugewiesen wurde und die Länge des OF-Vektors größer ist als die Hälfte der maximal erlaubten Vektorlänge. Sonst besteht das Risiko, diese Grenze auf der nächsten Ebene zu überschreiten. Bisher wird die maximale Vektorlänge auf σ_{\max} gesetzt. $\sigma_{\max} = 2\sqrt{2}$ und $i_{\max} = 3$ führen somit zu einer zuverlässigen Schätzung für eine Bewegung von bis zu 23 Pixel pro Bild, ohne dafür – wie bisher – die Parameter σ_{\max} und maximale Maskenbreite für einzelne Bildfolgen oder gar einzelne Objekte interaktiv anpassen zu müssen.

7.2.4 Experimentelle Ergebnisse

Das Verfahren ist inzwischen sowohl für Voll- als auch für Halbbilder implementiert. Allerdings wurden noch keine Versuche für Halbbilder durchgeführt.

³Je nach maximaler Maskengröße einige Pixel mehr oder weniger

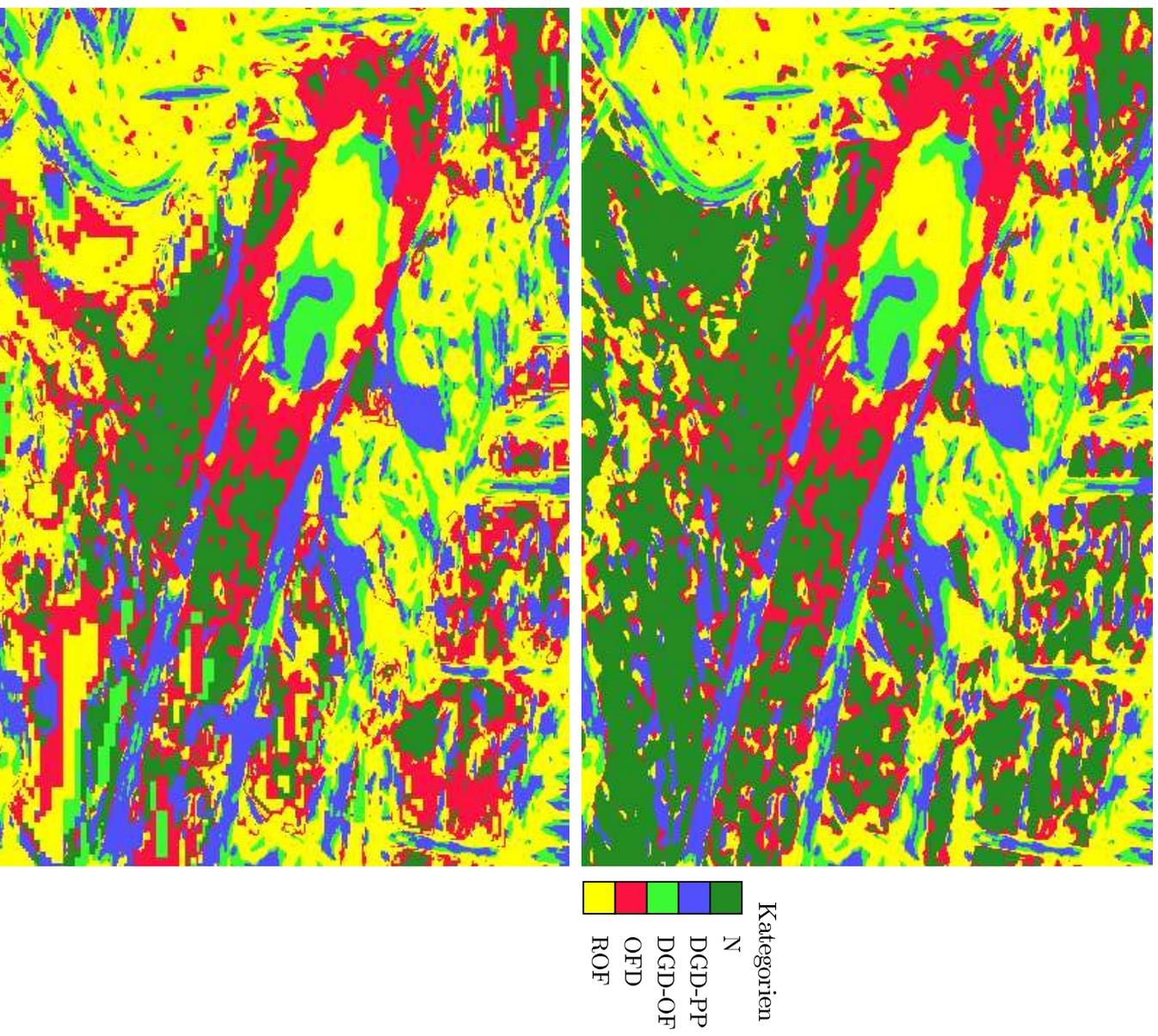


Abbildung 7.15: Vergleich der Ergebnisse von „konventioneller OF-Schätzung“ (oben) und Pyramiden-Ansatz (unten) für Bild 160 der Bildfolge *reading*, beschnitten auf den Bereich, in dem die Pyramide berechnet werden konnte. Der Pyramiden-Ansatz verringert die Anzahl der Punkte der Kategorie „Neutral“ erheblich, vor allem im unteren Bereich des Bildes, indem dort zuverlässigere Informationen aus verkleinerten Bildern verwendet werden, ohne dass dadurch die Umgebung des bewegten Fahrzeugbildes beeinträchtigt wird.

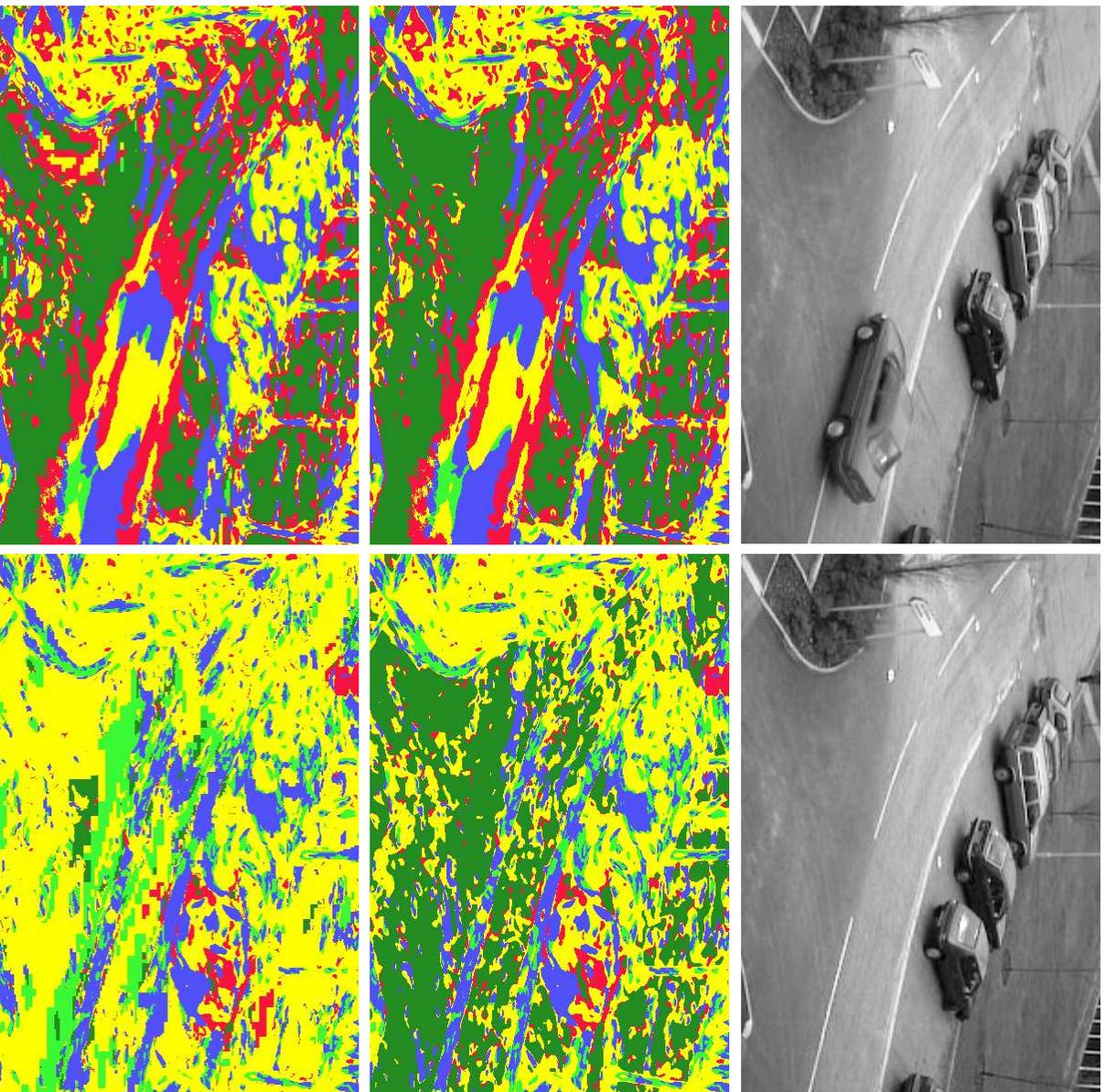


Abbildung 7.16: [Reading] Vergleich der Ergebnisse für Bilder 130 und 650 mit bisherigem Ansatz (Mitte) und mit Pyramiden-Ansatz (unten). Der Pyramiden-Ansatz reduziert die Anzahl der als „Neutral“ kategorisierten Vektoren erheblich, vor allem in Bild 650, indem auf zuverlässige Informationen aus verkleinerten Bildern zurückgegriffen wird. Dennoch bleibt die Diskontinuitätsmaner in Bild 650 erhalten und gestattet so eine Trennung des Fahrzeugabbildes vom Hintergrund allein durch Betrachtung des lokal geschätzten OF.

Kategorisierung auf Ebene i	Kategorisierung auf Ebene $i + 1$				
	N	DGD-PP	DGD-OF	OFD	ROF
N	K-W	K-W	K-A+1	K-W	K-A+1
DGD-PP	K-W	K-W	K-W	I-W	K-Komb
DGD-OF	K-W	K-W	K-W	K-W	K-Komb
OFD	K-W	I-W	K-W	K-W	K-W
ROF	K-W	K-W	K-W	K-W	K-W

Tabelle 7.5: Ergebnis des Kompatibilitätstests (1. Buchstabe; K=Kompatibel, I=Inkompatibel) und durch die Übergänge ausgelöste Handlungen (2. Buchstabe; W=Weitermachen, A+1=Akzeptiere Ebene $i + 1$, Komb=Kombination) je nach Kategorisierung auf Ebenen i und $i + 1$. Weitere Erläuterungen im Text.

7.2.5 Folgerungen

Auch wenn die Ergebnisse auf den ersten Blick kaum von den bisherigen Ergebnissen abweichen und für den Großteil der Bildpunkte die OF-Schätzung der feinsten Auflösung (Originalbild) übernommen wird, so bildet der Pyramiden-Ansatz dennoch die Basis für weitergehende Analysen. Die beiden Hauptvorteile sind, dass sehr große Verschiebungsraten bearbeitet werden können und dass die Adaptionsparameter nicht mehr für jede Bildfolge per Hand optimiert werden müssen.

7.3 Anwendungsbeispiele

Der Diskursbereich, in dem OF-Schätzung und Bild-Segmentierung am IAKS⁴ eingesetzt werden, ist u. a. die Verfolgung von Fahrzeugen im Bildbereich. Es ist daher naheliegend, den Fortschritt durch die in dieser Arbeit dargestellten Untersuchungen dadurch zu demonstrieren, dass Bildsegmentierung oder Fahrzeugverfolgung „erfolgreicher“ verlaufen als mit den zuvor verfügbaren Verfahren.

Bildsegmentierung Abbildungen 7.17, 7.18 und 7.19 zeigen zunächst die Verbesserung bei der GST-basierten Bildsegmentierung. Eine Möglichkeit zur Segmentierung in Bildfolgen (vgl. [Middendorf 99; Middendorf & Nagel 01]) besteht darin, Objekte durch sogenannte Diskontinuitätsmauern vom Hintergrund und von anderen Objekten zu trennen. Diese Diskontinuitätsmauern setzen sich aus Punkten der Kategorien ‘OFD’ und ‘DGD-PP’ zusammen. Um eine saubere Trennung der Objekte zu erreichen, ist es besonders wichtig, dass diese „Mauern“ keine Löcher aufweisen und gleichzeitig

⁴Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik der Universität Karlsruhe (TH)

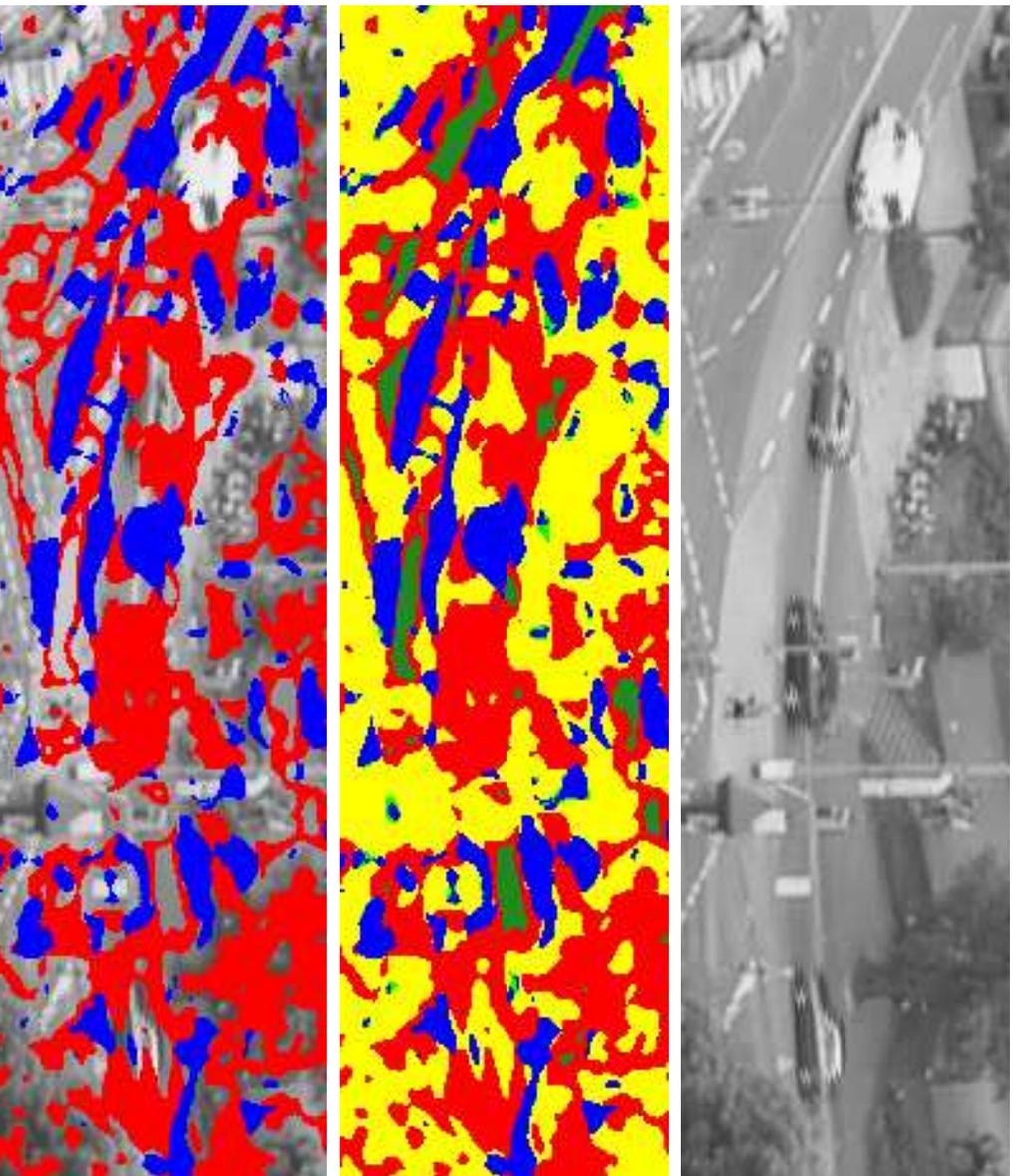


Abbildung 7.17: Kategorisierung des herkömmlich berechneten GSTs (Mitte) und sich daraus ergebende „Diskontinuitätsmauern“ (unten) in einem Ausschnitt von Halbbild 970 der Kreuzungsszene *krwb*. Im dargestellten Ausschnitt bewegen sich vier Fahrzeuge und ein Fußgänger (oben).

möglichst schmal sind, damit kleine Objekte nicht zu einem Großteil von der Diskontinuitätsmauer überlagert werden.

