

Untersuchung der Cepstrumtechnik
zur Querdisparitätsbestimmung
für die Tiefenschätzung
bei fixierenden Stereokonfigurationen

Kai-Oliver Ludwig

FBI-HH-B-153/92

März 1992

Universität Hamburg
Fachbereich Informatik
Bodenstedtstraße 16
D-2000 Hamburg 50
email: ludwig@kogs26.informatik.uni-hamburg.de

Kurzfassung

Bei der Abbildung unserer dreidimensionalen Welt auf ein zweidimensionales Bild geht die Tiefeninformation verloren. Der Mensch ist trotzdem in der Lage seine Umgebung dreidimensional wahrzunehmen.

Visuelle Informationen, die über das rechte bzw. linke Auge aufgenommen werden und nach mehreren Projektionsschritten zur primären Sehrinde gelangen, bilden dort in einer dünnen Schicht in Form alternierender Bänder mit rechter bzw. linker Augendominanz sogenannte okulare Dominanzstreifen. In jüngster Zeit wurde eine Reihe technischer Verfahren zur Disparitätsberechnung auf dieser Streifenarchitektur vorgestellt, von denen die Methode der Signalechobestimmung mit dem Cepstrumfilter [YS90] für diese Arbeit als Vorlage diente, weil sie sehr robust ist und das Korrespondenzproblem umgeht.

Nach einer Einführung über den Aufbau des Sehsystems und einer ausführlichen Darstellung der bisherigen Technik sowie der Untersuchung der geometrischen Abbildungsverhältnisse werden anhand eigener Untersuchungen einige Problemfälle des ursprünglichen Verfahrens identifiziert und Verbesserungen dazu vorgeschlagen (und implementiert). Es wird außerdem gezeigt, daß sich das Cepstrumfilter als Autokorrelationsoperation mit einer adaptiven, bandpaßähnlichen Vorfilterung verstehen läßt.

Bei gegebener Stereoanordnung läßt sich mit dieser Technik Disparität zuverlässig berechnen, die dann zu lokaler Tiefeninformation umgerechnet werden kann.

Abstract

Humans can perceive depth, despite the fact that 3D information is lost during perspective projection on a 2D surface.

Visual information is represented in the primate visual cortex in a peculiar structure of alternating bands of left and right eye dominance. Recently a number of algorithms have been proposed to compute local disparity using this architecture. Because of its simplicity and robustness I investigated the cepstral filtering method of Y. Yeshurun and Eric L. Schwartz [YS90] for fast disparity computation.

The biological background is comprehensively presented and the geometry of the stereo arrangement is investigated mathematically especially for fixation of simple surfaces (e.g. planes).

The current algorithm has been implemented and some special deficiencies are discussed. The robustness against image degradations like noise, rotation and scaling has been analyzed. Improvements to the present algorithm are presented. It is shown that the cepstral filter can be understood as the square autocorrelation of a bandpass filtered version of the original image.

The discussed framework is a reliable one-shot method for local depth estimation.

Gebrauchsnamen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen und dgl. in dieser Schrift berechtigt nicht zu der Annahme, daß solche Namen ohne weiteres von jedermann benutzt werden dürfen; oft handelt es sich um gesetzlich geschützte eingetragene Warenzeichen, auch wenn sie nicht alle explizit als solche gekennzeichnet sind.

Symbolische Mathematik.

Für einige der Berechnungen und Umformungen habe ich dankbar die Leistungen der Pakete MATHEMATICA 1.0* (Sun 3 Version) und MAPLE V** (Sun 4 Version) zur symbolischen Mathematik benutzt, wobei sich MAPLE im allgemeinen – außer bei ungewöhnlichen Integralen – als schneller und nützlicher erwies. Insbesondere bei größeren Produkten von Summen, wie z.B. den Invarianten in Kapitel 5, wird das sture Ausmultiplizieren und Zusammenfassen doch erheblich erleichtert. Ich habe solche Stellen, die nicht sofort nachvollziehbar sind, mit dem Zeichen \mathcal{M} – stellvertretend für MAPLE oder MATH – gekennzeichnet.

Englische Fachbegriffe.

In der Informatik und speziell in der Bildverarbeitung überwiegt die englischsprachige Fachliteratur. Einige Autoren entscheiden sich für eine Übersetzung ins Deutsche. Bekannte und gebräuchliche englische Fachausdrücke sind in diesem Text *kursiv* gesetzt.

*MATHEMATICA is a registered trademark of Wolfram Research Inc., Champaign, Illinois 61826-6059, USA.

**MAPLE is a registered trademark of Waterloo Maple Software, 160 Columbia Street West, Waterloo, Ontario, Canada N2L 3L3.

Inhaltsverzeichnis

1	Einleitung	1
2	Biologischer Hintergrund	5
2.1	Das menschliche Sehsystem	5
2.1.1	Die Retina	6
2.1.2	Die Kniekörper	8
2.1.3	Die Sehrinde	8
2.2	Die visuellen Verarbeitungskanäle	12
3	Die Cepstrumtechnik	15
3.1	Das Cepstrum	15
3.2	Cepstrum eines Doppelsignals	16
3.3	Verschiebungsbestimmung	19
3.3.1	Die Grundidee	19
3.3.2	Stand der Kunst	20
4	Das Modell zur lokalen Tiefenbestimmung	23
4.1	Modellbildung	23
4.1.1	Das Modell	24
4.1.1.1	Grundmodell	24
4.1.1.2	Biologisches Datenmaterial	26
4.1.1.3	Festlegung der Parameter	29
4.2	Modellgrenzen und Untersuchungsthemen	30
5	Untersuchungen zur Geometrie	31
5.1	Allgemeine Abbildungsgeometrie	31
5.1.1	Betrachtungen in der Ebene	31
5.1.1.1	Kreisförmige Abbildungsfläche	32
5.1.1.2	Ebene Abbildungsfläche	34
5.1.2	Betrachtungen im Raum	38
5.2	Abbildungsgeometrie bei lokaler Oberflächenapproximation	42
5.2.1	Approximation 1. Ordnung	43
5.2.2	Approximation 2. Ordnung	47
5.3	Disparität und relative Tiefe	48
5.4	Zusammenfassung	49

6	Untersuchung des 2D Cepstrums	51
6.1	Implementation	51
6.1.1	Allgemeines	51
6.1.1.1	Schnelle Transformationen (FFT und FHT)	51
6.1.2	Echobestimmung mit dem Cepstrumfilter	53
6.2	Untersuchung des 2D Cepstrums	54
6.2.1	E1 : Eingangsbeispiel	54
6.2.2	E2 : Auflösung	57
6.2.3	E3 : <i>random dot stereogram</i>	58
6.2.4	E4 : Inversion und unterschiedliche Helligkeiten	58
6.2.5	E5 : Gerade Kantensegmente	61
6.2.6	E6 : Periodische Muster	64
6.2.7	E7 : Rotation (Einzelbeispiele)	64
6.2.8	E8 : Skalierung (Einzelbeispiele)	66
6.2.9	E9 : Evaluierung für Rotation und Skalierung	69
6.2.9.1	Wahl der Testsignale	69
6.2.9.2	Versuchsdurchführung	69
6.2.10	E10 : Rauschinsensivität des Cepstrums	74
6.2.11	Zusammenfassung des bisherigen Verfahrens	75
7	Verbesserungen und Erweiterungen des 2D Cepstrums	77
7.1	Suchbereich	77
7.2	Unterschiedliche Helligkeiten	78
7.3	Blasse Streifen	79
7.4	Trägerfunktionen zur Signalextraktion	80
7.4.1	Gemeinsame Trägerfunktion	81
7.4.2	Unterschiedliche Trägerfunktionen	83
7.5	Bandpaßfilterung	85
7.6	Das Cepstrum als Autokorrelation mit adaptiver Vorfilterung	86
7.7	Zusammenfassung	88
8	Lokale Tiefenkarten	91
8.1	Disparitätskarten für generierte Bilder	91
8.2	Disparitätskarten für Kamerabilder	94
8.2.1	Tiefenkarten	95
8.2.2	Bemerkungen zu den berechneten Tiefenkarten	97
8.3	Zusammenfassung	98
8.4	Rohbilder	98
9	Zusammenfassung und Ausblick	101
A	Bildquellennachweis	105
	Literaturverzeichnis	106

Verzeichnis der Abbildungen

1.1	fixierende Stereoanordnung	3
2.1	Retina und Augapfel	6
2.2	Das Sehsystem	7
2.3	Die Schichten der primären Sehrinde	9
2.4	Okulare Dominanzstreifen	11
3.1	Prinzip der Verschiebungsbestimmung	19
4.1	Lokale Modellbildung	24
4.2	Das technische Modell	25
4.3	Lokale Tiefe am Fixationspunkt	26
4.4	Verteilung der Eingangssignale im visuellen Cortex	28
5.1	Was ist Querdisparität?	32
5.2	Querdisparität und Tiefe	33
5.3	Querdisparität bei ebener Abbildungsfläche	34
5.4	Isodisparitätslinien	35
5.5	Stereoanordnung (Bestimmungsgleichungen in 3D)	38
5.6	Disparitätsflächen für $s_y = const.$	41
5.7	Oberflächenapproximation 1. Ordnung am Fixationspunkt	43
5.8	Transformation bei Approximation 1. Ordnung	44
5.9	Isodisparitätslinien bei verschiedenen Öffnungswinkeln	45
5.10	Isodisparitätslinien bei verschiedenen Normalenvektoren	46
5.11	Oberflächenapproximation 2. Ordnung am Fixationspunkt	47
5.12	Transformation bei Approximation 2. Ordnung	48
6.1	Cepstrum eines Beispielsignals	55
6.2	Cepstrum eines Beispielsignals (2)	56
6.3	Cepstrum in unterschiedlichen Auflösungsstufen	57
6.4	<i>random dot stereogram</i>	58
6.5	Cepstrum für ein Stereopaar mit Inversion	59
6.6	Cepstrum bei additiver Konstante	60
6.7	Cepstrum bei geraden Kantensegmenten	61
6.8	Gerade Kantensegmente als Teilsignale ($b = a$)	62

6.9	Gerade Kantensegmente als Teilsignale ($b \neq a$)	63
6.10	Periodische Muster	64
6.11	Cepstrum bei Rotation eines Teilsignals (1)	65
6.12	Cepstrum bei Rotation eines Teilsignals (2)	65
6.13	Cepstrum bei Rotation eines Teilsignals (3)	66
6.14	Cepstrum bei Skalierung eines Teilsignals (1)	67
6.15	Cepstrum bei Skalierung eines Teilsignals (2)	67
6.16	Cepstrum bei Skalierung eines Teilsignals (3)	68
6.17	Evaluierung der Rotations- und Skalierungsinvarianz (1)	70
6.18	Cepstrum bei Rotation und Skalierung	71
6.19	Evaluierung der Rotations- und Skalierungsinvarianz (2)	71
6.20	Evaluierung der Rotations- und Skalierungsinvarianz (3)	72
6.21	Evaluierung der Rotations- und Skalierungsinvarianz (4)	72
6.22	Evaluierung der Rotations- und Skalierungsinvarianz (5)	73
6.23	Cepstrum bei weißem Rauschen	74
7.1	Eingeschränkter Suchbereich	78
7.2	Funktionale Deutung der blassen Streifen	79
7.3	Cepstrum mit Trennbereich	80
7.4	Modellbildung mit gleichartigen Trägerfunktionen	82
7.5	Cepstrum mit gaußförmigen Trägerfunktionen	83
7.6	Modellbildung mit unterschiedlichen Trägerfunktionen	84
7.7	Cepstrum mit unterschiedlichen Trägerfunktionen	84
7.8	Die Wirkung einer Bandpaßfilterung auf das Filterergebnis	85
7.9	Adaptiver Vorfilter: vorgefiltertes Bild und Filterkern	87
7.10	Autokorrelation bei vorhergehender Bandpaßfilterung	88
8.1	Teilsignale aus einem künstlich generierten Bild	92
8.2	Disparitätskarte für einen generierten Bildausschnitt	93
8.3	Stereoanordnung für die Kameraaufnahmen	94
8.4	Disparitätskarte für das Bildpaar „Buchseite“	95
8.5	Tiefenkarte für das Bildpaar „Buchseite“	95
8.6	Disparitätskarte für das Bildpaar „Maisdose“	96
8.7	Tiefenkarte für das Bildpaar „Maisdose“	96
8.8	Cepstrum an einer Zylinderrundung	97
8.9	Kamerabildpaar „Buchseite“	99
8.10	Kamerabildpaar „Maisdose“	100

Kapitel 1

Einleitung

- *Artificial Intelligence is the study of how to make computers do things, at which, at the moment, people are better.*[Ric83]
- *Artificial Intelligence is the branch of computer science which is concerned with the automation of intelligent behavior.*[LS89]
- *Artificial Intelligence is the study of mental faculties through computational models.*[CM85]

Wir sehen an diesen drei Definitionsversuchen, daß das Forschungsgebiet Künstliche Intelligenz mit verschiedenen Schwerpunkten gesehen wird.

Ohne jetzt den Begriff Intelligenz definieren zu wollen – unter anderem, weil dies einigermaßen schwierig ist – sind wir doch der Meinung, daß das Verstehen natürlicher Sprache, die Fähigkeiten zu Lernen, unsere Welt bildlich wahrzunehmen oder auch Schach zu spielen, Leistungen sind, die beim Menschen Intelligenz erfordern. Die verschiedenartigsten Anstrengungen, *Maschinen* oder *Programme* mit *intelligenter* Leistungsfähigkeit, zu erdenken oder zu konstruieren, lassen sich unter dem Begriff Künstliche Intelligenz fassen.

Einordnung

In dieser Arbeit geht es um Untersuchungen innerhalb des Forschungsgebietes der Künstlichen Intelligenz, die dem Teilbereich Bildverstehen zugeordnet sind. *Bildverstehen* ist nach B. Neumann [Neu86] als die Rekonstruktion und Deutung einer Szene anhand von Bildern definiert. Rekonstruktion und Deutung meint hier das Aufstellen und Verwenden einer computerinternen, symbolischen und mehrere Abstraktionsebenen umfassenden Szenenbeschreibung.

Passive Systeme

Bis vor wenigen Jahren wurden technische bildverstehende Systeme meist nur als passive – also nicht direkt am Bildaufnahmeprozess beteiligte oder den Prozess steuernde – Systeme aufgefaßt. Die Aufgabe dieser Systeme war die möglichst exakte

Wiedergewinnung der dreidimensionalen Struktur (Geometrie) aber auch anderer Eigenschaften der Szene aus den Bilddaten. Außer in industriellen Anwendungen ist es meistens das Ziel, eine möglichst vollständige, allgemeine und vielseitig verwendbare interne Beschreibung aufzubauen, um damit die gestellten Aufgaben zu lösen.

Unterstellen wir einmal, daß auch bei Lebewesen das Erzeugen und Verwenden einer internen Repräsentation der äußeren Welt den beobachtbaren Leistungen zugrunde liegt, so sind auch Lebewesen nach der Definition bildverstehende Systeme.

Ein Blick in die Natur zeigt uns aber, daß diese Systeme keineswegs im obigen Sinne passiv sind. Am Beispiel einfacher Lebewesen (Stubenfliege) hat uns V. Braitenberg [Bra73] gezeigt, daß Navigation in komplexer, natürlicher Umgebung mit vergleichsweise kleinen (neuronalen) Schaltkreisen und ohne eine vollständige geometrische Szenenbeschreibung möglich ist. Die gefundenen internen Verschaltungen – auch in anderen untersuchten Tieren (siehe z.B. [Nob90]) – scheinen vielmehr eng spezialisierte und an der zu lösenden Aufgabe orientierte Repräsentationen zu sein, sofern es sich überhaupt um explizite Repräsentationen handelt und nicht nur um einfachere Regelkreise, d.h. direkte Umverschaltungen der sensorischen Eingangsdaten in motorische Ausgangssignale [Mö90].

Aktive Systeme

In den letzten fünf Jahren haben einige Forscher, zum Teil auch wegen der prinzipiellen Probleme (z.B. [AWB87]) des bisherigen, passiven Ansatzes des rechnergestützten Bildverstehens, den Versuch unternommen, Bildverstehen mit aktiven – also in den Bildaufnahme-prozeß eingreifenden – Systemen zu betreiben (siehe z.B. [AWB87, Alo90]). Bisher ist dieses Forschungsgebiet, *Active Vision* genannt, noch recht klein, aber es gibt bereits interessante Ergebnisse: Es läßt sich z.B. zeigen, daß Probleme, die für einen unbewegten Beobachter schlecht gestellt (*ill-posed*) sind, gut gestellt (*well-posed*) werden, wenn man einen bewegten Beobachter einführt [AWB87], oder sogar geschlossene Lösungen besitzen [Ban86], wenn der bewegte Beobachter zusätzlich noch einen Szenenpunkt fixiert.

Auch wir Menschen verwerten aus dem enormen visuellen Angebot aus der sichtbaren Welt zum großen Teil nur kleine, für unsere derzeitige Aufgabe relevante Ausschnitte, indem wir diese Teile fixieren. Beim Autofahren beispielsweise wechselt der Blick von der Instrumententafel zu relevanten Punkten der Straßenszene, an denen sich Verkehrszeichen oder dgl. befinden, ständig. Dabei wird nur jeweils ein kleiner Ausschnitt um den Fixationspunkt scharf und detailliert wahrgenommen. Innerhalb dieses lokalen Ausschnittes werden Szeneneigenschaften wie z.B. Farbe, Form (Kon-turen) oder Tiefe sehr viel genauer wahrgenommen als im peripheren Bereich des Blickfeldes.

Möglicherweise sind einige höhere Lebewesen genau deshalb mit einer beweglichen, fixierfähigen und auf foveale Datenaufnahme ausgerichteten Stereoanordnung ausgerüstet, weil sich dies als vorteilhaft für die zu erledigenden Aufgaben im Rahmen der Konkurrenz und Auslese unter den evolutionären Bedingungen erwiesen hat.

In der vorliegenden Arbeit wird ein technisches Verfahren vorgestellt, mit dem bei einer solchen fixierenden Stereoanordnung in einem lokalen Ausschnitt um den Fixationspunkt Tiefe bestimmt werden kann (lokale Tiefenbestimmung).

Dieses Verfahren kann vollständig losgelöst von der Thematik *active vision* verstanden und auch angewendet werden. Das ist hier aber nicht unsere Zielsetzung und das wird den Einsatzmöglichkeiten des Verfahrens auch nicht gerecht.

Wir wollen nicht durch ein eventuell sehr rechenintensives Verfahren für ein gegebenes Bildpaar eine möglichst gute 3D Rekonstruktion des abgebildeten Ausschnittes erhalten. Diese Vorgehensweise hat ja bereits eine reichhaltige Menge an aufwendigen und durch viele z.B. an die Punkte der zu rekonstruierenden Objektoberfläche geknüpfte mathematische Randbedingungen eingeschränkte Rekonstruktionsalgorithmen geschaffen und entspricht dem im vorigen Abschnitt skizzierten Paradigma. Unsere Zielsetzung ist es vielmehr, einen einfachen, schnellen und biologisch plausiblen Algorithmus anzugeben, der eine vergleichsweise eher skizzenhafte aber robuste Rekonstruktion liefert, so daß das Verfahren für die lokale Tiefenbestimmung im Laufe einer Bewegung eingesetzt werden kann.

Die Problemstellung in dieser Arbeit

Aufbauend auf dem von Yeshurun und Schwartz [YS89] vorgeschlagenen Algorithmus zur Disparitätsbestimmung wird ein robustes Verfahren zur lokalen Tiefenbestimmung vorgestellt.

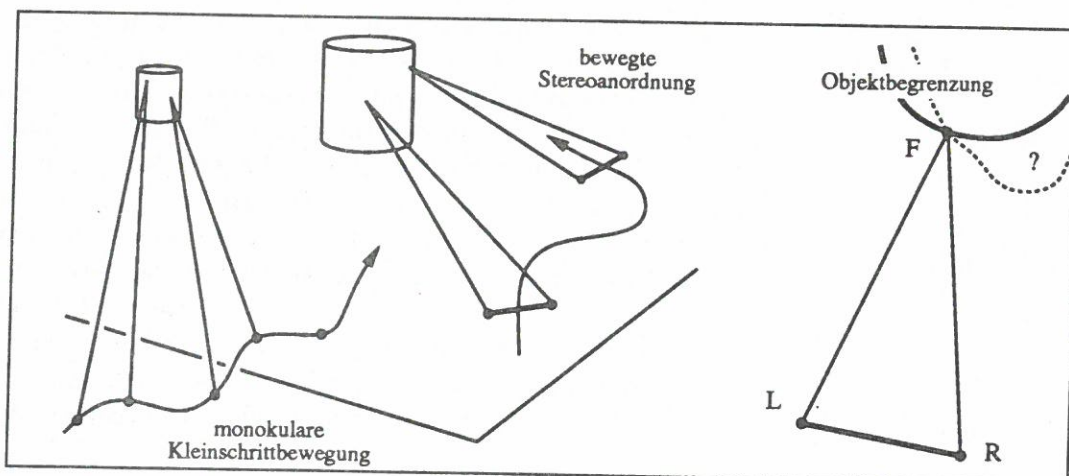


Abbildung 1.1: Links: Bewegungsverläufe zweier *Active Vision*-Systeme. Rechts: fixierende Stereoanordnung.

Der Ausgangspunkt ist dabei eine Stereoanordnung, bei der ein Objektpunkt einer Szene fixiert wird (siehe Abbildung 1.1). Wie schon erläutert, liegt der Schwerpunkt dabei nicht, wie in dem dargestellten passiven Ansatz, in einer möglichst exakten Rekonstruktion des abgebildeten Szenenteiles in allen Einzelheiten, sondern in einer schnellen und robusten Schätzung der Tiefenverhältnisse in der näheren Umgebung

des fixierten Punktes. Diese fixierende Anordnung sei im Laufe der Bewegung einer einzelnen Kamera oder eines Stereokamerakopfes entstanden.

Bei einer solchen Anordnung haben die auf der linken bzw. rechten „Netzhaut“ entstehenden Bildpunkte eines 3D Punktes aus der Umgebung des Fixationspunktes geringfügig voneinander abweichende retinale Koordinaten. Bereits Johannes Kepler (1571–1630) vermutete und Sir Charles Wheatstone (1802–1875) verifizierte 1838 im Versuch mit dem ersten Stereoskop, daß diese Verschiebung, Querdisparität genannt, – neben anderen Tiefenhinweisen, die wir hier nicht betrachten wollen – eine wichtige Information zur Tiefenwahrnehmung darstellt.

Ein aktives, bildverstehendes System benötigt natürlich sowohl eine Aufmerksamkeitssteuerung zur Auswahl geeigneter Fixationspunkte (im Beispiel des Autofahrers etwa für einen plötzlich auf die Straße rollenden Ball), als auch ein Steuerungssystem zur Fixation des ausgewählten Punktes. Diese beiden Teilprobleme werden hier nicht untersucht. Es gibt Vorschläge zur Wahl der Aufmerksamkeitssteuerung bei technischen Systemen [MvSG90, Ban86]. Untersuchungen zur Bewerkstelligung der Fixation mit zwei Kameras wurden an der Universität in Rochester (New York, USA) erfolgreich durchgeführt [OC90].

Kapitelübersicht

Die Architektur zur Verschiebungsbestimmung ist stark biologisch motiviert, und so wollen wir uns im nächsten Kapitel mit den notwendigen neurophysiologischen Grundlagen vertraut machen. Anschließend werden wir im dritten Kapitel die formale Grundlage einer Methode zur Disparitätsschätzung kennenlernen und eine kurze Übersicht über die bisherigen Anwendungen erhalten. Im vierten Kapitel werden die Inhalte aus den beiden vorhergehenden Kapiteln zusammengeführt und um weitere biologische Details ergänzt, um das Modell zur lokalen Tiefenbestimmung aufzustellen und zu erläutern. Die mathematischen Untersuchungen zur Geometrie der Stereoanordnung und dem Zusammenhang zwischen Disparität und Tiefe sind in Kapitel 5 zusammengefaßt. Die experimentellen Untersuchungen zur Cepstrumtechnik in Kapitel 6 und die Erweiterungen bzw. Verbesserungen des bisherigen Verfahrens in Kapitel 7 bilden den Schwerpunkt der Arbeit. Kapitel 8 illustriert die Anwendung der Cepstrumtechnik zur Bestimmung von lokalen Tiefenkarten für künstliche Bilder und Kameraaufnahmen. Abschließend gibt Kapitel 9 eine Zusammenfassung der Arbeit und einen Ausblick auf weitere sich anschließende oder vertiefende Forschungsthemen.

Kapitel 2

Biologischer Hintergrund

„I said in mine heart concerning the estate of the sons of men, that God might manifest them, and that they might see that they themselves are beasts. For that which befalleth the sons of men befalleth beasts; even one thing befalleth them: as the one dieth, so dieth the other; yea, they have all one breath; so that a man hath no preeminence above a beast ...“

King James Bible, Ecclesiastes 3:18-19

In diesem Kapitel wollen wir die biologischen Grundlagen für das im vierten Kapitel vorgestellte Verfahren kennenlernen. Da dieses Verfahren eine spezielle geometrische Anordnung der aufbereiteten visuellen Informationen innerhalb der Sehrinde ausnutzt, werden wir uns mit der Architektur des menschlichen Sehsystems beschäftigen und uns dabei insbesondere auf die spezifische Struktur einzelner, wichtiger Zwischenstationen des Informationspfades konzentrieren. Wir betrachten dazu kurz den Aufbau der Netzhaut und des Kniekörpers und ausführlicher den der primären Sehrinde. Wir benutzen dazu allgemeines Wissen aus Lexika und auch aus Arbeiten von Braitenberg, LeVay, Hubel und Wiesel.

Aufbauend darauf können wir dann die kurze Beschreibung der unterschiedlichen visuellen Informationskanäle beim Menschen verstehen. Wir interessieren uns dabei speziell für die verschiedenen Arten von Information, die die einzelnen Teilpfade von der Netzhaut bis zum visuellen Cortex transportieren. Wir greifen dazu auf Arbeiten von Livingstone zurück.

2.1 Das menschliche Sehsystem

Photonen, die von einer Szene in unser Auge einfallen, bilden in ihrer Gesamtheit unsere visuelle Information. Dieses Licht wird durch die Linse fokussiert und als scharfes Bild auf die lichtempfindliche, halbkugelförmige Auskleidung des Augapfels, die Netzhaut, projiziert. Die von der Netzhaut in elektrische Nervenimpulse umgewandelten Informationen werden über den Sehnerv weiter ins Gehirn geleitet.

2.1.1 Die Retina

Die Retina ist entwicklungsgeschichtlich als vorgeschobener Gehirnteil aufzufassen. Das bedeutet: hier werden nicht nur Photonen aufgenommen und in elektrische Impulse umgewandelt, sondern auch schon erste Verarbeitungsschritte durchgeführt. Insgesamt enthält die Netzhaut rund 130 Millionen lichtempfindliche Rezeptoren. Davon sind etwa nur 6 %, die sogenannten Zäpfchen (*cones*), farbempfindlich. Die anderen 94 %, die sogenannten Stäbchen (*rods*), sind zwar wesentlich lichtempfindlicher, können uns aber nur Grautöne vermitteln (vgl. Abbildung 2.1 links). Diese

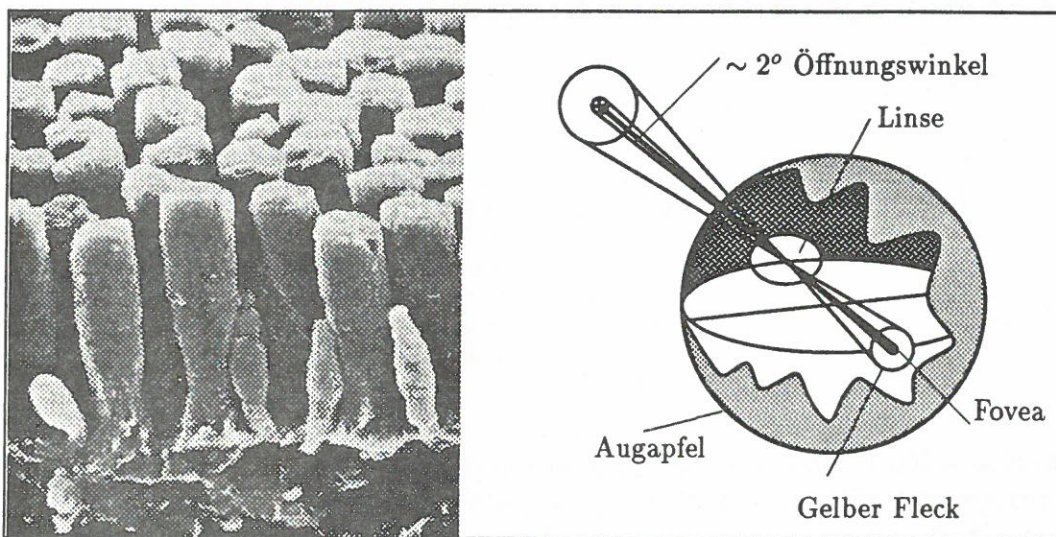


Abbildung 2.1: Links: Elektronenmikroskopische Aufnahme einer Retina. Die höheren, länglichen Rezeptorzellen sind Stäbchen, die kleineren, kegelförmigen nennt man Zapfen. Rechts: Schematische Zeichnung des Augapfels.

beiden Sorten von Zellen, deren Durchmesser etwa $2.5\mu\text{m}$ beträgt, sind innerhalb der Retina nicht gleichverteilt. Es gibt in der Nähe des Schnittpunkts der optischen Achse des Auges mit der Retina einen kleinen, gelben Fleck¹, in dem sich fast nur Zapfen befinden (siehe Abbildung 2.1 rechts). Wenn wir einen Punkt mit den Augen fixieren, liegt das Bild des fixierten Punktes im rechten und im linken Auge recht zentral in diesem Bereich, auf den sich ein Raumwinkel von rund 30° abbildet. Innerhalb dieses sog. Gelben Flecks befindet sich die Stelle schärfsten Sehens (Fovea). Die Auflösung in diesem Bereich beträgt etwa 0.5 Winkelminuten². Da die

¹Der Farbstoff, der die Rezeptoren gegen Blendung schützt und hohe Frequenzen filtert, ist gelb.

²Ich verwende in der ganzen Arbeit als Einheit für Winkel entweder Winkelsekunden ($''$), Winkelminuten ($'$) oder Grad ($^\circ$), je nach Größenordnung der betrachteten Werte. Eine Winkelminute ist $\frac{1}{60}^\circ$; eine Winkelsekunde $\frac{1}{3600}^\circ$. Der Faktor zwischen Strecke auf der Retina (in Millimetern) und Blickwinkel (in Minuten) ist ziemlich genau 200. So entspricht 1mm Retina 200 Minuten oder etwa 3.5° und ein einzelner Rezeptor mit etwa $2.5\mu\text{m} = 0.0025\text{mm}$ etwa 0.5 Minuten. Die hier genannten Werte sollen nur grobe Anhaltspunkte zur Vorstellung bieten. Eine detailliertere Diskussion wird in Kapitel 4 bei der Modellbildung vorgenommen.

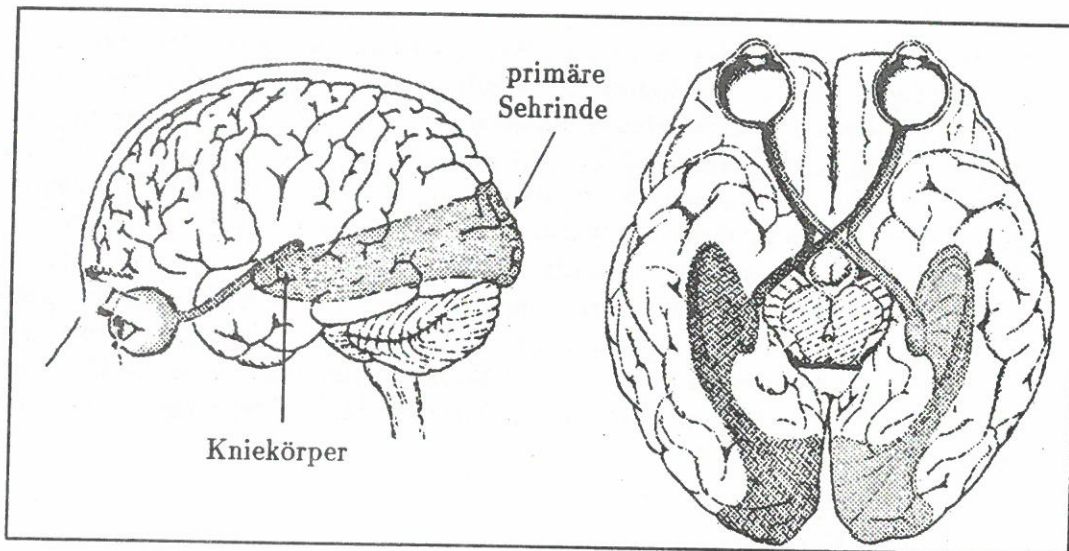


Abbildung 2.2: Das menschliche Sehsystem. Saggitalschnitt (links) und basale Ansicht (rechts).

innere Struktur der Netzhaut für uns nicht relevant ist, überspringen wir den weiteren Aufbau bis zu den Ganglienzellen am Ende der retinalen Verarbeitungskette. Eine einzelne Ganglienzelle bezieht ihre Informationen über mehrere Zellschichten hinweg von bis zu 500 Stäbchen oder Zäpfchen. Die Größe eines solchen kreisförmigen Einzugsbereiches, auch *rezeptives Feld* (*receptive field*) genannt, variiert von etwa $1'$ in der Fovea bis zu etwa 10° in der Peripherie.

Man kann zwei Typen von rezeptiven Feldern unterscheiden: *on-center-off-surround* und das Gegenstück *off-center-on-surround*. Wir wollen an dieser Stelle nur festhalten – ohne dies detaillierter zu begründen – daß mit dieser Verschaltungsweise näherungsweise eine diskrete Bandpaßfilterung auf den als diskrete Bildfunktion aufgefaßten retinalen Daten durchgeführt wird. Zum Vertiefen dieser Zusammenhänge sowie der aus den beiden folgenden Abschnitten sei das Buch von David Hubel [Hub88] empfohlen.

Die etwa 1 Million Ganglienzellen lassen sich in zwei unterschiedliche Typen einteilen: kleine und große Zellen. Die kleinen Zellen führen eine Art Subtraktion der Signale von verschiedenen Farbrezeptoren durch und können daher spezifische Farbinformationen weitergeben. Die großen Zellen dagegen addieren die Signale der Farbrezeptoren einfach auf und können, da sie keine Farbspezifität zeigen, als farbenblind angesehen werden [LH88, S.741]. Im nächsten Abschnitt knüpfen wir an diese Unterscheidung an. Nachdem nun die visuelle Information von etwa 130 Millionen Rezeptorsignalen auf etwa 1 Million Signale rund um den Faktor 100 komprimiert wurde, sammeln sich die Nervenstränge der Ganglienzellen an einer Stelle der Netzhaut – dem sogenannten Blinden Fleck, weil sich dort keine Rezeptoren befinden – und verlassen an diesem Ort gemeinsam den Augapfel gebündelt in einem dicken Strang, dem Sehnerv.

Bevor wir die Daten weiter auf ihrem Weg ins Gehirn verfolgen, ist noch eine für uns wichtige Tatsache festzuhalten: Die lokale, retinale Topologie der Informationen bleibt im Sehnerv und auch auf dem ganzen weiteren Weg bis zur Sehrinde im Prinzip erhalten. Oder anders ausgedrückt: Auf der Netzhaut benachbarte Daten bleiben benachbart, d.h. wir können uns vorstellen, daß das auf der Netzhaut entworfene Bild des Gesehenen zunächst weitgehend unverzerrt weitertransportiert wird.³

Bisher haben wir nur ein Auge betrachtet. Die Sehnerven beider Augen kreuzen sich noch vor den seitlichen Kniekörpern im sogenannten Chiasmus⁴ und verspleißen ihre Nervenfasern in einer speziellen Art und Weise: Der linke Kniekörper erhält alle Daten aus dem linken visuellen Feld beider Augen und der rechte Kniekörper entsprechend die Stränge aus dem rechten visuellen Feld beider Augen (siehe Abbildung 2.2).

2.1.2 Die Kniekörper

Die Kniekörper (*geniculate body, geniculate*) sind zwei etwa kirschkernegroße Ansammlungen von Nervenzellen tief im Gehirn.

Ein Kniekörper besteht aus sechs Schichten, deren Neuronen sich in der Art der Informationsverarbeitung und in der Größe unterscheiden. Die beiden inneren, unteren Schichten (*ventral layers*), die nur Informationen von den großen Ganglienzellen bekommen, bestehen aus größeren Zellen als die vier anderen oberen Schichten (*dorsal layers*), die nur von den kleinen Ganglienzellen gespeist werden. Dem entsprechend nennt man die unteren Schichten auch magno-zellulären Bereich, die oberen parvo-zellulären Bereich.

Die sechs Schichten enthalten abwechselnd Informationen aus dem linken und dem rechten Auge, mit einer seltsamen Vertauschung der Reihenfolge auf der Grenze zwischen magno- und parvo-zellulärem Bereich. Ein Kniekörper enthält also quasi sechs Bilder, sechs Karten des Gesehenen, drei aus dem linken und drei aus dem rechten Gesichtsfeld, und alle in präziser Übereinstimmung ihrer retinalen Koordinaten [Hub88, S.67].

2.1.3 Die Sehrinde

F. Gennari berichtet in seiner Schrift „De Peculiari Structura Cerebri“ von 1782 als erster von einem gestreiften Areal, das mehr als ein Jahrhundert später von S.E. Henschen (1847-1930) als Ort visueller Informationsverarbeitung identifiziert wurde. Der Streifenkortex (primäre Sehrinde, Area 17, V1, *primary visual cortex, striate cortex*)⁵ ist ein vielfach gefaltetes, ausgebreitet etwa $9 \times 5\text{cm}$ großes, 3mm

³„Oddly, between the retina and geniculate, in the optic nerve, they [the fibres] become almost completely scrambled, but they sort out again as they reach the geniculate.“ [Hub88, S.61]

⁴Der griech. Buchstabe Chi besteht aus zwei sich kreuzenden Linien.

⁵Der Name Streifenkortex rührt von dem von Gennari entdeckten helleren Streifen innerhalb der grauen Substanz her. Dieser hauptsächlich aus myelinisierten Axonen bestehende Streifen ist in der Hirnrinde besonders gut innerhalb des Sehrindenfeldes auszumachen.

dickes Gebiet⁶ der Großhirnrinde, das nach der Nomenklatur von J.S. Lund [Lun73] noch einmal in sechs unterschiedliche Schichten aufgeteilt werden kann (vgl. Abbildung 2.3). Neben der primären Sehrinde gibt es noch weitere visuelle Felder V2 (Area 18) bis V6, denen zum Teil auch schon Aufgabenbereiche zugeordnet werden konnten. Die Informationen werden von den Kniekörpern jeweils über ein breites Band, die Sehstrahlen (*optic radiations*), zur Sehrinde transportiert. Die Schicht 4C bildet dabei eine Art Empfangsschicht. Die Informationen werden von hier auf die anderen Schichten und Gebiete innerhalb der Schichten verteilt. Hierbei ist zu bemerken, daß nicht nur Daten tiefer ins Gehirn transportiert werden: Die Schicht 6 projiziert auch zurück zu den Kniekörpern.

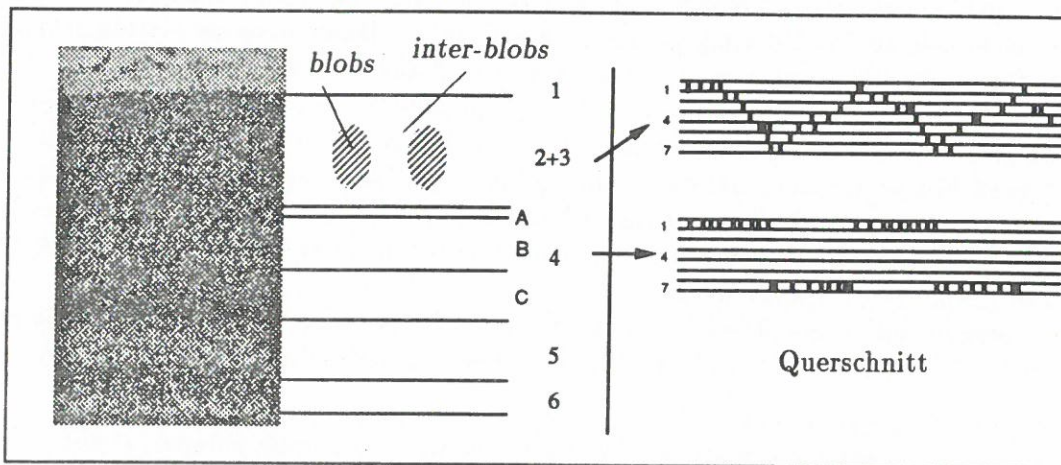


Abbildung 2.3: Links: Die Schichten der primären Sehrinde bei Rhesusaffen (*macaca mulatta*) nach der Notation von Lund (1973). Rechts: Darstellung der Augendominanz entlang zweier Querschnitte (Schicht 2+3 bzw. 4). In der Schicht 4 findet ein abrupter Wechsel zwischen kontralateralem (1) und ipsilateralem (7) Auge statt, während in Schicht 2 + 3 die Augendominanz in sanften Übergängen wechselt, so daß sich dort auch binokulare Zellen (4) finden.

Lokale Funktionseinheiten. Innerhalb der primären Sehrinde findet sich eine Vielzahl von Zellen, die auf unterschiedlichste Stimuli reagieren.

Es sind z.B. Zellen entdeckt worden, die auf einen statischen Reiz in Form einer Farbkontrastkante einer bestimmten Richtung reagieren. Für diese einfacheren Zellen hat man auch schon eine Einbettung in größere Strukturen entdeckt. So sind auf einem etwa 2mm^2 großen Abschnitt im Cortex in geordneter Weise⁷ Zellen für alle Kantenrichtungen enthalten.

⁶Faltet man die primäre Sehrinde tatsächlich aus, so nimmt sie in ihrer dreidimensionalen Gestalt, nicht die Form einer Ebene, sondern die der Oberfläche einer in Längsrichtung geteilten halben Birne an [Hub88, S.134]. Dabei wird der weitaus größte Teil des peripheren Gesichtsfeldes in der kleinen, unteren Halbkugel der Oberfläche der Birne repräsentiert, während die Fovea und alles andere innerhalb eines Öffnungswinkels von etwa 10° dem etwa doppelt so großen verbleibenden, oberen Teil entsprechen.

⁷Computereperimente zur Nachbildung und evtl. Erklärung selbstorganisierter Ordnung von orientierungssensitiven Zellen in Area 17 hat Chr. v. d. Malsburg [vdM73] durchgeführt.

In der zweiten und dritten Schicht findet sich eine andere Art von Untereinheit. 1978 entdeckte M. Wong-Riley etwa 0.2mm große, eiförmige Gebilde, die ihrer Form wegen *blobs* genannt werden (vgl. Abbildung 2.3). Im Gegensatz zu den diese *blobs* umgebenden Bereichen, den sog. *inter-blobs*, zeigten sich die *blobs* in späteren Untersuchungen hochselektiv für Farbe oder Helligkeit, aber völlig inaktiv bei Form- oder Bewegungsreizen. Abgesehen von solchen lokalen Funktionseinheiten ist der Aufbau der primären Sehrinde bemerkenswert homogen.

Okulare Dominanzstreifen. Wir interessieren uns speziell für die Schicht 4 der primären Sehrinde. Informationen aus dem linken bzw. rechten Auge wurden bisher getrennt transportiert. Während im Kniekörper „Karten“ vom linken bzw. rechten Gesichtsfeld noch vollständig getrennt waren, sind die Daten in dieser vierten Schicht auf spezielle Weise getrennt und doch vermischt⁸ angeordnet.

Die Daten aus dem linken bzw. rechten Auge sind abwechselnd in $0.3 - 0.45\text{mm}$ dünnen Streifen nebeneinander angeordnet [LHW75], wie etwa die ineinandergreifenden Finger betender Hände. Diese sog. okularen Dominanzstreifen (*ocular dominance columns*) – Streifen mit linker bzw. rechter Augendominanz – sind dabei in den Schichten 2 und 3 sowie 5 und 6 eher verwaschen in der Schicht 4 aber markant ausgeprägt (vgl. Abbildung 2.3).

Außerdem sind diese Streifen, zumindest in der Präparation von S. LeVay, D. Hubel und T.N. Wiesel [LHW75, S.562 ff], durch blasser Streifen (*pale bands*) getrennt:

„The pattern consisted of a system of dark-staining bands, each about $300\mu\text{m}$ wide, alternating with pale bands that were much thinner, about $50\mu\text{m}$ wide. The border between light and dark bands were not at all sharp, so that the apparent width of the light bands varied somewhat with different staining intensities and with different conditions of viewing the sections.“

sowie

„... varying degrees of shrinkage during the processing of the tissue ... so that the width of the columns in life would be about $300 - 450\mu$.“

Diese blassen Streifen liegen allerdings mit einer Breite von $25 - 50\mu\text{m}$ im Bereich der Auflösung der Elektroden:

„Finally, the physiology gives no suggestion of a no-man's-land between neighboring columns – the recordings indeed show an overlap of about $25 - 50\mu$, which may be real, but is probably produced by the use of electrodes whose uninsulated tip lengths are themselves $25 - 50\mu$.“

Die Meßwerkzeuge und die Meßergebnisse lassen meiner Meinung nach jedoch die Ansicht zu, daß die *pale bands* nicht nur in der speziellen Präparation, sondern auch real vorliegen könnten. Weiterhin schreiben S. LeVay, D. Hubel und T.N. Wiesel:

⁸Siehe zur geometrischen Struktur neuronaler Abbildungen auch den mathematischen Beitrag von H.A. Mallot, W. von Seelen und F. Giannakopoulos [MvSG90].

