

3D–Linienraster für die optische Formaufzeichnung

Diplomarbeit

von

Tobias Haist

Hauptberichter: Prof. Dr. H.J. Tiziani

Mitberichter: Prof. Dr. H.J. Bauer

Institut für Technische Optik

der

Universität Stuttgart

Juli 1996

Inhaltsverzeichnis

1	Einleitung und Motivation	6
1.1	Motivation	7
2	Beugung	10
2.1	Fresnel-Beugung	10
2.2	Fraunhofer-Beugung und Fourier-Hologramme	12
3	Fourier-Transformationen	13
3.1	Fourier-Transformation in der Optik	13
3.2	Diskrete Fourier-Transformation	15
4	Holographie	18
4.1	Grundprinzip	18
4.2	Einteilung von Hologrammen	19
4.2.1	Orts-Bandbreiten-Produkt	20
4.2.2	Fresnel-Hologramme	20
4.2.3	Fourier-Hologramme	21
5	Computergenerierte Hologramme	25
5.1	Vom Hologramm zum computergenerierten Hologramm	25
5.2	Einteilung computergenerierter Hologramme	28
5.3	Beurteilung computergenerierter Hologramme	31
5.3.1	Rekonstruktionsfehler	32
5.3.2	Beugungseffizienz	33
5.3.3	Testbilder und Wahl der Verfahren	34
5.4	Überblick CGH-Berechnungen	36

6	CGH–Kodierungen	38
6.1	Weitere punktorientierte Kodierungen	38
6.1.1	Kodierung nach Burch	38
6.1.2	Kodierung nach Huang und Prasada	39
6.2	Detour–Phase Hologramme	39
6.2.1	Lohmann–Kodierung	40
6.2.2	Lee–Kodierung	40
6.2.3	Kodierung nach Haskell	41
7	Diffusoren	42
7.1	Speckles	43
7.1.1	Auswertung der Rekonstruktion an diskreten Stellen	43
7.1.2	Parkettierung	44
7.1.3	Trigonometrische Interpolation und das Einschieben von Nullen	45
7.1.4	Statistische Mittelung	46
7.1.5	Simulation der Speckles	46
7.2	Random–Diffusor	47
7.3	Radar–Codes und Diffusoren	49
7.4	Linsen–Diffusor	50
8	Iterative–Fourier–Transformations–Algorithmen	51
8.1	Gerchberg–Saxton–Algorithmus	51
8.1.1	Anzahl der Iterationen	53
8.1.2	Optische Implementierung	54
8.1.3	Binarisierung	56
8.1.4	Fensterung im IFT–Algorithmus	58
8.1.5	IFT–Algorithmus mit Bandbreitenkontrolle	59
8.1.6	IFT–Algorithmus mit Amplituden–Phasen–Koppelung	60
8.1.7	Dreidimensionaler IFT–Algorithmus	61
9	Input–Output und Global–Iterative–Coding	67
9.1	Input–Output–Algorithmus	67

9.2	Global-Iterative-Coding	69
10	Direkte Suchalgorithmen	71
10.1	Direct-Binary-Search (DBS)	72
10.1.1	Grundidee	72
10.1.2	Fast-DBS-Algorithmen	74
10.1.3	Variationen	77
10.1.4	Beurteilung des DBS-Algorithmus	78
10.2	Simulated Annealing (SA) und Threshold Accepting (TA)	79
11	Hologramm-Quantisierung	82
11.1	PWM und PDM	83
11.2	Pulsbreitenmodulation mit variabler Schwelle	83
11.3	Nutzung des Babinet-Effekts	84
11.4	Interlacing und Iterative Interlacing	85
11.5	Neuronale Netze zur Binarisierung	86
11.6	Error-Diffusion	88
11.6.1	Komplexe Error-Diffusion	90
12	Weitere Verfahren zur CGH-Berechnung	91
12.1	Genetische Algorithmen	91
12.2	Ping-Pong-Propagation und andere einfache 3D-Verfahren	92
12.3	Analytische Verfahren	93
12.4	Superpositionen	94
12.5	Ray-Tracing	94
12.6	Brechung der Inversions-Symmetrie	94
12.6.1	Zusätzliches festes Filter	94
12.7	Nutzung von Symmetrien	95
12.7.1	Rotationssymmetrische Objekte	96
12.7.2	Separierung und eindimensionale Objekte (Streifen)	97
13	Phasenschiebung in doppelbrechenden Medien	98
13.1	Die Poincaré-Kugel	98

13.2 Jones–Darstellung	100
13.3 Pancharatnams Phase	101
13.4 Pancharatnams Phase und Phasenmessungen am LCD	102
13.5 Vergrößerung des Phasenschiebereichs bei LCDs	103
14 TN–LCDs als phasenschiebende, optische Elemente	108
14.1 Typischer technischer Aufbau eines TN–LCDs	108
14.2 LCD–Adressierungstechniken	110
14.2.1 Elektrische Adressierung	110
14.2.2 Optische Adressierung	111
14.3 Konkurrierende Lichtmodulatoren	112
14.4 Experimentelle Ergebnisse	113
14.4.1 Eigenschaften des VGA1–LCDs von CRL	114
14.4.2 Optische Rekonstruktion	119
14.4.3 Auswahl von LCDs als Hologrammedien	120
15 Computer, Rechenzeiten, Benchmarks	123
15.1 Grundsätzliche Beurteilung	123
15.2 Benchmarks	124
15.3 Fast–Fourier–Transformation	125
15.3.1 Beschleunigung der FFT	127
16 Beurteilung der verschiedenen Verfahren	132
17 Zusammenfassung und Ausblick	136
A Grundlagen der Beugungstheorie	138
A.1 Wellengleichung	138
A.2 Kirchhoff–Integral–Theorem	139
A.3 Fresnel–Kirchhoff–Formel	141
B Fourier–Transformation	143
C Abkürzungen und Symbole	145

Kapitel 1

Einleitung und Motivation

Ziel der vorliegenden Arbeit ist die Untersuchung des möglichen Einsatzes von Flüssigkristall-Anzeigen (LCDs) als diffraktive optische Elemente. Dabei werden folgende Schwerpunkte gesetzt:

- Implementierung und Untersuchung verschiedener Algorithmen zur Berechnung computergenerierter Hologramme (CGHs). Die Algorithmen sind insbesondere hinsichtlich der Verwendung von LCDs als Hologrammedien zu beurteilen.
- Erweiterung bestehender CGH-Algorithmen auf drei Dimensionen und Berücksichtigung besonderer LCD-Eigenschaften.
- Untersuchung praktischer Probleme beim Einsatz von LCDs als Medien für computergenerierte Hologramme.

Die Ergebnisse dieser Arbeit sollen als theoretischer und praktischer Ausgangspunkt für den Einsatz computergenerierter Hologramme zur optischen Formaufzeichnung dreidimensionaler Objekte dienen.

Wir erhoffen uns in nächster Zukunft eine deutliche Verbesserung der optischen Eigenschaften und Auflösungen von LCDs und einen weiteren Preisverfall dieser Elemente. In jedem Fall gehen wir davon aus, daß — angeregt durch weitere Fortschritte in der Computertechnik — kleine hochauflösende Lichtmodulatoren (*SLMs* = “*Spatial Light Modulators*”), seien es nun LCDs oder die in Abschnitt 14.3 besprochenen Elemente, deutlich leistungsfähiger und preiswerter werden, und so klassische Optiken immer stärker verdrängen bzw. unterstützen.

Die theoretischen Kernpunkte dieser Arbeit lassen sich direkt für den Einsatz konkurrierender Lichtmodulatoren übernehmen.

Im Laufe dieser Arbeit wird sich zeigen, daß die Möglichkeiten zur Berechnung computergenerierter Hologramme sehr vielschichtig sind. Es existieren sehr viele verschiedene Algorithmen, die untereinander kombiniert werden können und in denen Parameter frei gewählt werden können bzw. müssen. Den “besten” Algorithmus gibt es nicht. Je nach Anwendung ist sorgfältig zu untersuchen, wie die Hologramme berechnet werden sollen. Selbst auf eng umrissenen Teilgebieten werden in der Praxis sehr verschiedene Lösungsansätze eingesetzt und auf jedem Teilgebiete existieren Fachleute, die sich speziell mit den dort anwendbaren Algorithmen beschäftigen. Als Beispiel sei auf eine Arbeitsgruppe am Institut für Technische Optik (Universität Stuttgart) verwiesen, die sich ausschließlich mit der

Berechnung und Herstellung computergenerierter Hologramme für Strahlformungsaufgaben bei Hochleistungslasern beschäftigt.

Wir planen, computergenerierte Hologramme auf der Basis von LCDs für sehr verschiedene Aufgaben einzusetzen. Kurzfristig ist ein Einsatz im Rahmen des SYMAS¹-Projektes (siehe unten) vorgesehen; langfristig erhoffen wir uns aber einen breiten Einsatz der neuen Technik für eine Vielzahl von Anwendungen (s.u.).

Um den einerseits langfristigen und andererseits kurzfristigen Einsatz der neuen Technik zu ermöglichen, werde ich alle — aus meiner Sicht — interessanten Berechnungsverfahren kurz vorstellen und eine umfangreiche Literaturliste berücksichtigen. Bei der Vielzahl an verschiedenen Verfahren versteht es sich von selbst, daß nicht alle in der Fachliteratur vorgeschlagenen Algorithmen untersucht werden konnten. An den entsprechenden Stellen werde ich auf die Ergebnisse der Originalautoren verweisen. Auf die für den kurzfristigen Einsatz innerhalb des SYMAS-Projekts am geeignetesten erscheinenden Verfahren, insbesondere die *Iterativen Fourier-Transformations-Algorithmen*, werde ich entsprechend ausführlicher eingehen.

Die Darstellung der entsprechenden Theorie wird ergänzt durch mehrere Kapitel, die sich mit dem Einsatz von LCDs als Hologrammedien und experimentellen Ergebnissen befassen.

Mathematische und physikalische Grundlagen, die zum Verständnis der Arbeit erforderlich sind, werden der Vollständigkeit halber im Anhang beschrieben. Dort findet sich ebenfalls ein Verzeichnis aller verwendeten Abkürzungen und Variablen.

Fremdsprachliche Fachbegriffe wurden — sofern sie im Deutschen nicht sehr gebräuchlich sind — von mir nicht eingedeutscht. Da die Fachliteratur praktisch ausschließlich Englisch ist, scheint mir dies am angemessensten um dem Leser weiterführende Studien zu erleichtern. Hier gibt es natürlich Grenzfälle (z.B. “Speckles” kontra “Granulation”), die unterschiedliche Autoren unterschiedlich handhaben.

An einigen Stellen dieser Arbeit wird auf Adressen im Internet verwiesen. Ich verwende hierzu die Standard-URL-Adressen Nomenklatur. Dabei ist zu beachten, daß sich solche Adressen im Laufe der Zeit ändern können. Eine Suche (mit Standard-Search-Engines) führt aber im Zweifelsfall recht schnell zu dem gewünschten Dokument.

1.1 Motivation

Computergenerierte Hologramme (CGHs) wurden erstmals 1966 von Lohmann und Brown beschrieben ([brow66]). Damals noch eher eine Kuriosität, haben sich CGHs inzwischen als ernstzunehmende optische Elemente etabliert. Sie werden heute praktisch in allen Bereichen der Optik eingesetzt und ihre Bedeutung wird weiter (insbesondere in Verbindung mit Mikrooptik, Mikrosystemtechnik und Optoelektronik) stark zunehmen.

Da CGHs noch ein relativ junges Forschungsgebiet sind, sind die möglichen neuen Anwendungen noch nicht voll überschaubar. Mit CGHs wird es prinzipiell möglich, Punkten im dreidimensionalen Raum eine nahezu beliebige Lichtamplitude und Phase aufzuprägen.

Beispiele für bisherige Anwendungen, auf die hier leider nicht näher eingegangen werden kann, sind

¹ = “SYnthetische MASken”

- Strahlformung und Strahldiagnose von Hochleistungs-Lasern (z.B. [haup92], [hemb94])
- Optische Verbindungen und Multiplexing für digitale Hochgeschwindigkeits-Kommunikation und optische Computer (z.B. [kell93])
- Ersatz herkömmlicher, komplizierter Optiken (z.B. Strahlformung und Steuerung in kommerziellen CD-Playern ([kato91])) durch leichte, preiswerte Einzelemente
- Analogoptische Bildverarbeitung (z.B. [care91], [wyro91])
- Strahlsteuerung (z.B. in konfokalen Mikroskopen ([yin95]))
- Qualitätskontrolle asphärischer Elemente (z.B. [tizi80])
- Korrektur von optischen Systemen (z.B. [brow91]) und geometrische Transformationen (z.B. [bryn74])
- Display-Anwendungen (z.B. [luce92], [watl95])
- Scanning-Systeme (z.B. preiswerte Supermarkt-Scanner und Laser-Printer ([iwat92]))
- Meßtechnische Anwendungen (z.B. in Interferometern ([dein95]) oder zur Oberflächenanalyse ([silv93]))
- Mikrooptische Elemente (z.B. [spau91])
- Sicherheitstechnik und Identifikation (z.B. [gale91], [soar91])

Insbesondere bei der Strahlformung von Lasern und für optische Verbindungen haben die CGHs inzwischen viele alternative Konzepte ersetzt.

Die Vielseitigkeit computergenerierter Hologramme kann weiter gesteigert werden, wenn zeitlich variable Hologrammedien verwendet werden. Insbesondere kommen hierfür momentan LCDs (und in Zukunft auch die im Abschnitt 14.3 beschriebenen Elemente) in Frage. Das optische System kann in diesem Fall schnell und ohne erneute Justierung geändert werden. Auch hier sind die möglichen Anwendungen noch schwer zu überblicken.

Die meisten Systeme sind ihrem Charakter nach zweidimensional. Es wird eine Beziehung zwischen einer Objektebene und einer Bildebene — entsprechend der traditionellen Optik — hergestellt. Die Berechnungsverfahren für diese Fälle sind bekannt und werden beschrieben.

Es existieren aber auch Anwendungen, bei denen ein dreidimensionales, nicht an eine Bildebene gebundenes, Verhalten des optischen Elements gewünscht ist. Im Rahmen des SYMAS-Projektes soll untersucht werden, inwieweit entsprechende CGHs die optische Erfassung eines dreidimensionalen Objektes unterstützen können.

Bei klassischer Triangulation und Phasenschiebealgorithmen werden Muster (meist Streifen) auf ein Testobjekt projiziert und detektiert (siehe z.B. [wind95]). Bei dieser Projektion ergeben sich mit klassischen Optiken Probleme hinsichtlich der begrenzten Tiefenschärfe der Projektion (analog zu einem Diaprojektor, der nicht in der Lage ist, gleichzeitig mehrere verschieden weit entfernte Leinwände mit scharfen Bildern zu beleuchten). Im Rahmen der Streifenprojektion werden bisher Interferenzstreifen zur Beleuchtung bei großer Objektiefe verwendet. Bei Interferenzstreifen ergibt sich — prinzipbedingt — keine Begrenzung der

Tiefenschärfe. Allerdings ergeben sich praktisch Probleme (Stabilität) bei der Erzeugung der Streifen. Kompliziertere Muster, wie sie für einige Formaufzeichnungsverfahren² erforderlich sind, lassen sich nicht herstellen.

Mit den hier vorgestellten Algorithmen sind wir in der Lage, diese Probleme zu umgehen und mit einem optischen Element gleichzeitig in beliebigen Ebenen scharfe Abbildungen zu erzeugen.

Unabhängig von der geplanten Anwendung im SYMAS-Projekt sind neue Anwendungen z.B. in den folgenden Bereichen denkbar:

- Mikromarkierungen auf gekrümmten Oberflächen
- Display-Anwendungen
- Einsatz in scannenden Systemen (z.B. konfokale Mikroskopie und Kohärenzradar)

²siehe z.B. R.W.Malz: "Codierte Lichtstrukturen für 3-D-Meßtechnik und Inspektion", Doktorarbeit, ITO 1992

Kapitel 2

Beugung

Computergenerierte Hologramme gehören zur Klasse der *diffraktiven Optiken*. Ihr Verhalten wird durch die Beugungstheorie bestimmt.

Ausgehend von der Kirchhoff-Integral-Formel werden Fresnel und Fraunhofer-Beugung in einer für computergenerierte Hologramme sinnvollen Form entwickelt.

2.1 Fresnel-Beugung

Wie im Anhang ausgeführt, beschreibt die Kirchhoff-Integral-Formel (A.18) die Ausbreitung von gebeugtem Licht und stellt die exakte Formulierung des bekannten Huygensschen Prinzips dar.

Wir vernachlässigen in (A.18) den *Obliquity-Faktor* und betrachten den Sonderfall der Lichtausbreitung zwischen zwei parallelen Ebenen, die durch den Abstand z getrennt sind.

Während im Exponent eine exakte Verwendung des Abstandes r nötig ist, kann im Nenner der Integrand r durch den Ebenenabstand z ersetzt werden. Gleichzeitig ersetzen wir die transparente Apertur durch eine beliebige Apertur mit der komplexen Amplitude $O(x, y)$. Wir kommen damit zur vereinfachten Form

$$B(u, v) = \frac{1}{i\lambda z} \int \int_{-\infty}^{\infty} O(x, y) e^{ikr} dx dy \quad (2.1)$$

Dabei bezeichnet $B(u, v)$ das Lichtfeld in der Zielebene und $O(x, y)$ das Lichtfeld in der Ursprungsebene. u und v sind räumliche Koordinaten in der Zielebene.

Im Rahmen der *Fresnel-Näherung* ersetzt man den Abstand r zwischen zwei Punkten in der Ziel- und Ursprungsebene durch eine quadratische Näherung:

$$r = \sqrt{(u-x)^2 + (v-y)^2 + z^2} \approx z \left[1 + \frac{(u-x)^2}{2z^2} + \frac{(v-y)^2}{2z^2} \right] \quad (2.2)$$

Damit ergibt sich aus (2.1) direkt die *Fresnel-Beugung* als

$$B(u, v) = \frac{e^{ikz}}{i\lambda z} \int \int_{-\infty}^{\infty} O(x, y) e^{\frac{i\pi}{\lambda z} [(u-x)^2 + (v-y)^2]} dx dy \quad (2.3)$$

bzw. als die Faltung von O mit dem Fresnelpropagator $, (u, v, z)$:

$$B(u, v) = C [O(u, v) * , (u, v, z)] \quad (2.4)$$

$$, (u, v, z) = e^{\frac{i\pi}{\lambda z}(u^2+v^2)} \quad (2.5)$$

Die Konstante

$$C = \frac{e^{ikz}}{i\lambda z} \quad (2.6)$$

ist dabei für die meisten Anwendungen völlig belanglos¹.

