Untersuchung, Entwicklung und Implementierung von echtzeitfähigen Stereoverfahren in einem aktiven System

Diplomarbeit
am Lehrstuhl für Kognitive Systeme
des Instituts für Informatik und Praktische Mathematik
der Christian–Albrechts–Universität zu Kiel

vorgelegt von

Haye Hähne

Betreuer: Dipl. Phys. Michael Hansen
Prof. Dr. Gerald Sommer

Juli 1997

Zusammenfassung

In dieser Diplomarbeit wird die Entwicklung eines Verfahrens zur Ermittlung von Stereodisparität und Regelung der Vergenzbewegung eines Stereokamerakopfes vorgestellt, das als Komponente in ein aktives Bildverarbeitungssystem integriert werden kann.
Als Grundlage für die Entwicklung eines geeigneten Stereoverfahrens wird das Problem der Ermittlung von Disparität auf photometrischer Basis vorgestellt und in einer aktiven Philosophie interpretiert. Weil das Disparitätsproblem an sich eine schlecht gestellte Aufgabe darstellt, existieren vielfältige Lösungsansätze. In einer Literaturrecherche wird ein theoretischer Rahmen gesteckt, um einen strukturellen Vergleich von unterschiedlichen Stereoverfahren zu ermöglichen und einen Überblick über Grundprobleme und Lösungsstrategien des Disparitätsproblems zu gewinnen.
Da die Entwicklung des Stereoverfahrens laut Vorgabe auf der Korrelation oder dem Cepstrum-Filter basieren soll, werden die genauen Eigenschaften dieser zwei Ansätze untersucht. Je ein cepstrum- und korrelationsbasiertes Stereoverfahren werden in einer Standardvariante der Literatur implementiert und unter qualitativen Verbesserungen weiterentwickelt, wobei die Rahmenbedingungen eines aktiven Systems berücksichtigt werden. Diese betreffen insbesondere die Echtzeitfähigkeit des Stereoverfahrens, die durch die Implementierbarkeit auf dem vorliegenden Pipelinecomputer MV200 der Firma Datacube gegeben ist.
Die Standardverfahren und die darauf aufbauenden Eigenentwicklungen zur Messung von Stereodisparität werden anhand von synthetischen und realen Testbildern einem allgemeinen und einem qualitativen Vergleich unterzogen, auf dessen Grundlage die Auswahl des entwickelten korrelationsbasierten Stereoverfahrens für eine Implementierung auf dem MV200-System getroffen wird. Neben der Implementierung selbst wird auf Grundzüge des MV200-Systems eingegangen. Das Stereoverfahren findet schließlich in Verbindung mit einem Vergenzregler Einbettung in die Systemkomponente zur aktiven Disparitätsmessung.
Den Abschluß der Arbeit bildet eine Diskussion der betrachteten Stereoansätze basierend auf dem Cepstrum-Filter und der Korrelation. Theoretisch zeigt sich, daß der Cepstrum-Filter mathematisch eng mit der Korrelation verwandt ist, wobei der Cepstrum-Filter in Spezialfällen systematische Fehler produziert. Der Cepstrum-Filter scheint besser für eine reine Vergenzsteuerung als für ein umfassendes Stereoverfahren geeignet zu sein, während Korrelationsmaße nicht nur aufgrund der besseren Kenntnis ihrer Eigenschaften ein breiteres Anwendungsspektrum besitzen. Im Hinblick auf die Methodik eines Stereoverfahrens spielen beide Ansätze lediglich die Rolle eines photometrischen Vergleichsmaßes, während die Analysestrategie der betreffenden Vergleichsergebnisse einen wichtigeren Einfluß auf die Qualität der ermittelten Disparitätsinformationen besitzt.
Abschließend wird ein Ausblick auf konkrete Verbesserungen der untersuchten Stereoverfahren und auf andere aktuelle Stereostrategien gegeben.

Inhaltsverzeichnis

Kapitel 1

Einführung

Die gesuchte Systemkomponente soll die Aufgabe der Vergenzsteuerung und der Messung von Disparitätinformationen erfüllen.
Die Vergenzregelung kann als Spezialfall der Ermittlung von Stereodisparität ansehen werden, so daß in dieser Arbeit insbesondere die Entwicklung eines geeigneten Stereoverfahrens im Vordergrund steht.
Um den Grundstein für die Entwicklung des gesuchten Stereoverfahrens zu legen, folgt eine Beschreibung der klassischen Stereoproblematik sowie eine Interpretation der Problematik im Sinne einer aktiven Philosophie.

1.1 Beschreibung des Disparitätsproblems

Stereodisparität stellt für höhere Lebewesen unter anderen visuellen Hinweisen zur Wahrnehmung von räumlicher Tiefe aus Sterebildpaaren das wichtigste Merkmal dar. Deshalb ist die Gewinnung von Tiefeninformationen aus Stereobildern ein seit Jahrzehnten intensiv bearbeitetes Feld der Computer-Vision.

Mit *Disparität* wird die örtliche Verschiebung korrespondierender Bildstrukturen in einem Stereobildpaar (L,R) bezeichnet. Unter Korrespondenz der Bildstrukturen ist zu verstehen, daß sie auf die Projektion derselben physikalischen Szenestruktur in die Bildebene der linken und rechten Kamera zurückzuführen sind.
In Abbildung 1.1 ist dieser Sachverhalt am Beispiel der Verschiebung des Würfels zwischen L und R zu erkennen. Abhängig von der Stereoanordnung treten allgemein sowohl vertikale als auch horizontale Disparitäten auf.
Das klassische *Disparitätsproblem* besteht in der Aufgabe der vollständigen Detektion von Stereokorrespondenzen zwischen allen Pixeln aus L und R. Diese auch als Korrespondenzproblem bezeichnete Aufgabe der Zuordnung von Pixeln erfolgt ausschließlich auf Grundlage eines photometrischen Vergleichs von Grauwertstrukturen in L und R.
Als klassisches, universelles Korrespondenzproblem der Computer-Vision weist das Disparitätsproblem Querbezüge zu anderen Forschungsgebieten wie der Bild-

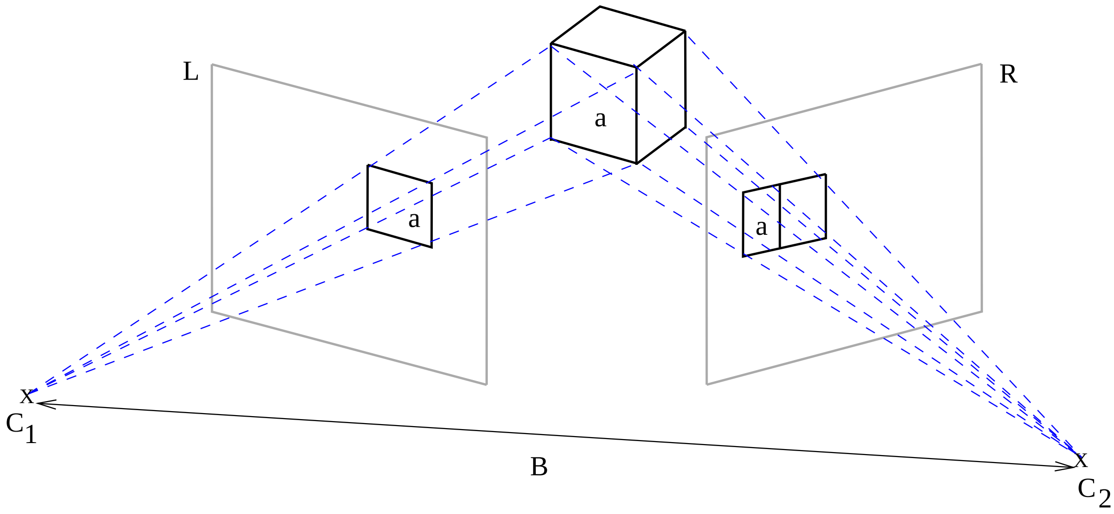


Abbildung 1.1: Durch Unterschiede in der Projektionsgeometrie von linker und rechter Kamera entstehen bei der Aufnahme eines Stereobildpaares (L,R) Disparitäten zwischen korrespondierenden Bildstrukturen in L und R.

folgenanalyse auf.
Typische Problemfälle der photometrischen Korrespondenzfindung werden In Abbildung 1.2 sichtbar. So existieren sowohl in L als auch in R Pixel ohne Korrespondenz im anderen Bild, beispielsweise in Bereichen von Verdeckungen von Bildstruktur (Bereich des Greifergelenks) oder aufgrund eines zu kleinen Überlappbereichs von L und R. Im Bereich des Robotergreifers besitzen korrespondierende Pixel in L und R durch Lichtreflexionen eine unterschiedliche Grauwertstruktur. In diesem Fall fällt die physikalisch begründete Korrespondenz nicht mit photometrischer Übereinstimmung zusammen. Dieser Effekt tritt ebenfalls im Zusammenhang mit Verzerrungen von Bildstruktur zwischen L und R auf, wie im Bildpaar im Fall des Klotzes deutlich wird. Verzerrungen sind ein typischer Bestandteil von realen Stereobildpaaren.
Zusätzlich können im Fall von periodischen Strukturen in L und R Alias-Probleme eine photometrische Zuordnung unmöglich machen, während das Aperturproblem in einigen Fällen eine eindeutige Zuordnung nur senkrecht zu einer Grauwertkante zuläßt. Fehlt bei der Zuordnung eines Pixels Bildstruktur im Sinne eines ausgeprägten Grauwertgradienten gänzlich, ist die Zuordnung überhaupt nicht möglich. Sie ist demnach grundsätzlich strukturgebunden.
Aus den geschilderten Schwierigkeiten wird deutlich, daß das Disparitätsproblem von der Aufgabe her schlecht gestellt ist.
Dieser Überlegung folgend stellt sich die Frage nach einer angemessenen Modellierung des Disparitätsproblems durch ein Stereoverfahren. In welchem Umfang sollen Problemfälle der Zuordnung betrachtet werden ? Beispielsweise kann in ei-

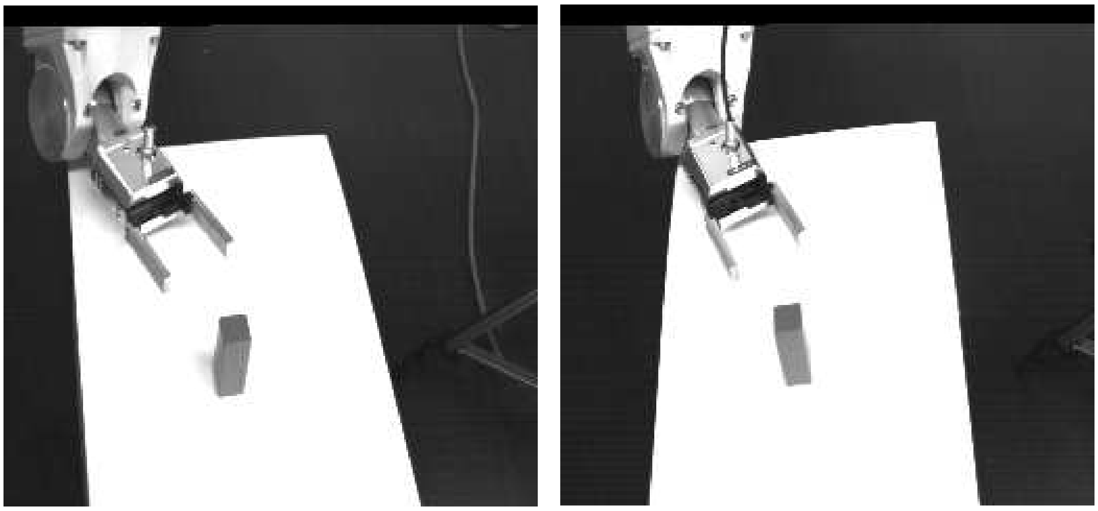


Abbildung 1.2: Bei der Aufnahme eines Stereobildpaars können durch die Projektion Verzerrungen und Verdeckungen von Bildstrukturen zwischen den Stereobildern auftreten, wie am Beispiel des Klotzes und des Greifergelenks deutlich wird. Diese und weitere Probleme machen die Korrespondenzsuche auf photometrischer Basis zu einer schlecht gestellten Aufgabe.

nem Stereoverfahren explizit nach Verdeckungen oder Glanzlichtern gesucht werden (siehe [Ching93]). Diese Vorgehensweise entspricht einem Regularisierungsansatz der schlecht gestellten Aufgabe des Disparitätsproblems.
In einem Stereoverfahren wird die Modellierung einen Kompromiß zwischen Komplexität und Laufzeit des Algorithmus darstellen, weil eine exakte Lösung des Disparitätsproblems allgemein nicht existiert.
Die Modellierung definiert die in einem Stereobildpaar vorhandenen Pixelkorrespondenzen, die ein Stereoverfahren möglichst vollständig und genau ermitteln sollte.
Ein Stereoverfahren liefert als Ausgabe eine *Disparitätskarte*, in der alle getroffenen Zuordnungen vermerkt sind.

Aus der Disparitätskarte lassen sich relative und bei Kenntnis der Kameraparameter durch Triangulation auch absolute Tiefeninformationen fur Pixel eines Stereobildpaars bestimmen. Ermittelte Tiefeninformationen sind nur im Nahbereich genau, weil die Disparität reziprok von der Tiefe abhängig ist. Fehler in der Disparitätsmessung fallen so bei einer großen Tiefe und folglich geringer Disparität stärker ins Gewicht als im umgekehrten Fall.

1.2 Das Disparitätsproblem im aktiven System

Das Disparitätsproblem in der obigen Beschreibung soll nun in einer aktiven Philosophie interpretiert werden. Die entstehende Sichtweise des Korrespondenzproblems weicht dabei in einigen Punkten von der klassischen Interpretation ab und gibt Rahmenbedingungen zur Modellierung des Disparitätsproblems durch ein Stereoverfahren vor, das für die Systemkomponente gesucht wird.

Eine Abschwächung soll das Disparitätsproblem hinsichtlich der Dichte der gesuchten Disparitätsinformationen erfahren. In einem aktiven System besteht nicht nach klassischem Vorbild das Ziel einer 3D-Rekonstruktion der kompletten Szene auf Basis einer dichten Disparitätskarte. Es sollen vielmehr aufgabenbezogene Disparitätsinformationen ermittelt werden. Fehlende Informationen sollen durch Aktivität des Systems gewonnen werden.
Deshalb muß das gesuchte Stereoverfahren keine aufwendigen und komplexen Strategien enthalten, um auf photometrischer Basis nicht durchführbare Zuordnungen von Bildstruktur wie beispielsweise im Fall von Verdeckungen oder Reflexionen zu detektieren und aufzulösen. Stattdessen soll eine im allgemeinen schwach besetzte Karte von sicheren Disparitätsinformationen erzeugt werden. Durch die Ausgabe eines Vertrauensmaßes für die getroffenen Zuordnungen sollen photometrische Problemfälle lediglich identifiziert werden können. Werden Disparitätsinformationen aus Bildbereichen benötigt, die durch das Vertrauensmaß gering bewertet werden, so muß ein übergeordneter Prozeß das System in Aktivität versetzen, so daß in einer Regularisierungsstrategie verläßliche Zuordnungen getroffen werden können. Im Fall einer Verdeckung beispielsweise kann diese Strategie durch eine Translationsbewegung des Kamerakopfes zum Erfolg führen. Diese Strategie findet sich auch in der Literatur wieder ([Fua93], [Vlee91]).
Eine aktive Vereinfachung des Disparitätsproblems entsteht durch eine dem gesuchten Stereoverfahren vorausgehende Vergenz der Kameras auf eine ROI, die durch die Blickrichtung der Stereokameras von einem übergeordneten Prozeß des aktiven Systems vorgegeben ist. Im Zustand der Vergenz werden Disparitäten im Bereich der ROI minimal.
Diese Vergenz ist biologisch motiviert. Der Mensch vergiert auf Dinge, die ihn interessieren. Das hat zum einen die Ursache, daß die Auflösung der Retina nicht gleichmäßig ist, und relevante visuelle Informationen in der Fovea liegen müssen, um sie zu analysieren. Zum anderen kann die menschliche visuelle Wahrnehmung erwiesenermaßen Tiefeninformationen nur aus relativ kleinen Disparitäten schöpfen ([Olson88]).
Durch die Minimierung der im Stereobildpaar vorkommenden Disparitäten kann in dem gesuchten Stereoverfahren die Korrespondenzsuche für Pixel in L und R eingeschränkt werden. Die Komplexität des Korrespondenzproblems sinkt.
Eine weitere Abweichung vom klassischen Disparitätsproblem ergibt sich aus der stärkeren Gewichtung der Ermittlung von relativen im Verhältnis zu absoluten

Tiefeninformationen. In einem aktiven System können relative Tiefeninformationen je nach gestellter Aufgabe ebenso wichtig sein wie absolute.
Diese Tatsache kann als Motivation für die Beschränkung auf die Ermittlung von horizontalen Disparitäten gewertet werden, die zur Gewinnung relativer Tiefeninformationen ausreichend sind, denn vertikale und horizontale Disparitäten hängen linear voneinander ab.
Ein weiteres Argument für diese in der Literatur standardmäßige Beschränkung auf die Detektion von Querdisparität liegt in der geringen Basislinie des vorliegenden Stereokamerakopfes, die eine vergleichsweise geringe Vergenz der Kameras erfordert. Relevante vertikale Disparitäten treten nur in Verbindung mit hohen Vergenzwinkeln auf (siehe Abschnitt 2.2.1).
In Abschnitt 2.2.1 wird deutlich, daß Epipolargeraden eine allgemeingültige Beschränkung für die Korrespondenzsuche in Stereobildpaaren darstellen. Dies ist eine weitere Rechtfertigung für eine eindimensionale Suche zur Ermittlung von Stereodisparität.
Die Beschränkung auf die Ermittlung von Querdisparität vereinfacht das kombinatorische Problem der Korrespondenzfindung erheblich. Dies betrifft neben Aspekten der Qualität von gemessenen Disparitätsinformationen insbesondere Aspekte der Laufzeit eines resultierenden Stereoverfahrens, die in der aktiven Sichtweise relevant ist. Ein aktive System legt seinen Einzelkomponenten Anforderungen an Echtzeitfähigkeit auf, weil es ständig im Dialog mit seiner Umwelt steht. Stereoinformationen müssen in einer vorgegebenen Zeitspanne, die die spezifische Aufgabenstellung im aktuellen Systemzustand vorgibt, ermittelt werden können. Diese Forderung deckt sich mit der aktuellen Definition von Echtzeitfähigkeit (DIN 44300/DIN 85):
"Echtzeitbetrieb ist ein Betrieb eines Rechensystems, bei dem Programme zur Verarbeitung anfallender Daten ständig betriebsbereit sind derart, daß die Verarbeitungsergebnisse innerhalb einer vorgegebenen Zeitspanne verfügbar sind. Die Daten können je nach Anwenderfall nach einer zufälligen zeitlichen Verteilung oder zu vorbestimmten Zeitpunkten anfallen." ([Knie91])
Orientiert man sich am biologischen Vorbild, so sollte das aktive System in einer festen, schnellen Taktrate Disparitätskarten durch das Stereoverfahren geliefert bekommen, die an die Videotaktrate der Bildaufnahme gebunden ist.

Das vorliegende Bildverarbeitungssystem, in das die Systemkomponente eingebettet werden soll, besteht zunächst aus dem Stereo-Kamerakopf "Bisight" der Firma TRC. Der Kamerakopf besitzt vier Freiheitsgrade, nämlich eine Drehung um die horizontale (*tilt*) und vertikale (*pan*) Achse des Kopfes sowie eine Drehung der einzelnen Kameras um eine vertikale Achse (*linke, rechte Vergenz*). Die Kameras besitzen einen Abstand (*Basislinie*) von 25 cm.
Weiterhin existiert zur Ansteuerung des Kamerakopfes eine Workstation mit Netzeinbindung. Um die Echtzeitfähigkeit des Systems zu gewährleisten, ist ein spezieller Bildverarbeitungs-Computer vorhanden. Es handelt sich um die

Pipeline-Architektur MaxVideo200 (MV200) der Firma Datacube. Das MV200-System ist über einen VME-Bus mit der Workstation als Host-Computer verbunden. Die Stereobilder des Kamerakopfes bilden ein Eingangssignal des MV200-Systems.

Kapitel 2

Das Disparitätsproblem in der Literatur

Die Grundlage für das gesuchte Stereoverfahren sollen Strategien der Literatur bilden. Dabei besteht die Vorgabe, Verfahren basierend auf Korrelation und dem Cepstrum-Filter auf ihre Eignung hin zu überprüfen.

In Stereoverfahren der Literatur findet sich das Korrespondenzproblem in verschiedenen Fassungen wieder. Korrespondenzen können in Stereobildpaaren ermittelt werden (statische Stereoanalyse) oder in einer zeitlichen Bildfolge mit nur einer Kamera, in der Disparität durch Bewegung der Kamera entsteht (dynamische Stereoanalyse). Auch aus der Aufnahme und Auswertung von Bildern einer statischen Szene mit einer Kamera, aber variierender Beleuchtung der Szene (photometrische Stereoanalyse), kann ein Stereoverfahren bestehen ([KleKoSchlü96]). Daneben gibt es neuere Ansätze, die ein Stereobildpaar aktiv mit einer eindeutig zu identifizierenden Textur ausleuchten und die in die Szene projizierte Textur zur Ermittlung von Korrespondenzen nutzen.
Als universelle Problematik findet sich das Korrespondenzproblem in Verfahren der Bildfolgenanalyse zur Bestimmung des optischen Flusses wieder. Auch in einer Bildfolge ist es das Ziel, zwischen zeitlich benachbarten Bildern die Positionen von korrespondierenden Pixeln zu ermitteln ([Ov92]). Deshalb weisen einige Autoren ausdrücklich auf die Tauglichkeit ihrer Verfahren sowohl zur Ermittlung von Stereodisparität als auch zur Berechnung des optischen Flusses in der Bildfolgenanalyse hin ([Weng92]).
Das Korrespondenzproblem ist als universelle Problematik in den letzten Jahrzehnten intensiv erforscht worden. Zu seiner Lösung existieren eine Vielzahl von Ansätzen in der Literatur. Insgesamt fällt es schwer, diese Verfahren konsistent in Klassen einzuteilen, da sie im allgemeinen eine Mischung aus heterogenen Strategien darstellen. Schon in der Modellierung des Disparitätsproblems klaffen die Verfahren je nach Aufgabenstellung weit auseinander, weil es als schlecht gestellte Aufgabe keinen eindeutigen theoretischen Rahmen zur Modellierung vorgibt.

Ebenfalls der qualitative Vergleich von Stereoverfahren entbehrt einer allgemeingültigen Grundlage. Bei der Beurteilung wie bei der Entwicklung von Stereoverfahren nimmt das Testprinzip neben theoretischen Modellüberlegungen eine wichtige Rolle ein. Dabei erschwert allein die Sensibilität vieler Stereoverfahren bezüglich von Eingabeparametern einen einheitlichen Vergleich. So existieren eine Vielzahl gleichberechtigter Lösungsverfahren.
Soll nun für eine konkrete Anwendung aus den vielen in der Literatur vorgestellten Stereoverfahren (nach [Kosch93] über 200 Artikel zu diesem Thema bis 1993) ein passendes ausgesucht werden, so gibt es dafür kein verläßliches Kriterium. Dies gilt insbesondere für korrelations- und cepstrumbasierte Stereoverfahren, die betrachtet werden sollen.

Um Kriterien für die Auswahl von Strategien der Literatur für das gesuchte Stereoverfahren zu gewinnen, soll ein theoretischer Rahmen gesteckt werden, der einen strukturellen Vergleich verschiedener Ansätze ermöglicht.
Zu diesem Zweck werden typische Komponenten von Stereoverfahren der Literatur vorgestellt. Die Komponenten beinhalten elementare Probleme und Lösungsstrategien der photometrischen Korrespondenzsuche.

2.1 Vergleichsmaße

Die Grundlage für eine Suche nach Pixelkorrespondenzen in einem Stereobildpaar (L,R) bildet in Verfahren der Literatur ein Maß zum Vergleich von Bildstrukturen.
Wichtige Vergleichsmaße der Literatur für Grauwert- oder Farbbilder sowie Bedingungen für ihre Anwendung sollen vorgestellt werden.

2.1.1 Arten von Vergleichsmaßen

Korrelation, quadratischer Fehler

Diese Maße stellen Matching-Verfahren für Templates $T_L(i,j)$ und $T_R(k,l)$ um Aufpunkte $L(i,j)$ und $R(k,l)$ dar. Die Templates werden aus L und R durch Anwendung eines Fensteroperators gewonnen und besitzen im allgemeinen die Form einer quadratischen Maske.
Ein hoher Korrelationswert sowie ein niedriger quadratischer Fehler zweier Templates sind ein Maß für hohe photometrische Übereinstimmung.
Die Maße können als klassische Vergleichsmaße eingestuft werden. Ein Beispiel für korrelationsbasierte Verfahren bietet [Ching93]. Die Eigenschaften der Korrelation als Vergleichsmaß werden unten näher betrachtet.

Cepstrum-Filter

Der Cepstrum-Filter vergleicht zwei Signale L,R global im Frequenzraum. Er beinhaltet jedoch im Vergleich zur Korrelation eine Vorfilterung von L und R, die das Signal-Rausch-Verhältnis der Filterung verbessert.
Die Antwort des Cepstrum-Filters besitzt die Dimension der Eingabebilder und bedarf einer Analysestrategie, um ein Vergleichsmaß für Bildstruktur in L und R abzuleiten.
Das Cepstrum-Verfahren stammt ursprünglich aus der Sonar-Technik ([OlCo91]) und wird zur Detektion von Echos in eindimensionalen Signalen verwendet. Erst seit jüngerer Zeit wird das Verfahren in einer zweidimensionalen Version zur Ermittlung von Stereodisparität verwendet.
Der Cepstrum-Filter existiert in einer komplexen und reellen Variante. Die reelle Variante wird unten näher betrachtet.

Phasenbasierte Maße

Diese Vergleichsmaße beruhen auf der Messung von lokalen Phasen mit einem analytischen Filterpaar in L und R. Aus der Differenz der lokalen Phase, die sich im Idealfall linear oder zumindest monoton verhält, kann direkt die Verschiebung von Bildstrukturen in L und R berechnet werden, ohne daß ein Matching durchgeführt werden muß. Die resultierende Disparitätskarte muß durch eine unterlegte Konfidenzkarte zusätzlich verifiziert werden.
Gerade phasenbasierte Verfahren gelten durch ihren direkten Algorithmus als geeignet für ein aktives System. Im Wesentlichen werden Gaborfilter ([Sa88], [Fle94] [JeJeTso89]) in ein oder zwei Dimensionen verwendet.

Kostenfunktionen

Globale Kosten- oder Energiefunktionen werden in unterschiedlichen Bereichen der Computer-Vision mit Erfolg verwendet ([HeiPeBou92]).
Eine Kostenfunktion ist global auf (L,R) definiert und besitzt im allgemeinen mehrere Anteile. Ein Anteil dient als photometrisches Vergleichsmaß, während weitere Anteile als Korrekturterme fungieren.
Mit den Korrekturtermen fließen Nebenbedingen (*constraints*) in das Vergleichsmaß ein, die Besonderheiten des Disparitätsproblems modellieren (siehe Abschnitt 2.2.1). Sie bewerten Zuordnungen von Pixeln in ihrem *Zusammenhang* und nicht *isoliert*. In den Korrekturtermen liegt so der Hauptunterschied zu obigen Vergleichsmaßen. Durch Einbeziehung von Nebenbedingungen für Zuordnungen wird eine flexiblere Modellierung des Disparitätsproblems erreicht.

Ein Beispiel sei mit der folgenden Kostenfunktion M gegeben, die eine Disparitätskarte D zu einem Stereobildpaar (L,R) der Dimension NxN bewertet. Die Kostenfunktion beinhaltet zunächst mit dem quadratischen Fehler (QF) einen

Anteil, der (L,R) nach einer pixelweisen Umordnung gemäß der Disparitätskarte vergleicht. Der Korrekturterm bewertet durch einen Ableitungsoperator $\triangle(D(i,j))$ kontinuierliche Disparitätsverläufe, wie sie typischerweise in Stereobildpaaren vorkommen, als positiv. Photometrischer Anteil und Korrekturterm werden in gewichteter Summe verknüpft :

$$M(L,R,D) = QF(L,R,D) + \lambda \sum_{i,j=0}^{N} \triangle(D(i,j))$$

Ein Vorteil von Kostenfunktionen bei einem Vergleich von Grauwertstrukturen liegt in der Auflösung des regelmäßigen Pixelrasters durch Miteinbezug der Disparitätskarte in den photometrischen Anteil. Bildstrukturen in (L,R) können so ohne die Verwendung von festen Templates einander angeglichen werden. Die Form eines Templates selbst nämlich besitzt einen ungeklärten Einfluß auf das Resultat eines Vergleichsmaßes. Weiterhin ist so eine Kostenfunktion besser für den Vergleich von verzerrten Bildstrukturen in L und R geeignet. (siehe Abschnitt 2.1.2).
Die Kostenfunktion im Beispiel kann in einem Optimierungsprozeß verwendet werden, der Bildstruktur von L an Bildstruktur von R unter minimaler topologischer Verformung von L anpaßt (siehe Abbildung 2.1). Die Wahl des Gewichtungsparameters λ ist wichtig, da photometrischer Anteil und Korrekturterm zum Teil ein gegenläufiges Funktionsverhalten besitzen.

Merkmalsbasierte Maße

Diese Maße sind auf aus Grauwertbildern abgeleiteten Merkmalen wie beispielsweise lokaler Orientierung definiert. Merkmalsbasierter Maße sind in L und R nur an informationsreichen Punkten definiert, wie beispielsweise an Orten lokaler Krümmung oder Kanten ([Weng92]). Auf diese Art wird versucht, die Menge der Eingabedaten L und R auf relevante Informationen zu beschränken.
Biologisch lassen sich merkmalsbasierte Maße durch die Tatsache motivieren, daß der Mensch in der Lage ist, aus *Random-Dot-Bildern* Tiefeninformationen zu extrahieren. Dort ist Struktur nur an wenigen, verstreuten Punkten kodiert ([Sta89]).
Das Problem eines optimalen Einzugsbereichs wird bei merkmalsbasierten Maßen auf die Vorverarbeitung verlagert (siehe Abschnitt 2.1.2).

Differentielle Verfahren

Differentielle Verfahren stammen aus der Bildfolgenanalyse. Ein Vergleich von Grauwertstruktur erfolgt dort in zeitlicher Richtung einer Bildfolge durch Bildung von örtlichen und zeitlichen Ableitungen. Bekannt ist das klassische Horn-Schunck-Verfahren (siehe [KleKoSchlü96]).

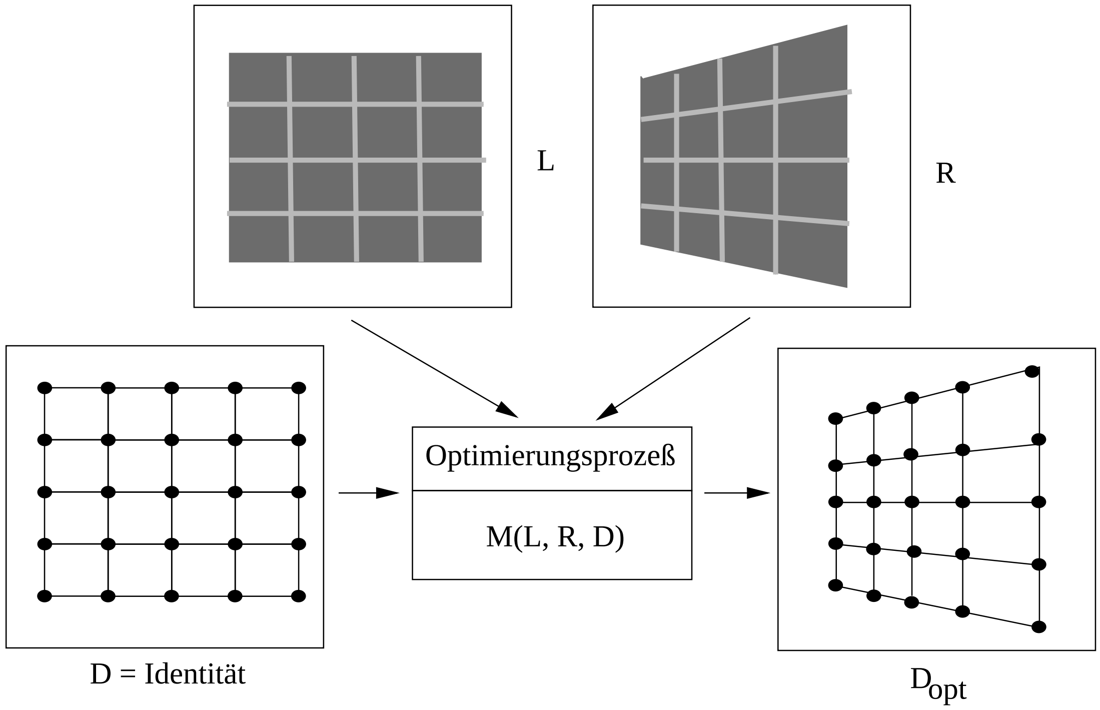


Abbildung 2.1: Eine Kostenfunktion M ermöglicht neben einem photometrischen Vergleich von Bildstruktur die Berücksichtigung von Nebenbedingungen für gesuchte Zuordnungen als Korrekturterme in einem globalen Vergleichsmaß. Beispielsweise kann M den globalen quadratischen Fehler von L und R bezüglich einer Disparitätskarte sowie die Stetigkeit der Disparitätskarte bewerten. Ein Optimierungsprozeß kann ausgehend von einer mit der Identität der Zuordnung initialisierten Disparitätskarte D nach der optimalen Disparitätskarte D_{opt} bezüglich M suchen.

Phasenkorrelation

Dieses auch als Dekonvolution bekannte Vergleichsmaß geht von einer konstanten Disparität von Stereosignalen aus, die unter Ausnutzung des Verschiebungstheorems der Fouriertransformation ermittelt wird. Eine Beschreibung der Phasenkorrelation findet sich in [Olson88].

Histogrammbasierte Maße

Unter Effizienzkriterien erweisen sich Maße als interessant, die Grauwertstrukturen auf Basis ihrer Histogramme vergleichen (siehe [KleKoSchlü96]). Durch den Bezug auf Histogramme ist der Rechenaufwand für einen Vergleich vom Einzugsbereich des Vergleichsmaßes entkoppelt, wenngleich auch der Übergang zu Histogrammen die Topologie von Bildstrukturen zerstört.

2.1.2 Der Einzugsbereichs eines Vergleichsmaßes

Wie für jeden Operator stellt sich auch für ein Vergleichsmaß die Frage nach einem passenden Einzugsbereich. Dabei wird der Einzugsbereich durch Wahl von Fensterfunktionen unterschiedlicher Ausdehnung verändert, die Bildstrukturen aus (L,R) extrahieren.
Betroffen sind Maße, deren Einzugsbereich nicht per Definition festgelegt ist. Bei anderen Maßen verlagert sich das Problem in Schritte der Vor- oder Nachverarbeitung von Signalen.
Nicht nur die Ausdehnung, sondern auch die Form der Fensterfunktion besitzt einen Einfluß auf die Vergleichsergebnisse. Die Fensterfunktion kann rechteckig, gaußförmig oder beispielsweise rund sein.
Im Folgenden soll jeweils der Vergleich eines Templates in einem Suchbereich betrachtet werden, aus dem eine Menge von Vergleichsergebnissen für die Suchpositionen resultiert (Template-Matching).

Optimaler Einzugsbereich

Man kann von der Annahme ausgehen, daß für jede Bildstruktur im Stereobildpaar ein optimaler Einzugsbereich für einen photometrischen Vergleich existiert. Im optimalen Fall liegen charakteristische Bildstrukturen im Einzugsbereich des betreffenden Templates, die eine eindeutige Zuordnung eines korrespondierenden Templates im Suchbereich ermöglichen.
Der Einzugsbereich sollte oberhalb von Rauscheinflüssen und periodischen Signalbestandteilen der Bildstruktur liegen. So werden lokale Extrema des Vergleichsmaßes im Suchbereich vermieden (Abbildung 2.2 a).
Auf der anderen Seite ist der Einzugsbereich so klein zu wählen, daß nur Regionen konstanter Disparität enthalten sind. Andernfalls können Strukturen am Rand des Templates, die Strukturen am Aufpunkt des Templates photometrisch

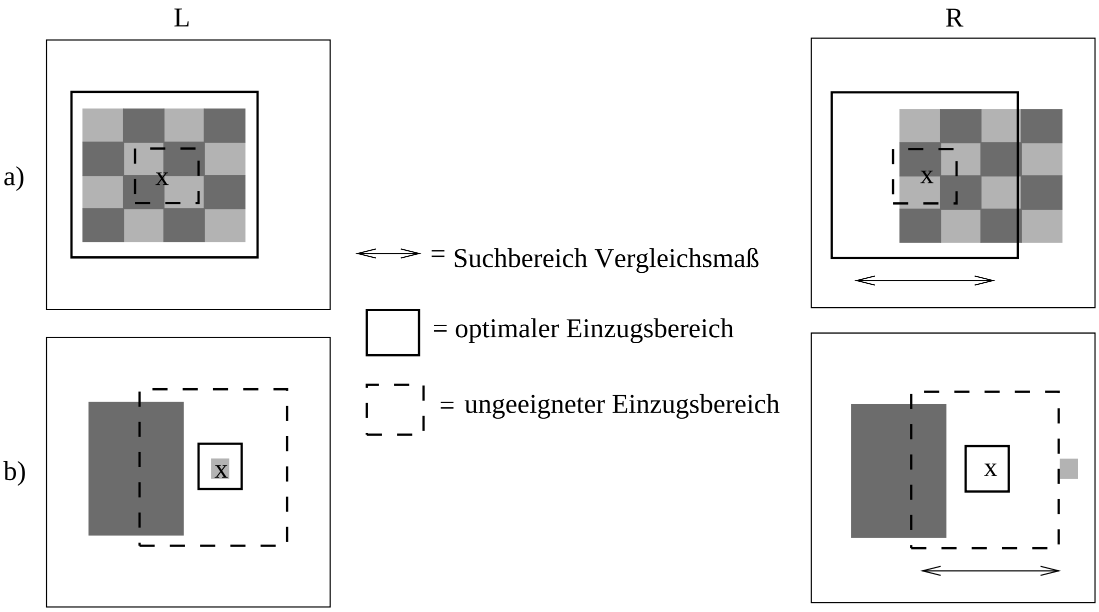


Abbildung 2.2: Für jede Bildstruktur in L und R läßt sich ein optimaler Einzugsbereich des Vergleichsmaßes festlegen, um eine eindeutige Korrespondenz zwischen L und R ermitteln zu können. Unterhalb des optimalen Einzugsbereichs können Alias-Fehler im Suchbereich entstehen (Fall a), darüber besteht die Gefahr der Beeinflussung von Vergleichsergebnissen durch Einbezug von Bildstrukturen in den Einzugsbereich, die eine stärkere photometrische Dominanz aufweisen als die Struktur am Aufpunkt (Fall b).

dominieren, die Vergleichsergebnisse beeinflussen. In einer anschließenden Korrespondenzsuche wird dem Aufpunkt dann die Disparität der dominanten Struktur zugewiesen.
In Abbildung 2.2 b) ist dargestellt, wie ein zu großer Einzugsbereich eines Templates zu einer falschen Zuordnung von Disparität führt, weil Bildstruktur, die nicht dem Aufpunkt zuzuordnen ist, das Matching dominiert. Ein zu großer Einzugsbereich führt im Grenzfall zu gleichförmig geglätteten Disparitätskarten.

Im optimalen Fall erfüllt der Einzugsbereich die beiden geschilderten Forderungen. In der Literatur gibt es mehrere Ansätze, um dem Problem eines optimalen Einzugsbereichs gerecht zu werden.
Einerseits kann man versuchen, für jede Bildstruktur vor einem photometrischen Vergleich den optimalen Einzugsbereich zu ermitteln ([OkKa92]). Alternativ kann man Vergleichsergebnisse in unterschiedlichen Einzugsbereichen generieren ([OlCo91], [LuBu93], [Ching93], [Weng92]) und die Suche nach einem optimalen Einzugsbereich auf eine Analyse der Vergleichsergebnisse verlagern.
Eine Pyramide bietet zu diesem Zweck den Vorteil, daß bei ihrer Erzeugung hochfrequente Signalanteile im Stereobildpaar unterdrückt werden, die für mehrdeutige Vergleichsergebnisse (Alias-Probleme) verantwortlich sind.

Verzerrungen im Einzugsbereich

Bei der Durchführung eines photometrischen Vergleichs besteht meist die Annahme, daß Bildstrukturen im Einzugsbereich nur gegeneinander verschoben, nicht aber gegeneinander verzerrt sind. Der erste Fall beinhaltet eine konstante Disparität im Einzugsbereich, während im zweiten Fall kontinuierliche Disparitätsverläufe auftreten. Verzerrungen von Bildstrukturen sind typisch für reale Stereobildpaare.
Um eine angenommene konstante Disparität zu gewährleisten, ist ein kleiner Einzugsbereich anzustreben, denn dort fallen Verzerrungseffekte weniger stark ins Gewicht.
Im Grenzfall des globalen Einzugsbereichs kann ein Vergleichsmaß undifferenzierbare Werte im Suchbereich liefern, die die Detektion kontinuierlicher Disparitätsverläufe unmöglich macht. Ein Beispiel hierfür findet sich in Abschnitt 3.3.3 im Fall des Cepstrum-Filters.

Vergleichbarkeit von Ergebnissen

Variiert man den Einzugsbereich zweier Templates, so liefert ein Vergleichsmaß im allgemeinen jeweils unterschiedliche Ergebnisse, weil neue Bildstrukturen in die Templates miteinbezogen werden.
Dieser Effekt ist bei der Analyse von Beträgen eines Vergleichsmaßes in unterschiedlichen Einzugsbereichen zu beachten.

Unabhängig von signalspezifischen Veränderungen tritt bei Variation des Einzugsbereichs eine Veränderung der Verteilung von Vergleichsergebnissen im Suchbereich auf, die auf generelle Signalcharakteristiken von realen Bildern zurückzuführen ist.
So werden die Ergebnisse eines Template-Matching bei einem kleinen Einzugsbereich eine größere Varianz im Suchbereich aufweisen als bei einem großen Einzugsbereich, bedingt durch die stärkere Glättungswirkung des Vergleichsoperators.

2.1.3 Normierung

Damit Vergleichsergebnisse ihrem Betrag nach nicht von der Signalenergie der Templates abhängen, besteht die Möglichkeit, diese auf ihre Signalenergie zu normieren und auf ein festes Intervall von Werten abzubilden.
Vergleichsergebnisse sind durch die Normierung nicht mehr von der Bildhelligkeit und dem Kontrast des Stereobildpaars abhängig.
Insbesondere ist bei einem photometrischen Vergleich unter Variation des Einzugsbereichs eine Normierung notwendig, falls die resultierenden Ergebnisse betragsmäßig verknüpft werden sollen, denn die Signalenergie eines Templates hängt vom Einzugsbereich ab.

2.2 Korrespondenzfindung als Optimierungsproblem

In Stereoverfahren basiert die Lösung des Korrespondenzproblems auf der Analyse von Ergebnissen eines Vergleichsmaßes in bestimmten Suchbereichen und Skalierungen. Durch die Suchbereiche für alle Pixel in L wird ein Raum aller möglichen Zuordnungen von L nach R aufgespannt.
Trifft man die Annahme, daß alle Pixel in L eine Korrespondenz in R besitzen, so läßt sich das Korrespondenzproblem als ein Optimierungsproblem von Zuordnungen bezüglich einer Zielfunktion M (siehe 2.1) zwischen L und R darstellen, die die Korrespondenz einer Zuordnung bewertet. Das Disparitätsproblem gibt Nebenbedingungen zur Einschränkung des Lösungsraums vor.
Gesucht ist eine optimale Zuordnung Z_{opt} im Lösungsraum, so daß $M(L, R, Z_{opt})$ ein Optimum einnimmt und Z_{opt} korrespondierende Zuordnungen enthält.