„We attach no functional significance to the fact that the column boundaries, as seen in these preparations, have an appreciable width.“

In Kapitel 7 wird gezeigt, wie das noch vorzustellende technische Verfahren durch Einführung solcher Trennstreifen bestimmte Problemfälle besser handhaben kann. Betrachten wir die okularen Dominanzstreifen nicht mehr im Querschnitt, sondern senkrecht dazu, mit Blick auf die Cortexoberfläche, so sehen wir eine Verteilung der Streifen wie in der linken Skizze von Abbildung 2.4. Fassen wir die Retina als Wertebereich und die Sehrinde als Bildbereich einer – neuronalen – Abbildung auf, so ist außer dem streifigen Bildbereich noch ein weiterer Punkt bemerkenswert: Die Abbildung hat einen ortsabhängigen Vergrößerungsfaktor. Im Bereich der Fovea wird 1° des Blickfeldes, also rund 0.3mm Gewebe der Retina auf etwa 6mm Gewebe des Cortex abgebildet. In der Peripherie bei 90° dagegen nur noch auf 0.15mm Cortexgewebe [Hub88].

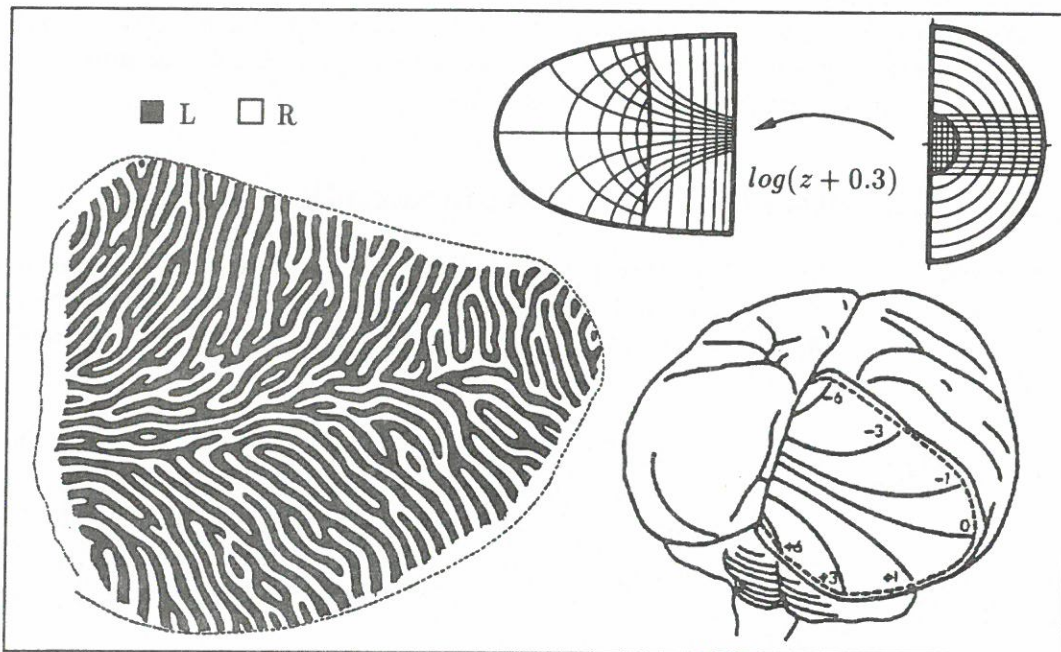


Abbildung 2.4: Rechts: Aufsicht auf die Cortexoberfläche Area 17 (V1) bei Rhesusaffen. Zum Vergleich: Skizze der Transformation eines Gitters unter $\log(z + 0.3)$ (Bezug siehe Text). Links: Verteilung der okularen Dominanzstreifen in diesem Gebiet (Schwarz: linkes Auge, Weiß: rechtes Auge).

Für diesen Vergrößerungsfaktor M , der von $0.5\text{mm}/^\circ$ bis $20\text{mm}/^\circ$ schwankt, haben P.M. Daniel und D. Whitteridge 1961 den Begriff *magnification*⁹ verwendet. Die obere Skizze der Abbildung 2.4 zeigt die genauere Geometrie¹⁰ der Verzerrung dieser Abbildung.

⁹Von Yeshurun und Schwartz mit *cortical magnification factor* bezeichnet.

¹⁰Als mathematisches Modell wird hier oft der komplexe Logarithmus in der Form $\log(z + a)$ mit variierendem a herangezogen. Die Wahl von $a = 1.7$ [HF77] bezeichnet Schwartz als inkorrekt und schlägt $a = 0.3$ vor [Sch86], welches besser mit den Experimenten von [Too82] übereinstimmt.

Im Licht dieser Fakten ist an der Verteilung der okularen Dominanzstreifen zweierlei bemerkenswert.

- Die Streifenbreite ist auffallend konstant über das gesamte Gebiet.
- Die Streifen verlaufen parallel zu Linien, die auf dem überwiegenden Teil der Retina Horizontalen, im fovealen Bereich aber Senkrechten, entsprechen.

Wir kommen darauf am Ende dieses Kapitels und in Kapitel 4 zurück.

Anknüpfend an die Vorstellung lokaler Funktionseinheiten von jeweils etwa 2mm^2 Größe in der gesamten primären Sehrinde stellen wir fest, daß eine solche lokale Verarbeitungseinheit Abschnitte von mindestens einem Paar okulärer Dominanzstreifen enthielte, und damit – wie wir noch sehen werden – genügend Informationen, um lokale retinale Verschiebung zu bestimmen, deren wichtige Rolle bei der Tiefenwahrnehmung die Einleitung bereits nannte. Im Gegensatz zu den vielen Untersuchungen über Orientierungskolumnen, sind mir bis jetzt allerdings noch keine neurophysiologischen Arbeiten über Disparitätskolumnen bekannt.

2.2 Die visuellen Verarbeitungskanäle

Mit den von M. Livingstone durchgeführten Untersuchungen wurde die Arbeitshypothese begründet, daß für die frühe visuelle Informationsverarbeitung getrennte Verarbeitungspfade existieren [Liv88, LH88]. Danach gibt es drei deutlich voneinander getrennte Kanäle.

- Der als parvo-interblob bezeichnete Kanal führt von den kleinen Ganglienzellen über die vier parvo-zellulären Schichten der Kniekörper, den Interblob-Bereich (V1) und die sog. blassen Streifen (V2) in noch unbekannte Bereiche des Gehirns. Dieser Kanal überträgt Informationen über Farbkontrastkonturen mit hoher Ortsauflösung. Er ist wahrscheinlich für die Fähigkeit von Bedeutung, ruhende Objekte genau zu erkennen.
- Eine sehr viel geringere Ortsauflösung hat der Kanal, der Informationen über Farbe und Grauschattierungen verarbeitet. Er benutzt Informationen aus allen sechs Schichten der Kniekörper und führt über die *blobs* in den Schichten 2 und 3 (V1) und die dünnen Streifen (V2) in das visuelle Feld V4, das für das Farbsehen eine wichtige Rolle spielt.
- Die Schiene von den großen Ganglienzellen über Schicht 4B und die dicken Streifen (Magno-Kanal) führt noch weiter in den mittleren Schläfen- oder Temporallappen (MT). Auf diesem Kanal werden Informationen über Bewegung und stereoskopische Tiefe übertragen. Dieser Kanal scheint farbenblind zu sein, denn im Gegensatz zum parvo-interblob Kanal, nimmt dieser Pfad keine Konturen wahr, die nur aufgrund eines Farbkontrastes zu erkennen sind.

Die Farbinsensitivität des Magno-Kanals legt nahe, daß zur Tiefenbestimmung keine Farbinformationen benötigt werden. Wir können uns daher zunächst bei unseren technischen Untersuchungen dahingehend unterstützt fühlen, daß die Beschränkung auf Grauwertbilder zur Tiefenbestimmung wahrscheinlich keine wesentliche Beschränkung darstellt. Die vergleichsweise grobe Ortsauflösung dieses Kanals paßt gut in das noch vorzustellende Modell.

Die biologischen Vorbetrachtungen sollen hiermit erst einmal beendet werden. Weitere Details werden erst bei der Modellbildung in Kapitel 4 wichtig werden.

Kapitel 3

Die Cepstrumtechnik

Betrachten wir zum Beispiel zwei lange, benachbarte Streifen einer Satellitenaufnahme der Erdoberfläche, die sich etwas überlappen, oder zwei Röntgenaufnahmen eines Organs zu verschiedenen Zeiten während einer medizinischen Untersuchung, aber unter ähnlichen Aufnahmebedingungen, oder – wie im Kontext dieser Arbeit – zwei Aufnahmen einer Szene von zwei leicht unterschiedlichen Standorten. Diesen Signalpaaren ist gemeinsam, daß die jeweiligen Einzelsignale in gewissen Teilen – idealisiert dargestellt – identische, aber verschobene Bildinformation enthalten.

In diesem Kapitel wollen wir die mathematische Grundlage für eine Technik kennenlernen, mit der die Verschiebung zwischen zwei solchen Signalen oder Teilsignalen bestimmt werden kann. Die Signale oder Teilsignale werden dazu in geeigneter Weise kombiniert und bilden dann das Eingangssignal für einen nicht-linearen Filter – das Cepstrum. Im Ausgangssignal befindet sich ein deutliches und daher leicht zu identifizierendes Maximum an einer Position, die der Verschiebung der Eingangssignale entspricht. Diese Vorgehensweise zur Verschiebungsbestimmung wollen wir Cepstrumtechnik nennen.

3.1 Das Cepstrum

Definition. Das Cepstrum einer Funktion $f(x)$ ist das Leistungsspektrum des Logarithmus des Leistungsspektrums des Signals:

$$\text{Cepstrum}\{f(x)\} = \|\mathcal{F}\{\log(\|\mathcal{F}\{f(x)\}\|^2)\}\|^2 \quad (3.1)$$

Hierbei und im folgenden sei mit

$$F(u, v) = \mathcal{F}\{f(x, y)\} = \frac{1}{2\pi} \cdot \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot e^{-i(ux+vy)} dx dy \quad (3.2)$$

die Fouriertransformierte von $f(x, y)$ bezeichnet.

Man nennt das Cepstrum auch *power cepstrum*, wenn es vom komplexen Cepstrum (*complex cepstrum*) – $\mathcal{F}\{\log(\mathcal{F}\{f(x)\})\}$ – unterschieden werden soll.

Historie. Das Cepstrum¹ wurde von Bogert et al. [BHT63] zur Bestimmung von Echos in eindimensionalen Signalen entwickelt. Das Einsatzgebiet reicht in der Folge von Anwendungen in der Seismologie [BHT63], der Sprachsignalverarbeitung [Nol64] bis hin zu zweidimensionalen Erweiterungen zur Lösung des Registrationsproblems in medizinischen Bildern [LMK89]. Zu dieser Anwendungsbreite hat unter anderem die Einfachheit und Robustheit des Cepstrums gegenüber traditionellen Techniken (wie z.B. Kreuzkorrelation) beigetragen (siehe [KC72, CSK77, OC90]). Einen zusammenfassenden Überblick über die Eigenschaften auch des komplexen Cepstrums geben D.G. Childers, D.P. Skinner und R.C. Kemerait [CSK77].

Ein Algorithmus zur Berechnung von Disparitäten mit dem zweidimensionalen Cepstrum wurde kürzlich von Y. Yeshurun und E.L. Schwartz [YS89] vorgestellt. Die Idee basiert auf der Wirkung des Cepstrums auf ein Doppelsignal.

3.2 Cepstrum eines Doppelsignals

Die folgende Rechnung findet sich kurz skizziert bei den Autoren – jedoch schon teilweise auf den biologischen Kontext bezogen, den wir erst im nächsten Kapitel behandeln wollen – und auch etwas ausführlicher bei D.Y. Lee, S. Mitra und T.F. Krile [LMK89]. Hier soll der mathematische Zusammenhang einmal etwas allgemeiner dargestellt werden.

Wir stellen uns dazu eine zweidimensionale, kontinuierliche Signalfunktion $s(x, y)$ vor. Wir überlagern dieser eine verschobene Kopie $s(x - x_0, y - y_0)$ von sich selbst und berechnen das Cepstrum des zusammengesetzten Signals:

$$f(x, y) = s(x, y) + s(x - x_0, y - y_0) \quad (3.3)$$

Fouriertransformation unter Berücksichtigung des Verschiebungssatzes (*shift theorem*) ergibt:

$$F(u, v) = S(u, v) + S(u, v) \cdot e^{-i(x_0 u + y_0 v)}$$

Das Leistungsspektrum lautet dann

$$\begin{aligned} \|F(u, v)\|^2 &= \|S(u, v)\|^2 \cdot \|(1 + e^{-i(x_0 u + y_0 v)})\|^2 \\ &= \|S(u, v)\|^2 \cdot 2 \cdot (1 + \cos(x_0 u + y_0 v)) \end{aligned} \quad (3.4)$$

¹Das Wort *cepstrum* ist ein Anagramm des Wortes *spectrum*. Auch die anderen Anagramme, wie *alanysis*, *quefreny*, *saphe*, usw. gehen auf Bogert et al. zurück. Bei der Beschreibung der Phänomene des Frequenzraumes des Frequenzraumes dienen sie der begrifflichen Trennung dieses Raumes vom Frequenzraum, unter Beibehaltung des gewohnten Argumentations- und Begriffskontextes des Frequenzraums (Frequenz, Phase, ...).

Nehmen wir jetzt, weiter Formel (3.1) folgend, den Logarithmus:

$$\log(\|F(u, v)\|^2) = \log(\|S(u, v)\|^2) + \log(2 \cdot (1 + \cos(x_0 u + y_0 v))) \quad (3.5)$$

Dieses Signal ist die Summe eines bildabhängigen Signalteils, über den wir im allgemeinen Fall keine besonderen Aussagen machen können und eines bildunabhängigen Signalteils, der im wesentlichen nur eine einzige, sehr dominante Frequenzkomponente $w_0 = x_0 u + y_0 v$ besitzt. Entsprechend erwarten wir, daß in dem Leistungsspektrum von $\log(\|F(u, v)\|^2)$ – d.h. dem Cepstrum von $f(x, y)$ – eine linienspektrumartige Komponente auf diesen Anteil hinweist.

Fouriertransformation liefert zunächst:

$$\mathcal{F}\{\log(\|F(u, v)\|^2)\} = \mathcal{F}\{\log(\|S(u, v)\|^2)\} + \mathcal{F}\{\log(2)\} + \mathcal{F}\{\log(1 + \cos(x_0 u + y_0 v))\} \quad (3.6)$$

Für $\|x\| < 1$ gilt die Reihenentwicklung

$$\log(1 + x) = \sum_{n=1}^{\infty} (-1)^{n+1} \cdot \frac{x^n}{n}$$

also kann (3.6) – mit $x \equiv \cos(x_0 u + y_0 v)$ – fast überall zu

$$\begin{aligned} \mathcal{F}\{\log(\|F(u, v)\|^2)\} &= \mathcal{F}\{\log(\|S(u, v)\|^2)\} + \mathcal{F}\{\log(2)\} \\ &\quad + \mathcal{F}\left\{\sum_{n=1}^{\infty} (-1)^{n+1} \cdot \frac{\cos^n(x_0 u + y_0 v)}{n}\right\} \\ &= \mathcal{F}\{\log(\|S(u, v)\|^2)\} + \delta(x, y) \\ &\quad + \sum_{n=1}^{\infty} (-1)^{n+1} \cdot \frac{\mathcal{F}\{\cos^n(x_0 u + y_0 v)\}}{n} \end{aligned} \quad (3.7)$$

umgeformt werden.

Da die Fouriertransformierte von $\cos(ax)$ nur aus einem Linienspektrum mit zwei Impulsen bei $u = -a$ und $u = a$ besteht, kann die Fouriertransformierte von $\cos^2(ax) = \cos(ax) \cdot \cos(ax)$ wegen des Faltungssatzes² für die Fouriertransformation leicht aus der von $\cos(ax)$ bestimmt werden. Entsprechend können auch die Fouriertransformierten aller höheren Potenzen $\cos^n(ax)$ gewonnen werden.

²Der Faltungssatz lautet: Die Fouriertransformierte eines Produktes ist gleich der Faltung der fouriertransformierten Faktoren:

$$\mathcal{F}\{f(x, y) \cdot g(x, y)\} = F(u, v) * G(u, v)$$

mit

$$F(u, v) * G(u, v) ::= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(s, t) \cdot G(u - s, v - t) ds dt$$

Damit wird (3.7) zu:

$$\mathcal{F}\{\log(\|F(u, v)\|^2)\} = \mathcal{F}\{\log(\|S(u, v)\|^2)\} + \sum_{n=-\infty}^{\infty} \frac{\delta(x - n \cdot x_0, y - n \cdot y_0)}{n}$$

Der erste Summand ist i.a. eine komplexe, der zweite eine reelle Funktion. Wegen

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2 + 2 \cdot f_{RG}R$$

für $f = f_R + if_I$ und $g = g_R$ gilt in unserem speziellen Fall,

$$\boxed{\text{Cepstrum}\{f(x, y)\} = \text{Cepstrum}\{s(x, y)\} + \sum_{n=-\infty}^{\infty} \frac{\delta(x - n \cdot x_0, y - n \cdot y_0)}{n}} \quad (3.8)$$

wobei die Deltaimpulse aus $\|g\|^2$ und aus $2 \cdot f_{RG}R$ zusammengefaßt wurden. Wir sehen also, daß die Koordinaten des stärksten Impulses der schnell in der Amplitude abfallenden Impulsfolge die Disparität der Eingangssignale kodieren. Verschattet der erste Term ($\text{Cepstrum}\{s(x, y)\}$) den zweiten Term nicht, so kann man dieses Maximum problemlos detektieren. Enthält das Cepstrum von $s(x, y)$ Komponenten in der Nähe von (x_0, y_0) , so müssen bereits im Spektrum Anteile entsprechender Frequenz vorliegen. Das wiederum bedeutet, daß $s(x, y)$ bereits Doppelstrukturen enthalten muß, wie sie normalerweise erst durch Bildung des Doppelsignals $f(x, y)$ entstanden wären. Dieser Fall tritt praktisch nicht auf. Leider können aber durch die konkrete Realisierung andere Spezialfälle entstehen, bei denen der stärkste Impuls der Impulsfolge dennoch überdeckt wird. Wir werden diese Fälle in Kapitel 6 bei der Untersuchung des diskreten 2D Cepstrum noch im einzelnen kennenlernen.

Phasenkorrelation. Den Zusammenhang der Signale aus Gleichung (3.3) können wir auch als Konvolution schreiben. Mit $f(x, y) = l(x, y) + r(x, y)$ gilt:

$$r(x, y) = l(x, y) * \delta(x - x_0, y - y_0) \Rightarrow R(u, v) = L(u, v) \cdot e^{-i(x-x_0, y-y_0)}$$

Auflösen nach der Exponentialfunktion, Erweitern mit $L^*(u, v)$ und Rücktransformation liefert die Grundformel für die bekannte Phasenkorrelationsmethode nach C.D. Kuglin und D.C. Hines [KH75],

$$\delta(x - x_0, y - y_0) = \mathcal{F}^{-1} \left\{ \frac{R(u, v)L^*(u, v)}{\|L(u, v)\|^2} \right\} = \mathcal{F}^{-1} \left\{ \frac{R(u, v)L^*(u, v)}{\|R(u, v)L^*(u, v)\|} \right\} \quad (3.9)$$

wobei in der letzten Umformung $\|RL^*\| = \|L\|^2$ wegen $\|L\|^2 = \|R\|^2$ und

$$\|RL^*\| = \sqrt{(RL^*)(RL^*)^*} = \sqrt{RL^*R^*L} = \|R\|\|L\| = \|L\|^2 \quad (3.10)$$

Autokorrelation. Den Zusammenhang von $Cepstrum\{f\}$ zur Autokorrelation von f kann man aus der folgenden Umformung ersehen:

$$\log(\|F\|^2) = \frac{\log(\|F\|^2)}{\|F\|^2} F F^* = \left\| \frac{\sqrt{\log(\|F\|^2)}}{\|F\|} F \right\|^2 =: \|H\|^2 \quad (3.11)$$

Der Logarithmus des Leistungsspektrums ist also das Leistungsspektrum einer gefilterten Version des Originalsignals. Wegen

$$\|\mathcal{F}\{\|H\|^2\}\|^2 = \|\mathcal{F}\{H^*H\}\|^2 = \|h^* * h\|^2 = \|h \circ h\|^2 \quad (3.12)$$

kann das Cepstrum als eine Art Autokorrelationsoperation mit einer adaptiven (nicht-linearen und bildabhängigen) Vorfilterung aufgefaßt werden (vgl. [OC90, S.28]). Wir werden in Kapitel 7 diesen Zusammenhang noch einmal aufgreifen.

3.3 Verschiebungsbestimmung

3.3.1 Die Grundidee

Die Formel (3.8) kann in der folgenden Weise ausgenutzt werden, um die unbekannte Disparität (x_0, y_0) zwischen zwei Signalen $s_1(x, y)$ und $s_2(x, y) = s_1(x - x_0, y - y_0)$ zu bestimmen.

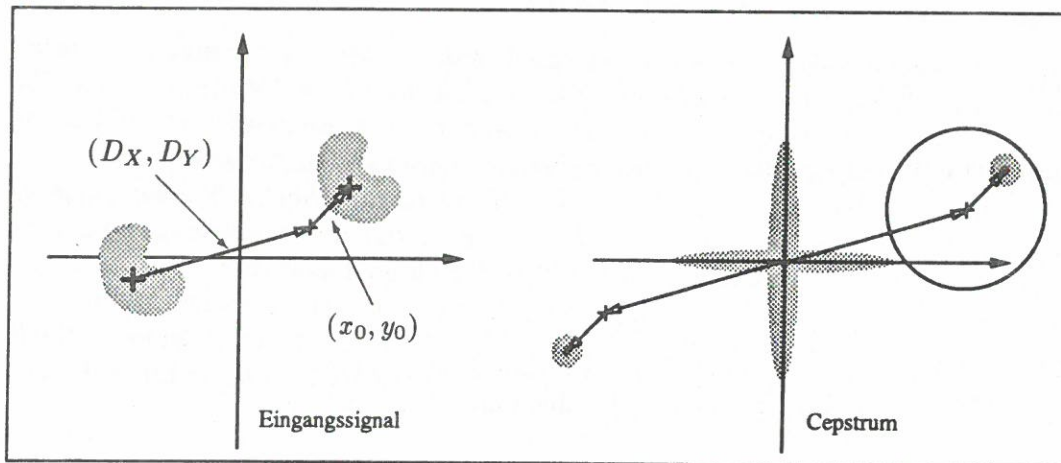


Abbildung 3.1: Das Prinzip der Verschiebungsbestimmung mit dem Cepstrumfilter. Schematische Skizze. Links: zusammengesetztes Eingangssignal. Rechts: gefiltertes Ausgangssignal (siehe Text).

Wir berechnen das Cepstrum von

$$\begin{aligned} f(x, y) &= s_1(x, y) + s_2(x - D_X, y - D_Y) \\ &= s_1(x, y) + s_1(x - (D_X + x_0), y - (D_Y + y_0)) \end{aligned} \quad (3.13)$$

Der zusätzliche Versatz (D_X, D_Y) ist ein uns bekannter Versatz.

Sind die beiden Teilsignale s_1 und s_2 nicht verschoben, also $(x_0, y_0) = (0, 0)$, so liegt nach Formel (3.8) ein Maximum bei (D_X, D_Y) vor. Wir können dieses Maximum durch Wahl von (D_X, D_Y) frei innerhalb der x - y -Ebene im Signal von $Cepstrum\{f(x, y)\}$ positionieren. Für Verschiebungen (x_0, y_0) , die klein gegenüber (D_X, D_Y) sind, gilt, daß sie sich in der Umgebung unseres ausgewählten „Referenzpunktes“ (D_X, D_Y) befinden. Besitzen die zu bestimmenden Disparitäten einen begrenzten Wertebereich und wählen wir (D_X, D_Y) entsprechend groß, so daß auch keine Überlagerung mit $Cepstrum\{s_1(x, y)\}$ stattfindet, können wir die unbekannte Verschiebung (x_0, y_0) durch Bestimmung des Maximums in einer lokalen Region um den „Referenzpunkt“ ermitteln.

3.3.2 Stand der Kunst

Neben den bereits genannten älteren und grundlegenden Arbeiten über das Cepstrum, gibt es die folgenden neueren Beiträge von Y. Yeshurun and E.L. Schwartz [YS89, YS90], D.J. Coombs, T.J. Olson and C.M. Brown [COB90, OC90] und D.J. Lee, S. Mitra and T.F. Krile [LMK89, LMK90], die spezielle Konkretisierungen der vorgestellten mathematischen Idee der Cepstrumtechnik enthalten. Ich will im folgenden den Inhalt der einzelnen Beiträge nur kurz skizzieren, um die Daten und Ergebnisse – soweit sie angegeben wurden – vorzustellen. Die genannten Ergebnisse werden in Kapitel 6 und 7 eigenen Ergebnissen gegenübergestellt.

Yehezkel Yeshurun and Eric L. Schwartz

Der Artikel von 1989 stellt einen biologisch motivierten Algorithmus zur Disparitätsbestimmung unter Verwendung des zweidimensionalen Cepstrums erstmalig vor. Der zweite Beitrag aus dem Buch „Computational Neuroscience“ ist bis auf kleinere Änderungen eine überarbeitete Version dieses ersten Artikels.

Die Autoren stellen den Algorithmus als ein Verfahren zur schnellen Stereosegmentation vor. Aus dem linken und rechten Bild werden dazu korrespondierende Quadrate von 32×32 Pixeln nebeneinander zu einem Doppelsignal angeordnet und dann mit dem Cepstrum gefiltert. Die Autoren geben einzelne Versuchsergebnisse an und machen Aussagen über das Verhalten bei künstlich herbeigeführten Störungen durch Glättung bzw. Histogrammtransformation eines Teilsignals, manuelle Eingriffe (*additional scribbling*), Rotation oder Skalierung.

„Size changes of up to 15 percent and rotations of ten degrees of one of the stereo frames can be routinely accepted by this algorithm. Considerable intensity changes can be applied to one of the stereo frames without disrupting the algorithm.“ [YS89, S.763]

Insgesamt stellen die Autoren die Methode unter Hinweis auf die prinzipielle Einschränkung durch die feste Ausschnittsgröße als sehr robust dar.

Sie verwenden fast durchgehend den Begriff *cepstral filtering*, den ich in der Form **Cepstrumfilterung** einfach übernehme, wenn ich auf den Filteraspekt des Verfahrens eingehen möchte. Dagegen werde ich an den Stellen, an denen die Autoren ein-

fach vom *algorithm* sprechen, den von mir eingeführten Begriff **Cepstrumtechnik** verwenden, um damit sowohl spezielle Algorithmen, als auch das generelle Prinzip der Wirkung des Cepstrums auf ein Doppelsignal zu bezeichnen.

Dah Jye Lee, Sunandra Mitra and Thomas F. Krile

Von diesen Autoren wird die Cepstrumtechnik sowohl zur Stereosegmentation in einem Motion-Stereo Algorithmus [LMK90] als auch zur Lösung des Registrationsproblems in medizinischen Bildern verwandt [LMK89].

Beim sog. **Registrationsproblem** soll die Verschiebung und Rotation eines neu aufgenommenen Bildes S gegenüber einem Archiv- oder Referenzbild R bestimmt werden. Die Autoren wählen zunächst korrespondierende Bereiche aus beiden Bildern aus, die markante und für die Registration wesentliche Strukturen enthalten. Dann werden beide Teilbilder auf vereinfachte Bilder (*simplified images*) reduziert, indem jeweils Hintergrundsubtraktion, Kantendetektion usw. durchgeführt wird. Aufgrund der Verschiebungsinvarianz des Amplitudenspektrums bestimmen die Autoren die Rotation über das Minimum der Maximumnorm der beiden Spektren unter dem Parameter Rotationswinkel; dieses Minimum wird dann angenommen, wenn die dominanten Merkmale im Bild parallel sind. Das Cepstrum wird dann auf das rotationskorrigierte Teilbild von S und das entsprechende Teilbild von R angewandt, um die verbleibende Translation zu ermitteln.

In eindimensionalen Einzelbeispielen wird die Überlegenheit des Cepstrums über die Phasenkorrelationsmethode vorgeführt und die beschriebene 2D Technik zur Untersuchung des Fortschreitens des „grünen Star“ angewandt. Zur Untersuchung der Stabilität wurden einzelne medizinische Bilder zusätzlich mit Gaußschem Rauschen gestört und so experimentell ein Signal-Rausch-Verhältnis³ (SNR) von 0.88 für die Cepstrumtechnik und ein SNR von 2 für die Phasenkorrelation festgestellt. Über die Größe der Bilder oder Ausschnitte werden leider keine Angaben gemacht.

In dem Beitrag zur **3D Rekonstruktion** bei parallelen Kameraachsen wird aus einer Bildsequenz von fünf, sieben oder neun Bildern das mittlere als Referenzbild ausgesucht. In einem hierarchischen Ansatz werden, von kurzer Stereobasis ausgehend (Verwendung der mittleren Bilder), grob Disparitäten ermittelt, die dann bei großer Stereobasis (Verwendung der äußeren Bilder der Sequenz) verfeinert werden können. Das Cepstrum wird hier verwendet, um mit Doppelsignalen aus Quadranten von 32×32 bis zu 8×8 Pixeln lokale Verschiebung in 256×256 Bildern zu messen.

David J. Coombs, Thomas J. Olson and Christopher M. Brown

Während bei den zuvor beschriebenen Arbeiten lokale Ausschnitte gefiltert wurden, verwenden diese Autoren ein vollständiges aber unterabgetastetes Bild. Das Ziel ist hier, die beiden unabhängig steuerbaren Kameras eines Stereokamerakopfes

³Man beachte, daß die Autoren ihr SNR als Quotienten der Varianzen von Signal und Störung definieren. Dies stimmt nur dann mit der üblichen Definition (vgl. z.B. [GW83]) überein, wenn Bildsignal und Störersignal beide verschwindenden Erwartungswert besitzen.

in eine fixierende Position zu bringen. Aufgrund der zunächst noch verschiedenen Blickrichtungen der Kameras enthalten sie unterschiedliche und – durch die Annahme partieller Überlappung der Blickkegel – auch gemeinsame Information. Aus den Einzelbildern (512×512) wird der 256×256 Ausschnitt aus der Mitte auf ein Quadrat mit 32 Pixeln reduziert.

Innerhalb von $51ms$ berechnet eine spezielle Hardware das Cepstrum aus dem Doppelsignal der beiden reduzierten Ausschnitte. Die ermittelte Disparität kann direkt in Motorsteuerungssignale umgesetzt werden. Es werden nicht direkt die Kameradaten gefiltert, sondern es wird noch ein *anti-aliasing filter* (Gaußfunktion mit $\sigma = 2.5$ Pixel) vorgeschaltet. Es wird versucht, das Maximum des Filterergebnisses mit Subpixelgenauigkeit zu bestimmen, indem nur noch die *scanline* mit dem Maximum weiterverwendet wird, um in dieser Zeile eine theoretisch und experimentell gestützte Modellfunktion anzupassen. Das ist notwendig, da die zentrale 256×256 Region recht grob unterabgetastet wurde.

Zusammenfassung und Vorausschau

Wir sehen, daß die Verschiebungsbestimmung mit der Cepstrumtechnik trotz der sehr unterschiedlichen Anwendungsdomänen bisher recht einheitlich folgendermaßen durchgeführt wird: Aus einem Bildpaar werden – mit unterschiedlichen Techniken – zwei korrespondierende quadratische Teilbilder entnommen. Diese werden entweder direkt oder nach einer Vorverarbeitung wie Glättung oder Extraktion markanter Strukturen zu einem Doppelsignal nebeneinander angeordnet. Dieses Doppelsignal entspricht – bei einer Kantenlänge D der Bildausschnitte – dem Spezialfall mit $(D_X, D_Y) = (D, 0)$ aus Abbildung 3.1.

In unserem Verfahren werden die Bilder eines Bildpaares mit jeweils 512×512 Pixeln zunächst in rechteckige Kacheln der Größe 32×64 Pixel unterteilt, damit wir dann durch die Zusammenstellung ein quadratisches Doppelsignal erhalten. Durch einfache Maximumsuche auf dem Filterergebnis erhalten wir 128 Disparitätswerte pro Bildpaar. Wir werden später sehen, daß diese Werte im wesentlichen schon die gesuchte Tiefeninformation darstellen. Als wesentliche Erweiterung gegenüber den bisherigen Anwendungen werden wir auch andere als rechteckige Bildausschnitte verwenden.

Im folgenden Kapitel werden wir das Modell zur lokalen Tiefenbestimmung genauer kennenlernen. Obwohl im strengen Sinne nicht von einer Modellbildung gesprochen werden kann, da die Funktionalität des Nachzubildenden noch nicht vollständig bekannt ist, habe ich für die Zusammenstellung und Auswertung des relevanten biologischen Datenmaterials und für die Wahl der Werkzeuge und Parameter zur Nachbildung eines Teils der Architektur der primären Sehrinde den Begriff Modellbildung verwendet, um den Vorgang der technischen Nachempfindung dieser biologischen Architektur zu bezeichnen. In diesem Sinne – als Subsumption neurophysiologischer Befunde in ein Modell mit Erklärungs- und Prädiktionscharakter – ist das technische Verfahren auch mit dem Begriff „technisches Modell“ bezeichnet worden.

Kapitel 4

Das Modell zur lokalen Tiefenbestimmung

In diesem Kapitel werden wir die biologischen Grundlagen und die mathematische Idee zur Verschiebungsbestimmung zu dem von Y. Yeshurun und E.L. Schwartz [YS89] vorgeschlagenen technischen Verfahren zusammenbringen. Hier wird allerdings nicht nur das knappe biologische Gerüst wiederholt, wie es die genannten Autoren vorstellen, sondern ich werde die gesamte verfügbare Palette biologisch relevanter – und zum Teil auch divergierender – Daten in einen gemeinsamen Zusammenhang stellen. Wir werden dann daraus die Größenordnung der Parameter des technischen Verfahrens ableiten.

4.1 Modellbildung

Wir wollen an die Idee lokaler Funktionseinheiten innerhalb der primären Sehrinde anknüpfen, wie sie in Kapitel 2 dargestellt wurde. Die tatsächliche Funktion der okularen Dominanzstreifen ist bislang noch ungeklärt. Ich habe dort angedeutet, daß eine lokale Funktionseinheit, deren Informationseinzugsbereich mindestens ein Paar okulärer Dominanzstreifen umfaßt, genügend Daten besitzt, um lokal Verschiebung ermitteln zu können.

Wie in der Einleitung bereits festgestellt, sind die Koordinaten eines 3D Punktes der Szene auf der linken bzw. rechten Netzhaut in der Umgebung des Fixationspunktes leicht verschieden. Da – wie wir in Kapitel 2 erfahren haben – die lokale Topologie auf dem ganzen Weg bis zur primären Sehrinde erhalten bleibt, entsprechen die Informationen benachbarter okulärer Dominanzstreifen immer noch im wesentlichen dem auf der jeweiligen Netzhaut entworfenen Bild; diese Teilbilder sind jedoch um kleine Beträge verschoben, da sie verschiedenen Augen entstammen.

Stellen wir uns nun eine hypothetische Funktionseinheit vor, die einen kurzen Abschnitt eines Paares benachbarter Streifen umfaßt, dann kann ihr Eingangssignal wie in Abbildung 4.1 dargestellt modelliert werden. Vernachlässigen wir jetzt nur vorläufig zur Motivation, da es sich um eine kleine, lokale Einheit handelt, per-

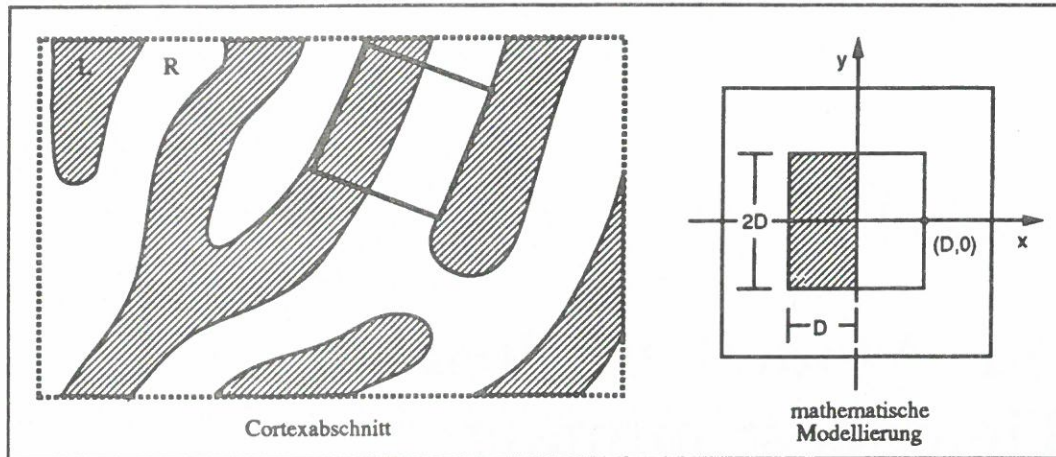


Abbildung 4.1: Lokale Modellbildung. Funktionseinheiten zur Disparitätsbestimmung.

spektivische Verzerrungen und andere Störeinflüsse, dann können linkes und rechtes Teilsignal als verschobene, aneinandergefügte Kopien eines Signals aufgefaßt werden. Anknüpfend an die Formelzeichen von Seite 19 bezeichnen wir die unbekannte Verschiebung der Teilsignale mit (x_0, y_0) , und der uns bekannte zusätzliche Versatz ergibt sich aufgrund der Streifenarchitektur zu $(D_X, D_Y) = (D, 0)$.

Mit Hilfe der Cepstrumtechnik kann die unbekannte Verschiebung geschätzt werden. Es soll damit keinesfalls gesagt werden, daß sich in der primären Sehrinde lokale Funktionseinheiten finden, die Cepstrumfilterung oder Fourieranalyse durchführen. Da aber mit einer solchen Architektur Disparitäten prinzipiell berechnet werden können, liefert dieses Verfahren eine klare algorithmische Begründung für die Existenz der Streifen in der primären Sehrinde [YS89].

4.1.1 Das Modell

Wir sehen uns jetzt kurz das technische Grundmodell an, diskutieren dann ausführlich das vorliegende biologische Datenmaterial und legen am Schluß damit die in der prinzipiellen Darstellung offengelassenen Parameter fest.

4.1.1.1 Grundmodell

Das Ausgangsdatum für eine Tiefenschätzung bildet ein Stereobildpaar einer fixierenden Stereoanordnung.¹ Die Einzelbilder sollen im wesentlichen foveale Daten enthalten, also mit einem geringen Aufnahmewinkel erzeugt werden. Wegen der Eigenschaften des Magno-Kanals (siehe Seite 12) benutze ich im folgenden nur Grauwertbilder. Denken wir uns die Einzelbilder jeweils in vertikale Streifen geschnitten, ordnen wir jeweils zwei korrespondierende Abschnitte davon (Rechtecke) in ein

¹Man kann Stereosegmentation mit parallelen optischen Achsen der einzelnen Kameras betreiben. Im folgenden meine ich mit Stereoanordnung immer eine fixierende Stereoanordnung.

Doppelsignal an und berechnen die Disparität, so erhalten wir insgesamt ein dünn besetztes Disparitätsfeld. Ist der Zusammenhang zwischen Disparität und Tiefe bekannt, können wir daraus ein grobes Tiefenfeld (Tiefenkarte) berechnen.

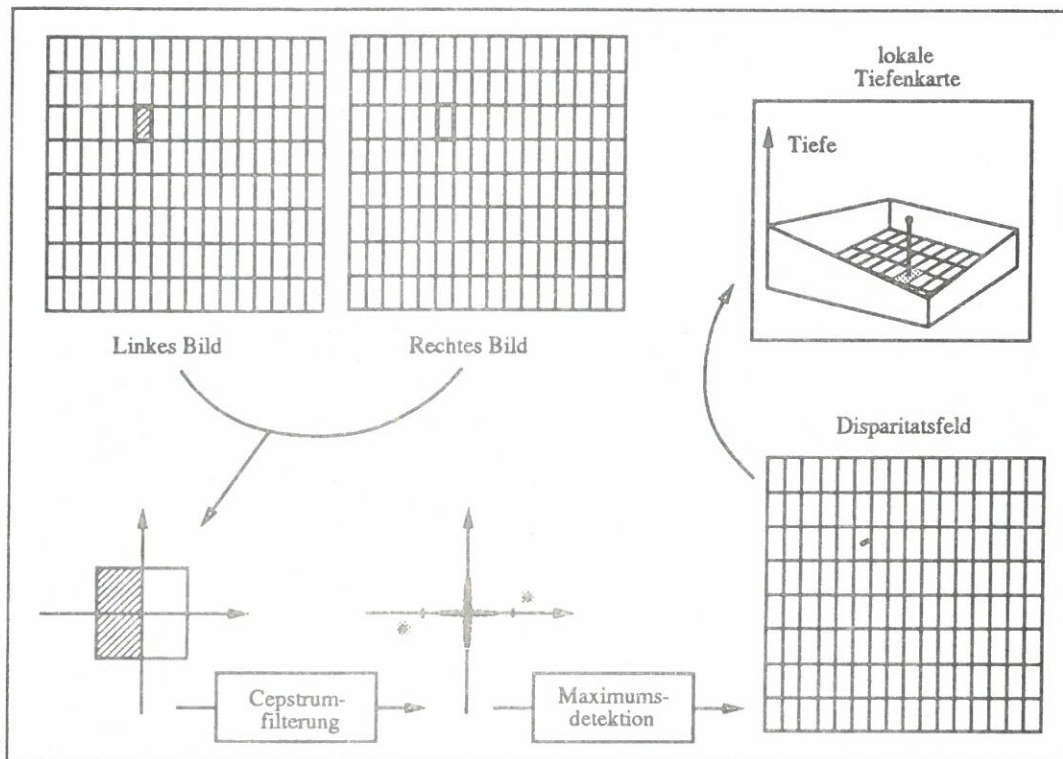


Abbildung 4.2: Das technische Grundmodell. Berechnungsschritte zur Bestimmung einer lokalen Tiefenkarte aus einem Stereobildpaar.

Die Abbildung 4.2 zeigt das generelle Vorgehen zur Tiefenbestimmung. Die Schritte über die Cepstrumfilterung und Maximumsuche bis zum Disparitätsfeld sind schon kurz beschrieben worden. Der Zusammenhang zwischen Querdisparität und Tiefe am Fixationspunkt wird in der folgenden Abbildung 4.3 (rechts) verdeutlicht.

Um diese Abbildung zu verstehen, stellen wir uns vor, daß die dreidimensionale Stereoanordnung auf eine zweidimensionale Anordnung durch Betrachtung eines horizontalen Querschnittes reduziert wurde, der den Fixationspunkt und die beiden Optischen Zentren enthält, wobei zur besseren Darstellung nur 4 statt der 16 vertikalen „Streifen“ aus Abbildung 4.2 verwendet wurden.

Aus Abbildung 4.3 (links) sehen wir zunächst, daß der (berechneten) Verschiebung zwischen den Teilsignalen zweier korrespondierender Bildausschnitte – in der Zeichnung für diesen zweidimensionalen Fall entsprechend als eindimensionales Binärmuster angedeutet – eine Tiefenstufe zugeordnet werden kann. Der linke Teil der Abbildung zeigt eine Ausschnittvergrößerung einer „Raute“, die sich aus einem Paar korrespondierender Bildausschnitte ergibt. Die rechte Seite der Abbildung zeigt, wie sich insgesamt innerhalb des abgebildeten Bereiches eine Objektkontur ergibt, wenn

ist eine Vertiefung im Zentrum dieses Gebietes, in der ausschließlich Zapfen liegen. Genauer unterscheidet man zwischen Fovea, Parafovea ($r = 1.25\text{mm}$) und Perifovea ($r = 2.75\text{mm}$) [ENC90, 27:S.176]. Gehen wir von einem Durchmesser von 0.5mm der Fovea aus [Hub88], so entspricht das – bei einer Brennweite von 17.3mm [Kor82, S.164] des menschlichen Auges – 1.7° oder etwa $100'$. Für die Parafovea ergibt sich der Durchmesser entsprechend zu $500'$.

Auflösungsvermögen und Rezeptorgröße. Man kann, je nach Verfahren, unterschiedliche Werte für das Auflösungsvermögen des Auges bestimmen. Um nur drei zu nennen: Punktauflösung (das Erkennen zweier Punkte als noch getrennt), Noniuauflösung (das Ablesen eines Versatzes zweier gegeneinander verschobener Linien) oder Frequenzauflösung (eine Rosette aus abwechselnd schwarzen und weißen Keilen, die zur Mitte hin dünner werden, verschmilzt an der jeweiligen Auflösungsgrenze zu einer grauen Scheibe).