Ich werde hier lediglich einige nützliche Eigenschaften des Fresnel–Propagators ohne Beweis (die Beweise sind mittels der Definition (2.5) leicht abzuleiten) angeben:

$$, *(u, v, z) = , (u, v, -z) \quad (2.7)$$

$$, (-u, -v, z) = , (u, v, z) \quad (2.8)$$

$$, (-u, -v, \infty) = 1 \quad (2.9)$$

$$\mathcal{F}[, (u, v, z)] = i\lambda , *(\lambda zu, \lambda zu, z) \quad (2.10)$$

Die Fourier–Transformation (siehe (B.1)) wird mittels des Operator \mathcal{F} geschrieben.

Die Fresnel–Propagation kann mittels der Fourier–Transformation ausgedrückt werden. Dies ist seit der Entdeckung der schnellen Fourier–Transformation (FFT) von erheblicher praktischer Bedeutung. Erst damit wird eine numerische Auswertung der Propagation komplizierter Lichtfelder, wie wir sie für die Berechnung von computergenerierten Hologrammen benötigen, praktikabel.

Gleichung (2.3) wird geschrieben als

$$B(u, v) = C e^{\frac{i\pi}{\lambda z}(u^2+v^2)} \int_{-\infty}^{\infty} O(x, y) e^{\frac{i\pi}{\lambda z}(x^2+y^2)} e^{-\frac{2\pi i}{\lambda z}[ux+vy]} dx dy \quad (2.11)$$

$$B(\mu, \nu) = C e^{\frac{i\pi}{\lambda z}(u^2+v^2)} \mathcal{F}[O(x, y) e^{\frac{i\pi}{\lambda z}(x^2+y^2)}] \quad (2.12)$$

Die Fourier–Transformation bezieht sich dabei auf *reduzierte Koordinaten* (*Ortsfrequenzen*), die einer einfachen Variablensubstitution entsprechen (siehe auch Abschnitt 3.1):

$$\mu := f_u := \frac{u}{\lambda z} \quad (2.13)$$

$$\nu := f_v := \frac{v}{\lambda z} \quad (2.14)$$

¹Der Faktor zeigt allerdings an, daß die gebeugte Welle eine Phasenschiebung um $\pi/2$ gegenüber einer ungebeugten Welle aufweist. Eine Tatsache, die z.B. zur Subtraktion von Bildern bei dynamischen Hologrammen in photorefraktiven Medien genutzt werden kann.

2.2 Fraunhofer–Beugung und Fourier–Hologramme

Die Berechnung des Beugungsbilds für große Abstände z zwischen Bild- und Ursprungsebene kann weiter vereinfacht werden.

Wenn

$$\frac{\pi}{\lambda z}(x^2 + y^2) \ll 1 \quad (2.15)$$

für alle Objektpunkte gilt, dann vereinfacht sich (2.12) zur *Fraunhofer–Beugung*

$$B(\mu, \nu) = C e^{\frac{i\pi}{\lambda z}(u^2 + v^2)} \mathcal{F}[O(x, y)] \quad (2.16)$$

Eng verwandt mit der Fraunhofer–Beugung sind die Fourier–Hologramme. Im einfachen Fall der Anordnung in Abbildung 2.1 berechnet sich die Lichtamplitude in der Bildebene über die Fresnel–Propagation.

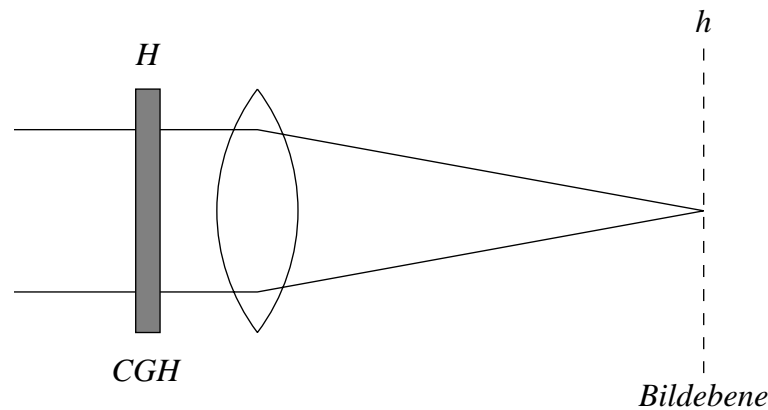


Abbildung 2.1: Rekonstruktion eines Fourier–Hologramms

Allerdings wird die Objektweite $O = h$ der Ursprungsebene mit dem Phasenglied

$$L(x, y) = e^{-i\frac{\pi}{\lambda f}(x^2 + y^2)} \quad (2.17)$$

einer Linse mit der Brennweite f multipliziert. Dieser Faktor hebt offensichtlich in Gleichung (2.12) gerade dann den zusätzlichen Phasenterm bei der Fourier–Transformation weg, wenn $z = f$ gilt. Damit ergibt sich für die Rekonstruktion dieses *Fourier–Hologramms* ebenfalls die Rekonstruktionsformel (2.16).

Kapitel 3

Fourier–Transformationen

Wir haben im letzten Kapitel gesehen, daß Beugungserscheinungen im wesentlichen durch Fourier–Transformationen beschrieben werden. Der Durchgang von Licht durch eine Linse mit nachfolgender Propagation im Raum kann ebenso durch eine Fourier–Transformation beschrieben werden. Es verwundert daher nicht, daß die Fourier–Transformation — und insbesondere ihre diskrete Version — entscheidend zur Berechnung computergenerierter Hologramme sein wird. Es ist daher unerlässlich, sich näher mit der zugrunde liegenden Theorie zu beschäftigen. Ohne Grundlagen zur diskreten und numerischen Behandlung von Fourier–Transformationen kann in der Praxis keine Berechnung computergenerierter Hologramme vorgenommen werden.

Für die — in diesem Zusammenhang weniger wichtigen — mathematischen Grundlagen der Fourier–Transformation sei auf die umfangreiche Literatur (siehe z.B. [bach92], [meyb91]) verwiesen.

3.1 Fourier–Transformation in der Optik

Die Fourier–Transformation wird in einer Vielzahl unterschiedlicher Anwendungsgebiete genutzt (z.B. allgemeine Signalanalyse, Quantenmechanik, Filtertheorie, Spektralanalysen, Kristallographie). Dabei werden natürlich unterschiedliche Variablen zur Beschreibung der physikalischen Größen verwendet. Beispielsweise wird in der Elektrotechnik zwischen einem eindimensionalen, zeitlichen Signal (z.B. Spannung in Abhängigkeit von Zeit) und der Frequenz des Signals mittels der Fourier–Transformation eine Verbindung hergestellt. Bei den in Kapitel 2 besprochenen Beugungserscheinungen wird dagegen zwischen räumlichen, zweidimensionalen Signalen, den komplexen Lichtamplituden, und sogenannten, ebenfalls zweidimensionalen, Raumfrequenzen transformiert. Die mathematische Struktur der Transformation wird von der Wahl dieser unterschiedlichen Variablen natürlich nicht beeinflusst.

In der Optik findet die Fourier–Transformation vielfältige Anwendungen. Außer zur Behandlung von Beugungserscheinungen und der Linsentransformation ist sie unerlässlich für die Beschreibung der Qualität optischer Systeme (Optische Transfer–Funktion (OTF), Modulationstransfer–Funktion (MTF) und Punktverwaschungs–Funktion), für die spektrale Analysen von Licht und die optische Informationsverarbeitung. Wir können auf diese vielfältigen Anwendungen hier nicht eingehen und verweisen auf die reichhaltige Literatur zur *Fourier–Optik*.

Für die beugungstheoretische Anwendung¹ haben wir eine Transformation zwischen einem Ortsraum (\vec{u}) und einem Ortsfrequenzraum (\vec{f}_u).

Der zweidimensionale Ortsraum bedarf keiner näheren Erläuterung. Der — entsprechend ebenfalls zweidimensionale — Ortsfrequenzraum muß dagegen erklärt werden. Raumfrequenzen werden über die bereits im letzten Kapitel angegebene Gleichung

$$\mu := f_u = \frac{u}{\lambda z} \quad (3.1)$$

bzw.

$$\nu := f_v = \frac{v}{\lambda z} \quad (3.2)$$

beschreiben.

Bei einer bestimmten Wellenlänge λ und einem festen Abstand z (z gibt dabei die Richtung der Lichtpropagation an) können wir somit Raumfrequenzen in gewöhnliche räumliche Koordinaten umrechnen. Ohne eine solche Umrechnungen hätten Raumfrequenzen keine anschauliche Bedeutung.

Wenn man also eine Beschreibung mit Raumfrequenzen – statt direkter räumlicher Koordinaten – wählt, gewinnt man den Vorteil einer allgemeinen – unabhängig von Wellenlänge λ und Abstand z – Beschreibung des physikalischen Systems.

Für unsere Anwendungen bei Fourier-Hologrammen heißt das zum Beispiel, daß wir die Rekonstruktion eines Hologramms einfach durch eine Fourier-Transformation ausdrücken können, wenn wir Raumfrequenzen für die Beschreibung des rekonstruierten Bildes verwenden. Die Beschreibung wird unabhängig von Wellenlänge und Brennweite f der Rekonstruktionsoptik (bei einem Fourier-Hologramm entspricht der Abstand z genau der Brennweite f der Linse), ohne Zweifel ein praktischer Vorteil.

Wir sehen andererseits daraus auch sofort, was passiert, wenn wir ein Hologramm, das für die Wellenlänge λ_1 berechnet wurde, mit einer anderen Wellenlänge λ_2 rekonstruieren. Nach (3.1) erhalten wir ein entsprechend größeres bzw. kleineres Bild².

Bei der Verwendung der reduzierten Koordinaten braucht man also die konkreten Systemparameter (Wellenlängen, Abstand, Abbildungsmaßstab) bei der Berechnung des Hologramms nicht berücksichtigen.

Bei der Besprechung der Algorithmen zur Berechnung computergenerierter Hologramme gehe ich von dem Hologramm als Ortsraum, repräsentiert durch die Koordinaten u, v aus. Die Rekonstruktion — gegeben durch die Fourier-Transformation — wird im Ortsfrequenzraum mit den reduzierten Koordinaten $f_x = x/(\lambda z)$ und $f_y = y/(\lambda z)$ dargestellt. Es muß darauf hingewiesen werden, daß es in der Fachliteratur oft üblich ist, in diesem Zusammenhang x bzw. y als reduzierte (!) Koordinaten zu verwenden um Schreibarbeit zu sparen und die Übersichtlichkeit zu erhöhen. Größen in der Hologrammebene werden von mir mit Großbuchstaben und Größen in der Rekonstruktionsebene mit Kleinbuchstaben bezeichnet.

¹und ebenso für die analytische Beschreibung linearer optischer Systeme mittels OTF und MTF

²Physikalisch ist das sehr anschaulich, da man sich eine Vergrößerung der Wellenlänge ebensogut als eine Verkleinerung der gesamten restlichen Anordnung (Hologramm und alle Abstände) denken kann. Ein feineres Hologramm bzw. Gitter führt aber automatisch zu einer größeren Ablenkung bei der Beugung.

3.2 Diskrete Fourier–Transformation

Unsere bisherige Beschreibung der Fourier–Transformation war kontinuierlich. Sowohl in der Eingangs– als auch in der Ausgangsebene der Fourier–Transformation gab es unendlich viele Punkte bzw. Werte.

In der Praxis werden wir aber nur eine endliche Anzahl von Meßpunkten registrieren und auch nur eine ebenfalls endliche Anzahl Amplitudenwerte innerhalb eines Computers verarbeiten können.

Wir müssen also von der kontinuierlichen Beschreibung zu einer diskreten Beschreibung übergehen³.

Wir gehen im folgenden dieser Arbeit davon aus, daß die Stützstellen, auf denen die Meßpunkte liegen, in äquidistantem Abstand Δx zueinander liegen. Das muß nicht immer der Fall sein. Meist wäre sogar eine andere Wahl der Stützstellen, z.B. in bestimmten Regionen eine erhöhte Anzahl (siehe [schr84] Kap. 2.2.3), sinnvoller. In der Regel weisen aber die in der Optik verfügbaren Aus– und Eingabemedien, insbesondere kommerzielle LCDs, ein gleichmäßiges Raster von Stützstellen auf.

Die einfachste und sinnvollste Möglichkeit, die Integration über die äquidistanten Meßpunkte durchzuführen, ist, sie einfach zu summieren. Wir wissen zunächst nicht, was sich zwischen den Meßpunkten abspielt und nehmen durch diese Summation einfach an, daß sich der Signalverlauf zwischen zwei Meßstellen nur wenig ändert.

Wir setzen also

$$a(\mu) = \int_{-\infty}^{\infty} A(x)e^{-2\pi i\mu x} dx \approx \sum_{n=-\infty}^{\infty} A(n\Delta x)e^{-2\pi i\mu n\Delta x} \quad (3.3)$$

Es stellt sich sofort die Frage, wieviele Stützstellen benötigt werden, um den durch die Approximation gegebenen Fehler nicht zu groß werden zu lassen.

Wir schreiben das abgetastete Signal x_n als (siehe B.7)

$$A_n := A(n\Delta x) = \text{comb}(x, \Delta x)A(x) \quad (3.4)$$

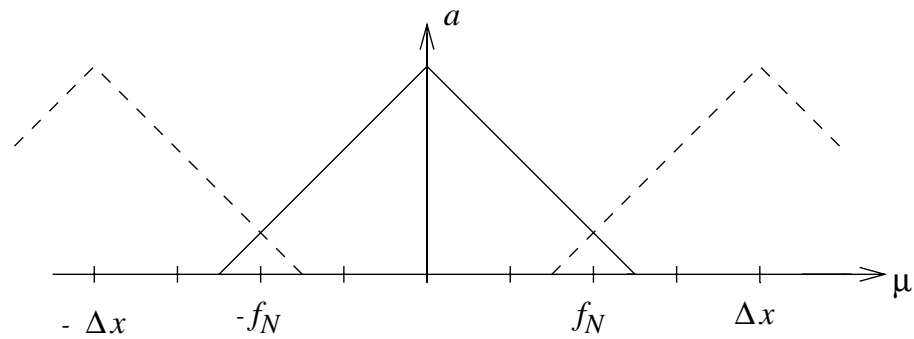
Mit (B.6) ergibt sich als Spektrum:

$$a_n = \text{comb}(\mu, 1/\Delta x) * a(\mu) \quad (3.5)$$

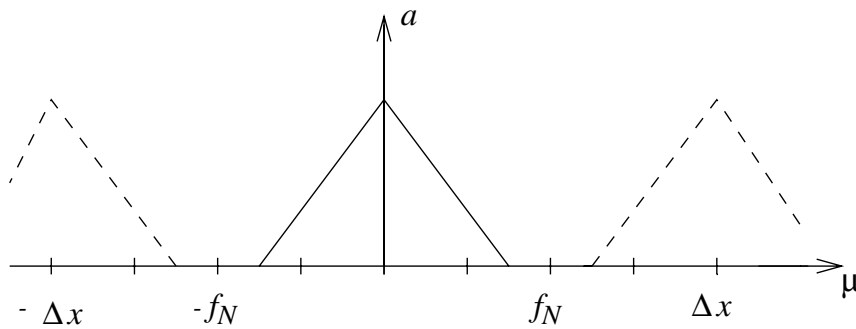
also die periodische Wiederholung des Spektrums von A . Die Wiederholungsperiode ist direkt durch $1/\Delta x$ gegeben. Es kommt ganz offensichtlich zu Fehlern wenn das Spektrum von a eine größere Ausdehnung als dieses $1/\Delta x$ hat. In diesem Fall überlagern sich dann die einzelnen spektralen Wiederholungen. Dies ist in Abbildung 3.1 für zwei verschiedene Signale demonstriert.

Der Fehler kann anschaulich (periodische Wiederholung des Spektrums) durch ein “Umklappen” des Spektrums an der sogenannten *Nyquistfrequenz* (s.u.) begriffen werden. Frequenzen mit $f + f_N$ erscheinen bei $f - f_N$. Man spricht von *Aliasing*. Wir weisen bereits

³Dasselbe Problem haben wir bei elektrischen Signalen: Wir können (und wollen es im übrigen auch nicht) einen Spannungsverlauf nicht unendlich genau vermessen. Wir haben uns auf eine begrenzte, ausreichende Anzahl von Meßpunkten zu beschränken



a.) Abtastbedingung verletzt: Aliasing



b.) Abtastbedingung erfüllt: kein Aliasing

Abbildung 3.1: Aliasing

hier darauf hin, daß der Aliasing-Fehler gleichbedeutend mit den “Speckles” bei computergenerierten Hologrammen ist (näheres siehe Abschnitt 7.1). Er kann theoretisch dadurch vermindert werden, daß keine äquidistante Abtastung erfolgt, sondern die Abtastung an zufälligen Punkten vorgenommen wird (“*Monte-Carlo-Abtastung*”, siehe [schr84] Kap. 2.2.3). Allerdings wird dadurch die Bildauflösung verringert.

Den wichtigsten Sonderfall haben wir bei einer sogenannten “Tiefpaß-Abtastung”: Das Signal wird als begrenzt hinsichtlich hoher Frequenzen angenommen (ein Fall, der sich sowohl in der Optik als auch in der Elektronik bei den meisten Systemen praktisch automatisch ergibt). Eine Begrenzung tiefer Frequenzen findet nicht statt. Wenn ein reelles Signal vorliegt, dann ist weiterhin zu beachten, daß das Spektrum symmetrisch bezüglich $f = 0$ ist. Insgesamt wird die Breite des Spektrums also bei einer maximalen Frequenz von f_g genau $2f_g$ sein.

Damit es zu keinen Überlappungen durch die Abtastung kommt, muß demnach $\frac{1}{\Delta x} \geq 2f_g$ bzw.

$$\Delta x \leq \frac{1}{2f_g} \quad (3.6)$$

gelten. Mit anderen Worten: Damit es zu keinen Fehlern bei der Abtastung kommt, darf das Signal nur Frequenzen unterhalb der sogenannten *Nyquistfrequenz* $f_N = 1/(2\Delta x)$ enthalten.

Ein solches Signal kann durch die sogenannte *Kardinalreihe* vollkommen aus den Abtastwerten restauriert werden (siehe z.B. [bach92] (7.6)):

$$A_r(x) = \sum_{n=-\infty}^{\infty} A(n\Delta x) \operatorname{sinc}(x/\Delta x - n) \quad (3.7)$$

Meist wird auf diese Art der Rekonstruktion verzichtet und einfach das Signal als annähernd konstant im Bereich um die Stützstelle angesehen.

Die Bandpaßabtastung (das Signal wird vor der Abtastung bandbegrenzt, es gibt also eine untere Grenzfrequenz $f_0 \neq 0$) erfordert in jedem Fall die Anwendung einer Rekonstruktionsformel⁴.