Die obige Modellierung des Disparitätsproblems ist hilfreich, um Stereoverfahren der Literatur in einen gemeinsamen theoretischen Rahmen einbetten zu können. Ein struktureller Vergleich wird so ermöglicht, auch wenn Stereokorrespondenz nicht exakt modelliert werden kann, da das Korrespondenzproblem an sich eine schlecht gestellte Aufgabe ist.
Die Modellierung zeichnet eine Kamera (L) als dominant aus, auf ihr Signal werden Zuordnungen bezogen. Es gibt in der Literatur auch eine allgemeinere

Definition von beiseitiger Zuordnung (siehe Abschnitt 2.2.1).
Das vorgestellte Optimierungsproblem ist nichtlinear, von hoher Komplexität und besitzt so keine konvexe Zielfunktion. In dieser Form kann es nur heuristisch bearbeitet werden. Es lassen sich allerdings zahlreiche Nebenbedingungen zur Einschränkung des Lösungsraums angeben, die sich aus dem speziellen Charakter des Disparitätsproblems ergeben.
In Verbindung mit einer linearen Zielfunktion kann man das Optimierungsproblem so weit vereinfachen, daß es mit probaten Algorithmen der Optimierungstheorie bearbeitet werden kann, zum Beispiel durch den Simplex-Algorithmus ([Vlee91]) oder dynamische Programmierung ([IntBo93]). Die Linearisierung des Disparitätsproblems stellt jedoch eine stark vereinfachte Modellierung dar.
Dominierend auf dem Gebiet der Suchstrategien zur Ermittlung einer optimalen Zuordnung sind deshalb Heuristiken, die strukturell an das Disparitätsproblem angepaßt sind und dessen Restriktionen ausnutzen. Sie liefern in der Praxis bessere Resultate und sind zudem schneller.
Viele Stereoverfahren stellen dem Optimierungsproblem Maßnahmen zur Regularisierung der schlecht gestellten Aufgabe der Korrespondenzsuche voran.
Im Folgenden sollen Beschränkungen des Lösungsraums, die in der Literatur der Vereinfachung des Optimierungsproblems dienen, sowie bestimmte Suchstrategien zur Ermittlung einer optimalen Zuordnung vorgestellt werden.

2.2.1 Beschränkung des Lösungsraums

Die Mächtigkeit des Lösungsraums aller möglichen Zuordnungen des obigen Optimierungsproblems ist extrem groß, denn eine Zuordnung von jedem Pixel in L zu jedem Pixel in R ist möglich. Offensichtlich gibt das Disparitätsproblem jedoch Nebenbedingungen für mögliche Zuordnungen vor, die aus der Stereogeometrie der Bildaufnahme resultieren und den Lösungraum einschränken.
Eine weitere Möglichkeit zur Reduzierung der Komplexität des Optimierungsproblems liegt in der Einschränkung des Definitionsbereichs möglicher Zuordnungen von L nach R durch Ausmaskierung von Zuordnungskandidaten.

Verdeckungen

In Folge von perspektivischen Verdeckungen oder nicht überlappenden Bildbereichen können Pixel in L nicht zugeordnet werden, weil korrespondierende Pixel in R nicht existieren. Solche Pixel sollen als *isolierte* Pixel bezeichnet werden.
In der Literatur gibt es Regularisierungsansätze zur Detektion isolierter Pixel ([Ching93], [LuBu93], [IntBo93], [Weng92]). Korrespondierende Zuordnungen werden unter Ausmaskierung isolierter Pixel auf einem somit eingeschränkten Definitionsbereich von L gesucht.
Zur Eliminierung isolierter Pixel gibt es in der Literatur verschiedene Strategien. Es werden beispielsweise sowohl Zuordnungen von L nach R als auch von R

nach L vorgenommen (*bidirectional matching* [Fua93]). Der Fall isolierter Pixel ist durch Differenzen in der beidseitigen Zuordnung detektierbar.
In einem korrelationsbasierten Verfahren wird beispielsweise die Qualität eines Template-Matching analysiert, um Rückschlüsse auf Verdeckungen zu ziehen ([Ching93]).
Liegt ein Optimierungsverfahren zugrunde, so können isolierte Pixel mit einem Kostenmaß belegt werden ([IntBo93]).

Stetigkeit der Disparität

In realen Stereobildern kann man von stetigen Disparitätsverläufen ausgehen. Aus diesem Grund werden in vielen Stereoverfahren kontinuierliche Disparitätsverläufe angenommen und somit mögliche Zuordnungen eingeschränkt.
Auf diese Tatsache hat sich auch die menschliche visuelle Wahrnehmung eingestellt, denn ihre Fähigkeit zur Gewinnung von Disparitätsinformationen aus einem Stereobildpaar ist durch den Gradienten des Disparitätsverlaufs beschränkt ([Kosch93]).

Epipolargeraden

Epipolargeraden E_L, E_R ergeben sich aus dem Schnitt einer Ebene E mit dem Neigungswinkel ϕ durch die Projektionszentren C_1, C_2 von linker und rechter Kamera einer Stereoanordnung mit L und R.
In Abbildung 2.3 ist der Verlauf von Epipolargeraden für eine bestimmte Ebene $E(\phi)$ dargestellt, wobei der Vergenzwinkel α der Kameras variiert wird. Ersichtlich ist, daß E_L und E_R für $\alpha = 0$ waagerecht verlaufen. Für $\alpha > 0$ erhalten E_l und E_r einen entsprechenden vertikalen Anteil.
E_L und E_R sind die Projektionen derselben Ebene $E(\phi)$ auf L und R. In E_L und E_R sichtbare Bildinformationen, die in $E(\phi)$ liegen, korrespondieren deshalb.
Zu einer gegebenen Epipolargeraden E_L von L genügt es deshalb, die Korrespondenzsuche in R auf E_R zu beschränken. Damit wird eine zweidimensionale Korrespondenzsuche auf eine eindimensionale reduziert.

Ordnungserhaltung

Liegen Bildstrukturen in einem Stereobild in einer bestimmten topologischen Ordnung zueinander vor, so findet sich diese Ordnung allgemein im anderen Stereobild wieder. Auf dieser Annahme beruhen einige Stereoverfahren, die Ordnungsbeziehungen für die Korrespondenzsuche im Lösungsraum ausnutzen ([Ching93]). Es lassen sich allerdings Fälle konstruieren, die die Annahme der Ordnungserhaltung verletzen ([Kosch93]).

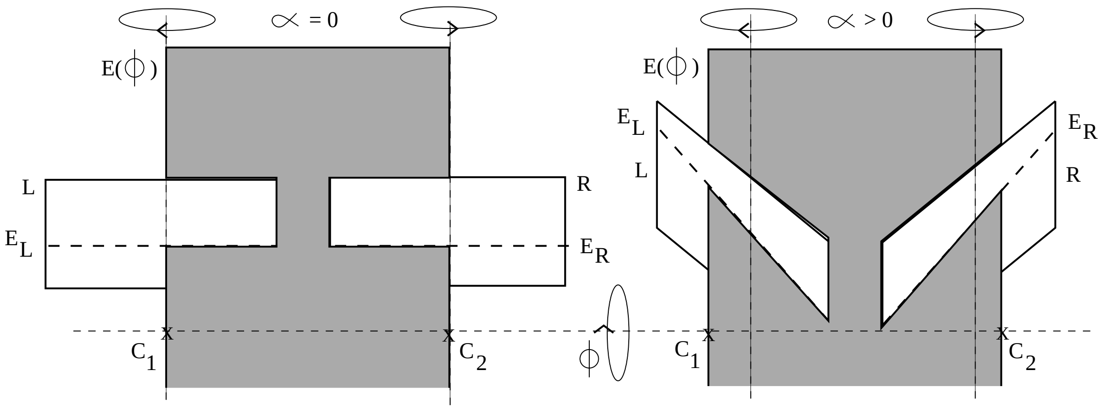


Abbildung 2.3: Epipolargeraden E_L, E_R werden durch die Kameraparameter festgelegt. Im Fall paralleler optischer Achsen linker und rechter Kamera ($\alpha = 0$) verlaufen sie horizontal, im Fall der Vergenz ($\alpha > 0$) schräg. Sie sind die Projektionen einer Ebene $E(\phi)$ durch die optischen Zentren einer Stereokamerakonfiguration auf L und R. E_L und E_R enthalten korrespondierende Bildstrukturen.

Annahmen über den Suchbereich

Für Pixel in L werden Annahmen über Bildbereiche zur Suche nach Korrespondenzen in R gemacht.
Dabei können große Suchbereiche zur Effizienzsteigerung in einer Pyramide verarbeitet werden. Zusätzlich wird die Dimension des Lösungsraums durch eine geringere Bildgröße reduziert.
Eine obere Schranke für den Suchbereich ergibt sich aus dem enstehenden Rand in R, in dem der Suchbereich die Bilddimension überschreitet.
Eine Beschränkung des Suchbereichs läßt sich ebenfalls biologisch motivieren ([Olson88], [Kosch93]).
Neben der Beschränkung von möglichen Zuordnungen wird durch den Suchbereich eine minimale Signalwellenlänge für Bildstrukturen vorgegeben, deren Disparität im Suchbereich eindeutig detektierbar ist. Nach dem Abtasttheorem liegt diese oberhalb des doppelten Suchbereichs (siehe auch [He97]). Mehrdeutigkeiten der Zuordnung entstehen durch hohe Frequenzen im Suchbereich mit zu großer Disparität.

Auswahl von Kandidaten

Ein Kriterium zur Einschränkung des Definitionsbereichs von möglichen Zuordnungen liegt in der Strukturgebundenheit von Disparitätsinformationen (siehe Abschnitt 1.1). So bietet sich eine Vorverarbeitung von L und R an, in der informationsreiche Punkte als Zuordnungskandidaten extrahiert werden ([IntBo93], [Gu90], [Weng92]). Merkmalsbasierte Verfahren verfolgen diese Strategie.

Ein Kriterium zur Einschränkung des Lösungsraums kann aus den Ergebnissen eines Vergleichsmaßes gewonnen werden, indem nur Positionen mit einem lokalen Extremum im Suchbereich betrachtet werden.

2.2.2 Suchstrategien

Nach der Einschränkung des Lösungsraums verfolgt jedes Stereoverfahren eine Suchstrategie zum Auffinden einer optimalen Zuordnung im Lösungsraum auf Grundlage von ermittelten Vergleichsergebnissen. Um dem Problem eines optimalen Einzugsbereichs gerecht zu werden (siehe Abschnitt 2.1.2), ist ein typischer Bestandteil der Strategie eine Skalensuche, die Vergleichsergebnisse in unterschiedlichen Einzugsbereichen analysiert.
In der Literatur orientiert sich die Skalensuche häufig an lokalen Extrema des Vergleichsmaßes in einer Skala, von denen im allgemeinen mehrere existieren. Schwellwertoperationen auf Vergleichsergebnissen sind kritisch zu bewerten, da sie einen zusätzlichen Eingabeparameter eines Stereoverfahrens darstellen.
Weiterhin ist die absolute Bewertung von Vergleichsergebnissen in einem schwer zugänglichen Rahmen vom jeweiligen Vergleichsmaß, dem gewählten Einzugsbereich und den vorliegenden Bildstrukturen in L und R abhängig (siehe Abschnitt 2.1.2).

Pyramidenverarbeitung

Pyramidenverarbeitung verbindet eine Skalensuche mit einer Effizienzsteigerung. Eine Gauß-Pyramide P besteht aus k Stufen, die sich aus einem Ursprungsbild durch einen Prozeß der sukzessiven Tiefpaßfilterung und Unterabtastung ergeben.
In einem Stereobildpaar (L,R) sind in der i. Stufe von P (mit P[1] = (L,R)) die Signale L und R der Dimension NxN auf die Dimension $\frac{N}{2^{i-1}}$x$\frac{N}{2^{i-1}}$ reduziert. Der Raum aller möglichen Zuordnungen erweitert sich im Falle der Pyramidenverarbeitung zu einen Lösungsbaum, wobei sich die Wurzel auf die k. Stufe bezieht. Im allgemeinen bildet eine Lösung für das Zuordnungsproblem der $i-1$. Stufe eine verfeinerte Lösung für das Problem der i. Stufe.
Die Ursache hierfür ist in der mit steigender Pyramidenstufen fortschreitenden Eliminierung hochfrequenter Signalanteile von L und R durch die Tiefpaßfilterung zu suchen, so daß in hohen Pyramidenstufen nur grobe Disparitäten gemessen werden können, weil auch nur grobe Bildstrukturen vorhanden sind.
Periodische Signalanteile im Suchbereich werden unterdrückt. Ein Vergleichsmaß kann so in höheren Pyramidenstufen ein eindeutiges Extremum liefern.

Bei der Realisierung einer Gauß-Pyramide ergeben sich jedoch einige Probleme, die teilweise systembedingt sind.

- Durch Quantisierungseffekte im Zusammenhang mit einer Gauß-Pyramide

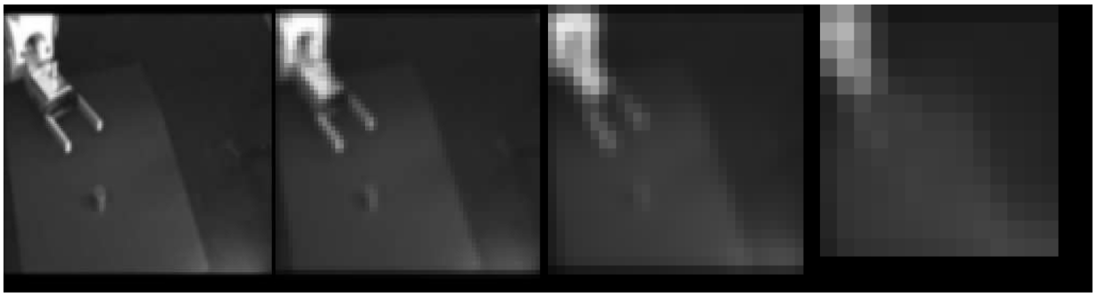


Abbildung 2.4: Durch Quantisierungseffekte können innerhalb einer Gauß-Pyramide feine Strukturen in hohen Pyramidenstufen gänzlich weggefiltert werden, wie im Falle des Klotzes sichtbar wird.

in einem digitalen Bildverarbeitungssystem ergibt sich ein relevantes Problem durch die vollständige Eliminierung von Bildstruktur in hohen Pyramidenstufen. In Abbildung 2.4 ist eine Szene mit einem Robotergreifer und einem Klotz dargestellt. In P(3) ist der Klotz vorhanden, während er durch Effekte der Tiefpaßfilterung und Quantisierung in P(6) nicht mehr vorhanden ist.

- Ein negativer Effekt der Diskretisierung bei der Bildung einer Gauß-Pyramide tritt ebenfalls auf.
  Eine Gauß-Pyramide setzt sich aus der k-maligen Hintereinanderausführung einer Gauß-Faltung G gefolgt von einem Unterabtastungsoperator $op_{sub[1,2,3,4]}$ zusammen :

  $$P(i) = op_{sub[1,2,3,4]}(G * P(i-1))$$

  Mit $op_{sub[1,2,3,4]}$ ergeben sich 4 verschiedene Arten der Unterabtastung (siehe Abbildung 2.5 a)).
  In Abbildung 2.5 b) ist zu sehen, daß op_{sub1} und op_{sub4} zu Pyramiden P(6), P'(6) von P(1) führen, die örtlich gegeneinander verschoben sind. Außerdem sind diese gegenüber P(1) verschoben, wie die Überblendung von jeweils P(1), P(6) und P(1), P'(6) zeigt.
  Stereoverfahren, die eine Pyramidenstruktur nutzen, stellen Beziehungen zwischen Pixeln aus P(i-1) und aus P(i) her. Durch den Effekt der Verschiebung von Bildstruktur durch Unterabtastung entsteht für diesen Prozeß ein systematischer Fehler.

- Faltungsoperationen etwa zur Berechnung eines Vergleichsmaßes oder zur Tiefpaßfilterung bei der Generierung einer Gauß-Pyramide erzeugen in hohen Pyramidenstufen erhebliche Randprobleme.
  Bei einer Dimension der Eingabebilder von 512x512 und einer Pyramide

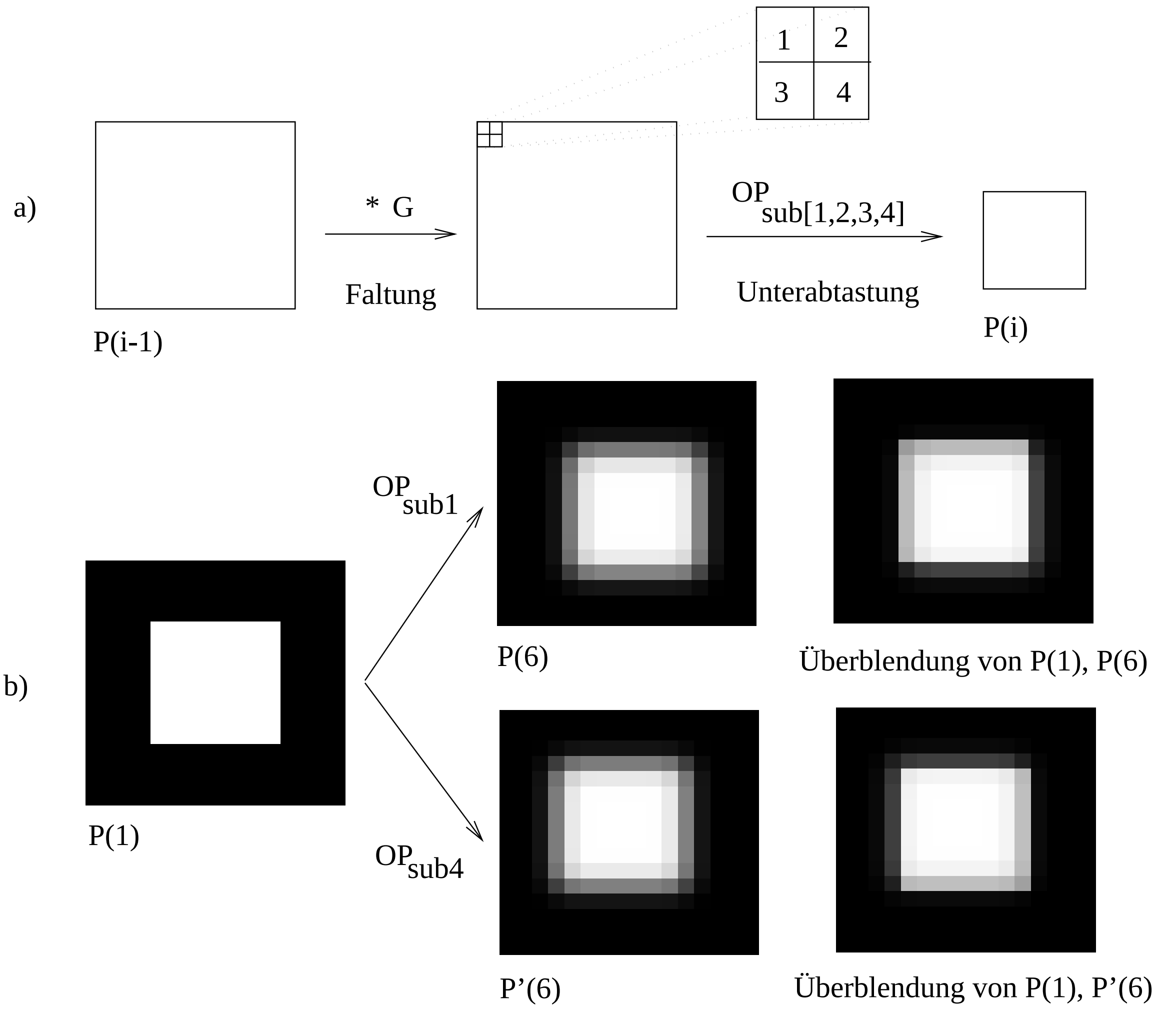


Abbildung 2.5: Ein negativer Effekt der Diskretisierung betrifft die Bildung einer Gauß-Pyramide. Durch die Entscheidung für eine Abtastposition während der Phase der Unterabtastung kann es zu einer systematischen Verschiebung von Bildstruktur zwischen den Pyramidenstufen kommen.

mit 5 Stufen (P(5) (32x32)) entsteht bei der Dimension einer Faltungsmaske von 9x9 Pixel bezogen auf P(1) ein 128 Pixel breiter Rand.

Eine Pyramidenverarbeitung zur Korrespondenzsuche ist in Stereoverfahren der Literatur ein probates Mittel.
Nach [HeiPeBou92] sind die genauen Effekte der Tiefpaßfilterung durch die Pyramide auf die Bildinformationen bezüglich einer Stereo-Korrespondenzfindung jedoch nicht auf theoretischer Grundlage untersucht worden. In der Literatur gibt es deshalb Vorbehalte gegenüber der Pyramidenverarbeitung, sie wird jedoch zumindest als notwendiges Mittel für eine hohe Effizienz angesehen.
Eine Alternative zur Gauß-Pyramide liegt in der Bildreduktion durch Zusammenfassung von einem Quadrat aus vier Pixeln in der Stufe i zu einem Pixel in der Stufe i+1. In dieser als *Quadtree* bezeichneten Pyramide kommt es zu Blocking-Effekten, sie reduziert jedoch analog zur Gauß-Pyramide hohe Frequenzanteile (Box-Filter statt Gauß-Filter).
Die Quadtree-Pyramide besitzt den Vorteil, bei der Generierung keine Randprobleme zu erzeugen. Vor allen Dingen tritt der Effekt der Verschiebung von Bildstruktur nicht auf, weil kein Subsampling-Operator verwendet wird. In der Vermeidung dieses systematischen Fehlers stellt die Quadtree-Pyramide eine ernstzunehmende Alternative zur Gauß-Pyramide dar.

Greedy-Strategie zur Korrespondenzsuche

In diesem Ansatz wird die Korrespondenzsuche im Lösungsraum für jedes Pixel isoliert durchgeführt, indem aus einem intuitiven Ansatz heraus die Position im Suchbereich mit extremaler Bewertung durch ein Vergleichsmaß ausgewählt wird. Diese einfache "Greedy-Strategie" besitzt den Nachteil, anfällig gegen Alias-Probleme zu sein.
In der Literatur wird das Alias-Problem deshalb üblicherweise durch eine *Coarse-to-Fine*-Skalensuche unterdrückt.
Die Strategie beinhaltet, in k Schritten ausgehend von einem großen hin zu einem kleinen Einzugsbereich des Vergleichsmaßes das Optimierungsproblem unter sukzessiver Verringerung des Suchbereichs nach obiger Greedy-Strategie zu lösen. Gewonnene Zuordnungen des i. Schritts dienen als Initialisierung für den Suchbereich des $i+1$. Schritts.
Analog wird im Fall einer Pyramidenverarbeitung von der k. Stufe ausgehend in der i. Pyramidenstufe eine Zuordnung generiert und mit ihr der Suchbereich der $i-1$. Stufe initialisiert.
Nach Abarbeitung der untersten Pyramidenstufe ist für jedes Pixel in einem Prozeß der sukzessiven Verfeinerung eine Zuordnung erzeugt worden, die genau einem von der Wurzel ausgehenden vollständigen Pfad im Baum aller möglichen Zuordnungen der Pyramide entspricht.
Da in einer Pyramide die Bilddimension geringer wird, kann mit steigender Py-

ramidenstufe auch der Suchbereich verringert werden, so daß der Suchbereich bezogen auf die höchste Auflösungsstufe konstant bleibt.

Stochastische Optimierung

Stochastische Optimierung beinhaltet das Durchsuchen des Lösungsraums nach einer optimalen Lösung mittels zufälliger sukzessiver Veränderung von Lösungen. Dabei werden in einem iterativen Prozeß temporäre Verschlechterungen der Zielfunktion in Kauf genommen.
Damit verwendet das Optimierungsverfahren eine Relaxationsstrategie, das heißt, eine gezielte Abschwächung des Optimierungsproblems.
Verfahren der stochastischen Optimierung stellen momentan das Mittel der Wahl dar, um komplexe Optimierungsprobleme von hoher Dimension wie das Korrespondenzproblem exakt zu lösen. Der Preis ist eine mit Echtzeitbedingungen unvereinbar hohe Rechenzeit.
Deterministische Optimierungsansätze scheitern daran, in lokalen Extrema der Zielfunktion zu terminieren oder innerhalb von suboptimalen Lösungen zu kreisen.
Simulated annealing beispielsweise ist ein Verfahren der Stochastischen Relaxation. Kern des Optimierungsverfahrens ist die Simulation des physikalischen Prozesses der Abkühlung von Molekülen in einem Temperaturmodell.
Diese Modellierung läßt sich auf das Disparitätsproblem übertragen ([Sta89], [Ba89]) und liefert in der Literatur anerkannt Disparitätsmessungen von hoher Qualitativ und Dichte.
Es sei angemerkt, daß es auch Ansätze für deterministische Relaxationsverfahren gibt, die nicht nur in Hinsicht auf Echtzeitforderungen anzustreben sind ([HeiPeBou92]).

2.3 Interpolation der Disparitätskarte

Aufgrund der Strukturgebundenheit einer photometrischen Korrespondenzsuche sind Disparitätskarten eines Stereoverfahrens nur an Orten von Struktur mit vertrauenswürdigen Zuordnungen besetzt.
Unter der Zielsetzung einer 3D-Oberflächenrekonstruktion wird in der Literatur eine Interpolation der Disparitätskarte an Orten mit geringer Vertrauenswürdigkeit der Zuordnung angestrebt ([Kosch93]). Diese Bereiche der Disparitätskarte sind auf strukturlose Bildregionen im Stereobildpaar zurückzuführen.
Eine Interpolation der Disparitätskarte ist jedoch ohne Apriori-Wissen nur auf heuristischer Basis möglich. Die Ausgabe eines Vertrauensmaßes bleibt erforderlich.

Ein typischer Interpolationsansatz beruht auf linearer Interpolation von jeweils benachbarten vertrauenswürdigen Disparitätsinformationen.

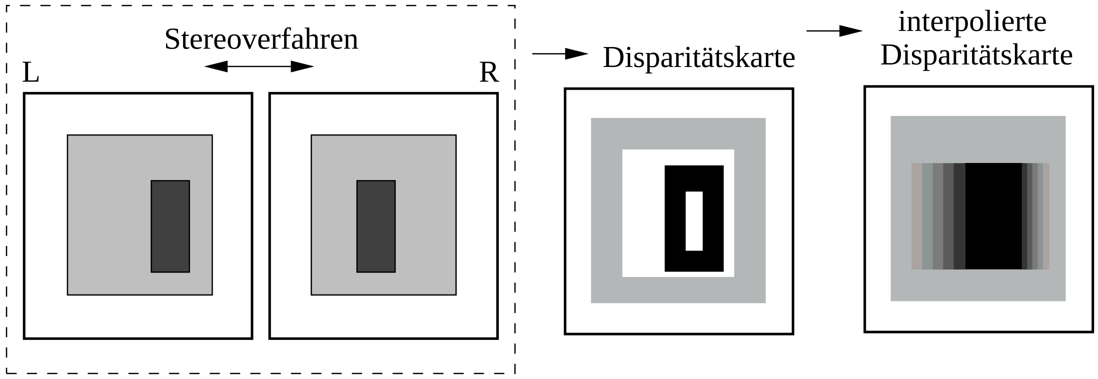


Abbildung 2.6: Die Interpolation von Disparitätskarten in strukturlosen Bildregionen bleibt eine Heuristik, wie das einfache Beispiel der linearen, horizontalen Interpolation zeigt. Nur durch Einbeziehung von Apriori-Wissen können bessere Interpolationsergebnisse erzielt werden.

Diese Strategie wird in Abbildung 2.6 verfolgt. In dem vorliegenden Stereobildpaar existieren in linkem und rechtem Bild zwei Strukturelemente, nämlich ein großes Rechteck und ein eingebettetes kleines Rechteck mit unterschiedlichen Disparitäten. Es ergibt sich eine schwach besetzte Disparitätskarte.

Interpoliert man strukturlose Regionen der Disparitätskarte in horizontaler Richtung linear durch angrenzende vertrauenswürdige Disparitätswerte, erhält man eine dicht besetzte Disparitätskarte, die jedoch in strukturlosen Regionen des großen Rechtecks falsche Werte aufweist.

Hinter dieser Interpolationsstrategie steckt die Annahme über einen kontinuierlichen Disparitätsverlauf (siehe Abschnitt 2.2.1).

An diesem einfachen Beispiel wird deutlich, daß eine Interpolation im allgemeinen keine verläßlichen Disparitätswerte in strukturlosen Bildregionen liefert.

In der Literatur wird die Interpolationsstrategie durch die Annahme über einen stückweise stetigen Disparitätsverlauf erweitert (siehe [Kosch93], [Fua93]). Oder es wird bei der Interpolation auf eine Konsistenz zum Grauwertverlauf im Stereobildpaar (L,R) geachtet ([Fua93]).

Es gibt Optimierungsverfahren, die in einem globalen, grauwertbasierten Zuordnungsprozeß von Pixeln implizit eine interpolierte Disparitätskarte erzeugen (siehe Abschnitt 2.2.2).

Einige aktive Stereoverfahren erzeugen durch eine gezielte Beleuchtung mit einem Streifenmuster oder durch einen Laserstrahl photometrische Bedingungen, die die Extraktion einer dichten, global vertrauenswürdigen Disparitätskarte ermöglichen, da die gesamte Szene mit einer projizierten Struktur versehen wird.

Die Interpolation ist im allgemeinen zeitaufwendig, da ein teilweise kompliziertes Interpolationsmodell zugrunde liegt. Deshalb sollte sie aufgabenbezogen durchgeführt werden.

Schließlich stellt eine dünn besetzte Disparitätskarte auch eine Reduktion auf relevante Daten dar, was eine vorteilhafte Eigenschaft bezüglich nachfolgender Verarbeitungsprozesse sein kann.

Kapitel 3

Untersuchung des Cepstrum-Filters

Vor dem Hintergrund der erfolgten allgemeinen Betrachtung des Disparitätsproblems in der Literatur werden nun konkret Stereoverfahren basierend auf dem Cepstrum-Filter untersucht.
Dies umfaßt eine Beschreibung und Untersuchung der Eigenschaften des Cepstrum-Filters. Standardmethoden der Literatur werden vorgestellt, strukturell analysiert und implementiert. Aus den gewonnenen Erkenntnissen aufbauend wird ein eigenes cepstrumbasiertes Stereoverfahren entwickelt, das Schwächen von Literaturansätzen vermeidet, wobei Aspekte der Integration in die Systemkomponente berücksichtigt werden. Im speziellen wird die Entwicklung eines zum Cepstrum-Filter analogen Operators notwendig, weil dieser durch Hardware-Restriktionen nicht auf dem MV200-System implementierbar ist.
Mit $\mathcal{F}$ sei im folgenden die Fouriertransformation bezeichnet.

3.1 Beschreibung des Cepstrum-Filters

Der Cepstrum-Filter ist auf einem Doppelsignal h definiert, das sich durch Zusammenfügen zweier Stereobilder L und R der Dimension NxN ergibt. Der Cepstrum-Filter von h ist definiert durch

$$Cepstrum(h) = \|\mathcal{F}(\log(\|\mathcal{F}(h)\|^2))\|^2.$$

Der Cepstrum-Filter ergibt sich als das Leistungsspektrum des Logarithmus des Leistungsspektrums des Doppelsignals h. Das Cepstrum ist ein nicht-linearer Filter.
Die Antwort des Cepstrum-Filters besitzt die Dimension des Doppelsignals der Eingabe.
Wie kann man den Cepstrum-Filter in einem Stereoverfahren verwenden ? Angenommen, in L und R befinden sich jeweils p korrespondierende Objekte,

die $d_i = (\triangle x_i, \triangle y_i)$ ($i \in \{1 \ldots p\}$) unterschiedliche Disparitäten besitzen. In $Cepstrum(h)$ sind diese p vorkommenden Disparitäten in Form von p linearen Impulsfolgen kodiert, die sich im Ursprung der Filterantwort $(0, \frac{N}{2})$ schneiden ([Ye89]). Die Impulsfolgen sind symmetrisch (zyklisch) zum Ursprung, sie fallen mit zunehmender Entfernung von ihm ab und ihre stärksten Impulse kodieren durch ihren Abstand vom Ursprung die Disparitäten d_i. Zur Detektion von Disparitäten reicht also die Impulsdetektion durch einen lokalen Maximumoperator in einem entsprechenden Suchbereich um den Ursprung der Filterantwort aus.
Die Symmetrie der Cepstrum-Antwort insgesamt ergibt sich aus Symmetrieeigenschaften der Fouriertransformation.
In Abbildung 3.1 existieren im synthetischen Stereobildpaar (L,R) korrespondierende Rechtecke mit konstanter Disparität. Beide Signale bilden das Doppelsignal h. Weiterhin sind die Filterantwort $Cepstrum(h)$ und die durch Schwellwertbildung und Extraktion lokaler Maxima extrahierten Impulse der Cepstrum-Antwort dargestellt. Im diesem Bild findet sich die gesuchte Disparität als Impulsfolge in einem Suchbereich um den Ursprung wieder.
Zu beachten ist, daß die Symmetrie der Cepstrum-Antwort um den Ursprung zu einer Mehrdeutigkeit hinsichtlich des Vorzeichens der in ihr kodierten Disparität führt. In Abbildung 3.1 kann man der Impulsfolge durch den Ursprung im zyklischen Suchbereich eine positive oder negative Disparität zuordnen, da zwei gleichwertige Impulse der Folge in zwei Quadranten um den Ursprung liegen.
Diese Mehrdeutigkeit läßt sich auflösen, indem man durch eine entsprechende Verschiebung von R eine "künstliche" Disparität zu R hinzuaddiert, so daß der Suchbereich nur noch in zwei Quadranten um den Ursprung der Cepstrum-Antwort liegt anstatt in vier Quadranten. Um aus einer detektierten Disparität die vorzeichenbehaftete wahre Disparität zu ermitteln, ist die künstliche Verschiebung zu subtrahieren.

3.2 Analyse und Wirkungsweise des Cepstrum-Filters

Nachdem das Prinzip der Detektion von Stereodisparität durch den Cepstrum-Filter vorgestellt wurde, soll diese Vorgehensweise theoretisch nachvollzogen werden. In [Ye89] und in [Lu91] findet sich eine Herleitung, die zeigt, warum sich Disparitäten in (L,R) durch Impulsfolgen in der Antwort des Cepstrum-Filters detektieren lassen.
Angenommen es liegt ein Stereobildpaar (L,R) vor, das eine konstante Disparität $(\triangle x, \triangle y)$ aufweist. Dann kann man R als verschobene Version von L betrachten. Interpretiert man weiterhin L und R als jeweils unendlich ausgedehnte Signale, so läßt sich das Doppelsignal h aus L und R als Überlagerung des Signals L mit

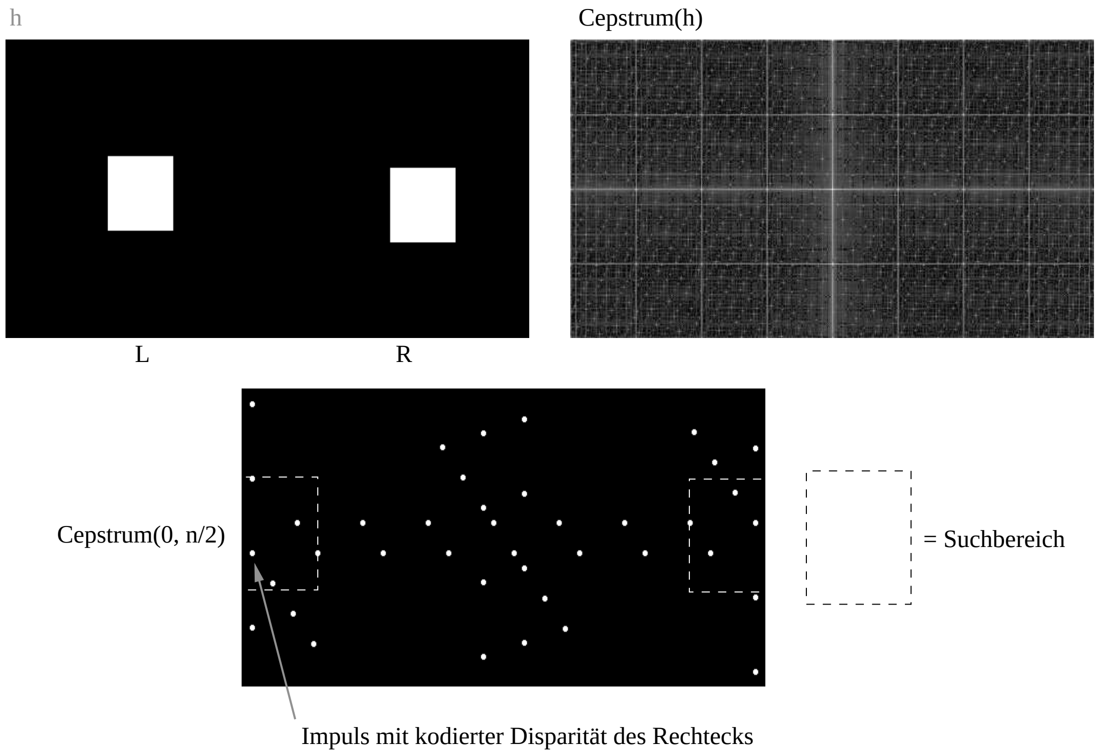


Abbildung 3.1: Der Cepstrum-Filter liefert zu einem Stereobildpaar eine Menge von Impulsen, deren Abstand vom Ursprung der Filterantwort Disparitätsinformationen kodiert. In dem vorliegenden synthetischen Stereobildpaar ergibt sich die Disparität der Quadrate als intensivster Impuls im Suchbereich.

einer verschobenen Version von sich selbst darstellen.

$$h(x,y) \quad = \quad L(x,y) + L(x + \triangle x, y + \triangle y) \tag{3.1}$$

Wendet man auf die linke und rechte Seite der Gleichung den Cepstrum-Filter an, so kann man unter Verwendung von Näherungen folgende Beziehung herleiten ([Lu91]) :

$$\begin{aligned} &Cepstrum(h(x,y)) = \\ &Cepstrum(L(x,y)) \quad + \quad \sum_{m \neq 0} \frac{\delta(x - m\triangle x, y - m\triangle y)}{m} \end{aligned} \tag{3.2}$$

In dem zweiten Summanden verbirgt sich die Disparität $(\triangle x, \triangle y)$ des Stereobildpaares in einer Impulsfolge kodiert, die mit wachsender Entfernung vom Ursprung der Cepstrum-Antwort abfällt. Der Abstand der Impulse beträgt $(\triangle x, \triangle y)$. Aus diesem Grund läßt sich die Disparität $(\triangle x, \triangle y)$ mit dem intensivsten Impuls der Folge identifizieren.

Es gibt unterschiedliche Philosophien, Gleichung 3.2 zur Detektion von Stereodisparität beziehungsweise Cepstrum-Impulsen zu interpretieren. Ludwig ([Lu91]) berechnet ausschließlich den Term $Cepstrum(h(x,y))$, obwohl dabei die Gefahr besteht, daß die gesuchte Impulsfolge durch den Term $Cepstrum(L(x,y))$ überlagert wird. Dieser Fall soll aber in der Praxis bis auf Ausnahmen nicht auftreten, so Ludwig, mit der Begründung, daß die Signalanteile von $Cepstrum(L(x,y))$ in $Cepstrum(h(x,y))$ nicht im Suchbereich für Disparitäten liegen. Dieser Argumentation kann man folgen, wenn man die Berechnung des Cepstrum-Filters als Autokorrelation von h interpretiert (siehe Abschnitt 3.2.1).Ludwig nimmt diesen systematischen Fehler in Kauf, um die Laufzeit des von ihm implementierten Stereoalgorithmus zu optimieren.

Yeshurun und Schwartz hingegen subtrahieren vor einer Impulsdetektion im Suchbereich den Term $Cepstrum(L(x,y))$ von $Cepstrum(h(x,y))$, obwohl Ludwig im Fall von realen Stereobildpaaren, die praktisch immer die Annahme konstanter Disparität verletzen, ein Versagen dieser Strategie festzustellen glaubt.

Neben dieser gespaltenen Auffassung über beide Zugänge zur Interpretation von Gleichung 3.2 in der Literatur ergeben sich in eigenen Praxisuntersuchungen keine relevanten Unterschiede zwischen beiden Vorgehensweisen. Deshalb werden beide Zugänge als gleichwertig eingestuft.

Unter obiger Annahme genau einer Disparität im Stereobildpaar wird in der Literatur als Alternative zum Cepstrum-Filter die verwandte Phasenkorrelation verwendet, um die Disparität zu detektieren ([Olson88]). Sie beruht auf der Anwendung des Verschiebungssatzes der Fouriertransformation.

Bei der Phasenkorrelation P kodiert ebenfalls ein Impuls der Filterantwort die Disparität.

$$P(L,R) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}(R)(u,v)\mathcal{F}^*(L)(u,v)}{||\mathcal{F}(R)(u,v)\mathcal{F}^*(L)(u,v)||}\right\} = \delta(x - \triangle x, y - \triangle y)$$

Die Phasenkorrelation berechnet eine normierte Kreuzkorrelation von L und R im Frequenzraum. Die Normierung schärft durch Equilisation von spektralen Werten den resultierenden Impuls der Phasenkorrelation. In der Interpretation als Kreuzkorrelation von L und R ist es plausibel, daß die Phasenkorrelation verschobener Signale einen Impuls an der Stelle der Verschiebung $(\triangle x, \triangle y)$ liefert.
Auch der Cepstrum-Filter läßt sich als Korrelation in Verbindung mit einer Funktion zur Verstärkung von Impulsen der Filterantwort interpretieren, wie in Abschnitt 3.2.1 gezeigt wird. So zeigt sich die Verwandschaft von Phasenkorrelation und Cepstrum-Filter. Laut [Olson88] bietet sie den theoretischen Vorteil, daß der Impuls der Phasenkorrelation nicht von der Filterantwort der Eingangssignale selbst überlagert wird, wie es beim Cepstrum-Filter der Fall ist.

3.2.1 Interpretation als Autokorrelation

Neben Gleichung 3.2 existiert eine weitaus anschaulichere Interpretation der Antwort des Cepstrum-Filters. Nach ([Olson88]) läßt sich der Cepstrum-Filter als Autokorrelationsfilter interpretieren, dessen Maxima durch einen adaptiven Vorfilter verstärkt werden. Auch [Ye89] erzielt bei Vergleichstests zwischen Cepstrum-Filter und Autokorrelation ähnliche Resultate, wobei der Cepstrum-Filter ein besseres Signal-Rausch-Verhältnis aufweist.
Die Interpretation als Autokorrelationsfunktion ist ein wichtiges Hilfsmittel, um Impulse der Cepstrum-Antwort zu verstehen.

Den Zusammenhang zwischen Cepstrum-Filter und Autokorrelationsfunktion (ACF) kann man theoretisch herleiten ([Lu91]). Zu diesem Zweck wird zuerst einmal die Rolle der Log-Funktion im Cepstrum-Filter untersucht.

$$\log(||H||^2) = \frac{log(||H||^2)}{||H||^2} HH^* = \left|\left|\frac{\sqrt{log(||H||^2)}}{||H||} H\right|\right|^2 = ||VH||^2 =: ||G||^2$$

mit H = $\mathcal{F}(h)$, $V := \frac{\sqrt{log(||H||^2)}}{||H||}$
Aus dieser Umformung wird deutlich, daß der Logarithmus des Leistungsspektrums des Doppelsignals h gleich dem Leistungsspektrum einer gefilterten Version G von h ist, die durch einen nicht-linearen, adaptiven Filter V aus h im Frequenzraum entsteht. Mit dieser Beziehung läßt sich die Wirkungsweise des Cepstrum-Filters neu interpretieren.

$$Cepstrum(h) =$$
$$||\mathcal{F}(\log(||\mathcal{F}(h)||^2))||^2 = ||\mathcal{F}(||G||^2)||^2 =$$
$$||\mathcal{F}(||G^*G||)||^2 = ||g^* * g||^2 = ||ACF(g)||^2$$

Ohne die Vorfilterung V des Doppelsignals berechnet der Cepstrum-Filter nach dem Korrelationssatz der Fouriertransformation lediglich die Autokorrelation. Die

Filterung durch die Log-Funktion dient dazu, die Autokorrelationsmaxima beziehungsweise Impulse des Cepstrum-Filters zu verstärken, indem periodische Signalbestandteile und Rauschen unterdrückt werden ([OlCo91]). So verbessert die Log-Funkion das Signal-Rausch-Verhältnis des Cepstrum-Filters.
Dies geschieht durch eine Equilisation von Frequenzen im Leistungsspektrum. Die Log-Funktion schwächt Signalbestandteile mit geringer Bandbreite aber hoher Intensität wie Rauschen und periodische Strukturen stärker ab als Signale großer Bandbreite und geringer Intensität. Letztere Signale sind verantwortlich für die nach erneuter Fouriertransformation entstehenden Impulse der Cepstrum-Antwort, eine Folgerung aus der Unschärferelation zwischen Orts- und Frequenzraum.
Verifizieren kann man diese Überlegung, indem man die Log-Funktion durch analoge Funktionen wie die Wurzel- oder Arctan-Funktion ersetzt. Diese Funktionen liefern im Cepstrum-Filter analoge Filterantworten ([Olson88]).
Auch die Interpretation des Cepstrum-Filters als Autokorrelation insgesamt läßt sich in praktischen Tests verifizieren. In Abbildung 3.2 sind die Cepstrum-Antwort eines Doppelsignals h sowie das Resultat der Autokorrelationsfunktion von h gegenübergestellt. Die Filter liefern eine strukturell ähnliche Antwort und Maxima beziehungsweise Impulse an identischen Positionen.
In seiner Arbeit, die auf der von Yeshurun-Schwartz (siehe [Ye89]) aufbaut, kommt Ludwig ([Lu91]) in einer weiterführenden Untersuchung der impliziten Vorfilterung des Cepstrum-Filters V zu dem Ergebnis, daß V im wesentlichen eine Bandpaßfilterung von L und R durchführt. Die alleinige Bildung der Autokorrelation mit einer vorausgehenden Bandpaßfilterung liefert einen zum Cepstrum nahezu identischen Filter.