Vergleicht man nun zunächst die in Abbildung 7.18 dargestellte GST-Kategorisierung der neuen Verfahren mit dem herkömmlichen Verfahren (Abbildung 7.17), so erkennt man, dass durch die neuen Verfahren zur GST-Bestimmung der Anteil an ‘OFD’-Vektoren zugunsten der ‘ROF’-Vektoren deutlich abnimmt.

Betrachtet man nur die Diskontinuitätsmauern (Abbildung 7.19), so kann man feststellen, dass die ‘OFD’-Vektoren die Fahrzeugabbilder weit weniger stark überdecken als zuvor, die Diskontinuitätsmauern aber in den beiden Fällen mit Berücksichtigung

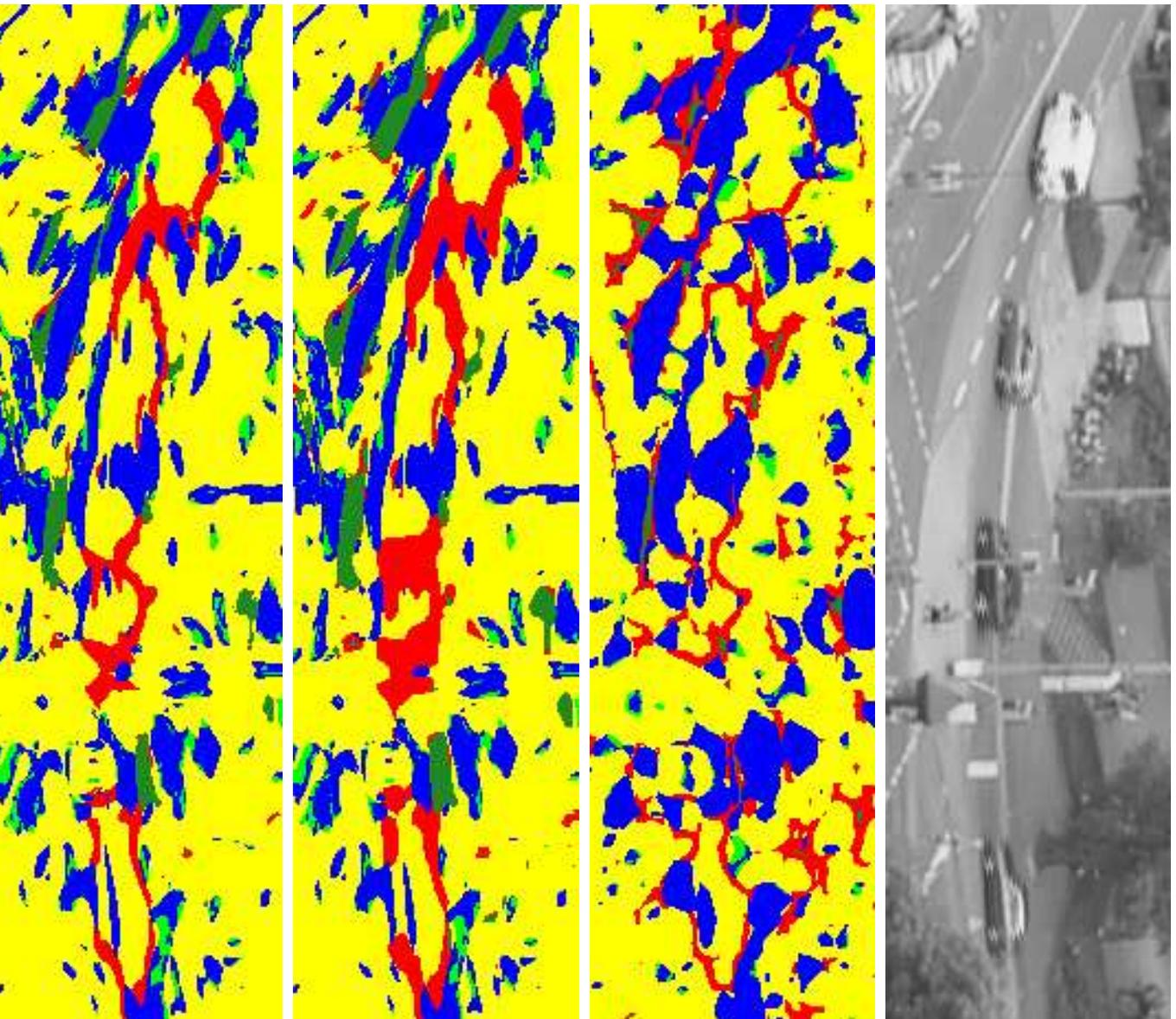


Abbildung 7.18: Kategorisierung des GST für die in dieser Arbeit neu eingeführten Verfahren, von oben nach unten: feste Kovarianzmatrix, aber Berücksichtigung der Diskontinuitätshinweise nach Variante 1 (100 Iterationen), adaptiv bestimmte Kovarianzmatrix, aber keine Berücksichtigung der Diskontinuitätshinweise (10 Iterationen), adaptiv bestimmte Kovarianzmatrix bei gleichzeitiger Berücksichtigung der Diskontinuitätshinweise (10 Iterationen).

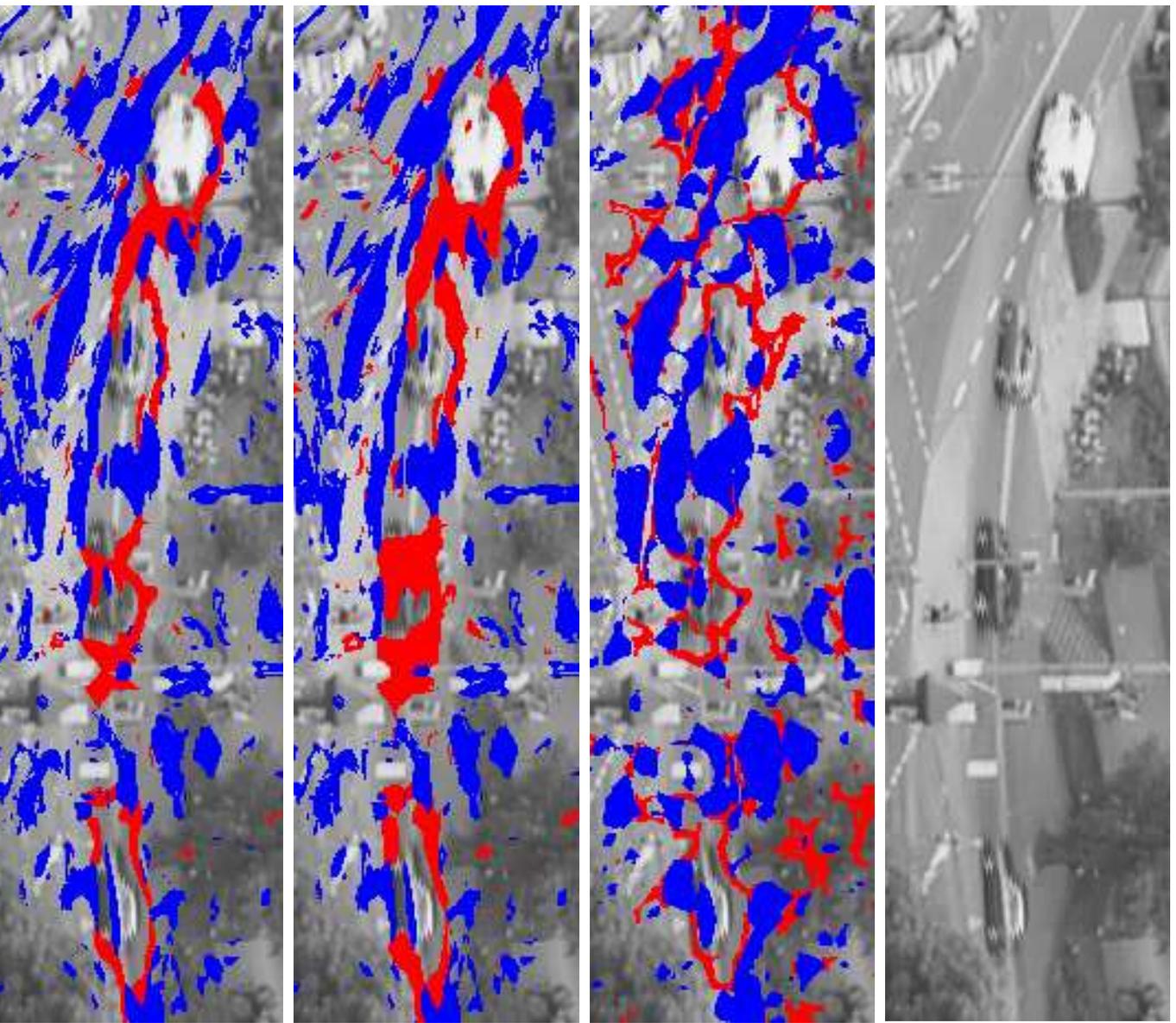


Abbildung 7.19: Wie Abbildung 7.18, aber Darstellung der „Diskontinuitätsmauern“.

der Diskontinuitätshinweise immer noch geschlossen sind. Insbesondere beim Abbild des dritten Fahrzeugs dieser Fahrzeugschlange, das zu diesem Zeitpunkt teilweise von einem Ampelnast verdeckt ist, zeigen sich erhebliche Unterschiede. Während das Fahrzeugabbild beim herkömmlichen Verfahren nahezu vollständig von Diskontinuitäten überlagert ist, ergeben die neuen Verfahren jeweils zwei große zusammenhängende Bereiche von 'ROF'-Vektoren im Front- und Heckbereich des Fahrzeugs, mit einer (sinvollen) Einschnürung bzw. Unterbrechung im Bereich des Ampelnasts. Bei adaptiv bestimmter Kovarianzmatrix ohne Berücksichtigung der Diskontinuitätshinweise fällt diese Unterbrechung aber vergleichsweise breit aus, so dass die anderen beiden neuen Verfahren hier Vorteile zeigen. In allen drei Fällen bleibt auch die Abgrenzung zu dem Fußgänger erhalten.

Fahrzeugverfolgung Diese soeben beschriebene Verbesserung der OF-Kategorisierung im Bereich des Abbilds des 3. Fahrzeugs wirkt sich auch bei einer Verfolgung dieses Fahrzeuges im Bildbereich aus. Dies soll am Beispiel einer „ROF-Flecken-basierten“ Fahrzeugverfolgung demonstriert werden. Dieses Verfahren (vgl. [Middendorf 99]) beruht im Wesentlichen auf einer Segmentierung des Bildes durch Ballung von ROF-Vektoren, die sich zueinander in einer 4er-Nachbarschaft befinden (ROF-Flecken). Nach halb-automatischer Initialisierung⁵ verläuft die Verfolgung durch Zuordnung von ROF-Flecken zu einem verfolgten Objektabbild, wenn sie mit dessen aus dem vorangehenden Bild präzidierter Position, Ausdehnung und mittlerer Bewegung kompatibel sind. Eine wesentliche Voraussetzung für das Gelingen einer solchen Verfolgung ist also das Vorhandensein von zuordenbaren ROF-Flecken.

Abbildung 7.20 vergleicht die Fahrzeugverfolgung ohne und mit Berücksichtigung der Diskontinuität bei der GST-Bestimmung. In beiden Fällen ist durch Verfahren, die nicht Gegenstand der Untersuchung sind, sichergestellt, dass im gesamten Bereich des Fahrzeugabbilds nach ROF-Flecken gesucht wird, die zur zuletzt geschätzten Bewegung des Fahrzeugabbilds kompatibel sind. Dennoch kann nur beim neuen Verfahren ein ROF-Flecken links der Verdeckung „aufgesammelt“ werden. Man erkennt, dass durch die Berücksichtigung der Diskontinuität (rechte Spalte) noch wesentlich näher an der verdeckenden Struktur Punkte als ROF-Vektoren kategorisiert werden, was entscheidend für die erfolgreiche Fortsetzung des dargestellten Verfolgungsprozesses ist.

Anmerkung In manchen Fällen sind schon wesentlich einfachere Veränderungen des Algorithmus oder der Parametrierung für Erfolg oder Mißerfolg einer Fahrzeugverfolgung verantwortlich. Wie bereits in Abschnitt 1.1 erwähnt, konnte schon beim Erstellen der in [Mück 00] verwendeten Trajektorien allein durch Veränderung der Maskengröße aufgrund der in Kapitel 3 vorgestellten Untersuchungen im Vergleich zu [Mück et al. 00] eine zusätzliche Trajektorie erzeugt werden. Bei Versuchen zur Fahrzeugverfolgung im

⁵Der Benutzer gibt einen Bildbereich vor, der nur ein Fahrzeugabbild enthält und bestimmt die diesem Fahrzeugabbild zuzuordnenden ROF-Flecken durch Angabe eines Bereichs für den mittleren OF-Betrag.