Wird die Abtastbedingung (3.6) nicht eingehalten, muß man mit einem Fehler in der Signalrekonstruktion rechnen. Dieser errechnet sich (siehe [baue93] 6.1.1-44) zu

$$\|A(x) - A_r(x)\| \leq 2 \int_{1/2\Delta x}^{\infty} [\|a(\mu)\| + \|a(-\mu)\|] d\mu = 4 \int_{1/2\Delta x}^{\infty} |a(\mu)| d\mu \quad (3.8)$$

Ein weiterer Fehler, die sogenannte *spektrale Diffusion* (*“Leakage”*), tritt dann auf, wenn wir nicht den gesamten Signalausschnitt bei der Integration berücksichtigen. Wir können in der Praxis bei der Bildung der Fourier-Transformation nicht von $-\infty$ bis ∞ integrieren. Viele Signale werden nur in einem begrenzten Bereich hinreichend von Null verschieden sein. In diesem Fall gibt dieser Bereich den sinnvollen Integrationsbereich an. Oft ist dieser Bereich aber viel zu groß. Man denke hier insbesondere an periodische Funktionen: Wählt man in diesem Fall als Integrationsbereich exakt eine Periode, dann ergeben sich keine Fehler. Bereits eine leichte Bereichsvergrößerung führt aber anschaulich sofort zu einem Fehler bei der Integration. In der Regel umgeht man solche Probleme durch sogenannte *Fensterfunktionen*, mit denen das Signal vor der Fourier-Transformation multipliziert wird. Es existieren eine Vielzahl von verschiedenen Verfahren zur Minimierung der Fehler aufgrund spektraler Diffusion, auf die wir hier nicht näher eingehen können (siehe [bach92], [gran89]).

⁴über das *Kohlenbergsche Abtasttheorem*

Kapitel 4

Holographie

Um computergenerierte Hologramme anzuwenden, sind Eigenschaften der “konventionellen” Hologramme nützlich. Für ein tieferes Verständnis sind zumindest Grundkenntnisse aus diesem Bereich unabdingbar. An diese Grundkenntnisse und die üblichen Bezeichnungen soll das vorliegende Kapitel erinnern. Es kann und will nicht das Wissen einer Monographie über “konventionelle” Hologramme (siehe z.B. [caul79], [hari84]) ersetzen.

4.1 Grundprinzip

Die Aufzeichnung von Lichtfeldern mittels herkömmlicher photographischer Methoden berücksichtigt lediglich die Amplitude des Lichtfeldes. Die Phaseninformation geht verloren. Um die Phaseninformation mit technischen Detektoren¹ aufzuzeichnen, bedient man sich des Tricks der kohärenten Überlagerung einer Referenzwelle \vec{r} mit dem Lichtfeld \vec{a} . Am Detektor wird damit die Intensität

$$I = (\vec{r} + \vec{a})(\vec{r}^* + \vec{a}^*) = |\vec{r}|^2 + |\vec{a}|^2 + \vec{r}^* \cdot \vec{a} + \vec{a}^* \cdot \vec{r} \quad (4.1)$$

gemessen. Der Aufzeichnungsvorgang ist in Abbildung 4.1 veranschaulicht.

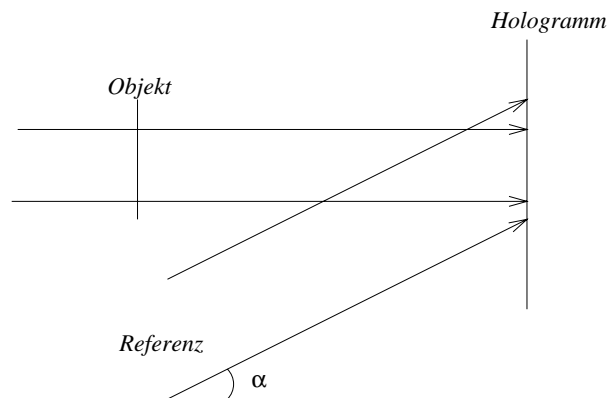


Abbildung 4.1: Aufnahme eines Fresnel-Hologramm

¹Herkömmliche Detektoren hoher räumlicher Auflösung sind zu langsam, um die hohen Frequenzen der elektromagnetischen Wellen im sichtbaren Bereich zeitlich — und damit ihre Phasenbeziehungen — aufzulösen.

Geht man von einem linearen Aufzeichnungsmedium aus, dann ergibt sich (nach der Entwicklung) unter der Voraussetzung der linearen Amplitudentransparenz des Speichers eine Amplituden-Transmission des Films (“Hologramms”) proportional zu I . Zur Rekonstruktion des Lichtfeldes beleuchtet man das Hologramm mit der ursprünglichen Referenzwelle \vec{r} . Direkt hinter dem Hologramm erhält man so die komplexe Lichtfeldamplitude

$$\vec{a}' = J\vec{r} = \vec{r}|\vec{r}|^2 + \vec{r}|\vec{a}|^2 + |\vec{r}|^2\vec{a} + \vec{a}^* \cdot \vec{r} \cdot \vec{r} \quad (4.2)$$

Wenn als Referenzwelle r eine ebene Welle verwendet wird, sind die Terme $|\vec{r}|^2$, $\vec{r} \cdot \vec{r}$ konstant und zunächst ohne Belang. Der Term $|\vec{r}|^2\vec{a}$ in (4.2) rekonstruiert das Lichtfeld. Entsprechend ergibt der Term $\vec{a}^* \cdot \vec{r} \cdot \vec{r}$ das sogenannte *pseudoskopische* Bild, also eine phasenkonjugierte Rekonstruktion.

Die übrigen beiden konstanten Terme sind als Störungen aufzufassen und werden als “Bias-Terme” bezeichnet.

Es gibt viele unterschiedliche Möglichkeiten, die Referenzwellen zu wählen. Die beiden einfachsten (und gebräuchlichsten) Varianten benutzen ebene Wellen, die entweder parallel (“on-axis”-Geometrie) oder schräg (“off-axis”-Geometrie) zu den Objektwellen einfallen.

Auch wir werden lediglich ebene Wellen als Referenz bzw. zur Rekonstruktion verwenden. Es könnte aber für einige Anwendungen im Bereich computergenerierter Hologramme durchaus von Vorteil sein, kompliziertere (vor allem rotationssymmetrische) Rekonstruktionswellen zu untersuchen.

4.2 Einteilung von Hologrammen

Hologramme können nach der Art der Aufnahme, nach dem Hologrammedium oder der Art der Rekonstruktion eingeteilt werden.

Das Hologrammedium wird in den meisten Fällen lediglich eine Beeinflussung der Amplitude zulassen (“*Amplitudenhologramm*”). Durch Bleichen des Amplitudenhologramms kommt man zu einem “*Phasenhologramm*”, bei dem die Amplitude konstant bleibt und die Phase eine räumliche Variation aufweist.

Nicht zu verwechseln sind “Amplituden-” bzw. “Phasenhologramme” mit den sogenannten “*Amplitude-Information*” bzw. “*Phase-Information*” Hologrammen. Bei selbigen wird von der zu speichernden Wellenfront (vor der Addition der Referenzwelle) lediglich Amplitude oder Phase verwendet. Amplitude-Information Hologramme haben einen sehr schlechten Beugungswirkungsgrad und werden daher kaum benutzt.

Hologrammedien stehen sowohl für Transmission als auch für Reflexion zur Verfügung. Auf die besonderen Eigenheiten verschiedener Aufzeichnungsmedien kann hier nicht eingegangen werden.

Von “*Fresnel-Hologrammen*” spricht man, wenn die Ausbreitung der Objektwelle zum Hologramm im wesentlichen durch eine Fresnel-Transformation bestimmt ist (siehe Kapitel 2).

Beim “*Fraunhofer-Hologramm*” erhält man durch Verwendung einer Linse, die die Fourier-Transformation des Objekts in der Hologrammebene bildet, ein “*Fourier-Hologramm*”. Wird nur die Intensität betrachtet, eignet sich auch eine Kugelwelle zur Beleuchtung; das

Quasi-Fourier-Hologramm erscheint im Fokus.

Wird das Objekt auf die Hologrammebene abgebildet, entfällt die übliche Forderung nach Monochromasie der Aufnahmequelle². Man spricht von einem “Bildebenen-Hologramm³”.

“Inkohärente Hologramme” bezeichnen Hologramme, die mit inkohärentem Licht aufgenommen wurden (siehe [good68]). “Weißlicht-Hologramme” werden mit weißem (polychromatischem) Licht rekonstruiert (siehe [hari84]).

4.2.1 Orts-Bandbreiten-Produkt

Hologramme enthalten Information über das gespeicherte Lichtfeld. Um eine gute Rekonstruktion des Lichtfeldes zu erhalten, ist eine gewisse minimale Speicherkapazität des Hologramms zu fordern. Üblicherweise verwendet man bei konventionellen und computergenerierten Hologrammen zur Beschreibung von Speicherkapazitäten das *Orts-Bandbreiten-Produkt* (*Space-Bandwidth-Produkt* (*SBP*)). Es gibt direkt die Anzahl der ansteuerbaren Punkte bzw. Photopigmente eines Hologrammediums an. Über die Dynamik der Ansteuerung (verschiedene Graustufen oder binäre Ansteuerung) wird keine Aussage getroffen.

Das SBP ist das Produkt aus Auflösung und Hologrammfläche⁴. Unterschiedliche Hologrammtypen benötigen unterschiedliche Speicherkapazitäten und haben dementsprechend unterschiedliche SBPs.

Übliche Hologramme mit z.B. 3000 Linien/mm und einer Größe von 25 cm² haben somit ein SBP in der Größenordnung von 30 · 10⁹.

Die Anzahl ansteuerbarer Pixel bei LCDs liegt dagegen lediglich im Bereich 10⁵ .. 10⁶. Hier liegt das Hauptproblem computergenerierter Hologramme begründet. Die extrem hohe Zahl von Punkten eines konventionellen Hologramms sprengt die Berechnungs- und Speicherkapazitäten heutiger Computer und führt zu enormen Problemen bei der Aufzeichnung des Hologramms. Selbst wenn der Computer in der Lage ist, ein sehr großes Hologramm zu berechnen, bleibt das Umsetzen der Computerdaten auf ein physikalisches Medium (z.B. Photoresist) schwierig, in angemessener Zeit zu bewerkstelligen.

Das geringe SBP computergenerierter Hologramme ist die Hauptursache für die im Vergleich zu konventionellen Hologrammen schlechte Qualität der CGHs.

4.2.2 Fresnel-Hologramme

Mit dem in Abbildung 4.1 dargestellten Aufbau werden *Fresnel-Hologramme* aufgenommen. Zu ihrer Betrachtung (Rekonstruktion) in endlicher Entfernung sind keine weiteren Hilfsmittel nötig.

Der Name “Fresnel-Hologramm” rührt daher, daß die Hologrammentstehung mittels der Fresnel-Beugungstheorie (siehe Kapitel 2) beschrieben wird.

²weil in diesem Fall auf der extrem kurzen Distanz vom Bild des Objekts bis zum Hologramm praktisch keine Divergenz der unterschiedlichen Farben auftreten kann

³=“image-plane hologram”

⁴Bei Volumen hologrammen ist es entsprechend das Produkt aus dreidimensionaler Auflösung und Hologrammvolumen.

Das eindimensionale Orts–Bandbreiten–Produkt⁵ SBP für den on–axis Fall ergibt sich nach dem in Abschnitt 4.2.1 gesagten ([caul79] (4.1)) zu

$$\text{SBP}_{1D} = 2\nu_{max}L \quad (4.3)$$

Dabei ist L die Hologrammlänge und ν_{max} die Cutoff–Frequenz des Hologrammediums (angegeben in lp/mm).

Fresnel–on–axis–Hologramme sind allerdings recht ungünstig. Die 0-te Beugungsordnung (Bias–Term in (4.2)) und die beiden rekonstruierten (virtuelle und pseudoskopische Rekonstruktion) Bilder überlagern sich. Sie sind nicht räumlich getrennt. Interessanter ist also das SBP für den off–axis Fall. Es ist wegen der zusätzliche Trägerfrequenz, die die Separation der Teilbilder bewirkt, deutlich verringert. Es ergibt sich ([caul79] 4.1)

$$\text{SBP}_{2D} = 2L \left(\nu_{max} - \frac{\sin \phi}{\lambda} \right) \quad (4.4)$$

Dabei ist ϕ der off–axis–Winkel.

Bei DeVelis und Reynold (siehe [caul79]) findet man detailliertere Informationen zu Fresnel–Hologrammen.

4.2.3 Fourier–Hologramme

Mit “Fourier–Hologrammen” werden alle Hologramme bezeichnet, deren Fourier–Transformation die Rekonstruktion eines gewünschten Objekts ergibt. Alle unterschiedlichen Typen (entsprechend verschiedenen Aufnahmegeometrien) können durch ein äquivalentes “linsenloses Fourier–Hologramm” dargestellt werden.

Ich werde die Fourier–Hologramme hier etwas ausführlicher besprechen da wir später sehen werden, daß sie für die uns interessierenden Anwendungen deutliche Vorteile gegenüber allen anderen Arten von Hologrammen besitzen.

Ein linsenloses Fourier–Hologramm wird durch den Aufbau in Abbildung 4.2 aufgenommen. (Zur Rekonstruktion wird das Hologramm wie gewöhnlich mit der Referenzwelle beleuchtet.)

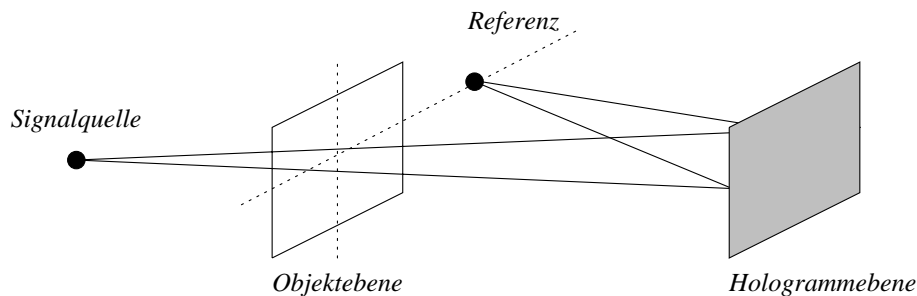


Abbildung 4.2: Aufnahme eines linsenlosem Fourier–Hologramm

⁵Zu 2 Dimensionen gelangt man einfach wenn man die SBPs der beiden Dimensionen miteinander multipliziert. Im Fall eines quadratischen Hologrammfilms wird also das Quadrat von (4.3) gebildet.

Die Referenzpunktquelle und das ebene — bzw. annähernd ebene — Objekt befinden sich in derselben Ebene. Das Objekt wird durch eine divergente Lichtquelle kohärent zur Referenzwelle beleuchtet. In [caul79] 4.3.3.1 findet der interessierte Leser eine genaue Analyse. Für uns entscheidend ist, daß ein Term der üblichen Superposition von Referenz- und Objektwelle in (4.2) bei der Rekonstruktion dieselbe *Intensität* ergibt wie die Intensität der Fourier-Transformation des aufgenommenen Hologramms. Die Phasen stimmen zwar nicht überein, da aber für die meisten Anwendungen nur Intensitäten von Belang sind, ist die Bezeichnung “Fourier-Hologramm” trotzdem angebracht.

Alle anderen Fourier-Hologramme lassen sich nun auf ein *äquivalentes* linsenloses Fourier-Hologramm zurückführen (siehe [caul79] 4.3.3.2). Dazu berechnet man die Bildpunkte R' und O' der Beleuchtungsquellen für Objekt O und Referenz R . Diese beiden Bildpunkte sind dann exakt die Referenz- bzw. Objektquellen im linsenlosen Fourier-Hologramm (siehe Abbildung 4.2). Einige typische Geometrien zeigt Abbildung 4.3.

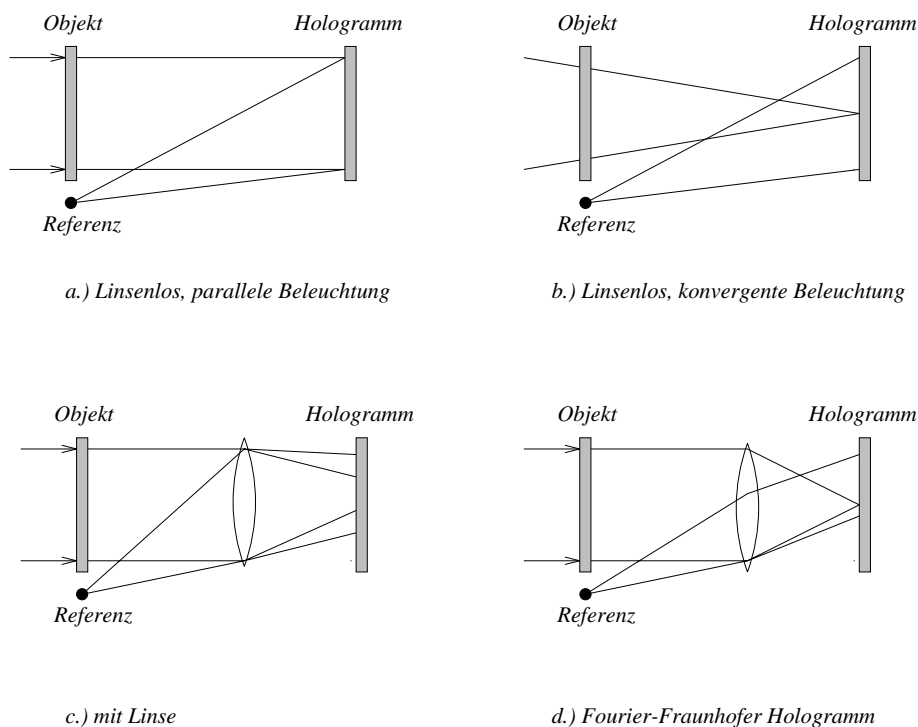


Abbildung 4.3: Geometrien für Fourier-Hologramme

Auf dem Hologrammfilm wird — wie gewöhnlich — die Intensität der Superposition von Objekt- und Referenzwelle aufgezeichnet. Die nach der Entwicklung des Films entstandene Transparenz ist rein reell. Bei der Rekonstruktion des Hologramms wird – per Definition — die Fourier-Transformation des Hologramms gebildet. Diese liefert aber für reelle Signale ein punktsymmetrisches Bild. Das Hauptbild des Objekts wird also von einem spiegelsymmetrischen Partnerbild begleitet. Wir werden auf diesen Punkt bei den computergenerierten Hologrammen noch ausführlicher zu sprechen kommen.

Einen Sonderfall der Fourier-Hologramme stellen die sogenannten *Fourier-Fraunhofer-Hologramme* dar (siehe Abbildung 4.3 d.). Nicht nur die Amplitude, sondern auch die Phase der Rekonstruktion stimmt mit der Objektamplitude überein.

Fourier-Hologramme haben mit einem Problem zu kämpfen: die in der Hologrammebene aufzuzeichnende Intensitätsverteilung ist sehr unregelmäßig. Es gibt weite Bereiche mit

einer nur relativ geringen Intensität und meist einen zentralen Bereich mit extrem hoher Intensität. Da die Dynamik des Hologrammediums aber in der Praxis zwangsläufig beschränkt sein muß, ist das sehr ungünstig. Ein Teil der Information in der Hologrammebene geht bei der Aufzeichnung verloren. Es gibt zwei einfache Lösungen, die dieses Problem vermindern: eine leichte Verschiebung der Hologrammebene entspricht einer Defokussierung der optischen Abbildung und einer Angleichung der in der Hologrammebene vorkommenden Intensitäten. Man spricht von *defokussierten Fourier-Hologrammen*. Eine andere Möglichkeit ist die Verwendung eines einfachen Diffusors zwischen Objekt und Objektbeleuchtung. Der Diffusor sorgt für eine annähernd zufällige Phase des Objekts und damit für eine annähernd konstante Objektamplitude in der Hologrammebene. Ich werde auf das wichtige Gebiet der Diffusoren im Rahmen der computergenerierten Hologramme noch ausführlich eingehen, so daß ich es an dieser Stelle bei dieser etwas dürftigen Beschreibung bewenden lassen will.