Die Autokorrelation selbst berechnet die Faltung eines Signals mit seinem konjugierten Signal. Dies kann man sich veranschaulichen, indem man das Signal h zyklisch verschoben mit sich selbst überlagert. Die Autokorrelationsfunktion für eine bestimmte Verschiebung ist ein Maß für die strukturelle Ähnlichkeit der überlagerten Signale. Mit dieser Sichtweise wird eine plausible Interpretation von Impulsen des Cepstrum-Filters möglich.
In Abbildung 3.2 sind ein Doppelsignal h eines Stereobildpaares (L,R) bestehend aus jeweils zwei gegeneinander verschobenen Kreisen, die Antwort des Cepstrum-Filters und die Autokorrelation von h sowie die in der Cepstrum-Antwort detektierten Impulse (A) dargestellt.
Für die zwei im Stereobildpaar vorkommenden Disparitäten wird das Doppelsignal in der entsprechenden zyklischen Verschiebung so mit sich selbst überblendet, daß jeweils die Eingabebilder L und R in den Doppelsignalen in einer entsprechenden Disparität übereinanderliegen. In den Überblendungen ist zu sehen, welche Bildbestandteile in L und R für die Impulse in der Cepstrum-Antwort verantwortlich sind. Im Fall der Verschiebung 1 liegen die beiden großen Kreise aus jeweils L und R genau übereinander, während bei der Verschiebung 2 die

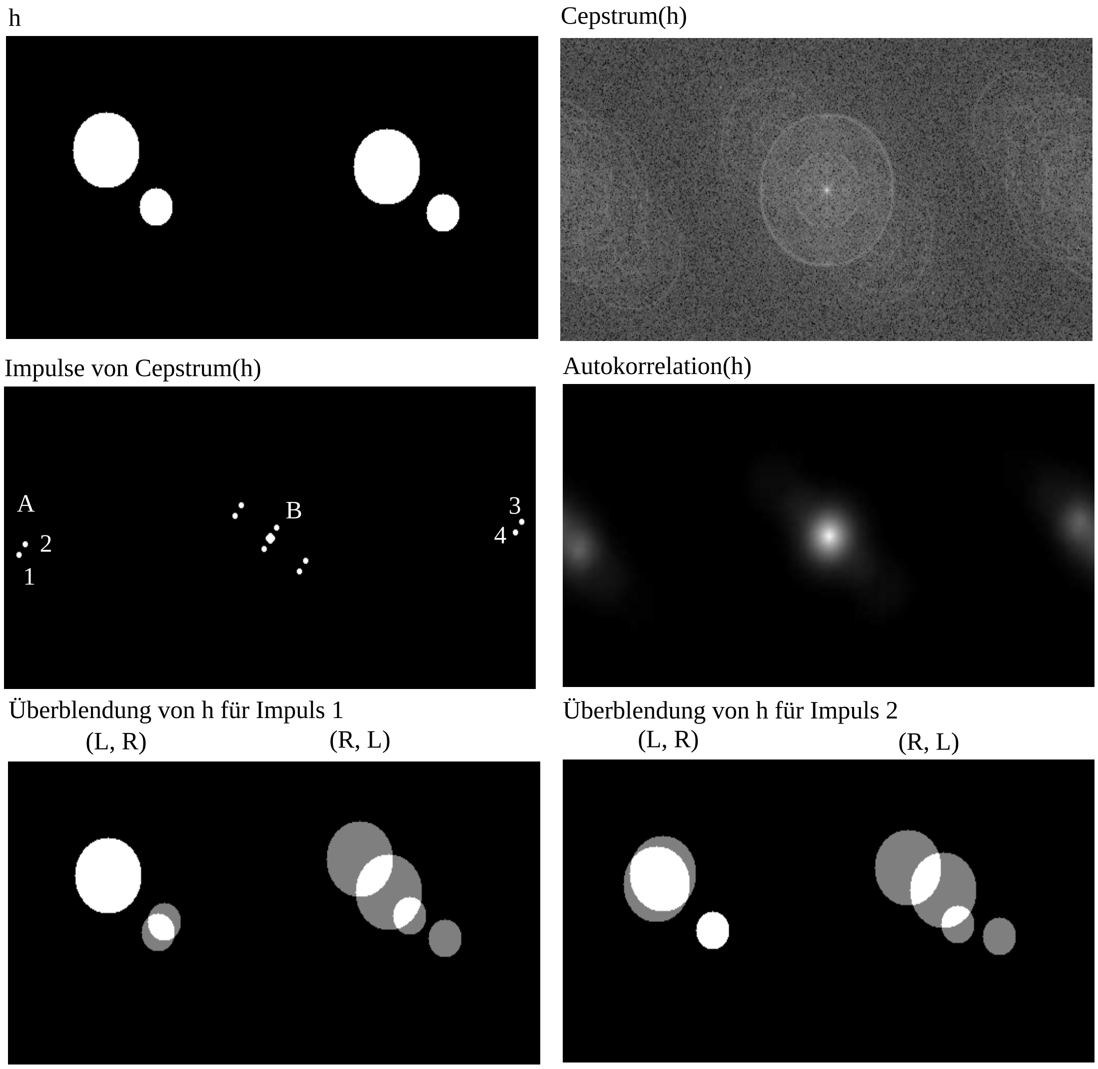


Abbildung 3.2: Die Antwort des Cepstrum-Filters läßt sich als Autokorrelation eines Doppelsignals h interpretieren, deren Maxima durch eine adaptive, bildabhängige Vorfilterung verstärkt werden. Die Überblendung von h mit sich selbst für detektierte Impulse beziehungsweise Disparitäten bietet ein Hilfsmittel zur Identifizierung der für die Impulse verantwortlichen Bildstrukturen in L und R. Dabei sind die Doppelsignale so gegeneinander zyklisch verschoben, daß die Eingabebilder L und R in entsprechender Disparität übereinanderliegen.

entsprechenden kleine Kreisen übereinander liegen.
Die Impulse im Zentrum der Filterantwort (B) stammen aus der Überlagerung der Eingabebilder L und R jeweils mit sich selbst. So erklärt sich auch der intensivste Impuls genau im Zentrum der Cepstrum-Antwort, der aus der Überlagerung des Doppelsignals h ohne Verschiebung mit sich selbst entsteht und damit maximale strukturelle Übereinstimmung widerspiegelt. Ein ausgeprägtes zentrales Maximum ist ein fester Bestandteil der Autokorrelationsfunktion.

3.3 Der Cepstrum-Filter als Vergleichsmaß

Aufgrund seines globalen Einzugsbereichs ist der Cepstrum-Filter als photometrisches Vergleichsmaß robust gegenüber Rauschen und Alias-Fehlern bedingt durch periodische Signalbestandteile (siehe 2.1.2). Olson [TaOlMa94] begründet die Robustheit gegenüber periodischen Signalbestandteilen mit der adaptiven Vorfilterung des Doppelsignals durch den Cepstrum-Filter.
Mit der Interpretation des Cepstrum-Filters als Autokorrelation sollen im Folgenden einige seiner Eigenschaften nachvollzogen werden. Die Grundlage der Untersuchung bilden verschiedene Testbilder.
Desweiteren werden Abweichungen des Cepstrum-Filters von seiner theoretischen Interpretation durch Gleichung 3.2 beschrieben.

3.3.1 Systematische Fehler

Im Prozeß der Anwendung des Cepstrum-Filters können als Sonderfälle Impulse entstehen, die verfahrensbedingt sind und keine Korrespondenzen zwischen L und R kodieren. Diese systematischen Fehler werden in [Lu91] beschrieben.
Besitzen L und R unterschiedliche Helligkeiten, so entsteht in der Cepstrum-Antwort ein Impuls im Ursprung und im Zentrum. Diese Impulse können intensiv sein, denn für ihre Entstehung ist die gesamte Signalenergie von L und R verantwortlich.
Der systematische Fehler entsteht bei der Bildung eines Doppelsignals aus einem Stereobildpaar. Besitzen L und R unterschiedliche Helligkeiten, so entsteht im Doppelsignal eine intensive Kante an der Nahtstelle von L und R. Im Doppelsignal werden verfahrensbedingt neue Frequenzen erzeugt.
Bei der Herleitung der durch Gleichung 3.2 (Seite 31) gegebenen Interpretation der Cepstrum-Antwort, aus der sich die Korrespondenz von Impulsen zu Disparitäten ergibt, wird eine Vereinfachung durchgeführt. In der Ursprungsgleichung 3.1 wird nicht die Wirkung einer Trägerfunktion berücksichtigt, die bei der Berechnung des Cepstrum-Filters zur Bildung des Doppelsignals Ausschnitte aus den unendlich ausgedehnten Eingangssignalen L und R extrahiert. So ist die Herleitung der Cepstrum-Technik mit einer systematischen Schwäche behaftet.
Bereits die Grundannahme einer konstanten Disparität, auf der Gleichung 3.2

beruht, wird in praktisch jedem realen Stereobildpaar verletzt. Ludwig ([Lu91]) beobachtet in diesem Fall das Auftreten von neuen Impulsen, die nicht mehr nach Gleichung 3.2 interpretiert werden können. In diesem Fall ist die theoretische Interpretation als Autokorrelation hilfreich.
Um die Annahme konstanter Disparität in den Eingangssignalen zu gewährleisten, bleibt nur der Ausweg, den Einzugsbereich des Vergleichsmaßes zu verkleinern (siehe Abschnitt 2.1.2). Allerdings wirkt sich bei dieser Vorgehensweise die verfahrensbedingte Koppelung von Einzugsbereich und Suchbereich des Cepstrum-Filters negativ aus, weil durch Reduzierung des Einzugsbereichs der Suchbereich, der auf Annahmen der Korrespondenzsuche basiert und somit fest vorgegeben ist, unerwünscht klein werden kann.

3.3.2 Beträge von Impulsen

In Abbildung 3.2 aus Abschnitt 3.2.1 sind Überblendungen eines Doppelsignals in durch den Cepstrum-Filter als Disparitäten ermittelten Verschiebungen mit sich selbst dargestellt. In den Überblendungen korrespondieren Impulse der Cepstrum-Antwort mit struktureller Übereinstimmung von Bildstruktur. In der Interpretation des Cepstrum-Filters als Autokorrelation können deshalb Beträge von Cepstrum-Impulsen als Beträge von lokalen Maxima der Autokorrelationsfunktion verstanden werden.
In den Überblendungen ist ersichtlich, daß ein Impuls wie im Fall der beiden großen Kreise auf eine hohe Korrelation von Bildstrukturen in einer bestimmten Disparität im Doppelsignal zurückzuführen ist. Daneben gibt es aber auch Bildstrukturen, die bei dieser Disparität das Korrelationsergebnis in unspezifischer Weise negativ beeinflussen, wie im rechten Teil der Überblendung erkennbar ist. Der Betrag der Autokorrelationsfunktion wird so durch Bildbereiche von lokal hoher und niedriger Korrelation des Doppelsignals beeinflußt.
Die Impulsbeträge des Cepstrum-Filters spiegeln somit die Dominanz der in einer entsprechenden Disparität vorliegenden Bildstruktur wider.
Allerdings kann der Fall eintreten, daß die hohe lokale Korrelation einer Bildstruktur in entsprechender Disparität durch niedrige lokale Korrelation von photometrisch dominanteren Bildstrukturen überlagert wird, so daß die Autokorrelationsfunktion kein entsprechendes Maximum aufweist. Der Impuls der Cepstrum-Antwort verschwindet. Dieser Fall der Dominanz von Bildstruktur ist auf einen zu großen Einzugsbereich des globalen Vergleichsmaßes Cepstrum-Filter zurückzuführen (siehe Abschnitt 2.1.2).
Der Effekt der Überlagerung von Cepstrum-Impulsen wird von Ludwig angesprochen ([Lu91]) und läßt sich auch theoretisch nachvollziehen (siehe Gleichung 3.2). Im umgekehrten Fall können Impulsbeträge durch Bildstrukturen mit lokal hoher Korrelation aber tatsächlich unterschiedlicher Disparität verstärkt werden. In diesem Fall handelt es sich um Alias-Fehler.

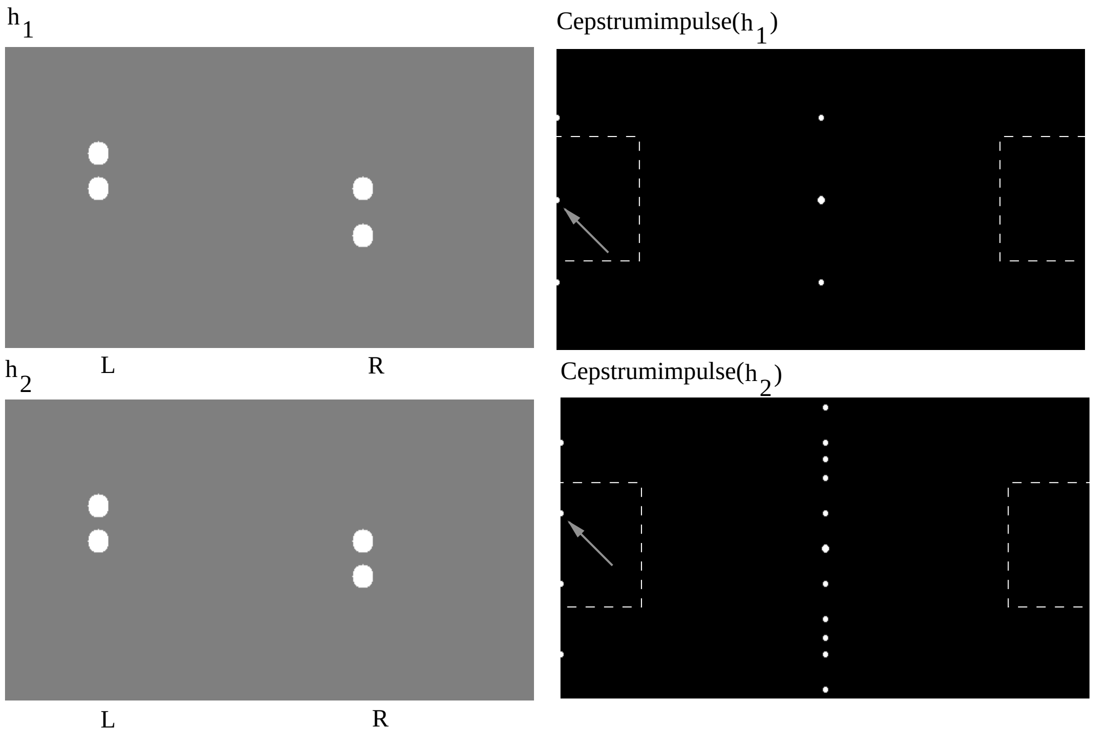


Abbildung 3.3: Vorsicht ist bei der Analyse der Beträge von cepstralen Impulsen geboten. Die strukturell identischen Doppelsignale h_1 und h_2 besitzen annähernd dieselbe Disparität, während die gekennzeichneten betragsstärksten Impulse im Suchbereich von h_1 und h_2 an vollkommen unterschiedlichen Positionen liegen. Die Impulse im zentralen Bereich der Cepstrum-Antworten sind auf die Autokorrelation der Eingabebilder L und R im Doppelsignal jeweils mit sich selbst zurückzuführen.

Aus diesen Überlegungen folgt, daß Impulsbeträge für identische Bildstrukturen, die jedoch unterschiedliche Disparitäten aufweisen, in zwei Stereobildpaaren unterschiedliche Werte annehmen können. Dabei bleibt im Korrelationsergebnis der positive Anteil zwar konstant, jedoch der negative Anteil schwankt. Die Beträge hängen also von den im Stereobildpaar auftretenden Disparitäten ab. So ändert sich bei mehreren Impulsen im allgemeinen deren Rangfolge im Betrag, wenn sich die im Stereobildpaar vorkommenden Disparitäten ändern.

Ein Beispiel für diesen Effekt stellt Abbildung 3.3 dar. Es sind zwei synthetische Doppelsignale h_1 und h_2 dargestellt, die bis auf geringfügig verschiedene Disparitäten identisch sind. Das Stereobildpaar h_2 könnte beispielsweise durch den Effekt einer perspektivischen Verzerrung entstanden sein.

Weiterhin sind die betragsmäßig stärksten Impulse der Cepstrum-Antwort extrahiert worden. Diese liegen an unterschiedlichen Positionen, obwohl sich die Disparitäten in h_1 und h_2 nur wenig unterscheiden und eine entsprechend kleine Impulsverschiebung zu erwarten wäre.

Eine einfache Strategie der Detektion des intensivsten Impulses in einem Suchbereich der Cepstrum-Antwort würde für h_1 und h_2 die Disparität unterschiedlicher Bildstrukturen ermitteln, im Fall von h_1 die des unteren Kreises in L, im Fall von h_2 die beider Kreise.

In Experimenten zeigt sich weiterhin, daß die Beträge von Cepstrum-Impulsen von der Signalenergie des Stereobildpaares abhängen, wie es nach dem Parsevalschen Theorem der Energieerhaltung der Fouriertransformation auch zu erwarten ist.
Das Fazit dieser Untersuchung ist, daß eine detektierte Menge von Cepstrum-Impulsen von ihrem Betrag her nur heuristisch bewertet werden kann, obwohl der Betrag als Maß für die Dominanz einer in entsprechender Disparität vorliegenden Bildstruktur gewertet werden kann.

3.3.3 Dichte von Disparitätsinformationen

Der Cepstrum-Filter liefert im allgemeinen nur eine Teilmenge der in einem Bildpaar vorkommenden Disparitätswerte in Form von Impulsen der Filterantwort. Diese Eigenschaft beruht auf dem globalen Einzugsbereich des Filters als Vergleichsmaß. Impulse des Cepstrum-Filters lassen sich als lokale Maxima der Autokorrelationsfunktion interpretieren. Dies schließt die Detektion von Disparitätswerten in unmittelbarer Nachbarschaft eines Impulses aus, weil sich Impulse gegenseitig überlagern. Es handelt sich um einen in Abschnitt 3.3.2 beschriebenen Fall. Kontinuierliche Disparitätsverläufe lassen sich so nur unterabgetastet detektieren.
Dieser Fall tritt insbesondere bei realen Stereobildern auf, die Verzerrungen enthalten (siehe Abschnitt 2.1.2). In Abbildung 3.4 sind ein Stereobildpaar h_1 einer realen Szene sowie die stärksten Impulse des Cepstrum-Filters im Suchbereich dargestellt, die noch eine nachvollziehbare Korrespondenz zu Disparitäten in Bildstrukturen besitzen. Der Cepstrum-Filter ordnet dem Bildpaar sieben Disparitäten zu, obwohl kontinuierliche Disparitätsverläufe vorliegen.
Im synthetischen Stereobildpaar h_2 von zwei verzerrten "Balken" ist der besagte Effekt am deutlichsten sichtbar. Auch hier detektiert der Cepstrum-Filter in einem kontinuierlichen Intervall nur wenige Disparitätswerte.
Der Prozeß der Unterabtastung von Disparitäten im Stereobildpaar ist analytisch ungeklärt. Ludwig ([Lu91]) befaßt sich mit dieser Frage, indem er den Vorgang des Zerfalls eines Impulses in mehrere benachbarte Impulse in Folge einer Rotation zwischen korrespondierenden Strukturen in L und R schildert. Die entsprechende Bildstruktur ist nun nicht nur zwischen L und R verschoben, sondern auch verzerrt, so daß kontinuierliche Disparitätsverläufe entstehen.

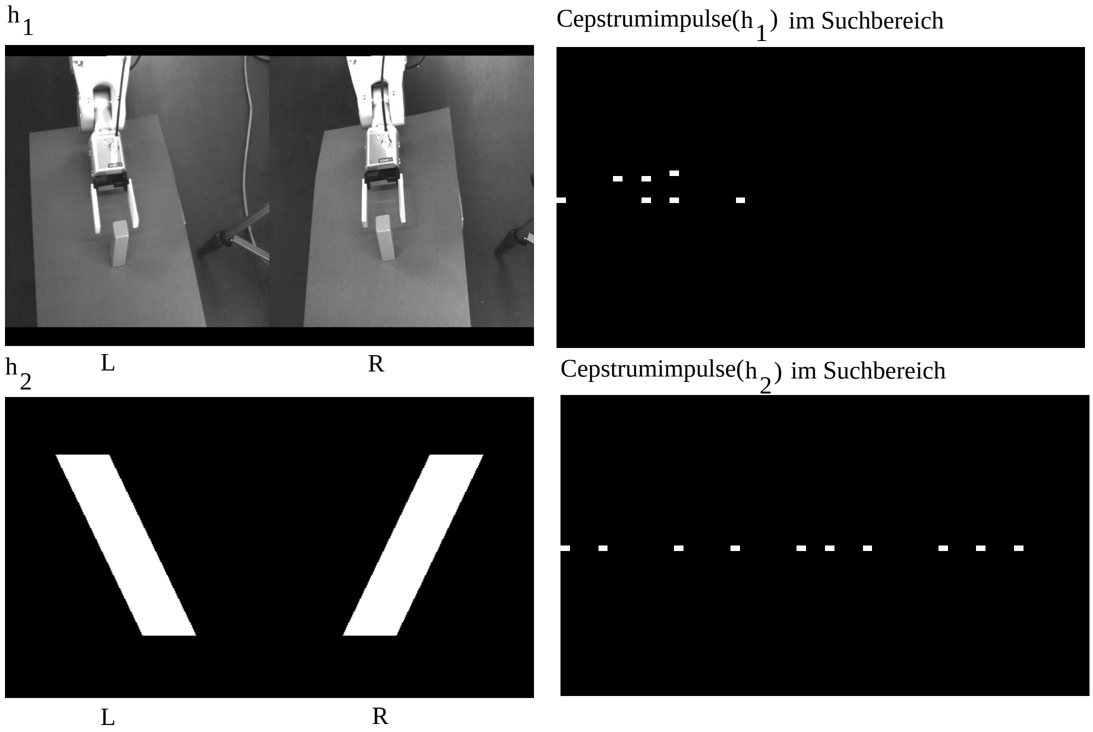


Abbildung 3.4: Der Cepstrum-Filter liefert keine kontinuierlich verteilten, sondern grob abgetastete Disparitätsinformationen, weil sich bezüglich der Suchposition benachbarte Impulse gegenseitig überlagern und so nicht mehr detektierbar sind.

3.4 Stereoverfahren auf Basis des Cepstrum-Filters

3.4.1 Der Cepstrum-Filter zur Disparitätsbestimmung

In der Literatur wird der Cepstrum-Filter in Stereoverfahren eingesetzt. Er liefert zu einem Stereobildpaar (L,R) ausschließlich eine Menge von Impulsen beziehungsweise kodierten Disparitäten, die im Bildpaar vorkommen, jedoch ohne Informationen darüber, welche Bildstrukturen für einen Impuls verantwortlich sind. Beträge von Impulsen beinhalten zwar Informationen über die Dominanz der verantwortlichen Bildstruktur, das globale Vergleichsmaß kann jedoch keine Informationen über ihre Lage liefern.
Diese zur Lösung des Disparitätsproblems elementare Frage stellt das eigentliche Problem bei der Analyse der Antwort des Cepstrum-Filters dar, wie unten deutlich wird (siehe Abschnitt 3.4.2).
Um diesen Zusammenhang herzustellen kann man von der plausiblen Annahme ausgehen, daß der intensivste Impuls des Cepstrum-Filters der dominantesten Bildstruktur im Stereobildpaar zuzuordnen ist. In einer Näherung kann man einem zentralen Aufpunkt von L und R die im intensivsten Cepstrum-Impuls kodierte Disparität zuordnen, weil mit hoher Wahrscheinlichkeit der Aufpunkt innerhalb der dominanten Bildstruktur liegt. Diese Strategie beinhaltet die Annahme konstanter Disparität im Einzugsbereich des Cepstrum-Filters. In diesem Fall kann der Cepstrum-Filter auch effizient durch die verwandte Phasenkorrelation ersetzt werden ([Olson88]).
In einem Stereoverfahren müssen jedoch unterschiedliche Disparitäten im Stereobildpaar detektierbar sein. Als Ausweg reduziert man den Einzugsbereich des Vergleichsmaßes, indem man jeweils L und R in eine Anzahl von Bild-"Kacheln" aufspaltet und jeweils zusammengehörige Kacheln aneinanderreiht. So erhält man in einer gleichmäßigen Aufteilung von L und R eine Anzahl von Doppelsignalen und berechnet deren Antworten des Cepstrum-Filters. Wieder besteht die Annahme, daß die Disparität innerhalb einer Kachel konstant ist. Dann wird der intensivste Impuls im Suchbereich der jeweiligen Cepstrum-Antwort ausgewählt und seine Disparität der Kachel zugeordnet.
Damit die Auflösung einer resultierenden Disparitätskarte nicht an die Kachelgröße gebunden ist, dürfen sich Kacheln überlappen. Es wird von einer fixierenden Stereokonfiguration ausgegangen, so daß die im Zentrum von (L,R) liegenden Disparitäten klein genug für die Suchbereiche in den einzelnen Kacheln sind.
Diese Vorgehensweise wird von Yeshurun-Schwartz und Ludwig ([Ye89], [Lu91]) mit der Streifenarchitektur im visuellen Wahrnehmungssystem höherer Lebewesen biologisch motiviert. Dort treten visuelle Informationen aus linkem und rechtem Auge alternierend in benachbarten, streifenförmigen Gehirnregionen auf (okulare Dominanzstreifen), wobei die Wahrnehmungsfähigkeit von Disparitäten

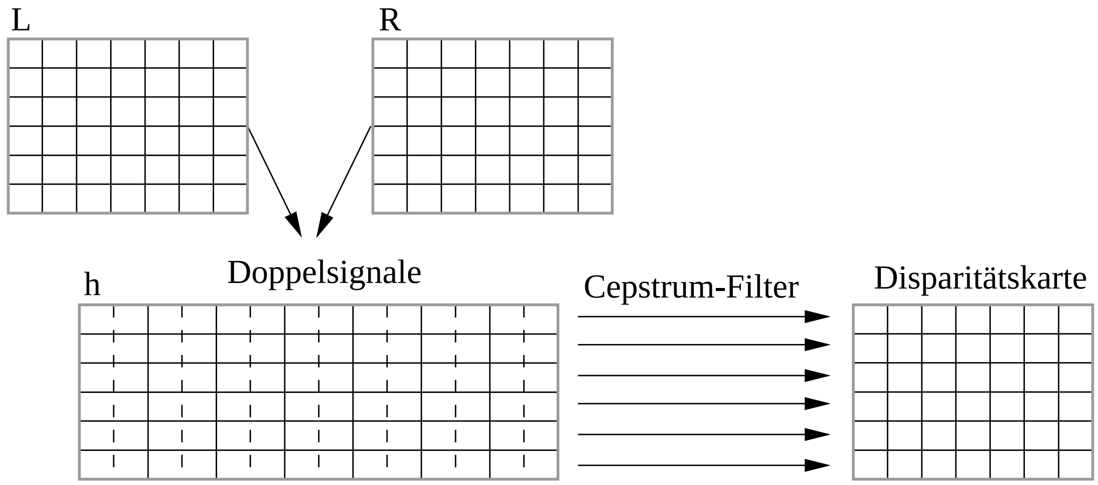


Abbildung 3.5: Das Stereoverfahren nach Yeshurun-Schwartz (YESA) modelliert biologische Sachverhalte. In Anlehnung an okkulare Dominanzstreifen im visuellen Cortex werden L und R in Kacheln aufgeteilt und zu entsprechenden Doppelsignalen zusammengefaßt. In diesen Doppelsignalen wird durch den Cepstrum-Filter eine angenommene konstante Disparität gemessen.

auf kleine Werte beschränkt ist. Außerdem wird die einfache, parallele Struktur des Algorithmus als angemessene Modellierung einer angenommenen direkten biologischen Disparitätsschätzung angesehen.

Das Stereoverfahren von Yeshurun-Schwartz

Das Stereoverfahren besitzt die oben beschriebene, einfache algorithmische Struktur gemäß Abbildung 3.5.
Zur Analyse der Cepstrum-Antworten wird Gleichung 3.2 exakt interpretiert. Das heißt, vor der Detektion von Impulsen wird das Cepstrum des Eingangssignals vom Cepstrum des Doppelsignals subtrahiert.
Zu beachten ist, daß wie bereits erwähnt die Dimension der Kacheln den Suchbereich des Vergleichsmaßes festlegt. Aufgrund von Randproblemen und systematischen Schwächen der Analyse der Cepstrum-Antwort durch Gleichung 3.2 liegt der Suchbereich in der Praxis unterhalb des Einzugsbereichs.
In der Praxis ergibt sich als Wert für die Dimension der Kacheln ein Kompromiß zwischen der Feinheit der resultierenden Disparitätskarte (Einzugsbereich) und den detektierbaren Disparitäten (Suchbereich).
Bei der Implementierung unter dem Bildverarbeitungssystem KHOROS wurde eine Kachelgröße von 32 Pixel und damit ein Suchbereich von ± 16 Pixel gewählt. Dieser Wert ergibt sich aus der Forderung nach Vergleichbarkeit zu anderen getesteten Verfahren. In [Lu91] wird als minimaler brauchbarer Wert eine Kachelgröße von 8 Pixel angegeben.

Weil angenommen wird, daß in realen Bildern hauptsächlich horizontale Disparitäten vorkommen, wurde in dieser Implementierung an Stelle einer vertikal rechteckigen Kachelform bei Ludwig und Yeshurun-Schwartz (siehe [Lu91], [Ye89]) eine quadratische Kachelform gewählt. Eine vertikal rechteckige Kachelform widerspricht dieser Annahme, denn die Dimensionen der Kachel und der Suchbereich für Disparitäten sind beim Cepstrum-Filter gekoppelt.
Abgesehen von einem unbearbeiteten Rand wird für jedes Pixel eine Kachel extrahiert. Das Verfahren liegt für eine Dimension der Stereobilder von NxN und einer Dimension der Kacheln von dxd in der Laufzeitklasse (O(N^2d^2)). Der Rechenaufwand für dieses Stereo-Algorithmus ist extrem hoch, da der Cepstrum-Filter N^2 mal berechnet werden muß. Auf einer Ultra-Sparc-Workstation fallen hierfür Rechenzeiten im Stundenbereich an.
Der Algorithmus ist zur Detektion horizontaler und vertikaler Disparität geeignet, jedoch wurde der Suchbereich auf horizontale Disparitäten beschränkt.
Das beschriebene cepstrumbasierte Stereoverfahren nach Yeshurun-Schwartz soll im Folgenden die Bezeichnung YESA erhalten.
Der Cepstrum-Filter wird in diesem Stereoverfahren in einem lokalen Einzugsbereich in L verwendet. Dadurch büßt der Filter weitgehend seine Robustheit gegenüber Rauschen und periodischen Bildstrukturen ein. Außerdem wird trotz des in Form der Kachelbildung reduzierten Einzugsbereichs des Cepstrum-Filters die Annahme von konstanter Disparität im Stereosignal verletzt. Ein weiterer Kritikpunkt ist die heuristische Identifizierung der Disparität einer Kachel mit dem intensivsten Impuls im Suchbereich (siehe Abschnitt 3.3.2).
Die angesprochenen Probleme werden in Kapitel 5 sichtbar, wo durch den YESA-Algorithmus erzeugte Disparitätskarten zu Testbildern den Ergebnissen anderer Stereoalgorithmen in einem Vergleich gegenübergestellt sind.

3.4.2 Der Cepstrum-Filter zur Vergenzbestimmung

Wie oben beschrieben läßt sich mit dem Cepstrum-Filter als Vergleichsmaß durch Detektion des intensivsten Impuls der Filterantwort in einfacher Weise eine Disparität für einen zentralen Aufpunkt eines Stereosignals ermitteln. Diese Tatsache prädestiniert den Filter als Vergenzregler eines Stereokamerasystems, das als Fehler für die Vergenzregelung nur einen einzigen Disparitätswert benötigt.
Olson verfolgt ([Olson88],[OlCo91]) den beschriebenen Ansatz und nutzt die ermittelte Disparität als Fehlermaß der Vergenz eines Stereokamerasystems.
Das Hauptproblem dieser Vorgehensweise liegt in der Tatsache, daß das Modell zur Disparitätsmessung zu eng gefaßt ist. Es treten im allgemeinen in L und R mehrere Disparitäten auf, so daß der Cepstrum-Filter mehrere Impulse mit unterschiedlichen Beträgen liefert, je nach Dominanz der verantwortlichen Bildstrukturen. Tritt der Fall auf, daß der Fixationspunkt im Zentrum von L nicht Teil der dominierenden Struktur von L und R ist, wird dem Fixationspunkt eine falsche Disparität zugeordnet. Dieses Problem eines zu großen Einzugsbereichs

wurde allgemein in Abschnitt 2.1.2 geschildert.
So besitzt das obige Vergenzverfahren denn auch die Eigenschaft, auf dominierende Bildstrukturen des Stereobildpaares zu vergieren, die im allgemeinen der Hintergrund darstellt, wie selbiger Autor in einem späteren Artikel einräumt ([TaOlMa94]).
Es wird deutlich, daß die einfache heuristische Impulsdetektion in der Cepstrum-Antwort nicht ausreichend ist, um das Korrespondenzproblem zu lösen beziehungsweise zu umgehen, wie [Lu91] behauptet.

Als Verbesserung schlagen Olson, Taylor und Lockwood ([TaOlMa94]) eine Suchstrategie über einen variablen Einzugsbereich des Vergleichsmaßes Cepstrum-Filter vor, sowie eine Zuordnung von Impulsen zu Bildstrukturen durch explizite Überprüfung einzelner Impulse durch ein neues lokales photometrisches Vergleichsmaß, um schließlich die in einem ausgewählten Impuls kodierte Disparität als Vergenzfehler zu nutzen. Damit weichen sie von der von Yeshurun und Schwartz geforderten, biologisch motivierten direkten Auswertung der Cepstrum-Antwort ab.
Die genannten Verbesserungen haben die Aufgabe, verschiedene Impulse der Cepstrum-Antwort den verantwortlichen Bildstrukturen zuordnen zu können. Die Annahme konstanter Disparität zwischen L und R wird demnach aufgehoben. Der resultierende Vergenzregler geht damit über die Bestimmung einer einzigen Disparität hinaus.
Insgesamt ergibt sich ein Stereoverfahren, das eine typische Skalenanalyse der Ergebnisse des Vergleichsmaßes beinhaltet. Das Verfahren stellt eine heterogene Strategie zur Vergenzregelung dar und reiht sich in Verfahren der Literatur ein.

Da das Problem der Zuordnung von Impulsen zu verantwortlichen Pixel eine zentrale Rolle bei der Analyse der Cepstrum-Antwort spielt, soll es genauer betrachtet werden.
Die Skalensuche des obigen Olson-Verfahrens zur Stereovergenz (siehe [TaOlMa94]) beinhaltet, den Cepstrum-Filter auf Fenster unterschiedlicher Größe um einen zentralen Aufpunkt anzuwenden. Impulse, die in den Cepstrum-Antworten für alle Fenstergrößen vorkommen, werden dem Aufpunkt zugeordnet, weil für den Impuls verantwortliche Bildstruktur um den Aufpunkt in allen Einzugsbereichen, insbesondere im kleinsten Fenster, enthalten sein muß.
Mit dieser Strategie lassen sich aber nur kleine Disparitäten für eine zentrale Bildregion ermitteln, weil die so detektierbaren Disparitäten vom minimalen Einzugsbereich des Cepstrum-Filters in der Skalensuche abhängen.
Um den Fall von größeren Disparitäten in der zentralen Bildregion detektieren zu können, werden bei Versagen der obigen Skalensuche die Impulse der Cepstrum-Antwort des Stereobildpaares nach fallendem Betrag geordnet untersucht. Es wird durch ein lokales Vergleichsmaß explizit überprüft, welche Pixel für den Impuls verantwortlich sind. Derjenige Impuls wird der zentralen Bildregion als Disparität zugeordnet, der in der entsprechenden Disparität bezüglich dem lokalen

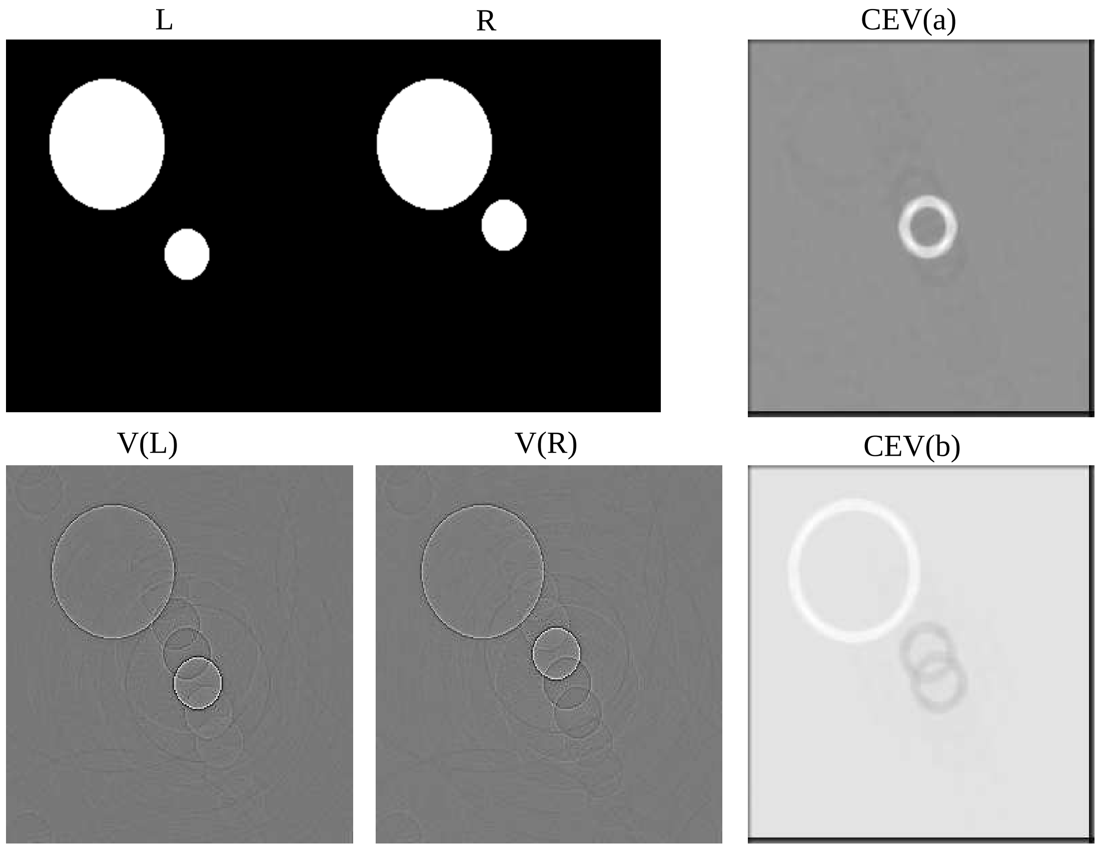


Abbildung 3.6: Der CEV-Filter des Olson-Algorithmus zur Vergenzsteuerung ermittelt für einen Impuls der Cepstrum-Antwort verantwortliche Pixel in L und R, wie am Beispiel der zwei intensivsten Impulsen a,b der Cepstrum-Antwort deutlich wird (helle Bildbereiche).Er besitzt diegleiche Wirkung wie eine lokale Kreuzkorrelation.

Vergleichsmaß am Aufpunkt einen signifikanten Wert aufweist (Schwellwertbildung).
Das lokale Vergleichsmaß trägt die Bezeichnung CEV (Cepstral-Equivalent-Prefilter) und beruht auf dem obigen Vorfilter V innerhalb des Cepstrum-Filters in der Interpretation als Autokorrelation (siehe Abschnitt 3.2.1). Der CEV-Filter berechnet die lokale Kreuzkorrelation von durch V gefilterten Signalen. Aus diesem Grund bezeichnet Olson den CEV-Filter als lokale Approximation des Cepstrum-Filters ([TaOlMa94]).

Das Prinzip der Zuordnung von Impulsen zu verantwortlichen Pixeln durch den CEV-Filter wird in Abbildung 3.6 deutlich. Dort ist ein synthetisches Stereobildpaar (L,R), die Antworten des Vorfilters $V(L)$, $V(R)$ sowie die Antwort des CEV-Filters (Einzugsbereich 8 Pixel) von $V(L)$ und $V(R)$ für die zwei stärksten Impulse a,b der Cepstrum-Antwort dargestellt. Helle Bereiche markieren die für

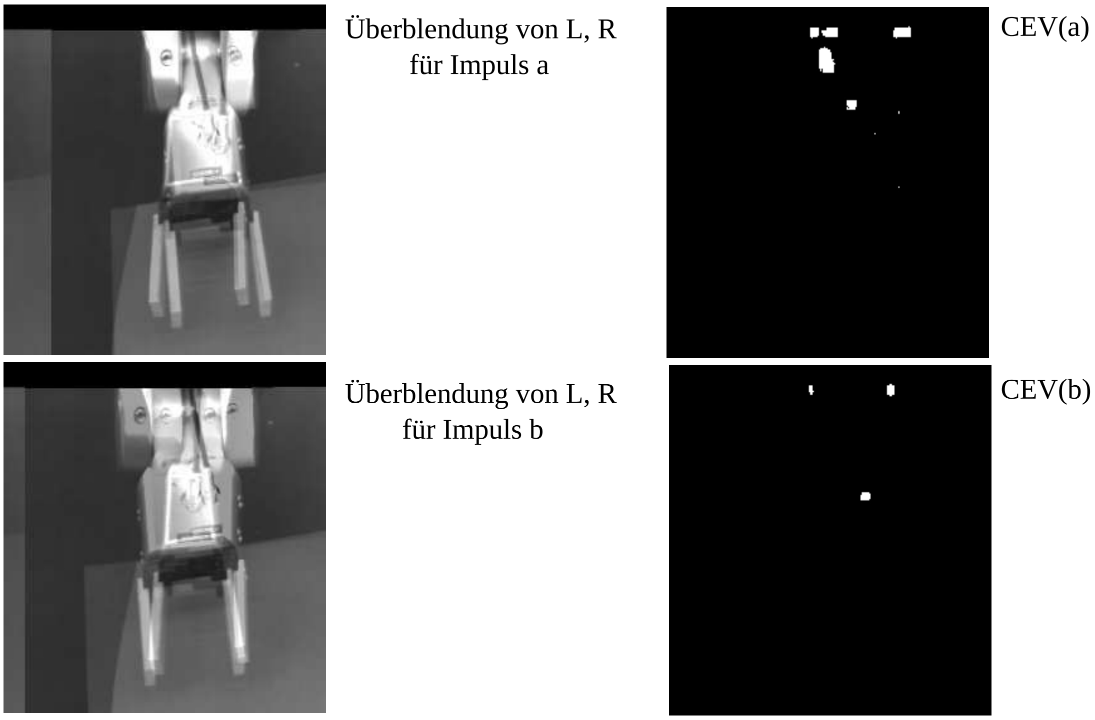


Abbildung 3.7: Dargestellt sind ein reales Stereobildpaar sowie die maximalen Beträge der CEV-Antwort für zwei ausgeprägte Impulse a,b der Cepstrum-Antwort. Im Gegensatz zu synthetischen Testbildern liefert der CEV-Filter unbefriedigende Resultate. In der CEV-Antwort weisen Pixel mit der Disparität b (Greiferzangen) eine geringere Filterantwort auf als Pixel mit einer tatsächlich anderen Disparität (Greiferbasis). Insofern versagt der CEV-Filter in seiner Funktion der Detektion von Pixeln, die für einen Cepstrum-Impuls verantwortlich sind.

einen Impuls verantwortlichen Pixel in entsprechender Disparität.
In Abbildung 3.7 wird analog von einem realen Stereobildpaar die Cepstrum-Antwort ermittelt. Für zwei ausgeprägte Impulse a,b sind die Überblendungen der Stereobilder sowie die betragsstärksten Antworten des CEV-Filters dargestellt.
Deutlich wird, daß der CEV-Filter im Fall des realen Stereobildpaars unerwünschte Antworten liefert. Im Fall der Disparität b werden für den Cepstrum-Impuls verantwortliche Bildstrukturen entsprechender Disparität (Greiferzangen) mit einer geringeren Filterantwort belegt als Bildstrukturen, die eine tatsächlich von b verschiedene Disparität besitzen wie die Greiferbasis. Dabei ist weniger der auftretende aber zu erwartende Alias-Fehler des lokalen Vergleichsmaßes negativ zu bewerten als die Bewertungsrangfolge von Bildstrukturen.
Auffällig ist, daß der CEV-Filter nur im Fall des synthetischen Bildpaars eine plausible Antwort liefert.