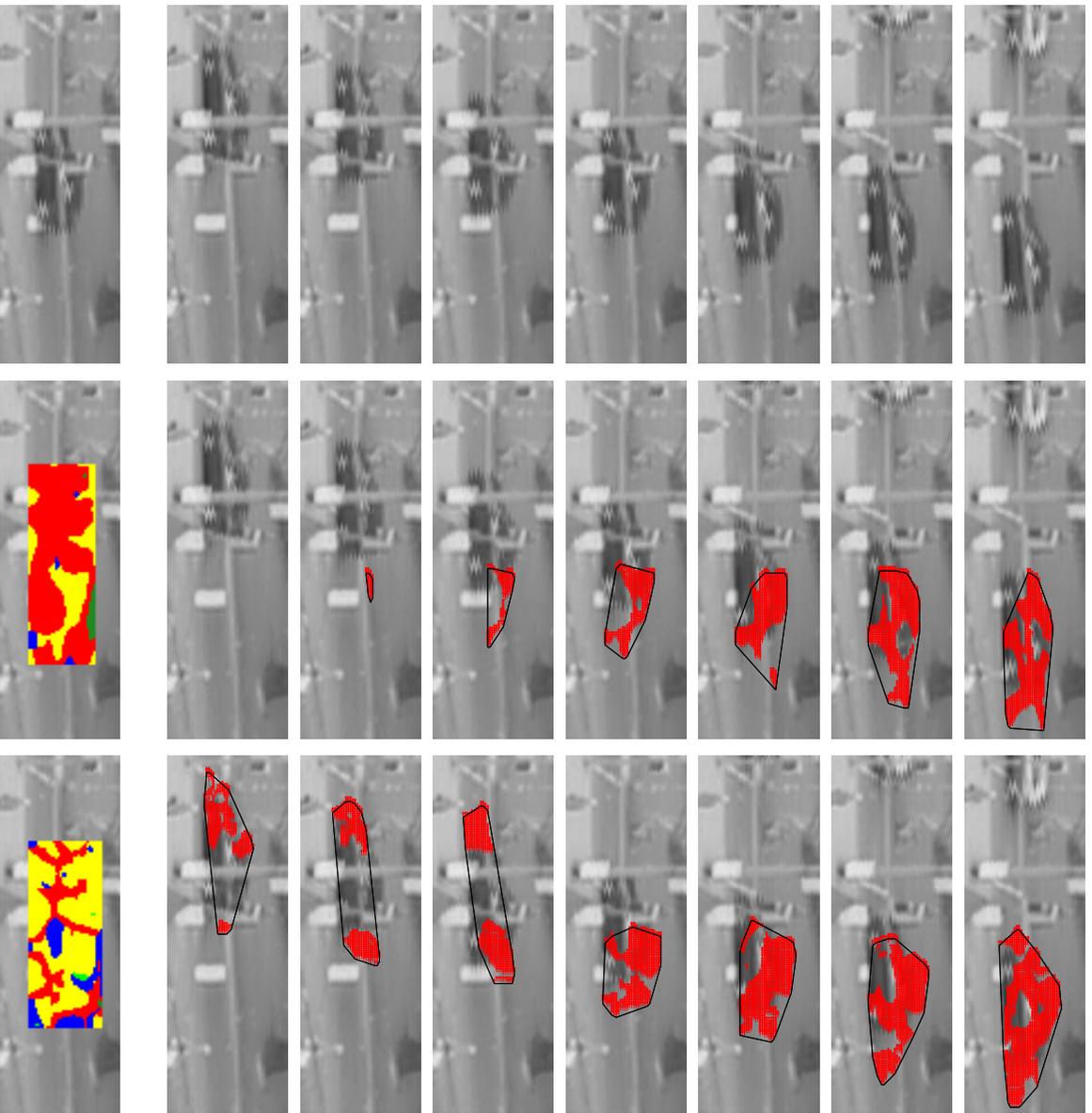


Abbildung 7.20: Vergleich der ROF-Flecken (rot) für das dritte Fahrzeug der Fahrzeugsschlange in den Halbbildern 925 bis 955 (Schrittweite 5) der Bildfolge kwbb beim Passieren mehrerer verdeckender Masten. Die mittlere Spalte zeigt das Resultat bei Verfolgung auf Grundlage des GST ohne Berücksichtigung der Diskontinuität, die rechte Spalte bei Berücksichtigung der Diskontinuität. In beiden Fällen ist durch Verfahren, die nicht Gegenstand der Untersuchung sind, sichergestellt, dass im gesamten Bereich des Fahrzeugabbilds nach ROF-Flecken gesucht wird, die zur zuletzt geschätzten Bewegung des Fahrzeugabbilds kompatibel sind. Dennoch kann nur beim neuen Verfahren ein ROF-Flecken links der Verdeckung „aufsummiert“ werden. Dies ist v. a. darauf zurückzuführen, dass ROF-Vektoren noch wesentlich näher an der verdeckenden Struktur geschätzt werden können. Dies wird auch bei Betrachtung der GST-Kategorisierung in Halbbild 940 (unten) deutlich.

Bildbereich ([Zacharias 03]) zeigte sich aber auch, dass die neu entwickelten Verfahren nicht in jedem Fall zu einer Verbesserung, sondern in manchen Fällen sogar zu einer deutlichen Verschlechterung der Verfolgungsergebnisse führten. Dafür konnten mehrere Gründe ausgemacht werden:

- Bei den erfolgreichen Verfolgungsversuchen mit den alten Verfahren wurden die Parameter z. T. interaktiv für jeden einzelnen Versuch optimiert. Es wäre sehr überraschend, wenn genau derselbe Parametersatz nach erheblichen Veränderungen des Algorithmus weiterhin zu einer erfolgreichen Verfolgung führt.
- Versucht man nun, auch für die neu entwickelten Verfahren, die Parameter zu optimieren, so bewegt man sich in einem relativ großen Suchraum. Schon in [Middendorf 99] wurde auf die Sensitivität des zugrundeliegenden Algorithmus auf Parameteränderungen hingewiesen.
- Es kann sogar vorkommen, dass eine *Verbesserung* der OF-Felder zu einer *Verschlechterung* der Verfolgungsergebnisse führt: In einem Schritt des Verfolgungsalgorithmus (Details in [Middendorf 99]) werden alle Punkte in einem bestimmten Bereich aufgesammt und geballt, die einen fest vorgegebenen Mahalanobis-Abstand von der für diese Bildpunkte geschätzten mittleren Bewegung unterschreiten. Der Mahalanobis-Abstand ist in diesem Zusammenhang – vereinfacht ausgedrückt – die Anzahl an Standardabweichungen, die ein Maßwert vom Mittelwert abweicht. Die notwendige Kovarianzmatrix wird gleichzeitig mit dem Mittelwert der OF-Vektoren über mehrere Bilder geschätzt. Nimmt nun die Varianz der OF-Schätzungen ab, so nimmt der Mahalanobis-Abstand eines OF-Vektors vom Mittelwert (bei gleichem euklidischen Abstand) zu.

Wie diese Überlegungen gezeigt haben, ist also das Gelingen oder Mißlingen einer Fahrzeugverfolgung nur bedingt ein geeignetes Kriterium, um die Qualität der OF-Schätzung oder der Bildsegmentierung zu beurteilen.

7.4 Fazit

In diesem Kapitel konnten einige Schwachstellen des Verfahrens erklärt werden. Die Berücksichtigung der Bewegungsdiskontinuität wurde in die adaptive GST-Schätzung integriert, was – wie durch Versuche gezeigt – zu einer Verbesserung der Ergebnisse sowohl der OF-Schätzung als auch der GST-Kategorisierung führte.

Kapitel 8

Zusammenfassung und Ausblick

8.1 Zusammenfassung

Die vorliegende Arbeit leistet mehrere Beiträge zur Forschung im Bereich Bewegungsschätzung und Segmentierung:

- Im ersten Teil der Arbeit wurden Implementierung und Parametrierung der Basisoperationen „Glättung“ und „Ableitung“ untersucht, z. T. unter Einbeziehung des sowohl Ableitungs- als auch Glättungsoperationen erfordernden GST.
 - Bei der Analyse der vorhandenen Implementierung und der damit verknüpften Korrekturverfahren stellte sich heraus, dass diese z. T. erhebliche Schwächen aufwiesen, auch wenn sie bei speziellen Testfällen sehr gute Ergebnisse lieferten. Davon ausgehend konnten neue Verfahren zur Korrektur des durch Diskretisierung und Beschränkung der Maskengröße verursachten Fehlers entwickelt werden. Die Wirksamkeit dieser neuen Verfahren wurde durch mehrere Versuche anhand klar definierter und naheliegender Kriterien nachgewiesen. Allerdings musste auch festgestellt werden, dass es im Zweifelsfall besser ist, größere Faltungsmasken zu verwenden als den durch die Beschränkung der Maskengröße verursachten Fehler durch ein (stets auf zusätzlichen Annahmen beruhendes) Verfahren zu korrigieren.
 - Die nachfolgende Untersuchung des Einflusses der Skala auf das Ergebnis der GST-Bestimmung – gemessen am Fehler der Gradientenrichtung und des OF bei einer Bildfolge mit bekannter Grundwahrheit – diente hauptsächlich dazu, den bei adaptiver GST-Bestimmung auftretenden Parameter „Skalierungsbereich“ vernünftig zu wählen. Gleichzeitig konnte gezeigt werden, dass die – im Sinne einer Minimierung des Fehlers eines mit der Grundwahrheit vergleichbaren Merkmals – optimale Skala nicht eindeutig sein muss.
- Im zweiten Teil der Arbeit – Kapitel 6 – konnte ein bisher heuristischer Ansatz

zur adaptiven GST-Bestimmung in einen theoretischen Rahmen eingebettet werden. Diese Einbettung erleichtert die Analyse der auftretenden „Problemfälle“ und ermöglicht es u. a., den sonst heuristisch gewählten Parameter „Anzahl der Iterationen“ fundierter zu wählen. Durch algorithmische Anpassungen konnte darüberhinaus eine Verbesserung der OF-Schätzung erzielt werden. Die somit erzielten Ergebnisse wurden mit den Resultaten anderer Autoren auf allgemein bekannten „Standardbildfolgen“ verglichen und erwiesen sich dabei als „konkurrenzfähig“.

- Im dritten Teil der Arbeit konnten zunächst einige Schwachstellen des Verfahrens erklärt und zum Großteil auf eine Verletzung der dem Verfahren zugrunde liegenden Annahmen zurückgeführt werden. Für den Problemfall „Bewegungsdiskontinuität“ wurde ein neues Verfahren umgesetzt, das die im GST enthaltene Hinweise auf das Vorliegen einer solchen Diskontinuität durch Rückkopplung schon während der adaptiven GST-Bestimmung nutzt. Dadurch konnte eine weitere Verbesserung der OF-Schätzung und vor allen der Bildsegmentierung erzielt werden. Auch in diesem Fall beruht das Vorgehen auf vergleichsweise wenigen Annahmen. Insbesondere handelt es sich um ein streng-lokales Verhalten. Dadurch wird bewusst auf die z. B. bei Niveauschnitt-Verfahren auf elegante Weise gegebene Möglichkeit verzichtet, die „Diskontinuitätsmauer“ bzw. Objektabbildkontur bei mangelnder lokaler Information aufgrund globalerer Information zu schließen. Trotzdem konnten mit diesen Verfahren Segmentierungsergebnisse erzielt werden, die zu einer Verbesserung der Ergebnisse beim Anwendungsbeispiel „Fahrzeugverfolgung im Bildbereich“ führten.

Insgesamt wurde darauf geachtet, die Zahl der Annahmen möglichst klein zu halten und diese explizit zu formulieren. Dies führt bei einem Vergleich der Ergebnisse zwar zu einem etwas schlechteren Abschneiden als bei Ansätzen, die mehr Annahmen über die Art der Bilder, den Abbildungsprozess oder die beobachtete Szene treffen. Die durchgeführten Vergleiche zeigen aber, dass auch auf diesem Wege akzeptable Ergebnisse erzielt wurden. Gleichzeitig ist eine leichtere Übertragbarkeit der Ergebnisse auf andere Bildfolgen zu erwarten.

8.2 Ausblick

Trotz der erzielten Fortschritte besitzt das verwendete Verfahren zur GST-Bestimmung und OF-Schätzung noch einige Schwächen, die in zukünftigen Arbeiten behandelt werden müssen:

- Sowohl das lineare als auch das quasi-inverse Adaptionungsverfahren führen zwar insgesamt zu einer Verbesserung, an einigen Punkten aber auch ohne Beeinträchtigung durch Bewegungsdiskontinuitäten zu einer Verschlechterung der OF-Schätzung. Durch die Einführung des quasi-inversen Adaptionsverfahrens wurde

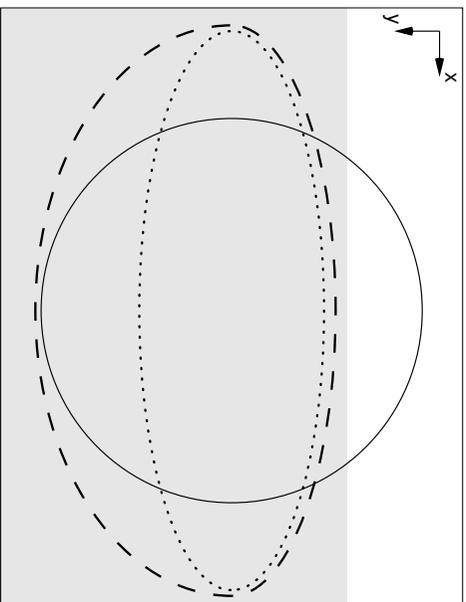


Abbildung 8.1: Schematische Darstellung der Probleme durch Punktsymmetrie der Gaußglocken: Der durchgezogene Kreis stellt den Bereich dar, der in der ersten nicht-adaptiven Iteration zur Schätzung des Gradienten herangezogen wird. Im Adaptionsschritt wird der betrachtete Bereich in Richtung des Gradienten gestaucht, senkrecht dazu ausgedehnt (gepunkteter Kreis). Wünschenswert wäre aber die Betrachtung eines Bereiches, wie er durch die gestrichelte Markierung angezeigt wird, da in positiver y -Richtung *keine* Kante vorliegt.

zwar ein nachweisbarer Fortschritt erzielt, offensichtlich ist aber das Adaptionsverfahren immer noch nicht optimal. Unklar bleibt auch, ob es ein solches für die Mehrzahl der Fälle optimales Adaptionsverfahren überhaupt gibt.

- Grundidee der Adaption – wie auf Seite 113 dargestellt – ist es, dass Ableitungs- und Mittelungsmasken groß ausgedehnt sind in Richtungen, in die sich die Grauwertstruktur wenig ändert, und nur geringe Ausdehnung in Richtungen besitzen, in die sich die Grauwertstruktur stark ändert. Die Gaußglocke kann diese Anforderungen aber aufgrund ihrer Punktsymmetrie nur unvollständig erfüllen. Abbildung 8.1 illustriert dies an einem Beispiel. Betrachtet wird die Auswirkung der Adaption an einem Punkt in der Nähe einer stark ausgeprägten geradlinigen Kante. Der durchgezogene Kreis symbolisiert den Bereich, der im Startschritt, also bei nicht-adaptiver GST-Bestimmung, betrachtet wird. Die Adaption in der bisherigen Form führt im ersten Iterationsschritt zu einer Streckung des berücksichtigten Bereichs entlang der Kante und zu einer Stauchung senkrecht dazu, gepunktet dargestellt. Diese „Stauchung“ führt aber auch zu einer Minderung des Einflusses des vom Aufpunkt gesehen weiter von der Kante entfernt liegenden Bereiches (im Bild unten). Wünschenswert wäre eine Form der Gewichtungsfunktion, wie sie durch die gestrichelte Figur angedeutet ist. Eine denkbare Lösung wäre eine Verschiebung des Maskenmittelpunkts von der Kante weg, wie es schon von [Nitzberg & Shioita 92] vorgeschlagen wurde.

- Im Gegensatz zu [Koenderink 88] und [Lindeberg 97], die zeitliche Kausalität

fordern, behandelt das Verfahren bisher die Zeit genau so wie die räumlichen Dimensionen. Auch wenn in der hier vorgelegten Untersuchung keine Echtzeit-Auswertung angestrebt wird, so könnte es auch beim Adaptionenverfahren Vorteile bringen, zwischen Zeit- und Raumachsen zu unterscheiden.

- In (nahezu) homogenen Bereichen kann aufgrund mangelnder Information kein zuverlässiger OF geschätzt werden. Dieses Problem kann nur durch zusätzliche Annahmen gelöst werden. Eine Möglichkeit dazu wurde durch den Pyramidenansatz bereits aufgezeigt.¹ Eine genauere Analyse des Pyramidenansatzes oder die Untersuchung von Alternativen stehen noch aus.
- Bei schneller Bewegung kann aufgrund mangelnder zeitlicher Kontinuität (Verstoß gegen Annahme 2) kein zuverlässiger OF geschätzt werden (vgl. [Middendorf & Nagel 00]). Auch in diesem Fall kann der Pyramidenansatz zu Verbesserungen führen. In diesem Bereich gibt es aber auch einige neuere Forschungsansätze anderer Autoren, die im Rahmen der vorgelegten Arbeit nicht berücksichtigt wurden, da dies nicht Hauptgegenstand der Untersuchungen war.

- Vor allem bei der Bewegung von texturierten Objekten wie PKWs auf nahezu homogenen Hintergrund ist der Bereich, in dem ein der Bewegung des Fahrzeugabbilds entsprechender OF geschätzt wird, oft deutlich über das Fahrzeugabbild hinaus ausgedehnt. Dies ist unerwünscht, aber bei genauer Betrachtung kein Fehler, da dies nur bedeutet, dass eine solche Bewegung konsistent ist mit dem beobachteten Ausschnitt der Bildfolge. Durch eine zeitliche Fortschreibung der in einem Bild gewonnenen Information ist eine erhebliche Verbesserung zu erwarten. Dies bedingt aber auch zusätzliche Annahmen über die beobachtete Szene.