Die Hauptvorteile von Fourier-Hologrammen sind:

- Die Position des rekonstruierten Bildes ist unabhängig von einer eventuellen lateralen Verschiebung des Hologramms. Der Grund hierfür ist mit (B.10) einsichtig. Die Verschiebung bewirkt in der Rekonstruktion lediglich eine Modulation der Phase. Die Intensität ändert sich also nicht.
- Einzelne, lokalisierte Fehler des Hologrammediums wirken sich kaum aus, da die Objekt-Information über das gesamte Hologramm verteilt gespeichert wird.
- Die Aberrationen (auch bei Änderung der Rekonstruktionswellenlänge λ) sind minimal.
- Fourier-Hologramme lassen sich problemlos in on-axis Geometrie verwenden. Das hat oft experimentell Vorteile und erhöht die effektiv speicherbare Informationsmenge. Die bei Fresnel-Hologrammen nötige Modulation zur Trennung der rekonstruierten Bilder kann entfallen.
- Eine weitere Steigerung der Speicherkapazität gegenüber Fresnel-Hologrammen kommt durch das Fehlen der Lichtbündelung zustande. Im Fresnel-Hologramm muß zusätzlich zur eigentlichen Objektinformation implizit die Information zur Bündelung des Lichts auf eine bestimmte Ebene – also eine Art mathematischer Linse – gespeichert werden. Dies ist bei Fourier-Hologrammen nicht nötig. Die Bündelung der Rekonstruktion wird durch eine zusätzliche externe Linse vollzogen. Caulfield (siehe [caul79] 4.3.4.5) gibt einen 4 bis 100 mal so großen maximal speicherbaren Informationsgehalt gegenüber Fresnel-Hologrammen an⁶.

Der letzte Punkt, das Vorhandensein einer zusätzlichen Rekonstruktionslinse, kann allerdings auch für viele Anwendungen einen finanziellen⁷ und praktischen Nachteil bedeuten. Die Rekonstruktionslinse sollte nach Möglichkeit mindestens doppelt so groß wie das Objektfeld sein. Wenn eine solche Linse verwendet wird und das Objekt keine Raumfrequenzen größer als $R/(2\lambda d)$ enthält (R ist der Linsenradius, d der aus Abbildung 4.2 bekannte Abstand), ergibt sich das Orts-Bandbreiten-Produkt zu ([caul79] 4.3.(32))

$$\text{SBP}_{2D} = \frac{R^4}{16\lambda^2 d^2} \quad (4.5)$$

⁶dieser Faktor hängt vom Objekt selbst und von der Stärke der impliziten Linse ab

⁷gute Rekonstruktionslinsen sind nicht billig und unter Umständen recht groß

Die Vergrößerung der maximal speicherbaren Datenmenge im Hologramm macht Fourier-Hologramme vor allem für computergenerierte Hologramme interessant. Dort wird man in aller Regel aus praktischen Gründen⁸ nur ein – im Vergleich zu konventionellen Hologrammen – sehr geringes SBP zur Verfügung haben.

⁸Berechnungskomplexität und Ausgabemedium

Kapitel 5

Computergenerierte Hologramme

Im vorangegangenen Kapitel wurden konventionelle Hologramme besprochen. Es soll nun im folgenden der Übergang zu computergenerierten Hologrammen vollzogen werden. Insbesondere werden unterschiedliche Typen computergenerierter Hologramme kurz besprochen. Auf Details und die konkrete Berechnung der Hologramme wird in den folgenden Kapiteln eingegangen.

Zunächst stellt sich aber die Frage, weshalb man überhaupt Hologramme berechnen soll, wenn man sie doch auch einfach aufnehmen kann. Als Gründe für die Berechnung seien angeführt:

- Berechnung von Hologrammen für Objekte, die nicht real sondern nur in digitaler Form (z.B. CAD Programm) vorliegen
- bessere Nutzung des Hologrammediums
- Hologramme für Transformationen aller Art (z.B. Laserstrahlformung für die Materialbearbeitung)
- Implementierung komplizierter optischer Funktionen (z.B. Kombination verschiedener optischer Grundelemente)
- Verwendung von Medien, auf die nicht direkt holographisch gespeichert werden kann

Kurz gesagt sind praktisch alle der in Abschnitt 1.1 aufgeführten Anwendungen Gründe für den Einsatz von computergenerierten Hologrammen. Die Mehrzahl dieser Anwendungen läßt sich nicht mit "konventionellen" Hologrammen verwirklichen.

5.1 Vom Hologramm zum computergenerierten Hologramm

Ein "konventionelles" Hologramm entsteht durch kohärente Überlagerung einer Referenzwelle mit der aufzuzeichnenden Objektwelle (siehe (4.2)). Dieser Vorgang kann einfach durch den Rechner nachgebildet werden (siehe Abbildung 5.1). Dazu berechnet man "einfach" die Objektwelle am Ort des Hologramms, also in der *Hologrammebene*, und addiert eine gewünschte Referenzwelle. Das Ergebnis ist dann zu quadrieren da der Film in der

“konventionellen” Holographie ja die Intensität und nicht die komplexe Amplitude des Lichtfeldes aufzeichnet.

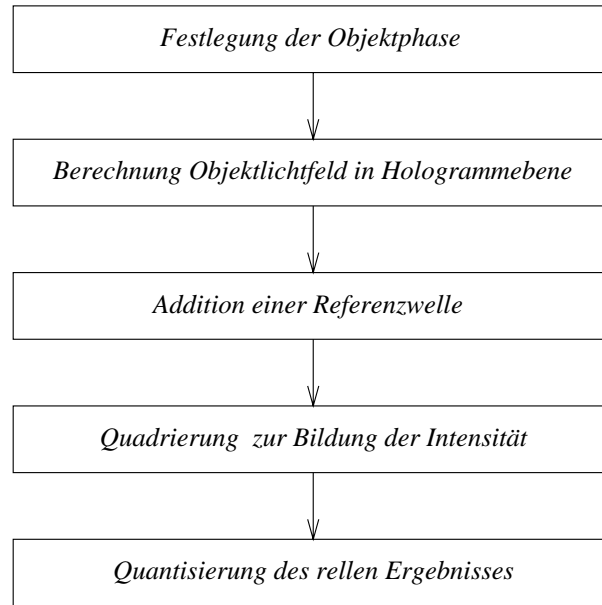


Abbildung 5.1: Simulation der optischen Hologrammaufnahme

Dieser physikalisch sehr naheliegende Ansatz bringt einige Probleme mit sich:

- Da bei einem “konventionellen” Hologramm mehrere Milliarden Bildpunkte auf dem Film gespeichert werden, ist die Berechnung der Objektwelle in der Hologrammebene sehr aufwendig.
- Die Ausgabe der errechneten Hologrammdaten auf ein Hologrammedium ist – ebenfalls wegen der enorm großen Datenmengen – problematisch.
- Die in (4.2) auftretenden zusätzlichen, störenden Terme (Bias-Terme und pseudoskopisches Bild) bleiben erhalten. Das Hologrammedium wird nicht optimal genutzt.

Nichtsdestotrotz haben wir hier also eine erste Möglichkeit zur Hologrammberechnung. Es bleibt lediglich die Frage, wie die Berechnung der Objektwelle in der Hologrammebene vorgenommen wird. Wie in Kapitel 2 beschrieben, erhält man die gesuchte komplexe Amplitudenverteilung in der Hologrammebene über eine Fresnel-Transformation aus der komplexen Amplitudenverteilung in der Objektebene. Dabei geht man der Einfachheit halber zunächst von einem ebenen, zweidimensionalen Objekt aus. Die Verallgemeinerung auf dreidimensionale Objekte erfolgt später, ist aber für viele Anwendungen nicht nötig.

Wenn man die Berechnung über die Fresnel-Transformation konkret durchführen will, stellt man fest, daß in der Objektebene meist nur die Amplitude der Objektwelle gegeben ist. Man stelle sich z.B. vor, daß ein Hologramm berechnet werden soll, daß in der Rekonstruktion das Bild eines Hauses ergibt. In diesem Fall liegt also das Haus in der Objektebene (das aufzuzeichnende Objekt in der “konventionellen” Holographie ist ein Haus) und man interessiert sich auch lediglich für die Amplitude dieses Objektes. Die Phase des Objektes bzw. der Hologrammrekonstruktion ist ohne Belang.

Mathematisch kann die Phase des Objektes, für welches das Hologramm berechnet werden soll also beliebig gewählt werden. Die Phase ist ohne Einfluß auf das Ergebnis der Hologrammrekonstruktion. Sie hat aber sehr wohl einen Einfluß auf das Hologramm selbst. Je nachdem, wie die Objektphase gewählt wird, ergeben sich völlig verschiedene Hologramme, die sich stark hinsichtlich Dynamik und räumlicher Verteilung relevanter Bereiche, d.h. Bereiche in denen die Intensität ungleich Null ist, unterscheiden.

Für unendlich große oder zumindest sehr große Hologrammedien mit einem extrem großen Dynamikbereich wäre das ohne Belang. In der Praxis werden die Hologramme aber auf einem räumlich begrenzten und dynamisch eingeschränkten Medium aufgezeichnet. Die ideal berechneten Hologrammdaten werden also bei der Aufzeichnung verfälscht. Es nimmt nicht wunder, daß dann aber auch die Rekonstruktion des Hologramms mehr oder weniger verfälscht sein muß.

Je nachdem wie man also die Phase des Objekts wählt, ergeben sich mehr oder weniger günstige Hologramme in dem Sinn, daß die Hologramme zu dem Hologrammedium passen.

Dieses Verhalten ist bereits von "konventionellen", optisch aufgenommenen Hologrammen, bei denen man in vielen Fällen eine Verbesserung der Bildqualität durch Diffusoren erreichen kann, bekannt.

Ich werde in späteren Kapiteln noch ausführlicher auf die Wahl der Objektphase eingehen. Vorerst werden wir — ganz der Simulation einer "herkömmlichen" Hologrammaufzeichnung entsprechend — die Phase entweder konstant oder zufällig wählen.

Damit hat man also die volle Information in der Objektebene, auf die nun eine Fresnel-Transformation angewendet werden soll. Die Fresnel-Transformation kann voll analytisch durchgeführt werden. Das ist für einfache Objekte wie z.B. einen Punkt durchaus möglich. Bei komplizierten Objekten (z.B. ein Gesicht oder eine überlagerte, komplizierte Phasenverteilung) wird eine analytische Berechnung praktisch unmöglich und der Einsatz eines Computers wird unumgänglich. Wir sind bei den *computergenerierten Hologrammen* angelangt.

In Kapitel 2 wurde gezeigt, daß die Fresnel-Transformation im wesentlichen auf eine Fouriertransformation reduziert werden kann. Das ist von erheblicher praktischer Bedeutung. Vom Ergebnis der Fresnel-Transformation interessieren uns nämlich nur die Amplituden an bestimmten Stellen. Und zwar sind das genau die Stellen, die wir bei der Hologrammausgabe (z.B. über einen Plotter) auch ansteuern können. Schließlich muß das berechnete Hologramm auch auf ein physikalisches Medium geschrieben werden. In der Regel werden diese Stellen einen äquidistanten Abstand zueinander haben, also auf einem zweidimensionalen Gitter liegen. Damit kann die kontinuierliche Fouriertransformation durch eine diskrete Fouriertransformation ersetzt werden.

Eine diskrete Fouriertransformation kann aber vergleichsweise schnell mittels der *schnellen Fouriertransformation*¹ ausgeführt werden.

Alle Hologrammberechnungen, die in den weiteren Kapiteln besprochen werden, stellen letztlich Abwandlungen bzw. Erweiterungen dieser einfachen Vorgehensweise dar.

¹Fast-Fourier-Transform (FFT)

5.2 Einteilung computergenerierter Hologramme

Bei der im letzten Abschnitt beschriebenen einfachen Hologrammberechnung sind wir direkt von einem Fresnel-Hologramm nach Abschnitt 4.2.2 ausgegangen. Nun haben wir aber im Kapitel 4 gesehen, daß es eine Vielzahl weiterer Hologramme gibt. Auch für dieses Hologrammtypen kann man – ganz analog zu Abschnitt 5.1 – Berechnungen vornehmen. Als Grundunterscheidung können die zwei Gruppen

- Fresnel-Hologramme
- Fourier-Hologramme

gewählt werden. Dabei kann man Fraunhofer-Hologramme als Sonderfall der Fourier-Hologramme (sehr großer Abstand) ansehen. Die beiden Hologrammtypen unterscheiden sich sowohl hinsichtlich ihrer Berechnung als auch in der optischen Anordnung zur Hologrammrekonstruktion.

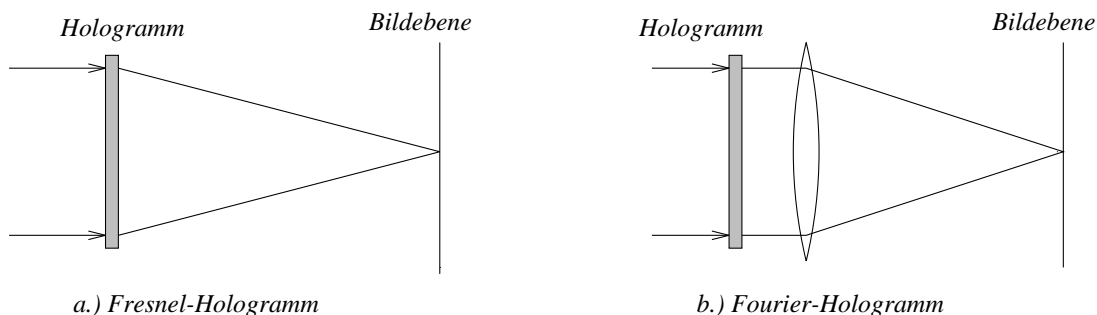


Abbildung 5.2: Rekonstruktion computergenerierter Hologramme

Fourier-Hologramme haben einige gewichtige Vorteile gegenüber Fresnel-Hologrammen (siehe näheres in Abschnitt 4.2.3) und werden daher öfter benutzt:

- Der Pixelabstand des Hologramms ist weitgehend ohne Belang²
- Eine beliebige Größenskalierung der Rekonstruktion durch einfache Wahl der Rekonstruktionslinse ist möglich
- Lokalisierte Defekte stören nicht
- Bessere Nutzung des Orts-Bandbreiten-Produkts
- Etwas einfachere mathematische Berechnung bzw. Handhabung

Der einschlägigen Literatur entsprechend gehe ich – wenn nicht ausdrücklich anders beschrieben – grundsätzlich von Fourier-Hologrammen aus. Eine Verallgemeinerung auf Fresnel-Hologrammen ist aber meist einfach möglich.

Weiterhin wird unterschieden zwischen

- on-axis Hologrammen

²Fresnel-Hologramme benötigen die für "herkömmliche" Hologramme nötige Auflösung von mehreren 100 bis 1000 lp/mm.

- off-axis Hologrammen

Im Abschnitt 4.2.2 haben wir gesehen, daß bei “herkömmlichen” Hologrammen praktisch kaum on-axis-Hologramme Verwendung finden. Der Grund hierfür lag in der störenden Überlagerung des rekonstruierten Bildes mit den in 4.2 angegebenen zusätzlichen Bildtermen.

Für die im letzten Abschnitt beschriebene physikalische Simulation der Hologrammaufzeichnung von Fresnel-Hologrammen bleibt dieses Problem. Es zeigt sich aber, daß man es bei computergenerierten Hologrammen vermeiden kann, weshalb hier meist on-axis-CGHs verwendet werden. Man beachte, daß man ein off-axis (Fourier-)Hologramm immer mit dem Formalismus des on-axis Hologramms betrachten kann. Dazu muß man nur das Objekt entsprechend verschieben.

Da Objekt und Hologrammebene im wesentlichen durch eine Fouriertransformation miteinander verbunden sind, bewirkt eine Verschiebung des Objekts in x-Richtung eine Modulation des Hologramms mit $e^{if_x \Delta x}$. Das entspricht aber genau der Addition einer schräg einfallenden Referenzwelle, wie es sein muß. Wir sehen hier sofort, wie beim “konventionellen” Hologramm einen gewaltigen Vorteil von on-axis Hologrammen gegenüber den off-axis Hologrammen. Bei off-axis Hologrammen findet eine Modulation statt. Diese Modulation ist um so stärker je stärker der off-axis-Effekt ist und führt zu der Forderung nach einer möglichst großen Hologrammauflösung. Dem on-axis Hologramm ist praktisch ein Gitter mit hoher Periode überlagert, das die off-axis Ablenkung bewirkt. Ein Teil der durch das Hologrammedium begrenzten Auflösung und Information wird also für das Gitter verwendet und “fehlt” zur Darstellung des eigentlichen Objektes. Wir gehen daher – wenn nicht ausdrücklich anders vermerkt – immer von on-axis Hologrammen aus und denken uns im Zweifelsfall einfach eine Verschiebung des Objekts.

Bisher wurden die Hologramme im Zusammenhang mit “gewöhnlichen” Filmen zur Hologrammaufzeichnung besprochen. Diese Filme stellen einfache Amplitudenfilter dar. Es sind aber auch Medien denkbar, die nicht die Amplitude der einfallenden Referenzwelle, sondern deren Phase oder unter Umständen Amplitude und Phase zugleich, modulieren. Man kann dementsprechend zwischen

- Amplitudenhologrammen (AH)
- Phasenhologrammen (“Kinofoms”) (PH)
- komplexen Hologrammen (KH)

unterscheiden. Reine Phasenhologramme (*Kinofoms*) haben eine deutlich höhere Beugungseffizienz (s.u.). Das ist kaum verwunderlich, werden doch bei Amplitudenhologrammen Teile des einfallenden Lichtes einfach ausgeblendet. Weiterhin vermeiden Phasenhologramme die oft störende (sowohl hinsichtlich einer Störung des rekonstruierten Bildes als auch hinsichtlich einer verminderten Intensität im rekonstruierten Bild) nullte Ordnung bei Fourier-Hologrammen und der nutzbare Bereich ist doppelt so groß wie bei Amplitudenhologrammen (siehe Abschnitt 4.2.3).