V. Braitenberg hat einige unterschiedliche Werte für das Auflösungsvermögen in einer Tabelle zusammengefaßt. Die Werte variieren danach von $1.57'$ bis $0.1'$ [Bra85, S.384]. Im allgemeinen geht man von einem Auflösungsvermögen von $1'$ aus. Dies entspricht auch, bei einem Rezeptordurchmesser von $2.5\mu\text{m}$, der Breite zweier benachbarter Rezeptoren (Damit zwei helle Punkte noch als getrennt wahrgenommen werden können, muß ein Rezeptor dazwischen dunkel bleiben). Insgesamt ergibt sich unter Hinzunahme verschiedenster Quellen ein Durchmesser von $2\mu\text{m}$ [ENC90, 27:S.177ff] bis $5\mu\text{m}$ [Kor82, S.164]. Die Werte um $2-2.5\mu = 0.5'$ sind dabei häufiger, als die größeren Werte, so daß ich von den niedrigeren ausgehen werde.

Mit diesem Wert hat man aber das Problem, die gemessene Auflösung von $0.1'$ zu erklären. Durch Interpolation des für diese Auflösung zu groben Datenrasters an anderer Stelle kann man meiner Meinung nach dieses *hyperacuity*-Phänomen erklären. Die Funktion des Kniekörpers ist noch nicht ganz klar, und die Frage stellt sich, ob seine Aufgabe nicht eine solche Interpolation sein könnte [Bra85, S.385].

Von diesem Begriff der Auflösung ist die Ortsauflösung für die Tiefenwahrnehmung zu unterscheiden. Nach C.W. Tyler [Tyl75] können Unterschiede in der Querdisparität vom Menschen nur mit etwa $10'$ Abstand aufgelöst werden.

Vergrößerungsfaktor. Wie bereits festgestellt, wird die Retina innerhalb der Fovea auf ein etwa 20fach größeres Gebiet im Cortex abgebildet. Leider gibt es auch hier stark variierende Meßwerte. Verwenden wir nach V. Braitenberg die Notation des inversen Vergrößerungsfaktors M^{-1} , so variieren bei Affen die Werte zwischen $2'/\text{mm}$ und $20'/\text{mm}$ nur für den engen Bereich der Fovea.

Es ist vernünftig anzunehmen, daß die Verteilung der Fasern, die vom Kniekörper zum Cortex führen, auf der Cortexoberfläche gleichmäßig ist, wenn wir die Konstanz der Breite der okularen Dominanzstreifen beachten und lokale Funktionseinheiten unterstellen, die alle die gleiche Aufgabe erfüllen. Wir können dann über die Auflösung δ und den inversen Vergrößerungsfaktor M^{-1} via $d = \delta \cdot M$ den Abstand der Eingangsfasern d im fovealen Bereich und – wegen unserer Überlegung – auf dem gesamten Cortexgebiet bestimmen [Bra85] (siehe Abbildung 4.4).

V. Braitenberg berechnet daraus – unter Annahme einer bekannten Cortexoberfläche von 1320mm^2 und hexagonaler Packungsstruktur der Fasern auf dem Cortex – die Anzahl der Eingangsfasern N auf dem Cortex. Wir wollen seine Überlegungen hier im Einzelnen nicht wiederholen, halten aber fest, daß für ihn N um 10^6 liegen sollte; er verwendet für seine Betrachtungen einen Bereich von 400,000 bis 3,600,000.

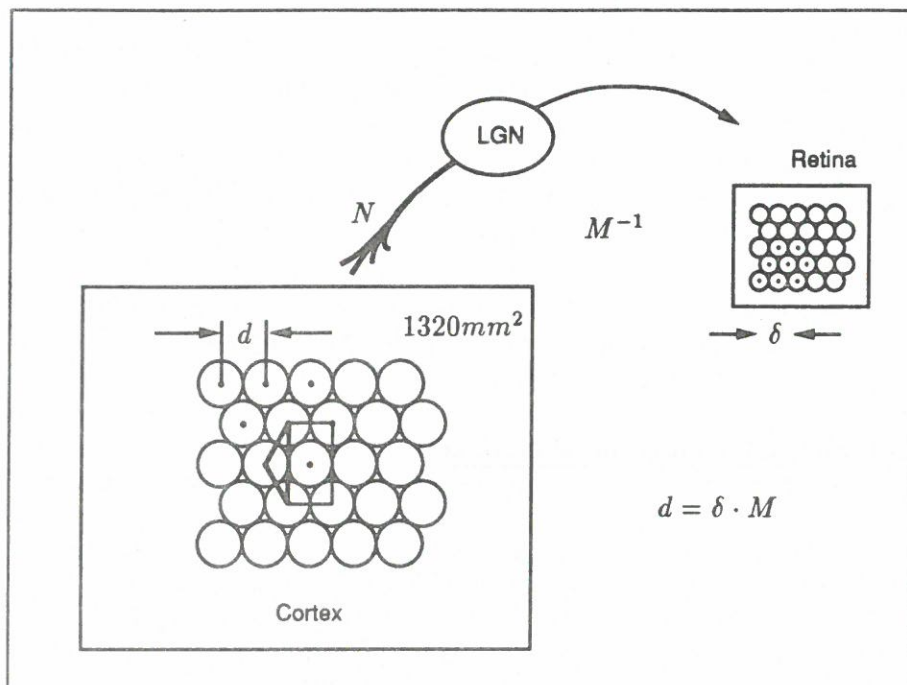


Abbildung 4.4: Verteilung der Eingangssignale auf den okularen Dominanzstreifen. Je nach Meßwerten ergeben sich – computertechnisch gesprochen – 6 bis 21 Eingangspixel orthogonal zur Richtung der Streifen. Die bei V. Braitenberg nicht angegebene Formel für $N(d)$ bei hexagonaler Packung kann man aus der Skizze als $N(d) = 1320/(\sqrt{12}d^2)$ ermitteln (Cosinussatz).

Panumbereich. Wir Menschen sehen nur innerhalb eines schmalen Tiefenbereiches um den Fixationspunkt einfach; davor und dahinter entstehen Doppelbilder. Obwohl auch Tiefenwahrnehmung bei Doppeltsehen (Diplopie) möglich ist, wollen wir uns hier nur auf diesen Fusionsbereich (Panumbereich) konzentrieren. Nach A. Korn [Kor82, S.163] werden im fovealen Bereich Punkte mit $\pm 3'$ Verschiebung noch fusioniert und einfach gesehen. Diese Beobachtung paßt auch gut zu der Einschränkung des technischen Verfahrens, bei gegebener Streifenbreite D nur Disparitäten mit einer Größe von $\pm \frac{D}{2}$ bestimmen zu können.²

²Spätere Untersuchungen haben gezeigt, daß diese Vorstellung eines Fusionsbereiches für absolute Disparitäten nicht voll zutreffend ist. Vielmehr spielt bei der Fusionfähigkeit zweier Punkte ihr Disparitätsgradient (*disparity gradient*) auch eine Rolle [BJ80]. Dennoch werden dadurch natürlich nicht die Messungen bezüglich eines absoluten Fusionsbereiches plötzlich falsch, sondern dieses neue Kriterium schränkt die Fusionsfähigkeit in besonderen Fällen lediglich noch weiter ein. S.B. Pollard et al. sprechen (vgl. [PMF85, S.450]) von einer wesentlichen Abweichung vom klassischen Konzept; obwohl der Disparitätsgradient die gemessenen absoluten Grenzen nicht erklärt.

Folgerungen. Die vorgenannten Daten und Zusammenhänge lassen sich nun insgesamt so auswerten:

- ① Aus dem Durchmesser der Fovea und der Annahme, daß der Effekt des Panumbereiches auf die Streifenbreite im Cortex zurückzuführen ist, ergibt sich wegen $\frac{100'}{6'} = 16\frac{2}{3}$ eine Anzahl von etwa 16 Streifen für die Fovea.
- ② Unter Annahme $M^{-1} = 5 - 10'/mm$ werden die 100' der Fovea auf 5 - 10mm Cortexoberfläche abgebildet. Das ergibt bei einer Breite D von 0.3 - 0.45mm der okularen Dominanzstreifen etwa 5 - 16 Streifen. (Man beachte die Doppelstruktur der Streifen im Cortex.)
- ③ Bei einer Cortexoberfläche von $1320mm^2$ und $N = 400,000 - 3,600,000.$, also der Annahme, daß der Kniekörper die Anzahl seiner Eingangsfasern nicht zu stark vergrößert oder verkleinert, erhalten wir $d = 0.21 - 0.62\mu m$; und mit $D = 0.3 - 0.45mm$ etwa 6 - 21 „Pixel“ pro Streifen.
- ④ In seiner Betrachtung [Bra85, S.383-386] argumentiert V. Braitenberg sehr überzeugend - innerhalb der von ihm betrachteten Größen - für die eher niedrigen Werte für M^{-1} . Verwendet man aber Werte von $M^{-1} = 2 - 5'/mm$, gelangt man bei der Breite der okularen Dominanzstreifen - die er selbst auch mit etwa 0.4mm angibt - jedoch zu Werten von 1 - 4 Streifen in der Fovea, was mir wenig erscheint; auch deshalb, weil dies etwa zu einer zweimal schlechteren Ortsauflösung als der von Tyler bestimmten führt. Deshalb habe ich zur Berechnung unter Punkt ② aus dem Bereich der gemessenen Vergrößerungsfaktoren ein mittleres Intervall verwendet.

4.1.1.3 Festlegung der Parameter

Obwohl damit die vorhandenen Daten keineswegs ausdiskutiert sind, wollen wir es hierbei belassen, da die Größenordnung der jeweiligen Werte zu erkennen ist.

Damit die Tiefenkarten einen aussagekräftigen Szenenausschnitt umfassen, darf dieser nicht zu klein sein; damit er noch zutreffend als „foveal“ bezeichnet werden kann, darf er aber auch nicht zu groß sein. Ich habe mich für einen Wert zwischen Fovea (100') und Parafovea (500') entschieden, der mit 200' deutlich dichter an der Fovea liegt. Damit wird ein Bereich von etwa 12cm in 2m Entfernung erfaßt.

Auf 200' entfallen 400 Rezeptoren, damit ist eine Pixelauflösung von 512×512 vertretbar. Wegen Punkt ② der Gesamtbetrachtung liegen in 200' 10 - 32 Streifen; 16 oder 32 sind also vertretbarere Werte. Wie man nachrechnen kann, wird keine der anderen Querbetrachtungen dadurch verletzt.

Der von mir gewählte Standardfall ist ein 512×512 Bild mit 16 Streifen (siehe Abbildung 4.2). Damit haben wir eine Ortsauflösung von 12.5' und eine Disparitätsauflösung von einer halben Minute.

4.2 Modellgrenzen und Untersuchungsthemen

Aus der vorgestellten Modellierung ergeben sich sofort drei Klassen von Fragestellungen:

- A** Inwieweit trifft die Annahme einer reinen Verschiebung der Teilsignale unter der perspektivischen Verzerrung tatsächlich zu?
- A1 Ist die Kachelung der Bildebene (siehe Abbildung 4.2) fein genug, so daß sich die Punkte innerhalb einer Kachel genügend gleichartig transformieren?
 - A2 Inwieweit ist dies von der lokalen Geometrie der Objektfläche am Fixationspunkt abhängig?
- B** Oder andersherum: Inwieweit kann die Cepstrumtechnik solche Verzerrungen robust verarbeiten (d.h. unter Verletzung der theoretischen Eingangsbedingung dennoch eine genügend genaue Verschiebung liefern) ?
- B1 Wie gut läßt sich das theoretisch vorhergesagte Maximum praktisch lokalisieren?
 - B2 Wie verhält sich die Lokalisationsleistung – immer noch eine reine Verschiebung unterstellt – bei Rauschen oder anderen Störungen, wie etwa unsymmetrischen Beleuchtungsverhältnissen?
 - B3 Wie werden perspektivische Verzerrungen toleriert? Bis zu welchem Grad sind Rotationen oder Maßstabsänderungen möglich?
 - B4 Was geschieht an Zylinderrundungen (Selbstverdeckung) oder anderen Verdeckungen, wo z.B. das rechte Auge mehr als das linke sieht?
- C** Wie gewinnen wir aus der berechneten Disparität die Tiefeninformation?
- C1 Was wollen wir unter Tiefe verstehen? Den Abstand von der linken Kamera, von der Rechten oder von der Mitte der Stereobasis? Gibt es noch eine andere Möglichkeit, Tiefe zu definieren?
 - C2 Gibt es außer dem Fixationspunkt noch andere Punkte, die Verschiebung Null besitzen, oder können wir – bei nicht ganz genau bekannten Parametern der Stereoanordnung – den Fixationspunkt daran identifizieren?

Zu den Problemstellungen unter B habe ich eine Reihe von Experimenten durchgeführt. In Kapitel 6 wird die Implementation der Cepstrumtechnik beschrieben. Dort werden die Untersuchungsergebnisse jeweils anhand von Einzelbeispielen vorgeführt und diskutiert. Um die Fragen unter A bzw. C zu beantworten, habe ich die Abbildungsgeometrie bei gegebener Stereoanordnung untersucht. Wir werden uns jetzt im nächsten Kapitel damit beschäftigen.

Kapitel 5

Untersuchungen zur Geometrie

„Bisher war ich bestrebt, für jedermann verständlich zu sein, aber von diesem Werke fürchte ich, daß es nur von solchen wird gelesen werden können, die sich das, was in den Büchern über Geometrie enthalten ist, angeeignet haben; denn da diese mehrere sehr gut bewiesene Wahrheiten enthalten, so schien es mir überflüssig, solche hier zu wiederholen, ich habe es aber darum nicht unterlassen, mich ihrer zu bedienen ...“

Nach René Descartes, La géométrie

In diesem Kapitel werden wir die mathematischen Zusammenhänge zwischen der Position eines 3D Punktes und der zugehörigen Querdisparität bei einer fixierenden Stereoanordnung genauer untersuchen. Es werden dazu die allgemeinen Formeln für runde und flache „Retina“ in der Ebene hergeleitet. Die Betrachtungen werden dann kurz auf den dreidimensionalen Fall ausgedehnt.

Auf den Zusammenhang zur Tiefe, wird an entsprechenden Stellen hingewiesen. Da ich in dieser Arbeit primär an relativer Tiefe interessiert bin, werde ich nur an einer Stelle beispielhaft durchrechnen, wie sich absolute Tiefeninformation gewinnen läßt. Wird die Objektoberfläche am Fixationspunkt durch einfache Funktionen approximiert, ergeben sich entsprechend einfache Formeln für die Disparität. Damit können wir uns z.B. ansehen, ob für alle Punkte innerhalb eines lokalen Rechteckes für die Verschiebungsbestimmung mit dem Cepstrum genügend ähnliche Disparitäten vorliegen.

5.1 Allgemeine Abbildungsgeometrie

5.1.1 Betrachtungen in der Ebene

Zu Beginn betrachten wir eine Ebene, in der die gesamte Stereoanordnung und auch der Objektbereich liegt. Linke bzw. rechte „Bildebene“ werden dann von Abschnitten auf Geraden oder Peripherien gebildet.

5.1.1.1 Kreisförmige Abbildungsfläche

Definition 1 (Querdisparität) Betrachten wir eine den Punkt F fixierende Stereoanordnung aus zwei idealen Lochkamas mit den Optischen Zentren L bzw. R . Für jeden weiteren Punkt P der Szene bezeichnen wir den von den Strecken \overline{PL} und \overline{PR} eingeschlossenen Winkel θ_P als absolute Disparität. Die Querdisparität oder relative Disparität $\delta_{P|F}$ (auch δ_P oder einfach δ) eines Punktes P bezüglich eines Fixationspunktes F ist definiert als die Differenz der absoluten Disparitäten:

$$\delta_{P|F} := \theta_P - \theta_F.$$

In der Literatur werden auch synonym die Begriffe Querdisparation, Disparität, retinal disparity, disparity verwendet.

Wir sehen rein qualitativ, daß für Punkte „hinter“ dem Fixationspunkt $\delta < 0$, für Punkte „vor“ dem Fixationspunkt $\delta > 0$ und für den Fixationspunkt selbst aufgrund der Definition trivialerweise $\delta = 0$ gilt.

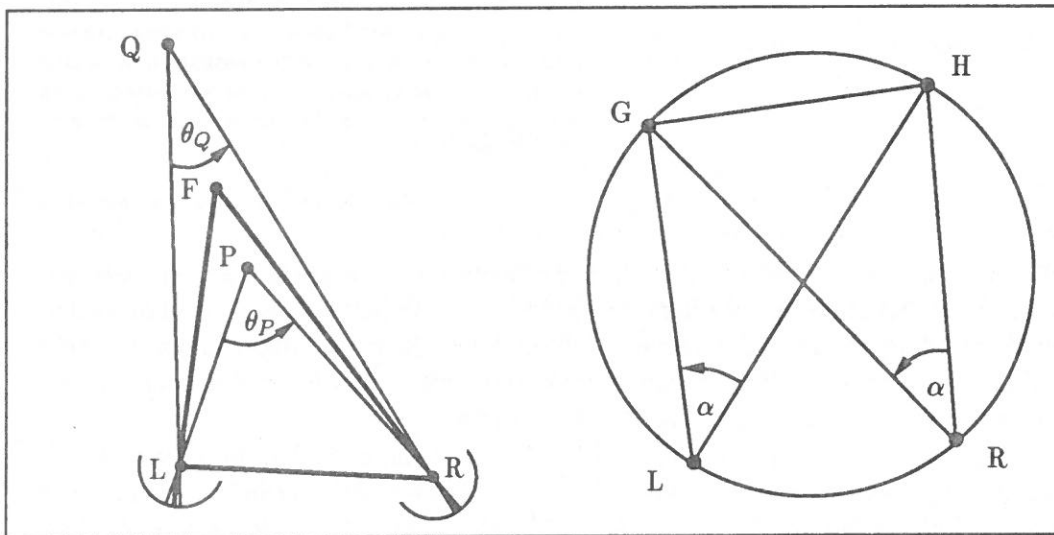


Abbildung 5.1: Links: Die Querdisparität bei runder Abbildungsfläche in der Ebene. Rechts: Zeichnung zu Satz 1.

Satz 1 Isodisparitätlinien sind Kreise.

Beweis. Eine Erweiterung des Satzes von Thales besagt, daß alle Peripheriewinkel über einer gemeinsamen Sehne gleich groß sind. Damit sehen wir (vgl. Abbildung 5.1, rechts), daß für je zwei Punkte G und H , die auf einem gemeinsamen Kreis durch die beiden Optischen Zentren liegen, $\delta_G = \delta_H$ gelten muß. Der Kreis für $\delta = 0$ wird Vieth-Müller-Kreis oder auch Horopter¹ genannt.

¹Beim Menschen ergibt sich durch Messung für $\delta = 0$ eine leicht vom idealen Kreis abweichende Kurve.

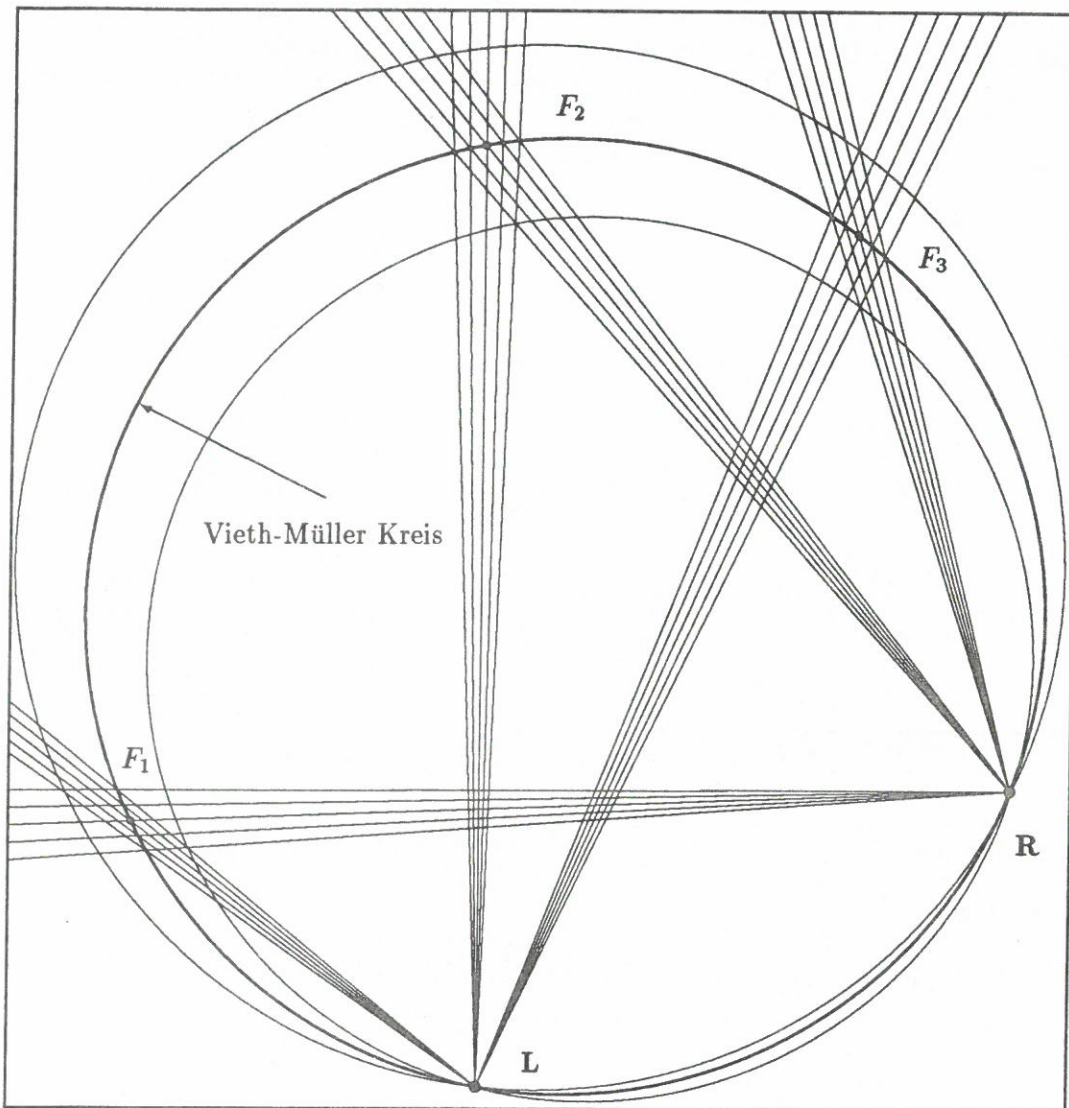


Abbildung 5.2: Kreise gleicher Querdisparität bei kreisförmiger Abbildungsfläche.

Bemerkungen: (1) Die zusätzlich dargestellten von L bzw. R ausgehenden Strahlen besitzen 1° Winkelabstand. Die an den Fixationspunkten entstehenden Rauten bilden also jeweils den Szenenbereich, der mit 4° Öffnungswinkel erfaßt wird. (2) Denken wir uns zu jedem der Fixationspunkte die Tangente an den Vieth-Müller Kreis, so zeigt sich, daß die Querdisparität für kleine Öffnungswinkel näherungsweise dem Abstand der einer bestimmten Querdisparität zugeordneten Kreise von dieser Tangente entspricht.

5.1.1.2 Ebene Abbildungsfläche

Haben wir – wie bei technischen Systemen fast immer – eine ebene Abbildungsfläche, ist der Zusammenhang zwischen Querdissparität und Tiefe nicht mehr ganz so einfach. Analog zu Definition 1 bezeichnen wir mit s_x die Querdissparität eines Punktes der Szene, die sich ebenfalls als Koordinatendifferenz zwischen dem linken und rechten Bildpunkt ergibt. Wir verwenden jetzt allerdings Kartesische Koordinaten.

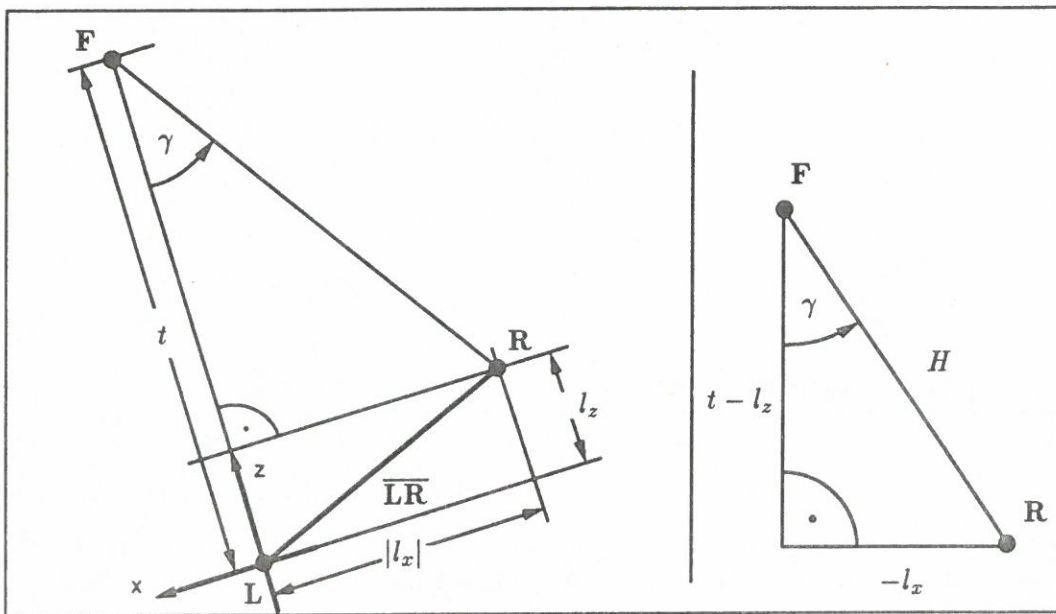


Abbildung 5.3: Links: Zur Berechnung der Querdissparität bei ebener Abbildungsfläche. Rechts: Zusammenhang für $\cos \gamma$ und $\sin \gamma$. Im folgenden wird immer $H < t$ und $\gamma < 90^\circ$ angenommen.

Ausgehend von einer Situation wie in Abbildung 5.3 links, berechnen wir jetzt die Verschiebung s_x und wählen dabei die Achsen des Koordinatensystems als x und z in dieser Weise, damit es nachher einfacher wird, die Formel für den dreidimensionalen Fall zu übernehmen. Der Ursprung des Koordinatensystems liege in L . Die Stereobasis sei \overline{LR} mit $R = (l_x, l_z)$ und es sei im folgenden immer $l_x < 0$, damit R immer „rechts“ von L liegt.

Für die linke Bildkoordinate x_L eines Punktes $P_L = (x, z)$ ergibt sich aus dem Strahlensatz sofort

$$x_L = f \cdot \frac{x}{z} \quad (5.1)$$

wobei f die Brennweite bezeichnen soll. Die rechte Bildkoordinate x_R erhält man auf entsprechende Weise, nachdem man das Koordinatensystem nach R verschoben und um $-\gamma$ gedreht hat. Verschieben ergibt

$$P' = (x - l_x, z - l_z) \quad (5.2)$$

Rotation um $-\gamma$ liefert

$$P_R = (\cos \gamma (x - l_x) - \sin \gamma (z - l_z), \sin \gamma (x - l_x) + \cos \gamma (z - l_z)) \quad (5.3)$$

und es ergibt sich analog zu Gleichung (5.1)

$$x_R = f \cdot \frac{\cos \gamma \cdot (x - l_x) - \sin \gamma \cdot (z - l_z)}{\sin \gamma \cdot (x - l_x) + \cos \gamma \cdot (z - l_z)} \quad (5.4)$$

Die Verschiebung definieren wir zu $s_x := x_R - x_L$. Setzen wir nun noch für $\cos \gamma$ bzw. $\sin \gamma$ die Terme $\frac{t-l_z}{H}$ und $-\frac{l_x}{H}$ ein (siehe Abbildung 5.3, rechts), so kürzt sich H heraus und die Bestimmungsgleichung für s_x lautet:

$$f \cdot \frac{(t - l_z)x + l_z z - t l_x}{-l_x x + (t - l_z)z + l_x^2 + l_z^2 - t l_x} - f \cdot \frac{x}{z} - s_x = 0 \quad (5.5)$$

Wir sehen durch Multiplikation mit f , daß sich die Brennweite unter $s_x^* = f \cdot s_x$ subsumieren läßt. Im folgenden bezeichnen wir s_x^* wieder mit s_x (führen also o.B.d.A die Berechnungen mit $f \equiv 1$ vor).

Wird Gleichung (5.5) vereinfacht, ergibt sich nach einigen Umstellungen^M:

$$\begin{aligned} l_x \cdot x^2 + (t l_z - l_x^2 - l_z^2) \cdot x + (l_x + s_x(l_z - t)) \cdot z^2 \\ + (s_x(t l_z - l_x^2 - l_z^2) - t l_x) \cdot z + s_x l_x \cdot x z = 0 \end{aligned} \quad (5.6)$$

Die Gleichung (5.6) gibt Anlaß zu folgendem Satz:

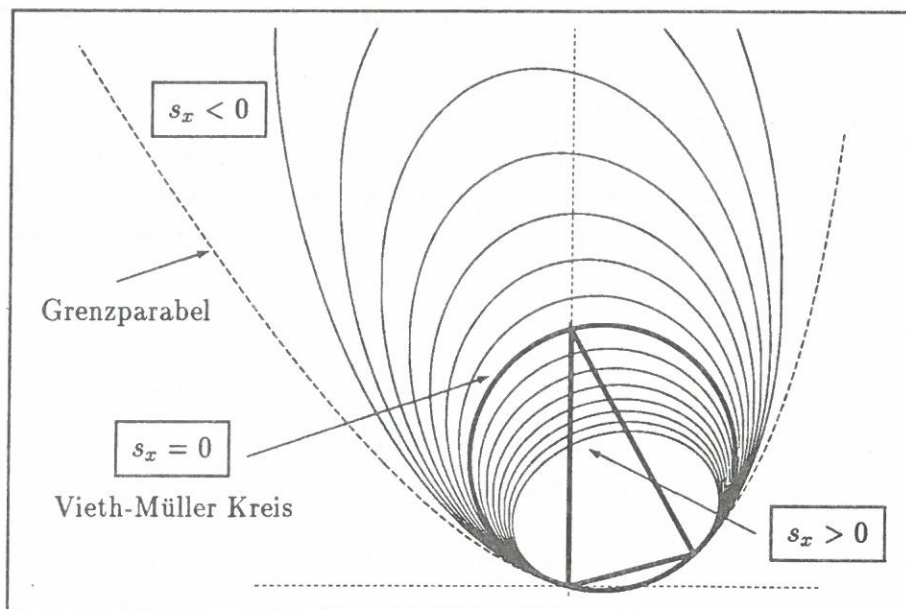


Abbildung 5.4: Isodisparitätätslinien nach Gleichung (5.5).

Satz 2 Isodisparitätätslinien sind Kegelschnitte.

Beweis. Die allgemeine Gleichung 2. Grades in x und y

$$F(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \quad (5.7)$$

beschreibt einen Kegelschnitt und besitzt die Invarianten

$$S_1 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad S_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \text{und} \quad S_3 = a_{11} + a_{22}$$

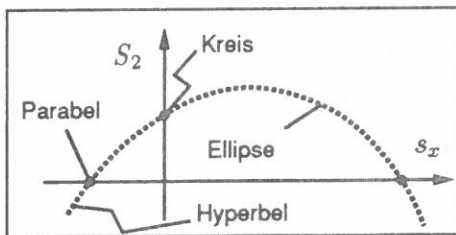
mit $a_{ik} = a_{ki}$ in den Determinanten. Ist $S_1 \neq 0$, so beschreibt die Gleichung einen eigentlichen Kegelschnitt. Unter dieser Bedingung stellt sie genauer für $S_2 < 0$ eine Hyperbel, für $S_2 = 0$ eine Parabel und für $(S_2 > 0) \wedge (S_1 \cdot S_3 < 0)$ eine reelle Ellipse dar.

Gleichung (5.6) ist eine solche Gleichung zweiten Grades mit^M

$$\begin{aligned} S_1 &= -\frac{1}{4} (l_x^2 + (l_z - t)^2) \cdot (l_x (l_x^2 + l_z^2) + s_x l_z (l_x^2 + l_z^2 - l_z t)) \\ S_2 &= -\frac{1}{4} l_x^2 \cdot s_x^2 + l_x (l_z - t) \cdot s_x + l_x^2 \\ S_3 &= 2l_x + s_x (l_z - t) \end{aligned}$$

Wir sehen am letzten Faktor von S_1 , daß Gleichung (5.6) sogar i.a. ein eigentlicher Kegelschnitt ist, denn nur für ein einziges s_x wird dieser Faktor Null.

Wir führen jetzt keine vollständige Kurvendiskussion durch, sondern wollen uns mit einer qualitativen Überlegung davon überzeugen, daß wir es in der Nähe des Fixationspunktes – also bei kleinen Verschiebungen – mit reellen Ellipsen zu tun haben. Dazu muß nach Satz 2 $(S_2 > 0) \wedge (S_1 \cdot S_3 < 0)$ gelten. Wir lesen die Werte der Invarianten für $s_x = 0$ aus den Gleichungen ab und erhalten $S_2 = l_x^2$ sowie $S_1 \cdot S_3 = \left(-\frac{1}{4} l_x (l_x^2 + (l_z - t)^2) \cdot (l_x^2 + l_z^2)\right) \cdot (2l_x) < 0$. Die Bedingungen sind also erfüllt. Da die Invarianten in s_x stetig sind, gilt das Ergebnis auch in einer genügend kleinen Umgebung von $s_x = 0$.



Qualitativer Verlauf der Invariante S_2 mit Bezeichnung des jeweils vorliegenden Kegelschnitttyps. Die Nullstellen von $S_2(s_x)$ liegen bei:

$$s_x = \frac{2 \cdot (l_z - t \pm \sqrt{l_x^2 + (l_z - t)^2})}{l_x}$$

Korollar 2.1 Für $s_x = 0$ ergibt sich ein Kreis, den wir wieder Vieth-Müller-Kreis nennen wollen.

Beweis. Durch Einsetzen wird Gleichung (5.6) zu

$$z^2 - tz + x^2 + \frac{(tl_z - l_x^2 - l_z^2)}{l_x} x = 0 \quad (5.8)$$

Das ist ein Kreis mit Mittelpunkt $M = \left(\frac{(l_x^2 + l_z^2 - tl_z)}{2l_x}, \frac{t}{2}\right)$.

Eine kurze Überlegung zur absoluten Tiefe wollen wir hier an die Kreisgleichung noch anschließen, bevor wir die Gleichungen für den Raum aufstellen.

Beim Betrachten von Abbildung 5.2 sehen wir, daß in der Nähe des Fixationspunktes die Isodisparitätlinien näherungsweise in gleichem Abstand und parallel zu der Tangente an den Vieth-Müller Kreis verlaufen. Zur Bestimmung der Steigung dieser Tangente („Referenzlinie“) formen wir die Kreisgleichung (5.8) nach z um^M, und erhalten

$$z(x) = \frac{t}{2} \pm \frac{1}{2} \sqrt{t^2 + 4 \left(\frac{l_x^2 + l_z^2 - l_z t}{l_x} \cdot x - x^2 \right)}$$

und für die Ableitung^M

$$z'(x) = \pm \frac{l_x^2 + l_z^2 - l_z t - 2l_x x}{l_x \sqrt{t^2 + 4 \left(\frac{l_x^2 + l_z^2 - l_z t}{l_x} \cdot x - x^2 \right)}}$$

also am Fixationspunkt ($x = 0$, erste Lösung)

$$z'(0) = \frac{l_x^2 + l_z^2 - l_z t}{l_x t}$$

Damit lautet die Geradengleichung der Normalen an diesem Punkt

$$g_{\bar{n}} : z(x) = \frac{l_x t}{l_z t - l_x^2 - l_z^2} \cdot x + t$$

Schneiden wir diese Gerade mit der Kegelschnittsgleichung, können wir die absolute Tiefe entlang der Normalen als Abstand des Schnittpunkts zum Fixationspunkt angeben. Wir erwarten zwei oder weniger Schnittpunkte zwischen einer Geraden und einem Kegelschnitt. Einsetzen der Geradengleichung in Formel (5.6) muß eine quadratische Gleichung in x liefern. Die Koeffizienten sind leider etwas umfangreichere Terme. Auflösen^M liefert zwei Werte x_1 und x_2 für x , von denen nur x_1 von s_x abhängt.

Einsetzen^M von x_1 in $g_{\bar{n}}$ liefert uns z_1 . Der zu einem bestimmten s_x gehörende Abstand d von $\mathbf{F} = (0, t)$ ergibt sich aus $d(s_x) = \sqrt{(x_1^2 + (t - z_1)^2)}$ nach Umformung^M zu

$$d(s_x) = \frac{\sqrt{l_x^2 + l_z^2} \cdot \sqrt{l_x^2 + (l_z - t)^2}}{l_x^2 - s_x l_x t + l_z^2} \cdot s_x t$$

Wenn wir noch, wie beobachtet, weitgehend lineares Verhalten in der Nähe des Fixationspunktes unterstellen, können wir nach s_x ableiten^M und bei $s_x = 0$ auswerten:

$$d = \frac{\sqrt{l_x^2 + (l_z - t)^2}}{\sqrt{l_x^2 + l_z^2}} \cdot t$$

Wie wir aus der Abbildung 5.2 sehen, ist der Fusionsbereich bei gleichschenkliger Anordnung größer als bei den anderen Konstellationen. Setzen wir also noch eine

gleichschenklige Anordnung voraus, d.h. $\mathbf{R} = (l_x, l_z)$ liege auf dem Kreis mit Mittelpunkt \mathbf{F} und Radius t . Dann gilt $l_x^2 + l_z^2 = 2l_z t$ und der absolute Tiefenbereich d_{symm} ist näherungsweise

$$d_{symm} = \frac{t^2}{\sqrt{2l_z t}} \cdot s_x \quad (5.9)$$

5.1.2 Betrachtungen im Raum

Analog zu den Betrachtungen im vorigen Abschnitt wollen wir jetzt beliebige Raumpunkte zulassen, und nicht nur Punkte, die in der Ebene der Stereoanordnung liegen. Dadurch ergibt sich außer der Verschiebung s_x in x -Richtung auch eine entsprechende Verschiebung s_y in y -Richtung. Wir wollen hier nicht mehr auf halbkugelförmige Retina eingehen, sondern gleich flache Retinae betrachten, wie wir sie in technischen Systemen vorliegen haben.

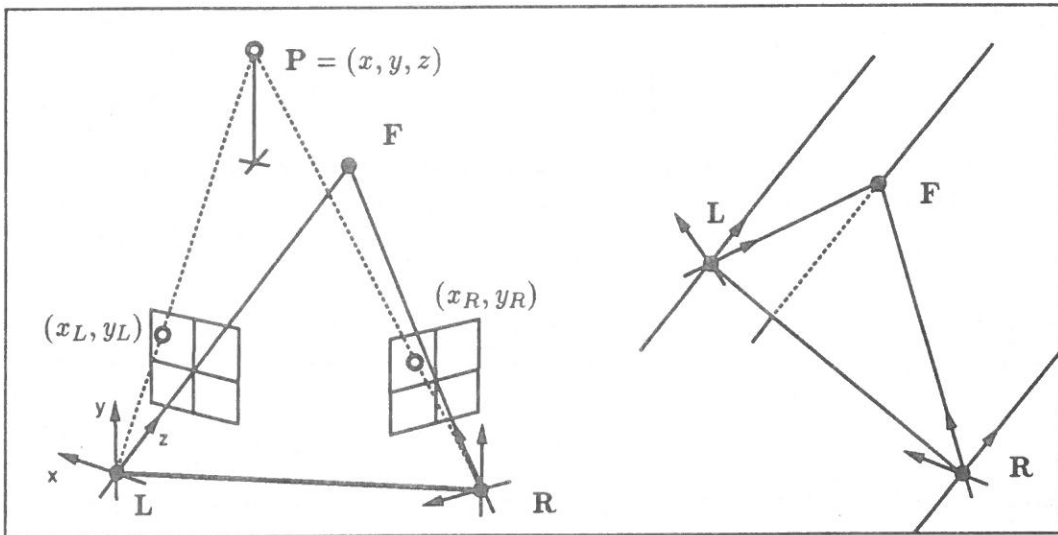


Abbildung 5.5: Links: Stereoanordnung zur Herleitung der Bestimmungsgleichungen für s_x und s_y . Skizze zur Illustration der im Text benutzten Bezeichnungen. Rechts: Da drei Punkte eine Ebene festlegen, kann jede beliebige Stereoanordnung (evtl. durch Drehung der Koordinatensysteme) in die links abgebildete Situation überführt werden.

Der Ausgangspunkt ist eine Situation mit den Bezeichnungen wie in Abbildung 5.5. Der 3D Punkt $\mathbf{P} = (x, y, z)$ hat (analog zu Gleichung (5.1)) im L-System die Bildkoordinaten

$$(x_L, y_L) = \left(f \cdot \frac{x}{z}, f \cdot \frac{y}{z} \right)$$

²Damit sind wir auf anderem Wege zu der gleichen approximativen Beziehung gekommen, wie sie von A. Korn [Kor82, S.166] für kreisförmige Retina bei symmetrischer Stereoanordnung angegeben wird.

Verschiebung und Drehung des Koordinatensystems nach \mathbf{R} ergibt analog zur Rechnung bis Gleichung (5.4)

$$\begin{aligned}x_R &= f \cdot \frac{\cos \gamma \cdot (x - l_x) - \sin \gamma \cdot (z - l_z)}{\sin \gamma \cdot (x - l_x) + \cos \gamma \cdot (z - l_z)} \\y_R &= f \cdot \frac{y}{\sin \gamma \cdot (x - l_x) + \cos \gamma \cdot (z - l_z)}\end{aligned}$$

und Einsetzen der Beziehungen für $\cos \gamma$ und $\sin \gamma$ sowie Umstellen nach x und z ergibt^M

$$\begin{aligned}x_R &= f \cdot \frac{l_x t + (l_z - t)x - l_x z}{(l_x^2 + l_z^2 - l_z t) - l_x x + (t - l_z)z} \\y_R &= f \cdot \frac{Hy}{(l_x^2 + l_z^2 - l_z t) - l_x x + (t - l_z)z}\end{aligned} \quad (5.10)$$

Aufstellen der Bestimmungsgleichung $y_R - y_L - s_y = 0$ und Umformen^M liefert, da sich $H = \sqrt{t^2 - 2l_z t - l_x^2}$ diesmal nicht herauskürzt, schließlich insgesamt

$$\begin{aligned}l_x x^2 + (tl_z - l_x^2 - l_z^2)x + (l_x + s_x(l_z - t))z^2 \\+ (s_x(tl_z - l_x^2 - l_z^2) - tl_x)z + s_x l_x xz = 0 \\s_y(l_z - t)z^2 + s_y(tl_z - l_x^2 - l_z^2)z + (tl_z - l_x^2 - l_z^2)y \\+ l_x xy + s_y l_x xz + (H + l_z - t)yz = 0\end{aligned} \quad (5.11)$$

Die Gleichungen für s_x und s_y haben jeweils die Form einer Hyperfläche 2. Grades. Auch diese Gleichungen kann man entsprechend nach ihren Invarianten klassifizieren. Wir wollen das etwas abkürzen, da wir die Gleichung für s_x schon kennen.

Die erste Gleichung beschreibt – wie schon erläutert – für $y = 0$ einen Kegelschnitt. Wir haben es im Dreidimensionalen deshalb einfach mit einem Zylinder zu tun: Je nach Kegelschnittsform mit einem elliptischen, parabolischen oder hyperbolischen Zylinder.

Die Gleichung für s_y ist etwas komplizierter. Der folgende Satz faßt die Aussagen zusammen.

Satz 3 Die Raumpunkte mit $s_y = 0$ liegen in einer Ebene und Raumpunkte mit ($s_y \neq 0 \wedge s_y = \text{const}$) auf einem Kegel.