Obwohl die vorliegende Arbeit einige wesentliche Fortschritte erzielt hat, so gibt es immer noch zahlreiche zu klärenden Punkte. Vor allem der letztgenannte Punkt verspricht weitere deutliche Fortschritte, erfordert aber aufgrund der dabei notwendigen Zusatzannahmen ein besonders sorgfältiges Vorgehen.

¹Die dabei verwendete Annahme besteht darin, dass sich Punkte ohne Bewegungsinformation so bewegen, wie der nächstgelegene Punkt mit zuverlässiger Bewegungsinformation.

Anhang A

Referenzbildfolgen

Bei den Untersuchungen zur Bestimmung des OFs wurden verschiedene Bildfolgen als Grundlage für Vergleiche verwendet. In diesem Abschnitt werden diese Bildfolgen vorgestellt und ihre Besonderheiten diskutiert.

A.1 `trans_tree`

Beschreibung

Die „Translating-Tree“-Bildfolge (Abbildung A.1) – erzeugt von David Fleet – ist ein Testfall für eine rein translatorische Bewegung der Kamera *senkrecht* zur ihrer Sichtachse. Die Szene ist planar – es handelt sich um eine Foto-Tapete – es gibt also keinerlei Bewegungsdiskontinuitäten. Der Betrag der VR liegt zwischen 1.73 bis 2.30 Pixeln pro Bild.



Abbildung A.1: Bilder 15 (links) und 25 (rechts) der „Translating-Tree“-Bildfolge `trans_tree`

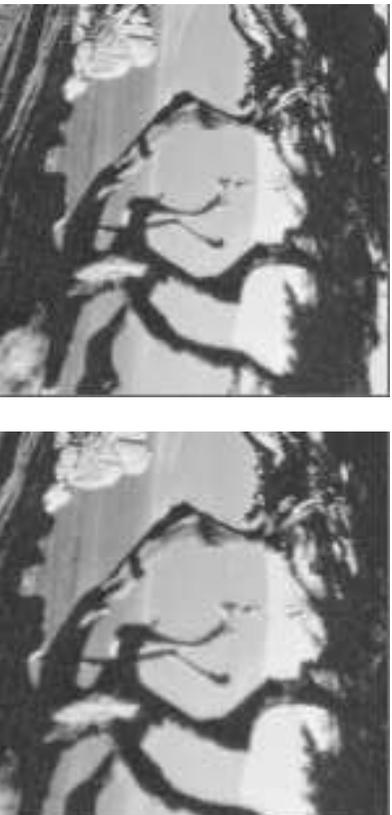


Abbildung A.2: Bilder 15 (links) und 25 (rechts) der „Diverging-Tree“-Bildfolge `div-tree`

Vorteile

- Die tatsächliche VR ist bekannt und liegt in einem Bereich, für den differentielle Verfahren noch gut geeignet sind.
- Es gibt keine Bewegungsdiskontinuitäten, auch wenn die abgebildete Szene dies auf den ersten Blick suggeriert.
- Die Bildfolge ist ein „Klassiker“: Es liegen Vergleichswerte für die Ergebnisqualität zahlreicher anderer Verfahren vor ([Barron et al. 94]), vgl. auch Abschnitt 6.2.

Nachteile

- Es handelt sich um eine semi-synthetische Bildfolge, d.h. es wurde eine kontrollierte Kamerabewegung in kalibrierter Umgebung durchgeführt. Mögliche – und in der Praxis selbst bei sorgfältigem Vorgehen nicht völlig zu vermeidende – Mängelgenauigkeiten verursachen unbekannte Fehler bei der Referenz-VR.
- Die Bildfolge ist nur wenige (41) Bilder lang.

Bezugsquelle

Bildfolgen-Archiv der University of Western Ontario ([CSD-UWO-Archiv]).



Abbildung A.3: Bilder 2 (links) und 12 (rechts) der „Yosemite“-Bildfolge *yosemite*

A.2 *div_tree*

Beschreibung

Die „Diverging-Tree“-Bildfolge (Abbildung A.2) – erzeugt von David Fleet – ist ein Testfall für eine rein translatorische Bewegung der Kamera *entlang* ihrer Sichtachse. Die Szene ist planar – es handelt sich um eine Foto-Tapete – es gibt also keinerlei Bewegungsdiskontinuitäten. Die VR liegt zwischen 0,0 (im Zentrum) und 2,0 Pixeln pro Bild.

Vor- und Nachteile

Wie bei der „Translating-Tree“-Bildfolge.

Bezugsquelle

Bildfolgen-Archiv der University of Western Ontario ([CSD-UWO-Archiv]).

A.3 *yosemite*

Beschreibung

Die Yosemite-Bildfolge (Abbildung A.3) – erzeugt von Lynn Quam – zeigt eine „Flugszene“, die durch Aufprojektion einer Aufsicht auf ein dreidimensionales Geländemodell erzeugt wurde, bei linearer Bewegung der „Kamera“ auf die Szene zu. Der Betrag der VR liegt zwischen 0.0 und 5.4 Pixeln pro Bild.

Vorteile

- Die Szene ist herausfordernder als die Translating-Tree- und Diverging-Tree-Bildfolge, da sie eine große VR und Bewegungsdiskontinuitäten enthält.
- Die tatsächliche VR ist bekannt.
- Die Bildfolge ist ein „Klassiker“: Es liegen Vergleichswerte für die Ergebnisqualität zahlreicher anderer Verfahren vor ([Barron et al. 94]).

Nachteile

- Die tatsächliche VR liegt nur mit deutlichen Quantisierungsfehlern (mit Quantisierungsstufen von ca. 0.03 Pixeln/Bild) vor.
- Ein Teil der Szene (Wolken) hellt sich von Bild zu Bild stark auf. Ein Vergleich nur des mittleren Fehlers verschiedener Ansätze unter Berücksichtigung dieses Bereichs lässt daher keine Rückschlüsse zu, ob der vermeintlich bessere Ansatz nur diese Aufhellung (besser) modelliert oder auch zu Verbesserungen in den übrigen Bildbereichen führt.
- Die Bildfolge ist sehr kurz (nur 15 Bilder).

Bezugsquelle

Bildfolgen-Archiv der University of Western Ontario ([CSD-UWO-Archiv]).

A.4 tr_bsb

Beschreibung

Bei der Bildfolge `tr_bsb` (Abbildung A.4) handelt es sich um die Anfang der 90er Jahre von M. Otte aufgenommene „Marmorblockszene“: Mehrere Marmorblöcke stehen auf einer ebenen Platte mit marmor-artiger Textur. Die Kamera wurde an einem Roboterarm rein translatorisch bewegt. Zusätzlich wurde einer der Blöcke von Bild zu Bild verschoben. Man könnte diese Bildfolge als semi-synthetisch bezeichnen, da sie eine Kombination aus kontrollierter Bewegung und realen Objekten darstellt.

Vorteile

- Die Szene ist repräsentativ für die Bewegung texturierter Objekte.

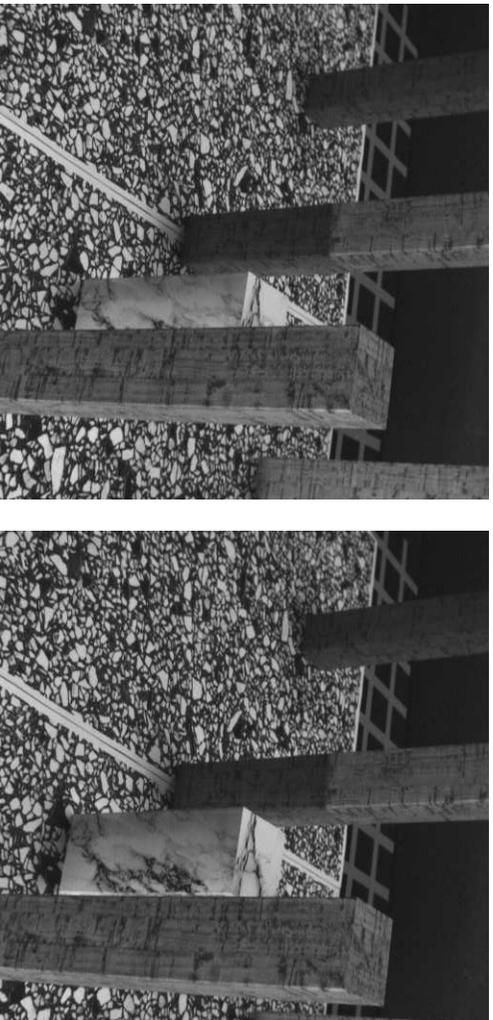


Abbildung A.4: Bilder 25 (links) und 45 (rechts) der Marmor-Bildfolge tr_bsb

- Es gibt große, zusammenhängende, gleichförmig bewegte Flächen, ideal für die Bestimmung von „glatten“ OF-Feldern.
- Otte hat aus der bekannten Szenengeometrie und der Kamerabewegung die „tatsächliche“ VR berechnet. Diese steht für Vergleiche zur Verfügung. Dabei wurden nur Pixel berücksichtigt, die eindeutig das Abbild *einer* Fläche in der Szene sind, also keine Pixel, bei denen Flächen aufeinandertreffen.
- Aus den Untersuchungen von [Otte 94] sind Vergleichswerte des OF-Fehlers verfügbar.

Nachteile

- Die Details der Methode zur Erzeugung der „tatsächlichen“ VR und der damit verbundene Fehler sind unbekannt.
- *Systematische* Unterschiede zwischen VR und geschätztem OF ließen Zweifel an der Korrektheit der berechneten VR aufkommen.
- Die Bewegung der Kamera und des Marmorblocks sind nicht exakt definiert; möglicherweise fand keine gleichmäßige Bewegung im Verlauf der Bildfolge statt.
- Die Vergleichswerte des OF-Fehlers stehen größtenteils nur als Grafiken ohne Maßstab zur Verfügung.
- Es gibt sehr dunkle, kontrastarme Bildbereiche.



Abbildung A.5: Habbilder 930 (links) und 2650 (rechts) der Bildfolge `kwbb`.

Bezugsquelle

Bildfolgen-Archiv des IAKS ([IAKS-Archiv]): „Marbled Block (old version) `tr_bsb`“.

A.5 `kwbb`

Beschreibung

Die Bildfolge `kwbb` (Abbildung A.5) ist eine insgesamt fast 3500 Halbbilder (ca. 70 Sek.) lange Aufnahme einer innerstädtischen Kreuzung in Karlsruhe, aufgenommen von einer temporär aufgestellten, an einem Stativ befestigten Kamera.

Vorteile

Die Bildfolge ist typisch für die vom IAKS untersuchten Verkehrsszenen:

- zahlreiche, kleine, unterschiedliche Objekte (PKW, LKW, Motorräder, Fußgänger),
- unkontrollierbare Beleuchtungsverhältnisse inkl. Spiegelungen und Schattenwurf,
- sich ändernder Bildausschnitt durch Vibrationen der Kamera und Synchronisationsfehler,
- zeilenversetzte Halbbilder.

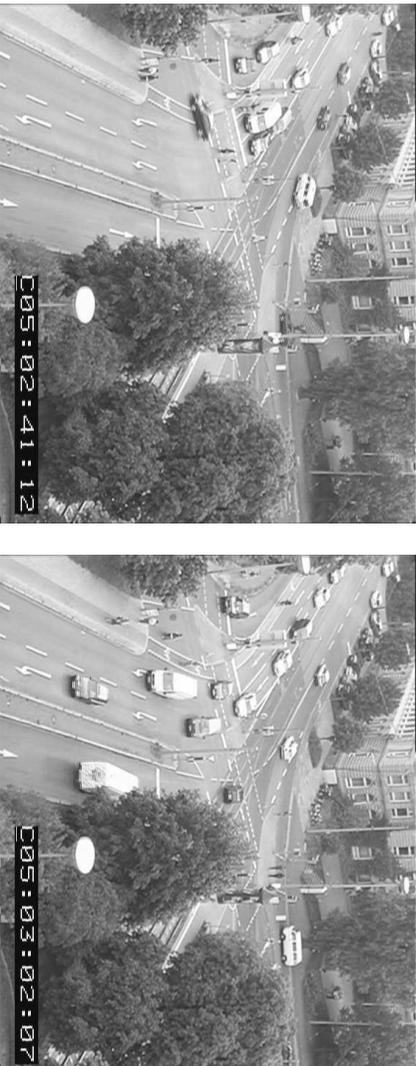


Abbildung A.6: Halbbilder 540 (links) und 1580 (rechts) der Bildfolge stau02.

Nachteile

- Es werden Halbbilder verwendet, was möglicherweise zu einer bisher nicht beachteten Beeinflussung der Ergebnisse führt.
- Es liegt keine Referenz-VR vor.

Bezugsquelle

Bildfolgen-Archiv des IAKS ([IAKS-Archiv]): „Karl-Wilhelm-Straße (normal conditions, old camera position) *kwbb*“.

A.6 stau02

Beschreibung

Die Bildfolge *stau02* (Abbildung A.6) ist insgesamt über 2000 Halbbilder lang. Da sie nicht öffentlich verfügbar ist, wurde sie nur dann herangezogen, wenn Effekte demonstriert werden sollen, die bei anderen Bildfolgen nicht beobachtet werden konnten.

Vor- und Nachteile

Bei dieser Bildfolge treten dieselben Vor- und Nachteile auf wie bei der Bildfolge *kwbb*.



Abbildung A.7: Halbbilder 550 (links) und 810 (rechts) der Bildfolge reading.

Bezugsquelle

Bildfolgen-Archiv des IAKS ([IAKS-Archiv]): „Karl-Wilhelm-Straße (normal conditions, old camera position) stau02“.

A.7 reading

Beschreibung

Die Bildfolge reading (Abbildung A.7) wurde als Test-Bildfolge des PETS 2000-Workshops veröffentlicht. Sie besteht aus 1452 JPEG-komprimierten Bildern. Insgesamt bewegen sich drei PKWs und zwei Fußgänger durch das Blickfeld der Kamera.

Vorteile

- Die Bildfolge ist sehr herausfordernd. Die PKWs bewegen sich mit bis zu 16 Pixel/Bild oder bremsen bis zum Stillstand.
- Es liegen Vergleichsergebnisse ([Ferrymann 00]) anderer Autoren vor.

Nachteile

- Durch die starke Komprimierung (PAL-Farbbilder mit 442368 Bildpunkten sind auf durchschnittlich 54 KB reduziert) wird vor allem bei bewegten Objekten die Annahme konstanter Helligkeit (Annahme 4) verletzt.

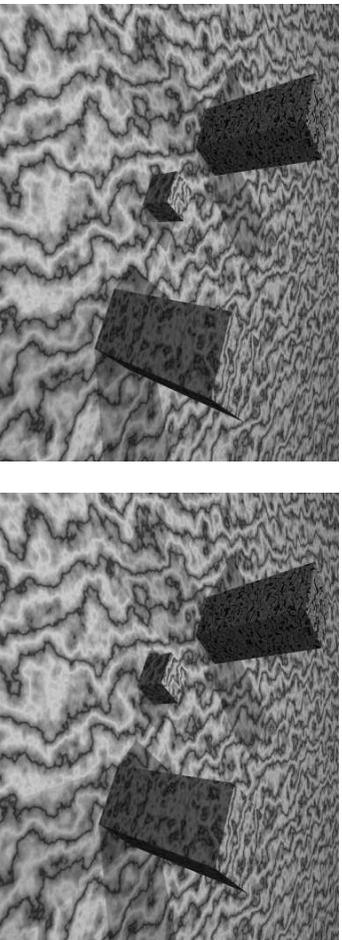


Abbildung A.8: Bilder 140 (links) und 150 (rechts) der Bildfolge `marmor_stat`.

- Durch die sehr schnelle Fahrzeugbewegung ist ein differentielles Verfahren eigentlich ungeeignet.