Das Hologrammedium wird noch zur weiteren Unterscheidung der Hologramme herangezogen. Unterschiedliche Medien haben unterschiedliche Dynamiken. Meist verwendet man die Anzahl unterscheidbarer Amplituden- oder Phasenstufen zur Kennzeichnung. Verwendet man also z.B. einen Photoplotter mit 8 möglichen Belichtungsstufen zur Hologrammaufzeichnung, dann ergibt sich ein 8-Stufen bzw. 3-Bit Hologramm. Hologramme mit nur zwei

verschiedenen Stufen (*binäre Hologramme*³) sind natürlich am einfachsten herzustellen und sind daher von besonderer Bedeutung.

Als letztes ist bei computergenerierten Hologrammen zwischen punkt- und zellenbasierten Hologrammen zu unterscheiden: Bisher wurden nur punkt-basierte Hologramme betrachtet. Es wird sich aber in Kapitel 6 zeigen, daß es nützlich sein kann, für jeden berechneten Hologrammwert n Ausgabepixel zu verwenden. Das heißt, die Anzahl der zu berechnenden Hologrammwerte ist um den Faktor n kleiner als die Anzahl der ansteuerbaren Hologrammpixel. Diese Freiheit kann zur Erhöhung der Dynamik oder zur Kodierung komplexer Amplituden (also Phase und Amplitude) in einem reellen Medium (z.B. einfacher Film) verwendet werden.

Es bleibt zu erwähnen, daß kein allgemeingültiges Einteilungsschema für CGHs existiert. Unterschiedliche Autoren verwenden unterschiedliche Grundeinteilungen. Eine Einteilung ist meist auch ohne Belang. Wichtig ist, die Unterschiede zwischen den möglichen Typen von CGHs zu kennen, um damit eine Auswahl der Berechnungsverfahren und Hologrammedien vornehmen zu können.

Welcher Hologrammtyp wird nun von uns verwendet? Offensichtlich ist die on-axis Fourier-Anordnung die für unsere Zwecke günstigste Geometrie. Phasenhologramme sind, sofern sie sich verwirklichen lassen, den Amplitudenhologrammen überlegen. Sie erhöhen die Beugungseffizienz, vermindern die oft störende nullte Ordnung ("DC-Anteil") und schränken die Rekonstruktion nicht auf symmetrische Objekte ein.

Wir werden also Fourier-Phasenhologramme verwenden und dementsprechend die Schwerpunkte dieser Arbeit an Fourier-Phasenhologrammen auslegen. Ein Übergang zu Amplitudenhologrammen ist dann interessant, wenn die phasenschiebenden Eigenschaften der LCDs sehr schlecht werden. Es sollte aber darauf hingewiesen werden, daß jedes binäre Amplitudenhologramm genauso als binäres Phasenhologramm verwendet werden kann und deswegen durchaus auch Techniken, die von Amplitudenhologrammen ausgehen, für uns wichtig sein können.

Das Umwandeln eines Amplitudenhologramms in ein Phasenhologramm ist in der "konventionellen" Holographie ein bekanntes Verfahren, um die Beugungseffizienz (siehe nächster Abschnitt) zu erhöhen. Es wird durch einfaches Bleichen des Hologramms durchgeführt. Ein binäres Amplitudenhologramm besteht aus den Werten 0 und 1. Ein binäres Phasenhologramm dagegen aus den Werten -1 und 1. Man sieht sofort, daß das entsprechende Phasenhologramm aus dem Amplitudenhologramm lediglich durch eine Verschiebung (nullte Ordnung fehlt!) und eine Skalierung (um den Faktor 4 höhere Beugungseffizienz!) hervorgeht.

³Innen ist die Bezeichnung "binäre Optik" zu verdanken. Der Begriff "binäre Optik" wird aber inzwischen oft allgemein für diffraktive, computergenerierte Optiken aller Art (unabhängig von der Zahl der Quantisierungsstufen) verwendet (siehe [veld92])

5.3 Beurteilung computergenerierter Hologramme

Die Beurteilung der Qualität computergenerierter Hologramme wird anhand der Rekonstruktion getroffen. Ausschlaggebend für die “Qualität” sind in der Regel zwei Parameter:

- Beugungseffizienz
- korrekte Lichtverteilung in der Rekonstruktion

Die Angabe der beiden entsprechenden Parameter in der Fachliteratur und bei eigenen Arbeiten ist mit großer Vorsicht und Sorgfalt zu beurteilen. Die üblichen numerischen Angaben täuschen eine exakte Beschreibung vor, die leider nur sehr eingeschränkt gegeben ist. Zunächst ist klar, daß man entsprechende Werte nur dann sinnvoll vergleichen kann, wenn sie auf denselben Definitionen beruhen. Sowohl bei dem Rekonstruktionsfehler als auch bei der Beugungseffizienz existieren unterschiedliche Definitionen weshalb die entsprechenden Werte nicht immer untereinander vergleichbar sind! Hinzu kommt, daß bei den Beurteilungen unterschiedliche Bereiche zur Auswertung verwendet werden können. Auch hier muß man sorgfältig prüfen, welche exakte Vorgehensweise zur Bestimmung von Beugungseffizienz und Rekonstruktionsfehler vom jeweiligen Autor gewählt wurde. (Einige Autoren scheinen sich dieser Problematik nicht bewußt zu sein und vermeiden es, die entsprechenden Details anzugeben).

Wurden alle diese Fallstricke beachtet, dann zeigt sich, daß zwar einzelne Hologramme, nicht aber gesamte Algorithmen, beschrieben werden können. Die Ergebnisse hinsichtlich Rekonstruktionsqualität und Beugungseffizienz, die mit verschiedenen CGH-Algorithmus erzielt werden, hängen stark von der konkret gewählten Definition von Beugungseffizienz und Rekonstruktionsfehler, aber auch von den verwendeten Testbildern ab.

Die Angabe konkreter Werte macht hier nur dann Sinn wenn für eine ganz spezielle Aufgabe, z.B. die Berechnung eines Fan-out Elements, ein spezielles Hologramm berechnet werden soll. In diesem Fall kann wirklich verglichen werden, welche Algorithmen eine Steigerung der Beugungseffizienz oder eine Verminderung des Rekonstruktionsfehlers bewirken. *Man muß sich dabei aber unbedingt bewußt sein, daß für eine andere Anwendung unter Umständen völlig andere Ergebnisse zu erwarten sind!*

Allgemeinere Aussagen über die Qualität der Algorithmen können hier leider nicht an einigen Werten festgemacht werden, sondern müssen qualitativ formuliert werden (siehe Kapitel 16).

In diesem Zusammenhang wäre es interessant ein allgemeines Meßverfahren zur Hologrammqualität einzuführen. Ein entsprechendes Testprogramm würde für verschiedene standardisierte Testbilder und Versuchsbedingungen (Hologrammgröße, Quantisierung usw.) jeweils den Rekonstruktionsfehler und die Beugungseffizienz ermitteln. Wenn klar ist, was für ein Hologramm konkret berechnet werden soll, kann die vollautomatisch generierte Tabelle für exakte Einzelbeurteilungen verwendet werden. Für allgemeinere Aussagen werden die Teilergebnisse zu einem Gesamtergebnis zusammengefasst, das eine sinnvolle Grobbeurteilung zuläßt.

Letztlich kann nur durch ein solches normiertes Beurteilungsverfahren langfristig eine sinnvolle Entscheidungshilfe erbracht werden. Bisher ist es leider meist erforderlich, die in Frage kommenden Verfahren zu implementieren um sie unter den gewünschten Versuchsbedingungen testen zu können. Mit den üblicherweise in der Fachliteratur angegebenen Ergebnissen kann selten auf Ergebnisse bei anderen Versuchsbedingungen geschlossen werden.

5.3.1 Rekonstruktionsfehler

Als Fehlermaß für die “korrekte Lichtverteilung” wird meist die mittlere quadratische Abweichung E der rekonstruierten Amplitude h von der gewünschten Amplitude f verwendet.

$$E = \sum_{f_x, f_y \in B} (|f(f_x, f_y)| - \lambda |h(f_x, f_y)|)^2 \quad (5.1)$$

Dabei gibt B den Bereich der gewünschten Intensitätsverteilung an und λ ist ein Skalierungsfaktor, der berücksichtigt, daß die Beugungseffizienz nicht 1 ist. Dieser Faktor wird so gewählt, daß der Rekonstruktionsfehler minimal wird. Dementsprechend werden die partiellen Ableitungen von E nach dem Real- und Imaginärteil von λ auf Null gesetzt, was zu

$$\lambda = \frac{\sum_{f_x, f_y \in B} |f(f_x, f_y)| |h(f_x, f_y)|}{\sum_{f_x, f_y \in B} |h(f_x, f_y)|^2} \quad (5.2)$$

führt (siehe [seld87] (10)). Es ist aber auch möglich, den Berechnungsaufwand für λ zu vermeiden und einen konstanten Wert zu verwenden. Dies ist unter Umständen innerhalb von Optimierungsroutinen ein einfacher Ansatz, um eine Beschleunigung zu erreichen. Leider verwenden einige Autoren dieses konstante λ auch zur allgemeinen Qualitätsbeurteilung ihrer Rekonstruktionen. Ein Vergleich mit anderen Arbeiten wird dann unmöglich.

Wenn der Fehler für unterschiedliche Testbilder beurteilt werden soll, empfiehlt sich eine Normierung gemäß

$$e := \frac{\sqrt{E}}{\max |f(f_x, f_y)| XY} \quad (5.3)$$

Dabei ist XY die Bereichsgröße, die für unterschiedliche Testbilder selbstverständlich auch unterschiedlich groß sein kann.

Es muß darauf hingewiesen werden, daß diese Definition der Rekonstruktionsqualität über den quadratisch gemittelten Fehler nicht unbedingt dem subjektiven Bildeindruck entspricht und außerdem stark mit den verwendeten Testbildern schwankt.

Das *Signal-Rausch-Verhältnis* (SNR⁴, siehe z.B. [buck79]) wird nur selten zur Beurteilung computergenerierter Hologramme verwendet. Es ergibt sich aus dem Verhältnis von Signalintensität zur Rauschintensität. Als Rauschintensität wird dabei der mittlere quadratische Fehler verwendet.

$$SNR := \frac{\sum_{f_x, f_y \in B} |f(f_x, f_y)|^2}{\sum_{f_x, f_y \in B} (|f(f_x, f_y)| - \lambda |h(f_x, f_y)|)^2} \quad (5.4)$$

Dabei wird jeweils über den Bereich B der Rekonstruktion summiert und λ gibt denselben Faktor wie in (5.1) an.

⁴=”Signal-to-Noise-Ratio”

Die Definition ohne diesen Faktor λ macht — aus den oben geschilderten Gründen — weniger Sinn, wird aber teilweise dennoch verwendet (z.B. [dall80] 3.155, [farh87]).

Für einige Anwendungen kann der *Kontrast*

$$K = \frac{h_{max}^2 - h_{min}^2}{h_{max}^2 + h_{min}^2} \quad (5.5)$$

wichtig sein. So ist es für meßtechnische Projektionsaufgaben ein Vorteil, wenn möglichst kontrastreiche Muster projiziert werden können.

5.3.2 Beugungseffizienz

Mit *Beugungseffizienz* wird das Verhältnis der in eine bestimmte Beugungsordnung gelenkten Intensität zur einfallenden Gesamtintensität bezeichnet. Bestimmte Algorithmen zur Erzeugung computergenerierter Hologramme optimieren nur Teilbereiche innerhalb einer Beugungsordnung. In diesem Fall ist dieser Teilbereich B für die Definition der Beugungseffizienz von Belang.

$$\eta = \frac{1}{\kappa} \int_{f_x, f_y \in B} |h(f_x, f_y)|^2 dx dy \quad (5.6)$$

κ ist dabei eine Normierungskonstante, die von der auf das Hologramm einfallenden Intensität abhängig ist.

Offensichtlich macht es keinen Sinn, wenn die einfallende Intensität in ein bestimmtes Gebiet gelenkt wird, in diesem Gebiet aber nicht das gewünschte Objekt rekonstruiert. In diesem Sinn macht die Angabe einer Beugungseffizienz ohne Angabe des Rekonstruktionsfehlers (s.o.) keinen Sinn. Aber auch die Angabe beider Parameter ist nicht unproblematisch. Einzelne falsche Intensitätsmaxima können durchaus den Fehler in einem sinnvollen Rahmen belassen, die Beugungseffizienz aber unatürlich überhöhen. Teilweise wird daher zur Bestimmung der Beugungseffizienz in (5.6) der Bereich B , über den summiert wird, nur als der Bereich gewählt, der einer Objektamplitude entspricht, die einen bestimmten Schwellwert übersteigt. In unserem ITO-Testbild (s. Abb. 5.3) würde dieser Bereich dem Dreieck und den enthaltenen Buchstaben entsprechen. Da diese Definition von der Strahlformungsgruppe am Institut verwendet wird, werde ich mich ihr anschließen.

Computergenerierte Hologramme werden besonders ausgiebig zur Strahlformung von leistungsstarken Lasern verwendet. Für solche Anwendungen ist es entscheidend, daß möglichst viel der Laserenergie in das gewünschte Zielgebiet gelenkt wird. Die Bedeutung dieser Anwendung erklärt warum die Optimierung von CGHs hinsichtlich der Beugungseffizienz in der Fachliteratur einen breiten Raum einnimmt.

Für unsere Anwendungen ist die Beugungseffizienz aus mehreren Gründen zweitrangig. Zunächst ist festzustellen, daß für einfache Projektionsaufgaben bereits ein kleiner 5mW HeNe-Laser stark abgeschwächt werden muß, um übliche Detektoren nicht zu übersteuern. Desweiteren wird die effektive Beugungseffizienz bei der Verwendung von LCDs als Hologrammedien primär durch das LCD selbst und nicht durch den CGH-Berechnungsalgorithmus bestimmt. Mehrere Punkte sind hierfür verantwortlich:

- Nur ein Teil des LCD-Pixels (oft nur ca. 25 Prozent) ist lichtdurchlässig

- Polarisatoren vor und hinter dem LCD schwächen die Lichtintensität weiter ab
- Die Mikrostruktur der Pixel führt bei der Beugung zu höheren Beugungsordnungen, in denen ein Teil der Intensität verloren geht

Die Mikrostruktur der einzelnen Pixel bedingt natürlich nicht nur höhere Beugungsordnungen, sondern führt auch zu einer Intensitätsvariation innerhalb der einzelnen Beugungsordnungen (z.B. bei rechteckigen Pixelaperturen zu einer sinc-förmigen Intensitätsvariation). Diese Effekte können — sofern sie stören — bei Kenntnis der Pixelgeometrie auf einfache Weise dadurch vermieden werden, daß das im Hologramm zu kodierende Objekt entsprechend entgegengesetzt gewichtet wird (im Beispiel mit rechteckigen Aperturen wird das Objekt durch die entsprechende sinc-Funktion dividiert). In der Praxis ist dies meist unnötig, da die Effekte im Vergleich zu anderen Störungen vernachlässigbar sind.

Die entsprechenden Störterme in horizontaler bzw. vertikaler Richtung werden als *Breitenstörterm* und *Höhenstörterm* bezeichnet. Eine genaue Analyse für verschiedene Kodierungen, bei denen die Aperturen künstlich auf dem Medium angebracht werden (siehe Kapitel 6), findet man bei Schreier ([schr84]).

Aus den genannten Gründen ist die Beugungseffizienz für uns hier kein Optimierungsziel und wurde auch nicht gemessen. Einige theoretische Werte für maximal erreichbare Beugungseffizienzen finden sich allerdings in der folgenden Tabelle (siehe [schr84] Kap. 2.2.7.).

diffraktive Struktur	Beugungseffizienz
kontinuierliche Amplitudengitter	6 %
kontinuierliche Phasengitter	34 %
binäre Amplitudengitter	10 %
binäre Phasengitter	41 %
geblazte Phasengitter (und allgemeine Phasenhologramme)	100 %

5.3.3 Testbilder und Wahl der Verfahren

Die Qualität eines Verfahrens zur Hologrammberechnung hängt stark vom zu kodierenden Objekt ab. Ich werde mich hier ausschließlich auf reine Intensitätsobjekte beschränken. Folgende Parameter bestimmen weiter ganz entscheidend die Rekonstruktionsqualität des Hologramms:

- wieviele Graustufen enthält das Objekt (ist es binär)?
- Spektrum des Objekts (viele feine, hochfrequente Details ?)
- Bereichsgröße und Bereichsposition des Objekts
- Eigenschaften des Hologrammediums

Je nach Anwendung kommt der Wahl der Testbilder eine große Bedeutung zu. Die Testbilder sollten der konkreten Anwendung möglichst nahe kommen. Die bekannten IEEE-Testbilder⁵, die insbesondere in der Bildverarbeitung eingesetzt werden, oder das "USAF

⁵<ftp://nic.funet.fi/pub/graphics/misc/test-images/INDEX>

1951 tribar resolution test target”⁶, sind für die meisten Anwendungen und den Stand der LCD-Technik viel zu kompliziert.

Im Rahmen des SYMAS-Projekts sollen binäre Strukturen projiziert werden. Als übliche Testmuster werden daher ebenfalls binäre Objekte verwendet. Es sind dies in erster Linie die in Abbildung 5.3 abgebildeten Strukturen, also das ITO-Institutslogo und ein Textausschnitt zur Untersuchung feinerer Details.

Als Testbild für komplizierte Objekte wurde ein Portrait von Denis Gabor verwendet.

Wenn nichts anderes angegeben ist, wurden diese Bilder für ein Hologrammedium mit einer Auflösung von 128×128 Pixeln bei 256 Phasenstufen verwendet.

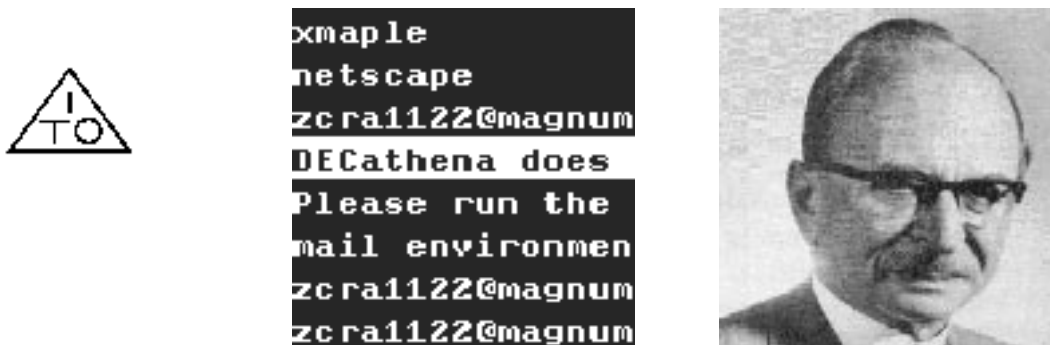


Abbildung 5.3: Verwendete Testmuster

Die Quantisierung auf 256 Phasenstufen orientiert sich dabei an dem von uns zunächst verwendeten videoadressierten LCD (siehe [schw96]). Die Auflösung von 128×128 Pixeln erlaubt auf neueren LCDs noch eine Parkettierung oder den Ausgleich von Diskretisierungsfehlern. Außerdem ermöglicht sie — im Vergleich zu höheren Auflösungen — eine schnellere Berechnung.