Beweis. Das allgemeinste Polynom 2. Grades in x , y und z

$$\begin{aligned}F(x, y, z) &= a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}xz + 2a_{23}yz + a_{33}z^2 \\&\quad + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} = 0\end{aligned}$$

beschreibt eine Hyperfläche und hat die Invarianten

$$S_1 = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}, \quad S_2 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$S_3 = a_{11} + a_{22} + a_{33} \quad \text{und}$$

$$S_4 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{33} & a_{13} \\ a_{31} & a_{11} \end{vmatrix}.$$

wieder mit $a_{ik} = a_{ki}$ in den Determinanten. Die Gleichung beschreibt einen Kegel für $(S_1 = 0) \wedge ((S_2 S_3 < 0) \vee (S_4 < 0))$. Außerdem zerfällt die durch $F(x, y, z)$ beschriebene Fläche in zwei Ebenen, falls

$$S_5 = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} & a_{14} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} = 0$$

Die Gleichung für s_y ist eine solche Fläche und die Invarianten berechnen sich^M zu

$$S_1 = 0$$

$$S_2 = \frac{H l_x^2 s_y}{4}$$

$$S_3 = s_y (l_z - t)$$

$$S_4 = -\frac{1}{4} \left((1 + s_y^2) l_x^2 + (H - l_z - t)^2 \right) < 0$$

Für $s_y = 0$ zerfällt die Gleichung wegen $S_5 = 0^M$, und wie man auch durch Einsetzen in (5.11) sehen kann, in zwei Ebenen zu:

$$\underbrace{y}_{E_1} \cdot \underbrace{(l_z t - l_x^2 - l_z^2 + l_x x + (H - l_z - t)z)}_{E_2} = 0$$

Die Ebene E_2 steht senkrecht auf E_1 und schneidet sie in einer Geraden, in die wir die x -Koordinate von \mathbf{R} einsetzen um zu sehen, daß E_2 „hinter“ \mathbf{R} liegt: d.h. sie ist wegen ihrer Lage hier nicht weiter relevant.

$$E_2^{xz}: \quad z(x) = -\frac{l_x}{H + l_z - t} \cdot x + \frac{l_x^2 + l_z^2 - l_z t}{H + l_z - t}$$

Nun ist

$$0 < H \Rightarrow l_z - t < H + l_z - t \Rightarrow \frac{l_z - t}{H + l_z - t} < 1 \Rightarrow$$

$$z(\mathbf{R}_x) = z(l_x) = \frac{l_z^2 - l_z t}{H + l_z - t} < l_z = \mathbf{R}_z$$

Entsprechend kann man für **L** und **F** argumentieren. Also liegen die Punkte mit $s_y = 0$ in jeder praktischen Stereoanordnung in der Ebene der Stereoanordnung. Die Gleichung stellt für $(s_y = \text{const}) \wedge (s_y \neq 0)$ einen Kegel dar, weil $S_4 < 0$ gilt und wegen

$$S_2 \cdot S_3 = \frac{H l_x^2 s_y^2 (l_z - t)}{4} < 0$$

die genannte Bedingung erfüllt ist.³

Man kann die Kegelgleichung nach $y(x, z)$ umformen und den Wert für $y(0, t)$ – also die Höhe eines Punktes mit Fixationspunktabszissen – ausrechnen und dann nach s_y zu

$$s_y = \frac{y(0, t)(t - H)}{Ht}$$

umformen^M. An dieser Gleichung kann man sehen, daß s_y in der Nähe des Fixationspunktes i.a. recht kleine Werte annimmt.

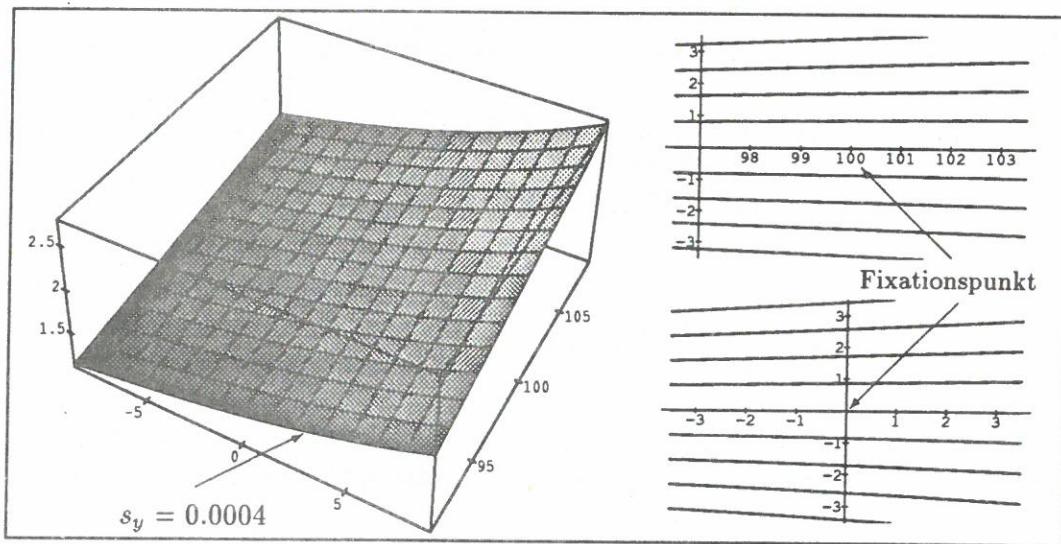


Abbildung 5.6: Links: Fläche der 3D Punkte mit $s_y = \text{const}$ in einer Umgebung (5° Öffnungswinkel) des Fixationspunktes. Rechts: Querschnitte am Fixationspunkt mit $\Delta s_y = 0.0002$ für $x = 0$ (oben) und $z = t = 100$ (unten, $\sim 2^\circ$).

Abbildung 5.6 zeigt $y(x, z)$ und Querschnitte für verschiedene s_y in der Nähe des Fixationspunktes. Der Kegelmantel verläuft, wie zu sehen, weitgehend waagrecht (Man beachte dazu in der linken Zeichnung, daß der dargestellte Abschnitt der z -Achse – um überhaupt etwas von der geringen Krümmung darstellen zu können – etwa 20 mal kleiner ist, als an den anderen beiden Achsen).

³Man kann hier auch die Eigenwerte der zu S_2 korrespondierenden Matrix ausrechnen (Hauptachsentransformation) und sich überzeugen, daß für $s_y \neq 0$ je zwei davon positiv sind und einer negativ ist; d.h. es liegt ein Kegel vor.

Die Menge der Raumpunkte mit gleicher Verschiebung (s_x, s_y) liegen also auf der Schnittkurve eines (in der Nähe des Fixationspunktes fast horizontal verlaufenden) Kegelmantels mit einem (senkrechten) Zylinder.

Schlußbemerkungen. Wir haben in Abschnitt 5.1.1 gesehen, daß sich im Falle einer ebenen Anordnung die (relative) Tiefe und die Querdisparität nur durch einen Faktor zusammenhängen. Verkürzt kann man also sagen: Querdisparität ist Tiefe. Es soll an dieser Stelle nicht unerwähnt bleiben, daß man Tiefe im dreidimensionalen Fall ganz anders definieren kann. Verwendet man als Tiefe eines 3D Punktes beispielsweise den Abstand vom Mittelpunkt der Stereobasis zu diesem Punkt, dann ergeben sich als Raumpunkte gleicher Tiefe dann Kugelschalen; D.h. insbesondere auch in y -Richtung gekrümmte Tiefenflächen.

Wenn wir dagegen die Verschiebung s_y in der Nähe des Fixationspunktes vernachlässigen, weil sie dort im Vergleich mit s_x sehr geringe Werte annimmt und sich auch praktisch mit Werten um einen Pixel⁴ im Bereich der Disparitätsauflösungsgrenze befindet, so gilt die Kurzformel 'Querdisparität ist Tiefe' auch für den dreidimensionalen Fall. Wie bereits festgestellt, liegen damit die Raumpunkte gleicher Tiefe auf zylindrischen Schalen. Die Punkte mit Verschiebung Null liegen auf einem Kreiszyylinder, die anderen auf entsprechenden elliptischen Zylindern.

5.2 Abbildungsgeometrie bei lokaler Oberflächenapproximation

Aus den bisher abgeleiteten Formeln läßt sich angeben, welche Koordinaten ein beliebiger einzelner 3D Punkt im rechten Bildkoordinatensystem hat, wenn seine Koordinaten im linken Bildkoordinatensystem bekannt sind; oder anders ausgedrückt, wie sich ein Bildpunkt vom linken System ins rechte transformiert.

Wir sind aber daran interessiert, etwas über das Transformationsverhalten größerer Punktmengen – nämlich des fixierten Oberflächenabschnitts – zu erfahren. Wir können hoffen, daß die Transformationsgleichung einfacher wird, wenn wir die Gleichungen für bestimmte Oberflächenformen spezialisieren.

Jede physikalische – also i.a. stetige und differenzierbare – Oberfläche läßt sich lokal in eine Taylorreihe entwickeln. Die Taylorapproximation 1. Ordnung am Fixationspunkt ist einfach eine Ebene, die Schmiegeebene an die Objektoberfläche an diesem Punkt.

Für uns ist es weniger wichtig, welche Bildkoordinaten (x_L, y_L) bzw. (x_R, y_R) ein gegebener 3D Punkt der Objektoberfläche besitzt. Wir wollen diese Abbildungssituation anders betrachten und wie bisher auch (vgl. Formel (5.10)) nach der Transformation \mathcal{T} fragen, die uns für einen Bildpunkt (x_L, y_L) im linken System, den – über einen gemeinsamen 3D Punkt korrespondierenden – rechten Bildpunkt (x_L, y_L) liefert: $(x_R, y_R) = \mathcal{T}\{(x_L, y_L)\}$

⁴Dieser Wert bezieht sich auf eine Gesamtauflösung von ± 16 Pixel für die Disparitäten (vgl. auch Abbildung 8.4 oder 8.6).

5.2.1 Approximation 1. Ordnung

Wir wollen jetzt sehen, wie sich die Formel für die Disparitäten s_x bzw. s_y vereinfacht, wenn die fixierte Objektoberfläche in erster Näherung durch eine Ebene beschrieben wird. Dazu gehen wir von einer Anordnung aus, wie in Abbildung 5.7 skizziert ist. Der Ursprung des Koordinatensystems liege wieder in L.

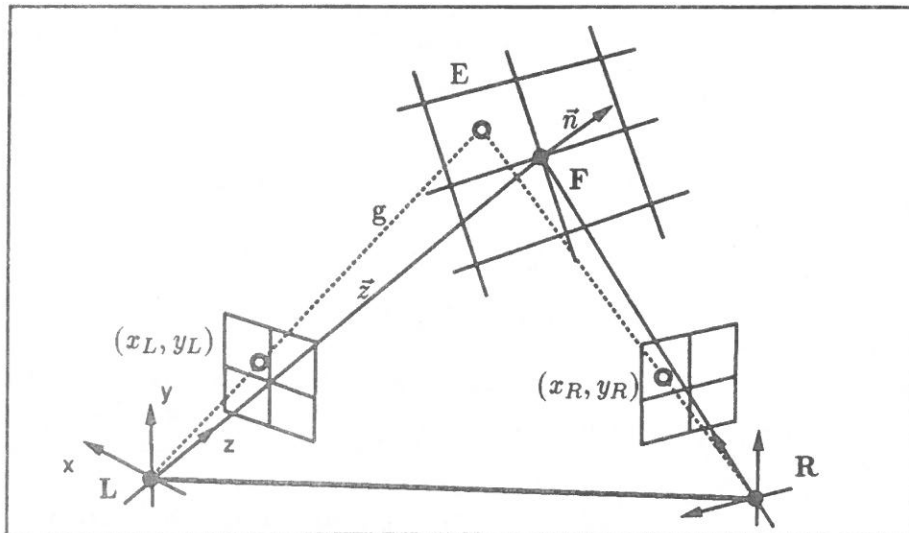


Abbildung 5.7: Stereoaufnahme mit linearer Oberflächenapproximation am Fixationspunkt.

Beschreiben wir die Ebene durch den Fixationspunkt $\vec{z} = (0, 0, t)$ mit ihrem Normalenvektor $\vec{n} = (n_x, n_y, n_z)$ und verlängern wir die Strecke durch den Ursprung des L-Systems und einen gegebenen Bildpunkt $\vec{s}_L = (x_L, y_L, f)$ weiter in Richtung der Szene, so erhalten wir als Durchstoßpunkt denjenigen 3D Punkt der Ebene, der sich bei Brennweite f auf (x_L, y_L) abbildet. Diesen Punkt können wir in Gleichung (5.10) einsetzen und die Formel vereinfachen.

Der Schnittpunkt ergibt sich zunächst z.B. aus der Ebenengleichung (in Punkt-Richtungsform) und der Geradengleichung (in Parameterform)

$$\begin{aligned} E &: \vec{n} \cdot \vec{x} - \vec{n} \cdot \vec{z} = 0 \\ g &: \vec{0} + \lambda \cdot \vec{s}_L \end{aligned}$$

durch Einsetzen eines Geradenpunktes in die Ebenengleichung und Auflösen nach λ_s

$$\begin{aligned} \vec{n} \lambda_s \vec{s}_L - \vec{n} \vec{z} &= 0 \quad \Rightarrow \\ \lambda_s &= \frac{\vec{n} \vec{z}}{\vec{n} \vec{s}_L} = \frac{n_z t}{n_x x + n_y y + n_z f} \end{aligned}$$

Setzen wir diesen 3D Punkt $\mathbf{P}_L = \lambda_s \cdot \vec{s}_L$ in (5.10) ein, so erhalten wir nach Umformung^M einen kompakten Ausdruck. Wir können das Ergebnis dieser Berechnung in den folgenden Satz fassen.

Satz 4 Bei Oberflächenapproximation 1. Ordnung transformieren sich die Bildkoordinaten nach

$$\begin{aligned} x_R &= \frac{a_1 x_L + a_2 y_L}{b_1 x_L + b_2 y_L + b_3} \\ y_R &= \frac{c \cdot y_L}{b_1 x_L + b_2 y_L + b_3} \end{aligned} \quad (5.12)$$

wobei die folgenden Größen a_i, b_j und c

$$\begin{aligned} a_1 &= ft(n_z(t - l_z) - n_x l_x), & b_1 &= n_x(l_x^2 + l_z^2) - t(n_x l_z + n_z l_x) \\ a_2 &= -ft n_y l_x, & b_2 &= n_y(l_x^2 + l_z^2 - l_z t) \\ c &= f n_z t H, & b_3 &= f n_z(l_x^2 + l_z^2 + t^2 - 2l_z t) \end{aligned}$$

Konstanten in Abhängigkeit der gewählten Stereoaordnung sind.

Die folgenden Abbildungen machen deutlich, wie sich die Punkte nach dieser Formel bei konkreten Stereoaordnungen transformieren. Zur Darstellung der Transformation gibt es mehrere Möglichkeiten. Man kann sich ansehen, wie ein quadratisches Gitter auf der linken Bildebene nach der Transformation auf der rechten Bildebene erscheint. Man kann auch die Disparität selbst darstellen; da dies ein Vektor ist also z.B. seine Komponenten s_x, s_y einzeln, seinen Betrag oder seine Richtung.

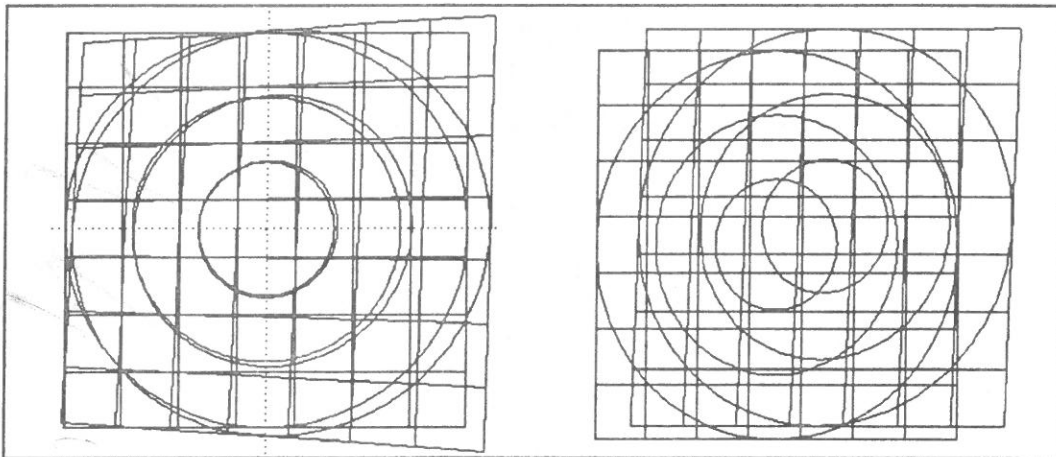


Abbildung 5.8: Verzerrung eines quadratischen Gitters mit konzentrischen Kreisen. Das auf der rechten Bildebene entstehende (transformierte) Gitter ist dem linken Gitter überlagert dargestellt. Die Parameter dieser Beispielanordnung entsprechen der Situation, daß ein Mensch ($\overline{LR} \approx 7\text{cm}$, $f = 1.7\text{cm}$) etwas in 1m Entfernung fixiert ($\vec{n} = (-0.1, 0.4, -0.7)$, $t = 100$, $l_x = -6.5$, $l_z = 2.5$). Links: Öffnungswinkel 100° . Rechts: Das gleiche Gitter wurde auf einen sehr kleinen (fovealen) Teilausschnitt ($\frac{1}{8}^\circ$, im ersten Quadranten) aufgetragen. Hier liegt im wesentlichen nur noch eine Verschiebung und kaum Verzerrung vor.

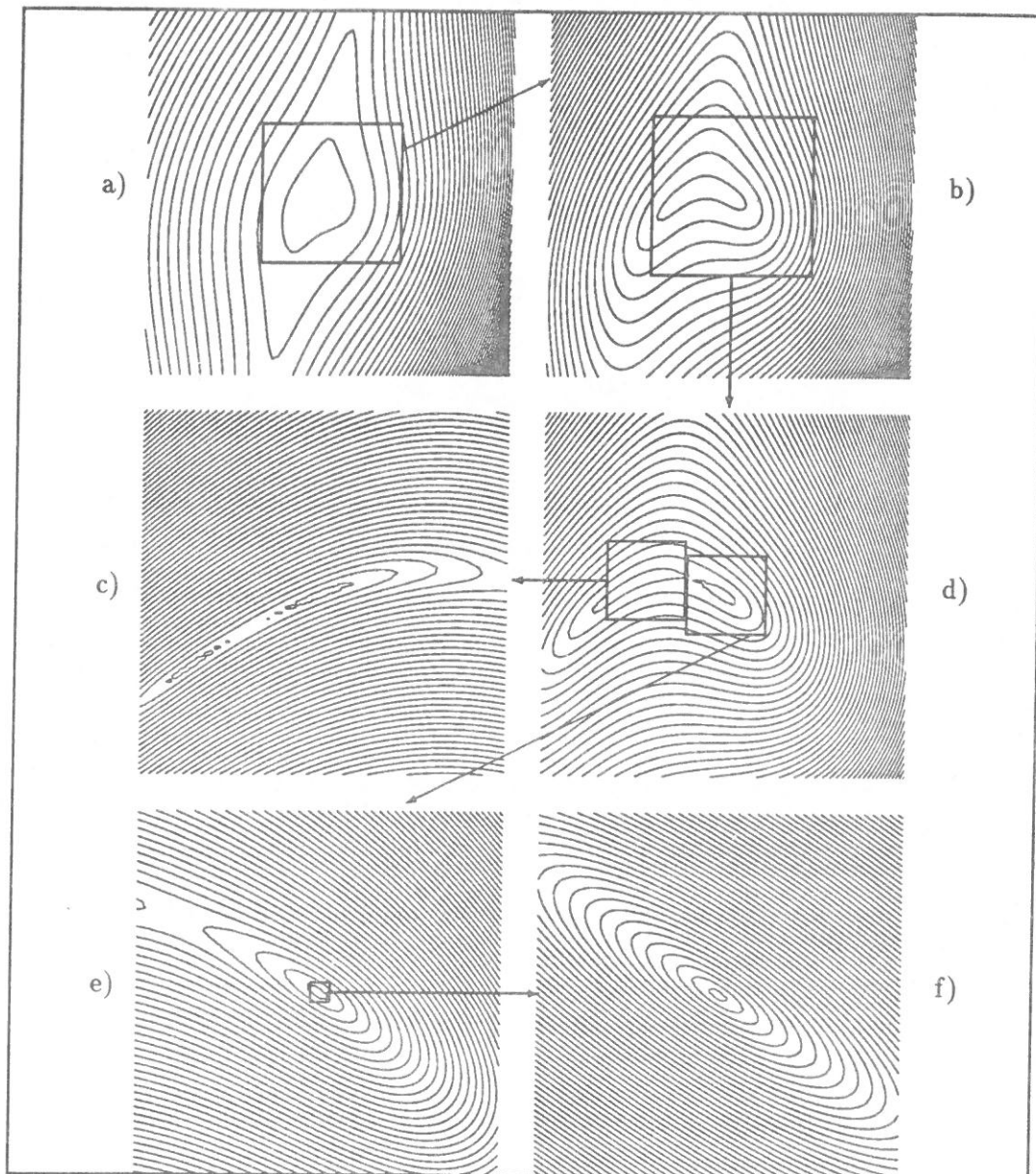


Abbildung 5.9: Isodisparitätlinien (Betrag) bei verschiedenen Öffnungswinkeln (a) 160° (b) 125° (c) 20° (d) 80° (e) 20° (f) 1° . Stereokonfiguration wie vorige Abbildung. Es wird deutlich, daß sich die Punkte bei großen Öffnungswinkeln in einer Kachel (vgl. Abbildung 4.2) für das Cepstrum sehr unterschiedlich transformieren können. Die Abbildungen zeigen weiter, daß im nicht-fovealen Bereich außer dem Fixationspunkt weitere Punkte mit Verschiebung Null existieren können (siehe zu diesen Abbildungen auch [KvD76]).

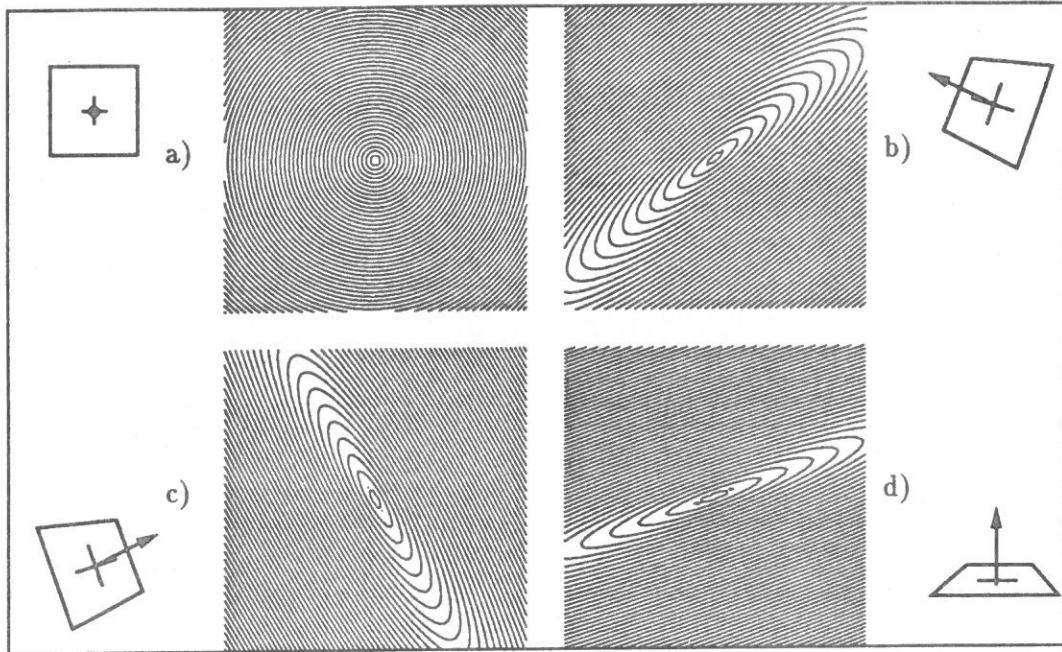


Abbildung 5.10: Isodisparitätslinien (Betrag) bei verschiedenen Normalenvektoren. Konfigurationsparameter der Stereoanordnung wie zuvor. Die Ebene ist bezüglich der Stereoanordnung frontoparallel (a), nach links gekippt (b), nach rechts gekippt (c) und nach hinten gekippt (d). (a) $\vec{n} = (-0.03073, 0.0, -0.99953)$ (b) $\vec{n} = (0.7, 0.5, -0.7)$ (c) $\vec{n} = (-0.7, 0.5, -0.7)$ und (d) $\vec{n} = (0.02284, 0.66896, 0.74294)$.

Die Abbildung 5.8 veranschaulicht, daß im fovealen Bereich kaum Verzerrungen, aber deutliche Verschiebungen vorliegen; speziell ist die Verschiebung in x -Richtung größer als die in y -Richtung.

Die Abbildungen 5.9 und 5.10 zeigen, daß sich die Punkte im nicht-fovealen Bereich innerhalb einer Kachel für die Verschiebungsbestimmung mit der Cepstrumtechnik derart unterschiedlich von links nach rechts transformieren können, daß die Disparitätsbestimmung hier aufgrund der vielen verschiedenen Disparitäten nur einen (unbrauchbaren) Mittelwert liefern kann.

5.2.2 Approximation 2. Ordnung

Ich habe auch versucht, die zu (5.12) korrespondierende Formel für die Approximation der Objektoberfläche am Fixationspunkt mit Hyperflächen 2. Grades (Flächen 2. Ordnung) zu ermitteln und auch in eine ebenso knappe Form zu bringen. Die Formel ist leider unhandlich lang und besitzt – da diese Flächen gekrümmt sind – auch einige Fallunterscheidungen bei Selbstverdeckung, so daß ich sie hier nicht detailliert diskutieren möchte.

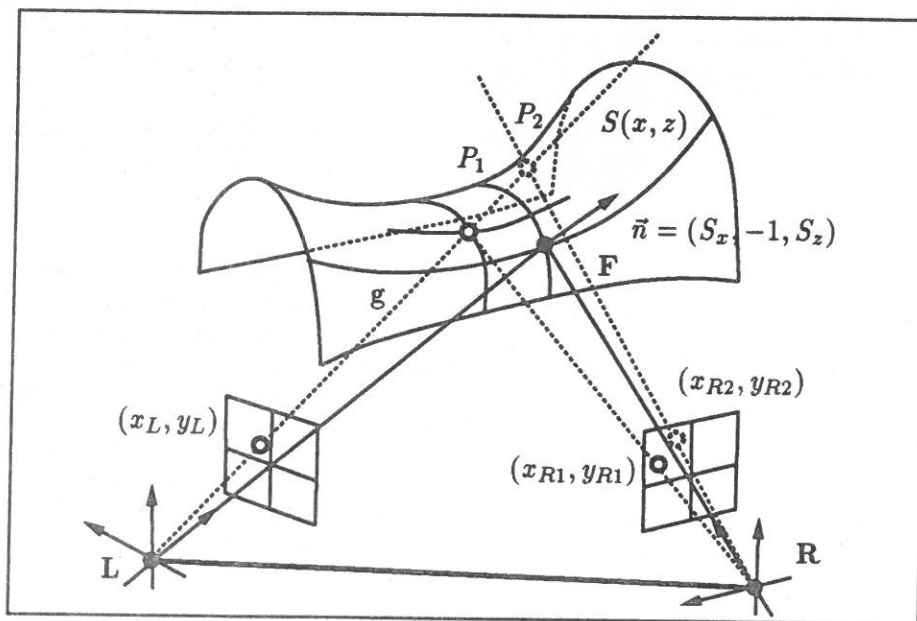


Abbildung 5.11: Stereoanordnung mit Oberflächenapproximation 2. Ordnung am Fixationspunkt.

Man kann die Formel ganz wie im Falle der Approximation 1. Ordnung aufstellen und vereinfachen. Beschreiben wir die Fläche S am Fixationspunkt mit Hilfe der Formel von Taylor z.B durch

$$S(x, z) = S_x x + S_z(z - t) + (S_{xz} + S_{zx}) \cdot x \cdot (z - t) + S_{xx}x^2 + S_{zz}(z - t)^2$$

wobei mit S_α bzw. $S_{\alpha\beta}$ die partiellen Ableitungen von S nach α bzw. α und β bezeichnet sind. Man kann wieder den Schnittpunkt der Geraden durch einen Bildpunkt mit der Fläche bestimmen. In diesem Fall ergeben sich i.a. zwei Schnittpunkte x_1 und x_2 oder auch keiner.

Unter günstigen Bedingungen kann man sich für einen der Schnittpunkte, sagen wir x_1 , entscheiden und x_1 dann auch in der gesamten, betrachteten Umgebung des Fixationspunktes verwenden und die Disparitätskarten wie im vorigen Abschnitt berechnen. Wählt man die Umgebung des Fixationspunktes so klein oder – was gleichwertig ist – die Hauptkrümmungen so gering, daß keine Schwierigkeiten auftreten, kann man praktisch wieder linear approximieren. Bei den von mir untersuchten Beispielen waren die Disparitätskarten für 1° Öffnungswinkel von denen bei

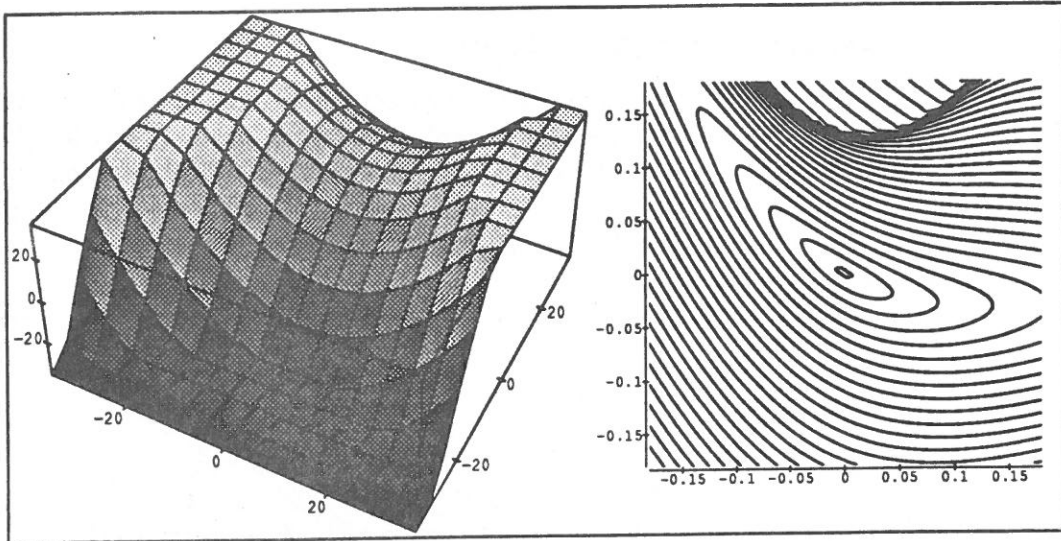


Abbildung 5.12: Transformation bei Oberflächenapproximation 2. Ordnung in größerer Umgebung des Fixationspunktes. Rechts: Darstellung der Isodisparitätätslinien (Betrags) bei 20° Öffnungswinkel. Zum direkten Vergleich mit Abbildung 5.9 (e) wurden die Parameter der Approximationsfläche so gewählt, daß sie als Approximation 1. Ordnung die dort gewählte Ebene besitzt: $S_x = \frac{1}{4}$, $S_z = \frac{7}{4}$, $S_{xx} = 1$, $S_{zz} = -1$, $S_{xz} = S_{zx} = 0$. Links: Darstellung von $S(x, z)$. Bemerkungen: (1) Bei der Darstellung von S hat MATHEMATICA Werte außerhalb des Darstellungswürfels auf den jeweiligen Extremwert gesetzt. Die Fläche hat also an diesen drei Schnittflächen tatsächlich größere bzw. kleinere Werte. (2) Die parallelen Linien in der Mitte oben (rechte Darstellung) sind keine echten Disparitätätslinien; hier ergeben sich nur mathematisch noch Werte, obwohl kein logisch sinnvoller Schnittpunkt mit der Sattelfläche mehr vorliegt (vgl. mit $S(x, z)$).

Approximation 1. Ordnung optisch nicht unterscheidbar, so daß Abbildung 5.9 (f) bei entsprechenden Werten für die partiellen Ableitungen auch für Approximation 2. Ordnung gilt.

Ich habe daher zur Illustration einen größeren Öffnungswinkel (20°) für die Abbildung 5.12 (vgl. Abbildung 5.9 (e)) gewählt, um die Unterschiede zu verdeutlichen.

5.3 Disparität und relative Tiefe

Bei einer Stereoanordnung können folgende Abstände als Tiefe für einen gegebenen Szenenpunkt dienen:

- Der Abstand vom linken Auge zum Szenenpunkt.
- Der Abstand vom rechten Auge zum Szenenpunkt.
- Der Abstand vom Mittelpunkt der Stereobasis (Zyklopische Tiefe).
- Die Horoptertiefe (Querdisparität evtl. mit einem Faktor multipliziert).

Die ersten drei Abstandsmaße sollten verständlich sein. Mit zunehmender Entfernung des Fixationspunktes unterscheiden sich die Werte dieser drei Maße immer weniger und können daher für große Entfernung als gleichwertig angesehen werden. Für kleine Entfernungen sind die ersten beiden Maße unsymmetrisch und man wird deshalb eher das dritte verwenden wollen.

Ich werde, da ich nur an relativer Tiefe interessiert bin, das vierte Abstandsmaß verwenden. Wie aus dem ganzen Kapitel klar geworden ist, besteht zwischen Tiefe am Fixationspunkt und Querdiparität sowohl bei kreisförmiger als auch bei ebener Retina näherungsweise ein 1:1 Zusammenhang. Ich bezeichne die unmittelbar an die Querdiparität nur durch einen Faktor gekoppelte Tiefe als Horoptertiefe – also Tiefe relativ zum Horopter. Falls absolute Tiefeninformation benötigt wird, kann man z.B. wie beschrieben vorgehen und Gleichung (5.9) benutzen.

5.4 Zusammenfassung

Es sind also folgende Dinge festzuhalten:

1. Die Querdiparität s_x hängt nur von der x - und z -Koordinate eines Punktes ab, nicht jedoch von y , während die Verschiebung s_y von allen drei Koordinaten abhängt. Letztere ist in der Nähe des Fixationspunktes i.a. gering. Das wesentliche Kennzeichen ist also s_x .
2. Wird die fixierte Oberfläche durch eine Ebene approximiert, ergibt sich eine einfache Funktion für die Verschiebung. Die Karten der Isodiparitätslinien veranschaulichen, daß bei fovealem Blickwinkel der Fixationspunkt der einzige Punkt ist, der die Verschiebung $(0, 0)$ besitzt.
3. Für die Anwendung der Cepstrumtechnik halten wir fest: Die Größenordnung der Verschiebungen variiert i.a. nur in einem engen Blickwinkel genügend wenig, um mit fester Streifenbreite bestimmt werden zu können. Die Rechtecke für die Teilbildextraktion dürfen nicht zu groß gewählt werden, da sich sonst die Punkte eines Rechteckes zu unterschiedlich transformieren und eine Verschiebungsbestimmung nur einen (wahrscheinlich unbrauchbaren) Mittelwert liefern kann.

Kapitel 6

Untersuchung des 2D Cepstrums

Dieses Kapitel gliedert sich in zwei Abschnitte. Da es mehrere Möglichkeiten der konkreten Implementation der Cepstrumfilterung gibt, beschreibt der erste Abschnitt die Implementation und begründet die spezielle Wahl der Umsetzung der genannten Formeln. Der zweite Abschnitt enthält eine Beschreibung experimenteller Untersuchungen zur Cepstrumfilterung. Ich führe die einzelnen Untersuchungsaspekte jeweils an prototypischen Experimenten vor und wir diskutieren die Ergebnisse im Vergleich mit den Angaben aus der Literatur.

6.1 Implementation

6.1.1 Allgemeines

Die Implementation erfolgte in der Programmiersprache C unter dem Betriebssystem Unix parallel auf einer Sun SPARCstation (Sun 4/370) und einem Personal Computer (AT 386). Zur Darstellung der Bilddaten und Ergebnisse, habe ich das X Window System (X11R4) verwendet.

6.1.1.1 Schnelle Transformationen (FFT und FHT)

Bei der Definition der kontinuierlichen Fouriertransformation gibt es einige Unterschiede. Ich habe insgesamt sechs verschiedene Definitionen angetroffen, die sich jeweils im Integrationsfaktor, im Vorzeichen der Exponentialfunktion, in der Frequenzvariablen (Frequenz oder Kreisfrequenz) oder in Kombinationen davon unterscheiden (vgl. Formel (3.2)). Da diese Unterschiede sich entsprechend bei den Algorithmen zur diskreten Fouriertransformation auswirken, muß man dies beachten, wenn man seine eigene Implementation mit den Ergebnissen anderer Implementationen oder Tabellen überprüfen oder vergleichen will. Ich gehe hier nicht mehr auf die Zusammenhänge zwischen kontinuierlicher Fouriertransformation, diskreter Fouriertransformation und schneller Fouriertransformation (FFT) ein. Eine meiner

Meinung nach immer noch ungeschlagene Einführung ist das Buch von Rafael C. Gonzales und Paul Wintz [GW83], das auch über die zahlreichen historischen Varianten der FFT Implementationen berichtet. Für weitere Einzelheiten sind auch die Bücher von R. Blahut [Bla87] oder R.N. Bracewell [Bra86a] zu empfehlen.

Die Ausführungszeiten lassen sich gegenüber einer FFT halbieren, indem man die Hartley Transformation [Bra86b] verwendet.

Die Hartleytransformierte einer kontinuierlichen Funktion $f(x, y)$ ist als

$$\mathcal{H}\{f(x, y)\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot \text{cas}(2\pi(ux + vy)) \, dx \, dy$$

definiert, wobei $\text{cas}(t) \equiv \cos(t) + \sin(t)$ gilt. Die Theoreme für die Fouriertransformation – wie z.B. *shift theorem*, *scaling*, *convolution*, *rotation*, ... – existieren auch für die Hartleytransformation. Einige der Theoreme sind identisch, andere gelten in einer leicht veränderten Form.

Bekanntermaßen gilt für die diskrete Fouriertransformation eines reellen Bildes, daß sie im Real- und Imaginärteil jeweils konjugiert-komplexe Werte enthält, so daß die tatsächliche Nettoinformation der Transformierten geringer ist. Dies gilt für die Hartleytransformation nicht. Die Hartleytransformierte ist wieder reell und läßt sich anschaulich als die Differenz von Real- und Imaginärteil der Fouriertransformierten beschreiben. Aus der Hartleytransformierten kann leicht der Real- und Imaginärteil der Fouriertransformierten wiedergewonnen werden.¹ Das Leistungsspektrum $\mathcal{S}(u)$ und die Phase $\mathcal{P}(u)$ ergeben sich ähnlich wie bei der Fouriertransformation:

$$\mathcal{S}(u) = F_{\text{real}}(u)^2 + F_{\text{imag}}(u)^2 = \frac{H(u)^2 + H(-u)^2}{2}$$

$$\mathcal{P}(u) = \text{atan} \left[\frac{F_{\text{imag}}(u)}{F_{\text{real}}(u)} \right] = \text{atan} \left[\frac{H(-u) - H(u)}{H(-u) + H(u)} \right]$$

Über eine Verwendung der Walsh-Transformation – als schnelle, aber auch nur approximative Berechnung der Fouriertransformierten – habe ich nachgedacht, aber in diesem Kontext auf eine Untersuchung verzichtet. Inwieweit also bei Verwendung der Walsh-Transformation die in dieser Arbeit bestimmten Eigenschaften des Cepstrums erhalten bleiben, muß noch untersucht werden.

¹R.N. Bracewell [Bra86b] gibt verschiedene Algorithmen für eine schnelle 1D Hartleytransformation (FHT) an. Ich habe den Grundalgorithmus implementiert und etwas verbessert. Der Vorteil der FHT ist, daß nicht wie bei der FFT erst noch ein mit Null gefüllter Imaginärteil mitberechnet werden muß und daß man die Information in kompakterer Form – und außerdem in der halben Zeit – erhält. Obwohl R.N. Bracewell dies nicht angibt, läßt sich – wenn man sich mit der Hartley Transformation etwas vertraut gemacht hat – auch eine 2D FHT finden, die sehr einfach durch 1D FHT's berechnet werden kann. (Es ist geplant diese Zusammenhänge und den Algorithmus in Kürze in einem weiteren Bericht oder einer Mitteilung zu beschreiben.)

6.1.2 Echobestimmung mit dem Cepstrumfilter

Es gibt im wesentlichen zwei Arten, die Echobestimmung mit dem Cepstrumfilter zu implementieren. Entweder man berechnet nach Formel (3.1) das Cepstrum oder man versucht die Impulsfolge, die sich am Ende ergibt, auf irgendeine Weise direkt zu ermitteln.

Impulsfolge. Beim Betrachten von Gleichung (3.8) könnte man auf die Idee kommen, einfach $Cepstrum\{s(x, y)\}$ von der Gleichung zu subtrahieren und damit nach der Impulsfolge aufzulösen und diesen Ausdruck zu berechnen. Diese an sich richtige Idee hat aber zwei entscheidende Nachteile. Erstens benötigt man 4 Fouriertransformationen, obwohl man mit der Hälfte auskommen kann, und zweitens haben die Experimente gezeigt, daß diese Form der Subtraktion schon bei geringfügig unterschiedlichen Teilsignalen keine guten Ergebnisse mehr liefert, weil die Leistungsspektren sich doch so unterscheiden, daß bei der Differenzbildung nicht allein die Impulsfolge übrig bleibt. Diese Variante wollen wir hier also nicht mehr betrachten.

D.J. Lee, S. Mitra und T.F. Krile [LMK90, S.202] bestimmen die Impulsfolge aus Gleichung (3.5). Dazu stellen sie die Formel nach dem Term um, der die Impulsfolge bildet, und erhalten

$$\begin{aligned} \mathcal{F}\{\log(2 \cdot (1 + \cos(x_0 u + y_0 v)))\} &= \mathcal{F}\{\log(\|F(u, v)\|^2)\} - \mathcal{F}\{\log(\|S(u, v)\|^2)\} \\ &= \mathcal{F}\left\{\log\left(\frac{\|F(u, v)\|^2}{\|S(u, v)\|^2}\right)\right\} = \mathcal{F}\left\{\log\left(\frac{\|L(u, v) + R(u, v)\|^2}{\|L(u, v)\|^2}\right)\right\} \quad (6.1) \end{aligned}$$

Das letzte Gleichheitszeichen folgt unter Benutzung der Bezeichnungen der Signale, wie in dem Absatz über die Phasenkorrelationsmethode (Seite 18). Die Autoren fügen noch eine additive Konstante -1 im Argument des Logarithmus ein, weil dies den Fehler in der Reihenentwicklung des Logarithmus verringert und das Auffinden des ersten Impulses erleichtert [LMK89, S.865].

Das stimmt zwar theoretisch, jedoch konnte ich am praktischen Verfahren keine Verbesserung feststellen. Im Gegenteil: Man hat sich nun zu überlegen, wie man den Logarithmus für Quotienten kleiner als eins sinnvoll berechnet. Die zweite rechen-technische Schwierigkeit ist, daß die Division nicht immer durchführbar ist (Division durch Null oder Werte um Null). Man muß also in diesen beiden Fällen durch eine Sonderbehandlung dafür sorgen, daß die normale Rechnung fortgeführt werden kann. Da die Autoren nicht angegeben haben, wie die Details konkret gelöst wurden, habe ich auf die zusätzliche Konstante verzichtet und das Ergebnis einer nicht durchführbaren Division durch eine große Zahl ersetzt. Mit dieser Abhilfe kommt man zu einem Verfahren mit nur noch drei Fouriertransformationen und verwertbaren Ergebnissen.

Cepstrum. Besser und unproblematischer ist der folgende Weg: Wie Y. Yeshurun und E.L. Schwartz [YS89] feststellen, ist das Maximum im Cepstrum so markant ausgeprägt, daß man nicht nach der Impulsfolge umstellen muß, sondern bei der

Implementation die Formel (3.1) direkt verwenden kann. Dann braucht man nur zwei Fouriertransformationen.

Aber auch hier gibt es zunächst ein rechentechnisches Problem, denn wird das diskrete Spektrum an einer Stelle identisch Null, kann der Logarithmus nicht mehr berechnet werden. Üblicherweise berechnet man daher $\log(x)$ als $\log(1+x)$ und ignoriert den Verfahrensfehler. Dieses Vorgehen kann man sich i.a. bei dem Cepstrumfilter nicht leisten, weil dadurch bei einem Signal mit geringer Amplitude kaum eine Wirkung des Logarithmus auf das Spektrum eintritt [GW83] und infolgedessen manchmal kein deutliches Echo mehr im Cepstrum erscheint.

In besserer Annäherung an $\log(x)$ habe ich also naheliegenderweise $\log(1+c \cdot x)$ verwendet. Bei sehr großem c berechnet man damit faktisch $\log(c)+\log(x)$, verändert also das diskrete Spektrum dieser Funktion, also das Cepstrum, bloß bei $(0,0)$ im Gleichanteil.

Damit nun die Güte der einzelnen Cepstrumfilterungen nicht bildsignalabhängig variiert, habe ich die Eingangssignale vor der Filterung auf einen großen Wertebereich normiert. Die Multiplikation des Eingangssignals mit einer großen Zahl α bedeutet $\alpha f(x) \Rightarrow \alpha \mathcal{F}\{f(x)\} = \alpha F(u) \Rightarrow \alpha^2 \|F(u)\|^2$. Wird also α entsprechend gewählt, ist die große Konstante c bereits im Eingangssignal enthalten. Man berechnet $\log(1+x)$ und erhält damit eine sehr gute Approximation an den tatsächlichen Logarithmus. Wenn nicht anders angegeben, ist in den beschriebenen Untersuchungen dieses „normierte“ Cepstrum verwendet worden.

6.2 Untersuchung des 2D Cepstrums

Die Untersuchung des 2D Cepstrums erfolgt experimentell und wird an wenigen Stellen durch analytische Betrachtungen ergänzt. Zunächst habe ich das Cepstrum auf künstliche Testsignale angewandt, um definierte Bedingungen für die Untersuchung zu schaffen. Die Koordinatenbezeichnungen beziehen sich auf Abbildung 4.1. Die mit **E1**, **E2**, ... durchnummerierten Experimente behandeln jeweils einen in sich abgeschlossenen Untersuchungsaspekt.