Bemerkung

Einige Versuche verwenden die verkleinerte Version `reading_ua`, die durch räumliche Tiefpassfilterung mit einer isotropen Gaußmaske ($\sigma = 0.5$) und anschließende $2 : 1$ -Unterabtastung gewonnen wurde.

Bezugsquelle

PETS-Workshop Bildfolgen-Archiv ([PETS-Archiv]).

A.8 `marmor_stat`

Beschreibung

Als Konsequenz aus der Tatsache, dass bei den bisher verwendeten Realwelt-Bildfolgen die tatsächliche Verschiebung nur mit unzureichender Genauigkeit bekannt ist, wurde eine neue synthetische Bildfolgen generiert. Die Erzeugung erfolgte mit Hilfe des „Persistence of Vision Ray-Tracers (POVRay)“. Es wurde eine Szene mit texturierten Oberflächen modelliert, die einer Marmor-Oberfläche ähnlich sehen. Die stationäre Kamera blickt schräg auf eine Marmorplatte, auf der sich ein stationärer und zwei gleichförmig bewegte Marmorquader befinden (Abbildung A.8). Die Bildfolge ist 200 Bilder lang und kann bei Bedarf beliebig in beide Richtungen verlängert werden.

Durch mehrere Untersuchungen wurde sichergestellt, dass die Bilddaten und die Referenz-VR tatsächlich fehlerfrei sind:

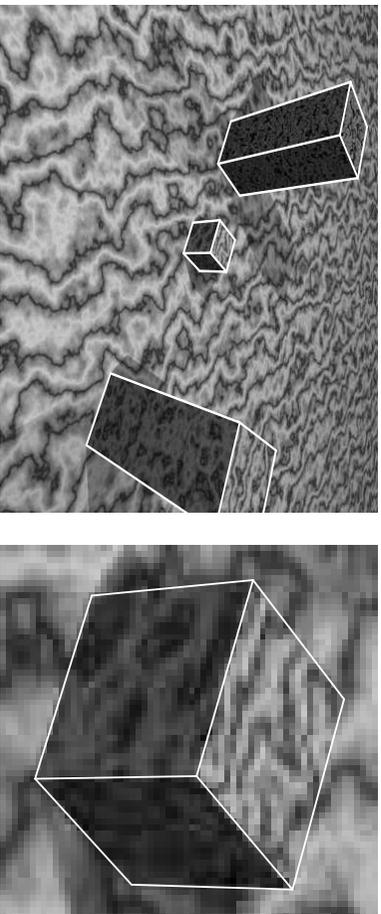


Abbildung A.9: Projektion des 3D-Szenenmodells in Bild 100 der Bildfolge *marmor-stat* (links) und Ausschnittsvergrößerung (rechts). Die überlagerten Modellkanten stimmen exakt mit den Bildkanten überein.

- Eine visuelle Begutachtung der Bilder ergab, dass aufeinanderfolgende Bilder „zueinander passen“, also weder Änderungen der Gesamthelligkeit noch unerklärliche Änderungen der Grauwertstruktur auftraten. Im stationären Teil der Bildfolge wurde darüber hinaus durch punktweisen Vergleich der Grauwerte sichergestellt, dass diese sich abgesehen von Schatten-Effekten nicht ändern.
- Um eine fehlerhafte Referenz-VR auszuschließen, wurde zunächst durch stichprobenartige Inspektion der Daten sichergestellt, dass die Werte der VR an benachbarten Punkten konsistent zueinander sind, dass also eine kleine, aber systematische Änderung der VR von Pixel zu Pixel erfolgt. Außerdem wurde kontrolliert, dass „Unstetigkeiten“ der VR exakt auf den Objektkanten auftreten und nicht etwa um einen oder mehrere Pixel verschoben. Versuche mit den geschätzten lokalen Tensoren $(\nabla g)(\nabla g)^T$ (siehe z. B. Winkel α_{v_i, E_0} in Abbildungen 5.2 und 5.4) zeigen außerdem, dass bei Verwendung der Referenz-VR Punkte mit kleinen und mit großem Fehler der lokalen Tensoren zumindest auf den ersten Blick unregelmäßig verteilt sind, was gegen eine systematisch zu groß oder zu klein angesetzte VR spricht.

Vorteile

- Die Bewegungen der Objekte in der Szene und ihrer Abbilder im Bildbereich sind exakt bekannt (siehe auch Abbildung A.9), ebenso die internen und externen Kameraparameter. Messungengenauigkeiten wie bei semi-synthetischen Bildfolgen sind somit ausgeschlossen.
- Alle Aufnahmeparameter (Belichtung, Spiegelungen, Schattenwurf, ...) können kontrolliert werden.

Nachteile

- Die deutlich texturierten Oberflächen des „Marmors“ sind untypisch für die sonst von uns beobachteten Objekte.
- „Realwelt-spezifische“ Probleme können nicht erkannt werden.

Bezugsquelle

Bildfolgen-Archiv des IAKS ([IAKS-Archiv]): „Marbled Block (new version)“.

Herleitung der Projektionsmatrix bei marmor_stat

Die Erzeugung der VR aus der bekannten Objektbewegung im Szenenbereich erfolgte mit Hilfe von `xtrack`. Dazu muss u.a. die Rotation des Kamerakoordinatensystems gegenüber dem Weltkoordinatensystem bekannt sein, während bei `POVRay` nur die Rotation um die einzelnen Achsen angegeben werden muss:

- Rotation um x -Achse um 45° :

$$R_x(45) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{pmatrix}$$

- Rotation um y -Achse um 120° :

$$R_y(120) = \begin{pmatrix} \frac{-1}{2} & 0 & \frac{1}{2}\sqrt{3} \\ 0 & 1 & 0 \\ -\frac{1}{2}\sqrt{3} & 0 & \frac{-1}{2} \end{pmatrix}$$

- Rotation um x -Achse um 90° :

$$R_x(90) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

Gesamtrotation:

$$R = \begin{pmatrix} \frac{-1}{2} & \frac{1}{2}\sqrt{3} & 0 \\ \frac{1}{4}\sqrt{6} & \frac{1}{4}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ -\frac{1}{4}\sqrt{6} & -\frac{1}{4}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix}$$

Literaturverzeichnis

Hinweis: Verweise auf Internet-Quellen sind in Schreibmaschinenschrift dargestellt.

- [Acton 90] F. S. Acton: *Numerical Methods that Work*. 2nd print. ISBN 0-88385-450-3, Mathematical Association of America, Washington/DC, 1990.
- [Almansa & Lindeberg 00] A. Almansa and T. Lindeberg: *Fingerprint Enhancement by Shape Adaptation of Scale-Space Operators with Automatic Scale Selection*. IEEE Trans. on Image Processing **9**:12 (2000) 2027–2042.
- [Alvarez et al. 93] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel: *Axioms and Fundamental Equation of Image Processing*. Archive for Rational Mechanics and Analysis **123**:3 (1993) 199–257.
- [Alvarez et al. 00] L. Alvarez, J. Weickert, and J. Sánchez: *Reliable Estimation of Dense Optical Flow Fields with Large Displacements*. International Journal of Computer Vision **39**:1 (2000) 41–56.
- [Anandan 89] P. Anandan: *A Computational Framework and an Algorithm for the Measurement of Visual Motion*. International Journal of Computer Vision **2**:3 (1989) 283–310.
- [Andersson & Knutsson 03] M. Andersson and H. Knutsson: *Transformation of Local Spatio-Temporal Structure Tensor Fields*. In: Proceedings of „IEEE International Conference on Acoustics, Speech, and Signal Processing“ (ICASSP ’03), 6–10 April 2003, Hongkong, published on CD-ROM.
- [Babaud et al. 86] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda: *Uniqueness of the Gaussian Kernel for Scale-Space Filtering*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **8**:1 (1986) 26–33.
- [Bab-Hadiashar & Suter 98] A. Bab-Hadiashar and D. Suter: *Robust Optic Flow Computation*. International Journal of Computer Vision **29**:1 (1998) 59–77.
- [Bachem 03] A. Bachem: *Hohe Grauwertdynamik für eine sichelsystem-gestützte Roboterführung*. Entwurf, März 2003.

- [Bakker 02] P. Bakker: *Image Structure Analysis for Seismic Interpretation*. Proefschrift, 4 June 2002, promotor: L.J. van Vliet, toegevoegd promotor: P. W. Verbeek, Technische Universiteit Delft, Nederlande, auf Englisch.
- [Barron et al. 94] J. Barron, D. Fleet, and S. Beauchemin: *Performance of Optical Flow Techniques*. International Journal of Computer Vision **12:1** (1994) 43–77.
- [Beghdadi et al. 03] A. Beghdadi, M. Mesbah, and J. Monteil: *A Fast Incremental Approach for Accurate Measurement of the Displacement Field*. Image and Vision Computing **21:4** (2003) 383–399.
- [Bergen et al. 92] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani: *Hierarchical Model-Based Motion Estimation*. In: Proceedings „Second European Conference on Computer Vision“ (ECCV ’92), Giulio Sandini (Ed.), 19–22 May 1992, Santa Margherita Ligure, Italy, LNCS 588, Springer-Verlag, Berlin/Heidelberg 1992, pp. 237–252.
- [Betke & Makris 01] M. Betke and N. C. Makris: *Recognition, Resolution, and Complexity of Objects Subject to Affine Transformations*. Intern. Journal Computer Vision **44:1** (2001) 5–40.
- [Bigün & Granlund 87] J. Bigün and G. H. Granlund: *Optimal Orientation Detection of Linear Symmetry*. In: Proc. „1st International Conference on Computer Vision“ (ICCV ’87), 8–11 June 1987, London, The Computer Society of IEEE, New York, pp. 433–438.
- [Brox et al. 03] T. Brox, M. Rousson, R. Deriche, and J. Weickert: *Unsupervised Segmentation Incorporating Colour, Texture, and Motion*. Rapport de recherche, INRIA, Sophia Antipolis, France, März 2003.
- [Bruhn et al. 02] A. Bruhn, T. Jakob, M. Fischer, T. Kohlberger, J. Weickert, U. Brünning, and C. Schnörr: *Designing 3-D Nonlinear Diffusion Filters for High Performance Cluster Computing*. In: L. Van Gool (Ed.): *Pattern Recognition*. 24th DAGM Symposium, 16–18 September 2002, Zürich, LNCS 2449, Springer-Verlag, Berlin/Heidelberg 2002, pp. 290–297.
- [Cannami & Verri 92] M. Cannami and A. Verri: *Motion Analysis from First Order Properties of Optical Flow*. Computer Vision, Graphics, and Image Processing (CVGIP): Image Understanding **56:1** (1992) 90–107.
- [Chen et al. 87] J. S. Chen, A. Huertas, and G. Medioni. *Fast Convolution with Laplacian-of-Gaussian Masks*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **9:4** (1987) 584–590.

- [Chen et al. 89] J. S. Chen, A. Huertas, and G. Medioni: *Author's Reply*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **11**:12 (1989) 1332.
- [Crowley 01] O. Chomat, V. Colin de Verdière, and J. L. Crowley: *Recognizing Goldfish? or Local Scale Selection for Recognition Techniques*. Robotics and Autonomous Systems **35**:3-4 (2001) 191–200.
- [Crowley et al. 02] J. L. Crowley, O. Riff, and J. H. Piater: *Fast Computation of Characteristic Scale Using a Half-Octave Pyramid*. In: L. van Gool, B. Schiele (Eds.); Proceedings „Workshop on Cognitive Vision“, 19–20 September 2002, Zürich; auch <http://www.vision.ethz.ch/cogvis02/finalpapers/crowley.pdf> .
- [CSD-UWO-Archiv] *Image Sequence Archive* of the University of Western Ontario: <ftp://ftp.csd.uwo.ca/pub/vision/TESTDATA> .
- [Dam & Nielsen 01] E. Dam and M. Nielsen: *Exploring Non-linear Diffusion: The Diffusion Echo*. In: [Kerckhove 01], LNCS 2106, pp. 264–272.
- [Deriche 87] R. Deriche: *Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector*. International Journal of Computer Vision **1**:2 (1987) 167–187.
- [Duda & Hart 73] R. O. Duda and P. E. Hart: *Pattern Recognition and Scene Analysis*. ISBN 0-471-22361-1, John Wiley & Sons, New York/NY, 1973.
- [ECVision] European Research Network for Cognitive Computer Vision Systems. <http://www.ecvision.info> .
- [Farnebäck 01] G. Farnebäck: *Very High Accuracy Velocity Estimation Using Orientation Tensors, Parametric Motion, and Simultaneous Segmentation of Motion Field*. In: Proceedings „Eight International Conference on Computer Vision“ (ICCV 2001), Vancouver/Canada, 9–12 July 2001, IEEE Computer Society, Los Alamitos, CA/USA, Volume I, pp. 171–177.
- [Felsberg & Sommer 01] M. Felsberg and G. Sommer: *Scale Adaptive Filtering Derived from the Laplace Equation*. In: B. Radig, S. Florczyk (Eds.): *Pattern Recognition*. 23rd DAGM Symposium, 12–14 September 2001, München, LNCS 2191, Springer-Verlag, Berlin/Heidelberg 2001, pp. 124–131.
- [Fermüller et al. 01] C. Fermüller, D. Shulman, and Y. Aloimonos: *The Statistics of Optical Flow*. Computer Vision and Image Understanding (CVIU) **82**:1 (2001) 1–32.
- [Ferryman 00] J. M. Ferryman (Ed.): *PETS 2000*. Proceedings „First International Workshop on Performance Evaluation of Tracking

- and Surveillance“, 31 March 2000, Grenoble/Frankreich. <http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2000.pdf> .
- [Fisher et al. 94] R. Fisher, S. Perkins, A. Walker, and E. Wolfart: *Hypermedia Image Processing Reference*. Department of Artificial Intelligence, University of Edinburgh/UK, http://www.cee.hw.ac.uk/hipr/html/hipr_top.html .
- [Florack 00] L. Florack: *Spatio-Temporal Trade-Off for Scale-Space Filtering*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **22**:9 (2000) 1050–1055.
- [Foley et al. 90] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes: *Computer Graphics: Principles and Practice*. Addison-Wesley Systems Programming Series, Addison-Wesley Publishing Company, Reading/MA 1990.
- [Golub & van Loan 96] G. H. Golub and C. F. van Loan: *Matrix computations*. 3rd edition. Johns Hopkins University Press, Baltimore/MID (u.a.), 1996.
- [Grimson & Hildreth 85] W. E. L. Grimson and E. C. Hildreth: *Comments On “Digital Step Edges From Zero Crossings Of Second Directional Derivatives”*. (vgl. [Haralick 84]) IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **7**:1 (1985) 121–127.
- [Gunn 99] S. R. Gunn: *On the Discrete Representation of the Laplacian of Gaussian*. Pattern Recognition **32**:8 (1999) 1463–1472.
- [Hager & Belhumeur 98] G. D. Hager and P. N. Belhumeur: *Efficient Region Tracking With Parametric Models of Geometry and Illumination*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **20**:10 (1998) 1025–1039.
- [Haralick 84] R. M. Haralick: *Digital Step Edges from Zero Crossings of Second Directional Derivatives*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **6**:1 (1984) 58–68.
- [Haralick 85] R. M. Haralick: *Authors Reply to [Grimson & Hildreth 85]*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **7**:1 (1985) 127–129.
- [Haubecker & Jähne 97] H. Haubecker and B. Jähne: *A Tensor Approach for Precise Computation of Dense Displacement Vector Fields*. In: E. Paulus and F. M. Wahl (Hrsg.) *Mastererkennung 1997*. 19. DAGM-Symposium, 15.–17. September 1997, Braunschweig; Informatik aktuell, Springer-Verlag, Berlin Heidelberg 1997, pp. 199–208.
- [Haubecker & Spies 99] H. Haubecker and H. Spies: *Motion*. In: B. Jähne, H. Haubecker, and P. Geißler (Eds.): *Handbook of Computer Vision*