Ergebnisse, die unter diesen Simulationsbedingungen gewonnen wurden, lassen sich *nicht* direkt auf andere Bedingungen (andere Bilder, andere Hologrammedien) übertragen!

Konkret heißt das, daß folgende Schritte durchgeführt werden müssen, wenn der Einsatz computergenerierter Hologramme für eine konkrete Anwendung geplant ist:

1. exakte Charakterisierung des Hologrammediums
2. Wahl von geeigneten Testbildern
3. grober Test aller in Frage kommenden Algorithmen
4. Feinabstimmung des favorisierten Algorithmus

Nur mit dieser Vorgehensweise läßt sich eine optimale Qualität erreichen. Leider ist diese Vorgehensweise sehr zeitaufwendig. Eine vollautomatische Durchführung des Verfahrens im Sinne eines “Expertensystems” ist denkbar, aber schwierig.

Die möglichen Verfahren zur Berechnung sind wegen der Kombinationsmöglichkeiten und freien Parameter der Algorithmen praktisch unbegrenzt. Am Ende dieser Arbeit werde

⁶<http://www.cs.berkeley.edu/~qtluong/photography/35mm/test.html>

ich zumindest grobe Orientierungshilfen zur Wahl der Verfahren anbieten, die Wahl eines Verfahrens erleichtern sollen.

5.4 Überblick CGH-Berechnungen

Bevor in den weiteren Kapiteln detailliert auf unterschiedliche Algorithmen zur Berechnung von CGHs eingegangen wird, sollen hier einige allgemeingültige Aussagen vorangestellt werden.

CGHs sind so zu berechnen, daß das Rekonstruktionsergebnis möglichst “optimal” ist. Die Beurteilung des Rekonstruktionsergebnisses erfolgt dabei anhand der in den letzten Abschnitten besprochenen Größen. Für die angestrebten Anwendungen ist der mittlere quadratische Fehler die wichtigste Beurteilungsgröße.

Zur Optimierung der wie auch immer definierten Qualität bieten sich bei gegebenem Hologrammedium insbesondere zwei Freiheitsgrade an:

- Objektphase⁷
- Bildbereiche außerhalb des eigentlich interessanten Bereiches (“windowing”)

Insbesondere hier müssen Berechnungsalgorithmen ansetzen, um eine optimale Qualität der Rekonstruktion zu erreichen. Bei der Optimierung sind natürlich die Eigenschaften des Hologrammediums als Nebenbedingungen zu berücksichtigen. Insbesondere sind dies

- begrenzte Hologrammfläche
- begrenzte Auflösung des Hologrammediums (Diskretisierung)
- begrenzte Dynamik des Hologrammediums (Quantisierung)
- meist nur Amplitude *oder* Phase des Hologramms frei beeinflussbar

Die Quantisierung wird oft nach der eigentlichen Berechnung des Hologramms in einem zweiten Schritt vorgenommen.

Ein wichtiger Parameter bei der Auswahl des Berechnungsalgorithmus ist außerdem die verfügbare Rechenzeit, da die sich der Rechenaufwand für die unterschiedlichen Algorithmen stark unterscheidet.

Bei der Berechnung sollte möglichst schon klar sein, wie die berechneten Hologrammwerte später in das Hologrammedium kodiert werden sollen (siehe folgendes Kapitel). Außerdem ist zu beachten, daß eine Vorgehensweise, die für einen bestimmten Typ von Objekten (z.B. einfache, binäre Gebilde) gute Ergebnisse liefert, mit anderen Arten von Objekten (z.B. komplizierte Gebilde mit vielen Graustufen) unter Umständen unbefriedigende Rekonstruktionen ergibt.

Insgesamt müssen bei der Herstellung eines computergenerierten Hologramms folgende Schritte durchgeführt werden:

⁷Die Objektphase ist natürlich nur dann ein Freiheitsgrad wenn die Intensität und nicht die komplexe Amplitude optimiert wird

- Festlegung des zu kodierenden Objekts, insbesondere der Objektphase
- Berechnung der komplexen Lichtamplitude in der Hologrammebene anhand der komplexen Objektamplitude
- Kodierung der komplexen Lichtamplitude in ein dem Hologrammedium entsprechendes Format
- praktische Umsetzung der Information in das Hologrammedium

Die ersten drei Punkte lassen sich nicht immer voneinander trennen. Viele (insbesondere moderne) Algorithmen zur Berechnung computergenerierter Hologramme führen die drei (oder zumindest die ersten beiden) Schritte gleichzeitig durch. Trotzdem werde ich mich zunächst an dieser Grobeinteilung orientieren und mich getrennt um die Kodierung und Wahl der Objektphase kümmern. Die zunächst getrennte Betrachtung ist didaktisch vorteilhaft und für die Beschreibung älterer Hologrammberechnungsverfahren sinnvoll.

Die Berechnung der Lichtamplitude in der Hologrammebene wurde bereits ausführlich in den vorangegangenen Kapiteln dargestellt und ist im wesentlichen durch diskrete Fourier-Transformationen zu erledigen. Auf den vierten Punkt, die Verwendung der LCDs als Hologrammedien, wird im zweiten Teil dieser Arbeit eingegangen werden.

Kapitel 6

CGH–Kodierungen

Bei der einfachen Berechnung eines Hologramms (z.B. mittels des in Abbildung 5.1 angegebenen Algorithmus) ergibt sich eine komplexe Amplitude in der Hologrammebene¹. Es stellt sich nun die Frage, wie diese komplexe Amplitude in einem reellen Hologrammedium gespeichert werden soll². Eine Art der Kodierung haben wir schon in 5.1 kennengelernt. Es handelte sich dabei um eine punktorientierte Kodierung. Jedem berechneten Punkt in der Hologrammebene wird ein Punkt auf dem Hologrammedium zugeordnet.

Eng verknüpft mit der Kodierung ist die *Quantisierung*. Das Hologrammedium ist nicht nur dahingehend begrenzt, daß es keine komplexen Amplituden speichern kann, sondern es kann oft auch nur *quantisierte* Amplituden oder Phasen speichern. Dies ist natürlich insbesondere bei LCDs von Belang. Die Quantisierung kann als eine Sonderform der Kodierung betrachtet werden. Ich werde ihr aber wegen der besonderen Bedeutung und der Komplexität der Quantisierungsverfahren später ein eigenes Kapitel widmen.

6.1 Weitere punktorientierte Kodierungen

6.1.1 Kodierung nach Burch

Die Kodierung nach Burch ist eine der gebräuchlichsten Kodierungen für Amplitudenhologramme überhaupt (siehe [lee78] Kap. 2.2.1., [farh87]).

Sie ist sehr einfach zu implementieren und lehnt sich direkt an die physikalische Hologrammaufzeichnung an. Sie wird auch heute noch in der Praxis verwendet (siehe z.B. [luce92]).

Kodiert werden soll $A(u, v)e^{i\phi(u, v)}$. Die Amplitudentransmission $H(u, v)$ des Hologramms ist dann gegeben durch

$$H(u, v) = \frac{1}{2}(1 + A(u, v) \cos[(2\pi\alpha u - \phi(u, v))]) \quad (6.1)$$

α gibt dabei den Winkel zwischen der Referenz- und Objektwelle — also die Trägerfrequenz

¹Man beachte, daß die Fouriertransformation einer Funktion in der Regel komplex sein wird (selbst wenn die Eingangsfunktion reell ist).

²In späteren Kapiteln werden Algorithmen vorgestellt, die direkt reelle Amplituden liefern und so dieses Kodierungsproblem umgehen.

— an. Für on-axis Hologramme wählt man diesen Winkel zu Null. In diesem Fall genügt es also, einfach den Realteil des komplexen Lichtfelds zur Kodierung zu verwenden.

6.1.2 Kodierung nach Huang und Prasada

Bei der Kodierung nach Huang und Prassada handelt es sich um eine Variation der Kodierung nach Burch, die einen verbesserten Kontrast erzielen soll (siehe [lee78] Kap. 2.2.2.). Diese Verbesserung wird allerdings mit einer starken Anfälligkeit der Kodierung hinsichtlich eventuellen Nichtlinearitäten des Mediums erkauft.

$$H(u, v) = \frac{A(u, v)}{2}(1 + \cos[(2\pi\alpha u - \phi(u, v))]) \quad (6.2)$$

Insgesamt waren in unseren Simulationen die Ergebnisse mit dieser Kodierung durchweg schlechter als mit der einfachen Kodierung von Burch. Das deckt sich mit der Tatsache, daß die Huang—Prasada—Kodierung in keiner neueren Arbeit zu CGHs verwendet wird.

6.2 Detour-Phase Hologramme

Die ersten computergenerierten Hologramme waren Detour-Phasen Hologramme (siehe [brow66], [lohm67]). Sie gehören zur Klasse der *zellenorientierten Hologramme*. Jede komplexe Amplitude in der Hologrammebene wird dabei von einer relativ großen und recht komplizierten Zelle beschrieben. Eine Zelle besteht dabei aus mehreren Pixeln.

Während bei punktorientierten Hologrammen die $N \times M$ ansteuerbaren Hologrammelemente direkt zur Kodierung von $N \times M$ Amplituden in der Hologrammebene genutzt werden, werden bei zellenbasierten Hologrammen lediglich $N \times M/m$ Amplituden kodiert. Dabei wurde mit m die Anzahl der Pixel pro Zelle bezeichnet.

In der Regel wird ein zellenbasiertes Hologramm dementsprechend eine wesentlich schlechtere Qualität als das punktorientierte Hologramm aufweisen, weil eben viel weniger komplexe Amplituden in der Hologrammebene gespeichert werden.

Warum kann trotzdem ein zellenbasiertes Hologramm eine gute Wahl für eine praktische CGH-Anwendung darstellen? Da nur $N \times M/m$ Amplituden kodiert werden, müssen (bei den meisten Kodierungsmethoden) auch nur $N \times M/m$ Amplituden berechnet werden. Der Rechenaufwand reduziert sich also ganz erheblich. Viele Hologrammedien (z.B. Laserprinter mit nachfolgender Photoreduktion) bieten 10.000×10.000 oder mehr ansteuerbare Pixel. Die Berechnung eines entsprechenden Hologramms ist in der Praxis extrem zeit- und speicherplatzintensiv. Hier bieten sich zellenbasierte Hologramme an.

Bei der Verwendung von LCDs als Hologrammedien ergibt sich dieses Problem allerdings kaum. Ganz im Gegenteil wünscht man sich dort mehr ansteuerbare Pixel. Detour-Phasen Hologramme eignen sich daher für den Einsatz im Zusammenhang mit LCDs nicht. Dementsprechend werden hier nur kurz und beispielhaft einige Verfahren angesprochen.

Bei der Wahl einer zellenbasierten Kodierung ist darauf zu achten, daß die Zellen möglichst einfach auf das Hologrammedium ausgegeben werden können.

Weitere, im folgenden nicht aufgeführte Kodierungen und interessante Abwandlungen findet man bei Urquhart ([urqu91]), Schreier ([schr84]), Cohn ([cohn94]). Bei Farhoosh et al. ([farh87]) und Allebach ([alle81]) findet sich ein Vergleich der wichtigsten Kodierungsmethoden.

Insgesamt gibt es eine sehr große Zahl möglicher Kodierungen, die alle heute nur noch Bedeutung für Ausgabemedien mit sehr großen Orts-Bandbreite-Produkten haben und daher hier nicht dargestellt werden. Allein für Amplitudenhologramme führt Schreier 19 verschiedene Grundvarianten auf. Dort finden sich auch ausführliche Besprechungen der Eigenschaften und Fehler dieser Kodierungen und weitere Literaturhinweise.

6.2.1 Lohmann-Kodierung

Innerhalb einer Zelle wird eine (rechteckige) Apertur angebracht. Die Größe der Apertur kontrolliert die Amplitude. Die Lage der Apertur in der Zelle bestimmt die Phase der Zelle. Dies ist unmittelbar einsichtig, da eine Verschiebung mit folgender Fourier-Transformation einer Phasenschiebung entspricht. In Abbildung 6.1 sind einige Beispielzellen dargestellt. Dieser Grundgedanke der Phasenschiebung, der allen Detour-Phasen-Hologrammen gemein ist, geht direkt auf Lord Rayleigh zurück und wird auch als “Detour-Phasen-Effekt” bezeichnet.

Offensichtlich können sich bei einer solchen Kodierung Probleme ergeben, da die Aperturen nebeneinanderliegender Zellen sich unter Umständen überlappen (“*gap-and-overlap problem*”).

Die Amplitude kann sowohl über die Aperturbreite, als auch über die Aperturhöhe kontrolliert werden.

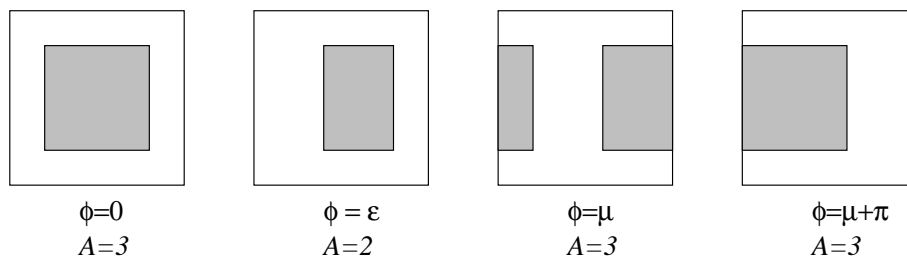


Abbildung 6.1: Lohmann-Kodierung: verschiedene Beispielzellen

Eine genaue Analyse der Kodierung und Methoden zur Vermeidung des *gap-and-overlap* Problems findet man bei Schreier ([schr84]) und Dallas ([dall80]).

Lohmann-Hologramme sind binär. Die im folgenden Abschnitt beschriebene Lee-Kodierung benötigt eine wesentlich geringere Auflösung des Hologrammediums, erfordert aber Graustufen.

6.2.2 Lee-Kodierung

Nach Lee wird jede Zelle aus drei gleichgroßen Unterzellen zusammengesetzt. Nach demselben Prinzip wie bei Lohmann (*Detour-Phase*) haben die Zellen zueinander eine Phasenschiebung von jeweils $2\pi/3$. In Abbildung 6.2 ist veranschaulicht, wie wir damit auf

einfache Weise eine beliebige komplexe Zahl kodieren können. Die drei Achsen im komplexen Kreisdiagramm entsprechen den drei Unterzellen.

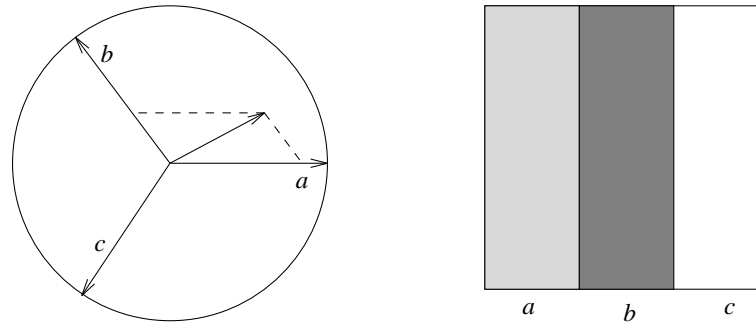


Abbildung 6.2: Lee-Kodierung: komplexe Darstellung und Zellengeometrie

Da die effektive Auflösung des Hologrammediums lediglich um den Faktor drei reduziert wird, schien ein Einsatz mit LCDs denkbar. In der Simulation zeigte sich aber, daß die Verbesserung der Repräsentationsmöglichkeiten (komplexe statt reeller Zahlen) nicht den Nachteil der Auflösungsver schlechterung ausgleichen kann.

6.2.3 Kodierung nach Haskell

Bei der Haskell-Kodierung ([schr84] 4.3.5.2) kann gegenüber einfacheren Kodierungen wie der Lohmann-Kodierung die Qualität etwas verbessert werden. Dazu werden in einem ersten Schritt Zellen (bestehend aus k Unterzellen) berechnet und in einer Tabelle abgelegt. Da es insgesamt 2^k verschiedene Zellen gibt, wird man in der Regel nicht alle möglichen Unterzellenkombinationen berücksichtigen. Stattdessen wählt man die Zellen, die den Bereich der nötigen komplexen Amplituden möglichst gut abdecken. Bei der eigentlichen Kodierung des Hologramms wird dann einfach für jeden komplexen Wert, der kodiert werden soll über die angelegte Tabelle die entsprechende Zelle verwendet. In Abbildung 6.3 ist beispielhaft eine Zelle dargestellt.

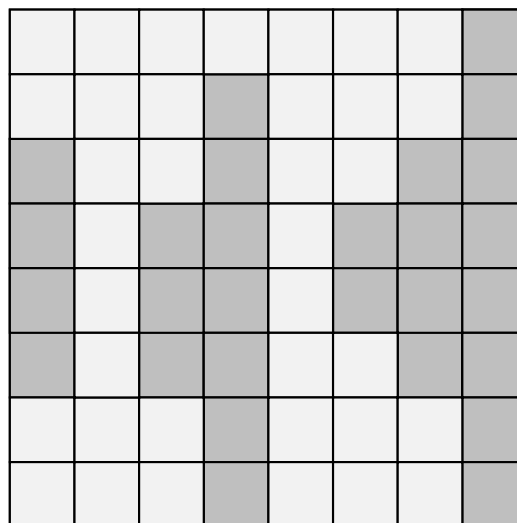


Abbildung 6.3: Haskell-Kodierung: eine Beispielizele

Kapitel 7

Diffusoren

Diffusoren dienen zur Festlegung der Objektphase. Über die Wahl dieser Objektphase ist die Qualität der Hologrammrekonstruktion ganz wesentlich zu beeinflussen. Eng verknüpft mit der Objektphase ist die Bandbreite des Objekts und damit Störungen der Hologrammrekonstruktion aufgrund von Speckles.

In diesem Kapitel werden einige einfache Diffusoren beschrieben. Die rechenintensiven, iterativen CGH-Algorithmien, die in späteren Kapiteln besprochen werden, legen ebenfalls die Objektphase fest und können dementsprechend ebenfalls als Diffusoren betrachtet werden. Diese Algorithmen werden aber in separaten Kapiteln besprochen.

Insgesamt existieren sehr viele “einfache” Diffusoren. Allgemein kann aber gesagt werden, daß die nicht-iterativen Hologrammberechnungen mit diesen Diffusoren zwar sehr schnell sind, aber auch nicht die Rekonstruktionsqualitäten iterativer Algorithmen erreichen. Einfache Diffusoren spielten vor allem in den Anfangsjahren der digitalen Holographie eine bedeutende Rolle. Heute sind sie praktisch vollständig durch iterative Verfahren ersetzt. Aus diesem Grund beschränke ich mich hier auch auf eine Besprechung der wichtigsten Diffusoren.

Ein idealer Diffusor sollte die folgenden drei Anforderungen erfüllen:

- Reduzierung der Dynamik des Objektspektrums (und damit des Hologramms)
- Verminderung von Speckles
- Verteilung der Objektinformation auf möglichst dem gesamten Hologramm

Die erste und die dritte Aufgabe sind eng miteinander verwandt. Wenn die in einem typischen Spektrum auftretenden scharfen Peaks vermieden werden und das Spektrum annähernd konstante Amplitude besitzt (“*spectrum levelling*”), dann wird auch automatisch die Objektinformation auf dem gesamten Hologramm gespeichert sein.