6.2.1 E1: Eingangsbeispiel

Als Eingangsbeispiel für die im folgenden benutzten Darstellungsweisen und um die Formeln des dritten Kapitels etwas zu beleben, betrachten wir ein Testsignal aus drei Quadraten mit unterschiedlichen Intensitäten.

Wie in Abbildung 4.1 bereits angedeutet und in Abbildung 6.1 zu sehen, werden die Doppelsignale vor der Cepstrumfilterung in ein größeres mit Null vorgefülltes diskretes Feld eingebettet (*null padding*), da sonst positive Disparitäten außerhalb des diskreten Cepstrums lägen. Nur zu Anschauungszwecken und weil sich quadratische Matrizen besser handhaben lassen ist in den folgenden Darstellungen der mit Null gefüllte Bereich noch etwas größer gewählt worden als für diesen Zweck notwendig.

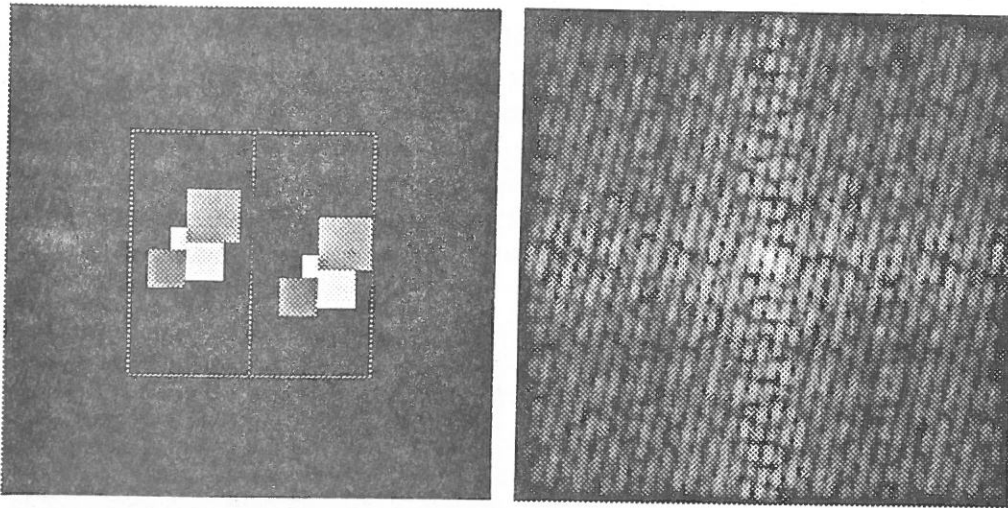
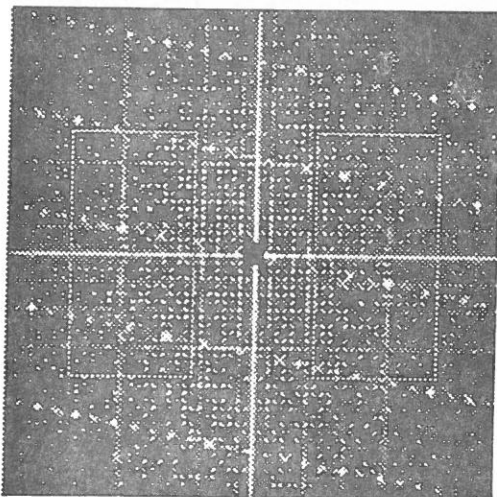


Abbildung 6.1: Links: Ein Beispielsignal aus drei Quadraten mit unterschiedlichen Intensitäten. Die Verschiebung beträgt $(x_0, y_0) = (3, 7)$. Die überlagerten Rechtecke markieren den Bereich der einzelnen Teilsignale. Rechts: Logarithmiertes Leistungsspektrum. Die Modulation durch den \cos -Terms ist deutlich zu erkennen. Unten links: Logarithmiertes Cepstrum. Ein Bereich um den Ursprung ist ausgespart (siehe Text). Die überlagerten Rechtecke markieren den Suchbereich nach Y. Yeshurun und E.L. Schwartz.



Da in den meisten Fällen die niedrigen Frequenzanteile so dominant vertreten sind, würde in einer Darstellung des reinen Leistungsspektrums i.a. außer einem hellen Punkt nichts weiter zu sehen sein. Deshalb werde ich in den Darstellungen immer das logarithmierte Leistungsspektrum zeigen, damit auch die schwächeren Strukturen sichtbar werden. Das logarithmierte Leistungsspektrum unseres Beispielsignals zeigt deutlich den mit Formel (3.4) vorhergesagten Cosinusanteil. Der Diracstoß im Ursprung des Cepstrums ist so dominant, daß er in einer Darstel-

lung des Cepstrums alle anderen Strukturen bei linearer Grauwertkennlinie unsichtbar macht. Damit wir nicht nur die beiden *peaks* sehen, die die Verschiebung kodieren, sondern auch die anderen Anteile des Cepstrums, habe ich in den Intensitätsdarstellungen das logarithmierte Cepstrum abgebildet und einen kleinen Bereich um den Ursprung, der uns nicht interessiert, ausgespart.

Die anderen regelmäßig verteilten Maxima ergeben sich durch die periodische Fortsetzung des an sich bandbegrenzten Spektrums bei der diskreten Berechnung des Cepstrums. Diese Maxima treten im wesentlichen nur bei idealen Testsignalen auf und verschwinden bei geringem Rauschen aufgrund ihrer kleinen Amplitude, so daß sie in praktischen Fällen nicht zu Fehlbestimmungen führen (vgl. Abbildung 6.2).

Wir sehen in Abbildung 6.2 außerdem, daß für Verschiebungen in der Nähe der tatsächlichen Verschiebung auch noch eine Signalantwort erkennbar ist. Stellen wir uns die Teilsignale in Gedanken übereinander verschoben vor, so herrscht bei Verschiebung Null perfekte Übereinstimmung vor; aber auch bei einer Verschiebung von einem Pixel nach rechts haben die Teilsignale noch in einem großen Teil des Rasters identische Werte, so daß auch hier eine Antwort des Cepstrums zu erwarten ist. Sie ist natürlich entsprechend geringer und wäre ohne die logarithmische Darstellung auch nicht auszumachen.

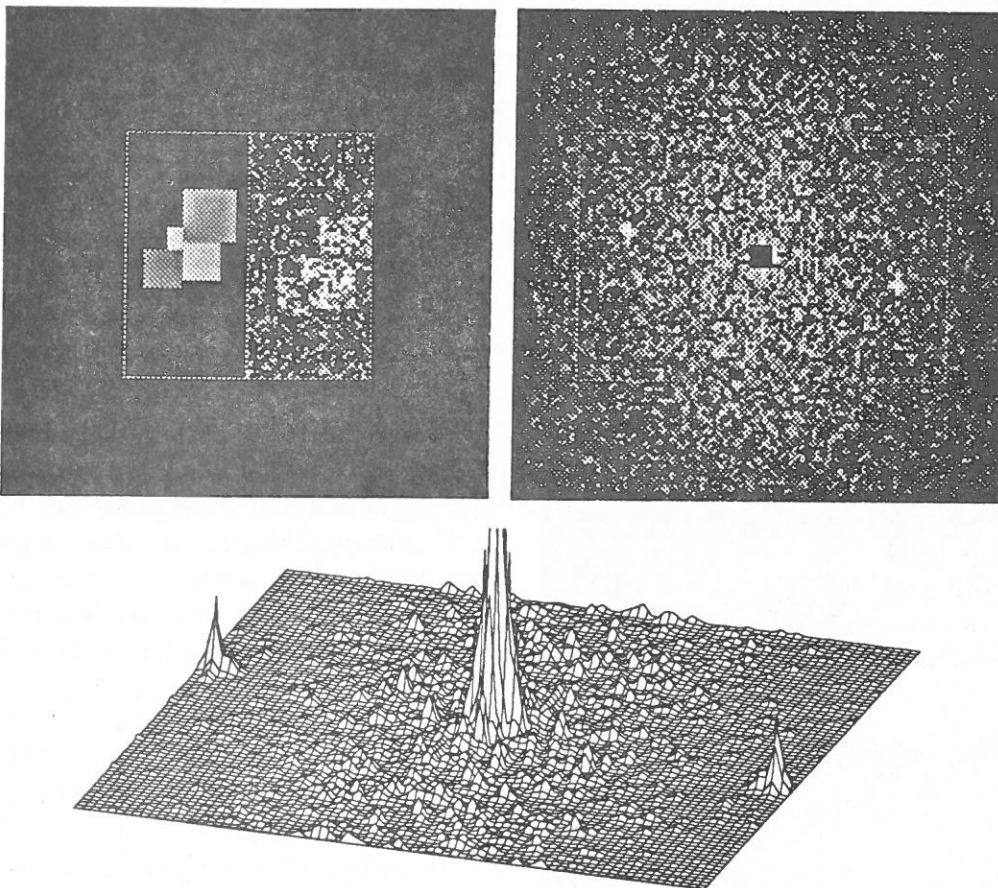


Abbildung 6.2: Links: Ein Beispielsignal. Das linke Teilsignal wurde mit weißem Rauschen (100% Signalamplitude) überlagert. Rechts: Cepstrum. Unten: Dreidimensionale Darstellung eines mittleren Bereiches des Cepstrums (nicht logarithmiert).

Zur Maximumdetektion schreiben Y. Yeshurun und E.L. Schwartz [YS89, S.761]:

„Then, as long as peak detection is restricted to a region of the cepstral plane within $[D/2, 3D/2]$, it is quite easy to locate this term as a “bright dot” or peak, with no competing areas of high intensity.“

Von einzelnen Ausnahmen abgesehen, die wir jetzt im folgenden kennenlernen werden, kann ich dieses Ergebnis durch eigene Experimente voll stützen.

6.2.2 E2: Auflösung

Berechnet man das diskrete Cepstrum für ein diskretes Signal, daß in einer groben Auflösung vorliegt, so ist auch das Cepstrum und damit die Position des Verschiebungsmaximums nur grob repräsentiert. Abbildung 6.3 zeigt ein Cepstrum in unterschiedlichen diskreten Auflösungsstufen. In den Darstellungen dieses Kapitels wird, wenn nicht anders angegeben, eine Kantenlänge von 128×128 Pixeln verwendet. Damit sind die Teilsignale 32×64 Pixel groß. Es entspricht auch den Erfahrungswerten aus der Literatur [OC90, YS90, LMK90], daß das Cepstrum mit weniger als 8 Pixeln Kantenlänge pro Teilsignal i.a. nicht mehr sinnvoll einzusetzen ist.

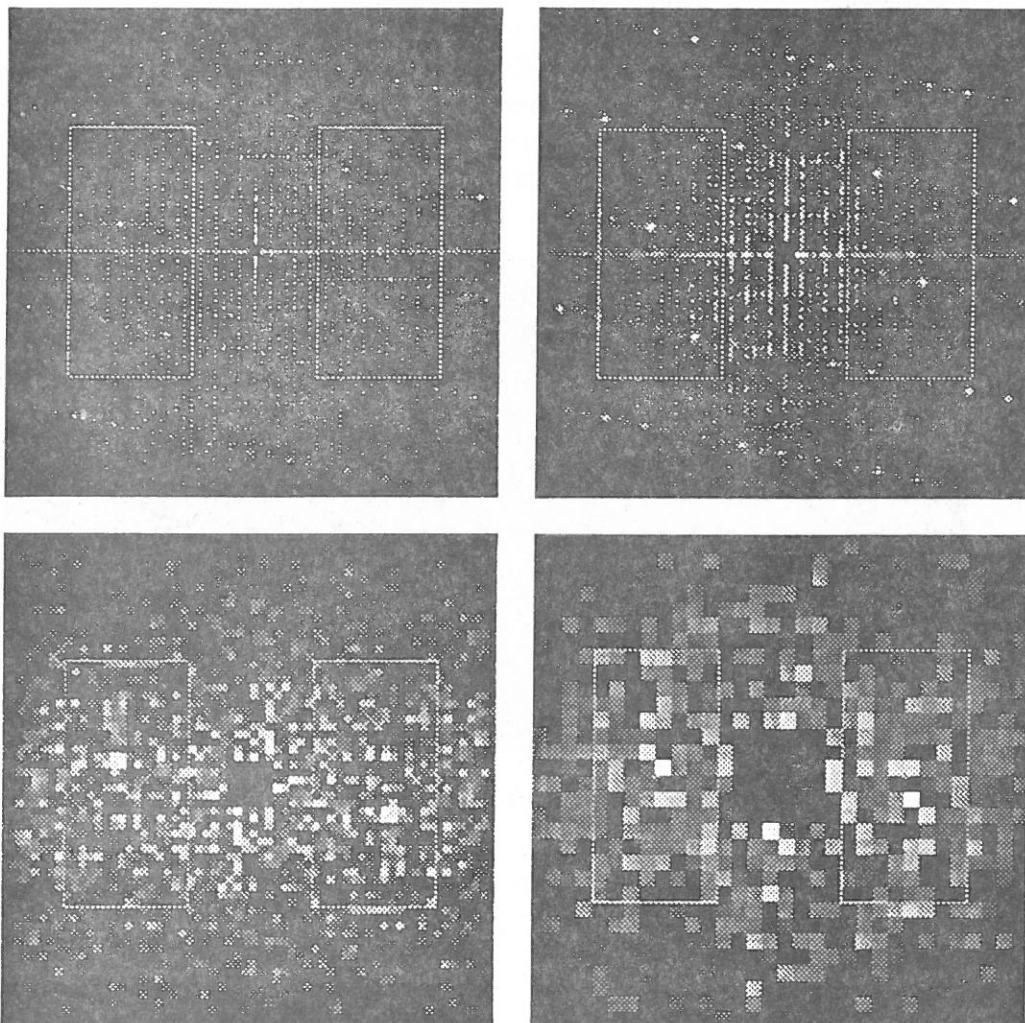


Abbildung 6.3: Diskretes Cepstrum eines Beispielsignals in unterschiedlichen Auflösungsstufen. Oben links: 256×256 , oben rechts 128×128 , unten links 64×64 und unten rechts 32×32 Bildelemente (Pixel). Wir sehen, daß ein Cepstrum mit einer Auflösung von 64×64 Pixeln gegenüber dem entsprechenden mit 256×256 Pixeln bereits deutlich an Aussagekraft verliert.

6.2.3 E3: random dot stereogram

B. Julesz hat gezeigt [Mar82], daß Menschen allein aufgrund der Verschiebung zweier Teilsignale Tiefe wahrnehmen können, auch wenn die Signale ansonsten keinerlei weitere monokulare Information enthalten. Dazu wurden zwei gleiche Bilder mit Zufallspunkten verwendet und in einem der Bilder ein Teilbereich verschoben und die entstehende Lücke wieder mit Zufallspunkten aufgefüllt. Obwohl beim Betrachten der Einzelbilder nichts zu erkennen ist, liegt bei stereoskopischer Betrachtung der beiden Bilder der verschobene Bildbereich in einer von der Größe der Verschiebung abhängigen anderen Tiefe als der Rest des Bildes.

Die Cepstrumtechnik kann auch für solche Zufallsdiagramme Verschiebung liefern. Im Gegensatz zu unseren bisherigen Testsignalen, bei denen das Maximum der Filterantwort (in der logarithmierten Darstellung) aus einem mehrere Pixel umfassenden „Hügel“ bestand, ergibt die Filterung hier eine punktuelle Antwort (ein Pixel), da eine Übereinstimmung der Teilsignale nur genau bei der zugrundeliegenden Verschiebung vorliegt.

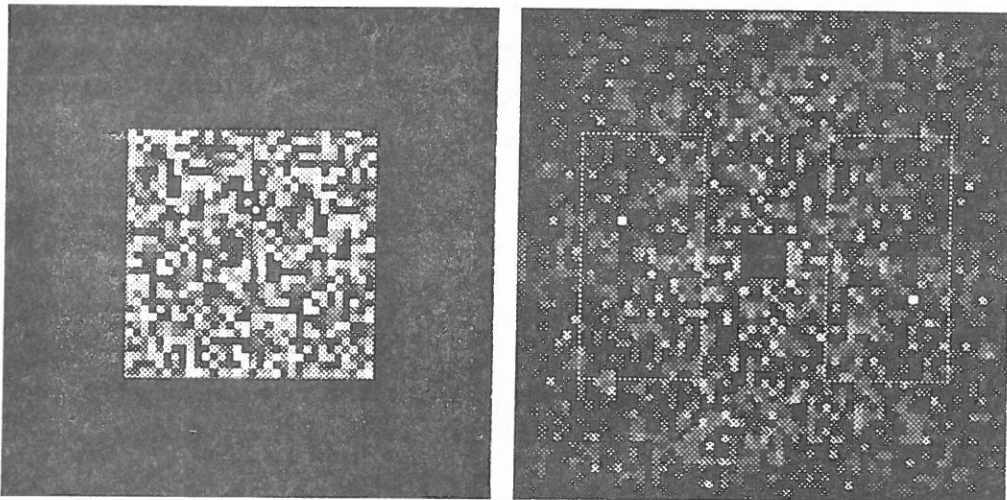


Abbildung 6.4: Links: Doppelsignal aus Zufallswerten. Das rechte Teilsignal ist gegenüber dem linken um $(x_0, y_0) = (3, 5)$ verschoben. Die entstehende Lücke wurde wieder mit zufälligen Werten aufgefüllt. Rechts: Cepstrum. Das Echomaximum ist hier so punktuell ausgeprägt (ein Pixel), daß es in der 128×128 Auflösung zwar auf dem Bildschirm erkennbar ist aber im Druck (auch wegen der zusätzlichen Reduzierung durch das Ditherverfahren) nicht mehr auszumachen ist. Dieses Beispiel ist deshalb mit einer Auflösung von 64×64 gerechnet.

6.2.4 E4: Inversion und unterschiedliche Helligkeiten

Y. Yeshurun und E.L. Schwartz berichten [YS89, S.763]:

„The algorithm [...] was not disturbed by this intensity difference, nor by simple additive intensity increments of 50 percent to one image of a stereo pair. In fact, positive and negative stereo pairs can be processed

with no difficulty, as is evident from the mathematical structure of the cepstral filter“

Dieser Aussage kann in dieser Allgemeinheit nicht zugestimmt werden. Zwar ist es theoretisch richtig, daß bei der ersten Fouriertransformation nur ein zusätzlicher Diracstoß bei $(0, 0)$ entsteht

$$\begin{aligned} f(x, y) &= s(x, y) + (1 - s(x - x_0, y - y_0)) \Rightarrow \\ F(u, v) &= \delta(u, v) + S(u, v) + S(u, v) \cdot e^{-i(x_0 u + y_0 v)} \end{aligned}$$

und deshalb das Cepstrum wiederum nur eine additive Konstante erhält, jedoch ergibt sich praktisch meistens ein weiteres lokales Maximum bei $(D, 0)$.²

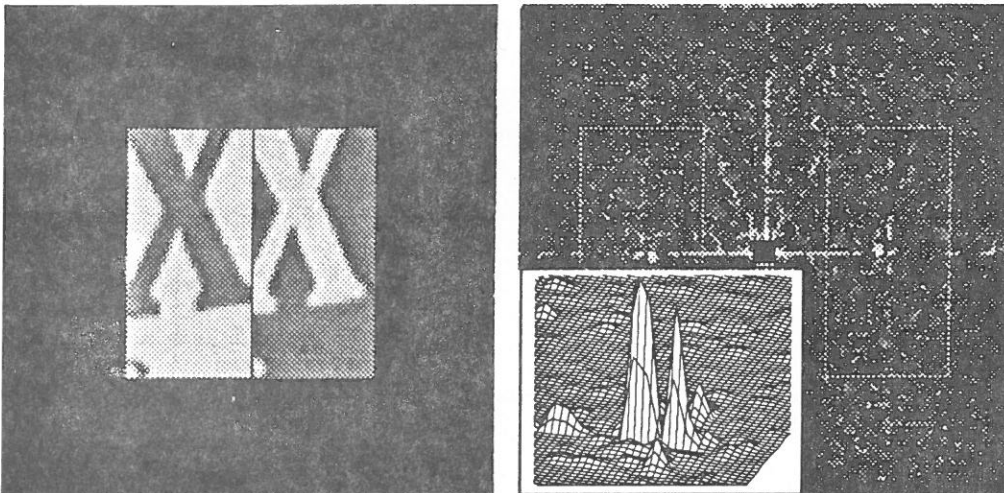


Abbildung 6.5: Cepstrum für ein Stereopaar mit Inversion. Links: Eingangssignal aus einem Kamerabild. Das rechte Teilsignal ist invertiert. Rechts: Cepstrum. Zur Verdeutlichung mit Ausschnittsvergrößerung (nicht-logarithmiert) für den Bereich um $(D, 0)$.

Dieses unerwünschte Maximum liegt an dem Ort, der Nulldisparität kodiert und kann die Erkennung des Maximums bei kleinen Verschiebungen unter ungünstigen Bedingungen stören.

Die Erklärung dieses Effektes ist relativ einfach und betrifft auch den Fall unterschiedlicher Helligkeiten. Denken wir uns vorläufig die Signale weg und betrachten nur noch die additive Helligkeit allein. Das linke Teilsignal sei also Null und das rechte habe einen konstanten Wert. Betrachten wir zur Vereinfachung einen Querschnitt entlang der x -Achse, so liegt als Signal ein Rechteckimpuls vor.

²Da das Cepstrum punktsymmetrisch um den Ursprung ist, spreche ich im folgenden zu einem Maximum im ersten oder vierten Quadranten das entsprechende korrespondierende Maximum im dritten oder zweiten Quadranten nicht mehr gesondert an. Da diese Symmetrie bei diskreter Berechnung streng genommen nicht gegeben ist, kann man bei der Berechnung des Maximums natürlich – gerade bei niedriger Auflösung oder dem Wunsch nach Subpixelgenauigkeit – beide Suchbereiche verwerten, um insgesamt eine genauere Lösung zu erhalten.

Cepstrum eines Rechtecksignals. Das Spektrum eines eindimensionalen Rechtecksignals der Breite D lautet mit $r_\alpha = -\frac{D}{2}$, $r_\beta = \frac{D}{2}$ und

$$f(x) = \begin{cases} 0 & : x < r_\alpha \\ 1 & : r_\alpha \leq x \leq r_\beta \\ 0 & : x > r_\beta \end{cases} \Rightarrow$$

$$F(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{r_\alpha}^{r_\beta} 1 \cdot e^{-iux} dx = \frac{e^{iur_\alpha} - e^{iur_\beta}}{\sqrt{2\pi}iu} \quad (6.2)$$

und Anwendung der Eulerschen Formel ($e^{i\varphi} = \cos \varphi + i \cdot \sin \varphi$):

$$F(u) = \sqrt{\frac{2}{\pi}} \cdot \frac{\sin(\frac{D}{2}u)}{u}$$

Durch die Betragsbildung wird die Frequenz verdoppelt und die dominante Frequenzkomponente dieses Signals erzeugt im Cepstrum jeweils ein Maximum bei $x = -D$ und $x = D$ (siehe Abbildung 6.6).

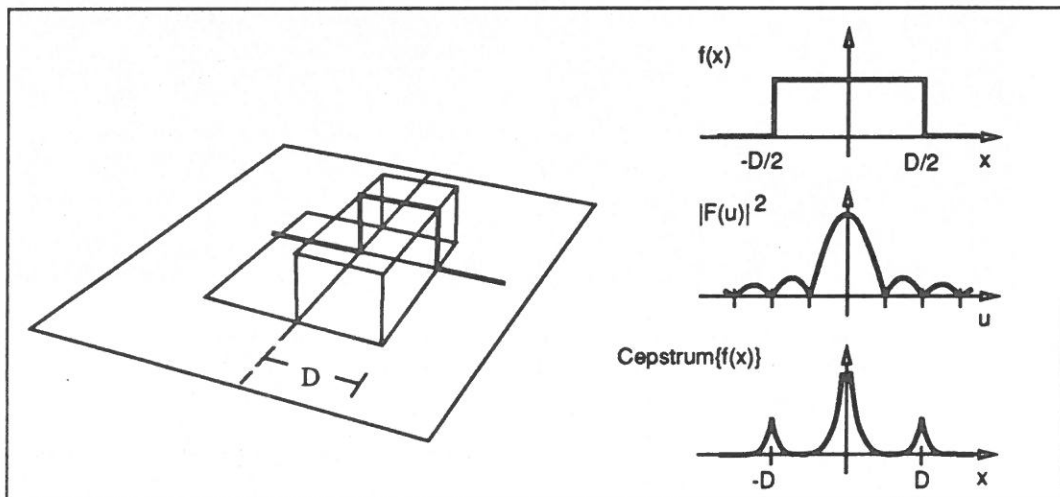


Abbildung 6.6: Links: Eingangssignal; Linkes Teilsignal Null, rechtes Teilsignal konstant. Mit Querschnitt bei $y = const$. Rechts: Skizze des Spektrums und Cepstrums für ein Rechtecksignal.

Besitzen also das Originalsignal und sein Negativ deutlich verschiedene Mittelwerte, prägt sich im Cepstrum – genau wie bei einer additiven Konstante – ein zusätzliches Maximum im Suchbereich aus. Es ist von der jeweiligen Bildfunktion abhängig, inwieweit sich das zusätzliche Maximum störend auswirkt. Man kann das Problem durch geeignete Normierung der Teilsignale umgehen. Im Falle einer Bandpaßfilterung entfällt das Problem natürlich auch, da der Gleichanteil der Signale eliminiert wird.

6.2.5 E5: Gerade Kantensegmente

Ein anderer möglicher Problemfall des bisherigen Verfahrens sind gerade Kantensegmente. Unter bestimmten Bedingungen enthält das Cepstrum in einem solchen Fall nicht nur das Maximum aufgrund der Verschiebung, sondern weitere lokale Maxima. Abbildung 6.7 zeigt ein Doppelsignal, bei dem die Teilsignale einen Abschnitt einer geraden Kante enthalten. Zur Erklärung der weiteren Maxima modellieren wir die

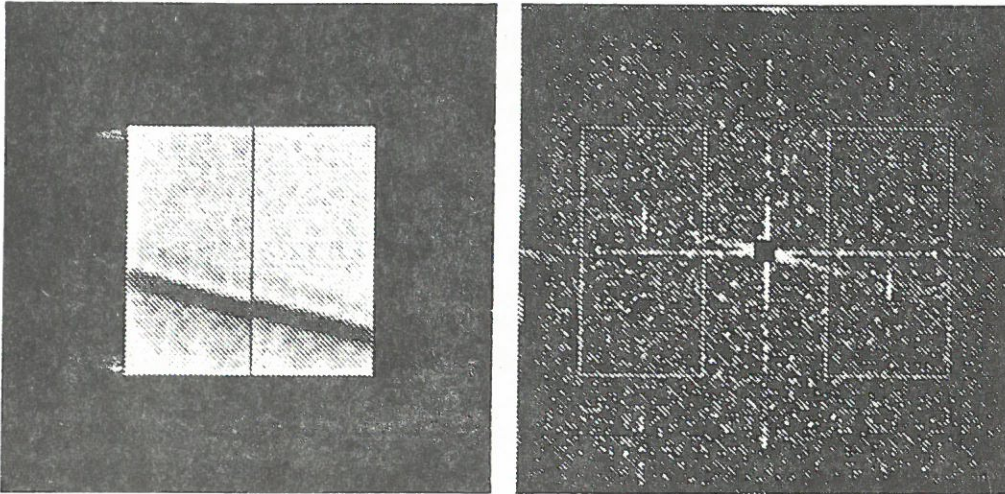


Abbildung 6.7: Cepstrum bei geraden Kantensegmenten. Links: Eingangssignal. Die Teilsignale enthalten jeweils ein gerades Kantensegment aus einer Kameraaufnahme eines Quaders. Rechts: Cepstrum. Im Suchbereich sind mehrere lokale Maxima.

geraden Kantensegmente als entsprechend geneigte Strecken (siehe Abbildung 6.8). Bei einer geeigneten Wahl des Koordinatensystems verlaufen die Strecken achsenparallel. Wir können dann das Doppelsignal $g(x, y)$ als Superposition zweier (eindimensionaler) Rechtecksignale der Breite a bzw. b auffassen. Zweidimensionale Fouriertransformation (vgl. Formel (3.2)) eines eindimensionalen Rechtecksignals (vgl. Formel (6.2) mit $r_\alpha = 0$ und $r_\beta = 1$) liefert

$g_1(x, y)$

1

$$G_1(u, v) = \frac{1 - e^{-iu}}{2\pi i u} \quad (6.3)$$

Wegen des Verschiebungssatzes und $\mathcal{F}\{f(\alpha x, \beta y)\} = \frac{1}{\|\alpha\beta\|} \cdot F\left(\frac{u}{\alpha}, \frac{v}{\beta}\right)$ (Skalierung) erhalten wir für die Fouriertransformierte des Doppelsignals mit den Bezeichnungen aus der Abbildung 6.9 und $a, b > 0$ insgesamt

$$G(u, v) = \frac{1 - e^{-iau}}{2\pi i u} + \frac{1 - e^{-ibu}}{2\pi i u} \cdot e^{-i(\tau_0 u + y_0 v)} \quad (6.4)$$

Das Leistungsspektrum ergibt sich nach einigen Umformungen^M zu:

$$\begin{aligned} \|G(u, v)\|^2 = & \frac{1}{2\pi^2 u^2} \cdot \left[2 - \cos(au) - \cos(bu) - \cos((a - x_0)u - y_0v) \right. \\ & - \cos((b + x_0)u - y_0v) + \cos((b + x_0 - a)u - y_0v) \\ & \left. + \cos(x_0u - y_0v) \right] \end{aligned} \quad (6.5)$$

Wir brauchen diese Terme jetzt nicht weiter zu transformieren um das Cepstrum zu ermitteln, da uns inzwischen klar ist, daß ein Cosinusanteil der Art $\cos(\alpha u + \beta v)$ im Leistungsspektrum zu einem Maximum im Cepstrum bei (α, β) führen muß.

Für das Beispiel mit der Quaderkante gilt $\boxed{b = a}$ und Formel (6.5) vereinfacht sich zu

$$\|G(u, v)\|^2 = \frac{1}{\pi^2 u^2} \cdot (1 - \cos(au))(1 + \cos(x_0u - y_0v)) \quad (6.6)$$

Wir sehen, daß außer dem Maximum bei (x_0, y_0) noch ein weiteres Maximum bei $(a, 0)$ zu erwarten ist. Wie man sofort sehen kann (vgl. Abbildung 6.8), liegt dieses Maximum im ungedrehten Koordinatensystem an einer Stelle mit $x = D$ – also i.a. recht zentral im Suchbereich –, falls der Betrag der Steigung des Kantensegments kleiner als 2 ist.

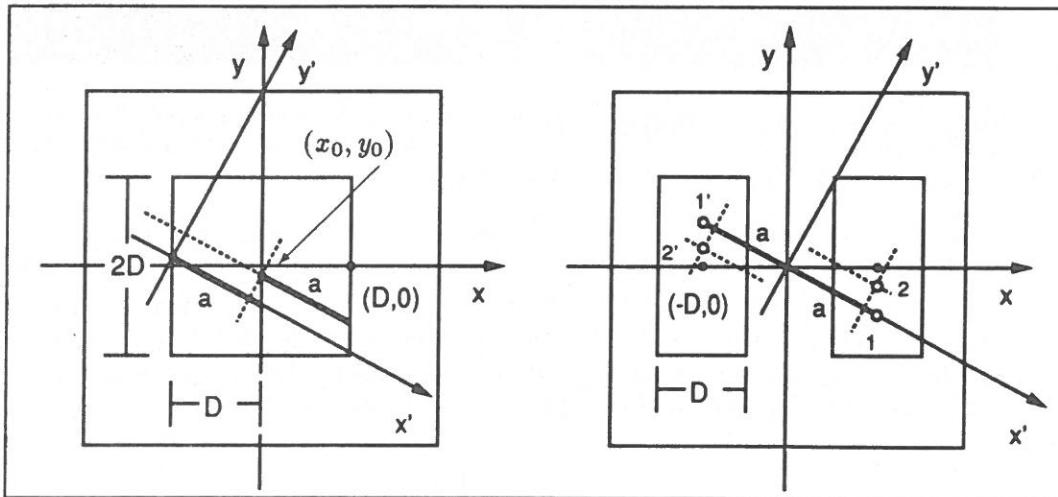


Abbildung 6.8: Gerade Kantensegmente als Teilsignal ($b = a$). Das gedrehte Koordinatensystem ist jeweils gestrichelt überlagert. Links: Skizze des Doppelsignals aus zwei geraden Kantensegmenten gleicher Länge. Rechts: Skizze der Maxima des Cepstrums.

Dieses Maximum kann man sich auch auf folgende Weise erklären: Nach Formel (3.8) setzt sich das Cepstrum eines Doppelsignals aus dem Cepstrum des Teilsignals und einer Impulsfolge zusammen. Das Teilsignal ist in diesem Fall ein (gedrehtes) eindimensionales Rechtecksignal, also kann man nach dem vorigen Abschnitt über das Cepstrum eines Rechtecksignals (vgl. Abbildung 6.6, rechts) erwarten, daß im Suchbereich ein Maximum an entsprechender Stelle auftritt.

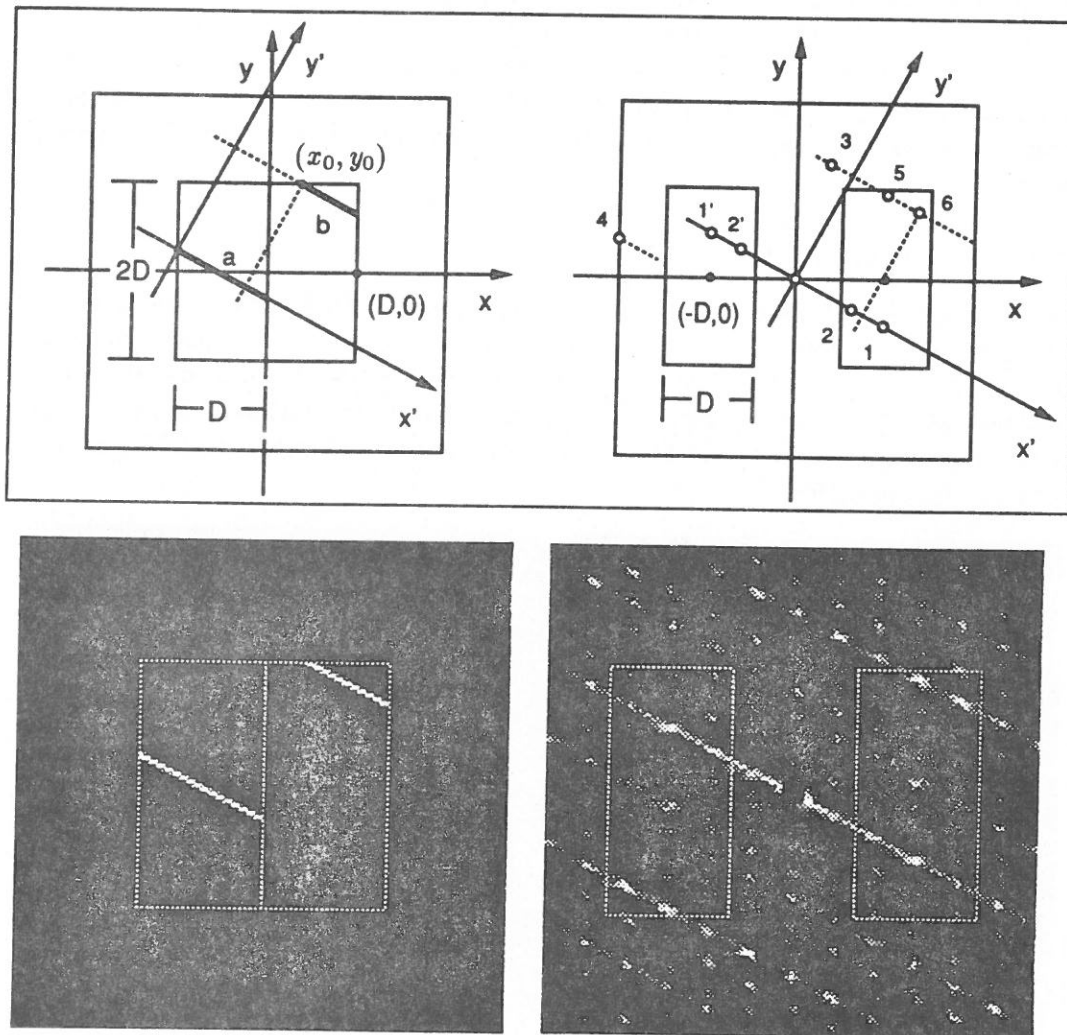


Abbildung 6.9: Gerade Kantensegmente als Teilsignale ($b \neq a$). Die Abbildung illustriert die Formel (6.5). Modellierung (oben) und praktische Rechnung (unten). Oben links: Doppelsignal aus eindimensionalen, parallelen Rechtecksignalen. Oben rechts: Skizze der zu erwartenden Maxima im Cepstrum nach Formel (6.5). Die Nummern in der Skizze bezeichnen den jeweiligen \cos -Term aus der Formel. Unten links: Diskretes Testsignal in Analogie zur gezeigten Modellsituation. Das diskrete Eingangssignal ist durch 13×13 Unterabtastung pro Pixel aus einer kontinuierlichen Funktion (homogene und 1 Pixel breite Gerade) errechnet. Unten rechts: Cepstrum. Es sind nicht alle Maxima gleich stark ausgeprägt. Die weiteren, schwächeren Maxima entstehen durch das *wrap-around* bei der diskreten Berechnung und die Symmetrie von $\cos(x)$.

Aus den schon erläuterten Gründen kann natürlich keine Verschiebung gemessen werden, wenn die beiden Kantensegmente nahtlos aneinander anschließen. Hier wird dann einfach nur das zweidimensionale Cepstrum eines (eindimensionalen) Rechtecksignals berechnet. Je nach Lage des Rechtecksignals liegt das (falsche) Maximum innerhalb oder auch außerhalb des Suchbereichs.

Dieses Experiment verdeutlicht außerdem, daß wir bei linearen Strukturen auf das bekannte Blendenproblem (*aperture problem*) treffen und – wie mit allen lokalen Verfahren – Verschiebung in solchen Fällen nur orthogonal zur Kante bestimmen können.

6.2.6 E6: Periodische Muster

Enthalten die Teilsignale ein periodisch wiederkehrendes Muster, kann mit einer lokalen Technik die wahre Verschiebung der Teilsignale nicht mehr ermittelt werden, sondern nur noch die Verschiebung modulo der Mustergröße ($V \bmod m$). Die Cepstrumtechnik liefert hier an der entsprechenden Stelle ein starkes Maximum. Es entstehen außerdem weitere schwache Maxima bei den Vielfachen von $V \bmod m$. (siehe Abbildung 6.10).

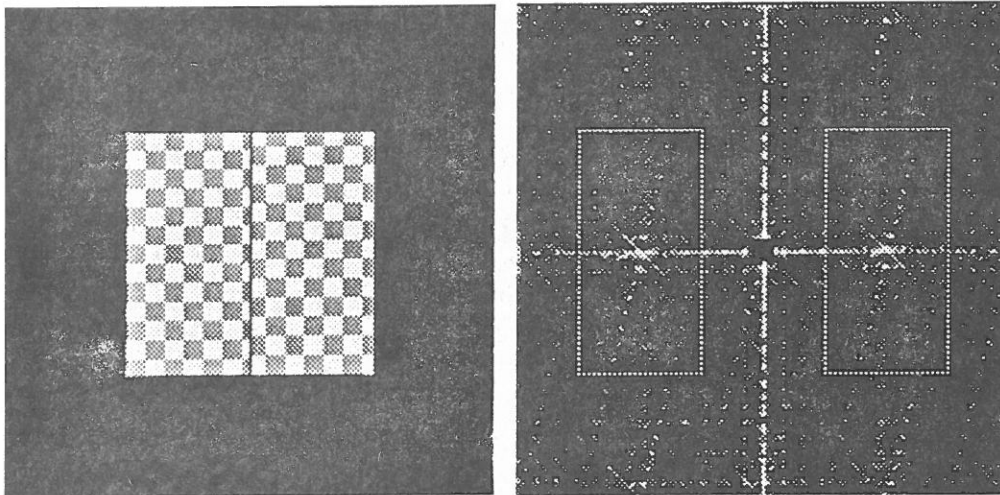


Abbildung 6.10: Schachbrettmuster (5×5 Pixel Quadrate) als Beispiel für ein periodisches Muster. Links: Eingangssignal mit $(x_0, y_0) = (-6, 9)$. Rechts: Cepstrum.

6.2.7 E7: Rotation (Einzelbeispiele)

Bisher haben wir ausschließlich Zusammensetzungen von Teilsignalen betrachtet, bei denen die Teilsignale nur verschoben waren. Im Hinblick auf die Verwendbarkeit der Cepstrumtechnik auch bei perspektivischen Verzerrungen sehen wir uns jetzt erst einige Einzelbeispiele an, bei denen ein Teilsignal rotiert oder skaliert wurde. Eine systematische Evaluierung wird dann in **E9** vorgenommen. Die Abbildungen in diesem Abschnitt sollen verdeutlichen, daß die Cepstrumtechnik bei Rotation eines

Teilsignale nicht immer erwartungsgemäß funktioniert. Abbildung 6.11 zeigt jedoch zunächst das Signal aus E1, ein positives Beispiel.

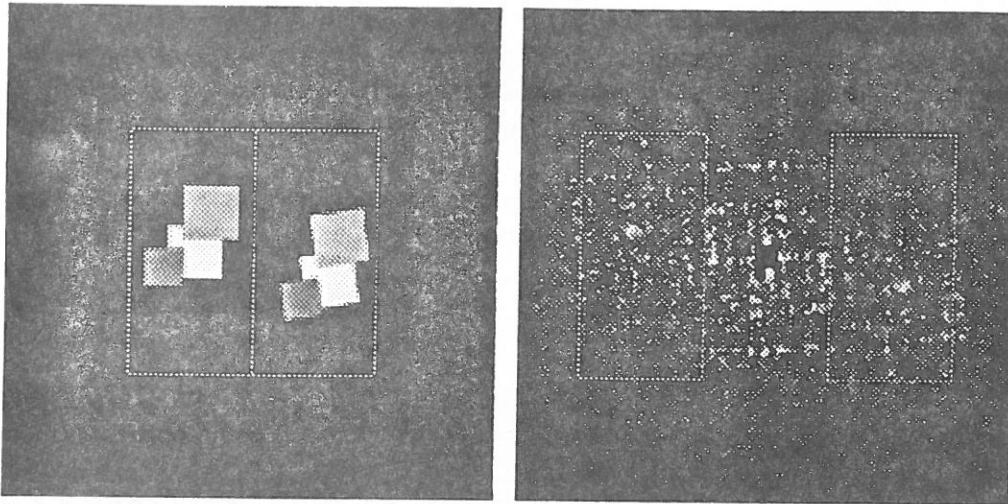


Abbildung 6.11: Cepstrum bei Rotation eines Teilsignals (1). Links: Eingangssignal. Die Teilsignale enthalten das Signal aus E1. Das rechte Teilsignal wurde um 8° gedreht. Ein Pixel des gedrehten Teilsignals wurde dabei mit 7×7 Unterabtastung aus dem linken Signal und arithmetischer Mittelung bestimmt. Rechts: Cepstrum. Trotz Rotation ist im Suchbereich ein erkennbares Echomaximum vorhanden.

Liegen dagegen Teilsignale ohne innere Struktur vor, so ergibt sich ein zunächst überraschendes Filterergebnis: Es entstehen mehrere Maxima.

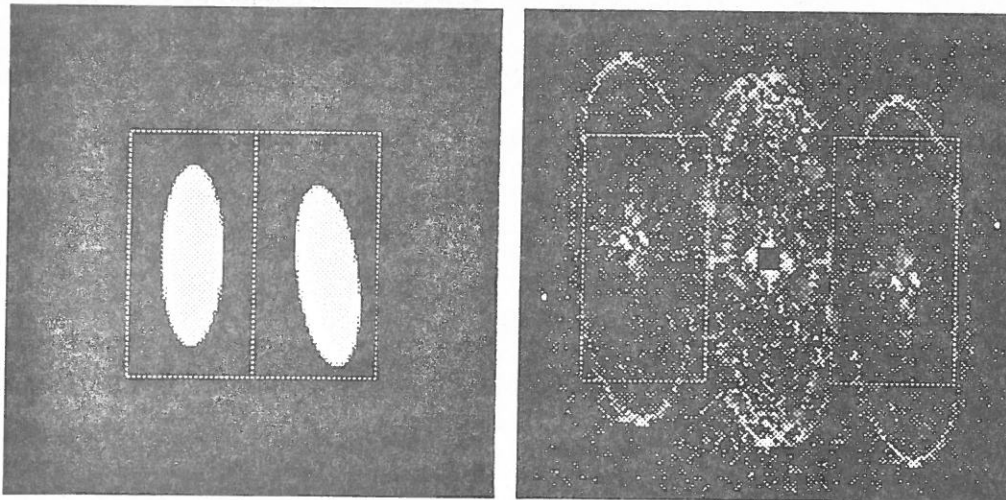


Abbildung 6.12: Cepstrum bei Rotation eines Teilsignals (2). Links: Eingangssignal. Die Teilsignale sind identisch Eins in einem elliptischen Bereich und außerhalb Null. Die Rotation ist $\varphi = 8^\circ$. Rechts: Cepstrum. Im Suchbereich befinden sich zwei Maxima.