- and Applications*. Volume 2, chapter 13. Academic Press, San Diego/CA, 1999.
- [Healey & Kondempudy 94] G. Healey and R. Kondempudy: *Radiometric CCD Camera Calibration and Noise Estimation*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **16**:3 (1994) 267–276.
- [Henze 95] N. Henze: *Stochastik I – Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Skriptum zur gleichnamigen Vorlesung, Institut für Mathematische Stochastik, Universität Karlsruhe (TH), 1995.
- [Heuser 00] H. Heuser: *Lehrbuch der Analysis, Teil 2*. 11. Aufl., ISBN 3-519-42234-4, B. G. Teubner-Verlag, Stuttgart, Leipzig, Wiesbaden 2000.
- [Hladůvka & Gröllner 01] J. Hladůvka and E. Gröllner: *Direction-Driven Shape-Based Interpolation of Volume Data*. In: T. Ertl, B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Eds.); Proceedings „Vision, Modeling, and Visualization 2001“ (VMV '01), 21–23 November 2001, Stuttgart, IOS Press, Amsterdam/NL, pp. 113–120.
- [Horn & Schunck 81] B. K. P. Horn and B. G. Schunck: *Determining Optical Flow*. Artificial Intelligence **17** (1981) 185–203.
- [Hummel 03] B. Hummel: *Vergleich von 2D- und 3D-Ansätzen zur Verfolgung von Fahrzeugen in Videoaufnahmen*. Diplomarbeit am Institut für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH), April 2003.
- [IAKS-Archiv] *Bildfolgen-Archiv* des Instituts für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH): <http://kogs.iaks.uni-karlsruhe.de/image-sequences>.
- [IEEE 754-1985] Institute of Electrical and Electronics Engineers: *754-1985 IEEE Standard for Binary Floating-Point Arithmetic*. IEEE Computer Society, New York, 1985.
- [Iijima 62] T. Iijima: *Basic Theory of Pattern Normalization (For the Case of a Typical One Dimensional Pattern)*. Bulletin of the Electrotechnical Laboratory **26** (1962) 368–388 (auf Japanisch).
- [Kerckhove 01] M. Kerckhove (Ed.): *Scale-Space and Morphology in Computer Vision*. Proceedings „3rd International Conference on Scale-Space“, 7–8 July 2001, Vancouver, Canada, LNCS 2106, Springer-Verlag, Berlin/Heidelberg 2001.
- [Kim et al. 01] E. Y. Kim, S. W. Hwang, S. H. Park, and H. J. Kim: *Spatiotemporal Segmentation Using Genetic Algorithms*. Pattern

- Recognition **34:10** (2001) 2063–2066.
- [Kirsch 71] R. A. Kirsch: *Computer Determination of the Constituent Structure of Biological Images*. Computers and Biomedical Research **4:3** (1971) 315–328.
- [Klinger 71] A. Klinger: *Pattern and Search Statistics*. In: J. S. Rustagi (Ed.): *Optimizing Methods in Statistics*. Academic Press, New York, 1971.
- [Koenderink 84] J. J. Koenderink: *The Structure of Images*. Biological Cybernetics **50** (1984) 363–370.
- [Koenderink 88] J. Koenderink: *Scale-Time*. Biological Cybernetics **58** (1988) 159–162.
- [Kollnig 95] H. Kollnig: *Ermittlung von Verkehrsgeschehen durch Bildfolgenauswertung*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Februar 1995; DISKI **88**; St. Augustin: Infix, 1995.
- [Korn 88] A. F. Korn: *Toward a Symbolic Representation of Intensifying Changes in Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **10:5** (1988) 610–625.
- [Lai & Vennuri 97] S.-H. Lai and B. C. Vennuri: *Reliable and Efficient Computation of Optical Flow*. International Journal of Computer Vision **29:2** (1998) 87–105.
- [Lindeberg 93] T. Lindeberg: *Discrete Derivative Approximations with Scale-Space Properties: A Basis for Low-Level Feature Extraction*. Journal of Mathematical Imaging and Vision **3:4** (1993) 349–376.
- [Lindeberg & Gårding 94] T. Lindeberg and J. Gårding: *Shape-Adapted Smoothing in Estimation of 3-D Depth Cues from Affine Distortions of Local 2-D Brightness Structure*. In: Proceedings „3rd European Conference on Computer Vision“ (ECCV '94), J. O. Eklundh (Ed.), 2–6 May 1994, Stockholm, LNCS 800, Springer-Verlag, Berlin/Heidelberg 1994, pp. 389–400.
- [Lindeberg 95] T. Lindeberg: *Direct Estimation of Affine Image Deformations Using Visual Front-End Operations with Automatic Scale Selection*. In: Proceedings „5th International Conference on Computer Vision“ (ICCV '95), 20–23 June 1995, Cambridge/MA, IEEE Computer Society 1995, pp. 134–141.
- [Lindeberg 97] T. Lindeberg: *Linear Spatio-Temporal Scale-Space*. In: B. M. Ter Haar Romeny et al. (Eds.): Proceedings „First International Conference on Scale-Space“ (Scale-Space '97), 2–4 July 1997, Utrecht, the Netherlands, LNCS 1252, Springer-Verlag, Berlin/Heidelberg 1997, pp. 113–127.

- [Lindeberg 98a] T. Lindeberg: *Feature Detection with Automatic Scale Selection*. International Journal of Computer Vision **30:2** (1998) 79–116.
- [Lindeberg 98b] T. Lindeberg: *Edge Detection and Ridge Detection with Automatic Scale Selection*. International Journal of Computer Vision **30:2** (1998) 117–154.
- [Lucas & Kanade 81] B. Lucas and T. Kanade: *An Iterative Image Registration Technique with an Application to Stereo Vision*. In: Proc. „7th International Joint Conference on Artificial Intelligence“ (IJCAI 81), 24–28 August 1981, Vancouver/BC, 674–679.
- [Marr & Hildreth 80] D. Marr and E. C. Hildreth: *Theory of Edge Detection*. Proceedings of the Royal Society of London **B207** (1980) 187–217.
- [Martin et al. 01] P. T. Martin, J. Perrin, and B. Hansen: *Incident Detection Algorithm Evaluation*. Report prepared for Utah Department of Transportation, Upper Great Plains Transportation Institute (UGPTI), North Dakota State University (NDSU), March 2001, http://www.ndsu.nodak.edu/ndsu/ugpti/MPC_Pubs/pdf/MPC01-122.pdf .
- [Ménin & Pérez 98] E. Ménin and P. Pérez: *Dense Estimation and Object-Based Segmentation of the Optical Flow with Robust Techniques*. IEEE Transactions on Image Processing **7:5** (1998) 703–719.
- [Mester & Hötter 95] R. Mester and M. Hötter: *Robust Displacement Vector Estimation Including a Statistical Error Analysis*. In: Proc. „5th International Conference on Image Processing and its Applications“, 3–6 July 1995, Edinburgh/UK, Institution of Electrical Engineers (IEE), London, 1995, pp. 168–172.
- [Middendorf 99] M. Middendorf: *Untersuchung von Schätzprozessen in der Bildfolgenauswertung*. Diplomarbeit am Institut für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH), August 1999.
- [Middendorf & Nagel 00] M. Middendorf and H.-H. Nagel: *Vehicle Tracking Using Adaptive Optical Flow Estimation*. In: Proceedings „First International Workshop on Performance Evaluation of Tracking and Surveillance“ (PETS 2000), 31 March 2000, Grenoble/Frankreich, pp. 38–45. Siehe auch [Ferryman 00].
- [Middendorf & Nagel 01] M. Middendorf and H.-H. Nagel: *Estimation and Interpretation of Discontinuities in Optical Flow Fields*. In: Proceedings „Eight International Conference on Computer Vision“ (ICCV 2001), Vancouver/Canada, 9–12 July 2001, IEEE Computer Society, Los Alamitos, CA/USA, Volume I, pp. 178–183.

- [Middendorf 02] M. Middendorf: *Untersuchungen zur bildgestützten Objektverfolgung in Videobildfolgen bei bewegter Kamera*. Interner Bericht des Instituts für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH), März 2002.
- [Mück 00] K. Mück: *Rechnergestützte Erkennung und Beschreibung innerstädtischer Straßenkrenzungen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juli 2000; Selbstverlag, 2000; auch <http://www.ubka.uni-karlsruhe.de/cgi-bin/psview?document=2000%2FInformatik%2F14>.
- [Mück et al. 00] K. Mück, H.-H. Nagel, and M. Middendorf: *Data-Driven Extraction of Curved Intersection Lanemarks from Road Traffic Image Sequences*. In: Proceedings „6th European Conference on Computer Vision“ (ECCV 2000), D. Vernon (Ed.), 26 June–1 July 2000, Dublin, Ireland, LNCS 1843, Springer-Verlag, Berlin/Heidelberg 2000, Part II, pp. 411–427.
- [Nagel 81] H.-H. Nagel: *What Can We Learn from Applications?* In: T. S. Huang (Ed.): *Image Sequence Analysis*, Information Sciences (Vol. V), Springer-Verlag, Berlin/Heidelberg 1981, pp. 20–213.
- [Nagel 95] H.-H. Nagel: *Optical Flow Estimation and the Interaction Between Measurement Errors at Adjacent Pixel Positions*. International Journal of Computer Vision **15**:3 (1995) 271–288.
- [Nagel & Gehrke 98a] H.-H. Nagel and A. Gehrke: *Spatiotemporally Adaptive Estimation and Segmentation of Optic Flow Fields*. In: Proceedings „5th European Conference on Computer Vision“ (ECCV 98), Vol. II, LNCS 1407, Springer-Verlag, Berlin/Heidelberg 1998, pp. 86–102.
- [Nagel & Gehrke 98b] H.-H. Nagel and A. Gehrke: *Bildbereichsbasierte Verfolgung von Straßenfahrzeugen durch adaptive Schätzung und Segmentierung von Optischen-Fluß-Feldern*. In: P. Levi et al. (Hrsg.) *Mustererkennung 1998*; Springer-Verlag, Berlin/Heidelberg 1998, pp. 314–321.
- [Nagel & Gehrke 98c] H.-H. Nagel and A. Gehrke: *Tracking Based on Spatiotemporally Adaptive Estimation and Segmentation of Optic Flow Fields*. Institut für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH), Entwurf, 1998.
- [Nagel & Haag 98] H.-H. Nagel and M. Haag: *Bias-Corrected Optical Flow Estimation for Road Vehicle Tracking*. In: Proc. „6th International Conference on Computer Vision“ (ICCV 98), 4–7 January

- ry 1998, Bombay, India, Narosa Publishing House New Delhi, India 1998, pp. 1006–1011.
- [Nagel 00] H.-H. Nagel: *Image Sequence Evaluation: 30 Years and Still Going Strong*. In: Proc. „15th Intern. Conf. Pattern Recognition“ (ICPR 2000), A. Sanfeliu, J.J. Villanueva et al. (Eds.), 3–7 Sept. 2000, Barcelona, Spain, Vol. 1, pp. 149–158.
- [Nagel 01] H.-H. Nagel: *Towards a Cognitive Vision System*. Vorläufiger Bericht zum Projekt „CogViSys“, September 2001, http://kogs.iaks.uni-karlsruhe.de/CogViSys/kogs_CogViSys_homepage_V22_1.in.pdf .
- [Nagel 02] H.-H. Nagel: *Kognitive Systeme, Materialien zur Vorlesung im SS 2002*. Vorlesungsskript, Juli 2002, <http://lernserver.ira.uka.de> (Zugang beschränkt).
- [Nathan 66] R. Nathan: *Digital Video-Data Handling*. Technical Report 32-877, Jet Propulsion Laboratory, California Institute of Technology, Pasadena/CA, 5 January 1966.
- [Nitzberg & Shiota 92] M. Nitzberg and T. Shiota: *Nonlinear Image Filtering with Edge and Corner Enhancement*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI **14**:8 (1992) 826–833.
- [Ortega & Rheinboldt 70] J. M. Ortega and W. C. Rheinboldt: *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, San Diego, 1970.
- [Otte 94] M. Otte: *Extraktion von linienförmigen Merkmalen und Ermittlung des optischen Flusses mit seinen Ableitungen aus Voll- und Halbbildfolgen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Februar 1994; DISKI **63**; Sankt Augustin: Infix, 1994.
- [Otte & Nagel 95] *Estimation of Optical Flow Based on Higher-order Spatiotemporal Derivatives in Interlaced and Non-interlaced Image Sequences*. Artificial Intelligence **78**:1 (1995) 5–43.
- [Ottlik & Nagel 03] A. Ottlik and H.-H. Nagel: *On Consistent Discrimination Between Directed and Diffuse Outdoor Illumination*. In: B. Michaelis, G. Krell (Eds.): *Pattern Recognition*. 25th DAGM Symposium, 10–12 September 2003, Magdeburg, LNCS 2781, Springer-Verlag, Berlin/Heidelberg 2003, pp. 52–59.
- [Pedersen & Nielsen 01] K. S. Pedersen and M. Nielsen: *Computing Optical Flow by Scale-Space Integration of Normal Flow*. In: [Kerckhove 01], LNCS 2106, pp. 14–25.
- [Penrose 55] R. Penrose: *A Generalized Inverse for Matrices*. Proceedings of the Cambridge Philosophical Society **51** (1955) 406–413.

- [PETS-Archiv] J. M. Ferryman: *PETS Test-data Sets*. Computer Vision Group of the University of Reading/UK, 2000–2002: <ftp://pets.rdg.ac.uk>.
- [Press et al. 90] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery: *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge Univ. Press, Cambridge/UK, 1990.
- [Pritchard 77] D. H. Pritchard: *U. S. Color Television Fundamentals – A Review*. IEEE Trans. on Consumer Electronics CE **23**:4 (1977) 467–478.
- [Roberts 65] L. G. Roberts: *Machine Perception of Three-dimensional Solids*. In: J. P. Tippett et al. (Eds.): *Optical and Electro-optical Information Processing*. MIT Press, Cambridge/MA 1965, pp. 159–197.
- [Rosenfeld & Thurston 71] A. Rosenfeld and M. Thurston: *Edge and Curve Detection for Visual Scene Analysis*. IEEE Transactions on Computers **C-20**:5 (1971) 562–569.
- [Rosenfeld 84] A. Rosenfeld (Ed.): *Multiresolution Image Processing and Analysis*. Springer-Verlag, Berlin/Heidelberg 1984.
- [Rudin 74] W. Rudin: *Functional Analysis*. McGraw-Hill, New York/NY, 1974, pp. 135–226.
- [Russ 95] J. Russ: *The Image Processing Handbook*. 2nd edition, CRC Press, Boca Raton/FL 1995.
- [Schäfer 00] Sven Schäfer: *Verwendung von Farbinformationen zur Segmentierung und Verfolgung von Fahrzeugen in Videobildfolgen*. Diplomarbeit am Institut für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH), Juni 2000.
- [Scharr et al. 97] H. Scharr, S. Körkkel und B. Jähne: *Numerische Isometrieop-
timierung von FIR Filtern mittels Querglättung*. In: E. Paulus and F. M. Wahl (Hrsg.), *Mustererkennung 1997*. 19. DAGM-Symposium, 15.–17. September 1997, Braunschweig; Informatik aktuell, Springer-Verlag, Berlin Heidelberg 1997, pp. 199–208.
- [Schwetlick 79] H. Schwetlick: *Numerische Lösung nichtlinearer Gleichungen*. 1. Auflage, ISBN 3-486-21731-3, Oldenbourg Verlag, München, Wien, 1979.
- [Sobel 70] I. Sobel: *Camera Models and Machine Perception*. AIM-21, Stanford Artificial Intelligence Lab, Palo Alto 1970. Siehe auch [Duda & Hart 73, S. 271f].
- [Sotak & Boyer 89] G. E. Sotak and K. L. Boyer: *Comments on „Fast Convolution with Laplacian-of-Gaussian Masks“*. IEEE Transactions