Grundsätzlich unterscheidet man zwischen *deterministischen Diffusoren* und *stochastischen Diffusoren*. Die deterministischen Diffusoren werden weiter eingeteilt in *allgemeine Diffusoren* (*general-purpose Diffusors*) und *bildabhängige Diffusoren* (*image-specific Diffusors*).

7.1 Speckles

Typischerweise entstehen Speckles bei kohärenter Beleuchtung einer optisch rauhen Oberfläche (für eine detaillierte Beschreibung siehe z.B. [tizi88], [gabo70]). Dabei muß die Apertur des abbildenden Systems (z.B. Auge) bzw. die kohärente Beleuchtung begrenzt sein. Tiefere Ursache für die Speckles ist die Bandbegrenzung (Apertur) in Verbindung mit einem nichtbegrenzten Spektrum (weißes Rauschen hat idealerweise ein konstantes, also völlig unbegrenztes Spektrum). Genau dieser Effekt tritt auch bei CGHs auf. Das Spektrum des zu rekonstruierenden Objekts, das im Hologramm aufgezeichnet werden soll, wird in der Regel die verfügbare Hologrammfläche überschreiten. Das Hologramm wirkt als Tiefpaß, genau wie die abbildende Apertur bei “herkömmlichem” Speckle. Dementsprechend ergeben sich auch bei CGHs Speckles (siehe [alle76]).

Man kann das Problem auch von einer anderen Seite betrachten: Ein nicht-bandbegrenztes Signal wird diskretisiert und abgetastet. Bei der Rekonstruktion über die Fourier-Transformation wird das Signal an den Stützstellen, an denen ursprünglich abgetastet worden war, korrekt rekonstruiert. Zwischen diesen Stützstellen, ergibt sich aber ein Signalverlauf, der mit dem Originalsignal nichts mehr gemein hat. Er entspricht nur dann dem Originalsignal, wenn dieses — entsprechend der Abtastfrequenz — bandbegrenzt war.

Dieser in der elektrischen Signalanalyse wohlbekannte Effekt (deswegen müssen beispielsweise akustische Signale vor einer Digitalisierung (z.B. CD-Recording) durch einen Tiefpaß gefiltert werden) wird dort “Aliasing” genannt. Wir haben ihn im Abschnitt 3.2 im Zusammenhang mit diskreten Fourier-Transformationen bereits angesprochen. Er ist völlig gleichbedeutend mit dem Speckle-Phänomen bei computergenerierten Hologrammen.

Wie können die Speckles vermindert werden? Zunächst ist festzustellen, daß das Problem um so größer ist, je größer die Bandbreite des im Hologramm aufzuzeichnenden Objekts ist. Da wir die Phase des Objekts frei wählen können, ist dies der natürliche Ansatzpunkt, um die Bandbreite und damit auch die Speckles zu reduzieren. Ich werde darauf insbesondere in den Abschnitten 7.2 und 8.1.5 eingehen.

Es existieren weitere — indirekte — Methoden zur Speckle-Reduktion, die im folgenden kurz besprochen werden. Insgesamt stehen folgende Möglichkeiten zur Verfügung:

- Bandbegrenzung des Objekts
- Auswertung der Rekonstruktion nur an diskreten Stellen (Stützstellen)
- Parkettierung
- Statistische Mittelung

Die Bandbegrenzung des Objekts wird durch spezielle Algorithmen zur CGH-Berechnung verwirklicht (s.u.).

7.1.1 Auswertung der Rekonstruktion an diskreten Stellen

Bei der *Auswertung der Rekonstruktion an diskreten Stellen* macht man sich den oben beschriebenen Effekt, daß die Speckles nur zwischen den Stützstellen auftreten, zunutze. Eine automatische Auswertung kann bei bestimmten Anwendungen zunächst die Position

dieser Stützstellen ermitteln (Referenzmuster) und dann bei weiteren Messungen nur noch die Intensität an diesen Stellen berücksichtigen. Alternativ kann vor der Rekonstruktion ein exakt passendes (Größe und Positionierung ist entscheidend) Gitter aufgestellt werden, das alle nicht gewünschten Raumbereiche ausblendet.

7.1.2 Parkettierung

Der Effekt der Ausblendung aller Zwischenpositionen kann auch auf eine vollkommen andere Art erreicht werden. Wenn das ursprüngliche Hologramm periodisch wiederholt wird (*Parkettierung*), dann wird das Spektrum dieses Hologramms, also die Rekonstruktion, einen diskreten Charakter aufweisen. Dies ist eine direkte Folge der Fouriertransformation¹ (siehe Anhang B). Je öfter wir das Hologramm wiederholen, desto diskreter wird die Rekonstruktion sein und desto stärker werden die Speckles reduziert.

Dabei ist allerdings nochmals darauf hinzuweisen, daß die Rekonstruktion dementsprechend von einem periodischen, dunklen Gitter überzogen zu sein scheint. Unter Umständen (Display-Anwendungen) kann dies störend sein. In diesem Fall darf die Parkettierung nicht zu groß gewählt werden. In unseren Experimenten erwies sich eine 4 bis 8 fache Parkettierung als günstig.

Die Parkettierung reduziert die Auflösung des Hologramms. Wenn unser CGH-Medium $N \times M$ ansteuerbare Pixel aufweist und wir m -fach parkettieren wollen, dann reduziert sich die Pixelanzahl des Grundhologramms auf $N \times M/m$. In der Praxis muß man dementsprechend abwägen, ob die Qualitätssteigerung durch Speckle-Reduktion per Parkettierung die Qualitätsminderung durch eine reduzierte Auflösung aufwiegt. Ein positiver Nebeneffekt der Speckle-Reduktion durch Parkettierung bzw. der Auflösungsverminderung ist die deutliche Verringerung der Rechenzeit (die Berechnungsdauer wächst — je nach Algorithmus — mindestens linear mit der Pixelanzahl des Hologramms).

Ob und wie stark parkettiert werden soll, kann nicht allgemein festgelegt werden, sondern orientiert sich stark an der konkreten Anwendung, dem Hologrammedium und dem Objekt. Für viele Anwendungen ist aber die bereits genannte 4-fache Parkettierung ein sinnvoller Anhaltspunkt.

Interessant sind Versuche, das Hologramm rein optisch zu vervielfältigen. In diesem Fall kann das Basishologramm die volle Anzahl ansteuerbarer Pixel verwenden und dementsprechend findet keine Reduktion der Auflösung statt. Neifeld (siehe [neif95]) schlägt — allerdings ohne Details anzugeben — einen Aufbau mit Strahlteilern vor.

Einen besonders interessanten Ansatz haben O'Brien, Wilkinson und Mears ([obri93]) verfolgt. Um das große SBP eines optisch adressierten ferroelektrischen LCDs zu nutzen, haben sie einen Array-Generator² verwendet. Vor dem Array-Generator befindet sich ein elektrisch adressiertes LCD, auf dem das Basis-Hologramm dargestellt wird.

Der Array-Generator vervielfältigt dieses Basishologramm und stellt damit auf dem nachfolgenden optisch adressierten FLCN das Gesamthologramm, das über einen polarisierenden Strahlteiler ausgelesen wird, dar (siehe Abbildung 7.1).

¹Anschaulich: eine periodische Funktion (entspricht unendlich-fache Wiederholung) wird durch eine Fourier-Reihe, also ein diskretes Spektrum, beschrieben

²Im wesentlichen ein Gitter oder besser ein weiteres CGH, das zu einer periodischen Wiederholung des Eingangsmusters führt

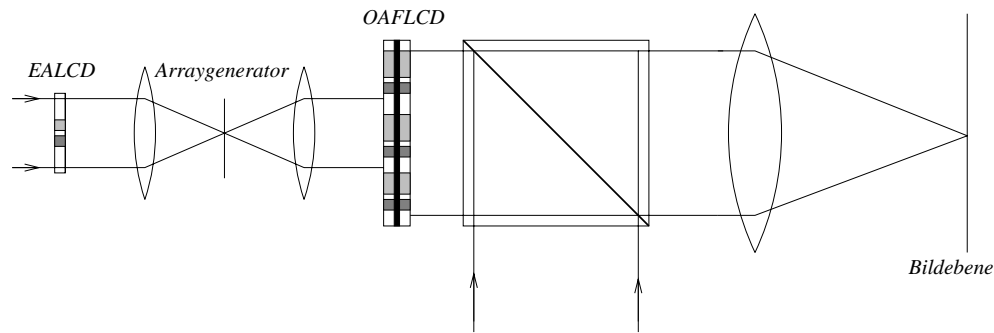


Abbildung 7.1: Einsatz eines optisch adressierten LCDs

Dieses Vorgehen ist sinnvoll, da das optisch adressierte FLCD ein wesentlich höheres Orts-Bandbreiten-Produkt (SBP) als ein elektrisch adressiertes LCD hat. Mit dieser Methode wird das SBP gegenüber dem rein elektrisch adressierten LCD demnach deutlich erhöht.

Nachteile sind die zusätzlichen Kosten für das zweite FLCD und den Arraygenerator.

Ganz entsprechend sollten auch andere Echtzeit-fähige Medien mit hohen SBP, wie z.B. photorefraktive Kristalle, so über ein elektrisch adressiertes LCD beschrieben werden können. Auch das elektrisch adressierte LCD kann durch ein anderes Element (z.B. DMD) ersetzt werden.

Mit entsprechenden Verfahren könnten unter Umständen noch deutliche Qualitätsverbesserungen im Bereich computergenerierter Hologramme ermöglicht werden. In diesem Zusammenhang ist auch eine Modifikation des Array-Generators zu überdenken (z.B. Array-Spots haben unterschiedliche, zufällige Phasen, so daß sich Speckles weiter reduzieren).

Die relativ schlechten Rekonstruktionsqualitäten bei O'Brien sind darauf zurückzuführen, daß der Array-Generator nicht passend zur Größe des Basis-Hologramms gewählt wurde (damit macht die periodische Wiederholung keinen Sinn!).

7.1.3 Trigonometrische Interpolation und das Einschieben von Nullen

Kirk et al. (siehe [kirk94]) verwenden eine andere Vorgehensweise, die sich an unseren Ergebnissen der Simulation von Speckles (siehe unten) anlehnt. Wenn man ohne eine Parkettierung eine annähernd diskrete Rekonstruktion will, kann man einfach das Objekt bereits diskret darstellen. D.h. zwischen jeweils zwei Objektpixeln werden j Nullen eingeschoben. Für dieses neue Objekt aus entsprechend mehr Pixeln wird mit einem der üblichen Hologrammalgorithmen das entsprechende Hologramm berechnet.

Es sei darauf hingewiesen, daß dies für ein ideales Hologrammedium ohne Quantisierung und sonstige simulierbare Fehler (bzw. ein Algorithmus, der kontinuierlich rechnet und erst später quantisiert) keinen Sinn macht! In diesem Fall kann das Einschieben von Nullen nämlich auch direkt durch eine trigonometrische Interpolation in der Hologrammebene mit einem deutlich verringerten Rechenaufwand durchgeführt werden (siehe z.B. [bach92] Kapitel 8.3.).

Der Rechenaufwand selbst ist ein weiterer Nachteil dieses Verfahrens. Während bei der Berechnung von j Hologrammen der Größe $N \times M$ lediglich j mal die Berechnungszeit

des Basishologramms zu Buche schlägt, ist bei der Berechnung eines $jN \times jM$ großen Hologramms ein wesentlich größerer, je nach Berechnungsalgorithmus eventuell sogar exponentiell wachsender, Aufwand nötig.

Extrem macht sich dies bei den von Kirk verwendeten DBS Methoden (ein 256x256-Hologramme statt vier 64x64-Hologramme) bemerkbar.

7.1.4 Statistische Mittelung

Die Reduktion von Speckles im Sinne einer statistischen Mittelung ist ein seit langer Zeit bekanntes Verfahren (siehe z.B. [tizi88] Kap. 7, [iwai96]). Insbesondere in der “konventionellen” Holographie und Interferometrie nutzt man rotierende Diffusoren. Bei einer Änderung der Objektphase ändert sich das Specklemuster in der Rekonstruktion. Mittelt man nun mehrere Rekonstruktionen, so wird das Specklemuster an Kontrast verlieren. In der konventionellen Holographie wird — über die schnell rotierenden Diffusoren — die begrenzte zeitliche Auflösung des Sensors (z.B. Auge) zur Mittelung genutzt.

Da die Specklemuster zwischen den Stützstellen einen zufälligen Charakter haben und sich bei einer Änderung der Objektphase ebenfalls ändern, mitteln sich die registrierten Intensitäten gegenseitig zu einem annähernd konstanten Hintergrund, der zwar den Kontrast herabsetzt, aber trotzdem zu einer verbesserten Gesamtqualität führt.

Entsprechend können wir die Speckles reduzieren, indem wir verschiedene CGHs mit jeweils unterschiedlicher Objektphase rekonstruieren und die Rekonstruktionen summieren. Das Orts-Bandbreiten-Produkt des Hologrammediums wird dabei vergrößert, indem die Anzahl auswertbarer Hologramme pro Sekunde vermindert wird. Die erzielten Ergebnisse mit dieser einfachen Methode waren befriedigend, werden hier aber aus Platzgründen nicht näher dargestellt, weil es natürlich unser Ziel ist, die Echtzeitfähigkeit der Hologramme möglichst gut zu erhalten. Für bestimmte Anwendungen kann die Mittelung sinnvoll sein und sollte daher im Auge behalten werden.

Auch eine Verbindung mit der Parkettierung ist möglich. Statt ein Hologramm mehrmals zu wiederholen, parkettiert man die Hologrammfläche mit verschiedenen Hologrammen, die zwar dieselbe Objektamplitude, aber eben verschiedene Objektphasen rekonstruieren (siehe [dall80] Kap. 6.4.2, [kirk94]).

7.1.5 Simulation der Speckles

Die Speckles kommen durch die Bandbegrenzung des Hologramms zustande und treten zwischen den eigentlichen Abtastpunkten auf. Wenn wir uns — in der Simulation — für diese Speckles interessieren, müssen wir lediglich die Rekonstruktion des Hologramms zwischen den eigentlichen Abtastpunkten berechnen.

Dies kann einfach dadurch erfolgen, daß wir das Hologramm vor der Rekonstruktion in einen Bereich aus Nullen einbetten (“*zero-padding*”) und dann die Rekonstruktion ausführen. Dieses Ergebnis ergibt sich direkt aus der mathematischen Theorie der diskreten Fourier-Transformation (siehe z.B. [bach92] Kap. 8.3), ist aber auch anschaulich verständlich da ja durch die umgebenden Nullen genau das physikalische System (Apertur \rightarrow außerhalb ist die Amplitude Null) simuliert wird. Je größer man den Bereich der Nullen wählt, desto mehr Zwischenstellen können betrachtet werden und um so klarer treten die Speckles hervor. Für

eine realistische Qualitätsabschätzung reicht aber bereits eine Einbettung des Hologramms in einen 4-mal so großen Bereich aus Nullen (2-fach zero-padding).

7.2 Random-Diffusor

Der einfachste und gleichzeitig grundlegendste Diffusor ist der *Random-Diffusor*. Die Objektphase wird rein zufällig (in der Regel weißes Rauschen) gewählt. Damit wird die Dynamik des Fourier-Spektrums, und damit des Hologramms, stark reduziert.

Als „*Dynamik*“ verwende ich mit Dallas (siehe [dal80] 6.118) das *peak-to-average* Verhältnis des Leistungsspektrums. Diese Dynamik sollte vom Hologramm möglichst korrekt gespeichert werden können, das heißt, sie sollte kleiner oder gleich der Dynamik des Hologrammediums sein.

Was macht man aber, wenn dies nicht der Fall ist? Offensichtlich muß ein Teil der Information verworfen werden und die Rekonstruktionsqualität nimmt ab. Üblicherweise wird das Spektrum einfach auf die Dynamik des Hologrammediums skaliert. Leider wird damit in vielen Fällen die Beugungseffizienz stark reduziert. Eine Verbesserung ergibt sich durch ein Abschneiden sehr großer Werte vor der Skalierung (*Top-Clipping*). Dazu werden alle Amplituden des Spektrums, die einen bestimmten Wert übersteigen auf diesen maximalen Wert gesetzt. Das Top-Clipping bewirkt indirekt eine Kantenverstärkung in der Rekonstruktion. Das Auftreten der Kantenverstärkung ist nicht verwunderlich, da bei üblichen Objekten insbesondere tieffrequente Terme (nullte Ordnung!) deutlich im Spektrum zutage treten und durch das Top-Clipping dann an Gewicht verlieren.

In Kapitel 11 werden noch einige weitere, kompliziertere Verfahren zur Speicherung von Signalen mit großer Dynamik besprochen.

Unmittelbar einsichtig ist aber, daß es günstig ist, wenn die Dynamik des Objektspektrums bereits möglichst klein ist.

Die Verwendung einer Zufallsphase zur Reduktion der Dynamik ist seit langem aus der „konventionellen“ Holographie bekannt (siehe z.B. [caul79] 8.4.) und wurde ursprünglich von Leith und Upatnieks (siehe [leit64]) eingeführt. Eine anschauliche Begründung zur Funktionsweise des Diffusors wurde von Burckhardt (siehe [burc70]) gegeben:

Die Autokorrelation des Objekts (unter Berücksichtigung des Diffusors) ist gegeben durch

$$, (f_x, f_y) = a(f_x, f_y)e^{ir(f_x, f_y)} * a(-f_x, -f_y)e^{-ir(-f_x, -f_y)} \quad (7.1)$$

Dabei ist $r(f_x, f_y)$ die zufällige Phase und $a(f_x, f_y)$ die Amplitude des Objekts. Anschaulich ist klar, daß für $f_x = f_y = 0$ eine große Autokorrelation zu erwarten ist, während alle anderen Autokorrelationswerte annähernd 0 ergeben sollten (die zufällige Phase führt ja letztlich zu zufälligen positiven und negativen Amplituden). Also:

$$, (f_x, f_y) \approx \delta(f_x, f_y) \quad (7.2)$$

Die Fouriertransformation der Autokorrelationsfunktion ergibt aber nach dem *Wiener-Khinchine-Theorem* (siehe z.B. [bach92] Kap. 10.3.3) gerade das Leistungsspektrum. Die Fouriertransformation einer Delta-Funktion ergibt eine Konstante, womit anschaulich klar ist, daß die Amplitude in der Hologrammebene annähernd konstant ist.

Eine statistische Analyse des Amplitudenspektrums von Objekten mit zufälliger Phase findet man bei Akahori (siehe [akah73]). Es zeigt sich, daß das Spektrum nicht vollständig konstant ist, sondern zu hohen Raumfrequenzen hin leicht abnimmt und somit ein gewisser Tiefpaßeffekt vorhanden ist.