3.4.3 Entwicklung eines cepstrumbasierten Stereoverfahrens

Auf Grundlage von Literaturansätzen wird ein eigenes cepstrumbasiertes Stereoverfahren entwickelt, das sich für die gesuchte Systemkomponente eignet.
Das Verfahren verwendet eine an die Hardware des MV200-Systems angepaßte Version des Cepstrum-Filters mit einem globalen und lokalen Einzugsbereich. Auf dieser Basis wird eine Strategie der Skalensuche durchgeführt, die Alias-Fehler bei der Zuordnung unter Ausnutzung einer Reihenfolge der Dominanz von Bildstrukturen vermeidet. Die Dominanzreihenfolge zur Betrachtung von Disparitäten ist durch eine Menge von im Suchbereich der Cepstrum-Antwort detektierten Impulsen vorgegeben. Der Algorithmus beinhaltet eine Komponente zur Unterdrückung des systematischen Problems der Überlagerung von Impulsen der Cepstrum-Antwort.

Ausgangspunkt der Entwicklung soll der beschriebene Olson-Algorithmus zur Stereovergenz sein. Er ermittelt zwar nur einen einzigen Disparitätswert, ist aber bezüglich der Analysestrategie der Cepstrum-Antwort interessanter und problemgerechter aufgebaut als der einfache YESA-Algorithmus zur Messung von Stereodisparität. Weiterhin geht die heuristische Olson-Strategie über das Maß eines einfachen Vergenzreglers hinaus, weil mehrere Disparitäten im Stereobildpaar analysiert werden.
Dies legt die Weiterentwicklung geeigneter Strategien des Olson-Algorithmus zur Analyse der Cepstrum-Antwort durch ein eigenes Stereoverfahren nahe.
Der erste Schritt besteht in der Portierung des Cepstrum-Filters auf die Architektur des MV200-System. Weil die benötigte Fouriertransformation dort nicht implementierbar ist, wird sie durch eine analoge globale Kreuzkorrelation ersetzt, die in einem horizontalen Suchbereich ermittelt wird. Cepstrum-Impulse finden ihre Entsprechung in lokalen Maxima der Korrelation. In gleicher Weise wird an Stelle des CEV-Filters des Olson-Algorithmus eine lokale Kreuzkorrelation verwendet.
Der Wegfall teilweise unplausibler Filterantworten im Fall des CEV-Filters und systematischer Schwächen bei der Interpretation der Cepstrum-Antwort erleichtern die Analyse von Ergebnissen der portierten Vergleichsmaße. Zu beachten ist, daß der Übergang zu analogen Korrelationsmaßen ebenfalls den Wegfall der Vorfilterung von Cepstrum- und CEV-Filter mit sich bringt. Da die Vorfilterung eine Verstärkung von Impulsen bewirkt, dürfen in einer Analysestrategie von Vergleichsergebnissen keine Beträge betrachtet werden, damit im gesuchten Stereoverfahren die Korrelationsmaße ein zu cepstralen Maßen äquivalentes Verhalten zeigen.
Als Suchstrategie werden Komponenten des Olson-Algorithmus aufgegriffen. Das Korrespondenzproblem verlagert sich im Fall des Cepstrum-Filters auf die Aufgabe der Auffindung von für einen Cepstrum-Impuls verantwortlichen Bildstruk-

turen.
Um diese Aufgabe zu lösen, werden im Olson-Algorithmus Impulse nach fallenden Beträgen geordnet betrachtet. Für jeden Impuls beziehungsweise jede Disparität werden verantwortliche Bildstrukturen durch hohe Beträge des lokalen Vergleichsmaßes identifiziert.
Diese Verfahrenskomponenten können in der Bindung eines lokalen Vergleichs an einen globalen Vergleich zu einer Skalensuchstrategie ausgedehnt werden, obwohl sie im Olson-Algorithmus eine andere Funktion erfüllen.
Die beschriebene Zuordnung von Pixeln ausschließlich auf Grundlage eines lokalen Vergleichs ist anfällig gegen Alias-Fehler beziehungsweise lokale Extrema des Vergleichsmaßes.
Die richtige Auswahl lokaler Extrema erfolgt in dieser Suchstrategie durch die Reihenfolge ihrer Betrachtung in der vom globalen Vergleichsmaß vorgegebenen Reihenfolge von Disparitäten, die sich nach dem Betrag der zugehörigen Korrelationsmaxima richtet. Die Ordnung nach fallenden Beträgen korrespondiert mit einer Rangfolge der Dominanz von Bildstrukturen in der jeweiligen Disparität (siehe 3.3.2). So wird für ein Pixel immer das lokale Korrelationsmaximum ausgesucht, das eine Entsprechung im globalen Kontext findet.
Das Kriterium für eine Zuordnung von Pixeln auf Grundlage der lokalen Korrelation soll an Stelle des Betrags im Olson-Algorithmus das Vorhandensein eines Korrelationsmaximums im Suchbereich sein. Eine Schwellwertbildung wird durch einen lokalen Maximumdetektor ersetzt. Implizit beinhaltet das Zuordnungskriterium ein Vertrauensmaß, denn ein Extremum des Vergleichsmaßes bedingt Struktur in dessen Umgebung.
Mit dieser Suchstrategie wird das Prinzip der Dominanz von Bildstruktur bei einem globalen Vergleich ausgenutzt. Die photometrische Dominanz von Bildstrukturen über schwächere Bildstrukturen jedoch kann bewirken, daß Extrema eines globalen Vergleichsmaßes für bestimmte Disparitäten überlagert werden (siehe Abschnitt 3.3.2). Aus diesem Grund ist der Cepstrum-Filter nicht in der Lage, kontinuierliche Disparitätsverläufe zu detektieren.
Es ist also bei einer Betrachtung von Disparitäten in einer Dominanzreihenfolge von Bildstrukturen das Ziel gegeben, nach Betrachtung einer bestimmten Disparität den Einfluß der zugehörigen dominanten Bildstruktur auf die globale Korrelation zu beseitigen. Durch die betreffende Bildstruktur überlagerte Impulse sind dann in den globalen Korrelationsergebnissen detektierbar.
Dies geschieht in einem Prozeß der Ausmaskierung von Pixeln, die durch die lokale Korrelation zugeordnet wurden. Sie werden nicht weiter betrachtet und bei erneuter Ermittlung der globalen Korrelation im Suchbereich nicht miteinbezogen. Man kann die Ausmaskierung als Ausdünnung des globalen Templates ansehen.
Dies hat zur Folge, daß bei der Neuberechnung der globalen Korrelation ein absolutes Maximum im Suchbereich, zu dessen Disparität Pixel durch lokale Korrelation zugeordnet wurden, verschwindet und stattdessen andere Maxima der

globalen Korrelation im Suchbereich entstehen.

Insgesamt wird so die Überlagerung von Impulsen der Cepstrum-Antwort vermieden, so daß kontinuierliche Disparitätsverläufe detektierbar sind.
Abbildung 3.8 bietet hierfür ein Beispiel. Dort ist für ein Stereobildpaar eine Sequenz von Ausmaskierungen von Bildstrukturen dargestellt, die durch das entwickelte Stereoverfahren zugeordnet wurden. Die Zuordnungen finden sich in der resultierenden Disparitätskarte wieder.
Wie beschrieben erfolgen die Zuordnungen von Pixeln zu Disparitäten sukzessiv durch das Zusammenspiel von globaler und lokaler Korrelation in einer Dominanzreihenfolge von Bildstrukturen unter Berücksichtigung ausmaskierter Pixel. Die Dominanzreihenfolge spiegelt sich in der Sequenz in einer von Bild zu Bild sinkenden Anzahl von Zuordnungen wider. Kontinuierliche Disparitätsverläufe im Bereich der Unterlage im Stereobildpaar werden detektiert.
Das Bildpaar enthält ebenfalls Strukturen, die die Gefahr von Alias-Fehlern bergen, wie beispielsweise die zwei Greiferzangen. Eine Vertauschung in der Zuordnung wird vermieden, weil die Greiferzangen durch die Suchstrategie im globalen Zusammenhang zugeordnet werden.

Der entwickelte Stereoalgorithmus besitzt einen hohen Rechenaufwand ($O(N^2)$), der sich mit der Kapazität der vorliegenden Pipeline-Architektur rechtfertigen läßt.
Die Komponenten des insgesamt entwickelten Stereoalgorithmus auf Basis des Cepstrum-Filters werden nun in ihrem Zusammenwirken vorgestellt. Das resultierende Stereoverfahren soll die Bezeichnung *CEPSIM* erhalten:

Gegeben sei ein Stereobildpaar (L,R), sowie Datenstrukturen $Maske_L, Maske_R$ zur Ausmaskierung von Pixeln und eine Disparitätskarte D derselben Dimension. Ein Suchbereich legt mögliche Disparitäten fest.

1. Für alle möglichen Positionen im Suchbereich wird die globale Korrelation berechnet, wobei ausmaskierte Pixel in L und R unberücksichtigt bleiben. Die Suchposition mit der maximalen globalen Korrelation $disp_{max}$ wird ausgewählt.

2. Für $disp_{max}$ wird bezüglich einer lokalen Korrelation überprüft, welche Pixel in (L,R) bezüglich benachbarter Suchpositionen ein lokales Maximum aufweisen. Betroffene Pixel werden in D eingetragen sowie in $Maske_L$ und $Maske_R$ ausmaskiert.

3. Abbruchbedingung: Wieviele Pixel wurden neu ausmaskiert ? Ist der Suchbereich vollständig abgearbeitet ?

Die Abbruchbedingung folgt aus der Tatsache, daß die Anzahl der neu ausmaskierten Pixel monoton fallend ist, weil sukzessive Bildstrukturen mit fallendem Strukturgehalt einander zugeordnet werden.

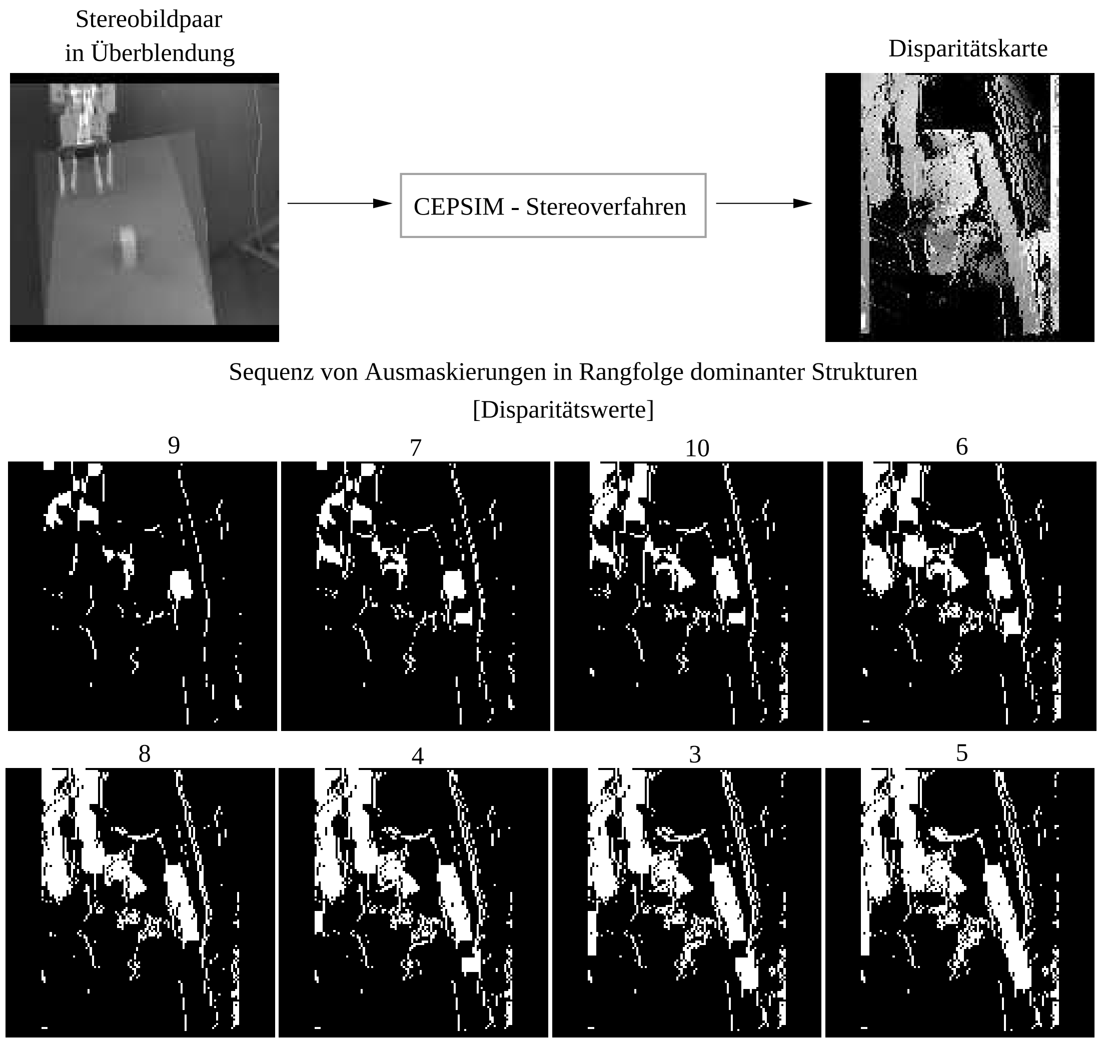


Abbildung 3.8: Die Suchstrategie des entwickelten cepstrumbasierten Stereoverfahrens beinhaltet eine Komponente zur Vermeidung der Überlagerung von Impulsen des Cepstrum-Filters. Sie beruht auf der sukzessiven Ausmaskierung von zugeordneten Pixeln einer betrachteten Disparität in einer Dominanzreihenfolge von Bildstrukturen. Ausmaskierte Bildstrukturen werden bei Betrachtung weniger dominanter Bildstrukturen einer anderen Disparität nicht in eine Korrelation miteinbezogen, so daß deren photometrische Dominanz vermieden wird.

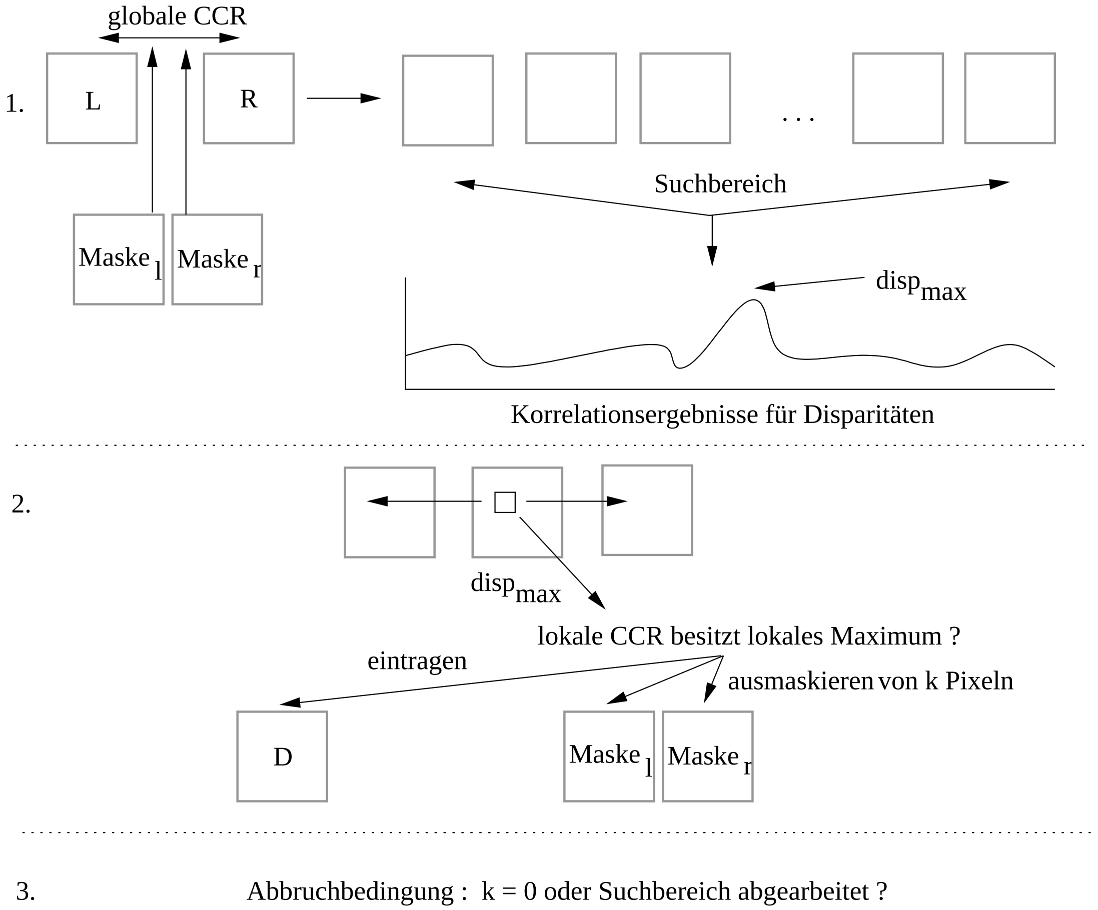


Abbildung 3.9: Der Aufbau des CEPSIM-Stereoverfahrens. Der Algorithmus gliedert sich zunächst in eine Phase der Berechnung von globaler Korrelation zwischen L und R unter Ausmaskierung betrachteter Pixel. In der Suchposition mit maximaler globaler Korrelation werden anschließend k korrespondierende Pixel durch lokale Korrelation ausmaskiert und in einer Disparitätskarte vermerkt. Der Prozeß der sukzessiven Ausmaskierung von Pixeln vermeidet Probleme der Überlagerung von Impulsen des Cepstrum-Filters.

Die Abbruchbedingung bewirkt, daß nur solange Positionen im Suchbereich betrachtet werden, wie Bildstruktur in entsprechender Disparität vorliegt. In der Praxis wird so im allgemeinen nur ein relevanter Teil des Suchbereichs abgearbeitet. So ist die Laufzeit des CEPSIM-Verfahrens strukturgebunden.
Keine Disparität wird mehrfach betrachtet, weil die Bildbereiche, die ein zugehöriges Maximum der globalen Korrelation hervorgerufen haben, ausmaskiert werden. So terminiert der Algorithmus spätestens nach Betrachtung aller möglichen Disparitäten. Diese Laufzeitschranke bildet die Grundvoraussetzung für die Einhaltung eines festen Taktes und damit die Echtzeitfähigkeit des Algorithmus.
In Abbildung 3.9 ist der Algorithmus grafisch dargestellt. Seine Struktur ist vergleichsweise einfach und regelmäßig. Dies ist ein Erfordernis für die Implementierung auf dem MV200-System.

Das entwickelte cepstrumbasierte Stereoverfahren beseitigt Schwächen von Verfahren der Literatur. Die wichtigste Verbesserung liegt in der Fähigkeit zur Detektion kontinuierlich verteilter Disparitäten, die eine Voraussetzung zur Verwendung des Cepstrum-Filters in einem Stereoverfahren darstellt.
Die Auswahl eines einzigen Cepstrum-Impulses nach dem Betrag hat sich im Fall des YESA-Algorithmus als nachteilhaft erwiesen, während die Analysestrategie des CEPSIM-Algorithmus der Problematik der Interpretation von Impulsbeträgen besser gerecht wird.
Es muß erwähnt werden, daß die verwendete Suchstrategie zwar elegant und praktikabel ist, wie sich unten herausstellen wird, daß sie jedoch eine Heuristik darstellt. Die heuristische Komponente kommt ins Spiel, wenn für ein Maximum der globalen Korrelation alle Pixel ausmaskiert werden, die an entsprechender Suchposition ein Maximum der lokalen Korrelation vorweisen. Im allgemeinen sind dies die Pixel, die das absolute Maximum der globalen Korrelation verursacht haben, sonst würde es an einer anderen Suchposition entstehen. Es kann jedoch der Fall eintreten, daß vereinzelte Pixel zugeordnet werden, die ihre Korrespondenz in einer anderen Suchposition besitzen. In diesem Fall wird durch einen Alias-Fehler eine falsche Zuordnung getroffen.
Die heuristische Komponente betrifft also das Problem der Zuordnung von Cepstrum-Impulsen zu verantwortlichen Pixeln und findet sich im beschriebenen Olson-Algorithmus zur Stereovergenz analog wieder.

Der Algorithmus wurde auf einer Workstation implementiert. In einem Takt der CEPSIM-Implementierung werden einmalig lokale (7x7) und globale (128x128) Korrelationsergebnisse in einem Suchbereich von ± 16 Pixel berechnet. Auf diese Ergebnisse wird während der sequentiellen Suche in den Suchpositionen zurückgegriffen, was eine erhebliche Effizienzsteigerung gegenüber der wiederholten Berechnung der Korrelation für jede Suchposition darstellt.
Auf einer Ultra-Sparc-Workstation ergibt sich zur Berechnung einer Disparitätskarte für ein Stereobildpaar der Dimension 128x128 eine Taktdauer von maximal 20 Sekunden. In Kapitel 5 sind durch das CEPSIM-Verfahren erzeugte Dispa-

ritätskarten von Testbildpaaren dargestellt.

3.5 Eignung des Cepstrum-Filters für Stereoverfahren

Die Betrachtung mehrerer Stereoverfahren auf Grundlage des Cepstrum-Filters zur Vergenzsteuerung oder Ermittlung von Stereodisparität beantwortet einige Fragen bezüglich dessen Eignung als Vergleichsmaß. Andere Resultate ergeben sich aus der vorausgegangenen Analyse der Eigenschaften des Cepstrum-Filters.
Im Fall des YESA-Algorithmus zeigt sich, daß die Koppelung von Einzugsbereich und Suchbereich des Cepstrum-Filters negativ zu bewerten ist. Mit zunehmendem Betrag der Disparität wird ihre Messung ungenauer, weil der Einzugsbereich des Filters wächst.
Ein wichtiges Ergebnis dieser Untersuchung liegt in der Erkenntnis, daß sich durch einfache Analyse von Impulsen des Cepstrum-Filters keine kontinuierlichen Disparitätskarten erzeugen lassen.
In Abschnitt 3.3 wurde gezeigt, daß die theoretische Interpretation der Cepstrum-Antwort in der Praxis systematische Schwächen zeigt.
Der Cepstrum-Filter eignet sich strukturell besser zur Vergenzsteuerung, bei der nur eine Disparität für einen Fixationspunkt ermittelt werden soll, als zur Messung von Stereodisparität.
Auch zur Vergenzsteuerung ist die Antwort des Cepstrum-Filters nicht ohne eine aufwendige Analysestrategie interpretierbar, wie die unterschiedlichen Heuristiken des Olson-Algorithmus zeigen.
Dabei liegt das Hauptproblem des Filters in der Zuordnung von Impulsen zu Pixeln, die eine entsprechende Disparität besitzen (siehe Olson-Algorithmus [TaOlMa94]). Das Problem des optimalen Einzugsbereichs beim Vergleich von Struktur wird auf die Zuordnung von Impulsen zu Pixeln verlagert.
Damit ist die einfache Struktur des Cepstrum-Filters allein als Vergleichsmaß in einem Stereoverfahren nicht ausreichend.

Kapitel 4

Untersuchung der Korrelation

Korrelationsbasierte Vergleichsmaße werden in gleicher Weise wie der Cepstrum-Filter auf ihre Eignung für das gesuchte Stereoverfahren hin untersucht.

4.1 Korrelationsvarianten

Die Korrelation hat als klassisches Vergleichsmaß ihre Eignung zur Verwendung in Stereoverfahren bewiesen. Es gilt, Eigenschaften verschiedener Varianten der Korrelation zu analysieren, um Kriterien zur Auswahl einer Variante für eine bestimmtes Stereoverfahren zu gewinnen.

Die Kreuzkorrelation zweier Signale ist eine statistische Maß für ihre Ähnlichkeit. Für zwei Templates $T_L(i,j)$, $T_R(k,l)$ der Dimension MxM um Aufpunkte $L(i,j)$ und $R(k,l)$ ist die Kreuzkorrelation (KK) gegeben durch :

$$KK(T_L(i,j), T_R(k,l)) \quad = \quad \sum_{x,y=0}^{M-1} T_L(i,j)(x,y) T_R(k,l)(x,y)) \qquad (4.1)$$

Die Variablen x, y seien lokale Templatekoordinaten bezüglich $T_L(i,j)$, $T_R(k,l)$. Um einen Einblick in die Eigenschaften von Korrelationsmaßen zu gewinnen, soll die in Abbildung 4.1 dargestellte Korrelation eines fest gewählten Templates $T_L(i)$ mit Templates $T_R(k)$ ($k \in [k_{min}, k_{max}]$) eines eindimensionalen Testsignals T_R betrachtet werden.
T_R beinhaltet mehrere Strukturen, von denen eine (mittelgroßes Rechteck) $T_L(i)$ entspricht, die in ihrer Struktur und Signalenergie von $T_L(i)$ bis hin zu einem Rauschsignal unterschiedlich stark abweichen. Zwischen den Strukturelementen wurde die Signalenergie auf Null gesetzt, um eine gegenseitige Beeinflussung unterschiedlicher Strukturen durch Überlappung in den Templates zu vermeiden.
Die einfache Kreuzkorrelation KK liefert ein lokales Maximum an der Position von $T_L(i)$ in T_R.

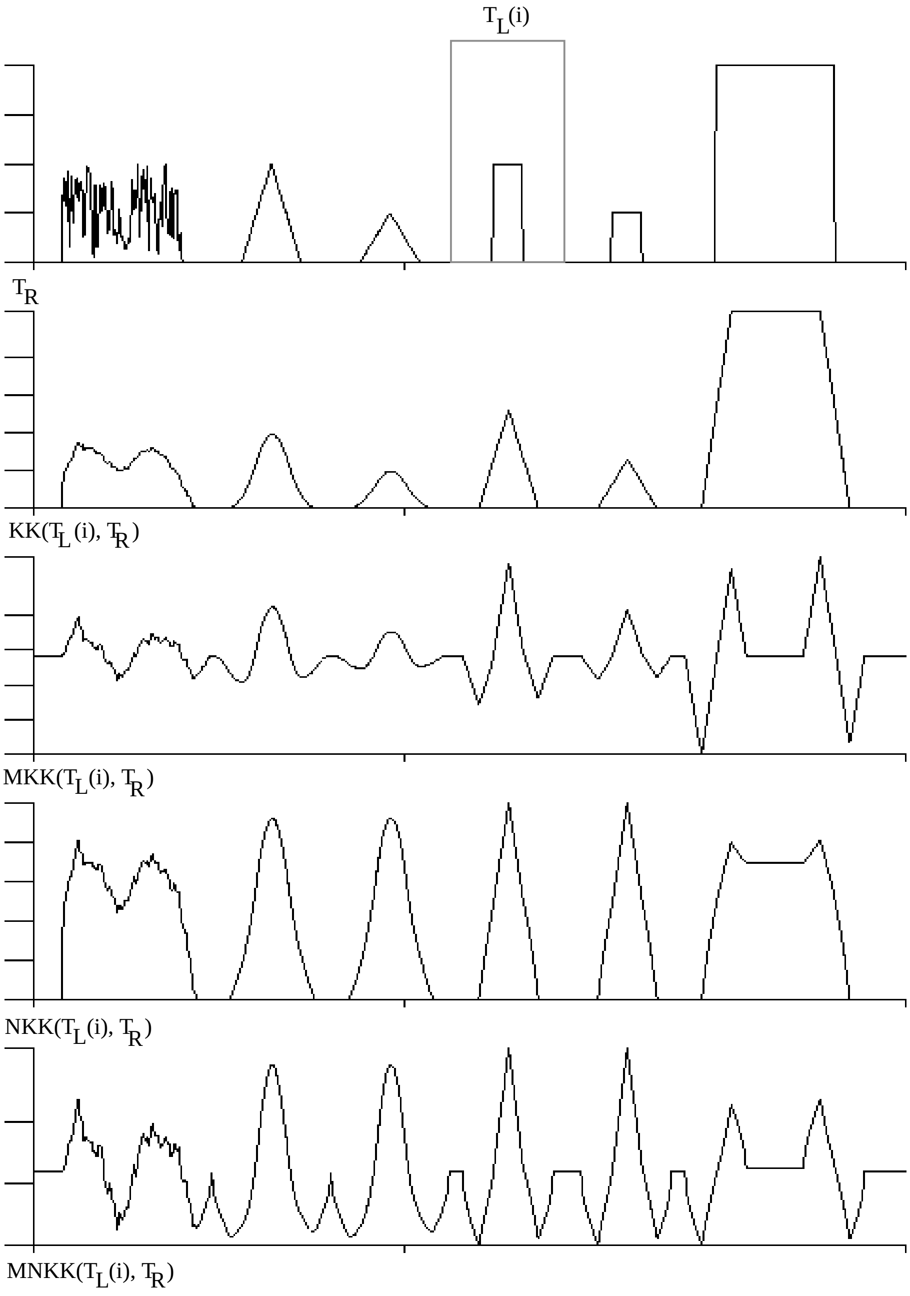


Abbildung 4.1: Dargestellt sind die durch Kreuzkorrelation ermittelten Vergleichsergebnisse eines Templates $T_L(i)$ mit einem eindimensionalen Testsignal T_R. Die Kreuzkorrelation (KK) liegt in den Varianten der Normierung (MKK), des Bezugs auf den mittleren Grauwert (NKK) sowie einer Kombination aus Normierung und Mittelwertbezug vor (MNKK).

Allerdings besitzt KK einige Eigenschaften, die intuitiven Anforderungen an ein Vergleichsmaß widersprechen.
Beispielsweise wird KK an der Position des Templates $T_R(k)$ mit maximaler Signalenergie im Suchbereich unabhängig von der Struktur von $T_L(i)$ ein absolutes Maximum einnehmen, was direkt aus Gleichung 4.1 ablesbar ist. Der Betrag von KK ist also in erster Linie von der lokalen Signalenergie in T_R abhängig.
Außerdem ist KK nicht in der Lage, Struktur in Templates von fehlender Struktur zu trennen. Dies ist von dem Fall zu unterscheiden, daß in beiden Templates Struktur fehlt und ein Vergleich generell nicht durchführbar ist (siehe Abschnitt 2.3). Nur bei Vorhandensein von Struktur sowohl in $T_L(i)$ als auch in $T_R(k)$ mit ähnlicher Signalenergie sind Beträge von KK vertrauenswürdig.
Im Fall von KK in Abbildung 4.1 sind die angesprochenen Defekte erkennbar. Die Insensibilität für Struktur wird daran deutlich, daß KK an der Stelle der Rauschstruktur in T_R eine höhere Antwort liefert als an der Stelle des kleinen Rechtecks, das strukturell $T_L(i)$ ähnelt. An der Stelle des großen, breiten Rechtecks liefert KK das absolute Maximum, weil seine Signalenergie maximal ist.
Die Ermittlung von KK entspricht arithmetisch der Faltung von $T_L(i)$ mit T_R.
Als Fazit bleibt, daß KK als Vergleichsmaß ungeeignet ist, falls bei einer Analyse von Vergleichsergebnissen in irgendeiner Form der Betrag eine Rolle spielt. Die einfache Beschaffenheit von KK reicht aus, falls nur lokale Maxima im Suchbereich analysiert werden.

Die unerwünschte Insensibilität gegenüber Struktur beseitigt der Bezug von KK auf den mittleren Grauwert in $T_L(i,j)$ und $T_R(k,l)$ durch die mittlere Kreuzkorrelation MKK:

$$MKK(T_L(i,j), T_R(k,l)) =$$

$$\sum_{x,y=0}^{M-1} (T_L(i,j)(x,y) - \overline{T}_L(i,j))(T_R(k,l)(x,y) - \overline{T}_R(k,l))$$

mit

$$\overline{T}_L(i,j) = \text{mittlerer Grauwert des Templates } T_L(i,j)$$

$$\overline{T}_R(k,l) = \text{mittlerer Grauwert des Templates } T_R(k,l)$$

Durch MKK werden nun strukturlose Regionen von T_R mit Null bewertet, weil dort das Template nicht von seinem Mittelwert abweicht. In Abbildung 4.1 wird dieser Effekt im Fall des großen Rechtecks deutlich. Der Bezug auf die Mittelwerte hat den Effekt einer der Korrelation vorausgehenden Ableitung der Grauwertfunktionen der Templates.
Durch die geringe Bewertung strukturloser Bildregionen generell benötigt MKK kein weiteres Vertrauensmaß.
Allerdings verletzt auch diese Bewertung von Templates intuitive Forderungen an ein Vergleichsmaß, denn auch bei fehlender Struktur sollte der Vergleich eines hellen Templates von einem dunklen anders ausfallen als der zweier dunkler

Templates.
Bei MKK bleibt der dominante Einfluß der Signalenergie von T_R auf den Betrag der Korrelation erhalten. Dieses Manko behebt die Normierung der Korrelationsbeträge auf ihre Signalenergie durch die normierte Korrelation NKK:

$$NKK(T_L(i,j), T_R(k,l)) =$$

$$\frac{\sum\limits_{x,y=0}^{M-1} T_L(i,j)(x,y) T_R(k,l)(x,y)}{\sqrt{\sum\limits_{x,y=0}^{M-1} T_L(i,j)(x,y)^2 \sum\limits_{x,y=0}^{M-1} T_R(k,l)(x,y)^2}} \tag{4.2}$$

Die Beträge von NKK liegen im Intervall [-1,1]. Sie hängen allein von der Übereinstimmung von Bildstruktur zwischen $T_L(i,j)$ und $T_R(k,l)$ und nicht mehr von der Ausprägung der Struktur ab.
In Abbildung 4.1 werden durch NKK die bis auf ihre Signalenergie identischen Dreiecke und Rechtecke in T_R auch identisch bewertet. In den Korrelationsergebnissen nimmt NKK beim Rechteck $T_L(i)$ nun ein globales Maximum ein. Damit macht eine Analyse der Korrelationsbeträge Sinn.
Allerdings verschlechtert sich die Konditionierung der Korrelationsergebnisse in Gegenwart von schwachen Strukturen in $T_L(i,j)$ und $T_R(k,l)$. Bei der Normierung wird der geringe Betrag der Korrelation in Folge der Division durch die geringe Signalenergie auf Dimensionen wie für ausgeprägte Strukturen erhöht. Kleine Veränderungen eines schwachen Signals, zum Beispiel durch Rauschen, bewirken starke Schwankungen in den Korrelationsergebnissen. Dieser Effekt ist in der Praxis bei Verwendung von NKK häufig zu beobachten. Die Korrelationsvariante benötigt ein strukturgebundenes Vertrauensmaß.

Verknüpft man die Eigenschaften des Bezugs auf den mittleren Grauwert und der Normierung, so erhält man die mittlere, normierte Kreuzkorrelation $MNKK$.

$$MNKK(T_L(i,j), T_R(k,l)) =$$

$$\frac{\sum\limits_{x,y=0}^{M-1} (T_L(i,j)(x,y) - \overline{T}_L(i,j))(T_R(k,l)(x,y) - \overline{T}_R(k,l))}{\sqrt{\sum\limits_{x,y=0}^{M-1} (T_L(i,j)(x,y) - \overline{T}_L(i,j))^2 \sum\limits_{x,y=0}^{M-1} (T_R(k,l)(x,y) - \overline{T}_R(k,l))^2}} \tag{4.3}$$

Das Vergleichsmaß besitzt die Eigenschaften von MKK und NKK in einer Kombination. In Abbildung 4.1 wird deutlich, daß darüberhinaus das Verhältnis der Beträge von $MNKK$ zwischen optimalem Wert für die richtige Struktur $T_L(i)$ und den übrigen Werten für falsche Strukturelemente wie Rauschen besser wird. Die Kombination beider Eigenschaften ist also nicht ohne gegenseitige Beeinflussung.

In der Literatur findet der quadratische Fehler (QF) ebenso Verwendung wie die Korrelation. Arithmetisch unterscheidet sich der quadratische Fehler von der Korrelation lediglich durch die Verknüpfung von Grauwerten zweier Templates durch Subtraktion an Stelle von Multiplikation.
Der quadratische Fehler wird analog zur Korrelation auf den mittleren Grauwert im Template bezogen sowie mit einer Normierung versehen. So besteht eine enge Verwandtschaft beider Vergleichsmaße. Deshalb wird auf eine analoge Untersuchung des quadratischen Fehlers verzichtet, denn die Unterschiede zwischen beiden Vergleichsmaßen sind im allgemeinen nicht qualitativer Natur.

4.2 Die Korrelation als Vergleichsmaß

In Abschnitt 4.1 wurden unterschiedliche Eigenschaften verschiedener Korrelationsvarianten vorgestellt. Die Auswahl einer der Varianten als Vergleichsmaß innerhalb eines Stereoverfahrens hängt von den zur Modellierung des Disparitätsproblems getroffenen Annahmen über die Beschaffenheit der Eingangssignale L und R sowie von der angestrebten Analysestrategie der Korrelationsergebnisse ab. Eine allgemeine Bewertung der Korrelationsvarianten muß so einer aufgabenbezogenen Bewertung weichen, weil alle Korrelationsvarianten die oben beschriebenen Vor- und Nachteile bezüglich intuitiver Kriterien besitzen.
Ein objektiveres Kriterium für eine Bewertung von Vergleichsmaßen könnte man in der Frage formulieren, ob aus der Korrelationsantwort für $T_L(i,j)$ in T_R die Position des korrespondierenden Templates $T_R(k,l)$ eindeutig ermittelbar ist, falls dies auf photometrischer Basis möglich ist. Auch bezüglich dieser objektiveren Anforderung an ein Vergleichsmaß lassen sich für die Korrelationsvarianten fatale Fälle konstruieren.
In [Fua93] werden Vergleichsmaße basierend auf der Korrelation und dem quadratischen Fehler bezüglich intuitiver Kriterien experimentell verglichen. Es handelt sich dabei um die vorgestellten Korrelationsmaße NKK, $MNKK$ und den quadratischen Fehler in den Varianten einer Normierung (NQF) sowie eines zusätzlichen Mittelwertbezugs ($MNQF$). Das Fazit ist, daß NQF, $MNKK$ und $MNQF$ ähnliche Ergebnisse liefern, während NQF wesentlich schlechter abschneidet.
In der Literatur wird der quadratische Fehler aber ebenso wie Korrelation als Vergleichsmaß verwendet. Es lassen sich auch Fälle konstruieren, bei denen der quadratische Fehler eine Zuordnung eines Templates im Suchbereich ermöglicht, während Korrelationsmaße versagen, wie im Fall von linearen Grauwertverläufen. In Abbildung 4.2 ist ein entsprechendes eindimensionales Testsignal T_R mit einem Template $T_L(i)$ aus T_R dargestellt. Die Korrelationsvarianten liefern undifferenzierbare Vergleichsergebnisse, während der quadratische Fehler QF ein eindeutiges Minimum im Suchbereich bei $T_L(i)$ liefert.

Aufgrund der geschilderten Eigenschaften der verschiedenen Korrelationsvarian-

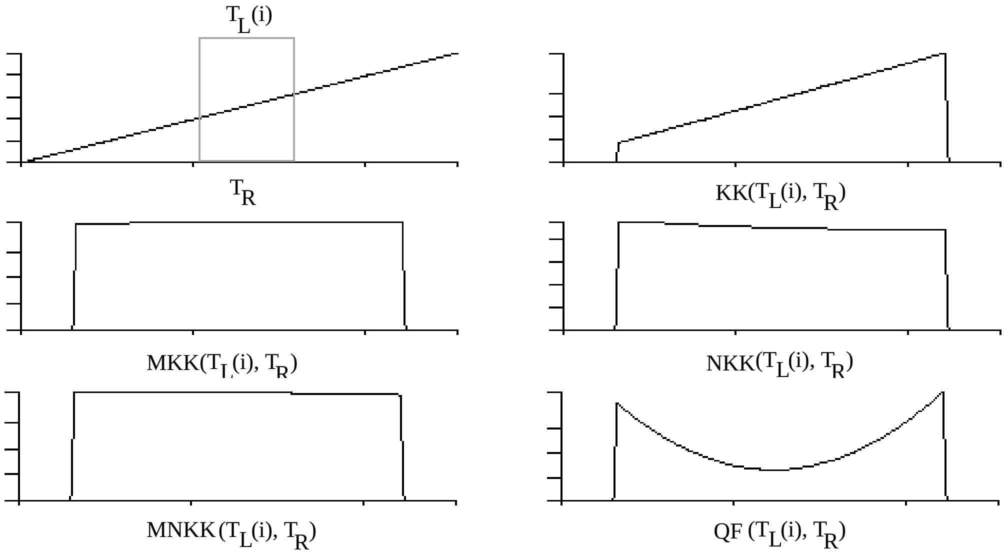


Abbildung 4.2: Lineare Grauwertverläufe (T_R) stellen einen Problemfall für Korrelationsmaße dar. In den Korrelationsantworten auf das Template T_L läßt sich kein Maximum detektieren, während der quadratische Fehler ein eindeutiges Minimum an der korrekten Suchposition liefert.

ten lassen sich ihre Anwendungsgebiete genauer spezifizieren.
Die Verwendung von KK oder QF kann beispielsweise für ein Stereoverfahren, das mit großen Templates arbeitet und örtlich gleichverteilte Signalintensitäten annimmt, ausreichend sein, weil es den Vorteil eines geringen Berechnungsaufwandes besitzt. Eine Analyse der Vergleichsergebnisse sollte sich auf eine einfache Detektion lokaler Maxima beschränken.
Eine Normierung der Korrelation (NKK) ist angemessen, falls beispielsweise unterschiedliche Beleuchtungsverhältnisse im Stereobildpaar angenommen werden oder falls mit variablen Fenstergrößen gearbeitet wird, deren Beträge verknüpft werden sollen.
Der Bezug auf den mittleren Grauwert ist erforderlich, falls eine nachfolgende Analysestrategie von Vergleichsergebnissen sensibel auf fehlende Struktur in Template und Suchbereich reagiert.

Grundsätzlich ist anzumerken, daß eine Verwendung von Schwellwerten für Korrelationsergebnisse vermieden werden sollte. Sie stellen zusätzliche Parameter eines Stereoverfahrens dar, und eine Analysestrategie basierend auf Korrelationsbeträgen erweist sich als weniger robust als beispielsweise die Orientierung an lokalen Maxima der Korrelation.
Insgesamt sind Korrelationsmaße insofern vielseitig, als daß sie in der Bildverarbeitung in Verbindung mit den meisten existierenden Strategien zur Korrespondenzsuche verwendet werden.

4.3 Stereoverfahren auf Basis der Korrelation

4.3.1 Das Greedy-Standardverfahren

In der Literatur ist die in Abschnitt 2.2.2 beschriebene Greedy-Strategie auf Basis von Korrelation ein Standardansatz.

Ein entsprechendes Stereoverfahren unter Verwendung von Gauß-Pyramiden P_l, P_r eines Stereobildpaars (L,R) wird wie folgt implementiert:
Für jedes Template $T_L(i,j)$ wird in der Pyramidenstufe i in einem Suchbereich das Template $T_R(k,l)$ mit maximaler Korrelation ausgewählt und mit der ermittelten Disparität der Suchbereich in der Stufe $i-1$ initialisiert. Der Suchbereich halbiert sich von Stufe zu Stufe, und die Disparitätsergebnisse werden wie in Abschnitt 2.2.2 beschrieben verfeinert. Die Suche wird von der maximalen Pyramidenstufe bis hin zur vollen Auflösung durchgeführt (Top-Down). Zur Interpretation einer Disparitätskarte als Ausgabe des Algorithmus bedarf es eines Vertrauensmaßes, weil in der Suchstrategie keine Bewertung der Zuordnung erfolgt.
Als Vergleichsmaß für das Template-Matching wird $MNKK$ verwendet. Der absolute Vergleich von Korrelationsbeträgen erfordert eine Normierung. Der Mittelwertbezug soll das Fehlen eines Vertrauensmaßes bei der Zuordnung in einer Pyramidenstufe insofern kompensieren, als das Vergleichsmaß sensibel gegenüber dem Vorhandensein von Struktur wird. Andernfalls wächst die Gefahr, daß in hohen Pyramidenstufen aufgrund nicht vertrauenswürdiger Korrelationsergebnisse ein falscher Pfad im Lösungsbaum der möglichen Zuordnungen eingeschlagen wird (siehe Abschnitt 2.2.2), so daß eine korrespondierende Zuordnung nicht mehr gefunden werden kann.
In der Interpretation der Suchstrategie als Optimierungsproblem (siehe 2.2) wird für jedes Pixel aus L im Baum aller möglichen Zuordnungen durch die Pyramidenstufen genau ein Pfad verfolgt, der in einer Top-Down-Strategie von der Wurzel bis hin zu einer Position im Suchbereich führt.
Der Algorithmus ist in Abbildung 4.3 dargestellt. In Bezug auf die verwendete Suchstrategie soll er die Bezeichnung *GREEDY* erhalten.