- on Pattern Analysis and Machine Intelligence PAMI **11:12** (1989) 1329–1332.
- [Sun 02] C. Sun: *Fast Optical Flow Using 3D Shortest Path Techniques*. Image and Vision Computing **20:13–14** (2002) 981–991.
- [Thompson 98] W. B. Thompson: *Exploiting Discontinuities in Optical Flow*. International Journal of Computer Vision **30:3** (1998) 163–173.
- [Tsin et al. 01] Y. Tsin, V. Ramesh, and T. Kanade: *Statistical Calibration of CCD Imaging Process*. In: Proceedings „Eight International Conference on Computer Vision“ (ICCV 2001), Vancouver/Canada, 9–12 July 2001, IEEE Computer Society, Los Alamitos, CA/USA, Volume I, pp. 480–487.
- [Uras et al. 88] S. Uras, F. Girosi, A. Verri, and V. Torre, *A Computational Approach to Motion Perception*. Biological Cybernetics **60** (1988) 79–87.
- [Van den Boomgaard & van der Weij 01] R. van den Boomgaard and R. van der Weij: *Gaussian Convolution. Numerical Approximations Based on Interpolation*. In: [Kerckhove 01], LNCS 2106, pp. 205–214.
- [Van Huffel & Vandewalle 91] S. Van Huffel and J. Vandewalle: *The Total Least Squares Problem: Computational Aspects and Analysis*. Frontiers in Applied Mathematics. SIAM, Philadelphia, 1991.
- [Venegas-Andraca 02] S. E. Venegas-Andraca: *Video Shot Segmentation*. MSc-Thesis, Robotics Research Group, Department of Engineering Science, University of Oxford/UK, 2002, <http://nodens.physics.ox.ac.uk/~venegas-andraca/msc-thesis.pdf>.
- [Vliet et al. 98] L. J. Vliet, I. T. Young, and P. W. Verbeek: *Recursive Gaussian Derivative Filters*. In: Proceedings „14rd International Conference on Pattern Recognition“ ICPR ’98, 16–20 August 1998, Brisbane (Australia), IEEE Computer Society Press, Los Alamitos, CA/USA, Vol. I, pp. 509–514.
- [Weber & Malik 93] J. Weber and J. Malik: *Robust Computation of Optical Flow in a Multi-Scale Differential Framework*. In: Proceedings „4th International Conference on Computer Vision“ (ICCV ’93), 11–14 May 1993, Berlin, pp. 12–20.
- [Weickert, S. Ishikawa, and A. Imiya: *Linear Scale-Space Has First Been Proposed in Japan*. Journal of Mathematical Imaging and Vision **10:3** (1999) 237–252.
- [Weickert & Brox 02] J. Weickert and T. Brox: *Diffusion and Regularization of Vector- and Matrix-valued Images*. In: M. Z. Nashed and O. Scherzer (Eds.): *Inverse Problems, Image Analysis, and*

- Medical Imaging*. Contemporary Mathematics, Vol. 313, American Mathematical Society (AMS), Providence/RI, 2002.
- [Wiskott 99] L. Wiskott: *Segmentation from Motion: Combining Gabor and Mallat-Wavelets to Overcome Aperture and Correspondence Problem*. Pattern Recognition **32:10** (1999) 1751–1766.
- [Witkin 83] A. P. Witkin: *Scale Space Filtering*. In: Proc. „Int. Joint Conference on Artificial Intelligence“ (IJCAI), Karlsruhe, 8–12 August 1983, pp. 1019–1021.
- [Xie et al. 02] B. Xie, V. Ramesh, and T. Boult: *Sudden Illumination Change Detection Using Order Consistency*. In: David Sutter (Chair), Proceedings „Workshop on Statistical Methods in Video Processing 2002“ (SMVP '02), 1–2 June 2002, Copenhagen/Denmark, available online <http://www.ds.eng.monash.edu.au/smvp> .
- [Yacoub & Davis 99] Y. Yacoub and L. S. Davis: *Temporal Multi-Scale Models for Flow and Acceleration*. International Journal of Computer Vision **32:2** (1999) 147–163.
- [Zacharias 03] F. Zacharias: *Untersuchung des Einflusses verschiedener Parameter auf die Detektion von verdeckenden Objekten in der Verfolgung von Fahrzeugen im Straßenverkehr*. Interner Bericht des Instituts für Algorithmen und Kognitive Systeme der Fakultät für Informatik der Universität Karlsruhe (TH), März 2003.

Verzeichnis der Annahmen

- Nr. 1** Die verarbeiteten Bilder entstehen in systematischer Art und Weise durch einen mathematisch zumindest annähernd beschreibbaren physikalischen Abbildungsprozess. (Seite 6)
- Nr. 2** Von Bild zu Bild erfolgen nur kleine, systematische Änderungen der Szene und der Kameraparameter. (Seite 7)
- Nr. 3** Bei sorgfältigem Vorgehen ist der geschätzte Optische Fluss eine gute Näherung an die tatsächliche Verschiebungsrate. (Seite 10)
- Nr. 4** Die Helligkeit des Abbilds eines Szenenpunkts bleibt in aufeinander folgenden Aufnahmen gleich. (Seite 11)
- Nr. 5** Die Bewegung eines Bildpunktes ist in kleinen Zeitintervallen gleichförmig. (Seite 12)
- Nr. 6** Die Bildfunktion g ist differenzierbar. (Seite 12)
- Nr. 7** Die Bildfunktion ändert sich in der lokalen Umgebung jedes Punktes näherungsweise linear. (Seite 13)
- Nr. 8** Alle Punkte in einer hinreichend großen räumlich-zeitlichen Umgebung des jeweils betrachteten Punktes (Aufpunkt) bewegen sich im Wesentlichen genauso wie der Aufpunkt. (Seite 13)
- Nr. 9** Alle Grauwerte sind mit einem additiven, normalverteilten Fehler (Rauschen) mit Mittelwert 0 und positions-, helligkeits- sowie zeitunabhängiger Varianz σ^2 behaftet, dessen Verteilungen voneinander unabhängig sind. (Seite 19)
- Nr. 10** Die Ableitungsschätzung ist mit einem normalverteilten Rauschen behaftet, das für alle partiellen ersten Ableitungen an allen Bildpunkten derselben Verteilung unterliegt und dessen Verteilungen voneinander unabhängig sind. (Seite 19)

Abkürzungsverzeichnis

A		
A+1	abgetastete Maskenkoeffizienten, Maske beiderseits um 1 Pixel vergrößert	53
A-	abgetastete Maskenkoeffizienten (ohne jede Korrektur)	53
AGR	abgetastete, gradienten-basiert randkorrigierte Maskenkoeffizienten	56
AG	abgetastete, gradienten-basiert skalierte Maskenkoeffizienten	56
AK	abgetastete Maskenkoeffizienten mit Skalierung nach [Korn 88]	56
AR	abgetastete Maskenkoeffizienten mit Randkorrektur	53
AS	abgetastete, skalierte Maskenkoeffizienten (Tiefpass)	53
AS _{σ}	abgetastete Maskenkoeffizienten, σ -basierte Skalierung (Ableitung)	56
AS _{mod}	abgetastete Maskenkoeffizienten, modifizierte σ -basierte Skalierung (Ableitung)	56
B		
BBE	Bild-Bereichs-Ebene	2
BBH	Bild-Bereichs-Hinweise	2
BPE	Begriffliche-Primitiven-Ebene	2
BSB	Bildsignalbeschreibungen	2
BSE	Bild-Signal-Ebene	1
C		
OCD	Charge Coupled Device, ladungsgelinkoppeltes Bauelement (gebräuchlichster Typ des Sensorchips in Digitalkameras)	19

D			
DGD-OF	Unterkategorie von DGD		23
DGD-PP	Unterkategorie von DGD		23
DGD	Kategorie „Dominierende Gradientenrichtung“		23
G			
GST	Grauwert-Struktur-Tensor		21
H			
HB	Halbbild (bei zeilenversetzten Bildern)		9
HSI	Hue-Saturation-Intensity (Farbsystem)		8
I			
I+1	integrierte Maskenkoeffizienten, Maske beiderseits um 1 Pixel vergrößert		53
I-	integrierte Maskenkoeffizienten (ohne jede Korrektur)		53
I _{GR}	integrierte, gradienten-basiert randkorrigierte Maskenkoeffizienten		56
I _G	integrierte, gradienten-basiert skalierte Maskenkoeffizienten		56
I _K	integrierte Maskenkoeffizienten mit Skalierung nach [Korn 88]		56
I _M	integrierte Maskenkoeffizienten mit Rand- und Mittelkorrektur		56
I _R	integrierte Maskenkoeffizienten mit Randkorrektur	53, 56	
I _S	integrierte, skalierte Maskenkoeffizienten (Tiefpass)		53
I _σ	integrierte Maskenkoeffizienten, σ -basierte Skalierung (Ableitung)		56
I _{S_{mod}}	integrierte Maskenkoeffizienten, modifizierte σ -basierte Skalierung (Ableitung)		56
J			
JPEG	Joint Photographic Experts Group		8
N			
NF	Normal-Fluss		69
NSE	Natürlich-Sprachliche-Ebene		2

NTSC	National Television System Committee (US-Fernseh-Standard)	8
N	Kategorie „Neutral“	23
O		
o. B. d. A.	ohne Beschränkung der Allgemeinheit	32
OA	Objektabbild	173
OFCE	Optical Flow Constraint Equation, OF-Beschränkungsgleichung	12
OFD	Kategorie „OF-Diskontinuität“	24
OF	Optischer Fluss	9
P		
PAL	Phase Alternating Line (europäischer Fernseh-Standard)	50
PDE	partial differential equation	16
R		
RGB	Rot-Grün-Blau (Farbsystem)	8
ROF	Kategorie „Regulärer Optischer Fluss“	24
S		
SAE	Sensor-Aktuator-Ebene	1
SBE	Szenen-Bereichs-Ebene	2
T		
TLS	Total Least Squares	20
U		
u. d. N.	unter der Nebenbedingung	20
V		
VRE	Verhaltens-Repräsentations-Ebene	2
VR	Verschiebungsrate	9

Symbolverzeichnis

$\bowtie_{\text{DGD-PP}}$	Winkelschwellwert „DGD-PP“	23
\otimes	Faltungsoperator	15
Δ	Laplaceoperator $\Delta u = u_{xx} + u_{yy}$	16
$\ \cdot\ _2$	Euklidnorm	75
$\ \cdot\ _F$	Frobeniusnorm	125
α_i	Ausdehnungsfaktoren	114
$\alpha_{a_+}, \mathbf{E}_0$	Winkel zwischen dem Vektor \mathbf{v}_+ und der Invarianzebene \mathbf{E}_0	68
$c_{1D}^{(i)}, c_{3D}^{(i)}$	erwarteter, theoretischer Maskenfehler	29
$\frac{\partial g}{\partial x}$	partiellen Ableitung der Funktion g nach x	12
η_A	Überlappungsfaktor für Ableitungen	90
η_M	Überlappungsfaktor für Mittelungsmasken	103
\mathbf{E}_0	Invarianzebene	67
\mathbf{e}_{\min}	Eigenvektors zum kleinsten Eigenwert λ_{\min}	21
erf	Fehlerfunktion: $\text{erf}(x) = 2/\sqrt{\pi} \cdot \int_0^x e^{-t^2} dt$	28
f_A	Adaptionfunktion	110
f_{sr}	Skaleniterationsfunktion	154
f_S	Schätzfunktion	153
f_{TI}	Tensor-Iterationsfunktion	154
$F^{(i)}(x_1, x_2, \sigma)$	Fläche unter der i -ten Ableitung der Gaußfunktion im Intervall $[x_1, x_2]$	28
$f^{(i)}(x, \sigma)$	i -te Ableitung der Gaußfunktion	27
g	Grauwertfunktion, Bildfunktion	11
g_{\max}	größter möglicher Grauwert (abhängig von interner Darstellung)	157
g_{\min}	kleinster möglicher Grauwert	157
g_x, g_y, g_t	abkürzende Schreibweise für die partiellen Ableitungen (Gradienten) der Bildfunktion $\frac{\partial g}{\partial x}$, $\frac{\partial g}{\partial y}$ und $\frac{\partial g}{\partial t}$	12

∇g	Gradientenvektor $(g_x, g_y, g_t)^T$	12
i_{\max}	maximale Verkleinerungsstufe bei der Bildpyramide	194
$I_{n \times n}$	Einheitsmatrix der Größe $n \times n$	34
$k_i^{(i)}(n, \sigma)$	Maskenkoeffizienten (eindimensionale Maske)	36
$\tilde{k}_i^{(i)}(n, \sigma)$	korrigierte Maskenkoeffizienten (eindimensionale Masken)	37
λ	Eigenwert	21
$\tilde{\lambda}_i$	normierter Eigenwert	114
λ_{\max}	größtmöglicher Eigenwert des GST	156
λ_{\min}	kleinster Eigenwert	21
μ	obere Intervallgrenze des Maskenbereichs	29
m	Maskenbereichsvektor	31
M_{GST}	Menge der möglichen GST	153
M_{Σ}	Menge der möglichen Kovarianzmatrizen	153
\mathcal{M}	Maskenbereich	27
N	Maskenausdehnung	27
\mathbb{N}	Menge der natürlichen Zahlen (ohne 0)	36
$N(\mu, \sigma^2)$	Normalverteilung mit Mittelwert μ und Varianz σ^2	60
ψ_{v_+, u_+}	Winkelmaß nach [Barron et al. 94]	74
$p(\nabla g, v_+)$	Fehlerprojektion	95
\mathbb{R}	Menge der reellen Zahlen	31
\mathbb{R}^n	Menge der reellen n -dimensionalen Vektoren	31
$\mathbb{R}^{m \times n}$	Menge der reellen $m \times n$ -Matrizen	153
σ	Standardabweichung der eindimensionalen Gaußfunktion	27
Σ	Kovarianzmatrix einer mehrdimensionalen Gaußfunktion	31
Σ_A	Ableitungs-Kovarianzmatrix	18
Σ_M	Mittlungs-Kovarianzmatrix	20
S_G	Gesamtschiefe	167
S_R	Richtungsschiefe	167
$\sigma_x, \sigma_y, \sigma_t$	Standardabweichung einer mehrdimensionalen Gaußfunktion mit Kovarianzmatrix in Diagonalgestalt in x -, y - bzw. t -Richtung	31
$\sigma_{\min}^2, \sigma_{\max}^2$	minimale und maximale Varianz in Hauptachsenrichtung der im Adaptionsschritt bestimmten Kovarianzmatrizen	114
$\tau_{\text{DGD-PP}}$	Kategorisierungsschwellwert „DGD-PP“	23
τ_{DGD}	Kategorisierungsschwellwert „DGD“	23
τ_N	Kategorisierungsschwellwert „Neutral“	23

TOPD	Kategorisierungsschwellwert „OFD“	24
t	Zeitpunkt	11
T_s	Skalenoperator	16
\mathbf{u}	OF-Vektor $(u_1, u_2)^T$	9
\mathbf{u}_+	erweiterter OF-Vektor $(u_1, u_2, 1)^T$	12
\mathbf{u}_{N+}	erweiterter NF-Vektor $(n_1, n_2, 1)^T$	69
\mathbf{u}_N	Normal-Fluss-Vektor	69
$\mathcal{U}(\mathbf{x}_0)$	räumlich-zeitliche Umgebung von \mathbf{x}_0	13
\mathbf{v}	VR-Vektor	9
\mathbf{v}_+	erweiterter VR-Vektor $(v_1, v_2, 1)^T$	68
$w(\Sigma_M, \mathbf{x}, \mathbf{x}_0)$	Wert einer am Punkt \mathbf{x}_0 zentrierten dreidimensionalen Mittelungsmaske am Punkt \mathbf{x} mit Mittelungskovarianzmatrix Σ_M	103
$w_x(\Sigma_A, \mathbf{x}, \mathbf{x}_0)$	Wert einer am Punkt \mathbf{x}_0 zentrierten dreidimensionalen Maske zur Berechnung der partiellen x -Ableitung am Punkt \mathbf{x} mit Kovarianzmatrix Σ_A	90
x, y	räumliche Koordinaten eines Bildpunktes, Pixelposition	11
\mathbb{Z}	Menge der ganzen Zahlen	30
$z^{(i)}(\sigma)$	der vom Ursprung aus gesehen am weitesten entfernt liegende nicht-negative Nulldurchgang der i -ten Ableitung	29