Zur Erklärung mache man sich klar, daß die zur Echobildung beitragende Information nur aus dem Rand des Signals besteht. Verschieben wir die Teilsignale gedank-

lich übereinander, so ergibt sich die beste Übereinstimmung der Teilsignale dann, wenn ein möglichst großer Bereich des Randes zur Deckung kommt. Bei gedrehtem Teilsignal ist dies an einem Punkt, der nicht mit dem Mittelpunkt der Teilsignale übereinstimmt. Aus Symmetriegründen ergeben sich also zwei Maxima, die sich bezüglich der wahren Verschiebung diametral gegenüberstehen. Je größer der Rotationswinkel, desto weiter entfernen sich die beiden peaks von dem Punkt der wahren Verschiebung.

Dieser Effekt tritt auch bei komplizierteren Formen in entsprechend verwickelter Form auf. Ich habe dieses einfache Beispiel gewählt, um das zugrundeliegende Prinzip deutlich zu machen.

Dieser „Randeffekt“ wird abgemildert, wenn das Signal eine ausreichende Textur besitzt, wie an der folgenden Abbildung beispielhaft demonstriert wird.

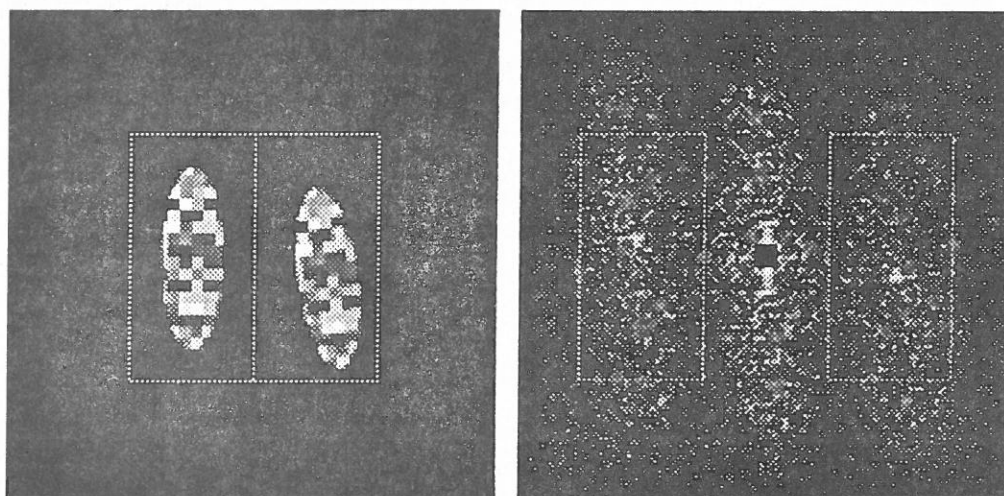


Abbildung 6.13: Cepstrum bei Rotation eines Teilsignals (3). Links: Eingangssignal. Die Teilsignale enthalten ein Zufallsmuster in einem elliptischen Bereich und sind außerhalb identisch Null ($\varphi = 8^\circ$). Rechts: Cepstrum. Im Suchbereich befindet sich ein abgeschwächter und verschmierter Echoanteil mit einem Maximum.

6.2.8 E8: Skalierung (Einzelbeispiele)

Ganz ähnliche Effekte wie bei der Rotation lassen sich bei der Vergrößerung oder Verkleinerung eines Teilsignals beobachten.

In diesem und dem nächsten Abschnitt wollen wir unter einer Skalierung s folgendes verstehen. Es soll z.B. $s = 4\%$ bedeuten, daß das linke Teilsignal 4% größer ist.

Für unser markantes Testsignal ergibt sich wieder ein eindeutiges und richtiges Echo-maximum, das allerdings – genauso wie im Falle der Rotation – gegenüber dem Fall mit reiner Verschiebung schon sichtbar abgeschwächt ist. Die Abbildungen 6.15 und 6.16 entsprechen den Abbildungen 6.12 und 6.13 des vorigen Abschnittes.

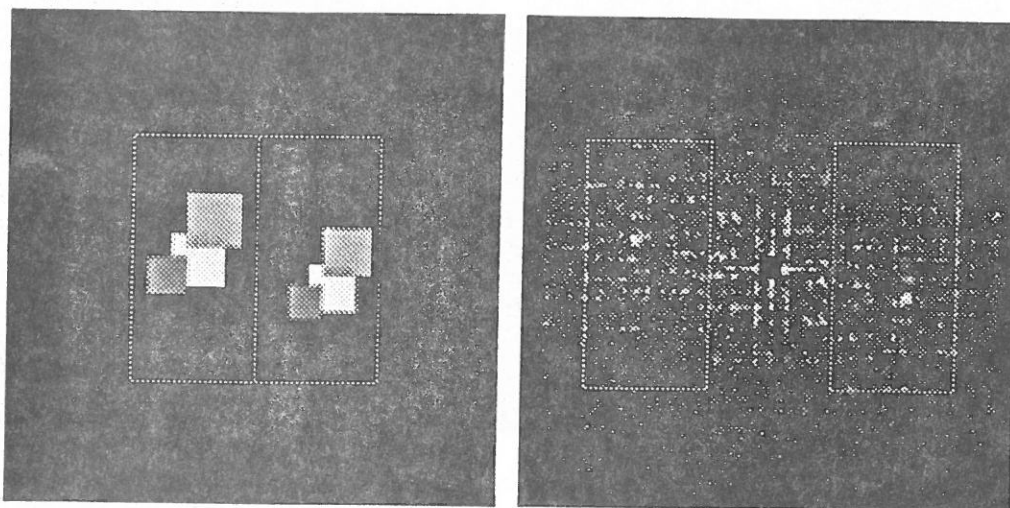


Abbildung 6.14: Cepstrum bei Skalierung eines Teilsignals (1). Links: Eingangssignal. Die Teilsignale enthalten das Signal aus E1. Das linke Teilsignal ist größer. Ein Pixel des skalierten Teilsignals wurde dabei wie bei den Beispielen unter E7 mit 7×7 Unterabtastung aus dem linken Signal und arithmetischer Mittelung bestimmt. Rechts: Cepstrum. Im Suchbereich ist trotz Skalierung ein erkennbares Echomaximum vorhanden.

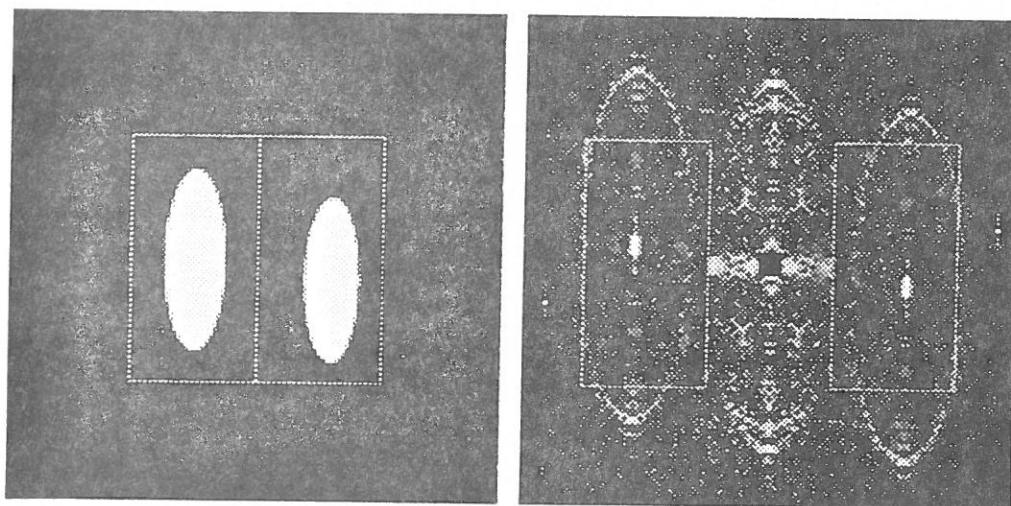


Abbildung 6.15: Cepstrum bei Skalierung eines Teilsignals (2). Links: Eingangssignal. Die Teilsignale sind identisch Eins in einem elliptischen Bereich und außerhalb Null. Mit Skalierung $s = 10\%$. Rechts: Cepstrum. Im Suchbereich befinden sich zwei Maxima an diametraler Position oberhalb und unterhalb des Punktes der wahren Verschiebung.

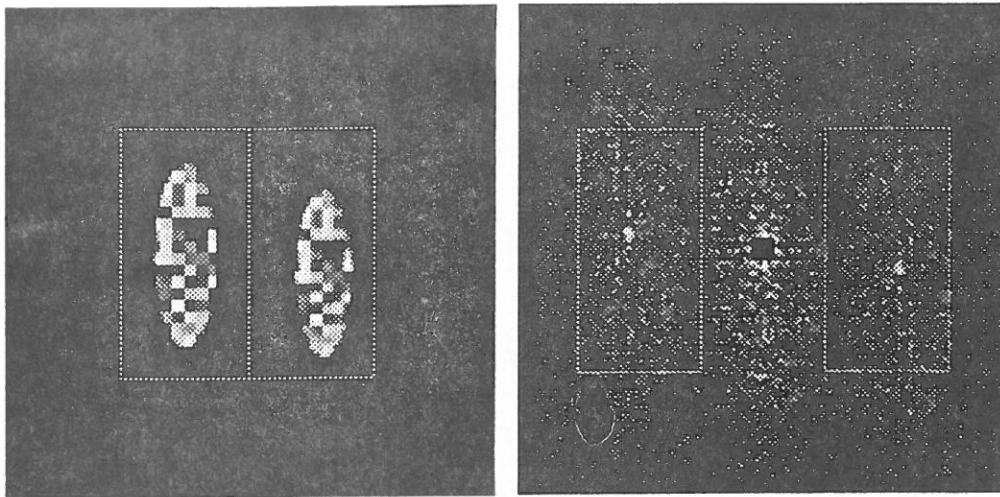


Abbildung 6.16: Cepstrum bei Skalierung eines Teilsignals (3). Links: Eingangssignal. Die Teilsignale enthalten ein Zufallsmuster in einem elliptischen Bereich und sind außerhalb identisch Null ($s = 10\%$). Rechts: Cepstrum. Im Suchbereich befindet sich wie im Falle der Rotation (Abbildung 6.13) ein abgeschwächter und verschmierter Echoanteil mit einem Maximum.

6.2.9 E9: Evaluierung für Rotation und Skalierung

Die im folgenden beschriebene Evaluierung hat gezeigt, daß die Verschiebungsbestimmung Rotation bzw. Skalierung eines Teilsignals um 3° bzw. 4% – ausreichende Textur vorausgesetzt – immer akzeptiert. Der Bereich wird bei Kombination von Rotation und Skalierung geringer.

6.2.9.1 Wahl der Testsignale

Bei der systematischen Untersuchung muß man sich entscheiden, welche Teilsignale benutzt werden sollen, um einerseits nur wenige Signale zu untersuchen und andererseits trotzdem damit möglichst allgemeine Aussagen zu erhalten. Um bei dieser Untersuchung die bereits vorgeführten störenden Effekte durch doppelte Maxima bei lokal begrenzten Signalen nicht mit der generellen Untersuchung der Resistenz gegen Rotation und Skalierung zu vermischen, habe ich mich gegen solche Signale entschieden.

Da *random dot stereograms* – wie in E3 gesehen – gewissermaßen Signale mit maximal punktueller Filterantwort sind, habe ich mich zur Evaluierung der Cepstrumfilterung für solche Zufallssignale entschieden.

Die Zufallstextur darf aber nicht zu grob und auch nicht zu fein gewählt werden: Wird sie zu grob gewählt (etwa ab 4 Pixel Kantenlänge für ein quadratisches Zufallselement), sind die Kanten einzelner Zufallsquadrate schon zu sehr eine eigene Struktur (vgl. auch E6), die auch bei geringfügiger Drehung nicht mehr gut genug korreliert. Wird die Textur dagegen zu fein gewählt, kann das korrespondierende „gedrehte“ Teilsignal nicht mehr genügend genau repräsentiert werden. Ich habe mich für 2×2 Zufallsquadrate entschieden.

Bei der Abwägung zwischen einer noch aussagekräftigen Auflösung einerseits und nicht zu hohem Berechnungsaufwand andererseits, habe ich mich bei diesem Experiment für eine Auflösung von 128×128 Pixeln entschieden. Damit ist das Doppelsignal mit 64×64 Pixeln repräsentiert. Die zu bestimmenden Disparitäten liegen dadurch im Bereich ± 16 Pixel.

6.2.9.2 Versuchsdurchführung

Unter Berücksichtigung der von Y. Yeshurun und E. L. Schwartz genannten Werte habe ich für einen bestimmten Versatz der Teilsignale $(x_0, y_0) = (5, 3)$ einzelne Vorversuche gemacht und den Untersuchungsbereich für Rotationswinkel (φ) und Skalierungsfaktoren (s) mit $(\varphi, s) \in [0^\circ, 8^\circ] \times [1.0, 1.15]$ festgelegt und mit $\Delta\varphi = \frac{1}{2}^\circ$ und $\Delta s = 0.1$ ausgewertet.

Für die Auswertung eines Punktes (φ, s) wurden aus statistischen Gründen 50 zufällige Einzelversuche herangezogen. Für ein einzelnes Doppelsignal wurde dabei jeweils das linke Signal zufällig erzeugt und eine gedrehte rechte Version berechnet. Ein Pixel des rechten Signals wurde dabei mit einer 4×4 Unterabtastung und arithmetischer Mittelung der 16 Werte gewonnen.

Die Auswertung an einer Stelle (φ, s) umfaßt mehrere Maßzahlen.

- Zur Bestimmung des Bereiches, in dem die Cepstrumtechnik noch zuverlässig eingesetzt werden kann, wurde der Prozentsatz exakter Disparitätsbestimmungen ermittelt.
- Um Aufschluß über die Art der Fehlbestimmungen zu erhalten, wurde ein Toleranzbereich von ± 5 Pixeln in x und y festgelegt. Nur Disparitäten, die innerhalb dieses Toleranzbereiches abgeliefert wurden, fanden bei der Berechnung von Mittelwert (μ) und Varianz³(σ^2) Berücksichtigung.
- Zur Ergänzung wurden in jedem Versuch die entsprechenden Werte für eine leicht geglättete Version⁴ des Cepstrums berechnet.

Diese (φ, s) -Evaluation wurde auch für die alternative Methode der Verschiebungsbestimmung über die Impulsfolge durchgeführt. Die Bilder auf den folgenden Seiten veranschaulichen die Ergebnisse des Versuchs. Die Abbildung 6.17 illustriert, daß

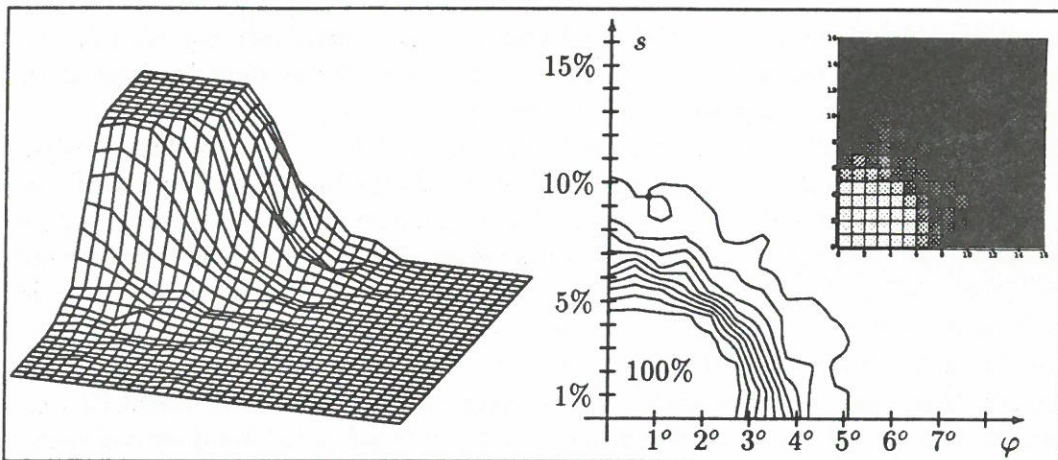


Abbildung 6.17: Evaluierung der Rotations- und Skalierungsinvarianz (1). Die Darstellungen zeigen den Anteil der exakten Disparitätsbestimmungen in Prozent. Links: Dreidimensionale Darstellung. Rechts: Höhenliniendarstellung. Die Höhenlinien liegen bei Vielfachen von 10%. Bemerkungen: (1) Zur besseren Darstellung hat die 3D Darstellung in diesem und in den folgenden Bildern ein feineres Raster als die ermittelten Datenpunkte. (2) Das Verfahren hat einen scharf begrenzten Bereich, in dem es exakt arbeitet. Außerhalb dieses Gebietes werden die Disparitätswerte schnell unzuverlässig.

das Verfahren für Rotationen bis 3° und Skalierungen bis 4% exakt arbeitet. Treten beide Störungen kombiniert auf, wird das Verfahren entsprechend schlechter. Der rechten Darstellung entnehmen wir, daß die Leistungsfähigkeit für $\sqrt{\varphi^2 + s^2}$

³Es wurde die für Stichproben übliche erwartungstreue $N - 1$ Statistik bei der Berechnung zugrunde gelegt.

⁴Es wurden zwei Glättungsschritte (8er Nachbarschaft) der Form

$$m_{xy} = \frac{1}{9} \left(8m_{xy} + \sum_{\substack{|x-i|=1 \\ |y-j|=1}} m_{ij} \right) \text{ durchgeführt.}$$

etwa konstant ist. Abbildung 6.18 zeigt einen Einzelfall aus der Evaluierung und gibt einen Eindruck von der Art der Cepstra im Grenzbereich. Die folgenden

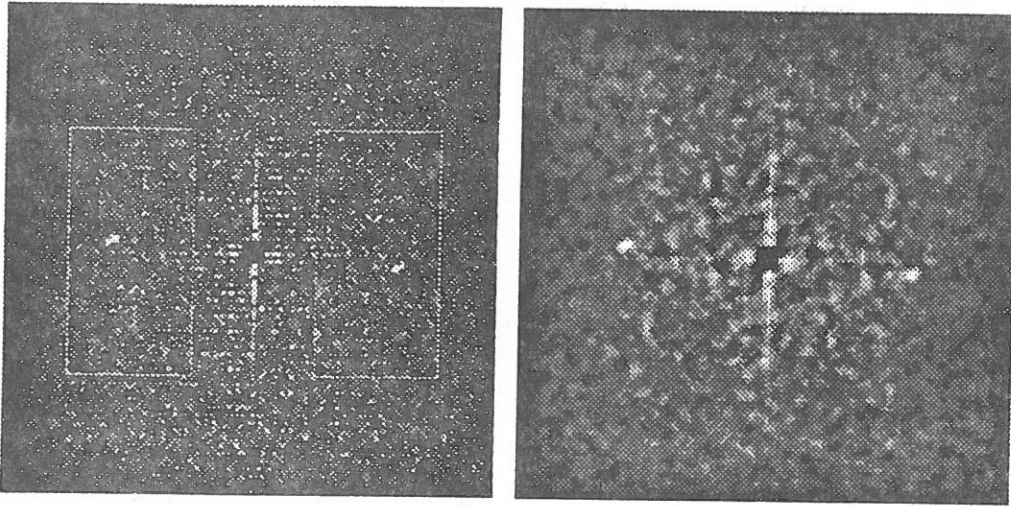


Abbildung 6.18: Cepstrum bei Rotation und Skalierung. Ein Einzelfall aus der Evaluierung mit $(\varphi, s) = (3^\circ, 3\%)$. Links: Cepstrum. Im Suchbereich ist das Maximum nicht mehr so stark (vgl. Abbildung 6.4) und auch leicht verwischt. Rechts: Geglättetes Cepstrum. Das Maximum erscheint deutlicher. Es stellt sich hier aber die Frage, ob die Koordinaten des Maximums noch gut genug die tatsächliche Verschiebung kodieren.

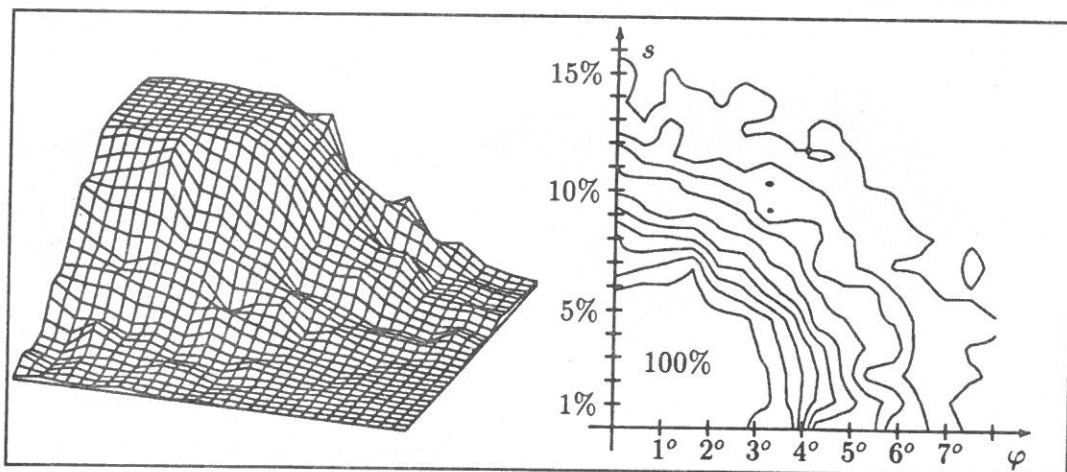


Abbildung 6.19: Evaluierung der Rotations- und Skalierungsinvarianz (2). Die Darstellungen zeigen den Anteil der exakten Disparitätsbestimmungen in Prozent für ein geglättetes Cepstrum. Links: Dreidimensionale Darstellung. Rechts: Höhenliniendarstellung. Die Höhenlinien liegen bei Vielfachen von 10%.

Abbildungen veranschaulichen einige der gewonnenen Daten bei Verwendung des zuvor beschriebenen geglätteten Cepstrums. Im Vergleich mit Abbildung 6.17 fällt auf, daß sich die Gesamtzahl exakter Disparitätsbestimmungen deutlich erhöht hat. Das Verfahren besitzt aber nun keinen scharf begrenzten Vertrauensbereich mehr;

die Disparitätswerte werden langsam immer unzuverlässiger. Die Abbildung 6.20 zeigt den Prozentsatz der Disparitätsbestimmungen, die noch innerhalb des für die Bestimmung von μ und σ^2 gewählten Toleranzbereiches liegen.

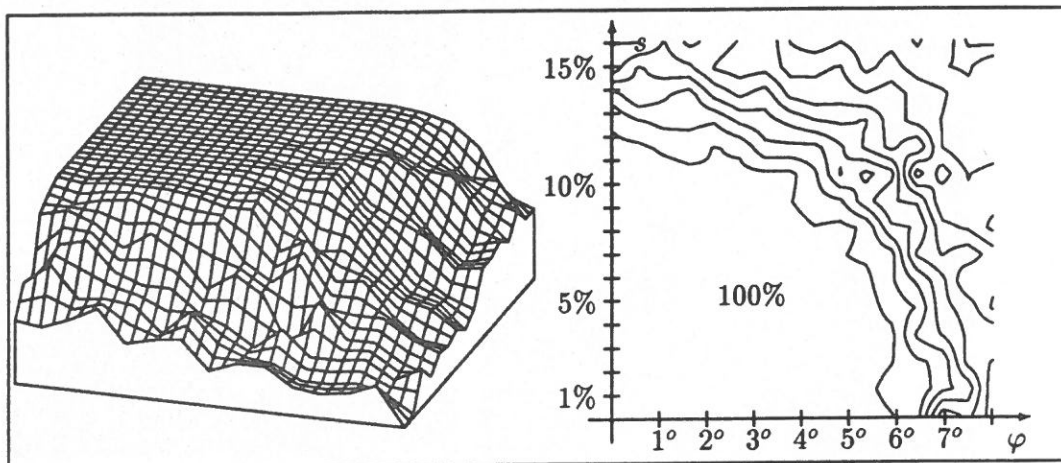


Abbildung 6.20: Evaluierung der Rotations- und Skalierungsinvarianz (3). Die Darstellungen zeigen den Anteil der Disparitätsbestimmungen in Prozent für ein geglättetes Cepstrum, die in den Toleranzbereich fallen (siehe Text). Links: Dreidimensionale Darstellung. Rechts: Höhenliniendarstellung. Die Höhenlinien liegen bei Vielfachen von 10%.

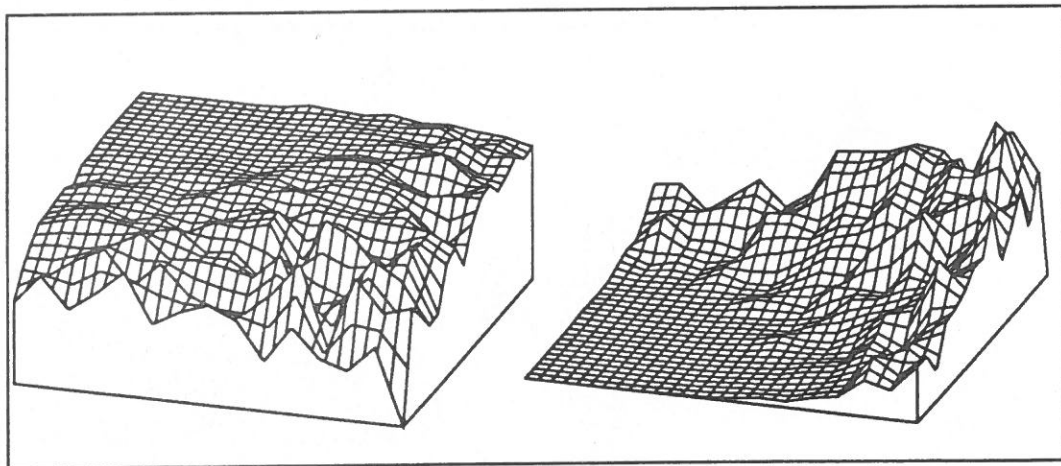


Abbildung 6.21: Evaluierung der Rotations- und Skalierungsinvarianz (4). Geglättetes Cepstrum. Die Darstellungen zeigen Mittelwert und Varianz für die Disparität s_x innerhalb des Toleranzbereiches. Die Daten für die Varianz legen nahe, daß die Fehlbestimmungen keine „Ausreißer“ sind, sondern durch die Glättung ein bis zwei Pixel in die Umgebung gewandert sind. Das läßt sich auch durch stichprobenartige Einzelversuche belegen. Die Daten für s_y und s_z unterscheiden sich nicht wesentlich. Links: Dreidimensionale Darstellung für μ ($\mu \in [0, 5.5]$). Rechts: Dreidimensionale Darstellung für σ^2 ($\sigma^2 \in [0, 13.5]$); Das Koordinatensystem wurde nicht – wie bei den anderen 3D Darstellungen – gedreht).

Wird die Verschiebungsbestimmung nicht über das Cepstrum, sondern über die Impulsfolge (siehe Abschnitt 6.1.2, Seite 53) durchgeführt, ergeben sich keine signifikanten Unterschiede, wie Abbildung 6.22 beispielhaft zeigt. Da die Verschiebungsbe-

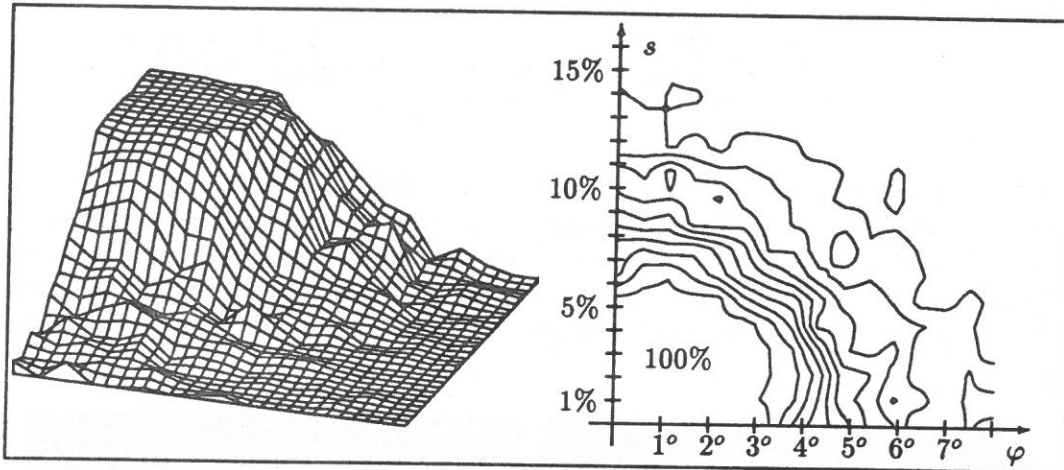


Abbildung 6.22: Evaluierung der Rotations- und Skalierungsinvarianz (5). Die Darstellungen zeigen den Anteil der exakten Disparitätsbestimmungen in Prozent für eine geglättete Version der Impulsfolge. Links: Dreidimensionale Darstellung. Rechts: Höhenliniendarstellung. Die Höhenlinien liegen bei Vielfachen von 10%.

stimmung bei Verwendung des Cepstrums bei gleicher Leistungsfähigkeit mit einer Fouriertransformation weniger auskommt, ist dies die günstigere Methode.

Schlußbemerkungen. Wie E3 deutlich gemacht hat, tritt bei Zufallsdiagrammen nur genau an den Koordinaten der tatsächlichen Verschiebung ein markantes Maximum auf und in der Umgebung nicht. Bei einem Testsignal, das bereits bei reiner Verschiebung kein so auf einen Pixel konzentriertes Maximum liefert, muß davon ausgegangen werden, daß die spezielle Struktur und Ausdehnung des Echos die Evaluierung entsprechend beeinflusst.

Es muß hier noch einmal deutlich darauf hingewiesen werden, daß das verwendete Testsignal mit Absicht als Zufallsdiagramm ausgewählt wurde, um sich diese punktuelle Struktur zunutze zu machen. Einerseits natürlich, um bei der Evaluierung ein möglichst bildunabhängiges Resultat zu erhalten. Andererseits aber auch, um nicht mit einem zu ausgedehnten „Maximumshügel“ ein zu gutes Resultat zu erhalten. Insofern konnte mit diesem Testsignal und seiner punktuellen Echostruktur in der vorgestellten Evaluierung die untere Grenze der Verwendungsfähigkeit des Cepstrums zur Verschiebungsbestimmung abgesichert werden.

Wie wir beispielhaft an Abbildung 6.11 gesehen haben und wie eine Reihe weiterer, hier nicht beschriebener Experimente gezeigt hat, schließt dieses Ergebnis nicht aus, daß bei entsprechend günstiger Struktur – es liege also keiner der bereits diskutierten problematischen Fälle vor – bis zu einem Faktor drei bessere Werte möglich sind. Dies ist jedoch bei realen Bilddaten nicht der Regelfall, hier liegen die Grenzwerte nach meiner Erfahrung mit etwa 6° und 6% zwischen diesen beiden Extremen.

6.2.10 E10: Rauschinsensivität des Cepstrums

Es werden üblicherweise folgende Definitionen für das Signal-Rausch-Verhältnis (*signal-to-noise ratio*) verwandt

$$(\text{SNR})_{rms} = \sqrt{\frac{\sum f_{xy}^2}{\sum r_{xy}^2}} \quad (\text{SNR})_p = \sqrt{\frac{(\max(f_{xy}) - \min(f_{xy}))^2}{\sum r_{xy}^2}}$$

wobei f das Bildsignal und r das Störsignal bezeichne.

Eine systematische Evaluierung der Cepstrumfilterung unter weißem oder auch gaußischem Rauschen ist nicht so einfach durchzuführen, da das Verhalten für verrauschte Signale wie bei unverrauschten Signalen stark von dem jeweiligen Signal selbst abhängig ist. Es läßt sich daher nur schwer ein allgemeingültiger SNR-Wert angeben.

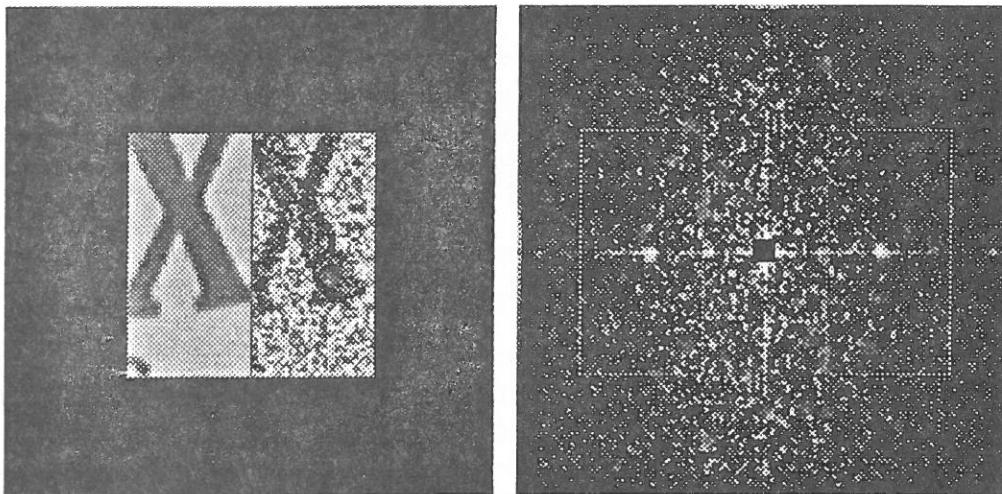


Abbildung 6.23: Demonstration der experimentell ermittelten Rauschinsensivität. Links: Teilsignal aus einem Kamerabild. Das rechte Teilsignal wurde mit weißem Rauschen (100% Signalamplitude, $\text{SNR}_{rms} = 2$) überlagert. Rechts: Cepstrum.

Das Echomaximum in dem in **E1** vorgestellten Beispiel (Abbildung 6.2) kann unter weißem Rauschen mit bis zu 170% Signalamplitude ($\text{SNR}_{rms} \approx 0.7$, $\text{SNR}_p \approx 14$) gefunden werden.

Eine Untersuchung des Zufallsmusters aus **E9** ergab dagegen ein ausgeprägtes Maximum bei bis zu 300% Signalamplitude ($\text{SNR}_{rms} \approx 0.6$, $\text{SNR}_p \approx 7$). Dieser Wert ist allerdings für typische Bilddaten viel zu hoch.

Zusammenfassend ergibt sich aus meinen Untersuchungen, daß der SNR-Wert nur einen groben Anhaltspunkt dafür geben kann, ob die Verschiebungsbestimmung mit dem Cepstrum versagen wird.

Genauere qualitative Aussagen kann man unter speziellen einschränkenden Bedingungen gewinnen. So hat die von J.C. Hassab und R. Boucher durchgeführte ma-

thematische Analyse für spezielle Spektren unter stationärem gaußischem Rauschen folgendes ergeben [HB76, S.444,453]:

„Thus the total signal power to noise power (S/N) alone is an insufficient measure for setting the cepstrum performance [...] The mean reduction in peak at τ is evaluated as a function of input total S/N and relative signal to noise bandwidths; it is shown that a decrease in S/N could be compensated for by an increase in relative bandwidth so as to maintain a constant reduction in peak at τ .“

Ich habe in einer Reihe von eigenen Untersuchungen festgestellt, daß für Kameradaten – die ja bereits mit gewissem realen Rauschen behaftet sind – zusätzliches weißes Rauschen eines Teilsignals bis 100% Signalamplitude ($\text{SNR}_{rms} \approx 2, \text{SNR}_p \approx 15$) die Maximumdetektion i.a. noch erfolgreich durchgeführt werden kann⁵ (vgl. Abbildung 6.2 und 6.23). Dieser Wert ist natürlich unter dem im vorigen Zitat zu Beginn geäußerten Vorbehalt zu verstehen.

6.2.11 Zusammenfassung des bisherigen Verfahrens

Ich will an dieser Stelle noch einmal darauf hinweisen, daß ich in den Grauwertdarstellungen die störenden Einflüsse in den dargelegten Problemfällen durch die Verwendung des logarithmierten Cepstrums überhöht habe, um die zugrunde liegenden Ursachen besser zeigen zu können.

Das Cepstrum enthält in den meisten Fällen nur eine zentrale Komponente und zwei starke Maxima, wie in der dreidimensionalen Darstellung in Abbildung 6.2 gezeigt. Bei einem Problemfall, der zu einer Fehlbestimmung führt, sieht das Cepstrum meistens nicht anders aus; die beiden Maxima liegen nur an der falschen Stelle und haben, was man ohne eine überhöhte Darstellung zunächst auch nicht sehen kann, eine andere Ursache.

Da mich diese „unmotivierten“ Fehlbestimmungen zu Beginn meiner Arbeit einigermaßen irritiert haben, bin ich dann zu einer Darstellung des logarithmierten Cepstrums übergegangen, bei der man bei kleiner Veränderung des Eingangssignals auch die entsprechenden Veränderungen im Filterergebnis beobachten kann. Bei einer Berechnung der Verschiebung über die Impulsfolge ist dieser eher binäre Übergang zur Fehlbestimmung noch ausgeprägter und (zunächst einmal) verwunderlicher.

Die Verschiebungsbestimmung mit dem Cepstrum kann – wie wir an einer Reihe von Beispielen gesehen haben – in gewissen Fällen problematisch sein. Davon abgesehen zeigt sie sich aber robust, auch in einer Reihe weiterer, hier nicht einzeln diskutierter Fälle (z.B. Glättung oder Histogrammtransformation eines Teilsignals).

Das Verfahren ist recht rauschinsensitiv ($\text{SNR}_{rms} \approx 2, \text{SNR}_p \approx 15$).

Die in der Literatur angegebenen Werte der Resistenz gegen Rotation und Skalierung sind nicht falsch, aber für den durchschnittlichen Fall um den Faktor 2 zu hoch. (vgl. Seite 20 und Seite 73).

⁵Umgerechnet auf den SNR Wert von D.J. Lee, S. Mitra und T.F. Krile (vgl. Seite 21), kann ich den Wert von etwa 0.88 bestätigen.

Insgesamt kann ich die Beurteilung der Cepstrumtechnik von Y. Yeshurun und E. L. Schwartz [YS89] stützen, obwohl sie, wie wir gesehen haben, teilweise zu positiv ausfällt und Problemfälle nicht erwähnt (Gerade Kantensegmente) oder anders darstellt (z.B. Inversion). Im folgenden Kapitel werden einige Möglichkeiten gezeigt, das bisherige Verfahren zu erweitern oder zu verbessern.

Kapitel 7

Verbesserungen und Erweiterungen des 2D Cepstrums

In diesem Kapitel wollen wir sehen, wie die im vorigen Kapitel besprochenen Problemfälle gelöst werden können. Dazu werden wir uns mit einigen Verbesserungen und Erweiterungen der Cepstrumtechnik beschäftigen.

Der Suchbereich zur Detektion des Echomaximums kann verkleinert werden. Des Weiteren kann das Problem bei unterschiedlichen mittleren Helligkeiten der Teilsignale durch eine einfache Maßnahme bei der Teilbildextraktion gelöst werden. Der Fall gerader Kantensegmente und bestimmter ungünstiger Zusammenstellungen der Teilsignale kann durch die technische Realisierung einer funktionalen Deutung der *pale bands* (vgl. Kapitel 2) gelöst werden.

Darauf aufbauend werden wir andere und i.a. günstigere als bisher in der Literatur angegebene Trägerfunktionen zur Teilsignalextraktion kennenlernen. Es wird außerdem demonstriert, daß eine Bandpaßfilterung der Rohbilder vorteilhaft ist. Schließlich werden wir auf einer Formel von T.J. Olson und D.J. Coombs aufbauend eine andere Sicht auf den Cepstrumfilter kennenlernen.

7.1 Suchbereich

Der von Y. Yeshurun und E.L. Schwartz [YS89] angegebene Suchbereich kann noch weiter eingeschränkt werden. Je größer die Verschiebung zwischen den Signalen ist, desto geringer ist der Flächenanteil des Signals, der in beiden Teilsignalen gemeinsam enthalten ist, und desto schwächer ist auch das Echo.

Die Linien konstanten Überlappungsgrades (oder einer gewissen Echomindeststärke) definieren Rauten um die Punkte $(-D, 0)$ und $(D, 0)$. Für alle Verschiebungen, die innerhalb der Raute mit den Eckpunkten $(D/2, 0)$, (D, D) , $(3D/2, 0)$ und $(D, -D)$ liegen, gilt, daß der gemeinsame Signalanteil mindestens 50% beträgt.

Andersherum ergibt sich daraus bei gegebenem Wertebereich für die zu bestimmenden Verschiebungen und einem gewünschten Überlappungsgrad – oder einer gewünschten Mindeststärke des Echomaximums – die zu wählende Streifenbreite D .

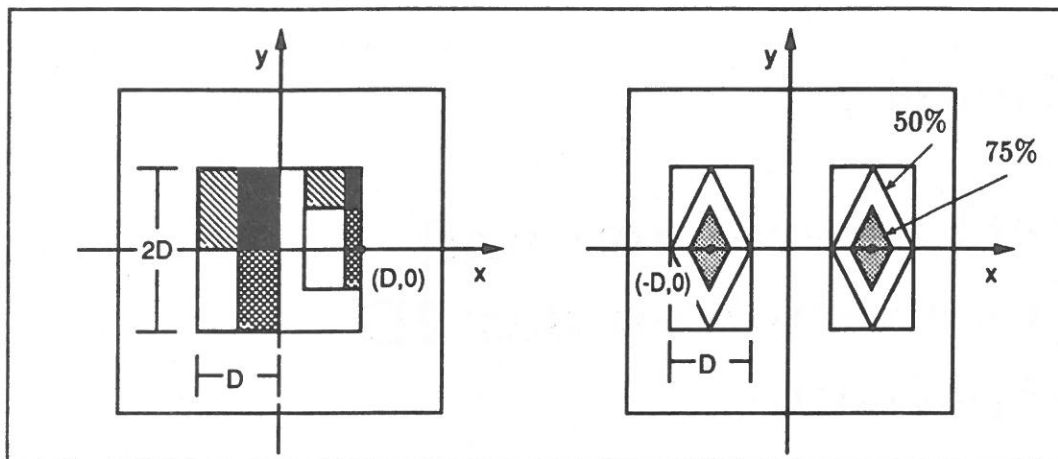


Abbildung 7.1: Eingeschränkter Suchbereich durch Vorgabe eines Mindestprozentsatzes für die gemeinsamen Signalanteil zwischen den Teilsignalen.

Außerdem braucht man im letzten Schritt der der Berechnung des Cepstrums natürlich nur noch die Spalten zu berücksichtigen, die in den Suchbereich fallen. Es entfallen also $\frac{3}{4}$ oder mehr der abschließenden Spaltentransformationen (vgl. [OC90]).

7.2 Unterschiedliche Helligkeiten

Wie bereits dargestellt, können unterschiedliche mittlere Helligkeiten der beiden Rechtecke zu unerwünschten lokalen Maxima bei $(D,0)$ und $(-D,0)$ führen, so daß kleine Verschiebungen nicht mehr sicher bestimmt werden können.

Rein technisch betrachtet können wir den „Referenzpunkt“ für Nulldisparität einfach verschieben (vgl. Abbildung 3.1), indem wir beispielsweise die Rechtecke aus dem linken Bild um die halbe Kantenlänge in x und y Richtung verschoben aus den Bildern entnehmen und bei der Verschiebungsbestimmung diesen Versatz wieder herausrechnen. Damit fallen die unerwünschten Maxima nicht mehr in den Suchbereich und kleine Verschiebungen können sicherer bestimmt werden.

Genau ein solcher Versatz scheint sich nach V. Braitenberg [Bra85, S.386] auch im biologischen Vorbild zu finden:

„... it seems that the cuts in one picture are halfway between the cuts in the other picture, so that each strip has overlapping information with the stripes on either side, belonging to the other eye ...“

7.3 Blasse Streifen

Die Suche nach einem Echomaximum im Cepstrum ist i.a. nicht erfolgreich, wenn Spezialfälle auftreten, wie sie in Abbildung 7.2 skizziert sind. Die Abbildung illustriert, daß sich nach der nahtlosen Zusammenstellung der Teilsignale in besonderen Fällen "künstliche Strukturen" ergeben, bei denen sich nicht mehr angeben läßt, welches die erzeugenden Teilsignale waren. Das abgebildete Doppelsignal könnte auch durch zwei sich zu einem Quadrat ergänzende dreieckförmige Teilsignale entstanden sein (gestrichelte Linie).