Abbildung 4.4 zeigt einen Fall auf, bei dem die Greedy-Suchstrategie versagt.
Dargestellt ist zu einen Template aus L der Suchbereich in R in vier Pyramidenstufen. Daneben befinden sich die Korrelationsantworten des Templates im Suchbereich in den jeweiligen Pyramidenstufen, wobei die Suchbereiche hoher Pyramidenstufen so auf den Suchbereich der untersten Pyramidenstufe gestreckt werden, daß entsprechende Suchpositionen direkt übereinander liegen. Die Struktur spiegelt den Lösungsbaum aller für das Template möglichen Zuordnungen wieder.
Als Bildstrukturen sind ein dominierendes großes Rechteck und ein kleines Rechteck im Suchbereich enthalten. Es gilt eine Korrespondenz für das kleine Rechteck im Template aufzufinden.

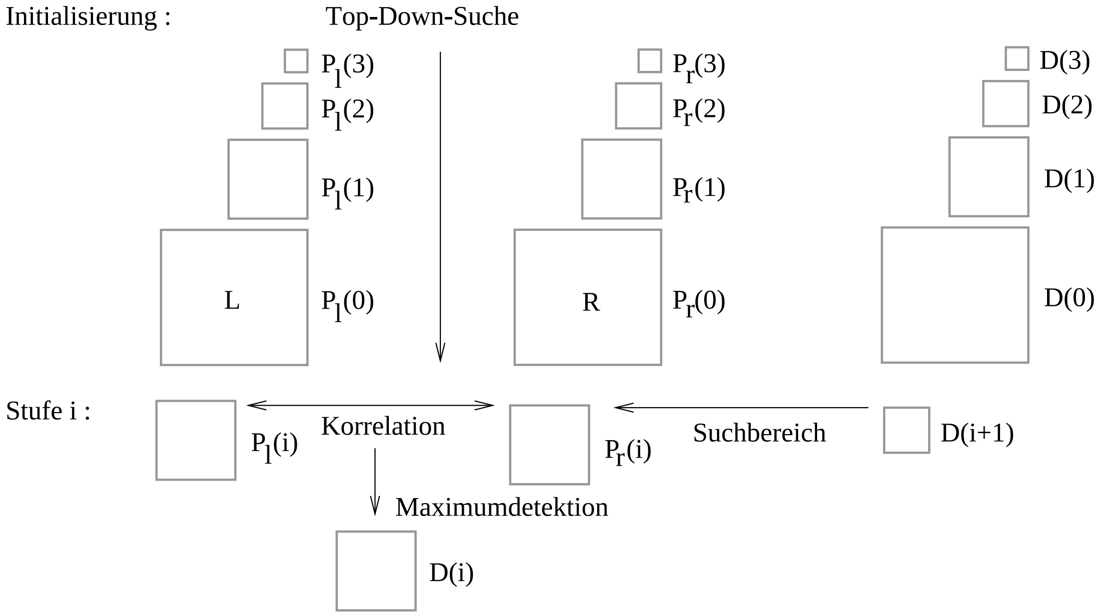


Abbildung 4.3: Das GREEDY-Stereoverfahren besteht aus einer Pyramidenverarbeitung, bei der in der i. Stufe jeweils die Position im Suchbereich mit optimalem Korrelationswert ermittelt wird. Mit den so gewonnenen Disparitätsinformationen wird in einer Top-Down-Strategie die Suche nach korrespondierender Bildstruktur der Stufe i-1 initialisiert.

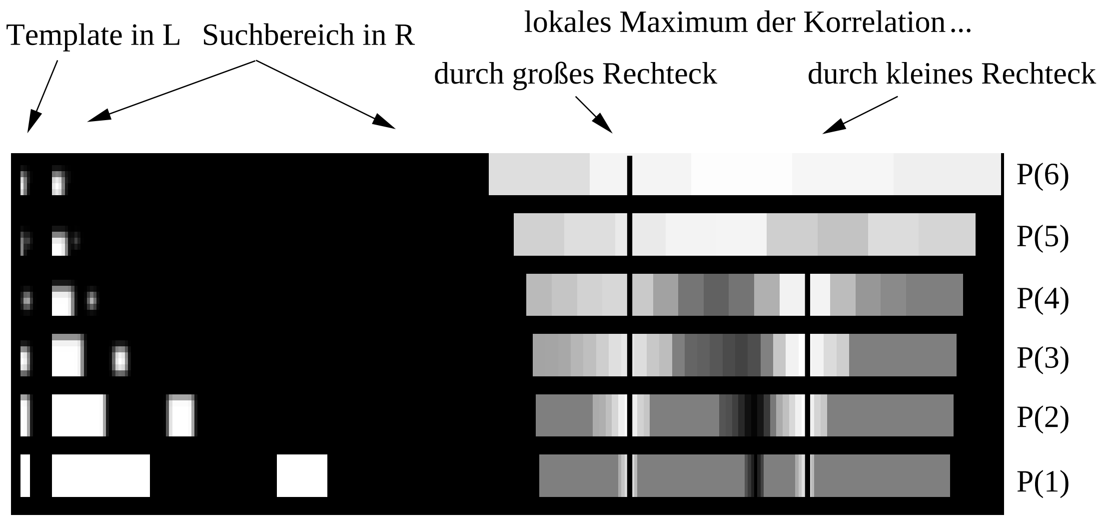


Abbildung 4.4: Für das vorliegende Template im Suchbereich versagt die Top-Down-Skalensuche der Greedy-Strategie in einer Pyramide von Korrelationsergebnissen. In der maximalen Pyramidenstufe P(6) wird der linke markierte Pfad im Lösungsbaum eingeschlagen, während der rechte markierte Pfad der korrespondierenden Zuordnung entspricht. Das kleine Rechteck im Template wird dem dominanten großen Rechteck im Suchbereich zugeordnet.

In einem Top-Down-Ansatz sucht die Greedy-Strategie ausgehend vom absoluten Korrelationsmaximum im Suchbereich der höchsten Pyramidenstufe durch die Skalen und verfolgt den gekennzeichneten linken Pfad zur Verfeinerung der Zuordnung. Dabei wird das kleine Rechteck im Template der Kante des großen Rechtecks zugeordnet, ein Fall der Dominanz von Bildstruktur. Die in der Suchposition der untersten Pyramidenstufe getroffene Zuordnung ist falsch. Der zur korrekten Zuordnung gehörige Pfad im Lösungsbaum ist rechts gekennzeichnet.
Außerdem kann der Fall aus Abschnitt 2.2.2 eintreten, daß feine Bildstrukturen der maximalen Pyramidenstufe durch Tiefpaßfilterung verschwunden sind. Dieser Effekt deutet sich in den Korrelationsantworten des Beispiels an, die in der maximalen Pyramidenstufe kein ausgeprägtes Maximum mehr aufweisen. Welchen Pfad die Greedy-Suchstrategie nun initial einschlägt, ist ungewiß. Eine vertrauenswürdige Zuordnung wäre in diesem Fall erst in einer tieferen Pyramidenstufe möglich.
Insgesamt zeigt dieses Beispiel, daß die Greedy-Suchstrategie durch die Top-Down-Verfolgung eines einzigen Pfades im Lösungsbaum an Effekten der Tiefpaßfilterung in der Gauß-Pyramide sowie der Dominanz von Bildstruktur (siehe Abschnitt 2.1.2) scheitern kann. Diese Fälle schildert auch [He97].

Der Greedy-Algorithmus wurde auf einer Workstation implementiert. Ausgehend von einem Stereobildpaar (L,R) der Dimension 128x128 werden Gauß-Pyramiden P_l, P_r der Tiefe 4 berechnet. Der Gauß-Filter besitzt die Dimension 7x7. Die Ergebnisse der Zuordnung in einer Pyramidenstufe werden in Disparitätskarten D_i vermerkt. Die Suche startet in $P_{l,r}(4)$ (Dimension 16x16) mit einer zu Null initialisierten Disparitätskarte D_4. Der Suchbereich beträgt jeweils $\pm$ 1 Pixel in unmittelbarer Nachbarschaft. Nach der Top-Down-Suche ergibt sich ein Suchbereich für Disparitäten von insgesamt ± 15 Pixel bezogen auf die unterste Pyramidenstufe. Die durch den Algorithmus gelieferten Disparitätskarten zu verschiedenen Testbildern sind in Kapitel 5 im Vergleich zu anderen Stereoverfahren dargestellt. Die erzeugten Disparitätskarten stammen von der Implementierung auf einer Ultra-Sparc-Workstation und wurden in einer Taktdauer von 8 Sekunden ermittelt. In Abschnitt 6.1 wird auf eine Implementierung der Standardstrategie auf dem MV200-System eingegangen.

4.3.2 Multi-Skalen-Analyse als Erweiterung der Standardmethode

In diesem erweiterten Stereoverfahren werden wie in Abschnitt 2.2.2 in einem Suchbereich Korrelationsergebnisse in einer Gauß-Pyramide ermittelt. Die Korrelationsergebnisse aus verschiedenen Pyramidenstufen werden verknüpft. Der Lösungsbaum aller möglichen Zuordnungen für ein Pixel wird vollständig, das heißt für alle möglichen Disparitäten betrachtet. Der Lösungsraum wird eingeschränkt, indem nur Positionen im Suchbereich betrachtet werden, die ein lokales

Korrelationsmaximum aufweisen. Die Suchstrategie verwendet so implizit ein Vertrauensmaß.
Der Hauptunterschied und Vorteil gegenüber obigem korrelationsbasierten GREEDY-Standardansatz liegt in dem vorliegenden Stereoverfahren in der Verwendung einer aufwendigeren, vollständigen Skalensuchstrategie in den Korrelationsergebnissen, die auf einer Modellierung des optimalen Einzugsbereichs der Korrelation beruht.
Da das Verfahren eine Eigenentwicklung darstellt, soll die Vorgehensweise durch eine Beschreibung der Entwicklung motiviert werden.

Ausgegangen wird von der Korrelation als Vergleichsmaß. Um dem Problem eines optimalen Einzugsbereichs gerecht zu werden, wird dieser variiert. Zu diesem Zweck werden Gauß-Pyramiden P_l, P_r gebildet. Es würde Vorteile besitzen, den Einzugsbereich bei voller Auflösung zu verändern, diese Möglichkeit kommt aber durch eine Beschränkung des MV200-Systems auf Templates der Dimension 8x8 Pixel nicht in Frage.
In Abschnitt 4.3.1 wurde auf Probleme der Standardstrategie hingewiesen, die durch die ausschließliche Betrachtung eines Pfades im Baum der möglichen Zuordnungen entstehen. Um diese Probleme zu umgehen, wird eine allgemeinere Modellierung der Entsprechung von Pixelkorrespondenzen in Korrelationsergebnissen angestrebt.
Zu diesem Zweck ist es hilfreich, nicht von der beschränkten Annahme auszugehen, daß sich eine Korrespondenz von Pixeln in einer Top-Down-Analyse eindeutig detektieren läßt. Es sollen also alle möglichen Korrelationsergebnisse der einzelnen Pyramidenstufen im Zusammenhang betrachtet werden, um ein Maximum an Informationen in die Analyse einzubeziehen. Da nicht von einer Verfeinerung einer Zuordnung von einer Pyramidenstufe zur nächsten ausgegangen wird, bleibt der Suchbereich über die Pyramidenstufen hinweg konstant.
Dies bedeutet allerdings einen hohen Rechenaufwand, da statt eines Pfades im Baum der möglichen Zuordnungen der ganze Baum betrachtet wird, jedoch läßt sich der Aufwand dieser vollständigen Analyse durch die hohe Pipeline- und Parallelkapazität der MV200-Systems rechtfertigen, wie sich zeigen wird.

Nach einer allgemeinen Spezifizierung der vollständigen Skalensuchstrategie soll nun auf die genaue Realisierung eingegangen werden. Hat man zu einem gegebenen Stereobildpaar (L,R) zwei Gauß-Pyramiden P_l, P_r berechnet und in einem Suchbereich in P_r bei fester Templategröße für alle Pyramidenstufen, Suchpositionen und Pixel in P_l die Korrelationsergebnisse in einer dreidimensionalen Datenstruktur KP abgelegt, so stellt sich die Frage nach einer angemessenen Analyse der ermittelten Korrelationspyramide KP.
Um das gesuchte Modell zur Identifikation von Pixelkorrespondenzen in KP zu erhalten, sollen einige Korrelationspyramiden KP für reale und synthetische Testbilder betrachtet werden. Die Testbilder beinhalten mit periodischen oder dominanten Strukturen eher problematische Bedingungen für eine Korrespondenzsu-

che.
In Abbildung 4.5 ist jeweils für ein einzelnes Template $T_L(i,j)$ ein zugehöriger Suchbereich T_R dargestellt, sowie resultierende Korrelationspyramiden KP ($MNKK$, 6 Stufen). Dabei wurden die Korrelationsantworten der horizontalen Suche zur besseren Vergleichbarkeit auf eine einheitliche Größe normiert, so daß entsprechende Suchpositionen in den Pyramidenstufen von KP übereinanderliegen.
Die isolierte Analyse von KP für ein einzelnes Template beziehungsweise Pixel, das zugeordnet werden soll, ist unumgänglich, weil auf dem MV200-System nur Suchstrategien mit einfachem, regelmäßigem Programmfluß implementierbar sind. Optimierungsstrategien, die Korrelationsergebnisse und Disparitäten im Zusammenhang betrachten, sind nicht realisierbar, da ihnen ein komplexer, iterativer Prozeß zugrunde liegt, der dem Pipeline-Prinzip widerspricht.
In Abbildung 4.5 erkennt man, daß zum Auffinden von einem zu $T_L(i,j)$ korrespondierenden Template $T_R(k,l)$ die Betrachtung lokaler Korrelationsmaxima in KP geeignet ist. Eine notwendige Bedingung für eine Korrespondenz stellt ein lokales Maximum an der Stelle (k,l) im Suchbereich von KP dar. Eine Ausnahme bildet Template f), in dem keine Struktur enthalten ist.
In einigen Pyramidenstufen liegt für $T_L(i,j)$ in KP ein lokales, in anderen ein absolutes Maximum im Suchbereich an der Stelle (k,l) vor.
Weiterhin ist ersichtlich, daß bei hoher Auflösung beziehungsweise niedriger Pyramidenstufe die Lage von lokalen Maxima immer mehr durch Rauscheinflüsse bestimmt wird, bedingt durch das sich verschlechternde Signal-Rausch-Verhältnis in den Templates. In hohen Pyramidenstufen kann sich durch Einbezug dominanter Strukturen in die Templates die Ausprägung eines Maximums bei (k,l) beliebig verschlechtern. Die Lage des absoluten Korrelationsmaximums in einer Pyramidenstufe verschiebt sich wie in Fall b) an eine andere Suchposition.
Zwischen diesen beiden Extremen in den Skalen zeigt sich an der Stelle von Korrespondenz (k,l) ein örtlich konstantes lokales Maximum, das sich als zusammenhängender "Streifen" durch benachbarte Pyramidenstufen zieht. Anzumerken ist, daß die örtlich konstante Lage im Fall einer Gauß-Pyramide durch die in Abschnitt 2.2.2 vorgestellte systematische Verschiebung von Bildstruktur geringfügig beeinflußt werden kann.
Im Fall a) existieren durch die periodische Bildstruktur im Suchbereich mehrere Streifen lokaler Maxima, von denen sich der korrespondierende durch über die Pyramidenstufen insgesamt höhere Korrelationsantworten auszeichnet.

Aus diesen Beobachtungen in den Korrelationspyramiden von Testbildpaaren läßt sich ein zentraler Effekt bei der Fortpflanzung lokaler Maxima durch Pyramidenstufen hindurch (Streifen) erkennen, der auf der einhergehenden Vergrößerung des Einzugsbereichs der Korrelation beruht. Betrachtet man die Korrelation von Templates an entsprechender Suchposition, aber mit unterschiedlichem Einzugsbereich, so bedingt das Ergebnis der niedrigeren Stufe rein numerisch das der

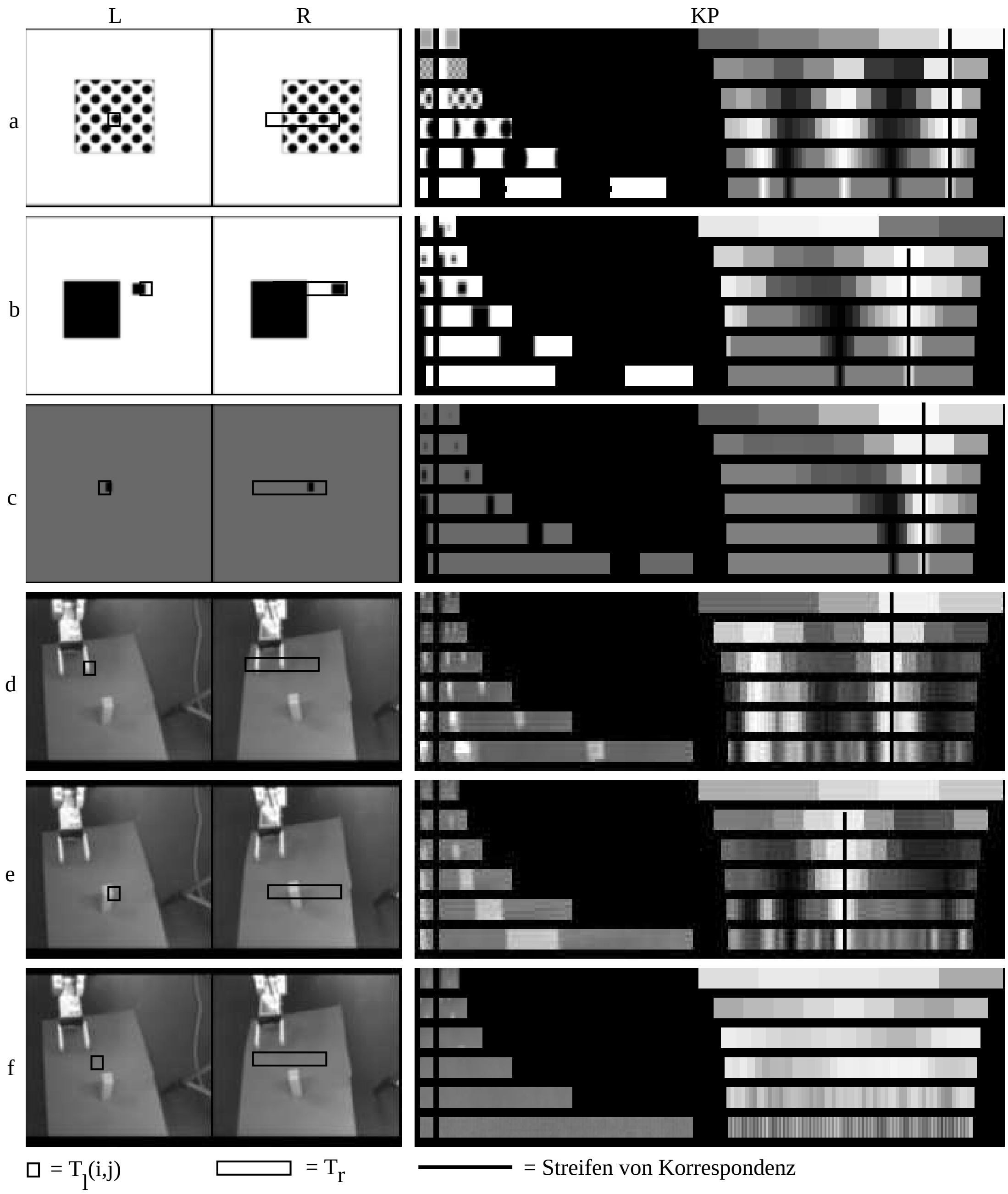


Abbildung 4.5: Dargestellt ist jeweils ein Stereobildpaar mit einem Template $T_L(i,j)$ und einem Suchbereich T_R. Daneben erkennt man die Lage von lokalen Maxima der Korrelation von $T_L(i,j)$ mit T_R in einer ermittelten Korrelationspyramide KP. Über die Pyramidenstufen zusammenhängende Korrelationsmaxima der Korrespondenz des Templates im Suchbereich sind markiert.

höheren Stufe, da in das Ergebnis der höheren Stufe die Signalbestandteile der niedrigeren als Teilmenge miteinfließen. Man kann also die Aussage treffen, daß sich ein lokales Maximum der Korrelation durch alle höheren Pyramidenstufen hindurchzieht.
Dieser besagte Effekt wird in KP in den Skalen nach unten durch Rauscheinflüsse und nach oben durch Dominanz von Bildstruktur durch einen zu großen Einzugsbereich des Vergleichsmaßes beschränkt. Weiterhin beeinflussen als Randeffekte die Filterung der Signale durch die Gauß-Pyramide sowie Quantisierungseffekte die Fortpflanzung lokaler Maxima durch die Pyramidenstufen.

Zusammenfassend soll das gesuchte Modell zur Detektion von Pixelkorrespondenzen in einer Pyramide von Korrelationsantworten KP in Form einer Skalensuchstrategie vorgestellt werden:
Unter einem Streifen von KP soll eine Folge von lokalen Korrelationsmaxima im Suchbereich verstanden werden, die sich durch benachbarte Pyramidenstufen zieht und eine konstante Lage aufweist. Zur Detektion der Korrespondenz eines Templates im Suchbereich werden in KP für jede Suchposition Streifen detektiert. Streifen können in den Pyramidenstufen in beliebiger Lage auftreten, weil ihre Ausdehnung nach unten durch hochfrequente Rauscheinflüsse und nach oben durch den Einfluß dominanter niederfrequenter Bildstrukturen sowie durch systematische Einflüsse der Gauß-Pyramide begrenzt wird. Die Ausdehnung des korrespondierenden Streifens über die Pyramidenstufen hinweg kann mit dem optimalen Einzugsbereich des Templates identifiziert werden.
Zur Trennung des korrespondierenden Streifens von anderen Streifen im Suchbereich wird jeder Streifen mit der Summe der enthaltenen Korrelationsmaxima bewertet.
Für die Suchposition k ergibt sich die Bewertung $W(k)$ eines Streifens, der p Korrelationsmaxima m_{ki} enthält, zu

$$W(k) = \sum_{i=1}^{p} m_{ki}$$

Die Position im Suchbereich k_{max} mit maximaler Streifenbewertung $W(k_{max})$ wird ausgewählt und dem Aufpunkt des Templates als Disparität zugeordnet.
Die gefundene maximale Bewertung stellt ein Vertrauensmaß für die gefundene Zuordnung dar, weil sie auf der Addition von Korrelationsmaxima basiert, die jedes für sich an Bildstruktur der korrelierten Templates gebunden sind.
In Abbildung 4.6 ist die Modellierung der Lage von Streifen lokaler Maxima zur Detektion von Pixelkorrespondenzen in KP dargestellt.
Bei dem Kriterium des Zusammenhangs eines Streifens durch die Pyramidenstufen sind Diskretisierungseffekte zu berücksichtigen. Wird beispielsweise für die Suchposition 11 (Pixel) in KP nach einem Streifen durch die Pyramidenstufen gesucht, so liegt die exakte Suchposition in den nächsten drei Pyramidenstufen bei 5.5, 2.75 und 1.375 (siehe auch Abbildung 4.5). Da die Suchpositionen in den

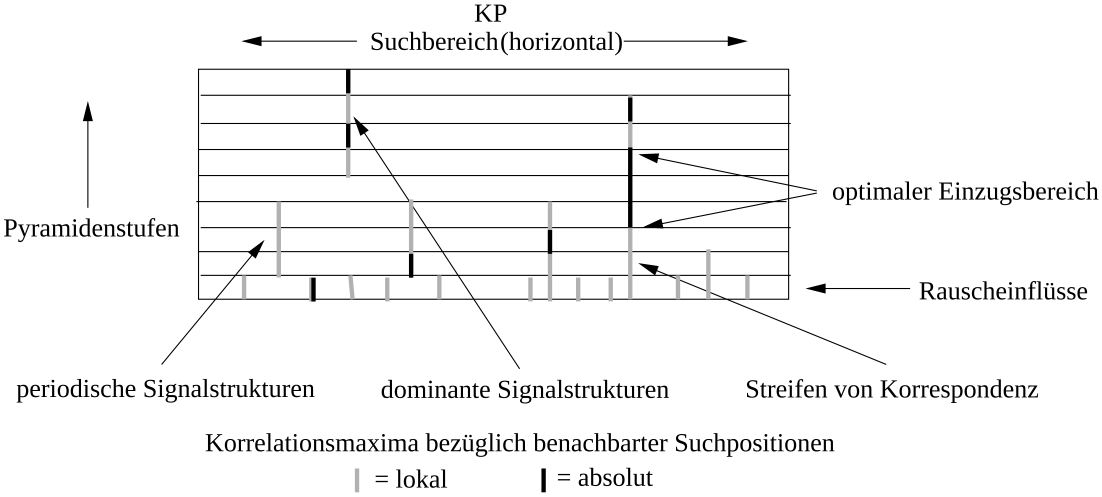


Abbildung 4.6: Das Modell zum Auffinden von Pixelkorrespondenzen in einer Korrelationspyramide KP geht von Streifen lokaler Korrelationsmaxima im Suchbereich aus, die über benachbarte Pyramidenstufen hinweg zusammenhängen und eine konstante Lage aufweisen. In einer Pyramidenstufe wird von der Existenz mehrerer Streifen für ein Pixel ausgegangen, deren Korrelationsmaxima durch Rauscheinflüsse, periodische Signalstrukturen, dominante Bildstrukturen oder die gesuchte Pixelkorrespondenz entstanden sein können. Der Streifen von Korrespondenz des betreffenden Pixels zeigt sich im Suchbereich durch die maximale Summe der enthaltenen Korrelationswerte.

jeweiligen Pyramidenstufen ganzzahlige Werte annehmen, wird der Wert der jeweiligen Subpixelposition in KP durch Werte benachbarter Suchpositionen linear interpoliert.
Durch die ausschließliche Betrachtung lokaler Maxima in KP wird der Raum möglicher Zuordnungen auf relevante Kandidaten eingeschränkt.
Das Stereoverfahren liefert als Ausgabe die durch die Suchstrategie detektierten Pixelkorrespondenzen in Form einer Disparitätskarte D und mit den bei der Korrespondenzsuche gefundenen maximalen Streifenbewertungen eine Karte CF für die Vertrauenswürdigkeit der Zuordnung.

Mit dieser Suchstrategie lassen sich die in Abbildung 4.5 dargestellten Korrespondenzen von Templates ermitteln, während obige Standardstrategie in den Fällen b) und e) eine falsche Korrespondenz liefert.
Die Suchstrategie beinhaltet eine Verknüpfung von Korrelationsbeträgen aus verschiedenen Einzugsbereichen. Aus diesem Grund ist eine Normierung der verwendeten Korrelation notwendig.

Es folgt eine Zusammenfassung der genannten Elemente des entwickelten Multi-Skalen-Stereoalgorithmus *MSA*.

1. Bilde zu einem Stereobildpaar (L, R) Gauß-Pyramiden $(P_l,\ P_r)$.

2. Bilde in einheitlichem, horizontalen Suchbereich in P_r aus der Korrelation (NKK) von $(P_l,\ P_r)$ mit festem Einzugsbereich eine Pyramide von Korrelationsergebnissen KP.

3. Detektiere und extrahiere in jeder Pyramidenstufe von KP für jedes Pixel lokale Maxima im Suchbereich.

4. Berechne für jedes Pixel und für jede Suchposition k in KP die Bewertung $W(k) = \sum\limits_{i=1}^{p} m_{ki}$ eines zugehörigen Streifens, der p Korrelationsmaxima m_{ki} enthält.

5. Weise jedem Pixel die Disparität k_{max} mit maximaler Streifenbewertung $W(k_{max})$ im Suchbereich zu. Vermerke k_{max} in einer Disparitätskarte D und $W(k_{max})$ in einer Vertrauenskarte CF.

In Abbildung 4.7 ist MSA graphisch dargestellt.
Die zwei Gauß-Pyramiden besitzen in Workstation- und MV200-Implementierung vier Stufen und werden mit einer 7x7-Gaußmaske aus Eingabebildern der Dimension 128x128 Pixel erzeugt. Die Pyramide KP wird durch Korrelation der Gauß-Pyramiden mit einer Templategröße von 7x7 in einem Suchbereich von ± 16 Pixel bezogen auf die volle Auflösungsstufe ermittelt.
Auf einer Ultra-Sparc-Workstation liefert MSA Disparitätskarten in einem Takt

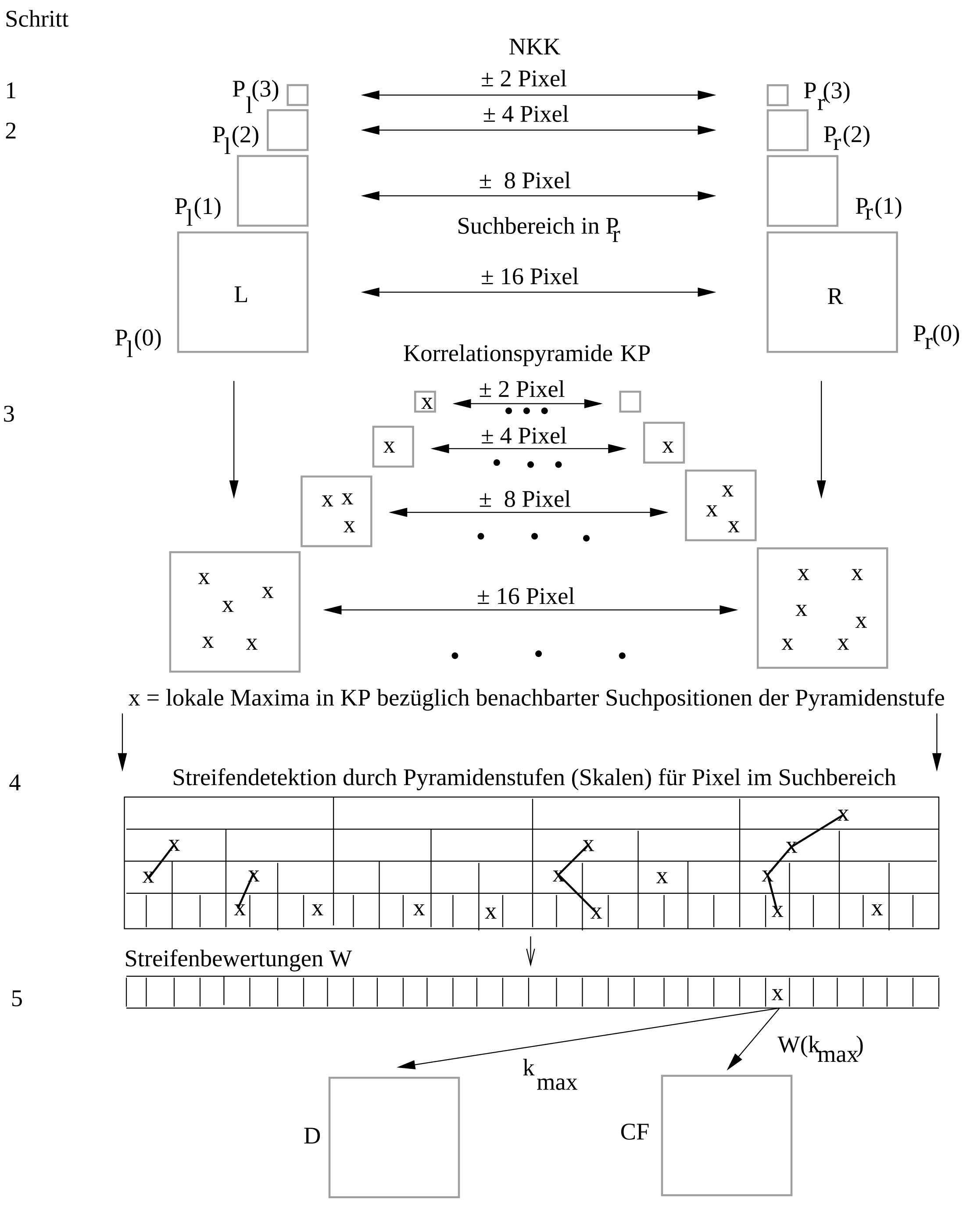


Abbildung 4.7: Das MSA-Stereoverfahren ermittelt Gauß-Pyramiden P_l und P_r der Eingabebilder, zwischen denen eine lokale Korrelation für alle Pixel, Pyramidenstufen und Suchpositionen ermittelt wird (KP). Es folgt eine Detektion von lokalen Maxima in KP bezüglich des Suchbereichs. In einer Skalensuche werden Streifen von lokalen Korrelationsmaxima für jede Suchposition extrahiert und bewertet. Aus den Bewertungen ergeben sich schließlich eine Disparitätskarte D und ein zugehöriges Vertrauensmaß CF.

von 14 Sekunden, in der MV200-Implementierung in einem Takt von 0.23 Sekunden.

Zu beachten ist, daß der MSA-Stereoalgorithmus keine Eingabeparameter benötigt (beispielsweise für Schwellwerte). Der Algorithmus läßt sich so ohne Wissen über seine Funktionsweise anwenden, was eine wichtige Eigenschaft für die Praxis darstellt.

Die durch den MSA-Algorithmus gelieferten Disparitätskarten werden in Kapitel 5 diskutiert.

Kapitel 5

Vergleich der Stereoverfahren

Auf der Grundlage eines allgemeinen und qualitativen Vergleichs wird unter den vorgestellten und implementierten Stereoverfahren eine Auswahl für die Systemkomponente getroffen. Untersucht werden als Standardansätze der Literatur das GREEDY- und YESA-Stereoverfahren sowie als eigene Weiterentwicklungen der Standardansätze das MSA- und CEPSIM-Stereoverfahren, indem die jeweilig erzeugten Disparitätskarten zu Testbildern verglichen werden. Die Ergebnisse werden aufgrund von strukturellen Eigenschaften der verantwortlichen Stereoverfahren sowie Grundzügen des Disparitätsproblems interpretiert.

5.1 Allgemeiner Vergleich

Der allgemeine Vergleich der vier Stereoverfahren soll generelle Eigenschaften und Schwächen zu Tage fördern, indem das Verhältnis der durch den jeweiligen Algorithmus falsch und richtig zugeordneten Pixel betrachtet wird.
Durch die Stereoverfahren werden grundsätzlich nur horizontale Disparitäten gemessen. Vergleichskriterien können sich allgemein in Ermangelung der Kenntnis der exakten Disparitätsinformationen der Testbildpaare nur an der Frage orientieren, ob die gemessenen Disparitäten in Relation zueinander Verhältnisse des jeweiligen Stereobildpaars widerspiegeln.
Beispielsweise ist zu beachten, wie kontinuierliche Disparitätsverläufe im Stereobildpaar detektiert werden, weil sie typisch für reale Stereobildpaare sind. Auch die Wiedergabe konstanter Disparitätsverläufe in den Karten ist auf allgemeiner Basis vergleichbar. In einigen Fällen wird explizit darauf hingewiesen, welcher Grauwert der jeweiligen Disparitätskarte der tatsächlichen Disparität entspricht. Aufgrund der Strukturgebundenheit von Disparitätsinformationen sind in den Karten diejenigen Bereiche ausmaskiert, an denen durch das Vorhandensein von Struktur im zugehörigen Testbildpaar vertrauenswürdige Zuordnungen möglich sind. Die Ausmaskierung erleichtert den Vergleich der Disparitätskarten, obwohl die Stereoverfahren teilweise auch in strukturlosen Bildbereichen sinnvolle Dispa-

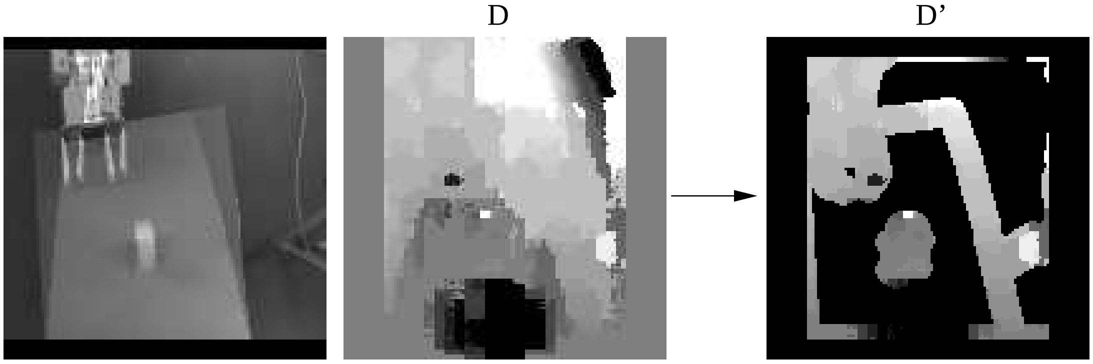


Abbildung 5.1: Das MSA-Stereoverfahren liefert eine Disparitätskarte D, die teilweise auch in strukturlosen Bildbereichen sinnvolle Disparitätsinformationen aufweist. Die Ausmaskierung vertrauenswürdiger Zuordnungen durch einen Bezug auf Bildstruktur im Stereobildpaar läßt die Disparitätskarte D’ entstehen.

ritätsinformationen liefern, wie in Abbildung 5.1 am Beispiel von MSA deutlich wird. Dort ist ein Stereobildpaar, eine durch das Verfahren erzeugte Disparitätskarte D und eine durch Ausmaskierung vertrauenswürdiger Zuordnungen entstehende Disparitätskarte D’ dargestellt. Durch die Skalensuche des Verfahrens werden Disparitäten in strukturlosen Bildbereichen durch benachbarte vertrauenswürdige Disparitätswerte interpoliert.

Als Testbilder werden synthetische und reale Stereobildpaare generiert beziehungsweise ermittelt. Jedes Bildpaar repräsentiert ein typisches Problem, das sich für Stereoverfahren unter realen Bedingungen ergeben kann.

- **Bildpaar A** enthält periodische Signalbestandteile von konstanter Disparität, die die Gefahr von Alias-Fehlern bergen.

- **Bildpaar B** enthält eine große, dominante Bildstruktur sowie in unmittelbarer Nachbarschaft eine kleine Bildstruktur. Bei einem zu großen Einzugsbereich des Vergleichsmaßes wird die kleine Bildstruktur photometrisch dominiert und erhält vom jeweiligen Stereoverfahren eine falsche Disparität zugeordnet (siehe Abschnitt 2.1.2).

- **Bildpaar C** enthält eine einzelne, schwach ausgeprägte Struktur. In einer Pyramidenverarbeitung kann diese Struktur in hohen Pyramidenstufen durch Tiefpaßfilterung und Quantisierungseffekte nicht mehr vorhanden sein. Für eine Top-Down-Suchstrategie wird die Auffindung von Korrespondenz problematisch.

- **Bildpaar D** stellt die reale Szene eines Robotergreifers mit einem Greifobjekt dar. Das Bildpaar enthält kontinuierliche Disparitätsverläufe. Das

Problem in diesem Bildpaar besteht in der Verzerrung von Bildstruktur. Die zwei Greiferzangen bergen als periodische Strukturen zusätzlich die Gefahr von Alias-Fehlern.

- **Bildpaar E** beinhaltet einen synthetisch erzeugten "Balken", der verzerrt wurde. Ähnlich wie in Bildpaar D widerspricht die Komponente einer starken Verzerrung von Bildstruktur der Annahme von Disparität als einer reinen Verschiebung, von der viele Stereoverfahren bei der Modellierung des Disparitätsproblems ausgehen.

In den Abbildungen 5.2 und 5.3 werden jeweils die Disparitätskarten zweier Stereoverfahren gegenübergestellt, wobei zur besseren Orientierung die zugrunde liegenden Stereobildpaare den Karten in einer Überblendung vorangestellt sind. Die Disparitätskarten beziehen sich jeweils auf das linke Stereobild.
Um möglichst identische Rahmenbedingungen für einen Vergleich zu schaffen, werden einheitlich Templategrößen von 7x7 Pixel und ein Suchbereich von ± 16 Pixel verwendet. Die Dimension der Stereobilder beträgt 128x128 Pixel. Diese Parameter werden mit Rücksicht auf die Beschränkungen des MV200-Systems gewählt. Eine Ausnahme bildet der YESA-Algorithmus, dessen Cepstrum-Filter durch Koppelung von Such- und Einzugsbereich einen größeren Einzugsbereich von 32 Pixeln benötigt.
Neben diesen Voraussetzungen benötigen die betrachteten Stereoverfahren keine weiteren Parameter. Dadurch ist der Vergleich gut konditioniert, weil Parameter wie Schwellwerte oder Gewichtungsfaktoren fehlen, deren geringfügige Veränderung Testergebnisse maßgeblich beeinflussen kann.
Die Disparitätskarten sind zur Visualisierung auf einen maximalen Grauwert normiert, so daß numerisch ähnliche Disparitätskarten durch Ausreißer im Grauwert dennoch eine unterschiedliche Gestalt besitzen können.

- **CEPSIM** : In Testbild A (Abbildung 5.2) ermöglicht die Skalensuchstrategie durch einen globalen Vergleich von Bildstruktur eine Zuordnung ohne Alias-Fehler.
  Testbild B und C führen zu korrekten Disparitätskarten. Testbild D stellt den interessantesten Fall dar. Hier liefert der Algorithmus wie erwünscht kontinuierliche Disparitätsverläufe im Bereich des Greifers und auf der Unterlage. Dies ist besonders zu beachten im Fall des Klotzes, der nur geringe Disparitätsdifferenzen aufweist und verzerrt ist.
  Die korrekt gemessenen Disparitätsverläufe sind allerdings durch vereinzelt falsche Zuordnungen überlagert, die auf die heuristische Komponente der Skalensuchstrategie des CEPSIM-Verfahrens zurückzuführen sind. Bei der Zuordnung der Greiferzangen jedoch verhindert die Skalensuchstrategie einen Alias-Fehler.
  In Testbild E zeigt sich, daß das CEPSIM-Verfahren im Gegensatz zum

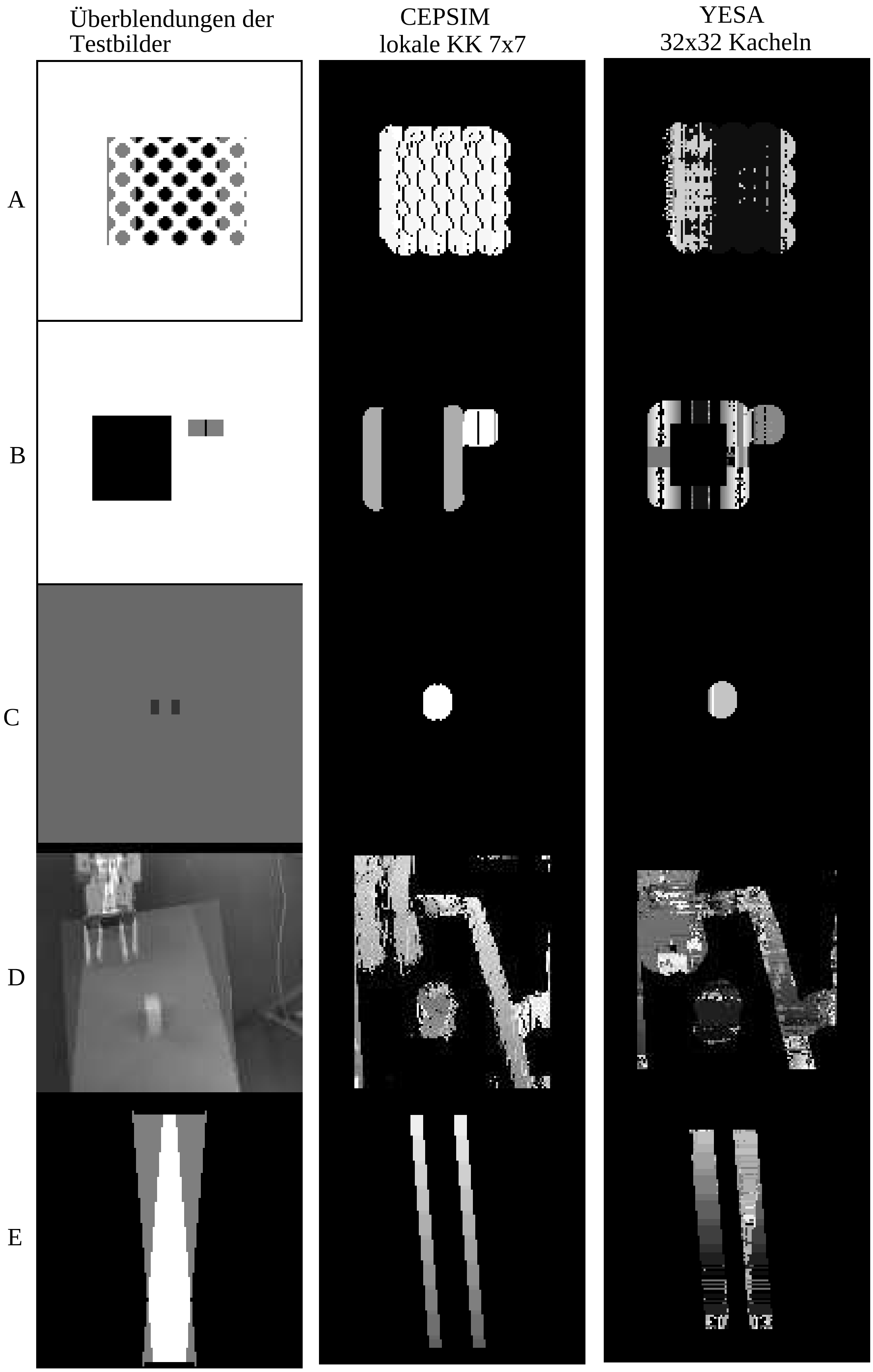


Abbildung 5.2: Die vom CEPSIM- und YESA-Stereoverfahren erzeugten Disparitätskarten für die Testbildpaare A-E.