Index

Symbole

- Δ , *siehe* Laplace-Operator
- ∇ , *siehe* Nabla-Operator
- \otimes , *siehe* Faltungsoperator

A

- Abbildungsprozess, 6, 9
- Abgeschlossenheit, 154
- Ableitungsbestimmung, 2
- Ableitungsmaske, 14
 - Gradienten-orientierte Randkorrektur, 45
- Gradienten-orientierte Skalierung, 43
- Mittelkorrektur, 41
- modifizierte σ -basierte Skalierung, 40
- Randkorrektur, 41
- σ -basierte Skalierung, 39
- Skalierung nach Korn, 40
- Ableitungsoperator, 15
- Abtastung, 28
- Adaption, 106
 - Grundidee, 113
 - lineare –, 114
 - quasi-inverse –, 116
 - ungekoppelte quasi-inverse –, 166
- Adaptionfunktion, 110, 154, 157
- anisotrope Faltungsmaske, 33
- Ansatz
 - Faltungs-, 17
 - Variations-, 17
- Ausdehnungsfaktoren, 114
- Auswahl, 106

B

- Bandpassfilter, 14
- Bauer-Fike

Theorem von –, 154

- BFE, *siehe* Bild-Bereichs-Ebene
 - BBH, *siehe* Bild-Bereichs-Hinweise
 - Begriffliche-Primitiven-Ebene, 2
 - Belichtung, 11
 - Bewegung, 11
 - Bewegungsschätzung, 3
 - Bild, 1, **6**
 - zeilenversetzt, 9, 50
 - Bild-Bereichs-Ebene, 2
 - Bild-Bereichs-Hinweise, 2
 - Bild-Signal-Ebene, 1
 - Bildauswertung, 6
 - Bildfolgen, 107
 - Bildprimitive, 6
 - Bildpyramide, 192
 - Bildsegmentierung, 4
 - Bildsignalbeschreibungen, 2
 - Bildverarbeitung, 6
 - Binomialfilter, 36
 - Box-Müller-Transformation, 66
 - BPE, *siehe* Begriffliche-Primitiven-Ebene
 - Brouwer
 - Fixpunktsatz von –, 154
 - BSB, *siehe* Bildsignalbeschreibungen
 - BSE, *siehe* Bild-Signal-Ebene
- ## C
- $c_{1D}^{(i)}$, $c_{3D}^{(i)}$, 29
 - CCD, 19
- ## D
- Dämpfung, 161
 - DGD, 23
 - DGD-OF, 23
 - DGD-PP, 23

- differentielle Verfahren, 10, 12
 - Diskontinuität, 24, 173
 - Diskontinuitätsbereich, 119, **173**
 - Diskontinuitätsflecken, 180
 - Diskontinuitätshinweis, 173
 - Diskontinuitätsmaske, 189
 - Diskontinuitätsmaß, 173, 187
 - Diskontinuitätsmauer, 182
 - diskrete Faltung, 14, 36
 - Diskretisierung, 9, 26
 - doppelt-adaptives Verfahren, 112
- E**
- Ebene
 - Begriffliche-Primitiven-, 2
 - Bild-Bereichs-, 2
 - Bild-Signal-, 1
 - Natürlich-Sprachliche-, 2
 - Sensor-Aktuator-, 1
 - Szenen-Bereichs-, 2
 - Verhaltens-Repräsentations-, 2
 - Eigenvektor, 21
 - Eigenwert, 21
 - problem, 67
 - Eigenwerte
 - normierte, 114
 - einfach-adaptives Verfahren, 110
 - Euklid-Norm, 149
- F**
- $f^{(\theta)}(x, \sigma)$, 27, 36
 - $F^{(\theta)}(x_1, x_2, \sigma)$, 28
 - f_A , *siehe* Adaptionsfunktion
 - f_S , *siehe* Schätzfunktion
 - f_{SI} , *siehe* Skaleniterationsfunktion
 - f_{TI} , *siehe* Tensor-Iterationsfunktion
 - Fahrzeugverfolgung, 3
 - Faltung
 - diskrete –, 14, 36
 - Faltungsansatz, 17
 - Faltungsformel, 15
 - Faltungsmaske, 14, 27
 - anisotrope, 33
 - Ausdehnung, 27
 - Bestimmung, 25
 - Breite, 27
 - endliche Ausdehnung, 27
 - Gradienten-orientierte Randkorrektur, 45
 - Gradienten-orientierte Skalierung, 43
 - Größe, 27
 - isotrope, 31
 - Mittelkorrektur, 41
 - modifizierte σ -basierte Skalierung, 40
 - Randkorrektur, 38, 41
 - σ -basierte Skalierung, 39
 - Skalierung, 38
 - Skalierung nach Korn, 40
 - Faltungsoperator, 15, 36
 - Farbbild, 8
 - Umrechnung in Grauwerte, 8
 - Farbkanal, 8
 - Farbsystem, 8
 - HSI-, 8
 - RGB-, 8
 - YIQ-, 8
 - Fehlerfortpflanzung, 60
 - Fehlerprojektion, 95
 - Fehlerquadrate
 - kleinste Summe der –, 47
 - kleinste Summe der absoluten –, 20
 - kleinster Median der –, 20
 - Filter, 14
 - Bandpass-, 14
 - Binomial-, 36
 - Hochpass-, 14
 - Tiefpass-, 14
 - Filterung
 - Tiefpass-, 2
 - Fixpunkt, 152
 - Fixpunktsatz von Brouwer, 154
 - Probenius-Norm, 150
 - Funktion
 - Adaptions-, 110, 154
 - Schätz-, 153
 - Skalen-Iterations-, 154
 - Tensor-Iterations-, 154

- G**
- Gauß'scher Skalenraum, 16
 - Gaußfunktion, 15, 18
 - Gaußglocke, 18
 - Gesamt-Überlappungsfaktor, 92
 - Gesamtschiefe, 167
 - Glättungsmaske, 14
 - Glättungsoperator, 15
 - Glättungswirkung, 15
 - Glätteheitsforderung, 13
 - Gleichung
 - Laplace-, 17
 - lineare homogene Diffusions-, 16
 - Wärmeleitungs-, 16
 - Gradienten-orientierte
 - Randkorrektur, 45
 - Skalierung, 43
 - Gradientenvektor, 12, 67
 - Grauwert, 8
 - Grauwert-Ableitungstensor, *siehe* Grauwert-Struktur-Tensor
 - Grauwertbild, 7
 - Grauwert-Struktur-Tensor, 5, 6, **21**, 68
 - Bestimmung
 - doppelt-adaptives Verfahren, 112
 - einfach-adaptives Verfahren, 110
 - fortschreibend einfach-adaptives Verfahren, 111
 - mehrfach-adaptives Verfahren, 113
 - Kategorisierung, **22**, 118, 173
 - Schätzfunktion, 153
 - Grauwertkante, 173
 - Grauwertsprung, 52, 53
 - Grauwertübergang, 52
 - linearer –, 53
 - GST, *siehe* Grauwert-Struktur-Tensor
- H**
- Halbbild, 9
 - Halbgruppen-Eigenschaft, 16, 26
 - Hauptachsenrichtung, 106, 153
 - Helligkeit, 8
 - Hochpassfilter, 14
- HSL-Farbsystem**, 8
- I**
- Integration
 - stückweise –, 28
 - Intensität, 8
 - Invarianz
 - Skalen-, 41
 - Invarianzebene, 67, 80
 - isotrope Faltungsmaske, 31
 - Iteration
 - gedämpfte, 161
 - Iterationsfunktion
 - Skalen-, 154
 - Tensor-, 154
- K**
- $\hat{k}^{(i)}(n, \sigma)$, 36
 - $\tilde{\hat{k}}^{(i)}(n, \sigma)$, 37
 - Kamera-Kalibrierung, 3
 - Kameraparameter, 1, 7
 - Kanal
 - Y-, 8
 - Kante
 - Grauwert-, 173
 - Objektbilds-, 173
 - Kirsch-Operator, 15
 - Kognitives Sichtsystem, 3
 - Konvergenz, 125
 - Konvergenzkriterium, 125, 149
 - Korrelationsverfahren, 10, 12
 - Kovarianz
 - Start-, 109
 - Kovarianzmatrix, 16, 18, 21, 81
- L**
- Laplace-Gleichung, 17
 - Laplace-Operator, 16, 26
 - Least Median of Squares, 20
 - Least Squares, 47
 - Total, 20
 - lineare Adaption, 114
 - lineare homogene Diffusionsgleichung, 16
 - linearer Diffusions-Skalenraum, 16

- linearer Grauwertübergang, 53
 - linearer Skalenoperator, 16
 - linearer Skalenraum, 16
 - LMedS, *siehe* Least Median of Squares
 - lokaler Tensor, 67
 - LS, *siehe* Least Squares
 - Luminanz, 8
- M**
- Marr-Hildreth-Koenderink-Witkin-Theorie, 16
 - Maske
 - Ableitungs-, 14
 - Glättungs-, 14
 - Tiefpass-, 14
 - Maskenausdehnung, 27
 - Fehler durch endliche –, 27
 - Maskenbereich, 27, 31, 81
 - Maskenbreite, 27, 53
 - Maskenfehler, 53, 71
 - erwarteter, 29
 - tatsächlicher, 29
 - theoretischer, 29
 - Maskenfunktion, 15
 - Maskengröße, 15, 27
 - Beschränkung der –, 26
 - Wahl der –, 27
 - Maskenkoeffizienten
 - abgetastet, 36
 - Bestimmung, 35
 - integriert, 37
 - Korrektur, 37, 81
 - vorläufige Bestimmung, 36
 - Mehr-Skalen-Ansatz, 15
 - mehrfach-adaptives Verfahren, 113
 - Minimierungsproblem, 20
- N**
- N, 23
 - Nabla-Operator, 12, 67
 - National Television System Committee, 8
 - Natürlich-Sprachliche-Ebene, 2
 - NF, *siehe* Normal-Fluss
- NF-Vektor, 69
 - nicht-linearer Skalenoperator, 16
 - nicht-linearer Skalenraum, 16
 - Norm
 - Euclid-, 149
 - Frobenius-, 150
 - Normal-Fluss, 69
 - NF-Vektor, 69
 - NSE, *siehe* Natürlich-Sprachliche-Ebene
 - NTSC, *siehe* National Television System Committee
 - NTSC-Standard, 8, 50
 - Nulldurchgang, 29
 - Nullstelle, 29
- O**
- Objekt, 1
 - Objektabbild, 1
 - Objektabbildskante, 173
 - OF, *siehe* Optischer Fluss
 - OFCE, *siehe* OF-Beschränkungsgleichung
 - OFD, 24, 173
 - Operator
 - Ableitungs-, 15
 - Faltungs-, 15, 36
 - Glättungs-, 15
 - Kirsch-, 15
 - Laplace-, 16, 26
 - Nabla-, **12**, 67
 - Roberts-, 15
 - Skalen-, 16, 41
 - Sobel-, 15
 - Optischer Fluss, 5, 6, **9**
 - OF-Beschränkungsgleichung, 12
 - OF-Bestimmung, 6
 - OF-Feld, 9
 - OF-Schätz-Zyklus, 132
 - OF-Vektor, 9
 - erweitert, 12
 - Orts-Frequenzraum, 14
 - Ortsraum, 14
 - Orts-/Zeitraum, 14
 - Oszillation, 131

- P**
 PAL-Standard, 50, 216
 Parameter
 Skalierungs-, 15
 partial differential equation, *siehe* PDE
 PDE, 16
 Pseudo-Inverse, 47
- Q**
 Quantisierung, 9
 quasi-inverse Adaption, 116
- R**
 Randkorrektur, 38
 Gradienten-orientierte, 45
 Rauschen, 4, 7
 Referenzfehler, 122
 Regression, 46
 RGB, 8
 Richtungsschiefe, 167
 Richtungsüberlappungsfaktor, 92
 Roberts-Operator, 15
 Robustheit, 139
 ROF, 24
 Rückkopplung, 1
- S**
 Σ , 31
 Σ_A , 18, 21, 81, 105
 Σ_M , 20, 21, 105
 σ_{\max} , 114
 σ_{\min} , 114
 SAE, *siehe* Sensor-Aktuator-Ebene
 SBE, *siehe* Szenen-Bereichs-Ebene
 Schätzfunktion, 153
 Schatten, 11
 Schichtenmodell, 1
 Schiefe
 Gesamt-, 167
 Richtungs-, 167
 Segmentierung, 2
 Sensor-Aktuator-Ebene, 1
 Sinus-Funktion, 53
 Skala, 5, 6, **15**, 27, 74, 85, 98, 105, 155
- Skalen-Iterationsfunktion, 154
 Skaleninvarianz, 41
 Skalenoperator, 16, 41
 linearer –, 16
 nicht-linearer –, 16
 Skalenraum, 15
 Gauß'scher –, 16
 linearer –, 16
 linearer Diffusions-, 16
 nicht-linearer –, 16
 Skalierung, 38
 Gradienten-orientierte, 43
 modifizierte σ -basierte, 40
 nach Korn, 40
 σ -basierte, 39
 Skalierungsbereich, 114, 153, 157
 Skalierungsparameter, 15
 Sobel-Operator, 15
 Spur, 23
 Standard
 NTSC-, 8, 50
 PAL-, 50
 Standardabweichung, 28
 Start-Kovarianz, 109
 stückweise Integration, 28
 Stufenkante, 52
 systematische Unterschätzung, 26
 Szene, 1, **6**
 Szenen-Bereichs-Ebene, 2
- T**
 T_s , *siehe* Skalenoperator
 Tensor
 Grauwert-Struktur-, 6, **21**, 68
 lokaler –, 67
 Tensor-Iterationsfunktion, 154
 Theorem von Bauer-Fike, 154
 Tiefpassfilter, 14
 Tiefpassfilterung, 2
 Tiefpassmaske, 14
 Randkorrektur, 38
 Skalierung, 38
 TLS, *siehe* Total Least Squares

- Total Least Squares, 20
- Trajektorie, 3
- U**
 - $\mathcal{U}_A(\mathbf{x}_0)$, 62, 153
 - $\mathcal{U}_{\text{GST}}(\mathbf{x}_0)$, 153
 - $\mathcal{U}_M(\mathbf{x}_0)$, 13, 20, 61, 153
 - Überlappungsfaktor, 90
 - bei Ableitungsmasken, 92
 - bei Mittelungsmasken, 103
 - Gesamt-, 92
 - Richtungs-, 92
 - ungekoppelte quasi-inverse Adaption, 166
 - Unterschätzung
 - systematische –, 26
- V**
 - Variationsansatz, 17
 - Verfahren
 - doppelt-adaptives –, 112
 - einfach-adaptives –, 110
 - mehrfach-adaptives –, 113
 - Verhaltens-Repräsentations-Ebene, 2
 - Verschiebungsrage, **9**, 173
 - VR-Vektor, **9**, 80
 - erweiterter, 68
 - VR, *siehe* Verschiebungsrage
 - VRE, *siehe* Verhaltens-Repräsentations-Ebene
- W**
 - Wärmeleitungsgleichung, 16
- Y**
 - Y-Kanal, 8
 - YIQ-Farbsystem, 8
- Z**
 - $z^{(i)}(\sigma)$, 29
 - Zeilenversatz, 9, 50
 - Zeitschritt, 9
 - Zufallsvariable, 60
 - Zufallszahlen, 66