Wie in Kapitel 2 besprochen, befinden sich nach den anatomischen Untersuchungen von S. LeVay, D. Hubel und T. Wiesel [LHW75] zwischen den okularen Dominanzstreifen dünnere Streifen, in denen die Anzahl tangentialer Fasern deutlich geringer ist als in den okularen Dominanzstreifen, so daß sie als trennende, blasse Streifen in der Präparation erscheinen. Ihre funktionale Bedeutung ist noch ungeklärt. Fassen wir die Verringerung der tangentialen Fasern im Bereich der *pale bands* – falls sie denn überhaupt existieren und nicht nur ein präparationstechnisches Artefakt sind (vgl. Seite 10) – technisch als Verringerung der Signalstärke in diesem Bereich auf, so ergibt sich insgesamt eine abgerundete Trägerfunktion, wie es der Querschnitt in

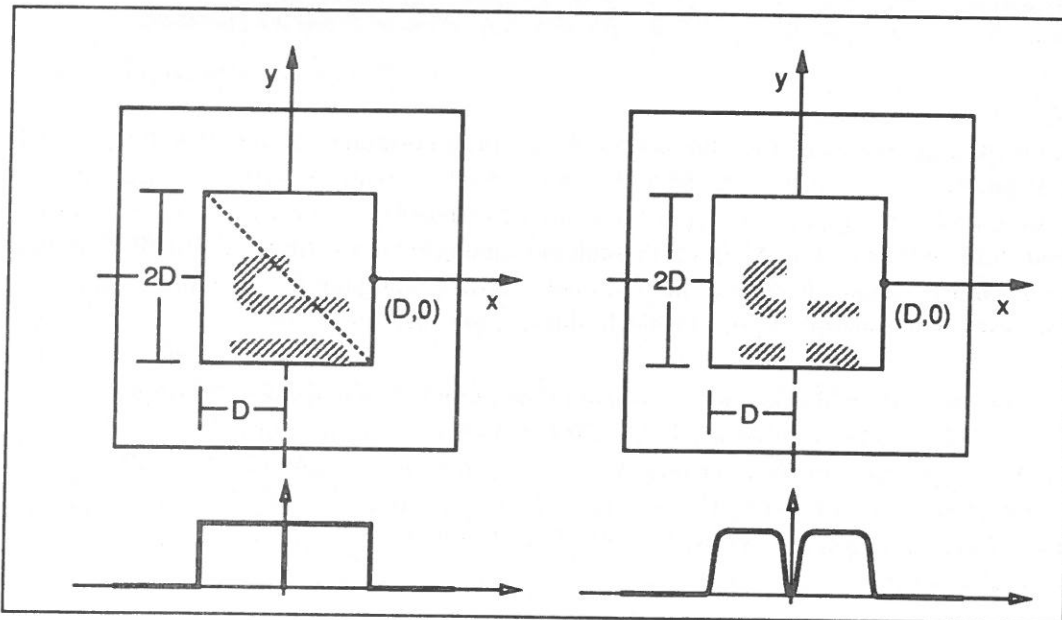


Abbildung 7.2: Links: bisheriges Verfahren. Die Trägerfunktionen sind im Querschnitt rechteckig. Rechts: Funktionale Deutung der Verringerung der tangentialen Fasern im Bereich der Blassen Streifen (*pale bands*) als Verringerung der Signalstärke der Teilsignale zur Beseitigung künstlich Strukturen.

Abbildung 7.2 andeutet. Ich habe bei den Experimenten mit der Cepstrumfilterung, die einen solchen Trennbereich berücksichtigt, mit $1/6 - 1/8^1$ der Breite der okularen

¹Die Werte lehnen sich an die in [LHW75] beschriebenen Messungen an.

Dominanzstreifen gearbeitet. Die quaderförmigen Trägerfunktionen zur Extraktion der Teilsignale erhalten dazu einen entsprechenden in x -Richtung stetig abgerundeten Verlauf (Gaußfunktion, siehe Abbildung 7.2). Mit dieser Modifikation² ist es nun möglich, in dem skizzierten Fall und in anderen Fällen „unglücklicher“ Zusammenstellung der Teilsignale doch noch Verschiebung zu bestimmen.

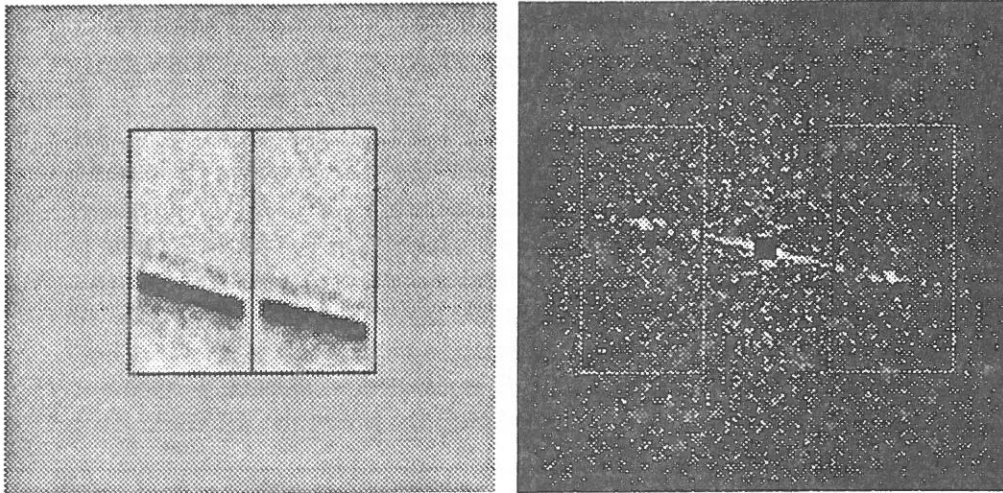


Abbildung 7.3: Cepstrum bei geraden Kantensegmenten mit einem Trennbereich. Links: Eingangssignal. Die Teilsignale enthalten jeweils ein gerades Kantensegment aus einer Kameraaufnahme eines Quaders. Rechts: Cepstrum. Der Vergleich mit Abbildung 6.7 zeigt, daß das Verschiebungsmaximum deutlicher und das störende Maximum schwächer geworden ist.

Der weiter vorne besprochene Problemfall gerader, bündig anschließender Kantensegmente wird damit auch gelöst.

Bei der Zusammenstellung der Teilsignale ist nach dem bisherigen Verfahren durch nahtloses Aneinanderfügen die Information darüber verlorengegangen, welches die Einzelsignale waren. Die Doppelsignale könnten auch durch Übereinanderanordnung oder andere nahtlose Zusammensetzungen erzeugt worden sein.

Die Verwendung eines Trennbereiches läßt sich dann als Wiedereinfügung dieser verlorengegangenen Information auffassen. Mit dieser Sichtweise können wir die Cepstrumtechnik noch etwas erweitern.

7.4 Trägerfunktionen zur Signalextraktion

Abschnitt 3.3.2 zusammenfassend stellen wir fest, daß bisher in der Literatur nur rechteckige Teilsignale verwandt wurden, obwohl Y. Yeshurun und E.L. Schwartz

²Man kann hier natürlich auch andere aus der Bildverarbeitung bekannte Fensterfunktionen benutzen, die bessere signaltechnische Eigenschaften haben. Y. Yeshurun und E.L. Schwartz schreiben, daß sie Fensterfunktionen wie *hanning windows* nicht für notwendig halten, da das Stereosignal extrem stark sei. [YS89, S.762]

ihren Algorithmus für beliebige Formen definieren [YS89, S.762]. Allerdings wird in ihrer gesamten restlichen Darstellung nur der Fall rechteckiger Teilsignale behandelt.

7.4.1 Gemeinsame Trägerfunktion

Wir stellen weiter fest, daß die technischen Anwendungen der Cepstrumtechnik in den genannten Beiträgen weder durch die Rechnungen in diesen Beiträgen, noch durch meine in Kapitel 3 dargestellte Rechnung umfassend korrekt modelliert worden sind. Die dargestellte Rechnung stellt lediglich eine mathematische Motivation dar, die nur in Spezialfällen der technischen Anwendungen in vollem Umfang zutrifft.

Tatsächlich wurden die Ausgangsbilder vor der Filterung – mathematisch gesprochen – noch mit einer quaderförmigen Trägerfunktion multipliziert, um die Ausschnitte zu bilden. Diese modifizierten Signale wurden danach erst mit bekannter Verschiebung addiert, so daß sie nebeneinander zu liegen kommen. Dieses Signal wurde dann gefiltert. Anknüpfend an Gleichung (3.13) heißt das:

$$\begin{aligned} f(x, y) &= s_1(x, y) \cdot t(x, y) + s_2(x - D_X, y - D_Y) \cdot t(x - D_X, y - D_Y) \\ &= s(x, y) \cdot t(x, y) + s(x - x_0 - D_X, y - y_0 - D_Y) \cdot t(x - D_X, y - D_Y) \end{aligned}$$

wobei (x_0, y_0) die unbekannte Verschiebung und (D_X, D_Y) die bekannte Verschiebung ist. Die Trägerfunktion zur Extraktion der Teilsignale ist $t(x, y)$. Es sollte außerhalb einer kleinen Umgebung um den Ursprung natürlich $t(x, y) \equiv 0$ gelten, damit wirklich nur lokale Ausschnitte der beiden Ausgangsbilder zur Disparitätsschätzung herangezogen werden. Wegen der einfachen mathematischen Behandelbarkeit habe ich – wie auch die anderen Autoren – den Fall mit $t(x, y) \equiv 1$ durchgerechnet und damit implizit genügend beschränkte Signale $s(x, y)$ unterstellt, die trotz Verschiebung noch vollständig innerhalb des Rechteckes liegen. Das ist natürlich i.a. nicht erfüllt.

Wie an der Fouriertransformierten zu sehen ist ($\mathbf{u} = (u, v)$),

$$F(\mathbf{u}) = S(\mathbf{u}) * T(\mathbf{u}) + \left(S(\mathbf{u}) e^{-i(x_0 u + y_0 v)} * T(\mathbf{u}) \right) \cdot e^{-i(D_X u + D_Y v)} \quad (7.1)$$

lassen sich die beiden Exponentialfunktionen mit (x_0, y_0) bzw. (D_X, D_Y) diesmal nicht so leicht aufsplitten und zu einem cos-Term zusammenfassen.

An dieser mathematischen Schwierigkeit kann man aber sehen, daß die Cepstrumtechnik einen von der Fensterfunktion abhängigen Verfahrensfehler haben muß: Stellen wir uns, bei festem $(D_X, D_Y) = (D, 0)$, den Durchmesser³ des Trägers von $t(x, y)$ immer kleiner werdend vor, so ist klar, daß ein immer dominanteres Maximum im Cepstrum bei $(D, 0)$ entstehen muß.

³Der Durchmesser ist anschaulich für unregelmäßig berandete Mengen als die kleinste umschreibende Kugel definiert.

Im Grenzfall nur noch eines Punktes bestimmen wir die Verschiebung zweier Diracstöße, die außer der Intensität an diesem Ort keine Information mehr über $s(x, y)$ enthalten. Den anderen Grenzfall einer unendlich ausgedehnten Trägerfunktion haben wir schon berechnet. Es stellt sich also die Frage nach einer günstigen Wahl der Trägerfunktion $t(x, y)$, die zwischen diesen beiden Extremen liegt.

Wegen der wohldefinierten Begrenztheit im Orts- und Frequenzraum, habe ich mir die Gaußfunktion

$$g_{\sigma_x, \sigma_y}(x, y) = g_{\sigma_x}(x) \cdot g_{\sigma_y}(y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)}$$

mit

$$g_{\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \frac{x^2}{\sigma^2}}$$

als alternative Fensterfunktion gewählt.

Meine praktischen Versuche haben gezeigt, daß damit die störenden Effekte bei geraden Kantensegmenten und „künstlichen Strukturen“ weitgehend von selbst entfallen, da die Information über die Art der Fensterbildung noch im Signal enthalten ist.

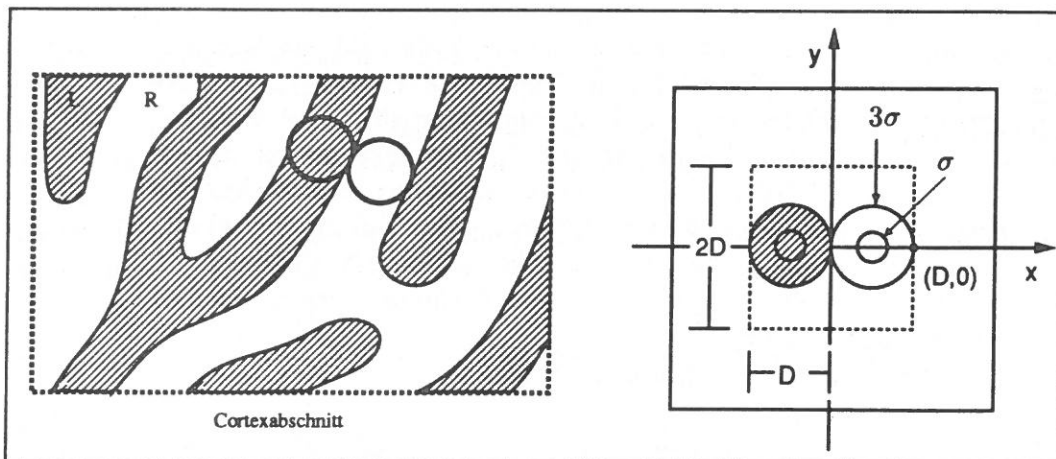


Abbildung 7.4: Modellbildung mit gleichartigen Trägerfunktionen. Links: zwei runde Einzugsbereiche für eine hypothetische Disparitätszelle. Rechts: mathematische Modellierung mit sehr gering überlappenden gaußförmigen Trägerfunktionen.

Um die Ergebnisse bei gaußförmiger Trägerfunktion mit denen bei rechteckiger Trägerfunktion vergleichen zu können, muß die Gaußfunktion noch geeignet parametrisiert werden. Damit die Gaußfunktion an das bisherige Rechteck angrenzt und außerhalb praktisch Null ist, muß $3\sigma_x = D$ und $3\sigma_y = 2D$, also $\sigma_y = 2 \cdot \sigma_x$ gelten. Mit dieser Wahl haben wir etwa vergleichbare Eingangssignale geschaffen. Die folgende Abbildung demonstriert die Überlegenheit gaußförmiger Trägerfunktionen gegenüber quaderförmigen Trägerfunktionen.

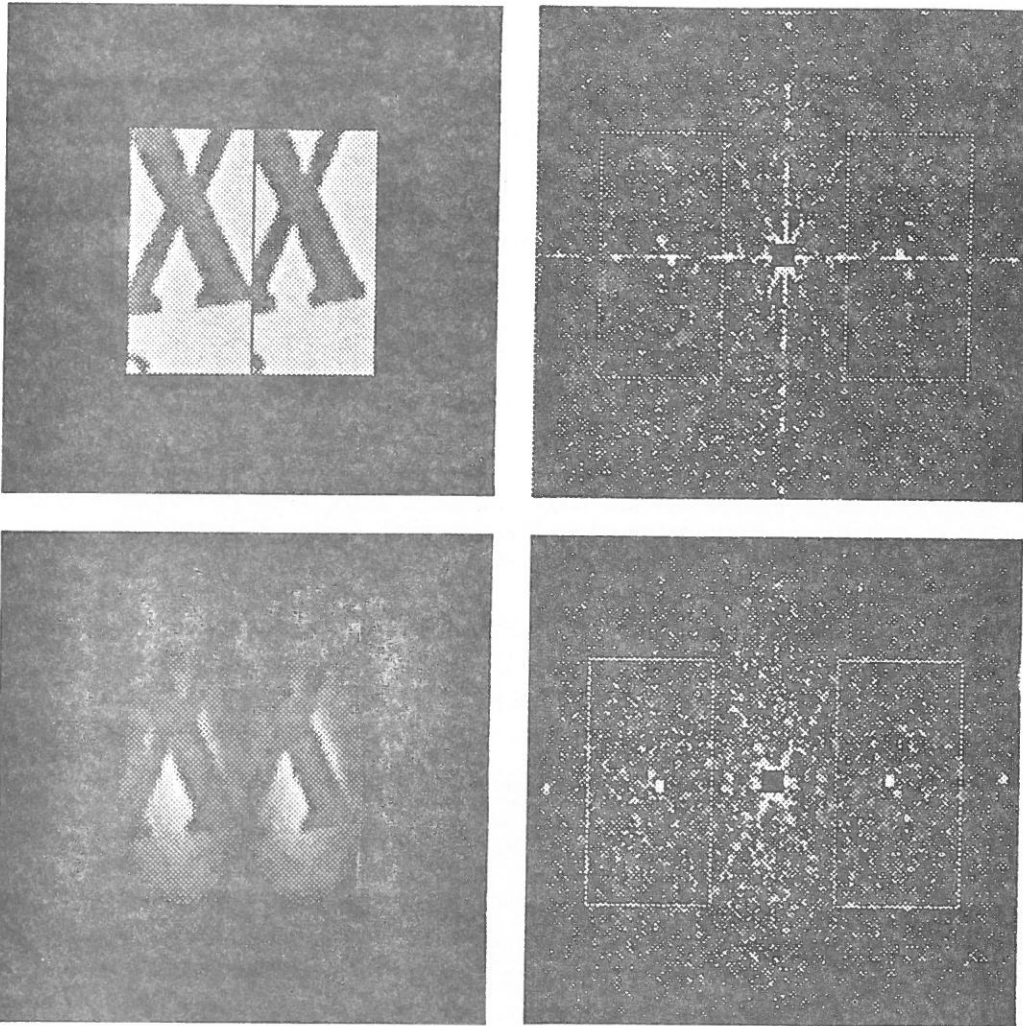


Abbildung 7.5: Cepstrum mit gaußförmigen Trägerfunktionen. Teilbilder aus einer Kameraaufnahme, wie zuvor. Es ist jeweils links das Eingangssignal und rechts das Cepstrum abgebildet. Oben: Das bisherige Verfahren mit quaderförmigen Trägerfunktionen. Unten: Das Doppelsignal wurde mit gaußförmigen Trägerfunktionen ($\sigma_x = 10\frac{2}{3}$, $\sigma_y = 21\frac{1}{3}$) gebildet.

7.4.2 Unterschiedliche Trägerfunktionen

Noch einen Schritt weitergehend, kann man sich auch unterschiedliche Trägerfunktionen zur Teilbildextraktion vorstellen, d.h.

$$\begin{aligned}
 f(x, y) &= s_1(x, y) \cdot t_1(x, y) + s_2(x - D_X, y - D_Y) \cdot t_2(x - D_X, y - D_Y) \\
 &= s(x, y) \cdot t_1(x, y) \\
 &\quad + s(x - (x_0 + D_X), y - (y_0 + D_Y)) \cdot t_2(x - D_X, y - D_Y)
 \end{aligned}$$

wobei hier die t_i verschiedene Trägerfunktionen bedeuten sollen.

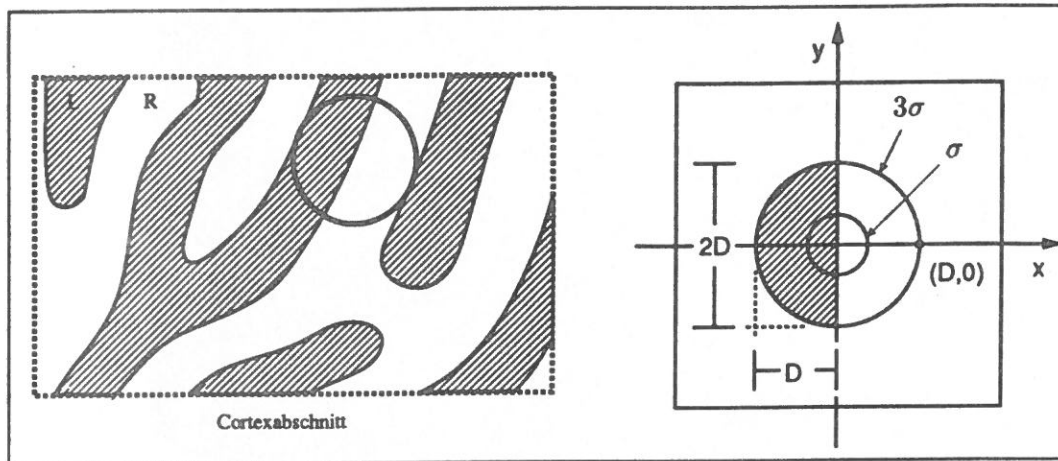


Abbildung 7.6: Modellbildung mit unterschiedlichen Trägerfunktionen. Links: Skizze für eine hypothetische Disparitätszelle mit nur einem gaußförmigen Einzugsbereich. Rechts: Mathematische Modellierung mit zwei sich ergänzenden gaußförmigen Trägerfunktionen.

Abbildung 7.6 zeigt eine biologisch orientierte Wahl der Trägerfunktionen. Damit trifft man allerdings wieder – wegen der nahtlosen Zusammenstellung – in entsprechenden Fällen auf die bereits besprochenen Effekte. Man kann auch hier einen Trennstreifen verwenden um die Probleme zu vermeiden. Ich habe Versuche mit

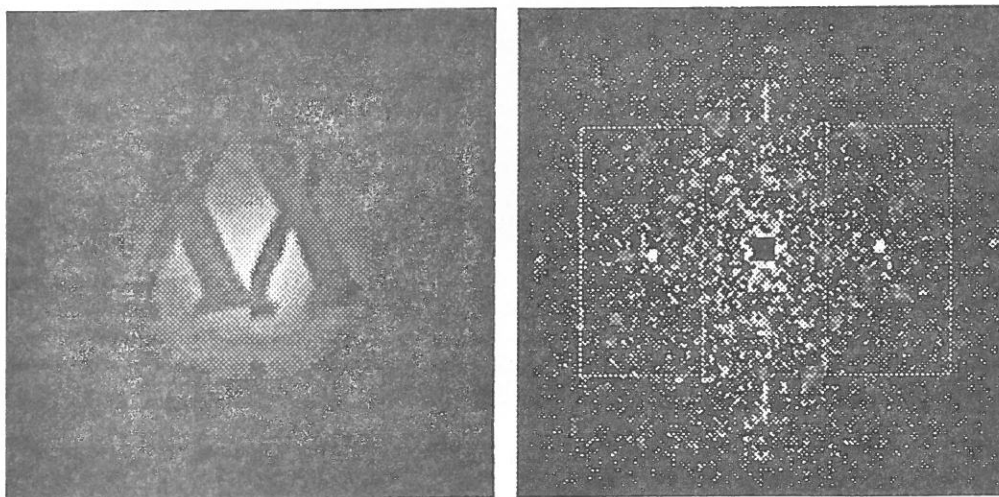


Abbildung 7.7: Zur Demonstration des Cepstrums mit unterschiedlichen Trägerfunktionen. Der verwendete Bildausschnitt entspricht dem aus der vorigen Abbildung. Links: Ein Doppelsignal bei dem die Teilbilder mit verschiedenen gaußförmigen Trägerfunktionen aus dem linken bzw. rechten Bild extrahiert wurden. Für die Gaußfunktionen wurde jeweils $\sigma = 21\frac{1}{3}$ gewählt. Rechts: Cepstrum.

gleichartigen und verschiedenartigen gaußförmigen Trägerfunktionen bei variierendem σ_x und σ_y durchgeführt. Es hat sich ergeben, daß bei vergleichbarer Trägergröße die Wahl gaußförmiger und speziell gleichartiger Träger zur Teilbildextraktion bei Realweltbildern bessere Ergebnisse liefert als die Wahl quaderförmiger Träger.

Natürlich sind hier auch ganz andere Variationen unterschiedlicher Trägerfunktionen denkbar. Wenn es gelänge in diesem Fall, den zu Formel (7.1) entsprechenden Ausdruck theoretisch auszuwerten, könnte man daraus evtl. eine den angesprochenen systematischen Fehler minimierende Trägerfunktion entwickeln.

7.5 Bandpaßfilterung

Y.Yeshurun und E.L. Schwartz schreiben [YS89, S.763]:

„We emphasize that the input to this algorithm is not necessarily a “gray scale” image, as we have used in the figures. Naturally, any preprocessing (e.g. high pass filtering, edge enhancement, etc.) is compatible with this approach. However it is important to emphasize that we obtain good performance without any preprocessing.“

Diese Erfahrungen kann ich stützen. Wenn man die Rohbilder durch eine geeignete

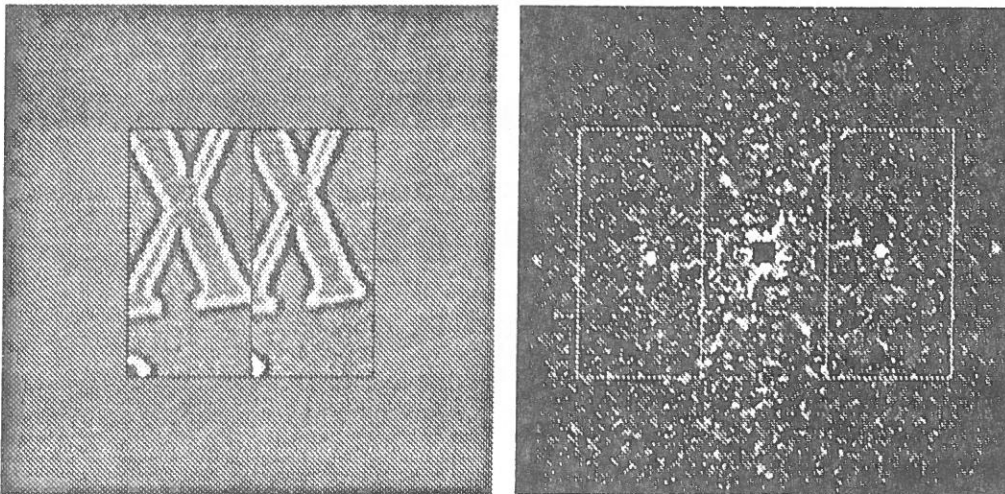


Abbildung 7.8: Die Abbildung demonstriert die Auswirkung einer Bandpaßfilterung auf das Filterergebnis am Beispiel des bekannten Ausschnittes. Links: Doppelsignal. Die Eingangsbilder sind vor der Bildung des Doppelsignals mit einem LoG-Filter ($\sigma = 0.71$) gefiltert worden. Rechts: Cepstrum. (vgl. Abbildung 7.5, oben)

Bandpaßfilterung aufbereitet, wird das Echo jedoch deutlicher. Tatsächlich sind die Daten in der primären Sehrinde – durch die *center surround* Verschaltung retinaler Ganglienzellen – bereits vorverarbeitet. Die Wirkung der rezeptiven Felder wird oft rechentechnisch durch eine LoG-Bandpaßfilterung (*mexican hat* Filterung)

$$\text{LoG}_\sigma[f(x, y)] := f(x, y) * m_\sigma(x, y) = f(x, y) * \Delta g_\sigma(x, y)$$

der retinalen Eingangsdaten angenähert. Dabei bezeichnet hier Δ den Laplaceoperator

$$\Delta f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2}$$

$LoG_\sigma[f]$ berechnet wegen

$$f(x, y) * \Delta g_\sigma(x, y) = \Delta(f(x, y) * g_\sigma(x, y))$$

– vereinfacht dargestellt – eine Ableitung von dem mit g_σ geglätteten f . Mit dieser Sichtweise ist klar, daß wir das Echo im Filterergebnis nur „verschmieren“, wenn wir σ zu groß wählen. Dagegen können wir durch Wahl eines geeignet kleinen σ erhoffen, das Echo bis zu einem gewissen Grad „schärfen“ zu können.

In Anlehnung an die Größe biologischer rezeptiver Felder (vgl. Seite 7) und aus rein rechentechnischen Überlegungen liegt der sinnvolle Bereich für σ etwa in dem Intervall $[0, 2]$.

Die Abbildung 7.8 veranschaulicht die Wirkung einer *mexican hat* Filterung auf das Verschiebungsmaximum im Cepstrum. Die Implementation von $LoG_\sigma[f]$ wurde in diesen Beispielen noch nicht mit ausgeglichenen Ortsraumfiltern durchgeführt, sondern erfolgte einfach durch Multiplikation von $F(u, v)$ mit $M_\sigma(u, v)$ im Frequenzraum.

7.6 Das Cepstrum als Autokorrelation mit adaptiver Vorfilterung

Wie bereits in Kapitel 3 angedeutet ergibt sich der Zusammenhang des Cepstrums zur Autokorrelation wegen

$$\log(\|F(u, v)\|^2) = \left\| \frac{\sqrt{\log(\|F(u, v)\|^2)}}{\|F(u, v)\|} F(u, v) \right\|^2 = \|H\|^2 \quad (7.2)$$

und

$$\|\mathcal{F}\{\|H\|^2\}\|^2 = \|\mathcal{F}\{H^*H\}\|^2 = \|h^* * h\|^2 = \|h \circ h\|^2 \quad (7.3)$$

zu

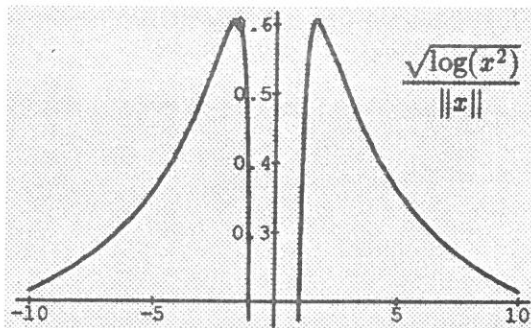
$$Cepstrum\{f\} = \|h \circ h\|^2 \quad (7.4)$$

mit $h(x, y) = k_F(x, y) * f(x, y)$. Das Cepstrum kann also im wesentlichen als eine Autokorrelation mit einer nicht-linearen und bildabhängigen Vorfilterung aufgefaßt werden.

Die Fouriertransformierte des adaptiven Vorfilters lautet

$$K_F(u, v) = \frac{\sqrt{\log(\|F(u, v)\|^2)}}{\|F(u, v)\|} \quad (7.5)$$

Es interessiert uns natürlich, wie das vorgefilterte Bild $f(x, y) * k_F(x, y)$ aussieht und welche Form der bildabhängige Filterkern $k_F(x, y)$ besitzt. Die Argumentation von T.J. Olson und D.J. Coombs zu der Formel finde ich nicht richtig verständlich, und das in Anhang A [OC90, S.29] gezeigte „vorgefilterte Bild“ scheint mir nicht korrekt zu sein. Ich habe deshalb eigene Untersuchungen angestellt.



Die nebenstehende graphische Darstellung des Ausdrucks für die Fouriertransformierte des Filterkerns veranschaulicht, daß hohe Amplituden bis zu einer Grenzamplitude von 1 stark gedämpft werden. Amplituden kleiner als 1 werden dagegen enorm verstärkt, ändern jedoch ihr Vorzeichen.

Für einfache Funktionen $f(x, y)$ kann man den Filterkern auch noch analytisch berechnen, aber das hilft wegen der amplitudenabhängigen – und nicht frequenzabhängigen – Struktur für allgemeine Aussagen nicht sehr viel weiter.

Für eine bandbegrenzte Funktion – so könnte man natürlich argumentieren – sieht das Spektrum im wesentlichen wie das einer Gaußschen Glocke aus. Dabei unterstellen wir dann, daß die abweichenden Modulationen gering genug sind, so daß sie keine wesentliche Änderung in der Dämpfung des Filters bewirken. Unter dieser Voraussetzung kann man z.B. den Filterkern für ein Doppelsignal aus zwei Gaußschen Glocken ausrechnen und ihn als Prototypen für eine bestimmte Menge bandbegrenzter Signale ansehen.

Meine praktischen Versuche haben gezeigt, daß $k_F(x, y)$ für eine Reihe von Beispielen die Form eines Bandpaßfilters wie $m_\sigma(x, y)$ mit kleinem σ hat. Umgekehrt bedeutet

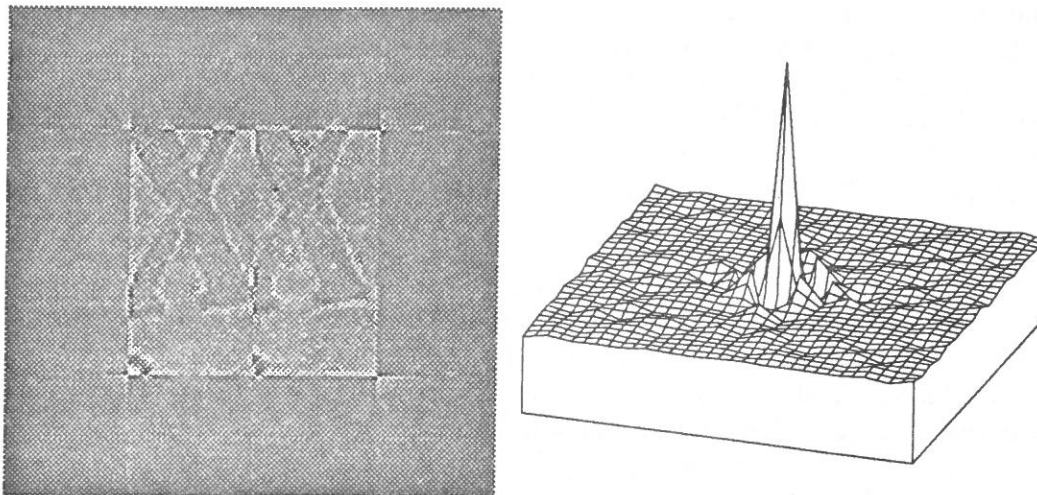


Abbildung 7.9: Das Cepstrum als Autokorrelation. Links: Adaptiv vorgefiltertes Bild $f(x, y) * k_F(x, y)$. Rechts: Darstellung des zugehörigen Filterkerns $k_F(x, y)$. Das Gitter ist zur besseren Darstellung um den Faktor zwei feiner als die Datenpunkte gewählt worden.

dies – wenn diese experimentell gestützte Vermutung stimmt –, daß die Cepstrumsberechnung ohne den Zwischenschritt der Logarithmierung des Spektrums für bandpaßgefilterte Daten auch ein gutes Maximum liefern müßte. Oder anders formuliert:

Eine Disparitätsbestimmung müßte – bei *mexican hat* gefilterten Bildern – allein mit der Autokorrelation durchgeführt werden können.

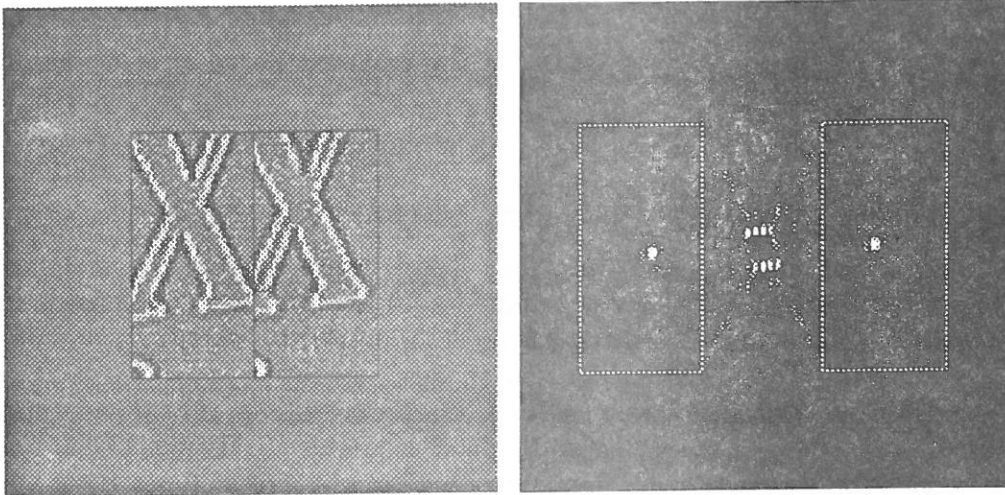


Abbildung 7.10: Autokorrelation bei vorhergehender Bandpaßfilterung. Links: Gefiltertes Eingangssignal g (LoG Filter, $\sigma = 0.35$). Rechts: Darstellung von $\|g \circ g\|^2$ (nicht logarithmiert, da die – vergleichsweise schwachen – peaks in einer logarithmierten Darstellung nicht mehr zu erkennen wären).

Diese Sichtweise ist auch biologisch viel plausibler. Gehen wir davon aus, daß die Funktion der rezeptiven Felder retinaler Ganglienzellen auf eine statische Bandpaßfilterung der retinalen Daten hinausläuft, dann wäre für disparitätssensitive Zellen der Sehrinde oder tieferer Schichten eine neuronale Implementation einer lokalen Autokorrelation bereits ausreichend. Rein vom technischen Standpunkt – das soll hier abschließend noch einmal betont werden – ist das Cepstrum zur Verschiebungsbestimmung nachweislich besser geeignet als die Autokorrelation. Der bildabhängige Vorfilter ist also nur bandpaßähnlich. Die Stärke des Cepstrums scheint gerade in dieser Bildabhängigkeit zu liegen.

7.7 Zusammenfassung

Die Cepstrumtechnik ist ein robustes und schnelles Verfahren zur Disparitätsbestimmung. Das Verfahren arbeitet ausschließlich lokal und ist damit leicht parallelisierbar. Wir haben im vorigen Kapitel einige in der Literatur nicht genannte Spezialfälle kennengelernt, in denen die bisherige Cepstrumtechnik unter bestimmten, bildabhängigen Bedingungen nicht mehr so brilliant funktioniert. Es wurde in diesem Kapitel gezeigt, wie die Problemfälle eliminiert werden können. Neben den eher trivialen Verbesserungen bezüglich des Suchbereichs und der Extraktion nicht korrespondierender Bildausschnitte (Abschnitt 7.2) zur Umgehung des verfahrensbedingten Maximums bei $(D, 0)$, wurde – ausgehend von einer biologisch motivierten Änderung der Trägerfunktion – auf andere, i.a. günstigere Trägerfunktionen zur Teilbildextraktion hingewiesen.

Es ist illustriert worden, das eine LoG Bandpaßfilterung der Rohdaten mit kleinem σ vorteilhaft sein kann.

Abschließend wurde eine alternative und biologisch wesentlich plausiblere Sichtweise auf die Wirkung der Cepstrumfilterung vorgestellt: Praktische Versuche haben gezeigt, daß sich das Cepstrum als Quadrat der Autokorrelation des Doppelsignals mit bildabhängiger, aber bandpaßähnlicher Vorfilterung verstehen läßt. In der Gegenprobe wurde illustriert, daß das Quadrat der Autokorrelation bei LoG-gefilterten Daten erwartungsgemäß verhält. Das Echomaximum ist jedoch deutlich schwächer als bei der Cepstrumfilterung. Der Vorteil des Cepstrums scheint unter anderem in seiner Bildabhängigkeit zu liegen.

Kapitel 8

Lokale Tiefenkarten

Wir wollen jetzt sehen, wie die mit der Cepstrumtechnik bestimmten lokalen Tiefenschätzungen aussehen. Ich habe die Rekonstruktionsfähigkeit für einzelne Beispiele künstlich generierter Szenen überprüft, weil hierbei alle relevanten Parameter – und damit auch die Ergebnisse – bekannt sind. Im zweiten Abschnitt stelle ich die lokalen Tiefenschätzungen für zwei Kamerabildpaare vor.

8.1 Disparitätskarten für generierte Bilder

Man kann sich gedanklich komplizierte Szenen aus mathematischen Objekten (Kugeln, Polyedern, Fraktalen etc.) mit bestimmten Oberflächeneigenschaften (reflektierend, streuend, u.a.m.) und verschiedenen, unterschiedlich hellen Beleuchtungsquellen vorstellen. Für eine ideale Lochkamera mit einer bestimmten Kameraposition und festgelegten Kameraparametern wie Neigungswinkel, Brennweite etc. ist unter diesem exakten Aufbau auch das Bild mathematisch bestimmt, das diese Kamera von der aufgebauten Szene entwirft. Ein *raytracer* ist ein Programm, das ein solches Bild numerisch bestimmt.

Ich habe den frei verfügbaren *raytracer* MTV von Mark van de Wetering [vdW88] verwendet,¹ um die Bildpaare bei gegebener Stereoanordnung und speziellen fixierten Objektoberflächen zu berechnen.

Ich habe bei den Experimenten mit künstlichen Szenen keine umfassenden Testreihen durchgeführt, weil das generieren vieler Bildpaare entsprechend aufwendig ist und allein beispielsweise für den Fall einer fixierten Ebene ein systematisches Durchgehen aller Normalenvektoren unter vertretbarem Aufwand nicht möglich ist.

Für Einzelbeispiele mit geneigter Ebene wurden die Disparitätskarten mit der Cepstrumtechnik berechnet und die so gewonnenen Disparitäten mit den exakten mathematischen Werten verglichen. Die künstliche Ebene wurde dazu so generiert, daß sie einen Blickwinkel von etwa 1° ausfüllt.

¹Der *raytracer* bestimmt den Wert für ein Bildelement aus dem Strahl aus der linken oberen Ecke des Bildelements. Ich habe meine Kopie des Programmes dahingehend modifiziert, daß der Strahl aus der Pixelmitte verwendet wird.

Beispiel. Für das im folgenden betrachtete Beispiel wurde die Ebene in 80×80 bzw. 160×160 Quadrate unterteilt, um eine Textur anzubringen.² Als Textur wurden hier Zufallsmuster (starke Textur) bzw. Kamerabildausschnitte (eher „schwache“ Textur) verwendet. Die Stereoanordnung entspricht der im Text von Abbildung 5.8 (Seite 44) angegebenen. Dadurch ergeben sich bei einer Teilsignalgröße von 32×64 Pixeln etwa 5×5 bzw. $2\frac{1}{2} \times 2\frac{1}{2}$ Pixel für ein Texturquadrat und pro Teilsignal insgesamt etwa 5×10 bzw. 10×20 Texturquadrate. Bei einer Auflösung von 512×512 Pixeln pro Bild ergibt sich eine Aufteilung der Bildebene in 16×8 Kacheln. Die Abbildung 8.1 gibt einen Eindruck von den entstehenden Teilsignalen.

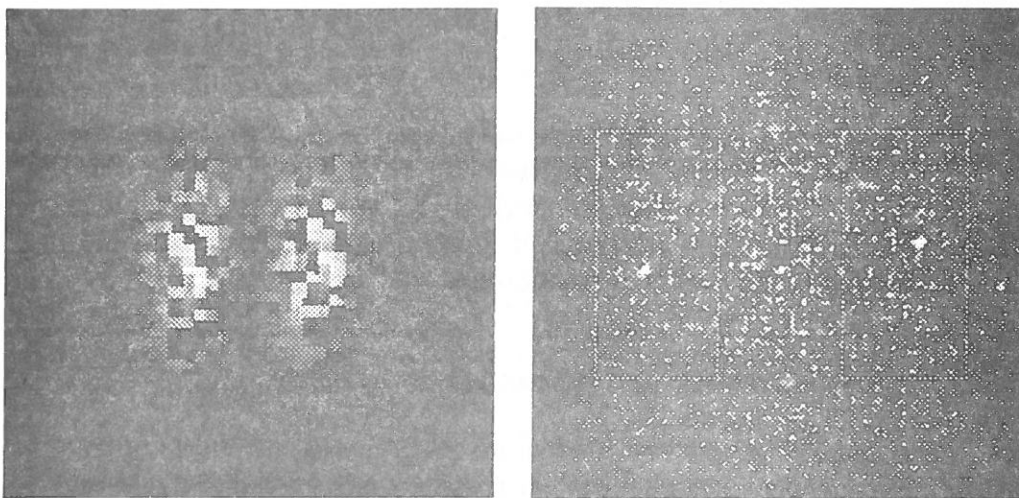
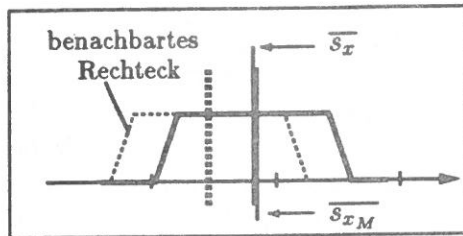


Abbildung 8.1: Die Abbildung veranschaulicht die Art der Teilsignale und Cepstra für die im Text beschriebenen generierten Bildern. Links: Eingangssignal bei 160×160 Texturquadraten auf der Ebene. Die Teilsignale entstammen der Kachel (5, 1) in Abbildung 8.2. Rechts: Cepstrum.

Bei dem Vergleich mit Hilfe der exakten Formel (5.12) für die Disparitäten bei fixierter Ebene muß man sich noch entscheiden, welchen Wert man zum Vergleich heranziehen will. Bei der herkömmlichen Cepstrumtechnik hat man mindestens die Wahl zwischen der Verschiebung am Mittelpunkt des Rechteckes $s_{\vec{M}} = (s_{x_M}, s_{y_M})$ oder dem Mittelwert der Verschiebungen innerhalb des gesamten rechteckigen Teilgebietes $\vec{s} = (\bar{s}_x, \bar{s}_y)$.

Unterstellt man wieder rechteckige Ausschnitte, dann kann man die Verschiebung nach Formel (5.12) auf einem solchen rechteckig begrenzten Gebiet ohne größere Schwierigkeiten integrieren^M und erhält eine umfangreichere Formel. Ich will hier nicht weiter auf die Formel eingehen. Um eine anschauliche Vorstellung von der Verteilung der Disparitäten zu bekommen, kann man die Formel (5.12) verwenden, um mit einer numerischen Rechnung und genügend feiner Abtastung des Rechteckes ein Histogramm der Disparitäten zu erstellen.

²Für die höhere Unterteilung wurden Konstanten im raytracer geändert. Dank Mark van der Wattering (keine quadratischen Algorithmen) gab es auch bei dieser großen Anzahl von Objekten keine Probleme bei der Berechnung der Bilder.



Die nebenstehende Skizze zeigt die typische Verteilung der Werte für s_x innerhalb eines Rechteckes. Die Werte für \bar{s}_x und s_{x_M} unterscheiden sich nur geringfügig. Außerdem haben die Werte benachbarter Rechtecke einen größeren Überlappungsbereich.

Für die genannte Konfiguration unterscheiden sich die Werte der Disparitäten innerhalb eines der 128 Rechtecke maximal um $1\frac{1}{2}$ Pixel.