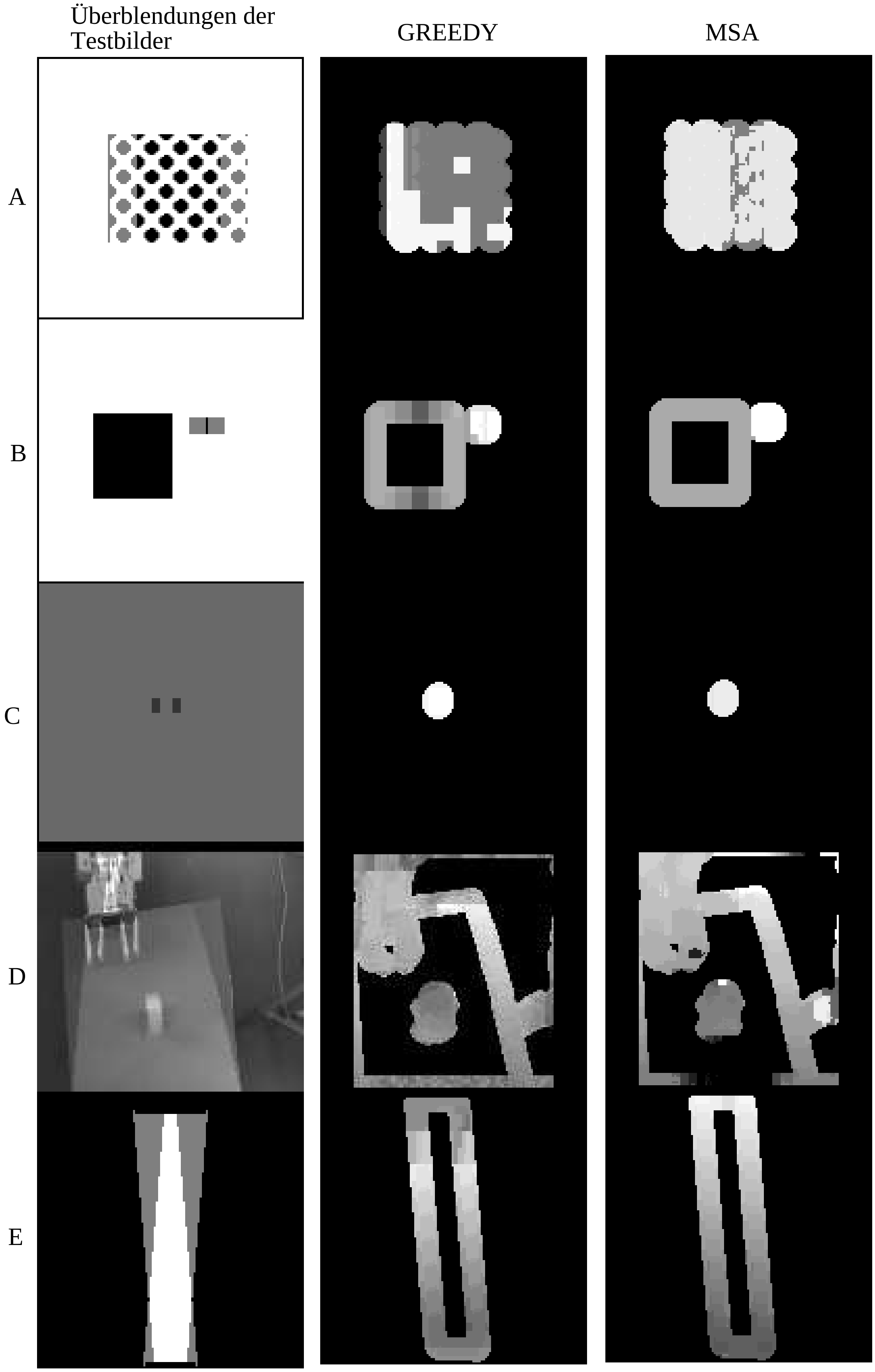


Abbildung 5.3: Die vom GREEDY- und MSA-Stereoverfahren erzeugten Disparitätskarten für die Testbildpaare A-E.

Cepstrum-Filter kontinuierliche Disparitätsverläufe im globalen Einzugsbereich des Vergleichsmaßes detektieren kann.

- **YESA** : Testbild A werden links und rechts in einem vertikalen Streifen die korrekten Disparitätsinformationen zugeordnet. Im Bereich dazwischen werden falsche Zuordnungen getroffen.
  Diese Alias-Fehler sind nicht anders zu erwarten, weil der Algorithmus einen festen Einzugsbereich verwendet. Am Rand der periodischen Struktur ergeben sich dann wieder brauchbare Ergebnisse, weil der Rand als Struktur eindeutig zuzuordnen ist.
  Unabhängig davon sind die korrekten Disparitätswerte von Störeinflüssen überlagert.
  Zu Testbild B liefert das Stereoverfahren im Bereich des großen Rechtecks nur in unmittelbarer Umgebung zu Kanten die korrekte Disparität (schwarz). In der Mitte des großen Rechtecks gibt es eine horizontale Zone (mittelgrau), die durch ein Aperturproblem entsteht. Im Einzugsbereich des Vergleichsmaßes existiert in diesem Fall nur eine durchgehende vertikale Kante, für die der Cepstrum-Filter keine Impulse liefern kann, was in der Interpretation als Autokorrelation plausibel wird. Das kleine Rechteck besitzt eine korrekte Disparität.
  Zu Testbild C werden korrekte Werte geliefert.
  In Testbild D ergeben sich viele falsche Zuordnungen, die nicht nur auf Alias-Fehler zurückzuführen sind. In den falschen Zuordnungen läßt sich keine Systematik erkennen, was auf eine Überlagerung mehrerer Ursachen hinweist.
  Grundsätzlich läßt sich jedoch die korrekte Wiedergabe kontinuierlicher Disparitätsverläufe feststellen.
  Eine ähnliche Aussage kann man zur Disparitätskarte von Testbild E treffen.

  Insgesamt sind die Ergebnisse des YESA-Algorithmus von Fehlern durch Alias-Probleme sowie einer Überlagerung unterschiedlicher Fehlerquellen gestört. Hier zeigt das Verfahren seine Schwächen in der einfachen Analysestrategie von Cepstrum-Impulsen, die durch systematische Probleme bei der Interpretation der Cepstrum-Antwort beeinträchtigt wird (siehe Abschnitt 3.3).

- **GREEDY** : Randprobleme entstehen zum einen bei der Bildung der Gauß-Pyramide (siehe Abschnitt 2.2.2), zum anderen bei der Berechnung der Korrelation. Sie wirken sich stärker als bei den anderen betrachteten Stereoverfahren aus, weil eine Top-Down-Strategie durch die Pyramidenstufen verfolgt wird und der Rand in der höchsten Pyramidenstufe am breitesten ist.
  Ein genereller Nachteil der Greedy-Strategie in Verbindung mit einer Top-

Down-Suche liegt in der Entstehung von Blocking-Effekten, die auf falsche Zuordnungen in hohen Pyramidenstufen im Prozeß der absteigenden Verfeinerung der Disparitätsergebnisse zurückzuführen sind (siehe Abschnitt 4.3.1).

In Testbild A trifft der GREEDY-Algorithmus in hohen Pyramidenstufen falsche Zuordnungen, wie an den Blocking-Effekten erkennbar ist, obwohl er gegen Alias-Fehler durch die Pyramidenverarbeitung unempfindlich ist.
Hier schränkt der ausgeprägte Rand das Stereobildpaar bereits auf einen Bereich ein, in dem nur noch periodische Strukturen vorkommen und Alias-Probleme unvermeidlich sind. Diese Vermutung bestätigt sich bei Anwendung des Stereoverfahrens auf das Bildpaar A mit einem erweiterten Rand. In diesem Fall werden korrekte Zuordnungen getroffen, so daß sich die falschen Zuordnungen für Testbild A als Randprobleme relativieren lassen.
Zu Testbild B und C werden korrekte Disparitätskarten generiert. Die Empfindlichkeit des Algorithmus gegenüber der Dominanz von Bildstrukturen wirkt sich in Testbild B nicht aus. Das kleine Rechteck scheint noch zu weit vom großen Rechteck entfernt zu sein, als daß eine photometrische Dominanz durch das große Rechteck eintreten würde.
Auch für Testbild D besitzt die Disparitätskarte am Ort des Greifers, am Würfel und auf der Unterlage erwünschte kontinuierliche Disparitätsverläufe. Allerdings wird das Ergebnis an der rechten oberen Ecke der Unterlage durch einen Blocking-Effekt verfälscht.
Mit einer ähnlichen Aussage läßt sich die zu Testbild E gelieferte Disparitätskarte charakterisieren.

- **MSA** : In Testbild A liefert MSA weitgehend korrekte Disparitätswerte. Falsche Zuordnungen werden vereinzelt im zentralen Bereich der Disparitätskarte getroffen, wo der Algorithmus bei der Zuordnung mehrere lokale Korrelationsmaxima zur Auswahl hat. Davon abgesehen verhindert aber die Skalensuche erwartungsgemäß Alias-Fehler bei der Zuordnung der vorliegenden periodischen Struktur.
Zu bemerken ist, daß der MSA-Algorithmus robust gegenüber Randproblemen ist, obwohl eine Pyramidenverarbeitung verwendet wird. Die Ursache liegt in der Suchstrategie, die im Gegensatz zu einer Top-Down-Suche nicht an die maximale Pyramidenstufe gebunden ist, in der Randprobleme am stärksten ausgeprägt sind. Weiterhin besitzt der MSA-Algorithmus eine Randbehandlung bei der Generierung von Pyramiden und Korrelationswerten, so daß Vergleichsergebnisse auch im Randbereich ermittelt werden können (siehe Abschnitt 6.1.1).
Die Zuordnungen für Testbild B und C sind vollkommen korrekt.
In Testbild D werden die zu erwartenden kontinuierlichen Disparitäts-

verläufe am Greifer und auf der Unterlage detektiert. Die rechte Greiferzange hat im unteren Teil erkenntlich am dunklen Bereich in der Disparitätskarte eine falsche Zuordnung erhalten.
Für Testbild E ergibt sich eine korrekte Disparitätskarte.

5.2 Qualitativer Vergleich

Nach einem allgemeinen Vergleich folgt nun ein qualitativer Vergleich der vier Stereoverfahren auf numerischer Basis. Dabei soll das Vergleichskriterium die Genauigkeit der durch die einzelnen Verfahren getroffenen Zuordnungen sein.
Zu diesem Zweck ist es notwendig, zu einem gegebenen Stereobildpaar die tatsächlichen Disparitäten zu kennen. Dieses Wissen ist bei den synthetischen und realen Stereobildpaaren aus Abschnitt 5.1 generell nicht vorhanden, da der Prozeß der Erzeugung des jeweiligen Stereobildpaares nur in einfachen Fällen analytisch bekannt ist.
Als Ausweg wurde eine Simulation verwendet, in der sich Stereobildpaare aus einer synthetischen Szene mitsamt zugehörigen Disparitätskarten erzeugen lassen. Es werden geeignete Stereobildpaare ausgewählt, zu denen die vier in dieser Arbeit vorgestellten Stereoverfahren Disparitätskarten liefern. Abschließend werden die gemessenen Disparitätskarten mit den erzeugten Disparitätskarten der Simulation statistisch verglichen.

5.2.1 Erzeugung von Testbildern in der Simulation

In der Simulation wird eine einfache Szene modelliert, in der sich ein Stereokamerakopf befindet. Die Szene besteht aus einer verwinkelten Wandstruktur von ebenen Segmenten. Auf die Wandsegmente läßt sich eine Grauwertstruktur projizierten, entweder Sinuswellen oder lineare Grauwertverläufe. So wird gewährleistet, daß ein durch das Kameramodell gewonnenes Stereobildpaar an jedem Punkt Struktur aufweist, was eine Grundvoraussetzung zur Generierung dichter Disparitätskarten ist.
Der Stereokamerakopf besitzt drei Freiheitsgrade der Bewegung (Schwenkung, linke und rechte Vergenz) und läßt sich frei in der Szene positionieren. Brennweite und Länge der Stereobasis sind veränderbar.
Die Simulation kann aus den Parametern des Kameramodells und der Szene ein Stereobildpaar sowie eine zugehörige, dichte Disparitätskarte ermitteln, die die tatsächlichen Disparitätsinformationen beinhaltet.
Auf der Basis von Stereobildpaaren und exakten Disparitätskarten läßt sich schließlich ein Vergleich zu Disparitätskarten ziehen, die durch die Stereoverfahren gemessen werden.

Um auf dieser Grundlage einen einheitlichen Vergleich zwischen den Stereoverfahren ziehen zu können, sollten die verschiedenen Verfahren Bildpaare vorfinden, zu

denen sie unter Berücksichtigung ihrer Beschränkungen, die sich aus der jeweiligen Modellierung des Disparitätsproblems ergeben, in gleicher Weise korrespondierende Zuordnungen treffen können, weil in diesem Abschnitt die Genauigkeit der Zuordnung analysiert werden soll und nicht das Problem der Zuordnung selbst.
Es wurde oben bereits erwähnt, daß die Wahl von Testbildern entscheidenden Einfluß auf Vergleichsergebnisse haben kann. So sind die in der Simulation erzeugten Stereobildpaare mit den vorkommenden Projektionen von Sinuswellen für den Vergleich von phasenbasierten Stereoverfahren vorgesehen, für die sie strukturell optimal geeignet sind.
Die Forderung nach einheitlichen Rahmenbedingungen für einen Vergleich schränkt die Variationsmöglichkeiten der Bildgenerierung durch die Simulation ein. Zum einen gibt es die Wahlmöglichkeit einer sinusförmigen und einer linearen Grauwertstruktur auf den Wänden der Simulationsszene. Eine lineare Struktur ist jedoch nach Abschnitt 4.2 ungeeignet für korrelationsbasierte Stereoverfahren. Zum anderen muß eine sinusförmige Wandstruktur eine derartige Wellenlänge besitzen, daß die Stereoverfahren im Suchbereich weniger als eine halbe Sinuswelle vorfinden, um Alias-Fehler auszuschließen. Da der Suchbereich aller betrachteten Stereoverfahren ± 16 Pixel und die Dimension eines Kamerabildes 128x128 Pixel beträgt, dürfen folglich in einem geeigneten Stereobildpaar maximal vier Wellenzüge enthalten sein.
Die Wandstruktur der Simulationsszene besteht aus ebenen Segmenten. So können Stereobilder erzeugt werden, die ein oder mehrere Wandsegmente als Projektion beinhalten. Der Fall von mehreren Wandsegmenten stellt von der Grauwertstruktur in der Projektion einen qualitativen Unterschied gegenüber nur einem Wandsegment dar, weil im Stereobildpaar an der Nahtstelle der Segmente im allgemeinen eine dominierende Kante entsteht, die für die Stereoverfahren eine zusätzliche Information darstellt.
Desweiteren bietet die Variation von Freiheitsgraden der Bewegung des Kamerakopfes in der Simulationsszene die Möglichkeit der entsprechenden Beeinflussung von Verzerrungseffekten in den erzeugten Stereobildern.

Aus diesen Überlegungen heraus werden vier Stereobildpaare generiert, die die oben genannten relevanten Klassen von in der Simulation erzeugbaren Stereobildpaaren abdecken. Die Bildpaare enthalten dementsprechend die Kombinationen der Fälle von einem oder mehreren Wandsegmenten im Bild sowie einer planaren oder schrägen Aufsicht auf die Wandsegmente, die eine Veränderung der in den Bildpaaren entstehenden Verzerrungseffekte bewirkt.
Die vier Stereobildpaare sowie eine Momentaufnahme der Simulationsszene zu ihrer Erzeugung (Aufsicht) sind in Abbildung 5.4 dargestellt. Zusätzlich sind die erzeugten tatsächlichen Disparitätskarten abgebildet, die sich jeweils auf das linke Stereokamerabild beziehen.

Bildpaar 1 ergibt sich aus der schrägen Aufsicht der Stereokameras auf eine sinusförmige Wandstruktur in der Simulationsszene. Die Entfernung der Stereo-

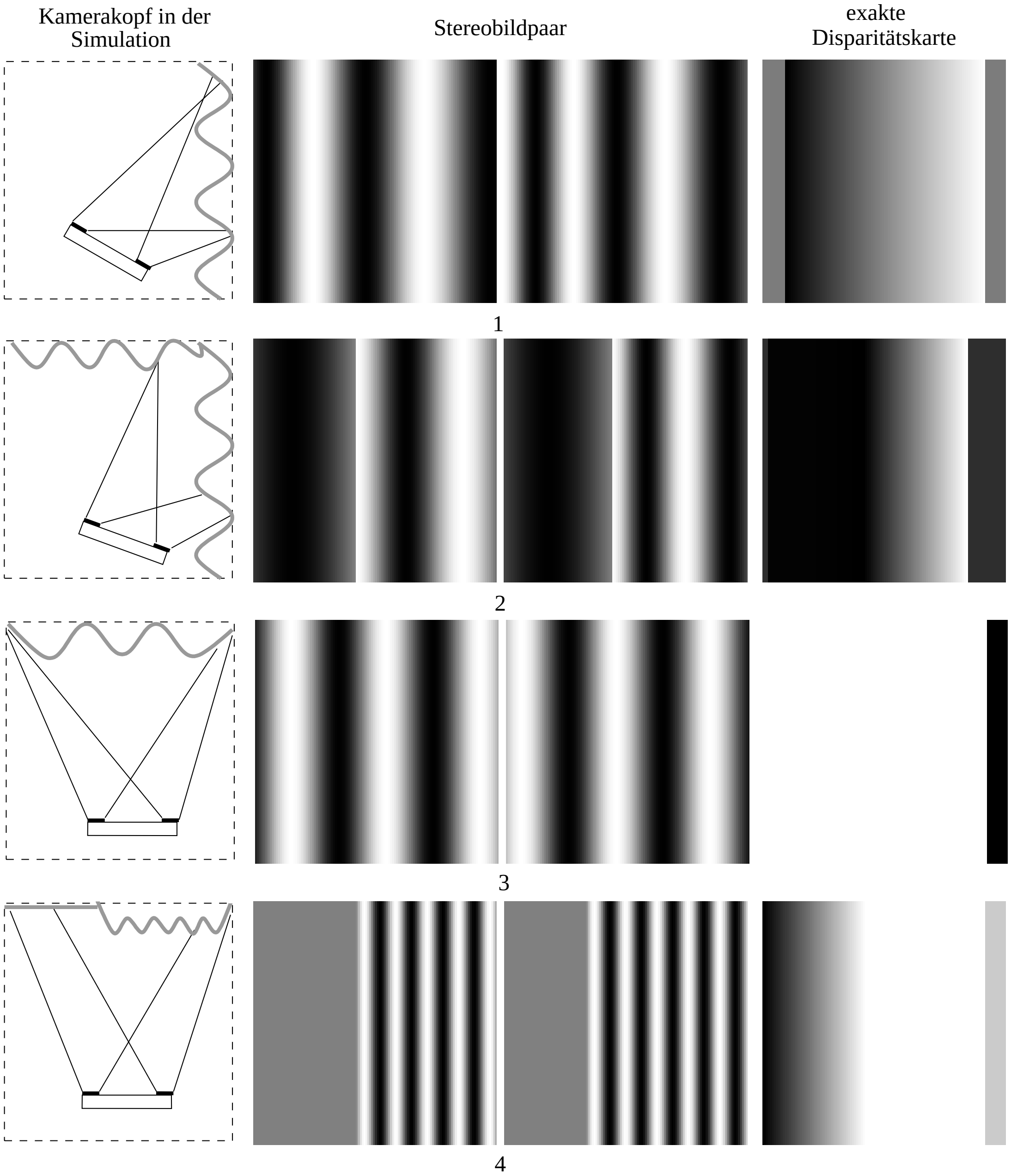


Abbildung 5.4: Die Simulation erzeugt zu entsprechenden Parametern der Szene und des Kameramodells ein Stereobildpaar und eine exakte Disparitätskarte.

kameras zur Wand ist gering, wobei die rechte Kamera näher zur Wand steht. Dadurch besitzen die Sinuswellen im linken und rechten Kamerabild eine erhebliche horizontale Verzerrung gegeneinander.
Bildpaar 2 wurde in derselben Konfiguration der Szene ermittelt wie Bildpaar 1, jedoch mit dem Unterschied, daß der Kamerakopf so weit geschwenkt wurde, daß die Nahtstelle zu einem benachbarten Wandsegment im Bildpaar enthalten ist. Das ursprüngliche Wandsegment besitzt im Bildpaar eine starke Verzerrung, während das benachbarte Wandsegment nahezu parallel zu den Bildebenen der Kameras liegt und relativ weit entfernt ist, so daß dort keine Verzerrungen von Bildstruktur auftreten und die Disparität konstant ist.
Bildpaar 3 beinhaltet einen Fall konstanter Disparität im Bildpaar, welches sich aus einer frontoparallelen Aufsicht auf ein Wandsegment ergibt.
Bildpaar 4 ergibt sich ebenfalls aus frontoparalleler Aufsicht auf eine Wand. Ein benachbartes Wandsegment, das keine Grauwertstruktur aufweist, schließt sich an. Die Wand selbst besitzt eine sinusförmige Grauwertstruktur, deren hohe Frequenz zu Alias-Problemen führen würde, wenn die Nahtstelle beider Wandsegmente nicht einen Anhaltspunkt für eine eindeutige Zuordnung bieten würde.
So stellen die vier Bildpaare unterschiedlich schwierige Testkandidaten für die betrachteten Stereoverfahren dar. Bildpaare 1 und 2 beinhalten das Problem von Verzerrungen. Bildpaar 4 birgt die Gefahr von Alias-Fehlern bei der Zuordnung, während Bildpaar 3 den einfachsten Fall konstanter Disparität darstellt, zu dem alle betrachteten Stereoverfahren eine korrekte Disparitätskarte liefern sollten.

5.2.2 Auswertung der Versuchsergebnisse

Die vorgestellten Stereoverfahren YESA, CEPSIM, GREEDY und MSA werden auf die obigen vier Stereobildpaare der Simulation angewendet. Es ergeben sich die Disparitätskarten der Abbildung 5.5.
Diese gemessenen Disparitätskarten werden mit den tatsächlichen Disparitätskarten statistisch verglichen. Für ein Bildpaar wird jeweils der mittlere Fehler der Disparitätsmessungen sowie die Varianz der Fehler bestimmt. Die Fehler werden durch Bildung der absoluten Differenz von gemessenen und tatsächlichen Disparitäten ermittelt.
Der statistische Vergleich wird unter dem Bildverarbeitungssystem KHOROS durchgeführt. Um der in Abschnitt 5.2.1 angesprochenen Forderung nach Vergleichbarkeit von Testergebnissen gerecht zu werden, wird aus den erzeugten Disparitätskarten ein vertrauenswürdiger Bereich ausmaskiert, auf dem die statistische Auswertung erfolgt. Die Maskierung berücksichtigt die Randbehandlung der vier Stereoverfahren sowie einen Bezug auf Struktur im Stereobildpaar (Schwellwertbildung).

So ergibt sich für jedes Stereobildpaar und Stereoverfahren ein Wert für den mittleren Fehler und die Varianz der Fehler in Pixel, jeweils dargestellt in Abbildung

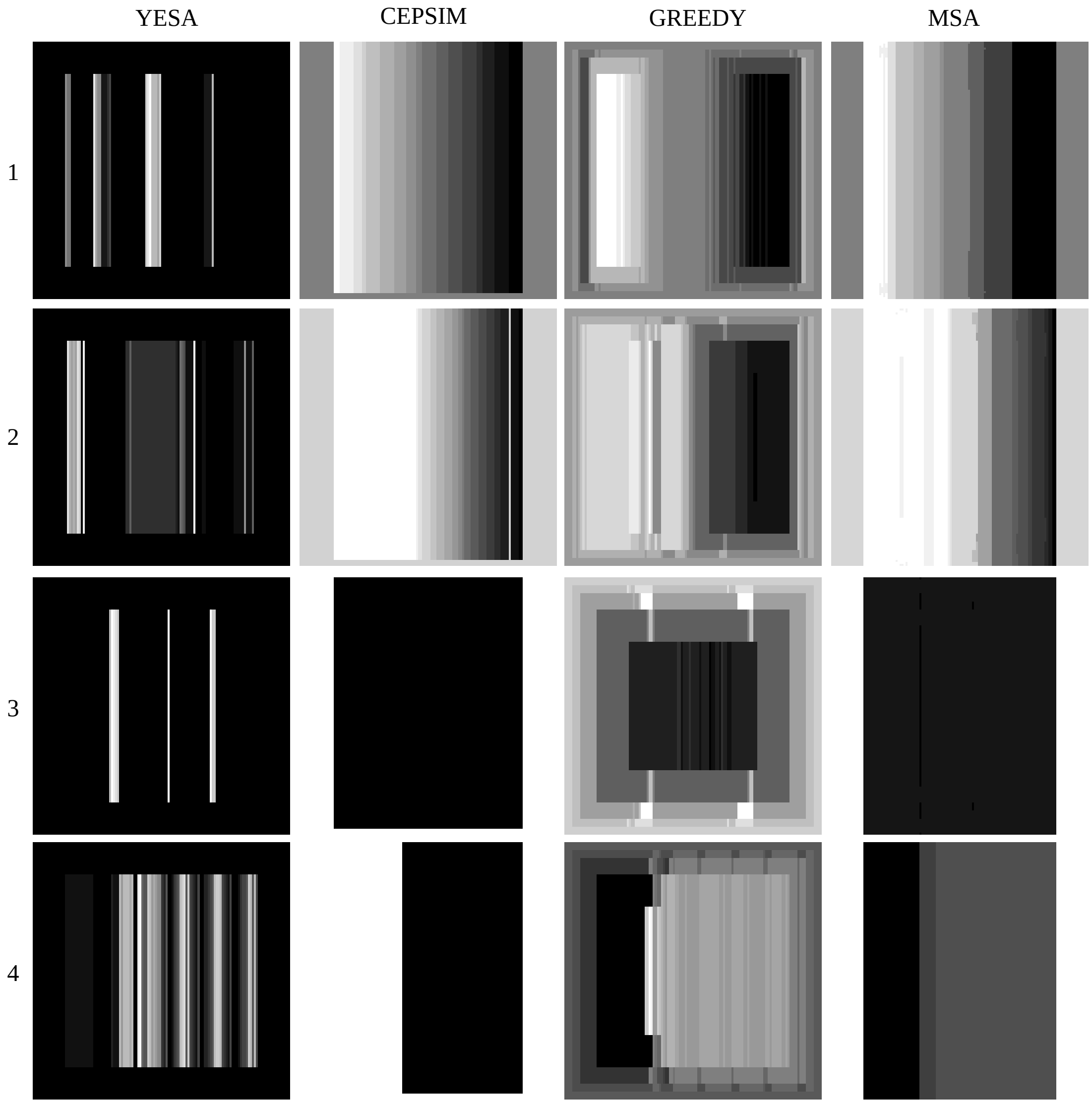


Abbildung 5.5: Die Anwendung der Stereoverfahren YESA, CEPSIM, GREEDY und MSA auf die synthetischen Stereobildpaare 1-4 der Simulation liefert die obigen Disparitätskarten.

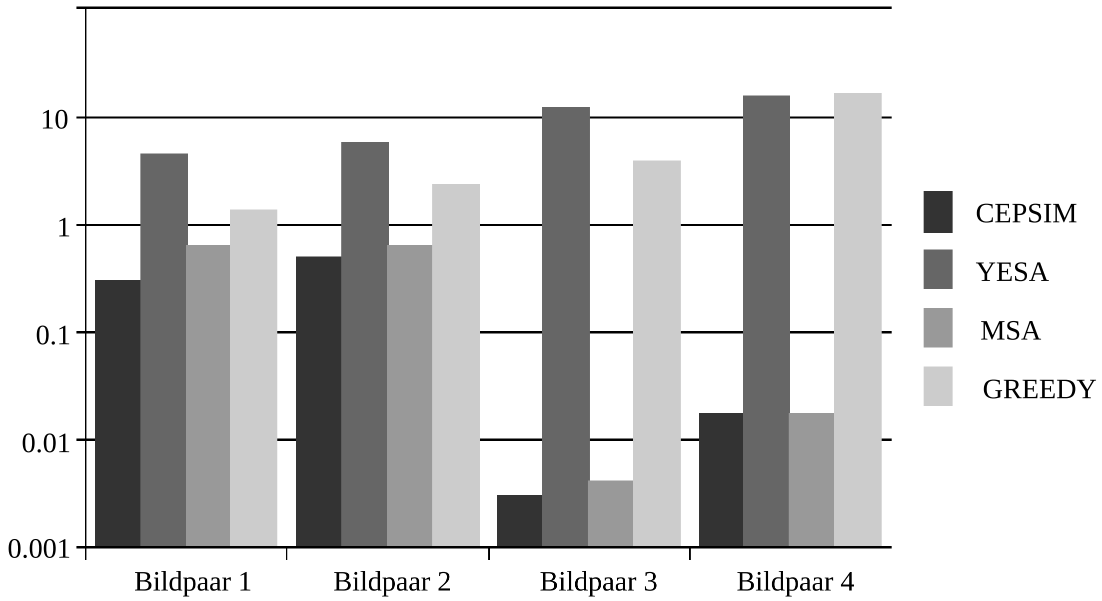


Abbildung 5.6: Der mittlerer Fehler (in Pixel, logarithmisch) der durch die Stereoverfahren gemessenen Disparitäten im Vergleich für die vier synthetischen Stereobildpaare.

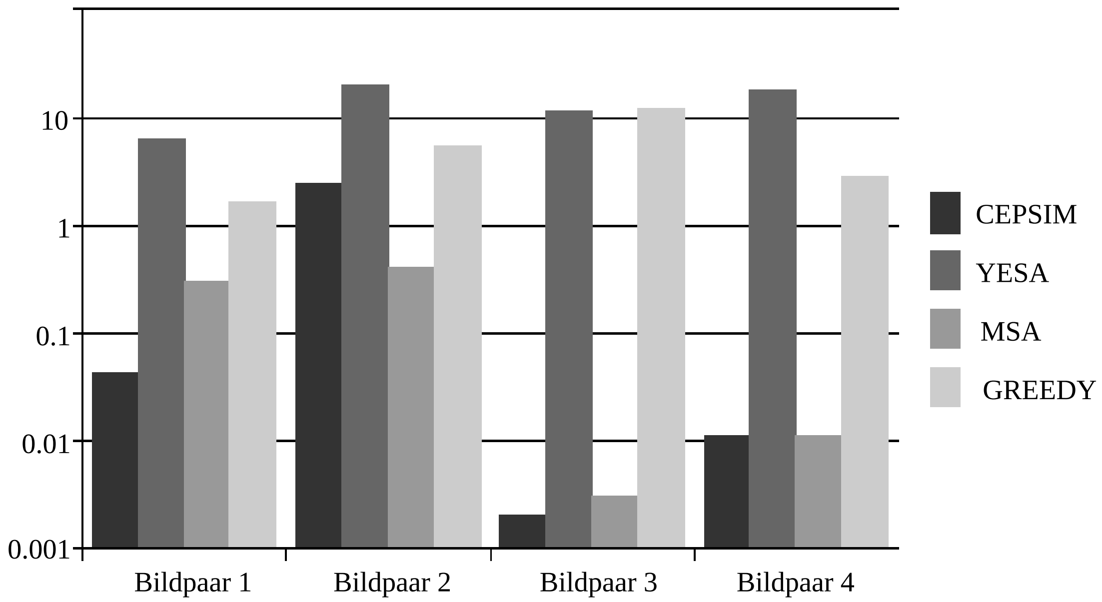


Abbildung 5.7: Die Varianz der Fehler (in Pixel, logarithmisch) der durch die Stereoverfahren gemessenen Disparitäten im Vergleich für die vier synthetischen Stereobildpaare.

5.6 und in Abbildung 5.7.
Weil die unterschiedlichen Verfahren starke Differenzen in den statistischen Werteausprägungen aufweisen, wurde eine logarithmische Darstellung gewählt.

Ein Zeichen von Robustheit des betreffenden Stereoverfahrens stellt eine Konstanz in den statistischen Werten über alle Stereobildpaare dar, wobei die Varianz als abgeleitete Größe im Zusammenhang mit dem mittleren Fehler betrachtet werden muß. Niedrige Werte in der Varianz sind nur im Zusammenhang mit einem geringen mittleren Fehler als positiv zu bewerten. Andernfalls weisen sie lediglich auf eine konstant falsche Disparitätsmessung des Stereoverfahrens hin.
Es ist zu erwarten, daß die statistischen Werteausprägungen für ein bestimmtes Stereoverfahren den oben beschriebenen Schwierigkeitsgrad der photometrischen Korrespondenzfindung in den einzelnen Testbildpaaren widerspiegeln.
Bei einer vollständig korrekten Zuordnung von Pixeln für ein Stereobildpaar ist ein mittlerer Fehler der Disparitätmessung von weniger als 0.5 Pixel zu erwarten. Diese Schranke ergibt sich aus einem systematischen Fehler durch Quantisierungseffekte, denn die gemessenen Disparitätskarten der Stereoverfahren sind ganzzahlig, während die analytisch ermittelten Disparitätskarten der Simulation auf Subpixelgenauigkeit exakt sind. Es folgt eine genauere Betrachtung der statistischen Werteausprägungen.

- **CEPSIM** : Das CEPSIM-Verfahren weist für die vier Stereobildpaare einen mittleren Fehler der Disparitätsmessung im Subpixelbereich auf. Wie zu erwarten liegt der mittlere Fehler in den Fällen der verzerrten Bildpaare 1 und 2 über den Werten für konstante Disparitätskarten 3 und 4. Bemerkenswert ist, daß im Fall 4 der mittlere Fehler und die Varianz von CEPSIM und MSA-Verfahren identisch sind. Erklären läßt sich dies aus der Tatsache, daß die Algorithmen identische Vergleichsmaße, Einzugsbereiche und Templategrößen verwenden. Weiterhin orientiert sich die genaue Zuordnung von Pixeln in beiden Algorithmen an der maximalen Auflösungsstufe der Stereobilder.
  Die Werte für die Varianz der Fehler besitzen nicht die Konstanz des MSA-Verfahrens. Im Fall des Bildpaares 2 liegt die Varianz bei 2.4 Pixel. Der Wert zeigt, daß in diesem Fall nicht ungenaue, sondern falsche Zuordnungen vorgenommen wurden. Diese falschen Zuordnungen zeigen sich als durchgehende Spalte in der betreffenden Disparitätskarte des CEPSIM-Stereoverfahrens aus Abbildung 5.5. Verantwortlich für die systematisch falschen Zuordnungen ist die heuristische Komponente der Skalensuchstrategie im Zuordnungsprozeß.
  Abgesehen von diesem Ausreißer jedoch liegt die Varianz der Fehler bei der Disparitätsschätzung durch CEPSIM im Subpixelbereich, ein Indiz auf konstant richtige Zuordnungen, wie sie für die Bildpaare 1-3 zu erwarten sind. In Bildpaar 4 erweist sich die verwendete Skalensuchstrategie des Algorithmus' als robust gegenüber falschen Zuordnungen durch Alias-Fehler.

- **YESA** : Die mittleren Fehler ebenso wie die Varianzen lassen in ihrer Höhe keinen sinnvollen Zusammenhang zwischen gemessenen und tatsächlichen Disparitätsinformationen erkennen. Dieser Zusammenhang ist nur bei allgemeiner Betrachtung der Disparitätskarten des YESA-Verfahrens in der Relation der Disparitätswerte zueinander zu erkennen (Abbildung 5.5).

  Als Ursache für die schlechten Ergebnisse des Verfahrens sind wiederum systematische Schwächen in der einfachen Analysestrategie von Antworten des Cepstrum-Filters anzuführen. In Bildpaar 1 und 2 wird zudem durch Verzerrungen von Bildstruktur die Annahme konstanter Disparität im Einzugsbereich als Grundlage der verwendeten theoretischen Interpretation der Cepstrum-Antwort verletzt. Dabei bergen Verzerrungen hauptsächlich die Gefahr der Überlagerung von Cepstrum-Impulsen.
  In Bildpaar 4 sind falsche Zuordnungen aufgrund von Alias-Problemen durch den festen Einzugsbereich des Cepstrum-Filters unvermeidlich, so daß hohe Werte für den mittleren Fehler und die Varianz zu erwarten sind.

  Insgesamt kommen bei den sinusförmigen Stereobildern systematische Schwächen des Algorithmus' stärker zum tragen als im Fall der Testbildpaare aus Abschnitt 5.1.
  Es zeigt sich deutlich, daß das YESA-Verfahren für eine praxistaugliche Implementierung zur Disparitätsmessung eine dem Cepstrum-Filter besser angepaßte Analysestrategie der Filterantwort benötigt, die systematische Probleme der Cepstrum-Technik ausschließt oder in eine Suchstrategie miteinbezieht.
  Zu diesem Schluß kommt ebenfalls [Lu91] in seiner Analyse des Cepstrum-Filters.

- **MSA** : Das MSA-Verfahren zeigt ein stabiles Verhalten in Bezug auf beide statistische Größen. Der mittlere Fehler wie die Varianz des Fehlers liegen vollständig im Subpixelbereich. Die durch den Algorithmus vorgenommenen Zuordnungen sind bis auf einzelne Ausreißer vollständig korrekt, denn die Werte für den mittleren Fehler kommen der oben beschriebenen optimalen Schranke durch Quantisierung von Disparitätswerten von 0.5 Pixel nahe. Gestützt wird diese Aussage durch die konstant geringe Varianz der Fehler, die unterhalb von 0.5 Pixel liegt.
  Auch liegen die statistischen Werte in einer erwarteten Rangfolge gemäß der Schwierigkeit der gestellten Aufgabe, die in der Ermittlung von Disparitätsinformationen aus den vier unterschiedlichen Stereobildpaaren liegt. Die verzerrten Bildpaare 1 und 2 führen zu höheren Fehlern und Varianzen der Disparitätsmessung als Bildpaar 4. Für Bildpaar 3 ergibt sich der geringste Fehler, wie es angesichts einer konstanten Disparität im Stereobildpaar auch zu erwarten ist.
  In dieser durchgehenden Plausibilität der Ergebnisse zeigt sich der Unter-

schied zum CEPSIM-Stereoverfahren. Er ist auf den Unterschied zwischen einer im jeweiligen Stereoverfahren verwendeten exakten beziehungsweise heuristischen Suchstrategie zur Ermittlung von Korrespondenzen zurückzuführen.

- **GREEDY** : Bei einer qualitativen Untersuchung der durch das GREEDY-Verfahren erzeugten Disparitätskarten hinterläßt der Algorithmus ein anderes Bild als bei allgemeiner Betrachtung der Disparitätskarten aus Abbildung 5.5. Dort erweisen sich die ermittelten Disparitäten zumindest in ihren Relationen als weitgehend zutreffend. In Bildpaar 4 werden für die Sinuswellen mit hoher Frequenz gleichförmige Zuordnungen getroffen. Durch die Pyramidenverarbeitung werden Alias-Fehler unterdrückt.
  In qualitativer Hinsicht weist aber gerade der Fall von Bildpaar 4 einen hohen mittleren Fehler auf, denn die getroffenen Zuordnungen sind zwar relativ gleichförmig aber mit einem konstanten Fehler behaftet. Dieser Effekt ist auf eine falsche Zuordnung in einer hohen Pyramidenstufe zurückzuführen.
  Generell verursachen derartige Blocking-Effekte höhere mittlere Fehler und Varianzen, als sie beim MSA oder CEPSIM-Verfahren auftreten, insbesondere bei den Bildpaaren 3 und 4.
  Für Bildpaar 1 ergeben sich wider Erwarten die besten statistischen Werte des Verfahrens. Im Vergleich zu Bildpaar 3 ist anzunehmen, daß die exaktere Messung kontinuierlicher als konstanter Disparitätsverläufe auf den Effekt der vollständigen Eliminierung von feinen Bildstrukturen (siehe Abschnitt 2.2.2) durch Tiefpaßfilterung in der Gauß-Pyramide zurückzuführen ist. Unter einer entsprechenden Tiefpaßfilterung weisen verzerrte Sinuswellen einen höheren Kontrast auf als die gleichmäßigen Sinuswellen aus Bildpaar 3. So werden Zuordnungen im Fall des Bildpaares 1 in hohen Pyramidenstufen korrekt vorgenommen, während bei Bildpaar 3 die Struktur soweit von hohen Frequenzanteilen befreit ist, daß in den gleichförmigen Grauwerten keine eindeutigen Zuordnungen mehr möglich sind.

Kapitel 6

Integration eines Stereoverfahrens in die Systemkomponente

Im allgemeinen und qualitativen Vergleich (siehe Kapitel 5) zeigt das MSA-Stereoverfahren das stabilste Verhalten.
Deshalb wird in diesem Kapitel die Implementierung von MSA auf dem MV200-System und die Koppelung des Stereoverfahrens an einen Vergenzregler beschrieben, so daß die Entwicklung der Systemkomponente zur aktiven Berechnung von Disparitätskarten abgeschlossen wird.

6.1 Implementierung von Stereoverfahren im MV200-System

In diesem Kapitel soll auf Fragen der Implementierung von Stereoverfahren auf dem MV200-System eingegangen werden. Im speziellen wird die Implementierung des MSA-Stereoverfahrens vorgestellt. Als Alternative wird eine Implementierung des CEPSIM-Verfahrens beschrieben, das durch seine Struktur einfacher auf dem MV200-System implementierbar ist. Aufgrund des Implementierungsaufwands liegt das CEPSIM-Verfahren nur in einer Planung durch Abbildung des Algorithmus' auf die Pipelinearchitektur vor, so daß Implementierungshindernisse ausgeräumt wurden.
Zum Verständnis der Implementierungsfragen werden Grundzüge des MV200-Systems und der Softwareschnittstelle Imageflow 2.5 vorgestellt.
Auch auf die Implementierung komplexerer Programmstrukturen auf dem MV200-System wird eingegangen. Dies verdient nicht zuletzt deshalb Beachtung, weil die zu Imageflow 2.5 gelieferte Dokumentation viele Fragen offenläßt. Sekundärliteratur ist rar, da die Spezialhardware keine weite kommerzielle Verbreitung besitzt.

Desweiteren wird auf Strategien zur effizienten Ausnutzung des MV200-Systems und Probleme unter Imageflow 2.5 eingegangen.
Als elementare Komponente der auf dem MV200-System implementierten Stereoverfahren wird die Realisierung eines Template-Matching geschildert. Am Beispiel der Implementierung des GREEDY-Stereoverfahrens werden die dabei auftretenden Probleme im MV200-System deutlich.

Aufbau des MV200-Systems und Elemente der Pipelineverarbeitung

Ein Pipelinecomputer ist ein Datenflußrechner. Er arbeitet nach dem Fließbandprinzip:
Ein komplexer Arbeitsgang wird in eine Vielzahl elementarer Einzelschritte zerlegt, die parallel von verschiedenen Verarbeitungsstationen ausgeführt werden. In sequentieller Verknüpfung ist die Ausgabe von Daten einer Station Eingabe der nächsten Station. Die Sequenz von Verarbeitungsstationen bildet eine Pipeline.
Ist die Pipeline (das Fließband) einmal voll durchlaufen, so werden Ausgaben des komplexen Arbeitsganges im Takt der einzelnen Verarbeitungsstationen ausgegeben. Dies ist ein Standardansatz zur Beschleunigung von Berechnungen in der Datenverarbeitung allgemein.
Im MV200-System sind die bearbeiteten Daten Bilder, beziehungsweise Pixel. Sie durchlaufen einzeln von einem Ursprungsbild (Bild = *Surface*) ausgehend sequentiell eine Pipeline (*Pipe*) und werden in einem Zielbild abgelegt. Der Takt beträgt dabei 20 Mhz, in Ausnahmen 40 Mhz. In den Pipelines sind Operationen erlaubt, die von der Bestückung des MV200-Boards mit Erweiterungsmodulen abhängen.
Das MV200-Board selbst (Ab) besteht hauptsächlich aus einem Multiplexer, der Leitungen zwischen Erweiterungsmodulen verschaltet. Die Erweiterungsmodule bestehen in der vorliegenden Bestückung aus Speichermodulen (Am), die die Bilder beinhalten, einer arithmetisch-logischen Einheit (Au) mit linearem, nichtlinearem und statistischem Teil, einem Faltungsmodul (Ap), einem Warper (Wx) und Modulen zur Digitalisierung (As) und Analogisierung (Ag) von Bildern zur Ein- und Ausgabe.
Der Datenfluß innerhalb einer Pipe von einem Speichermodul durch Erweiterungsmodule zu einem anderen Speichermodul wird durch ein Bussystem des MV200-Boards organisiert. Drei unabhängige Busse ermöglichen die Aktivität von drei Pipes auf einem MV200-Board parallel.
Das vorliegende MV200-System verbindet zwei MV200-Boards in obiger Erweiterung sowie vier MV200-Boards mit je einem Faltungsmodul.
Das System ist mit einem Host-Computer über einen VME-Bus verbunden, der zur Kommunikation während des Betriebs dient. Allerdings kann das MV200-System auch autark agieren.
Abbildung 6.1 vermittelt einen schematischen Eindruck vom MV200-System.

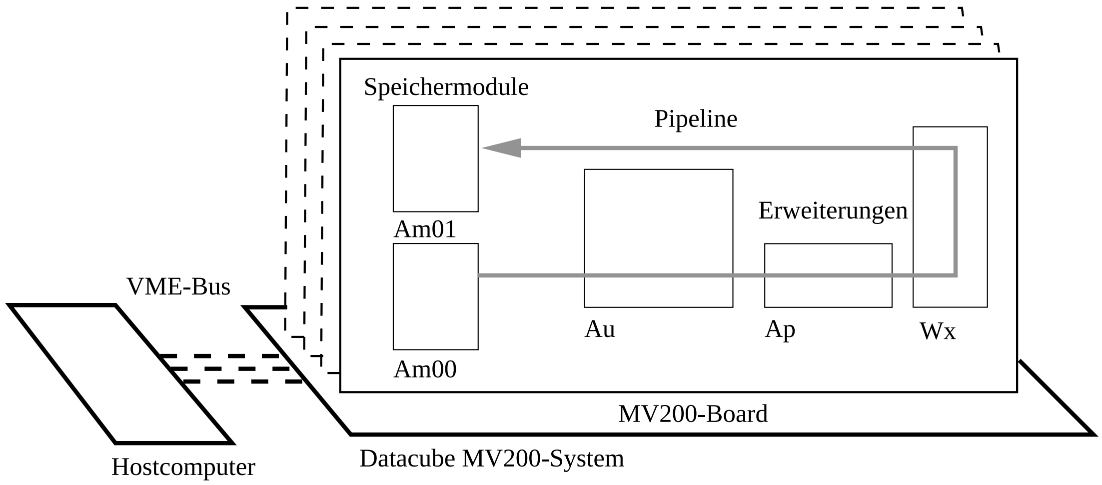


Abbildung 6.1: Dargestellt ist der schematische Aufbau des vorliegenden MV200-Pipeline-Systems von Datacube mit Anbindung an einen Host-Computer.

Softwareentwicklung unter Imageflow 2.5

Das Softwarepaket Imageflow 2.5 dient als Programmierschnittstelle der MV200-Hardware. Es spiegelt strukturell den Hardwareaufbau wieder, ist also relativ maschinennah.
Die Programmierung des MV200-Systems läßt sich auf die Implementierung von Pipestrukturen zurückführen. Um eine Pipe zu definieren, bedarf es der Festlegung des Ursprungsbilds, Zielbilds und des Datenwegs zwischen Speichermodulen, MV200-Board und Erweiterungsmodulen.
Von der Hardware aus betrachtet stellt eine Pipestruktur ein Schaltwerk dar. Der Datenweg in diesem Schaltwerk muß unter Imageflow 2.5 in atomaren Berechnungselementen (Gatter, Multiplexer, Alus, ...) vollständig festgelegt werden. Dabei verbirgt Imageflow ein Mindestmaß an Hardware vor dem Anwender. Anders gesprochen kommt der Anwender nicht darum herum, sich den Aufbau des MV200-Systems von der Hardware her genau zu eigen zu machen.
Neben dem Datenweg sowie Ursprungs- und Zielbild müssen in einer Pipe Datentypen und Berechnungsfenster (*Processing-Rectangles*) in den Bildern definiert werden. Ein Berechnungsfenster stellt den Bereich eines Bildes dar, dessen Pixel verarbeitet werden sollen.
Auch die Vergabe eines Datenbusses auf dem betreffenden MV200-Board und die Tiefe des Datenweges einer Pipe, die nicht beliebig groß sein kann, müssen implizit berücksichtigt werden. Imageflow 2.5 verbirgt die letztgenannten Arbeitsschritte zwar vor dem Anwender, kann aber die Abbildung der durch den Anwender vorgegebenen Pipestrukturen auf die Hardware nicht in jedem Fall durchführen. In diesem Fall muß der Anwender auftretende Hardware-Konflikte explizit von

Hand lösen.
Die Aufrufe verschiedener Pipes lassen sich unter Imageflow 2.5 eventgesteuert (Interrupts) in Abhängigkeit bringen. Pipe A ruft nach Beendigung Pipe B auf, Pipe B wiederum eine andere Pipe. Mit diesem Mechanismus lassen sich Sequenzen von Pipe-Aufrufen bilden, die ohne Kontrolle der CPU des Host-Computers ablaufen können.

Die Definition eines vergleichsweise komplexen Algorithmus erfordert eine Vielzahl von Pipestrukturen. Dies führt zu Ressourcenkonflikten und durch Seiteneffekte zum Verlust an Übersicht über das Zusammenwirken der Pipes. Imageflow bietet keine Schutzmechanismen. Das MV200-System verhält sich in der Programmierung dann wie ein komplexes Assemblerprogramm, das eine Vielzahl an globalen Variablen besitzt, die von mehreren Programmteilen überschrieben werden.
Um insgesamt nicht den Überblick zu verlieren, gibt es mehrere Möglichkeiten. Man kann vor dem Aufruf jeder Pipe den kompletten Datenweg der Pipe und alle erforderlichen Rahmenbedingungen explizit setzen. Dies erfordert eine große Anzahl von überflüssigen Imageflow-Befehlen, da der Datenweg meist aus früheren Pipeaufrufen oder aus Standardeinstellungen heraus bereits korrekt festgelegt ist. Diese Methode ist so aus Effizienzgründen nachteilig, außerdem sprengt sie in einem komplexeren Programm unter Verwendung von Pats (siehe Abbildung 6.1) unweigerlich den Speicherplatz, den Imageflow dem Anwender für diesen Zweck im Arbeitsspeicher des Host-Computers einräumt.
Deshalb hat sich eine Strategie bewährt, die für komplexere Algorithmen zu lauffähigen, effizienten Imageflowprogrammen führt. Dabei wird vor dem Aufruf eines Blocks von zusammengehörigen Pipes von einem Standard-Initialisierungszustand des MV200-Systems ausgegangen, der nach Ausführung des Blocks wiederhergestellt wird. So sind zur Ausführung einer Pipe nur wenige Imageflowanweisungen nötig, da diese im Datenweg allgemein Standardwerte verwendet.

Es gibt Programmierschnittstellen, die stärker von der Maschine des MV200-Systems abstrahieren als Imageflow 2.5. Sie beinhalten teilweise Möglichkeiten des interaktiven Tests von Pipeausführungen sowie dringend notwendige Werkzeuge zum Debugging. Dabei existieren graphische Oberflächen zur Implementierung von Pipestrukturen (*WITFLOW, MERLIN*). Die Oberflächen erleichtern die Programmierung des MV200-Systems, der Entwickler muß jedoch aus Effizienzgründen die Abbildung auf die Hardware meist explizit vornehmen, um diese geschickt nutzen zu können ([OlTaLo95]), weil sich komplexe Algorithmen durch einen Compiler nicht einfach auf die hochgezüchtete Hardware abbilden lassen.

Effizienz der Implementierung

Unter Imageflow gibt es die Möglichkeit, Blöcke von Pipeaufrufen gewissermaßen zu compilieren. Der resultierende Code wird in einer sogenannten *Pat* abgelegt. Imageflow führt bei der Pat-Generierung die Pipes in einer Art Simulation aus. So werden Parameter des Datenweges der Pipes im voraus berechnet und als Momentaufnahme in der Pat gespeichert. Diese Parameter betreffen beispielsweise die Bilder, Berechnungsfenster, Parameter der Berechnungselemente und Datenbuskonfigurationen. Durch diese Parameter ist in einer Anwendung die Zahl der notwendigen Pats unmittelbar von den gewünschten Parameterausprägungen der untergeordneten Pipes abhängig.
Aus Effizienzgründen besteht ein Imageflow-Programm auf höchster Ebene aus Pat-Aufrufen. Auch diese lassen sich komplett eventgesteuert ausführen, so daß keine CPU-Kontrolle des Host-Computers notwendig ist.
Das Ausmaß an Effizienzgewinn beim Übergang von Pipe- zu Pataufrufen läßt sich am Beispiel der Implementierung von MSA absehen. Die Zeit zur Berechnung einer Disparitätskarte sinkt infolgedessen von 0.33 Sekunden auf 0.23 Sekunden.

Um eine effiziente Implementierung eines Algorithmus auf dem MV200-System zu erreichen, sollte das Pipelineprinzip in einer möglichst großen Tiefe genutzt und die Systemkapazität zur parallelen Ausführung von Pipes ausgeschöpft werden.
In vielen Algorithmen stellt die Effizienz einen Kompromiß aus Laufzeit und Speicherbedarf dar. Im Falle des vorliegenden MV200-Systems hat sich gezeigt, daß das Kriterium des absoluten Speicherbedarfs eine untergeordnete Rolle spielt (insgesamt 72 MB Bildspeicher auf den zwei MV200 Boards).
Die Fähigkeiten der Speichermodule (arithmetisch, statistisch) sollten weitestgehend genutzt werden, weil sie gegenüber anderen Erweiterungsmodulen den Vorteil einer geringeren Pipeline-Tiefe bieten. Außerdem lassen sie sich intern mit doppelter Taktfrequenz betreiben.
Bei der Verwendung der Speichermodule ergibt sich allerdings ein Flaschenhals im Datenfluß. Sie können zu einem Zeitpunkt nur je einen 8-Bit-Datenstrom aufnehmen und abgeben. Möchte man in einer Pipe Operationen auf mehreren Datenströmen (z.B. Addition) durchführen, so kann man als Ausweg die Daten eines vorhergegangenen Verarbeitungsschrittes in verschiedenen Speichermodulen mehrfach ablegen (siehe Abschnitt 6.1.2). Bei zweistelligen Operationen in einer Pipe durch ein Speichermodul stammt dann ein Operand aus dem Modul selbst und der andere als Eingangsdatenstrom aus einem entsprechenden anderen Speichermodul.

Beschränkungen und Probleme

Typischerweise entstehen bei der Entwicklung eines MV200-Programms immer wieder Fehler aus Inkonsistenzen in den verwendeten Datentypen einer Pipe.

Einige Berechnungselemente des MV200-Systems lassen die Verarbeitung aller möglichen 8-Bit-Datentypen zu, andere nur die Verarbeitung bestimmter, beispielsweise vorzeichenbehafteter oder vorzeichenloser Datentypen.
Rundungsfehler durch die 8-Bit-Arithmetik sind zu beachten. Daher sollte man prüfen, ob sich Berechnungen, die den 8-Bit-Wertebereich kraß überschreiten, durch alternative Berechnungen ohne Rundungsfehler approximieren lassen (beispielsweise absoluter Betrag an Stelle des quadratischen Fehlers). Der dabei entstehende Fehler kann geringer sein als der alternative Rundungsfehler.
Eine Möglichkeit zur Reduzierung von Rundungsfehlern ist durch die Simulation einer 16-Bit-Arithmetik gegeben, die durch das Zusammenfassen von zwei 8-Bit Datenströmen entsteht. Diese Erweiterung der Rechengenauigkeit ist jedoch nicht im MV200-System vorgesehen und erfordert so einen unverhältnismäßig hohen Implementierungsaufwand.

In einem Imageflow-Programm ist der Entwickler aus Effizienzgründen darum bemüht, möglichst viel Code in Form von Pats zu implementieren. Bei der Softwareentwicklung unter Imageflow 2.5 ergibt sich ein Problem mit dem für Pat-Code reservierten Hauptspeicher des Host-Computers. Bei komplexeren Programmen steigt die Zahl an Pat-Strukturen und damit der Bedarf an Pat-Speicherplatz rapide an, so daß dieser gegebenenfalls überläuft.
Da die Anzahl der Pats von möglichen Parameterausprägungen untergeordneter Pipe-Aufrufe abhängt, ist es möglich, auf maschinennaher Ebene innerhalb einer Pat-Struktur mehrere Parameterausprägungen zu verarbeiten (*Pat-Variables*), den implementierten Algorithmus zu vereinfachen oder einfach Pat-Strukturen zum Preis einer höheren Laufzeit des jeweiligen Programms in Pipes und Codesegmente des Host-Computers aufzulösen, um einen Speicherüberlauf zu vermeiden.

Template-Matching im MV200-System

Eine weitere Beschränkung der MV200-Hardware liegt in der Einhaltung eines festen, rechteckigen Rasters in der Verarbeitungsreihenfolge von Pixeln zwischen Ursprungsbildern und Zielbildern einer Pipe. Die Topologie aller beteiligten Pixel bleibt erhalten.
Im Fall der Implementierung eines Template-Matching zwischen zwei Bildern, bei dem verschiedene Templates des einen Bildes mit Suchbereichen im anderen Bild verglichen werden, wird die Abbildung der unregelmäßig angeordneten Templates und Suchbereiche auf ein festes Pixelraster notwendig, um das Template-Matching in einem einzigen Pipetransfer durchführen zu können.
Zu zwei Eingabebildern (L,R) wird in der Literatur der Ansatz verfolgt, für ein Template $T_L(i,j)$ aus L und einen Suchbereich T_R aus R, der aus einer Anzahl von Matching-Kandidaten besteht, einerseits $T_L(i,j)$ in Wiederholung zu einem Bild L', andererseits T_R zu einem Bild R' derselben Dimension zusammenzufas-

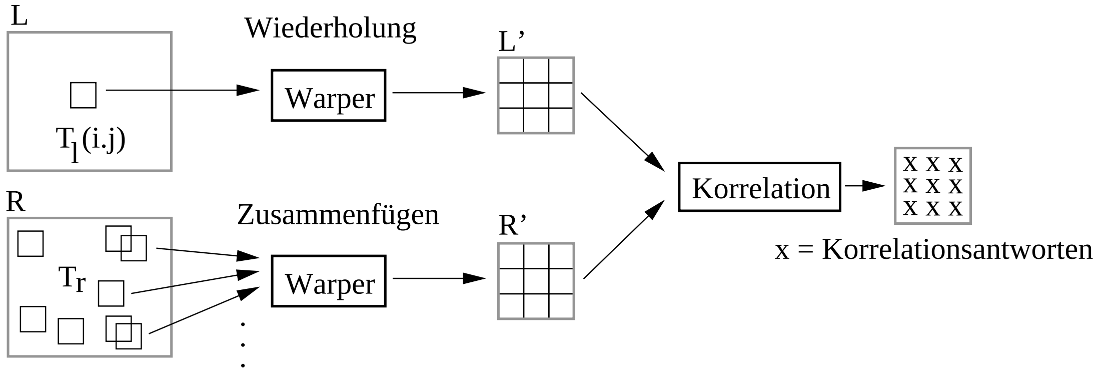


Abbildung 6.2: Um Korrelationsantworten eines Templates im Suchbereich effizient auf dem MV200-System berechnen zu können, bedarf es der Abbildung der unregelmäßig angeordneten Templates auf ein festes Pixelraster. Diese Aufgabe kann im MV200-System das Warpermodul erfüllen.

sen (siehe Abbildung 6.2). L' und R' können nun in festem Pixelraster verglichen werden, beispielsweise durch Korrelation.
Im MV200-System bietet allein das Warper-Erweiterungsmodul die Möglichkeit, die Abbildung auf ein festes Pixelraster durchzuführen. Von diesem Modul machen denn auch MV200-Entwickler zur Implementierung eines korrelationsbasierten Template-Matching Gebrauch ([IDUC94]).

Der Nachteil dieser Strategie liegt in der entstehenden Expansion der Bilder. Eine um den Faktor der Templategröße und die Anzahl von Suchpositionen höhere Zahl an teilweise redundanten Pixeloperationen ist notwendig. Das resultierende Bild von Korrelationsergebnissen ist schwach besetzt.

Bei einer MV200-Implementierung des GREEDY-Stereoverfahrens kommt zur Realisierung eines Template-Matching allerdings eine andere Strategie zur Anwendung, weil die mit obiger Strategie einhergehende Expansion der Bilder in der vorliegenden Anwendung die Kapazität des MV200-Systems überschreitet.
Im GREEDY-Stereoverfahren (siehe Abschnitt 4.3.1) wird ein korrelationsbasiertes Template-Matching zwischen Stufen $P_l(k)$, $P_r(k)$ einer Pyramide der Eingabebilder durchgeführt, wobei für jedes Template von $P_l(k)$ ein unterschiedlicher Suchbereich in $P_r(k)$ verwendet wird. Der Suchbereich ist in der Pyramidenstufe k durch die Disparitätskarte der vorausgegangenen Stufe D_{k+1} gegeben.
Die unterschiedlichen Suchbereiche müssen wie beschrieben auf ein festes Pixelraster abgebildet werden, um Korrelationsantworten aller Templates in ihren Suchbereichen durch eine einzige Pipe berechnen zu können.
Als besagte Alternative zu obiger Strategie wird wie in vielen Verfahren der Literatur von der Annahme eines stetigen Verlaufs der Disparitätskarte ausgegangen (siehe Abschnitt 2.2.1). Nach dieser Annahme kann man für benachbarte Pixel in

$P_l(k)$ und $P_r(k)$ von ähnlichen Disparitätswerten in der vorausgegangenen Pyramidenstufe und damit ähnlichen Suchbereichen in der aktuellen Pyramidenstufe ausgehen.
Die Idee ist nun, die in $P_r(k)$ durch D_{k+1} vorgegebenen unterschiedlich verteilten Suchbereiche durch eine entsprechende Umordnung von $P_r(k)$ mit Hilfe des Warper-Moduls so auf ein Bild $P_r(k)'$ abzubilden, daß die Suchbereiche in demselben regelmäßigen Pixelraster wie die Templates in $P_l(k)$ vorliegen (siehe Abbildung 6.3).
Dabei wird durch die Zuordnungen der vorausgegangenen Stufe $P_r(k)$ gewissermaßen durch Verzerrung an $P_l(k)$ angeglichen, so daß eine Verfeinerung der Zuordnung durch Korrelation in festem Pixelraster möglich ist.
Im Gegensatz zum vollständigen Kopieren von Templates und Suchbereichen bleibt bei dem Ansatz der Umordnung von Bildstruktur die Dimension der zu korrelierenden Bilder konstant.

Der beschriebene Ansatz für ein Template-Matching im GREEDY-Stereoverfahren besitzt jedoch unterschiedliche Schwächen.
Zum einen ruft die Verzerrung von Bildstruktur gemäß D_{k+1} Blocking-Effekte in $P_r(k)'$ hervor.
Im Fall des künstlichen Signals B in Abbildung 6.3 sind die Blocking-Effekte noch moderat, während sie im Fall A der realen Szene fatale Ausmaße annehmen (von Randeffekten ist abzusehen). Fatal deshalb, weil Blocking-Effekte dominante Korrelationsziele darstellen. Die Korrelationsantworten der Stufe k werden so stärker durch den systematischen Fehler der Blocking-Effekte beeinflußt als durch die Bildstrukturen in $P_r(k)$ selbst. Blocking-Effekte werden durch die Stufen der Top-Down-Suche in den Disparitätskarten stärker, weil mit fallender Pyramidenstufe der Kontrast der Bilder zunimmt.
Zum anderen überschneiden sich in $P_r(k)'$ die Suchbereiche der Templates. Wie in Abschnitt 2.1.1 beschrieben, stellt bei der Korrespondenzsuche die Einbeziehung von Disparitätsinformationen der Nachbarschaft grundsätzlich eine qualitative Verbesserung der Suchstrategie dar. Im vorliegenden Ansatz jedoch können sich in $P_r(k)'$ Suchbereiche überschneiden, die aus vertrauens- und nicht vertrauenswürdigen Zuordnungen von D_{k+1} resultieren. So werden vertrauenswürdige Zuordnungen der Stufe k durch nicht vertrauenswürdige Zuordnungen der Stufe $k+1$ beeinflußt. Die Miteinbeziehung eines Vertrauensmaßes schon während der Top-Down-Suche des GREEDY-Verfahrens im MV200-System wäre erforderlich, was allerdings eine einschneidende Veränderung des Algorithmus gegenüber der Workstation-Implementierung darstellen würde.

Wegen der geschilderten Probleme bei der Realisierung eines Template-Matching wurde die Implementierung des GREEDY-Verfahrens auf dem MV200-System nicht weiter verfolgt, obwohl die Strategie der Umordnung von $P_r(k)$ zu $P_r(k)'$ interessante Eigenschaften aufweist. In Abbildung 6.3 wird deutlich, daß sich aus der Abbildungsgeometrie resultierende Verzerrungen zwischen den Bildern redu-

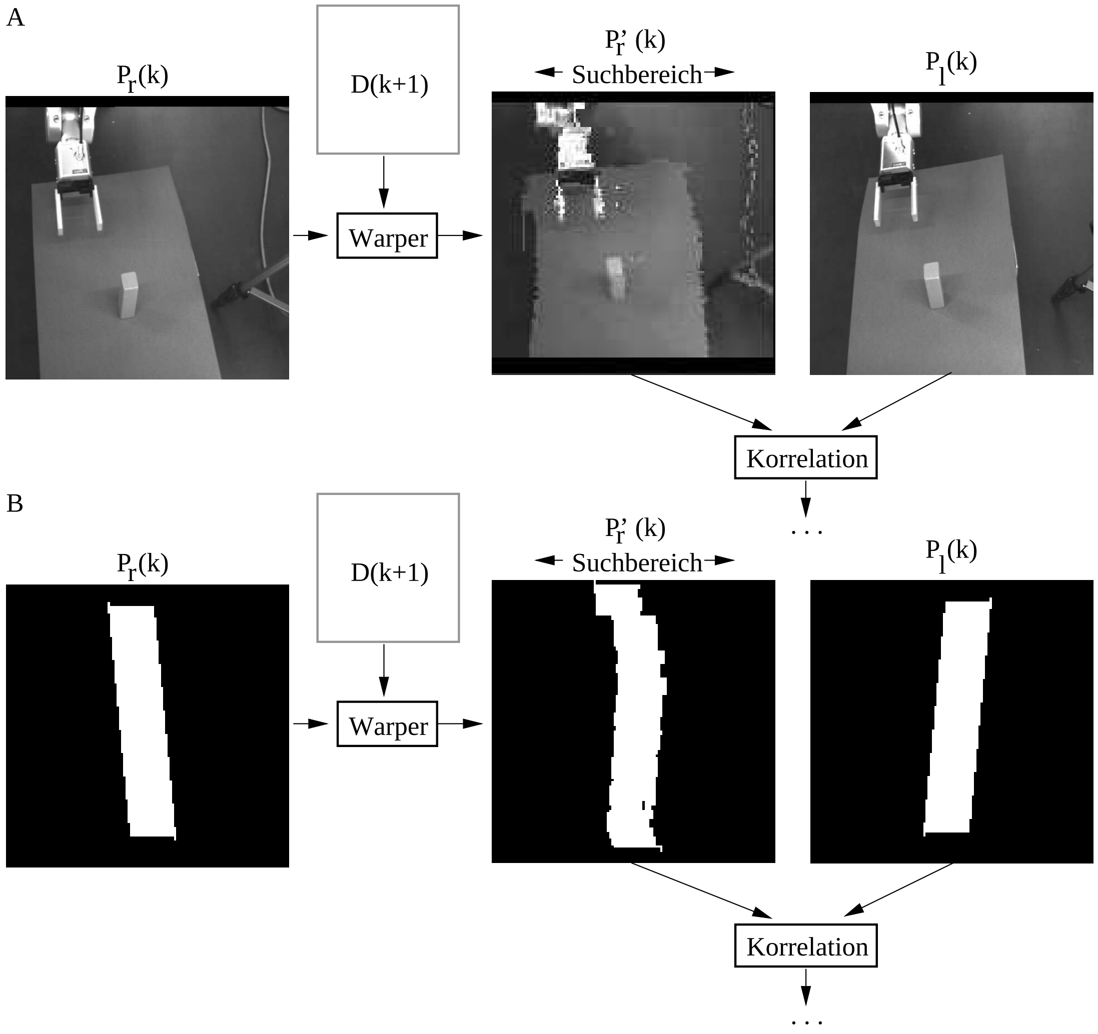


Abbildung 6.3: Das GREEDY-Stereoverfahren initialisiert Suchbereiche für Templates mit der Disparitätskarte der vorausgegangenen Pyramidenstufe. Um die ungleichmäßig verteilten Suchbereiche auf regelmäßige Datenstrukturen des MV200-Systems abbilden zu können, wird unter der Annahme stetiger Disparitätsverläufe das gesamte Eingabebild, in dem die Suchbereiche liegen, entsprechend verformt. Dabei ergeben sich jedoch fatale Blocking-Effekte wie im Fall der zwei Testbildpaare.

zieren, so daß ein korrelationsbasiertes Matching beim Vergleich von Templates zwischen $P_l(k)$ und $P_r(k)'$ gegenüber $P_r(k)$ bessere Vergleichsbedingungen vorfindet.

6.1.1 MSA auf dem MV200-System

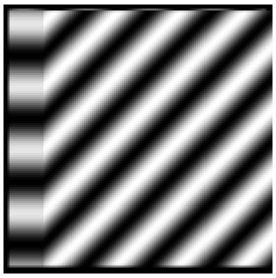
Der Aufbau des MSA-Stereoverfahrens wurde in Abschnitt 4.3.2 vorgestellt. Die Implementierung auf dem MV200-System verlangt allerdings einige Modifikationen, die aus Restriktionen der Hardware des MV200-Systems und Imageflow 2.5 resultieren. Diese Modifikationen zu obigem Algorithmus werden nun vorgestellt.

1 : Bildung der Gauß-Pyramiden
Es werden vier Stufen (Dimension 128x128 ... 16x16) einer 6-stufigen Gauß-Pyramide verarbeitet. Randprobleme durch die Tiefpaßfilterung bei der Bildung der Pyramiden werden durch Auffüllen mit Nullen behandelt. Dies hat zur Folge, daß Pixel im Einflußbereich des Randes dunkler werden. Der Einfluß dieser Ungleichbehandlung des Randes wird durch die Normierung der folgenden Korrelation kompensiert.

2 : Bildung einer Pyramide von Korrelationsergebnissen
Korrelationsergebnisse werden durch ein Template-Matching ermittelt. Die Dimension der Templates verursacht einen Rand in vertikaler Richtung, der mit Nullen aufgefüllt wird. Korrelationswerte bleiben davon unbeeinflußt, weil nur eine horizontale Kante entsteht.
In horizontaler Richtung ergibt sich ein breiterer Rand, der wiederum durch die Dimension des Templates, aber hauptsächlich durch den horizontalen Suchbereich entsteht. Als Randbehandlung wird eine geeignete horizontale Erweiterung der zu korrelierenden Stereobilder benötigt, weil das Auffüllen von Nullen im allgemeinen eine die Korrelation dominierende vertikale Kante erzeugen würde. Stattdessen wird das Signal auf dem undefinierten Rand durch Fortsetzung mit der äußersten Spalte erweitert, so daß keine dominierenden vertikalen Strukturen entstehen (Abbildung 6.4).

Das Vergleichsmaß NKK (siehe Abschnitt 4.1) besitzt einen reellen Wertebereich im Intervall [-1,1]. Die Implementierung von NKK erfordert eine Anpassung des reellen Wertebereichs an die 8-Bit-Integer-Arithmetik des MV200-Systems.
Bei der Berechnung von NKK im MV200-System treten weiterhin Probleme bei der Implementierung der benötigten Divisions- und Quadratwurzelfunktion auf, die nicht in der Hardware vorgesehen sind.
Unter Imageflow 2.5 erwiesen sich die durch Simulation nicht vorhandener Hardwarefunktionen komplexen Patstrukturen für die Implementierung von NKK als zu umfangreich, um nicht im Zusammenspiel mit dem restlichen Programmcode von MSA den Pat-Speicherplatz zu überschreiten.
Als Ausweg wird ein Vergleichsmaß NKK' für Templates $T_L(i,j)$ und $T_R(k,l)$

Erweiterung von P_l



Erweiterung von P_r

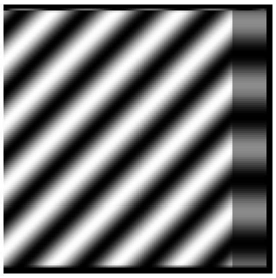


Abbildung 6.4: Durch einen horizontalen Suchbereich beim Template-Matching entsteht ein entsprechender undefinierter Rand in den Stereobildern. Im MSA-Stereoverfahren wird eine horizontale Randbehandlung durch Wiederholung der äußersten Spalten durchgeführt.

mit einer vereinfachten, linearisierten Normierung in Betracht gezogen, die sich von NKK ableiten läßt:

$$NKK'(T_L(i,j),T_R(k,l)) = \frac{\sum_{x,y=0}^{M-1} T_L(i,j)(x,y)T_R(k,l)(x,y)}{\sum_{x,y=0}^{M-1} T_L(i,j)(x,y) \sum_{x,y=0}^{M-1} T_R(k,l)(x,y)}$$

Diese an die Hardware des MV200-Systems angepaßte Normierung liefert zwar für Testsignale die gewünschte Normierungseigenschaft, gibt aber unerwünschte Bewertungen von Templates ab. Auf das eindimensionale Testsignal aus 4.1 antwortet NKK' mit identischen Korrelationsbeträgen für unterschiedliche Templates im Suchbereich (siehe Abbildung 6.5). Diese Tatsache ist unerwünscht, weil die Suchstrategie von MSA Korrelationsbeträge analysiert. Weil die modifizierte, normierte Korrelation jedoch konform mit den von Imageflow 2.5 vorgesehenen System-Ressourcen ist, wird NKK' in der Implementierung von MSA verwendet. Die Normierung der Korrelation erfolgt auf das Integer-Intervall [-31 ... 31], das zur Vermeidung eines Überlaufs im Wertebereich 8-Bit-Signed-Data durch die Suchstrategie 4. vorgegeben ist.
In Abschnitt 6.1.2 wird beispielhaft die Implementierung des exakten Vergleichsmaßes NKK durch Pipestrukturen vorgestellt. Das modifizierte Maß NKK' führt zu einer einfacheren Implementierung.

3 : Extraktion lokaler Maxima
Ein lokales Maximum der Korrelation wird als schwaches ($\geq$) Maximum eines Pixels in der Pyramide von Korrelationsergebnissen KP bezüglich benachbarter Suchpositionen definiert. Schwach deshalb, weil durch den Wertebereich von KP in 6-Bit Auflösung Maxima in Form eines Plateaus auftreten können. Liegt das Pixel am Rand des Suchbereichs, so wird Null als Nachbarwert angenommen. Alle Pixel in KP, die ein solches horizontales Maximum im Suchbereich aufweisen, bleiben unberührt, alle anderen werden auf Null gesetzt. Es ergibt sich eine

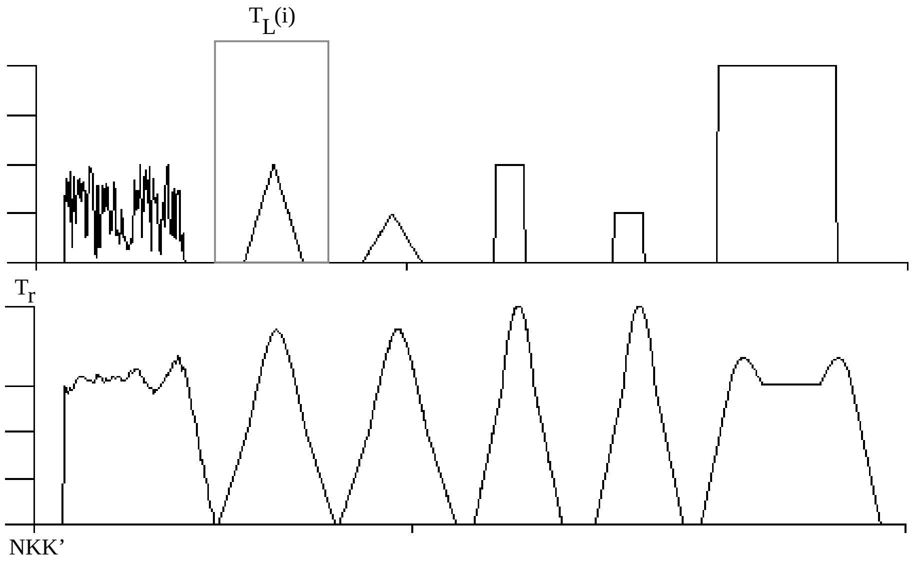


Abbildung 6.5: Augrund von Systembeschränkungen wurde auf dem MV200-System eine modifizierte normierte Korrelation implementiert. Sie erfüllt zwar gewünschte Normierungseigenschaften, trennt aber vom Betrag der Korrelationsantwort her ähnliche Signalstrukturen nur unbefriedigend.

Datenstruktur KP', in der nur noch lokale Korrelationsmaxima vorkommen.

4 : Suchstrategie

Zum Verständnis der Implementierung der Suchstrategie ist es hilfreich, sich das in Abbildung 4.6 vorgestellte Modell zum Auffinden von Pixelkorrespondenzen in KP' in Erinnerung zu rufen. Eine Korrespondenz ist dort durch einen Streifen von Korrelationsmaxima im Suchbereich mit maximaler Bewertung definiert.

In der MV200-Implementierung wird die Suchstrategie modifiziert. Streifen können laut Modell in einem beliebigen Bereich von Pyramidenstufen KP beziehungsweise KP' erscheinen. Eine direkte Umsetzung der Suchstrategie würde erfordern, daß Streifen unabhängig von ihrer Lage in den Pyramidenstufen detektiert werden, wie im Fall 1 in Abbildung 6.6. Diese Strategie wurde implementiert, sie führte jedoch in ihrer Komplexität im Stereoverfahren wiederum zu einem Überlauf des Pat-Speicherplatzes.

Eine modifizierte, einfachere Suchstrategie war hingegen verträglich mit den Systemressourcen. Sie beinhaltet die Annahme, daß jeder Streifen bis hin zur untersten Pyramidenstufe (maximale Auflösung) vorhanden ist (Fall 2 in Abbildung 6.6). Nur solche Streifen werden detektiert und mit der Summe der beteiligten lokalen Maxima gewichtet. Besitzt eine Suchposition keinen derartigen Streifen, wird ihr die Bewertung Null zugewiesen.

Insgesamt ergibt sich eine Bottom-Up-Suche nach Streifen in KP'. Bewertun-

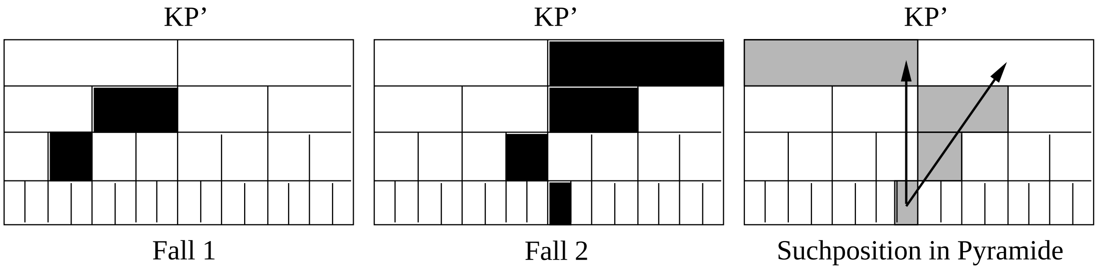


Abbildung 6.6: Die Bottom-Up-Suchstrategie des MSA-Verfahrens im MV200-System fordert den Zusammenhang eines Streifens lokaler Korrelationsmaxima bis hin zur vollen Auflösung der Pyramide. Fall 1 wird durch die Skalensuche mit Null bewertet, Fall 2 mit der Summe der Korrelationsmaxima über die Pyramidenstufen hinweg. Durch Mittelwertbildung benachbarter Suchpositionen wird in der Bottom-Up-Skalensuche der Korrelationswert der exakten Suchposition in der jeweiligen Pyramidenstufe angenähert.

gen von Suchpositionen beziehen sich so generell auf Pixel der untersten Pyramidenstufe und sind aufgrund der Strukturgebundenheit eines Streifens nur im Einzugsbereich von 7x7 Pixel um Bildstrukturen ungleich Null. Die insgesamt resultierenden Disparitätskarten des Stereoverfahrens sind entsprechend schwach besetzt.
Den Unterschied zwischen Disparitätskarten des MSA-Stereoverfahrens in der MV200- und Workstation-Implementierung zeigt Abbildung 6.7. Für dasselbe Stereobildpaar ist die Disparitätskarte der MV200-Implementierung schwächer besetzt. Aufgrund der Strukturgebundenheit der Detektion von Stereodisparität auf photometrischer Basis allgemein ist dies allerdings kein Nachteil.
Die modifizierte Suchstrategie wurde nach Erprobung mehrerer Alternativen ausgewählt. Ein Vergleich zwischen der exakten und der modifizierten Suchstrategie förderte für Testbilder keine Unterschiede zu Tage. Erklären läßt sich diese Tatsache mir der geringen Tiefe der Pyramide von vier Stufen. Außerdem ergibt sich die Stufe maximaler Auflösung in KP' für die Suchstrategie aus Signalen, die ausgehend von der Bildaufnahme die zweite Stufe einer Gauß-Pyramide darstellen und somit von Rauscheinflüssen weitgehend befreit sind.
Auch theoretisch läßt sich die ausschließliche Bottom-Up-Suche aus der Überlegung heraus rechtfertigen, daß sich lokale Maxima der Korrelation innerhalb eines Streifens bedingen. Ein lokales Maximum in einer Pyramidenstufe von KP' bedingt aus numerischen Gründen (siehe Abschnitt 4.3.2) ein Maximum an entsprechender Suchposition der darunterliegenden Stufe. Man kann demnach die Annahme aufstellen, daß der Streifen von Korrespondenz im Suchbereich bis auf die unterste Stufe beziehungsweise bis zur maximalen Auflösung von KP' hinabreicht.

Zur Detektion von Streifen wird das Kriterium des Zusammenhangs von Korrelationsmaxima über Pyramidenstufen hinweg verwendet, die an entsprechender

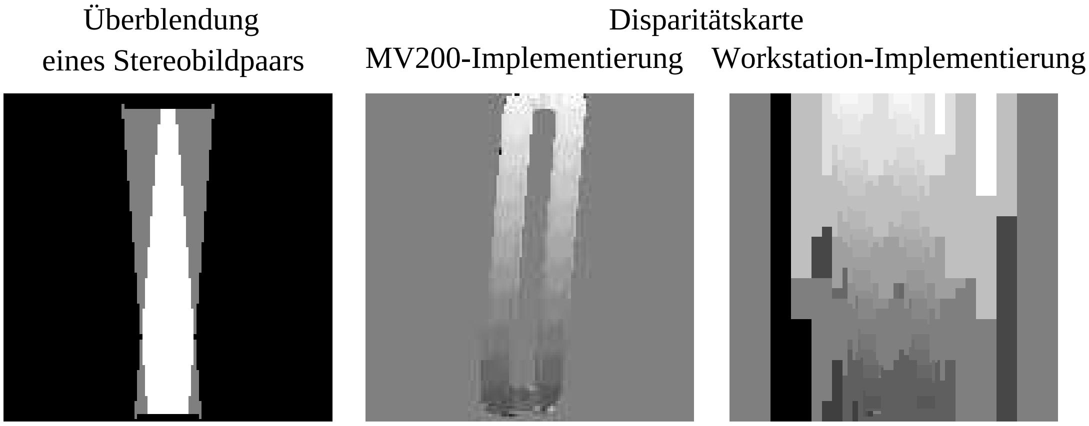


Abbildung 6.7: Das MSA-Stereoverfahren liefert in der MV200-Implementierung schwächer besetzte Disparitätskarten als die Workstation-Implementierung. Ursache für diesen Effekt sind eine modifizierte Skalensuche sowie Rundungsfehler der 8-Bit-Arithmetik des MV200-Systems. Zu beachten ist der Bezug der beiden Disparitätskarten auf das jeweils rechte oder linke Kamerabild.

Suchposition liegen. In Abschnitt 4.3.2 wurde dabei auf die Notwendigkeit der Betrachtung von Diskretisierungseffekten hingewiesen. Eine lineare Interpolation von Korrelationswerten an Suchpositionen, die der exakten Position benachbart sind, wird in der MV200-Implementierung durch einfache Mittelwertbildung ersetzt (Abbildung 6.6).

Die Bewertungen der Suchpositionen für alle Pixel liegen als Summe der auf das Intervall [-31 ... 31] normierten Korrelationswerte über maximal vier Pyramidenstufen im Intervall [-124 ... 124] (8-Bit-Signed-Data).

5 : Erzeugung von Disparitätskarte und Vertrauensmaß
Für jedes Pixel wird die Position mit maximaler Bewertung im Suchbereich ausgewählt. Existiert kein strenges Maximum, wird Null als Disparität und Vertrauensmaß ausgegeben.

MSA wurde unter Verwendung der zwei mit mehreren Erweiterungsmodulen bestückten MV200-Boards (Ab00, Ab01) implementiert. Die Laufzeit des Algorithmus wurde nach Gesichtspunkten aus Abschnitt 6.1 optimiert. Dies betrifft Strategien einer möglichst hohen Pipelinetiefe sowie der gleichmäßigen Ausnutzung der Parallelkapazität des MV200-Systems.
So werden in der vorliegenden Implementierung zu einem Zeitpunkt mindestens zwei Bilder parallel verarbeitet. Zu diesem Zweck werden die vorhandenen Datenstrukturen wie beispielsweise Pyramiden im Suchbereich mittig auf die einzelnen Boards aufgeteilt. Für jeden der Schritte 1.-5. von MSA sind infolgedessen identische Pipestrukturen auf beiden MV200-Boards vorhanden, die getrennte Hälften

der Datenstrukturen bearbeiten. Es existieren spezielle Pipes, die einen zuweilen notwendigen Informationsaustausch zwischen beiden Boards ermöglichen.
MSA ist ohne Kontrolle des Host-Computers autark auf dem MV200-System lauffähig. So demonstriert die Implementierung die an heutigen Maßstäben gemessen hohe Rechenleistung der zwei MV200-Boards. Die Implementierung ist in der Lage, einen konstanten Takt von der Bilddigitalisierung bis zur fertigen Disparitätskarte von 0.237 s einzuhalten (Mittelwert über einen Zyklus von 100 Takten). In einem Takt werden zum Teil aufwendige Berechnungen (NKK, Skalensuchstrategie) auf schätzungsweise 187 Bildern der Dimension 128x128 Pixel durchgeführt. Diese Zahl ergibt sich aus der Berechnung der zwei Gauß-Pyramiden ausgehend von Eingabebildern der Dimension 512x512 Pixel (40 Bilder), der Berechnung der Korrelationspyramide KP (38 Bilder), der Maximumdetektion in KP (38 Bilder), der Skalensuche in KP' (38 Bilder) und schließlich der Generierung von Disparitätskarte und Vertrauensmaß (33 Bilder).

6.1.2 Beispiel einer Pipestruktur

Exemplarisch für die Implementierung des MSA-Stereoverfahrens auf dem MV200-System soll die Berechnung der normierten Korrelation NKK aus Abschnitt 6.1.1 durch eine Kombination von Pipes/Pats vorgestellt werden. Das MV200-System ist für derartige Berechnungen ausgelegt. Kompliziertere Programmflußstrukturen erfordern einen ungleich höheren Implementierungsaufwand.
Im Zustand vor der Berechnung der Korrelation befinden sich die berechneten Gauß-Pyramiden P_l, P_r der Eingabebilder in je einem entsprechenden Speichermodul (Am02 , Am12) der zwei verwendeten MV200-Boards.
In der folgenden Ermittlung der Korrelationspyramide KP wird die angesprochene Trennung von Datenstrukturen zur Parallelisierung durchgeführt. Board Ab00 ist für die Berechnung der einen Hälfte von KP im Suchbereich $0 \ldots -16$ Pixel zuständig, Board Ab01 für die Berechnung der anderen Hälfte von $0 \ldots 16$ Pixel. Aus diesem Grund sind die verwendeten Pipes für beide MV200-Boards identisch, sie werden in der folgenden Beschreibung nicht unterschieden. Die Pipes berechnen jeweils NKK in der k. Stufe von KP für die Suchpositionen $\pm i$ parallel, wie in Abbildung 6.8 zu sehen ist. Dort sind die zur Berechnung von NKK verwendeten Pipes dargestellt. Sie spiegeln jeweils Teile der Formel 4.2 aus Abschnitt 4.1 wider.