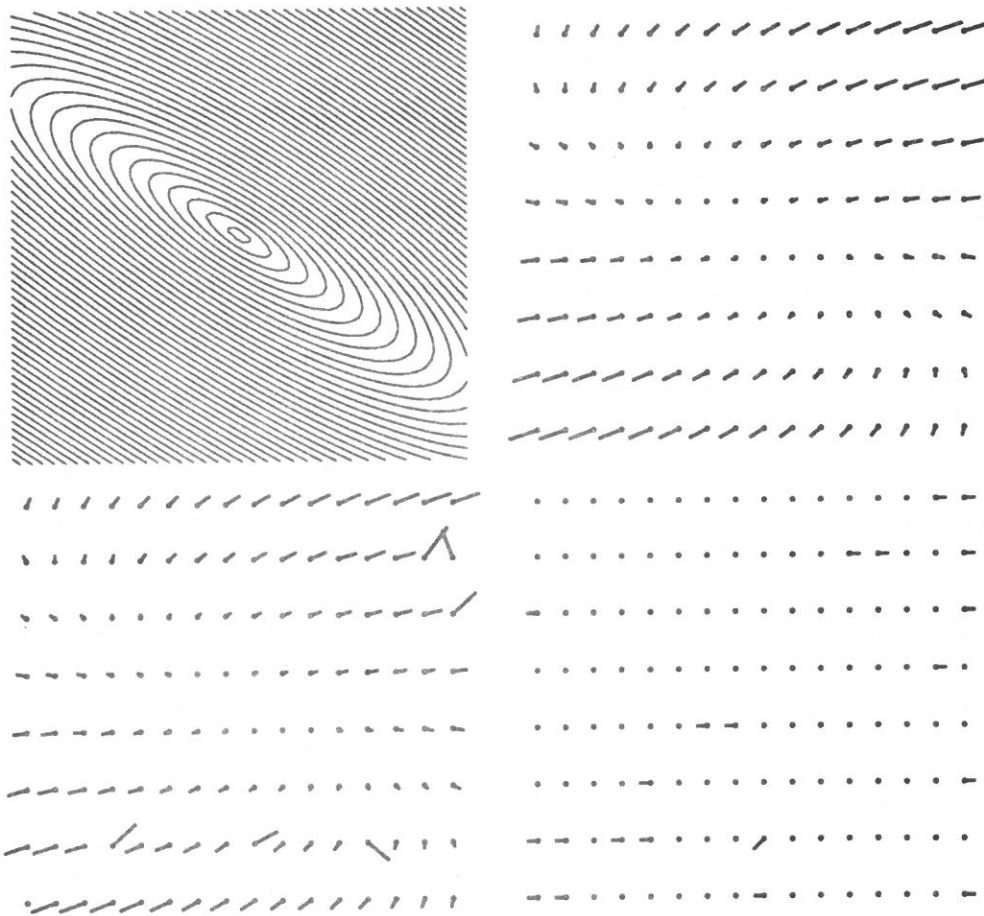


Abbildung 8.2: Disparitätskarten für einen generierten Bildausschnitt. Oben Links: Isodisparitätslinien (Identisch mit Abbildung 5.9 f). Oben Rechts: Exakte Disparitätsvektoren \vec{s} nach Formel (5.12) für 16×8 Kachelung. Unten Links: Berechnete Disparitäten für 80×80 Texturquadrate. Es ergeben sich trotz der sehr groben Bildstruktur innerhalb eines Teilsignals nur einzelne Abweichungen von den theoretischen Werten. Unten Rechts: Berechnete Disparitäten für 160×160 Texturquadrate. Es sind die Differenzvektoren zwischen den berechneten und exakten Werten vergrößert dargestellt. Die Abweichungen betragen maximal einen Pixel.

Zusammengefaßt gilt: Für größere Texturquadrate (mehr als 5 Pixel Kantenlänge) entstehen noch einige Fehlstellen. Für zu feine Unterteilungen (ab 2 Pixel Kantenlänge) werden die Texturquadrate nicht mehr gut genug im Bild repräsentiert, und es ergeben sich wenige Fehlstellen um einen Pixel. Bei mittlerer Texturquadratgröße (3 – 4 Pixel pro Quadrat) sind Fehlstellen am seltensten. Die Wahl anderer Normalenvektoren und anderer Texturen ergaben ähnliche Ergebnisse. Insgesamt ergaben sich für die von mir untersuchten künstlichen Bildpaare bei mittlerer Texturquadratgröße meistens die exakten Disparitätswerte.

8.2 Disparitätskarten für Kamerabilder

Für die Aufnahme der Bilder stand eine HITACHI CCD Kamera mit $\frac{2}{3}$ Inch Sensor, $574(H) \times 581(V)$ Bildpunkten und einem FUJINON Zoomobjektiv (A13 \times 10 BMD-D8, 10 – 130mm) zur Verfügung. Die Digitalisierung erfolgte mit dem DVS-Bildspeichersystem (FB 768 P).

Für die genaue Ausrichtung und Positionierung der Kamera konnte eine xyz -Verstelleinheit mit $\frac{1}{10}mm$ Feineinstellung in allen drei Raumachsen sowie entsprechenden Einstellmöglichkeiten für Schwenk- und Neigungswinkel verwendet werden. Die Fixation erfolgte für das linke bzw. rechte Bild nach Verschiebung um die gewählte Stereobasis einzeln durch manuelle Zentrierung einer sehr kleinen Markierung (etwa 1.5×1.5 Pixel) auf dem Objekt. Für die beiden folgenden Beispiele habe ich die Brennweite zu 130mm und die Entfernung zur Kamera zu $t = 2m$ gewählt. Die Stereobasis wurde mit 7.00cm in Anlehnung an den menschlichen Augenabstand gewählt.

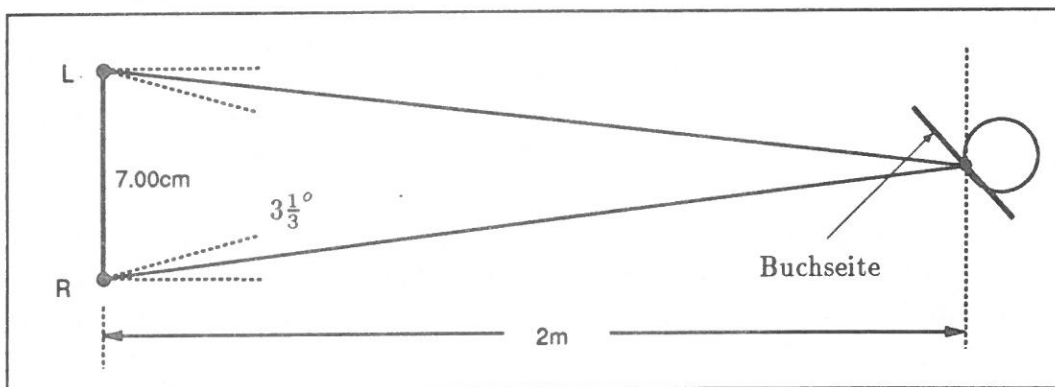


Abbildung 8.3: Skizze der Stereoanordnung für die Kameraaufnahmen von der Maisdose und der Buchseite. Die Buchseite ist links dem Betrachter zugewandt und rechts abgewandt. Die Maisdose ist nicht zentral, sondern etwas rechts von der Mittellinie fixiert worden, um auf einer Seite die Selbstverdeckungskante und den Hintergrund im Ausschnitt abbilden zu können.

Die Kameradaten für die beiden folgenden Beispiele sind jeweils Grauwertbildpaare mit 512×512 Pixeln Auflösung und 256 Graustufen.

8.2.1 Tiefenkarten

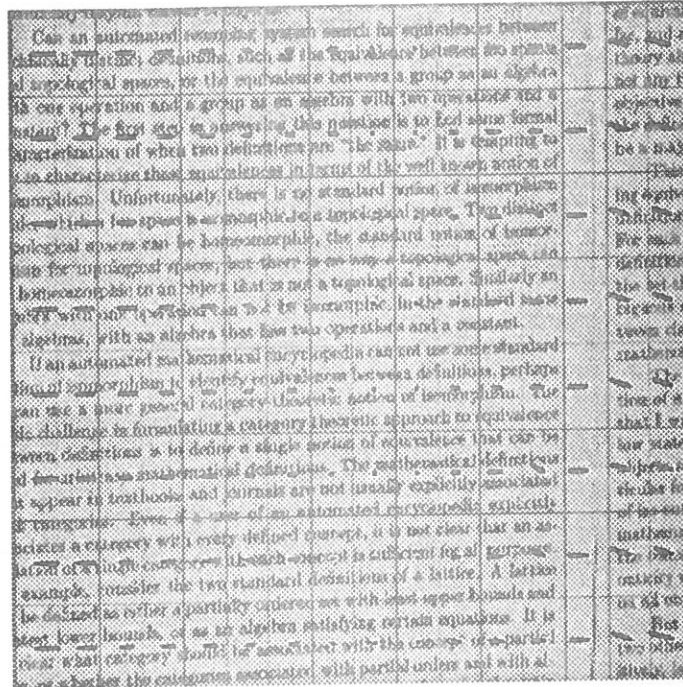


Abbildung 8.4: Disparitätskarte für das Bildpaar „Buchseite“. Die Disparitätsvektoren sind um den Faktor 4 vergrößert dargestellt. Die Rechtecke deuten nur die Unterteilung an; die Teilsignal-extraktion erfolgte mit gleichartigen gaußförmigen Trägerfunktionen ($\sigma_x = 10\frac{2}{3}$, $\sigma_y = 21\frac{1}{3}$).

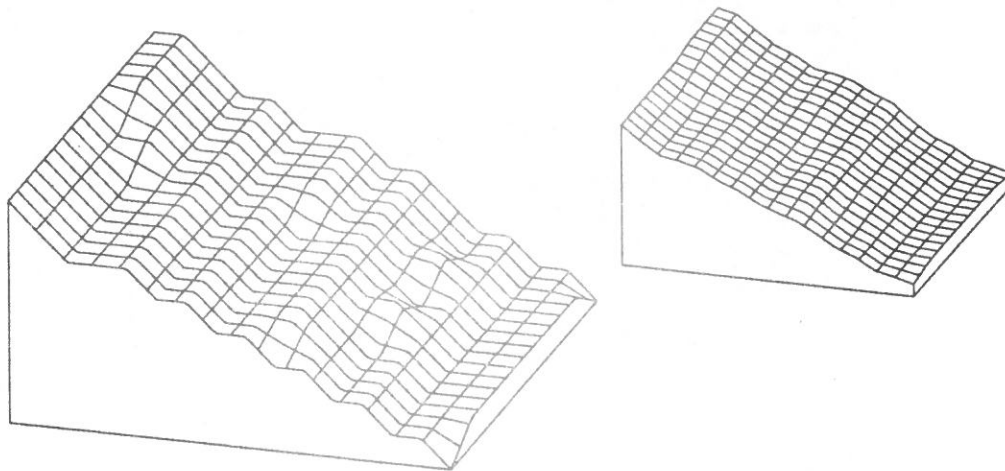


Abbildung 8.5: Lokale Tiefenkarten des Bildpaar „Buchseite“. Nach Formel (5.9) entspricht die Tiefe in der Nähe des Fixationspunktes bis auf einen Faktor der Querdisparität. Links: Darstellung der Querdisparitäten. Rechts: Darstellung der geglätteten Daten (arithmetische Mittelung). Bemerkung: Das Gitter ist zur besseren Darstellung quadratisch, also in y -Richtung um den Faktor 2 feiner als die Disparitätsdaten.

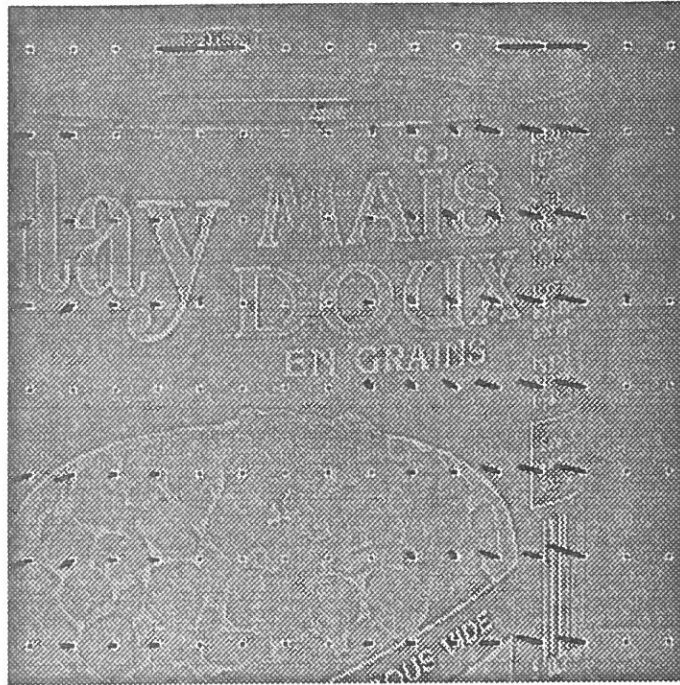


Abbildung 8.6: Disparitätskarte für das Bildpaar „Maisdose“. Die Bilder wurden zuvor mit einem LoG -Filter ($\sigma = 0.71$) bandpaßgefiltert. Die Disparitätsvektoren sind um den Faktor 4 vergrößert dargestellt. Die Teilsignalextraktion erfolgte mit gleichartigen gaußförmigen Trägerfunktionen ($\sigma_x = \sigma_y = 10\frac{2}{3}$).

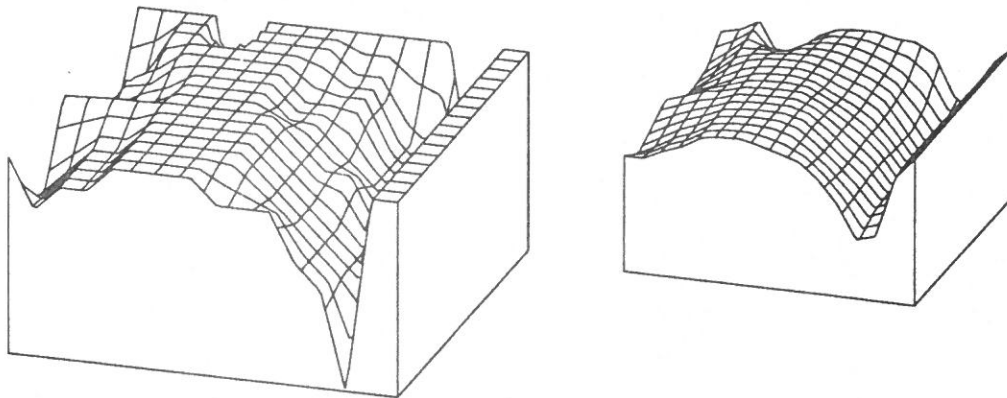


Abbildung 8.7: Lokale Tiefenkarten des Bildpaar „Maisdose“. Links: Darstellung der Querdisparitäten. Rechts: Darstellung der geglätteten Daten (arithmetische Mittelung).

8.2.2 Bemerkungen zu den berechneten Tiefenkarten

Kamerabildpaar „Buchseite“. Wie an dem Bildpaar zu erkennen ist, gibt es keine Schwierigkeiten bei genügend texturierter Oberfläche. Die Kacheln (13, 0) bis (13, 7) illustrieren das *aperture problem* bei geraden Strukturen. Bei der Berechnung dieses Bildpaares mit 32×16 Kacheln erhält man ein analoges Ergebnis wie bei der 16×8 Kachelung, bis auf eine Ausnahme: Für die linken Hälften der eben genannten Kacheln versagt das Verfahren aufgrund zu geringer Struktur der Teilsignale.

Kamerabildpaar „Maisdose“ Die Kacheln (0, 4) bis (4, 4) enthalten keine Textur, also versagt das Verfahren hier. In den Kacheln (14, 0) bis (14, 7) und (15, 0) bis (15, 7) bildet sich der – zu weit entfernte – Hintergrund ab. Die Disparitäten sind zu groß, um noch bestimmt werden zu können. Außerdem ist die Maserung des Hintergrundes recht schwach.

Diese beiden Arten von Fehlerquellen ließen sich durch einen Anwendbarkeitsindikator unterscheiden, der ein Maß für die in einem lokalen Abschnitt enthaltene Textur liefert: Bei guter Textur und Versagen der Echobestimmung liegt das Objekt wahrscheinlich außerhalb des Fusionsbereiches. Als Anwendbarkeitsindikator ist die Varianz der Bildwerte in einem Teilsignal denkbar oder auch Fourierdeskriptoren für Textur. Ein Nachteil solcher Indikatoren liegt natürlich in der mehr oder weniger willkürlichen Entscheidungsschwelle.

Zum Schluß soll noch am Beispiel dieses Bildpaares illustriert werden, daß sich auch in realen Bildern mehrere deutliche Maxima ergeben können. Wir betrachten dazu ein Teilsignal an der Zylinderrundung.

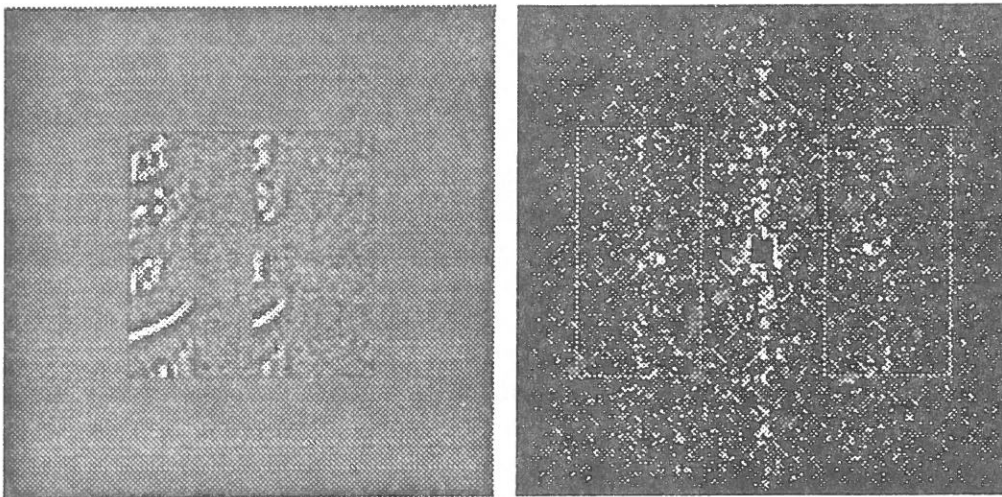


Abbildung 8.8: Teilsignal an der Zylinderrundung (Kachel (12, 5)). Links: Für eine bessere Sichtbarkeit ist hier das Doppelsignal für quaderförmige Trägerfunktionen und nicht für die zur Berechnung herangezogenen kleineren gaußförmigen Funktionen dargestellt. Es wird deutlich, daß zwei dominante Verschiebungen in diesem Teilsignalpaar existieren: Die vertikale Zylinderkante ist nach links verschoben und die Beschriftung hat einen vertikalen Versatz. Rechts: Das Cepstrum weist demnach zwei (schwache) Maxima auf.

8.3 Zusammenfassung

In diesem Kapitel wurde an Beispielen demonstriert, daß das vorgestellte Verfahren zur lokalen Tiefenbestimmung für künstliche generierte Bilder die mathematisch exakten Disparitätswerte bis auf Abweichungen von einem Pixel bestimmen kann.

Die Anwendung des Verfahrens auf Kamerabilder ergibt gute, aber noch nicht in allen Einzelheiten befriedigende Ergebnisse: Es wäre wünschenswert für diejenigen Fälle einen Anwendbarkeitsindikator für das Cepstrum zu besitzen, in denen Fehlbestimmungen aufgrund zu geringer Textur oder eines sich nicht im Fusionsbereich befindenden Oberflächenabschnittes ergeben.

Wir haben außerdem gesehen, daß sich nicht nur bei künstlichen Teststimuli (siehe E7, E8), sondern auch bei realen Szenenausschnitten zwei Maxima durch die zugrundeliegende Geometrie der gesamten Anordnung ergeben können. Die Berechnung einer mehrwertigen Disparitätskarte mit einem abgleichenden Nachbearbeitungsschritt, der zur Festsetzung der Disparität einer Kachel, die Disparitätswerte jeweils benachbarter Kacheln heranzieht, könnte für solche Fälle eine Lösung sein.

8.4 Rohbilder

Die beiden folgenden Seiten zeigen die die verwendeten Rohbilder.



Abbildung 8.9: Kamerabildpaar „Buchseite“ mit 512 × 512 Pixel Auflösung und 256 Graustufen. Oben: Linkes Bild. Die überlagerten Rechtecke deuten eine 16 × 8 Unterteilung an. Unten: Rechtes Bild.

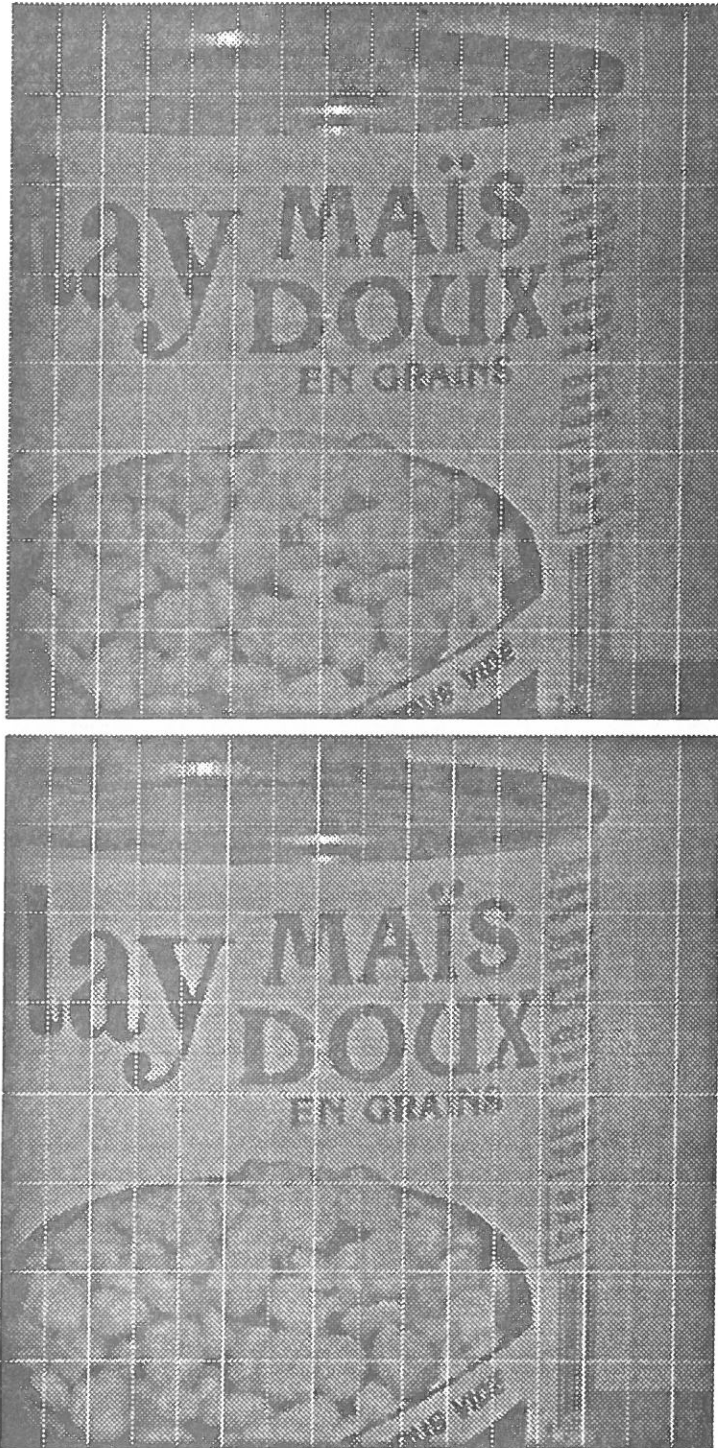


Abbildung 8.10: Kamerabildpaar „Maysdose“ mit 512×512 Pixel Auflösung und 256 Graustufen. Oben: Linkes Bild. Die überlagerten Rechtecke deuten eine 16×8 Unterteilung an. Unten: Rechtes Bild.

Kapitel 9

Zusammenfassung und Ausblick

In dieser Arbeit haben wir ein Verfahren kennengelernt, mit dem bei fixierender Stereoanordnung Tiefeninformation in einem lokalen Ausschnitt um den Fixationspunkt gewonnen werden kann.

Wir haben uns zunächst mit dem Aufbau des menschlichen Sehsystems beschäftigt und dabei erfahren, daß visuelle Informationen auf drei spezifischen, getrennten Kanälen von der Retina zur primären Sehrinde gelangen: Speziell gilt, daß ein Kanal nur Informationen über Bewegung und stereoskopische Tiefe überträgt. Dieser Pfad hat sich in entsprechenden Untersuchungen von M. Livingstone [Liv88, LH88] als weitgehend farbinsensitiv gezeigt und er besitzt eine vergleichsweise grobe Ortsauflösung. Von den beiden weiteren Kanälen überträgt der eine Farbkontrastkonturen mit hoher Ortsauflösung und der andere Informationen über Farbe und Grauschattierungen, allerdings mit recht grober Auflösung.

Diesen drei Informationspfaden ist jedoch eines gemeinsam: Auf der Retina benachbarte Daten bleiben auch auf dem ganzen Weg bis zur primären Sehrinde benachbart [Hub88], so daß man sich vereinfacht gesagt vorstellen kann, daß die primäre Sehrinde ein weitgehend unverzerrtes Bild von dem erhält, was auf der Retina abgebildet wird. Mit einer wichtigen Ausnahme allerdings: Die Informationen des linken und rechten Auges sind dort abwechselnd in dünnen Streifen nebeneinander abgelegt.

Aufbauend auf dieser markanten Streifenarchitektur [LHW75, HF77, Too82, Sch86, Bra85], deren genaue funktionale Bedeutung allerdings bislang noch ungeklärt ist, wurde ein technisches Verfahren vorgestellt, daß sich diese Architektur zunutze macht, um Tiefeninformation zu gewinnen. Dabei wird die bekannte Tatsache ausgenutzt, daß die Bilder auf der linken bzw. rechten Retina lokal – also in der Umgebung eines retinalen Ortes – sich nur durch eine geringfügige Verschiebung (Querdisparität) unterscheiden, jedenfalls sofern man von speziellen Problemen durch Verdeckung, unterschiedlicher Linsen im linken und rechten Auge und Ähnlichem absteht, und außerdem die schon historische Erkenntnis berücksichtigt, daß genau diese

Verschiebung direkt mit der relativen Tiefe der abgebildeten Objektoberfläche in Bezug auf die Tiefe am Fixationspunkt zusammenhängt.

Von der Vorstellung hypothetischer disparitätssensitiver Zellen ausgehend, die jeweils einen kurzen Abschnitt zweier benachbarter Streifen – also ein Doppelsignal – als Eingangssignal erhalten, haben wir ein technisches Modell zur Tiefenbestimmung kennengelernt. Wir haben Vielzahl biologischer Größen zusammengestellt, und im Zusammenhang ausgewertet, um günstige Parameter für das technische Verfahren festzulegen. Es ist interessant, daß sich analog zu den in dem Modell unterstellten disparitätssensitiven Zellen auch in entsprechenden neurophysiologischen Untersuchungen der Feinstruktur der primären Sehrinde in dem nachgeschalteten visuellen Feld V2 Zellen auffinden lassen, die für binokulare Tiefe empfindlich sind.

Eine mathematische Überlegung hat uns zuvor nahegebracht, daß ein spezieller nichtlinearer Filter (Cepstrum) [BHT63] auf ein Doppelsignal in einer Weise wirkt, die eine einfache und unproblematische Extraktion der Verschiebung der beiden Teilsignale des Doppelsignals erlaubt, wenn diese Signale vor der Filterung geeignet angeordnet werden. Diese Idee bildet die Grundlage für den Algorithmus von Y. Yeshurun und E. L. Schwartz [YS89, YS90] zur Disparitätsbestimmung.

Das bisherige Verfahren wurde für verschiedene (technische) Varianten implementiert und untersucht. Bei der Beschreibung der Implementation und speziell der Transformation in den Frequenzraum wurde darauf hingewiesen, daß es günstiger ist, statt der üblichen Fouriertransformation, die reellwertige und für den eindimensionalen Fall um den Faktor zwei schnellere Hartleytransformation [Bra86b] zu verwenden.

Die Ergebnisse von Y. Yeshurun und E.L. Schwartz bezüglich der Robustheit und Stabilität des Cepstrums konnten teilweise bestätigt werden. Obwohl die Autoren ihren Algorithmus für beliebig geformte Trägerfunktionen definiert haben, wird von ihnen nur der Fall quaderförmiger Trägerfunktionen vorgestellt. Auch die übrige Literatur berichtet weitgehend nur von dieser Art der Trägerfunktion. Durch diese spezielle Wahl ergeben sich bei der Verschiebungsbestimmung spezifische Probleme, wenn die Teilbilder beispielsweise gerade Kantensegmente enthalten oder unterschiedliche mittlere Helligkeiten besitzen.

Wir haben gesehen, daß diese Probleme abgemildert bzw. gelöst werden können, wenn man einer neurophysiologischen Feinstruktur (*pale bands*) eine funktionelle Bedeutung zuweist. Darauf aufbauend haben wir außerdem i.a. günstigere Trägerfunktionen kennengelernt bei denen die Information über die spezielle Bildung des Doppelsignals im Doppelsignal erhalten bleibt. Es wurden weitere Verbesserungsvorschläge dargelegt; unter anderem ist eine LoG-Filterung der Rohdaten vorteilhaft.

Bei den Betrachtungen zur Geometrie haben wir die näheren Zusammenhänge zwischen Querdisparität und relativer Tiefe erläutert. Die Abbildungsverhältnisse – speziell bei lokaler Oberflächenapproximation – wurden schrittweise anhand entsprechender Formeln für runde und ebene Abbildungsfläche entwickelt und durch

Graphiken veranschaulicht. Dadurch wurde z.B. klar, daß die in die Cepstrumsfilterung eingehenden Punkte recht unterschiedliche Verschiebungen innerhalb des Bereiches der Teilsignale besitzen können. Die Verschiebungsbestimmung liefert also eine mittlere Verschiebung aller Punkte in einem Teilsignal, die bei zu großem Durchmesser der Trägerfunktionen nur ein unbrauchbarer Mittelwert sein kann. Es wurde weiter deutlich, daß außerhalb eines fovealen Ausschnittes um dem Fixationspunkt, außer der Neigung der Objektoberfläche auch ihre spezielle Form die Verschiebung der Punkte erheblich beeinflußt.

Wir haben uns abschließend davon überzeugt, daß die mit dem vorgeschlagenen Verfahren berechneten lokalen Disparitätskarten bei Ausschnitten künstlicher Bilder, die durch einen raytracer erzeugt wurden und bei denen dadurch alle relevanten Parameter und damit auch die zugehörigen Disparitätskarten bekannt sind, bis auf Abweichungen von einem Pixel korrekt sind. Anhand zweier Beispiele wurde demonstriert, daß das Verfahren auch für Realweltbilder die gewünschten Ergebnisse liefert.

An die in dieser Arbeit vorgestellten Untersuchungen können sich noch folgende Betrachtungen anschließen.

- Das Cepstrum hat einen von den Trägerfunktionen zur Teilbildextraktion abhängigen Verfahrensfehler. Praktisch zeigt sich dieser Fehler unter vernünftigen Bedingungen nicht. Eine tiefere mathematische Untersuchung der Formeln für spezielle Trägerfunktionen könnte den Verfahrensfehler in Abhängigkeit des Trägers ermittelbar machen, so daß die Wahl und Parameterisierung der Fensterfunktionen auch theoretisch gestützt werden könnte.
- Die Sichtweise auf das Cepstrum als Autokorrelationsoperation mit adaptiver, bandpaßähnlicher Vorfilterung ist interessant. Zwar ist dieser Filter stark bildabhängig, aber vielleicht läßt sich das experimentell ermittelte Bandpaßverhalten bei einer tieferen Untersuchung der mathematischen Struktur des Filterkerns trotz dieser Tatsache untermauern oder für den allgemeinen Fall nicht bestätigen.
- Wir haben in dieser Arbeit eine Methode kennengelernt, schnell und robust lokale Tiefeninformation für eine einzelne fixierende Anordnung zu erhalten. Für ein aktives bildverstehendes System muß das Problem der Kombination und Integration benachbarter lokaler Tiefenschätzungen gelöst werden. Unterstellen wir das Ziel oder den Zweck, über ein gegebenes Objekt mehr zu erfahren, über das schon ein bestimmtes, aus einer Reihe von fovealen Teilausschnitten bestehendes Wissen vorliegt, so könnten die bereits vorhandenen lokalen Tiefenschätzungen zum Ableiten einer Vorzugsbewegungsrichtung oder einem neuen Fixationspunkt herangezogen werden.
- Die Tatsache, daß die diskrete 1D Hartleytransformation um den Faktor zwei schneller ist, als die entsprechende Fouriertransformation erzeugt den Wunsch

nach einer analogen Transformation für den zweidimensionalen Fall. Da die 2D Hartleytransformation wegen

$$\text{cas}(\alpha + \beta) \neq \text{cas}(\alpha) \cdot \text{cas}(\beta)$$

(mit $\text{cas}(x) \equiv \sin(x) + \cos(x)$) nicht separierbar ist, kann man die 2D Transformation nicht wie im Falle der Fouriertransformation nur durch 1D Transformationen aller Zeilen und anschließend aller Spalten berechnen. Es ist aber dennoch möglich die 2D Hartleytransformation auf einfache Weise auf 1D Hartleytransformationen zurückzuführen, so daß der Rechenzeitgewinn um den Faktor zwei abzüglich geringer *overhead*-Zeiten bestehen bleibt.

Anhang A

Bildquellennachweis

Soweit nicht anders angegeben sind die Abbildungen von mir mit dem Programm *idraw* aus dem X11 Toolkit *interviews* der Universität Stanford, Kalifornien erstellt worden oder mit eigenen Programmen bzw. MATHEMATICA berechnet worden. Die Beschriftung ist je nach Zeichnungstyp mit *idraw*, einzelnen \LaTeX -Kommandos oder gemischt hinzugefügt worden.

Abbildung 2.1: Elektronenmikroskopische Aufnahme aus der Zeitschrift *Spektrum der Wissenschaft* [SB87, S.116].

Abbildung 2.2: Links: In Anlehnung an die Zeichnung von M. Glickstein [Gli88, S.115]. Rechts: An die Abbildung aus dem Buch 'Eye, Brain and Vision' [Hub88, S.60] angelehnt.

Abbildung 2.3: Ebenfalls aus dem Buch von David Hubel [Hub88, S.97, S.107] entnommen.

Abbildung 2.4: Links: Verteilung der Okularen Dominanzstreifen aus dem Artikel von D. Hubel und D.C. Freeman [HF77, S.337]. Rechts: Abbildungsgeometrie bei Makaken aus dem Artikel von S. LeVay, D.H. Hubel und T.N. Wiesel [LHW75, S.570]. Oben: Zur Abbildung der Funktion $\log(z + 0.3)$: Der dargestellte Wertebereich (Halbkreise und Gitter) ist $z \in [0, 12] \times [-12, 12] \subset \mathbb{C}$ und der Bildbereich etwa $\log(z + 0.3) \in [-1.26, 2.54] \times [-1.7, 1.7] \subset \mathbb{C}$. Die Skizze im Anhang des Berichtes von E.L. Schwartz [Sch86] diene als Vorbild.

Abbildung 4.4: In Anlehnung an die Zeichnungen von V. Braitenberg [Bra85, S.387].

Abbildung 5.4: Die Isodisparitätslinien sind für $s_{x1} \approx -0.51974$ (Parabel) und $s_x \in [7a, -7a]$ paarweise in äquidistanten Schritten der Weite a mit $a = s_{x1}/10$ bei einer Stereoanordnung mit $t = 14, lx = -6.8, lz = 1.8$ berechnet.

Literaturverzeichnis

- [Alo90] John (Yiannis) Aloimonos. Purposive and Qualitative Active Vision. In *Proc. Image Understanding Workshop, Pittsburgh, PA, USA*, pp. 816–828, 1990. September 11-13.
- [AWB87] John Aloimonos, I. Weiss, and Amit Bandopadhyay. Active Vision. In *Proc. First Int. Conf. on Comp. Vision, London*, pp. 34–54, 1987.
- [Ban86] Amit Bandyopadhyay. *A computational study of rigid motion perception*. PhD thesis, University of Rochester (Department of Computer Science), 1986.
- [BF90] S.T. Barnard and M.A. Fischler. Computational and Biological Models of Stereo Vision. In *Proc. Image Understanding Workshop, Pittsburgh, PA, USA*, pp. 439–448, 1990. September 11-13.
- [BHT63] B.P. Bogert, M.J.R. Healy, and J.W. Tukey. The quefreny alalysis of time series for echoes: cepstrum, cross-cepstrum, and saphe cracking. In *in Proceedings: Symposium on Time Series Analysis*, pp. 209–243, 1963.
- [BJ80] Peter Burt and Bela Julesz. A Disparity Gradient Limit for Binocular Fusion. *Science*, Vol. 208, pp. 615–617, 1980.
- [Bla87] Richard E. Blahut. *Fast Algorithms for Digital Signal Processing*. Addison-Wesley, Reading, Mass., 1987.
- [Bra73] Valentin Braitenberg. *On the Texture of Brains*. Springer-Verlag, Berlin – Heidelberg – New York, 1973.
- [Bra85] Valentino Braitenberg. Charting the Visual Cortex. In A.Peters and E.G. Jones, editors, *Cerebral Cortex*, Vol. 3, Chapter 10, pp. 379–414. Plenum Press, New York and London, 1985.
- [Bra86a] Ronald N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, New York, 1986.
- [Bra86b] Ronald N. Bracewell. *The Hartley Transform*. Oxford University Press, 1986.

- [CM85] Eugene Charniak and Drew McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, Mass., 1985.
- [COB90] D.J. Coombs, T.J. Olson, and C.M. Brown. Gaze Control and Segmentation. Technical report, Department of Computer Science, Univ. of Rochester, 1990.
- [CSK77] D.G. Childers, D.P. Skinner, and R.C. Kemerait. The Cepstrum: A Guide to Processing. In *Proc. IEEE*, Vol. 65, pp. 1428–1442, 1977.
- [CT65] J.W. Cooley and J.W. Tukey. An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. of Comput.*, Vol. 19, pp. 297–301, 1965.
- [DA89] U.R. Dhond and J.K. Aggarwal. Structure from Stereo - A Review. *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 19, No. 6, pp. 1489–1510, 1989.
- [ENC90] ENCYCLOPÆDIA BRITANNICA. *The University of Chicago*. Peter B. Norton (President), Philip W. Goetz (Editor in Chief), and Robert P. Gwinn (Chairman; Board of Directors), Chicago, 1990.
- [Gen82] F. Gennari. *De Peculiari Structura Cerebri*. 1782. Gedruckt von Giambattista Bodoni.
- [Gli88] Mitchell Glickstein. Die Entdeckung der Sehrinde. *Spektrum der Wissenschaft*, S. 112–119, Nov 1988.
- [GW83] Rafael C. Gonzales and Paul Wintz. *Digital Image Processing (2.Ed)*. Addison-Wesley, Reading, Mass., 1983.
- [HB76] J.C. Hassab and R. Boucher. A probabilistic analysis of time delay extraction by the cepstrum in stationary gaussian noise. *IEEE Trans. on Information Theory*, Vol. 22, pp. 444–454, 1976.
- [HF77] D.H. Hubel and D.C. Freeman. Projection into the visual field of ocular dominance columns in macaque monkey. *Brain Research*, Vol. 122, pp. 336–343, 1977.
- [Hub88] David Hubel. *Eye, Brain and Vision*. Scientific American Library, NY, 1988.
- [Jä89] Bernd Jähne. *Digitale Bildverarbeitung*. Springer-Verlag, Berlin – Heidelberg – New York, 1989.
- [KC72] R.C. Kemerait and D.C. Childers. Signal detection and extraction by cepstrum techniques. *IEEE Trans. on Information Theory*, Vol. 18, pp. 745–759, 1972.

- [KH75] C.D. Kuglin and D.C. Hines. The Phase Correlation Image Alignment Method. In *Proc. Int. Conf. on Cybernetics and Society*, pp. 163–165, 1975.
- [Kor82] Axel Korn. Bildverarbeitung durch das visuelle System. In: *Fachberichte Messen · Steuern · Regeln*, Herausgeber: M. Syrbe und M. Thoma, Band. 8, Abschnitt 5.4, S. 161–167. Springer-Verlag, 1982.
- [KvD76] J.J. Koenderink and A.J. van Doorn. Geometry of Binocular Vision and a Model for Stereopsis. *Biological Cybernetics*, Vol. 21, pp. 29–35, 1976.
- [LH88] Margaret Livingstone and David Hubel. Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception. *Science*, Vol. 240, pp. 740–749, May 1988.
- [LHW75] S. LeVay, David Hubel, and Thorsten N. Wiesel. The Pattern of Ocular Dominance Columns in Macaque Visual Cortex Revealed by a Reduced Silver Stain. *Journ. Comp. Neur.*, Vol. 159, pp. 559–576, Jan 1975.
- [Liv88] Margaret S. Livingstone. Art, Illusion, and the Visual System. *Scientific American*, pp. 68–75, Jan 1988.
- [LKM88] Dah Jye Lee, Thomas F. Krile, and Sunanda Mitra. Power cepstrum and spectrum techniques applied to image registration. *Appl. Opt.*, Vol. 27, pp. 1099–1106, 1988.
- [LMK89] Dah Jye Lee, Sunanda Mitra, and Thomas F. Krile. Analysis of sequential complex images, using feature extraction and two-dimensional cepstrum techniques. *Jour. Opt. Soc. Am. A*, Vol. 6, No. 6, pp. 863–870, 1989.
- [LMK90] Dah Jye Lee, Sunanda Mitra, and Thomas F. Krile. Dense Depth Map from 2-D Cepstrum Matching of Image Sequences. In *Int. Workshop on Robust Computer Vision, Seattle, WA, USA*, pp. 200–207, 1990. October 1-3.
- [LS89] George F. Luger and William A. Stubblefield. *Artificial Intelligence and the Design of Expert Systems*. Benjamin/Cummings Publishing Company, 1989.
- [Lun73] J.S. Lund. Organisation of neurons in the visual cortex, area 17, of the monkey (*Macaca Mulatta*). *Journ. Comp. Neur.*, Vol. 147, pp. 455–496, 1973.
- [Mar82] David Marr. *Vision*. W.H. Freeman and Company, San Francisco, 1982.
- [Mö90] Bernhard Möhl. Die Flugsteuerung der Wanderheuschrecken. *Spektrum der Wissenschaft*, S. 66–75, Jul 1990.

- [MvSG90] H.A. Mallot, Werner von Seelen, and F. Giannakopoulos. Neural Mapping and Space-Variant Image Processing. *Neural Networks*, Vol. 3, pp. 245–263, 1990.
- [Neu86] Bernd Neumann. Vorlesung Bildverstehen. *Sommersemester*, 1986.
- [Nob90] Suga Nobuo. Neuronale Verrechnung: Echoortung bei Fledermäusen. *Spektrum der Wissenschaft*, S. 98–106, Aug 1990.
- [Nol64] A.M. Noll. Short-time spectrum and cepstrum techniques for vocal-pitch detection. *Journ. Acoust. Soc. Amer.*, Vol. 36, pp. 296–302, 1964.
- [Obe57] Fritz Oberhettinger. *Tabellen zur Fouriertransformation*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete, Band XC. Springer-Verlag, Berlin – Heidelberg – New York, 1957. Das Tabellenwerk enthält zwei große Teile mit analytischen sin- bzw. cos-Transformationspaaren.
- [OC90] T.J. Olson and D.J. Coombs. Real-Time Vergence Control for Binocular Robots. Technical Report 348, Department of Computer Science, Univ. of Rochester, 1990.
- [PMF85] Stephen B. Pollard, John E.W. Mayhey, and John P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, Vol. 14, pp. 449–470, 1985.
- [Ric83] Elaine Rich. *Artificial Intelligence*. McGraw-Hill, New York, 1983.
- [SB87] Julie L. Schnapf und Denis A. Baylor. Die Reaktion von Photorezeptoren auf Licht. *Spektrum der Wissenschaft*, S. 116–123, Jun 1987.
- [Sch86] E.L. Schwartz. Novel Architectures for Image Processing Based on Computer Simulation and Psychophysical Studies of Human Visual Cortex. AFOSR - Technical Report 86-0059, New York University Medical Center, 1986.
- [Too82] R.B. Tootel. Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, Vol. 218, pp. 902–904, 1982.
- [Tyl75] C.W. Tyler. Spatial organization of binocular disparity sensitivity. *Vision Research*, Vol. 15, pp. 583–590, 1975.
- [vdM73] Chr. von der Malsburg. Self-organization of orientation sensitive cells in the striata cortex. *Kybernetik*, Vol. 14, pp. 85–100, Jan 1973.
- [vdW88] Mark van de Wattering. **Programmpaket:** mtv-beta.shar. **Herkunft:** usenet (*comp.graphics*), 1988. markv@cs.uoregon.edu.

- [YS89] Yehezkel Yeshurun and Eric L. Schwartz. Cepstral Filtering on a Columnar Image Architecture: A Fast Algorithm for Binocular Stereo Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 759–767, Jul 1989.
- [YS90] Yehezkel Yeshurun and Eric L. Schwartz. Neural Maps as Data Structures: Fast Segmentation of Binocular Images. In Eric L. Schwartz, editor, *Computational Neuroscience*, Chapter 20, pp. 256–266. The MIT Press, 1990.