Zunächst wird in `TransmittPipe` das Bild $P_l(k)$ horizontal um i Pixel verschoben und im Bild `Aop` abgelegt. Dadurch wird die Suchposition für die anschließende Korrelation festgelegt. Eine Verschiebung wird durch die Veränderung von Berechnungsfenstern der Bilder erreicht. $P_r(k)$ gelangt ohne Verschiebung nach `Bop`. Zur parallelen Weiterverarbeitung werden die Ergebnisse des Pipetransfers auf beiden MV200-Boards abgelegt, jedoch mit entgegengesetzten Verschiebun-

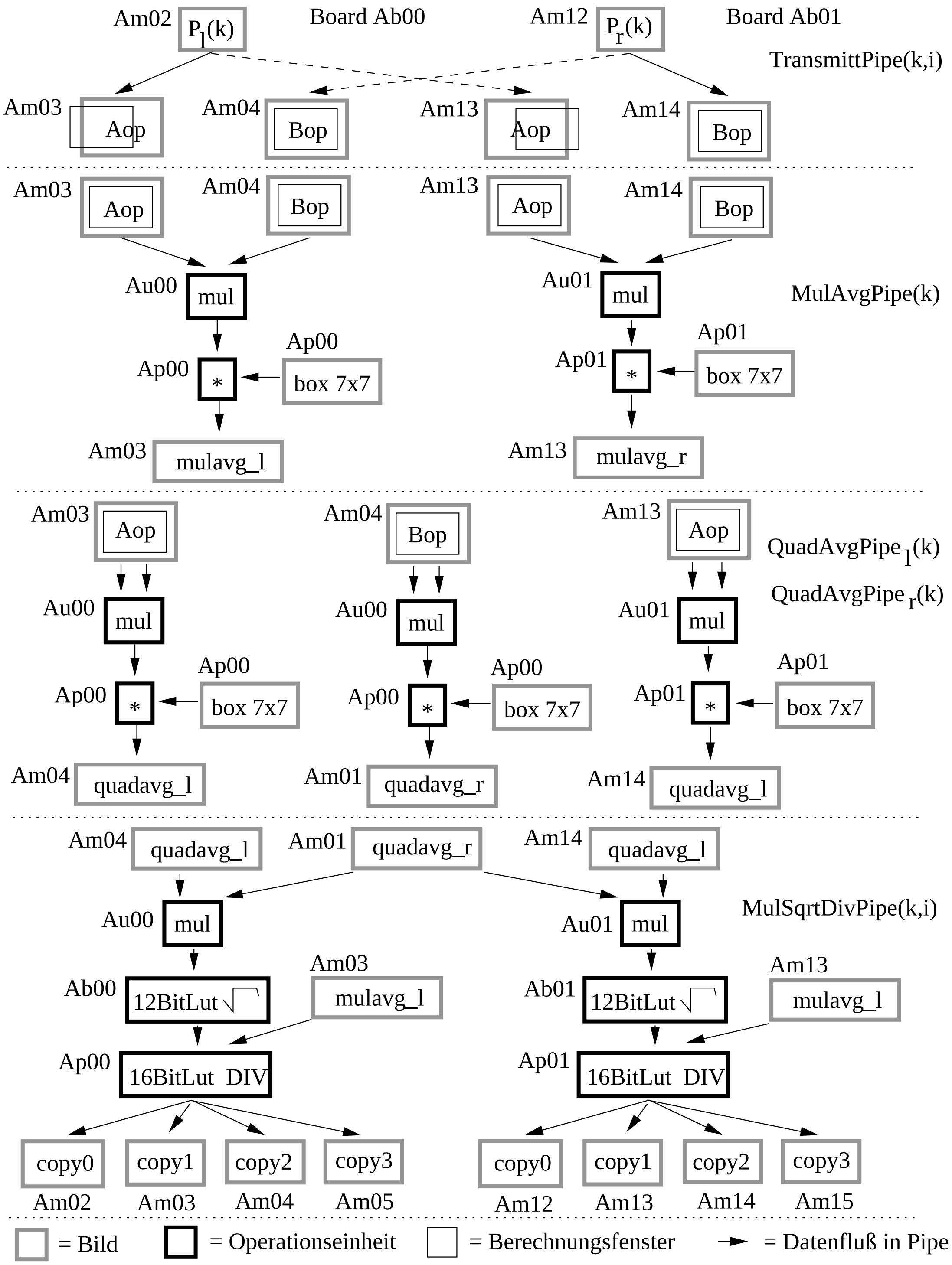


Abbildung 6.8: Die Berechnung der normierten Korrelation auf dem MV200-System bietet ein Beispiel für die Implementierung des MSA-Stereoverfahrens insgesamt. Sie wird parallel für zwei Suchpositionen durch eine vierstufige Sequenz von Pipe-Aufrufen realisiert. Die Berechnungen umfassen jeweils die kompletten Bilder der Pyramidenstufe $P_l(k)$ und $P_r(k)$.

gen $\pm i$.
Bei diesem Transfer wird gleichzeitig die Randbehandlung für die folgende Korrelation durchgeführt. Dies geschieht zum einen dadurch, daß `Aop` und `Bop` einen drei Pixel breiten Rand besitzen, der Nullen beinhaltet. Die in Abschnitt 6.1.1 erläuterte horizontale Randerweiterung entsteht durch Pipetransfers von $P_l(k)$ nach `Aop` in pixelweise steigender Verschiebung, wobei das Bild vom vorausgegangenen Pipetransfer in `Aop` überschrieben wird.
In Pipe `MulAvgPipe` wird daraufhin der Zähler

$$\sum_{m,n=-3}^{3} (P_l(k)(x+m \pm i, y+n) P_r(k)(x+m, y+n))$$

von NKK für alle $(x, y) \in P_r(k)$, $P_l(k)$ berechnet. Der Pixelstrom von `Aop` und `Bop` wird in der arithmetischen Einheit multipliziert, danach mit einem 7x7 Boxfilter im Faltungsmodul gefaltet und gelangt schließlich als Resultat in das Bild `mulavg`. Die Faltung entspricht dabei der Summation über ein 7x7 Template. Zu beachten ist, daß bei der Multiplikation ein Rundungsfehler von 8 Bit entsteht.
In Pipe `QuadAvgPipe` werden die Normierungsterme

$$\sum_{m,n=-3}^{3} P_l(k)(x+m \pm i, y+n)^2 \text{ und } \sum_{m,n=-3}^{3} P_r(k)(x+m, y+n)^2$$

im Nenner von NKK für alle $x, y \in P_r(k)$, $P_l(k)$ ermittelt. Die Berechnung erfolgt analog zur `MulAvgPipe`, mit dem Unterschied, daß das in `Aop` beziehungsweise `Bop` vorliegende Bild $P(k)$ der Pyramidenstufe in der arithmetischen Einheit mit sich selbst multipliziert wird. Das Ergebnis wird in `quadavg` abgelegt.
Die Pipe `MulSqrtDivPipe` bewerkstelligt die Bildung der Division von Zähler (`mulavg`) durch die Wurzel aus dem Produkt der Normierungsterme `quadavg_l`, `quadavg_r`. Da das Ziehen einer Wurzel sowie die Division nicht hardwaremäßig vorgesehen sind, werden diese Operationen durch zwei Look-Up-Tables (LUT) nachgebildet. Die Look-Up-Tables realisieren jeweils eine einstellige Funktion zwischen 12-Bit beziehungsweise 16-Bit Werten. Numerisch stellt die 12-Bit LUT zur Berechnung der Quadratwurzel einen Engpaß dar, es gibt aber keine zweite 16-Bit-LUT auf dem MV200-Board. Da die Division eine zweistellige Funktion ist, wird der 8-Bit Pixelstrom von `mulavg` mit dem 6-Bit Ergebnis der Wurzeloperation zu einem 16-Bit Strom zusammengefaßt, wobei die 6 Bit der Wurzeloperation als höherwertige Bits in einen 8-Bit Strom eingefügt werden. Die 16-Bit-LUT wird so als zweistellige 8-Bit Funktion programmiert.
Das Ergebnis der Division wird dann in vier Bildern parallel abgespeichert. Diese Strategie gilt der Effizienzsteigerung nachfolgender Berechnungsschritte unter Ausnutzung der Speichermodule (siehe Abschnitt 6.1).

Ein Block aus den vier Pipeaufrufen `TransmittPipe(k,i)`, `MulAvgPipe(k)`,

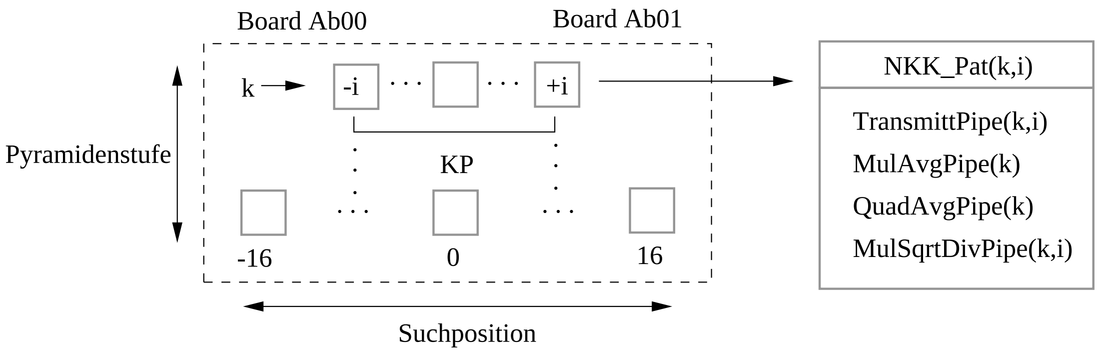


Abbildung 6.9: Eine aus vier Pipeaufrufen bestehende Sequenz zur Berechnung der Korrelationsantworten für die Suchpositionen ± i in der Pyramidenstufe k von KP wird zu einer Pat-Struktur zusammengefaßt.

`QuadAvgPipe`$_l$`(k)` und `MulSqrtDivPipe(k,i)` zur Berechnung von NKK in KP für die Suchpositionen $\pm i$ in der Pyramidenstufe k wird jeweils in einer PAT-Struktur `NKK_Pat(k,i)` zusammengefaßt. Zur Initialisierung der Berechnung einer Pyramidenstufe k von KP wird eine Pat `Init_Pat(k)` gebildet, die aus den Pipeaufrufen `TransmittPipe(k,i)` und `QuadAvgPipe`$_r$`(k)` besteht und den von der Suchposition unabhängigen Normierungsterm `quadavg`$_r$ von $P_r(k)$ einmalig für die jeweilige Pyramidenstufe berechnet.
Die Abfolge der Pipeaufrufe in einer Pat `NKK_Pat(k,i)` kann in Abbildung 6.9 nachvollzogen werden. Dort ist eine Pyramide von Korrelationsergebnissen KP dargestellt, sowie die zur Generierung eines Eintrags KP(k,i) notwendige Sequenz von Pipeaufrufen.
Die Korrelationsantworten KP(k,0) werden doppelt berechnet, was jedoch keine Einbuße an Effizienz darstellt, weil die Berechnungen parallel auf zwei MV200-Boards ablaufen. Die doppelte Berechnung wird aufgrund der in folgenden Verarbeitungsschritten benötigten symmetrischen Aufteilung von KP zur Parallelisierung durchgeführt, um Randprobleme an der Schnittstelle Null im Suchbereich zu verhindern.
Insgesamt ergibt sich KP aus dem sequentiellen Aufruf von `Init_Pat(k)`, `NKK_Pat(k,i)` mit $k = 0 \ldots 3$, $i = 0 \ldots \frac{16}{2^k}$. Die Sequenz von Pat-Aufrufen kann ohne Kontrolle des Host-Computers ausgeführt werden.

6.1.3 CEPSIM auf dem MV200-System

Es sollen die Unterschiede zwischen der Implementierung des CEPSIM-Stereoverfahrens auf einer Workstation sowie der geplanten Implementierung auf dem MV200-System vorgestellt werden.
Die Struktur des CEPSIM-Verfahrens aus Abschnitt 3.4.3 läßt sich in numerischer Hinsicht einfacher auf das MV200-System abbilden als die des

MSA-Stereoverfahrens, da alle benötigten Rechenoperationen in der Hardware des Systems existieren. Dagegen ist das CEPSIM-Verfahren in der MV200-Implementierung von der Kontrolle und Rechenleistung des Host-Computers abhängig, während die MSA-Implementierung aufgrund einer regelmäßigeren Programmstruktur ohne Rechenleistung des Host-Computers auskommt.

Dennoch wurden numerische Modifikationen vorgenommen. Erfahrungen mit der MSA-Implementierung haben gezeigt, daß Rundungsfehler der 8-Bit-Arithmetik des MV200-Systems bei der Berechnung der normierten Korrelation erheblich sind. Die Überschreitung des 8-Bit-Wertebereichs entsteht durch die verwendete Produktbildung. Bei wenig Kontrast im Stereobildpaar liefert die Korrelation keine unterscheidbaren Ergebnisse.
Deshalb wurde die Korrelation des CEPSIM-Verfahrens durch die absolute Differenz von Grauwerten ersetzt. Diese nicht unerhebliche Vereinfachung läßt sich durch die Vermeidung relevanter Rundungsfehler rechtfertigen. Es wird eine unverhältnismäßig effizientere Berechnung des Vergleichsmaßes erreicht, weil sich die absolute Differenz auf dem MV200-System nicht nur erwartungsgemäß schneller, sondern auch eleganter implementieren läßt. Die Parallelkapazität des MV200-Systems wird besser ausgenutzt. Für die einfache Analyse von Vergleichsergebnissen durch die Suchstrategie des CEPSIM-Verfahrens reicht die absolute Differenz aus, weil jeweils nur lokale Maxima in einem festen Einzugsbereich betrachtet werden.
Insgesamt ist zu erwarten, daß das CEPSIM-Verfahren aufgrund der geschilderten effizienteren Implementierbarkeit auf dem MV200-System eine geringfügig höhere Taktrate als das MSA-Stereoverfahren (4 Bilder/Sekunde) erreicht.

6.2 Koppelung von Stereoverfahren und Vergenzregler

In Abschnitt 1.2 wurden in einer aktiven Philosophie Rahmenbedingungen für die Einbettung eines Stereoverfahrens in die Systemkomponente gesteckt. Dort wird eine der Ermittlung einer Disparitätskarte vorausgehende Vergenz der Stereokameras gefordert.
Die Vergenzbewegung umfaßt die Fixation einer dominanten Kamera (L) auf einen zentralen Szenepunkt einer ROI, den ein übergeordneter Prozeß durch Vorgabe einer Blickrichtung bestimmt, im allgemeinen auf einen Punkt mit einem hohen Maß an Strukturinformationen. Die nicht-dominante Kamera (R) führt die Vergenzbewegung aus.
Diese einseitige Vergenz stellt eine Anpassung an das zugrunde liegende MSA-Stereoverfahren dar, das sowohl die Vergenz steuert als auch im Zustand der Vergenz Disparitätskarten mit Bezug auf die dominante Kamera liefert.
Implementiert wird die Vergenzsteuerung durch einen Regler auf dem Host-

Computer, sowie durch das MSA-Stereoverfahren, das parallel auf dem MV200-System läuft. Das Stereoverfahren liefert den Vergenzfehler für den Regler. Diese Lösung ist eleganter als die Auswahl eines separaten Vergenzreglers aus einer Vielzahl von Alternativen in der Literatur. Der unten beschriebene Regelungsprozeß ist in Abbildung 6.10 zusammengefaßt.

Der Vergenzregler findet als Ausgangsposition die linke Kamera in der festen Vergenz $\Phi_l = 0$ und die rechte Kamera in einem Vergenzwinkel $\Phi_{r\_init} > 0$ vor. In diesem Zustand können Disparitäten in (L,R) auftreten, die bei Objekten im Nahbereich der Kameras den angenommenen Suchbereich des MSA-Stereoverfahrens (± 16 Pixel) überschreiten. Deutlich wird an dieser Stelle, daß der Suchbereich beschränkt sein muß, weil er einen Kompromiß aus detektierbaren Disparitäten und einem Rand im Stereobildpaar darstellt. Große Disparitäten lassen sich nur in einer zentralen Bildregion detektieren.
Aus dieser Überlegung heraus wird der Regler in zwei separate Phasen unterteilt. In der ersten Phase wird der Vergenzwinkel Φ_r auf Grundlage eines globalen Vergleichs zwischen L und R grob auf einen Startwinkel Φ_{r*} für eine genaue Vergenz geregelt, indem die rechte Kamera des Kopfes unter Veränderung von Φ_r aktiv über einen weitaus größeren Suchbereich $\Phi_{r\_init} \pm \triangle\Phi$ als den des Stereoverfahrens geschwenkt wird. Für den vorliegenden Kamerakopf Bisight ermöglicht der gewählte aktive Suchbereich eine grobe Vergenz auf Szenestrukturen, die mindestens 1 Meter von den Kameras entfernt sind. Für jede Suchposition wird eine globale Bewertung der Vergenz durch den Vergleich des zugehörigen Stereobildpaares (L,R) ermittelt. Die Suchposition Φ_{r*} mit optimaler Bewertung wird angesteuert. In diesem Zustand liegen die in der ROI auftretenden Disparitäten im Suchbereich des MSA-Stereoverfahrens. L und R überlappen sich ausreichend für eine statische Korrespondenzsuche.
Ausgehend von dem Startwert Φ_{r*} wird in einer zweiten Regelungsphase die Vergenz durch ein lokales Kriterium bezüglich der ROI von L auf einen exakten Vergenzwinkel Φ_{r_opt} geregelt, so daß im optimalen Fall linke und rechte Kamera genau auf einen Szenepunkt innerhalb der ROI fixieren (Abbildung 6.10).
Diese zwei Phasen der groben und feinen Vergenz stellen eine Coarse-To-Fine-Strategie bezüglich des Einzugsbereichs der Messung des Vergenzfehlers dar, so daß Alias-Fehler bei der Disparitätsmessung durch das MSA-Stereoverfahren unterdrückt werden. Schließlich wird die Kontrolle an einen übergeordneten Prozeß zurückgegeben.

Beide Reglerphasen messen den Vergenzfehler im Host-Computer auf Basis von Ausgaben des parallel laufenden MSA-Stereoverfahrens auf dem MV200-System. Das MSA-Stereoverfahren liefert einige zur Messung des Vergenzfehlers geeignete Ausgaben, die ohne Beeinflussung des Stereoverfahrens durch den Host-Computer abrufbar sind.
Zum einen die Disparitätskarte, deren globale Disparität im Zustand der Vergenz minimal sein sollte. Zweitens das Vertrauensmaß, das getroffene Zuordnungen

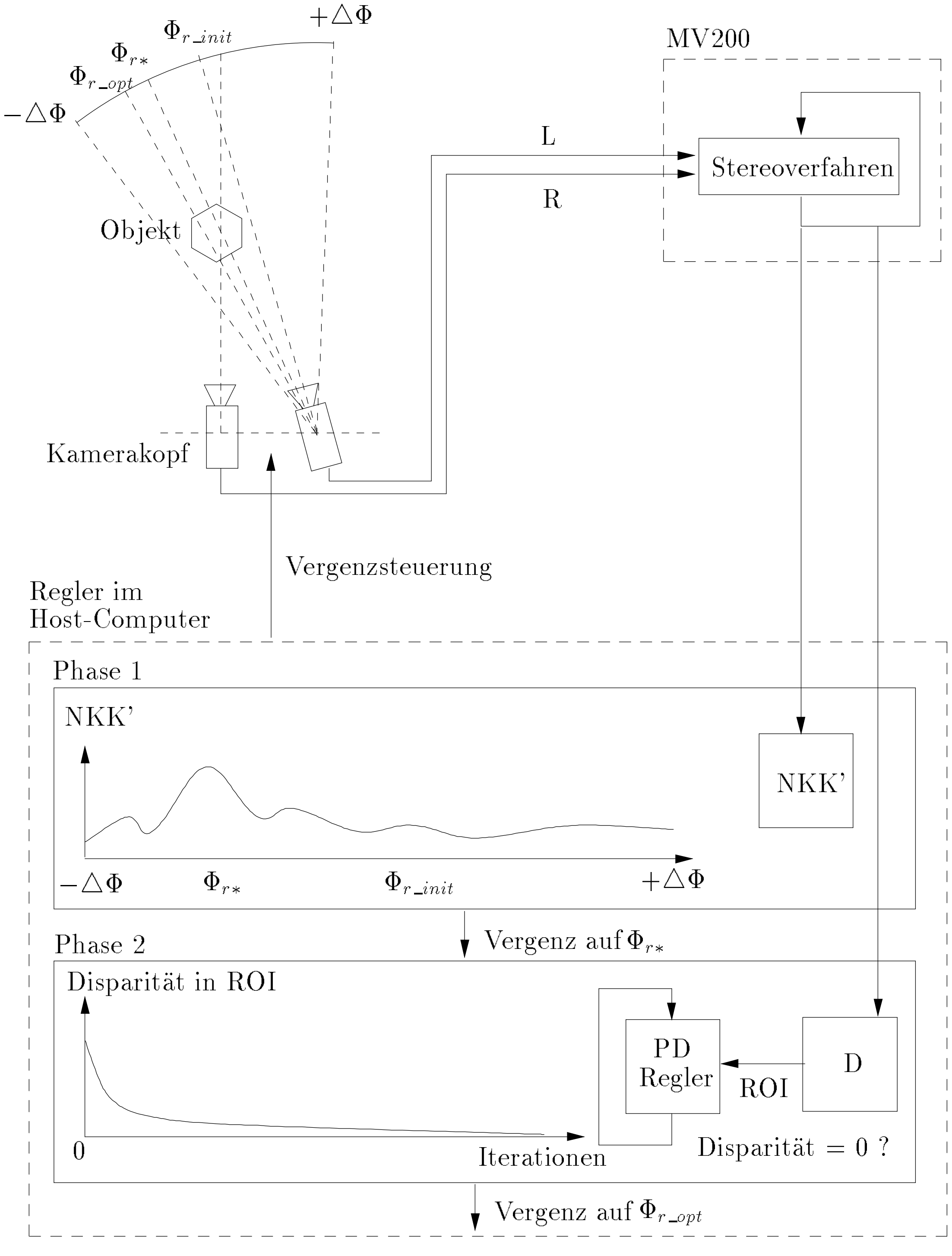


Abbildung 6.10: Die Vergenzsteuerung des Kamerakopfes erfolgt in zwei Phasen. In Phase 1 wird die rechte Kamera aktiv über einen Suchbereich $\Phi_{r\_init} \pm \triangle\Phi$ bewegt, wobei der Vergenzfehler durch globale Korrelation grob auf Φ_{r*} herabgeregelt wird. In einer zweiten Phase führt ein PD-Regler eine feine Vergenzregelung auf Basis der durch das Stereoverfahren gelieferten Disparitätskarten auf Φ_{r_opt} durch.

bewertet. Im Zustand der Vergenz lassen sich theoretisch die Pixel in L mit einer maximalen Konfidenz zuordnen, die globale Summe des Vertrauensmaßes wäre demnach maximal. Als dritte Möglichkeit bieten sich direkte Vergleichsergebnisse zwischen L und R im Suchbereich des Stereoverfahrens an, die durch Korrelation ermittelt werden.
Für die erste Reglerphase zur Bestimmung eines globalen Vergenzfehlers wurde die globale Korrelation zwischen L und R als bewährtes photometrisches Vergleichsmaß ausgewählt. Die schwache Besetzung von Disparitätskarte und Vertrauensmaß schwankt im aktiven Suchbereich, was ihre globale Verwendung als Vergenzmaß erschwert. Eine geeignete Normierung auf die Besetzung der jeweiligen Datenstruktur wäre erforderlich.
Zum Transfer der globalen Korrelationsergebnisse des Stereoverfahrens auf den Host-Computer bietet Imageflow 2.5 spezielle ereignisgesteuerte Mechanismen.
Da zu Beginn der zweiten Reglerphase die rechte Kamera im Vergenzwinkel Φ_{r*} grob vergiert hat, kann der im Suchbereich des Stereoverfahrens liegende Vergenzfehler verläßlich einer Disparitätskarte D von MSA entnommen werden. Nur der Vergenzfehler in einer zentralen ROI der dominanten Kamera ist relevant, da nach erfolgter Vergenz hier die Disparität auf Null geregelt sein soll. Der ROI wird also in einer zentralen 6x6 Umgebung der Disparitätskarte der Vergenzfehler der zweiten Reglerphase als medianer Disparitätswert entnommen.
Ein PD-Regler regelt in einem iterativen Prozeß der Fehlermessung und der Veränderung des Vergenzwinkels ausgehend von Φ_{r*} zu Φ_{r_opt} die Disparität der ROI auf Null, was gleichzeitig das Abbruchkriterium der zweiten Regelungsphase darstellt. Die Herabregelung der Disparität verläuft bei einem PD-Regler zeitlich idealerweise ohne Überschwinger (siehe Abbildung 6.10).
So zeichnet sich ein genaues algorithmisches Bild der gesamten Vergenzregelung ab.

Zu bemerken ist, daß die aktive Schwenkung der nicht dominanten Kamera in der ersten Reglerphase durch einen beispielsweise fokusbasierten separaten Vergenzregler beschleunigt werden kann.
Man kann sich eine andere aktive Strategie der Vergenzregelung überlegen, die auf nur einer einzigen homogenen Reglerphase beruht. Die Idee ist, den Suchbereich des jeweiligen Stereoverfahrens aktiv auf das notwendige Maß zu erweitern. Ermittelt werden dann nicht Korrelationsergebnisse zwischen festen Stereobildern durch Verschiebungen der Bildinhalte, so daß Randprobleme entstehen, sondern aus und einer durch aktive Bewegung der nicht dominanten Kamera gewonnenen Folge von Stereobildpaaren. Ein Vergenzfehler ergibt sich dann allein aus der nun verläßlichen Disparitätskarte des Stereoverfahrens.

Kapitel 7

Diskussion und Ausblick

7.1 Diskussion der Stereoverfahren

Das Ziel der Entwicklung der gesuchten Systemkomponente wird durch die Koppelung des MSA-Stereoverfahrens an einen Vergenzregler erreicht.
Als Kernpunkt der Systemkomponente wurden Stereoverfahren basierend auf dem Cepstrum-Filter und der Korrelation auf ihre Eignung hin untersucht.
Zu diesem Zweck wurden Standardansätze der Literatur sowie aufbauende Weiterentwicklungen implementiert, namentlich die Stereoverfahren GREEDY, YESA, MSA und CEPSIM.
Der allgemeine und qualitative Vergleich der Stereoverfahren auf der Basis von Testbildern zeigt Tendenzen in der Leistungsfähigkeit auf. Es bleibt eine Abhängigkeit der Vergleichsergebnisse von den ausgewählten Testbildern. In den besagten Tendenzen zeichnet sich jedoch eine Rangfolge ab, die grundsätzlich an die Komplexität der den Stereoverfahren zugrunde liegenden Algorithmen gebunden ist.

Die schlechtesten Disparitätsmessungen liefert das YESA-Stereoverfahren. Erklären läßt sich diese Tatsache aus einer einfachen, restriktiven Modellierung des Disparitätsproblems sowie einer ungeeigneten Analysestrategie von Antworten des Cepstrum-Filters, die zudem keine Skalensuche beinhaltet.
Restriktive Modellannahmen betreffen eine konstante Disparität im Einzugsbereich des Vergleichsmaßes, zudem macht sich die Koppelung von Einzugsbereich und Suchbereich des Cepstrum-Filters in dieser Implementierung negativ bemerkbar. Systematische Schwächen in der verwendeten Analyse von Cepstrum-Antworten fallen insbesondere durch die relativ kleinen Fenstergrößen (Kacheln) stärker ins Gewicht. Weitere Fehler in Disparitätsmessungen entstehen durch Alias-Probleme.
Insofern sind die teilweise falschen Zuordnungen des Stereoverfahrens zu erwarten.
Die Berechnung des Cepstrum-Filters für jedes Pixel eines Stereobildpaars macht

den Algorithmus extrem rechenaufwendig. An dieser Stelle muß bemerkt werden, daß bei der Entwicklung des YESA-Stereoverfahrens das Hauptgewicht auf der Modellierung eines angenommenen biologischen Sachverhaltes im menschlichen visuellen System lag. Für eine Praxisanwendung als Stereoverfahren erweist sich der Algorithmus als eher ungeeignet.

Das CEPSIM-Stereoverfahren analysiert Antworten des Cepstrum-Filters in einer stabileren Strategie. Dabei rechtfertigt sich die Verwendung der zum Cepstrum-Filter analogen globalen Korrelation nicht nur aus der Notwendigkeit heraus, das Vergleichsmaß auf dem MV200-System implementieren zu können, sondern auch aus der vollständigen Nachvollziehbarkeit der Ergebnisse der Korrelation als Vergleichsmaß gegenüber dem Cepstrum-Filter, der in Sonderfällen systematische Fehler produziert.
Die Suchstrategie des CEPSIM-Verfahrens ermöglicht die Detektion kontinuierlicher Disparitätsverläufe. Dies stellt einen Fortschritt gegenüber Analysestrategien der Literatur dar. Im Gegensatz zum YESA-Verfahren existiert eine heuristische Skalensuche.
Das CEPSIM-Stereoverfahren entwickelt insofern elementare Ansätze der Cepstrum-Technik weiter und erfüllt strukturelle Anforderungen für die Implementierung im MV200-System. Die Laufzeit des CEPSIM-Verfahrens liegt in einem Zeitintervall, weil sie adaptiv von den im jeweiligen Stereobildpaar vorliegenden Disparitätsinformationen abhängt.
Im Vergleich zum YESA-Stereoverfahren erweist sich das CEPSIM-Stereoverfahren als robuster und genauer (siehe Abschnitt 5.2).

Das GREEDY-Stereoverfahren stellt einen Standardansatz der Literatur dar. Die Hauptprobleme der Suchstrategie bilden Blocking-Effekte in den ermittelten Disparitätskarten, die durch die Top-Down-Suche in einer Pyramide entstehen. Dies drückt sich deutlich in den qualitativen Vergleichsergebnissen aus.

Das MSA-Stereoverfahren wurde aus Grundprinzipien des GREEDY-Stereoverfahrens heraus entwickelt. Mängel in der Suchstrategie stellen den Ansatzpunkt zur Verbesserung des Verfahrens dar. Die im MSA-Verfahren verwendete vollständige Analyse von Korrelationsergebnissen in den Skalen betrachtet alle möglichen Zuordnungen im Stereobildpaar und wird dem Disparitätsproblem besser gerecht als die GREEDY-Suchstrategie, die nur an eine Top-Down-Suche gebundene Zuordnungen betrachtet. Das MSA-Stereoverfahren zeigt im Vergleich zu allen anderen Verfahren das stabilste Verhalten. Der Algorithmus beinhaltet die aufwendigste Modellierung des Disparitätsproblems, die im Gegensatz zum CEPSIM-Verfahren keine heuristischen Elemente besitzt. Die qualitativen Unterschiede zum CEPSIM-Verfahren schlagen sich in geringfügig höherem Rechenaufwand nieder.
Der absolut gesehen hohe Rechen- und Speicheraufwand des MSA-Stereoverfahrens läßt sich durch die Kapazität des MV200-Systems rechtfertigen.

Allgemein kann mangelnde Rechenleistung in der Vergangenheit als Ursache für die Entstehung einfacher Suchstrategien wie dem GREEDY-Verfahren als Standardansätze der Literatur angesehen werden, weil Stereoverfahren wie oben erwähnt einen Kompromiß zwischen Güte der Lösung und Laufzeit des Algorithmus darstellen.
Die Forderung nach Echtzeitfähigkeit gemäß obiger Definition (siehe 1.2) wird durch die Stereoverfahren MSA und CEPSIM in gleicher Weise erfüllt. Mit dem MV200-System erreichen die Algorithmen in der gegenwärtigen Implementierung eine Taktrate von vier Bildern pro Sekunde.

Aus der Untersuchung von Cepstrum-Filter und Korrelation als Vergleichsmaß läßt sich das allgemeine Fazit ziehen, daß der Cepstrum-Filter besser zur Vergenzregelung geeignet scheint als zur Disparitätsmessung in einem Stereoverfahren, da die theoretische Interpretation der Filterantwort nur die Detektion einer einzigen Disparität zuläßt. Die Detektion mehrerer Disparitäten in der Antwort des Cepstrum-Filters kann nur auf heuristischer Basis erfolgen, wobei die Interpretation als Autokorrelation ein Hilfsmittel darstellt. Systematische Schwächen des Cepstrum-Filters erschweren die Analyse.
Ansonsten zeigt sich der Cepstrum-Filter als theoretisch und praktisch eng verwandt mit Korrelationsmaßen, die die erwähnten systematischen Schwächen vermeiden, jedoch ein schlechteres Signal-Rausch-Verhältnis als der Cepstrum-Filter besitzen.
In einem Stereoverfahren spielen der Cepstrum-Filter und die Korrelation lediglich die Rolle eines photometrischen Vergleichsmaßes. Es zeigt sich, daß die Komplexität von Algorithmen zum Auffinden von Pixelkorrespondenzen, die von der verwendeten Analysestrategie der jeweiligen Vergleichsergebnisse abhängt, einen wichtigeren Einfluß auf die Güte der gemessenen Disparitäten besitzt als die Art des verwendeten Vergleichsmaßes. Zu diesem Schluß kommt auch [Olson88].
Eine qualitative Einstufung von Vergleichsmaßen wie dem Cepstrum-Filter oder Korrelationsmaßen, wie er zum Teil in der Literatur zu finden ist, ist demnach kritisch zu bewerten, weil es hauptsächlich auf eine dem Vergleichsmaß angepaßte Analysestrategie der Ergebnisse ankommt. Die Funktion von photometrischen Vergleichsmaßen ist grundsätzlich austauschbar. Vergleich und Einteilung von Stereoverfahren sollten von der Verwendung eines bestimmten Vergleichsmaßes entkoppelt werden.

7.2 Ausblick

Die bestehenden Stereoverfahren CEPSIM und MSA lassen sich im Hinblick auf die in Abschnitt 1.2 gesteckten Ziele weiterentwickeln.

Für das aktuell verwendete Stereoverfahren MSA besteht die Möglichkeit, das Pyramidenkonzept für eine Verbesserung der Echtzeitfähigkeit zu nutzen. Ein

aktives Konzept verlangt von einem Stereoverfahren nicht Disparitätskarten konstanter Güte, sondern von einer Qualität, die aufgabenspezifisch ist. So wäre es in einfacher Weise möglich, die Lage der niedrigsten Pyramidenstufe des MSA-Stereoverfahrens beziehungsweise die maximale Auflösungsstufe der Eingabebilder variabel zu gestalten. Dadurch wäre das MSA-Verfahren in der Lage, eine höhere mittlere Taktrate zu erzielen.
Durch Koppelung des CEPSIM-Stereoverfahrens an eine Pyramidenverarbeitung wäre eine analoge Möglichkeit zur Variation von Genauigkeit der gelieferten Disparitätskarten und Laufzeit des Algorithmus gegeben.
Eine weitere Aufgabe, die das CEPSIM-Verfahren betrifft, liegt in der Klärung der Frage, ob sich die heuristische Suchstrategie des Algorithmus durch eine effiziente, exakte Suchstrategie ersetzen läßt, ohne das algorithmische Grundprinzip der Cepstrum-Technik fallenzulassen. Eine exakte Suchstrategie sollte das fundamentale Prinzip der Cepstrum-Technik in der Dominanzreihenfolge von Bildstrukturen (Folge von Cepstrum-Impulsen) im Stereobildpaar weiterverfolgen. Der Algorithmus CEPSIM zeigt, daß der Dominanzbegriff von Struktur einen eleganten, effizienten Zugang zur Vermeidung von Alias-Problemen bei der Korrespondenzsuche darstellt.

Aus dieser Diplomarbeit soll ebenfalls ein Ausblick auf die allgemeine Weiterentwicklung von Stereoverfahren gegeben werden, wie er sich aus der Entwicklung und Untersuchung von eigenen Verfahren und Verfahren der Literatur ergibt.
Momentan liefern aktuelle Algorithmen der Optimierungstheorie hinsichtlich Qualität und Dichte der gewonnenen Disparitätsinformation die besten Ergebnisse.
Die Ursache liegt einfach darin, daß im Zuordnungsprozeß von Pixeln keine isolierten Zuordnungen getroffen werden, sondern Zuordnungen, die globale Informationen über die Zuordnung von anderen Pixeln in Wechselwirkung berücksichtigen. Es fließen so neue Nebenbedingungen in den Optimierungsprozeß der Korrespondenzfindung ein, die die Optimierung vereinfachen.
Stochastische Optimierungsalgorithmen wie Simulated-Annealing besitzen zwar den Nachteil, nichtdeterministisch und bezüglich ihrer Eingabeparameter schlecht konditioniert zu sein, sie verdienen aber eine weitere Erforschung.
Bei der Anpassung dieser Verfahren an das Disparitätsproblem besteht in der Literatur das Bestreben, den Nichtdeterminismus zu beseitigen ([HeiPeBou92]). Besonders hinsichtlich einer angestrebten Echtzeitfähigkeit ist Determinismus zwingend notwendig. Die sehr hohe Laufzeit der angesprochenen Optimierungsverfahren fällt angesichts einer dynamischen Entwicklung der Kapazität von Bildverarbeitungssystemen zukünftig weniger ins Gewicht.
Aber auch andere aktuelle Ansätze der Literatur betrachten die Zuordnung von Pixeln in einem Zusammenhang ([He97]), ohne sich auf die Optimierungstheorie zu stützen, die durch ihren iterativen Prozeß als nicht biologisch motiviert kritisiert wird.

In aktiven Stereosystemen, bei denen Echtzeitfähigkeit das oberste Gebot ist, stellen phasenbasierte Stereoverfahren den Stand der Kunst dar. Die Subpixelgenauigkeit verbunden mit einer hohen Taktrate der Disparitätsmessung (bis zu 25 Hz) lassen in verhaltenbasierten Systemen eine aktive Generierung von Disparitätskarten zu. Die Kombination von hierarchischem, phasenbasiertem Stereo mit einer Regelung der Blickrichtung und der Vergenz des aktiven Systems führen zu einer robusten Lösung des Zuordnungsproblems ([Ha96]).

Literaturverzeichnis

[Ba89] Stephen T. Barnard. *Stochastic Stereo Matching over Scale.* International Journal of Computer Vision, 3. 17-32 (1989), Kluwer Academic Publishers, 1989.

[Ching93] Wee-Son Ching, Peng-Seng Toh, Kap-Luk Chan and Meng-Hwa Er. *Robust Vergence with Concurrent Detection of Occlusion and Specular Highlights.* Fourth International Conference on Computer Vision, IEEE Computer Society Press, 1993.

[Fle94] D.J. Fleet, A.D. Jepson, M.R.M. Jenkin. *Phased-Based Disparity Measurement.* CVGIP: Image Understanding, Vol.53, No.3, March, pp.343-358, Academic Press, 1994.

[Fua93] Pascal Fua. *A parallel stereo algorithm that produces dense depth maps and preserves image features.* Machine Vision and Applications 1993, 6:35-49, Springer Verlag, 1993.

[Gu90] Chuang Gu, Li-de Wu. *Structural Matching of Multiresolution for Stereo Vision.* 10th international Conference on Pattern Recognition, IEEE Computer Society Press, 1990.

[Ha96] M. Hansen, G. Sommer. *Active depth estimation with gaze and vergence control using gabor filters* . In Proc. 13th ICPR, Vienna, Vol. A, 344-348, 1996.

[He97] Rolf Henkel. *Stereovision by Coherence Detection.* Proceedings CAIP 97 to appear, Springer Verlag.

[HeiPeBou92] F. Heitz, P. Perez, P. Bouthemy. *Multiscale Minimization of Global Energy Functions in Some Visual Recovery Problems.* CVGIP: Image Understanding, Vol.59, No. 1, January, pp.125-134, Academic Press, 1994.

[IDUC94] *International Datacube User's Conference.* Dublin, Ireland, October 11-13, 1994. Österreichisches Forschungszentrum Seibersdorf G.m.b.H. 1994.

[IntBo93] Stephen S. Intille, Aaron F. Bobick. *Disparity Space Images and Large Occlusion Stereo.* Computer Vision ECCV 94, Springer Verlag, 1994.

[JeJeTso89] Michael R. M. Jenkin, Allen D. Jepson, John K. Tsotsos. *Techniques for Disparity Measurement.* Computer Vision and Image Understanding, Academic Press 1991.

[Knie91] T. Knieriemen. *Autonome mobile Roboter.* BI-Wissenschaftsverlag, 1991.

[KleKoSchlü96] Reinhard Klette, Andreas Koschan, Karsten Schlüns. *Computer Vision, räumliche Informationen aus digitalen Bildern.* Vieweg & Sohn Verlagsgesellschaft mbH, 1996.

[Kosch93] Andreas Koschan. *What is New in Computational Stereo Since 1989: A Survey on Current Stereo Papers.* Technische Universität Berlin, Technischer Bericht 93-22, August 1993.

[Lu91] Kai-Oliver Ludwig. *Untersuchung der Cepstrumtechnik zur Querdisparitätsbestimmung für die Tiefenschätzung bei fixierenden Stereokonfigurationen.* Diplomarbeit, Fachbereich Informatik, Universität Hamburg, 1991.

[LuBu93] A. Luo, H. Burkhardt. *An Intensity-Based Cooperative Bidirectional Stereo Matching with Simultaneous Detection of Discontinuities and Occlusions.* International Journal of Computer Vision, 15, 171-188 (1995), Kluwer Academic Publishers, 1995.

[OkKa92] Masatoshi Okutomi, Takeo Kanade. *A Locally Adaptive Window for Signal Matching.* International Journal of Computer Vision, 7:2, 143-162(1992), Kluwer Academic Publishers, 1992.

[OlCo91] Thomas J. Olson, David J. Coombs. *Real-Time Vergence Control for Binocular Robots.* International Journal of Computer Vision, 7:1, 67-89(1991), Kluwer Academic Publishers, 1991.

[Olson88] Thomas J. Olson, Robert D. Potter. *Real-Time Vergence Control.* University of Rochester, Computer Science Department, Technical Report 264, November 1988.

[OlTaLo95] Thomas J. Olson, John R. Taylor, Robert J. Lockwood. *Programming a Pipelined Image Processor.* Computer Vision and Image Understanding, Volume 64, Number 3, Academic Press, 1996.

[Ov92] Ian Overington. *Computer Vision.* Elsevier Science Publishers B.V., 1992.

[Sa88] T.D. Sanger. *Stereo disparity computation using Gabor filters.* Biol.Cybernetics 59, 405-418, 1988.

[Sta89] J.P.Pascal Starink, Eric Backer. *Finding Point Correspondences Using Simulated Annealing.* Pattern Recognition, Vol.28, No.2, pp 231-240, Elsevier Science Ltd, 1995.

[TaOlMa94] John Taylor, Thomas Olson, W. N. Martin. *Accurate Vergence Control in Complex Scenes.* Proceedings 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, 1994.

[Vlee91] D. De Vleeschauwer. *An Intensity-Based, Coarse-to-Fine Approach to Reliably Measure Binocular Disparity.* CVGIP: Image Understanding Vol.57, No. 2. March, pp.204-218, Academic Press, 1993.

[Weng92] Juyang Weng, Narendra Ahuja, Thomas S. Huang. *Matching Two Perspective Views.* IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL 14, NO. 8, AUGUST 1992, IEEE Computer Society Press, 1992.

[Xie93] Ming Xie. *Stereo and Motion Matching: a Hough-Transform inspired Method.* Pattern Recognition Letters 15 1143-1150, Elsevier Science Ltd, 1994.

[Ye89] Yehezkel Yeshurun, Eric L. Schwartz. *Cepstral Filtering on a Columnar Image Architecture: A Fast Algorithm for Binocular Stereo Segmentation.* Transactions on Pattern Analysis and Machine Intelligence, VOL. 11, NO.7, July 1989, IEEE Computer Society Press, 1989.

Danksagung:

Hiermit möchte ich meinem Betreuer Dipl. Phys. Michael Hansen sowie meinen Kollegen für die hilfsbereite und geduldige Unterstützung bei der Erstellung dieser Diplomarbeit danken. Mein weiterer Dank gilt Anne-Kathrin Häfner für das Korrekturlesen dieser Arbeit.

Erklärung:

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig verfaßt und ausschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Kiel, den