Nachteil des Random-Diffusors ist, daß das Phasenrauschen zu einer starken Vergrößerung der Bandbreite des Objekts und damit zu Speckles führt. Glücklicherweise ergibt sich ein sehr zufällig aussehendes Specklemuster (man spricht von “black spaghetti stew”), das unter Umständen nicht stark stört. Ein Beispiel ist in Abbildung 7.3 dargestellt. Zum Vergleich ist in Abbildung 7.2 die Rekonstruktion eines Objekts mit konstanter Phase dargestellt. Man sieht die deutliche Verbesserung, die bereits mit dem einfachen Random-Diffusor erzielt werden kann.

Die Simulation der Speckles wurde über ein 2-fach zeropadding (siehe Abschnitt 7.1) durchgeführt.

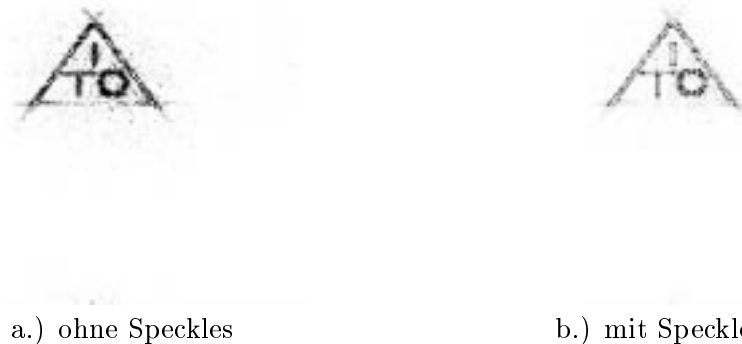


Abbildung 7.2: Speckle-Simulation mit konstanter Objektphase

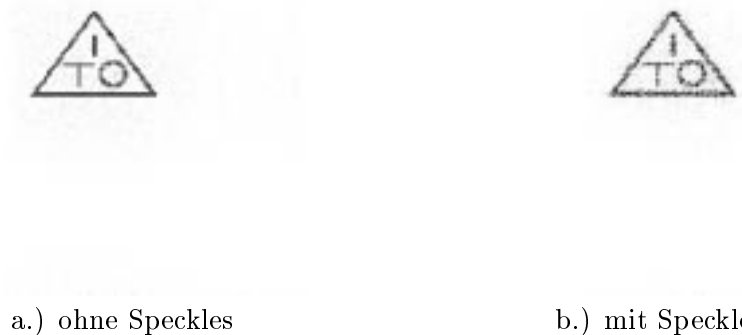


Abbildung 7.3: Speckle-Simulation mit Random-Diffusor

Noch deutlicher ist die Verbesserung bei komplizierten Bildern (siehe Abbildung 7.4).

Abwandlungen des einfachen Random-Diffusors, die zwar zu Hologrammen mit größerer Dynamik führen, die Speckles aber vermindern sollen, sind bekannt (siehe [lohm71]). Dazu wird z.B. lediglich jeder zweite Phasenwert zufällig gewählt. Die Zwischenwerte werden jeweils aus dem Mittel der benachbarten Werte gebildet. In der Simulation ergaben sich allerdings praktisch keine Änderungen im Vergleich zum Random-Diffusor.

Man sollte noch darauf hinweisen, daß die Stärke des Random-Diffusors natürlich einfach variiert werden kann, indem nur Phasenwerte in einem bestimmten Bereich berücksichtigt werden. Dementsprechend ist der Diffusor-Effekt mehr oder weniger “stark”.



a.) konstante Phase



b.) Zufällige Phase

Abbildung 7.4: Random-Diffusor bei komplizierten Objekten

7.3 Radar-Codes und Diffusoren

Zwischen Diffusoren und Radar-Codes zur Impulskompression besteht ein enger Zusammenhang. Im letzten Abschnitt wurde gezeigt, daß eine Phasenverteilung, deren Autokorrelation einer Delta-Funktion entspricht, zu einer günstigen Dynamik in der Hologrammebene führt.

Die Forderung nach einer deltaförmigen Autokorrelation ist aber auch aus der Impulskompression beim Radar bekannt. Dort hat man das Problem, daß die Sendeleistung nicht beliebig gesteigert werden kann. Um trotzdem eine hohe Reichweite bei gutem Signal-Rausch-Verhältnis zu erreichen, bedient man sich des Tricks der *Impulskompression*. Dazu sendet man ein zeitlich ausgedehntes Signal $x(t)$ und filtert das empfangene Signal $y(t)$ mit

$$H(f) = kX^*(f)e^{-i\omega\tau} \quad (7.3)$$

Diese Filter hat die Impulsantwort

$$h(t) = kx(\tau - t) \quad (7.4)$$

Und führt zu einem

$$y'(t) = h(t) * y(t) = kx(t) * x(-t) \quad (7.5)$$

also der Autokorrelation von x . Offensichtlich läßt sich diese dann besonders gut detektieren, wenn y' bzw. die Autokorrelation des gesendeten Signals $x(t)$ eine Deltafunktion ist³.

Bei dieser Parallele nimmt es nicht wunder, daß die eindimensionalen Codes der Radartechnik in zweidimensionaler Form auch als Diffusoren für computergenerierte Hologramme verwendet werden können.

Auf die unterschiedlichen Codes im Radarbereich soll hier nicht weiter eingegangen werden. Der Leser sei auf die reichhaltige Fachliteratur verwiesen.

Im Rahmen dieser Arbeit wurden Schröder- und Heimiller-Codes zur Hologrammberechnung getestet.

³Unter Umständen muß die Geschwindigkeit des zu detektierenden Objekts in die Überlegungen einbezogen werden. In diesem Fall muß dann von der Autokorrelation zur Kreuzkorrelation übergegangen werden

Stellvertretend für die vielen unterschiedlichen Diffusoren sei hier der (bildabhängige) zweidimensionale Schröder-Code angegeben:

$$p(u, v) := \frac{f(u, v)^2}{\sum_{n,m=0}^{N,M} f(n, m)^2} \quad (7.6)$$

$$\phi(u, v) = -2\pi \sum_{n,m=0}^{N,M} \left(\frac{M+N}{2} - n - m \right) p(u, v) + \frac{\pi nm}{NM} \quad (7.7)$$

Dabei ist f_{nm} die (diskrete) Objektamplitude. N bzw. M bezeichnet die Anzahl der Objektpunkte in den beiden Objektdimensionen.



Abbildung 7.5: Rekonstruktion mit zweidimensionalem Schröder-Diffusor

Weitere Diffusoren findet man in [dall80], [chu72], [chu72b], [yone76], [tori78], [lohm71].

7.4 Linsen-Diffusor

Der einfachste deterministische general-purpose Diffusor ist der *Linsen-Diffusor*. Man denkt sich einfach eine defokussierende Linse direkt hinter dem Objekt angebracht. Die Verschmierung durch die Linse führt anschaulich zu einer Reduktion der Dynamik und zu einer (wenn auch nur geringen) Verteilung der Objektinformation. Leider wird allerdings auch die Rekonstruktion leicht unscharf bzw. verschwommen, was sich insbesondere bei feinen Details negativ bemerkbar macht.

In der Praxis wird die Linse durch die Objektphase festgelegt. Man wählt:

$$\phi(u, v) = 2\pi \left(\frac{u^2}{c_x} + \frac{v^2}{c_y} \right) \quad (7.8)$$

Die Konstanten c_x und c_y geben dabei die Stärke des Effekts an.

Auch dieses Verfahren wurde aus der "konventionellen" Holographie übernommen (siehe z.B. [caul79] 8.4.).

Kapitel 8

Iterative–Fourier-Transformations– Algorithmen

Die überwiegende Mehrzahl von Phasen–CGHs wird heute mit *Iterativen Fourier–Transformations Algorithmen (IFTA)* berechnet. Die Algorithmen bieten eine gute Rekonstruktionsqualität bei einigermaßen moderaten Anforderungen an die Rechenkapazität. Schnellere Verfahren — wie z.B. die einfachen Diffusoren aus dem vorangegangenen Kapitel — führen zu deutlichen Qualitätseinbußen während die direkten Suchverfahren und genetischen Algorithmen einen unverhältnismäßig hohen Rechenaufwand erfordern.

Es hat sich im Laufe der Diplomarbeit gezeigt, daß auch für uns diese IFT–Algorithmen — beim momentanen Stand der Computertechnik — das günstigste Verhältnis von Rechenzeit zu Rekonstruktionsqualität liefern. Daher wurde speziell die Klasse der IFT–Algorithmen um eine neue dreidimensionale Variante ergänzt.

8.1 Gerchberg–Saxton–Algorithmus

Der erste IFT–Algorithmus wurde von Gerchberg und Saxton 1972 (siehe [gerc72]) vorgeschlagen. Er wurde zur Bestimmung der Phase eines Lichtfeldes aus den Intensitäten des gebeugten und des ungebeugten Lichts verwendet.

Der Algorithmus läßt sich aber praktisch unverändert auch zur Optimierung von Phasen–CGHs verwenden (siehe [gall73]), da Hologramm und Rekonstruktion ebenfalls über eine Fourier–Transformation miteinander verbunden sind.

Bereits in den Kapiteln 5 und 7 wurde darauf hingewiesen, daß der wichtigste Freiheitsgrad zur Optimierung der Rekonstruktionsqualität von Hologrammen die Objektphase ist. Die verhältnismäßig einfachen Diffusoren aus Kapitel 7 werden nun durch eine iterative Berechnung der optimalen Objektphase ersetzt. Dazu wird ausgehend von einer Startphase (meist zufällige Phasenverteilung) laufend zwischen Bild– und Hologrammebene gewechselt und versucht, die Zwangsbedingungen in den beiden Ebenen möglichst gut zu erfüllen. Zwangsbedingung in der Bildebene ist sicher, daß die Intensität (bzw. die Amplitude) der gewünschten Rekonstruktion entspricht. In der Hologrammebene ist die Zwangsbedingung,

daß die Amplitude konstant sein soll (Phasenhologramm).

In Abbildung 8.1 ist der Algorithmus komplett dargestellt.

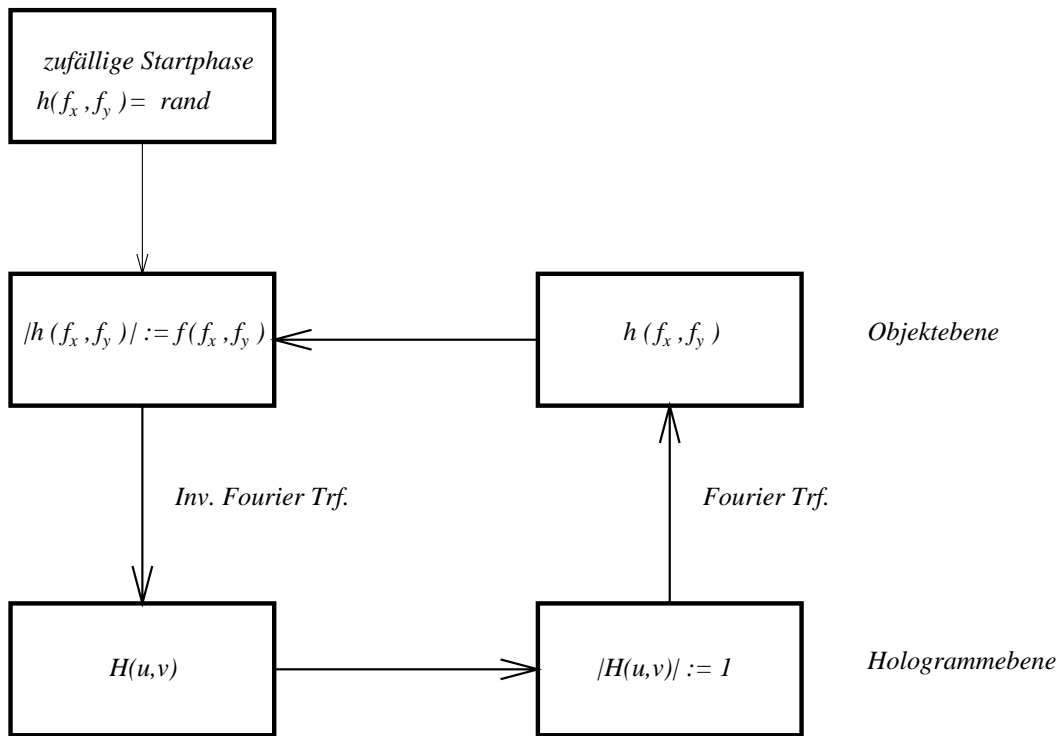


Abbildung 8.1: Gerchberg-Saxton-Algorithmus

Einen Beweis für die Konvergenz des Algorithmus zum Fehler 0 (für praktisch alle Fälle) findet man ebenfalls bei [gerc72]. Es ist aber zu beachten, daß dabei kontinuierlich gerechnet wird. Quantisierungseffekte, wie man sie zwangsläufig bei einem CGH hat, werden nicht berücksichtigt. Das nach obigem Algorithmus berechnete CGH ist dementsprechend nach der nötigen Quantisierung in der Regel nicht das im Sinne eines minimalen Fehlers optimale Hologramm.

Im folgenden werde ich eine kurze Erklärung zur Funktionsweise des Algorithmus angeben bzw. den Zusammenhang zu einem alten mathematischen Verfahren aufzeigen:

Mit \hat{B} und \hat{G} werden die Operatoren des IFT-Algorithmus im Raum- bzw. Frequenzraum-bereich, also

$$\hat{B} |h| e^{i\phi} := |f| e^{i\phi} \quad (8.1)$$

$$\hat{G} H e^{i\theta} := e^{i\theta} \quad (8.2)$$

bezeichnet.

Dabei ist $|f|$ die gewünschte Objektamplitude, die das CGH später rekonstruieren soll, und ϕ die aktuelle Objektphase ($h = |h|e^{i\phi}$). \mathcal{F} bzw. \mathcal{F}^{-1} ist der Operator der Fourier-Transformation bzw. der inversen Fouriertransformation.

Offensichtlich lautet in dieser Schreibweise der IFT-Algorithmus

$$h_{n+1} := \hat{N}h_n = \hat{B}\mathcal{F}^{-1}\hat{G}\mathcal{F}h_n \quad (8.3)$$

Im Falle der Konvergenz ($n \rightarrow \infty$) ergibt sich damit das Eigenwertproblem

$$\hat{N}h_n = kh_n \quad (8.4)$$

Numerisch können solche Eigenwertprobleme mit dem Verfahren von Mises (siehe [enge87] Kapitel 5.3) gelöst werden. Dazu wird ein beliebiger Startvektor h_0 gewählt und laufend die Iterationsvorschrift (8.4) durchgeführt. Für $n \rightarrow \infty$ ergibt sich nach Mises der Eigenvektor des Problems a_n zum betragsgrößten Eigenwert über die Iterationsvorschrift (8.4).

Man sieht also, daß die IFT-Algorithmen im Kern lediglich das seit langer Zeit in der Mathematik bekannte Verfahren von Mises verwenden. Das schmälert die Leistung von Gerchberg und Saxton allerdings nicht, da die Wahl der entsprechenden Operatoren nicht durch das Verfahren selbst gegeben ist und die Konvergenz für einen Satz von Operatoren jeweils noch gezeigt werden muß.

8.1.1 Anzahl der Iterationen

Die Konvergenz ist für übliche Testbilder von 256x256 Pixeln so gut, daß nach ca. 50 Iterationen keine nennenswerten Verbesserungen mehr erreicht werden. Dabei hängt die Konvergenz stark von den konkreten Versuchsbedingungen, also von dem zu kodierenden Objekt und dem Hologrammedium, ab. So kann es im Einzelfall durchaus passieren, daß auch nach mehreren hundert Iterationen noch Verbesserungen möglich sind (siehe Kapitel 16).

Durch leichte Abwandlungen des Basis-Algorithmus kann die Konvergenz weiter verbessert werden (siehe z.B. "Input-Output-Algorithmus" in Abschnitt 9.1, [kot191], [kot194]). Diese Beschleunigung geht allerdings oft auf Kosten der Rekonstruktionsqualität und/oder Beugungseffizienz.

Insbesondere hängt die Konvergenz bei diesen Variationen des Grundalgorithmus teilweise stark von zusätzlichen Parametern innerhalb der Optimierung ab. Ein günstiges Verhalten der Algorithmen ergibt sich nur, wenn diese Parameter "richtig" gewählt werden. Für unterschiedliche Anwendungen sind diese "richtigen" Parameter jeweils erst zu ermitteln, was einen erheblichen Nachteil für die Praxis bedeutet.

Für Strahlformungsaufgaben, bei denen viel Aufwand in ein ganz spezielles CGH gesteckt wird, ist dieser Aufwand zur Ermittlung günstiger Parameter unter Umständen gerechtfertigt. Bei den von uns angestrebten Anwendungen mit vielen verschiedenen CGHs, die auf das zeitlich variable Hologrammedium LCD ausgegeben werden, ist dagegen die Stabilität des Algorithmus eine wichtige Eigenschaft. Die Algorithmen sollen automatisch, ohne die externe Hilfe eines Experten, Ergebnisse liefern. Eigene Experimente mit den von Kotlyar et al. vorgeschlagenen Abwandlungen des IFT-Algorithmus bestätigten die schnellere Konvergenz, aber auch die Probleme bei der Wahl geeigneter Parameter, die oft zu einer Zerstörung der Konvergenz führen¹.

Üblicherweise erzielt der Gerchberg-Saxton-Algorithmus in den ersten 20 bis 50 Iterationen eine deutliche Verbesserung der Bildqualität und Beugungseffizienz. Durch viele weitere Iterationen (ca. 100 bis 500) läßt sich die Qualität meist um einige weitere Prozent steigern. Die Berechnung ist dann aber in einem lokalen Optimum gefangen.

¹insbesondere scheint das Verfahren eher für sehr einfache Objekte (entwickelt wurde der Algorithmus für Strahlformungsaufgaben) brauchbare Ergebnisse zu liefern

Bei einer erneuten kompletten Berechnung mit unterschiedlicher Startverteilung stagniert die Optimierung in einem anderen Extremum und kann dementsprechend eine bessere oder schlechtere Qualität erreichen. Man kann daher z.B. den IFT-Algorithmus 10 mal anwenden und von den 10 berechneten Hologrammen das Günstigste als Lösung des Optimierungsproblems ansehen.

Eine andere Lösung des Stagnation-Problems besteht in einer Änderung der Operatoren in der Objektebene bzw. der Rekonstruktionsebene im Laufe des Algorithmus (siehe z.B. [háje92], [wyro88], [wyro90]). Damit soll ein Entkommen aus lokalen Optima ermöglicht werden. Der hohe Rechenaufwand nach ca. 50 Iterationen in Verbindung mit diesen Verfahren lohnt wegen der nur minimal erzielbaren Verbesserung der Hologrammqualität für die von uns angestrebten Anwendungen nicht. Für Strahlformungsaufgaben, bei denen jedes Prozent der Beugungseffizienz wichtig ist, und die Berechnungsdauer fast ohne Belang ist, sind die Verfahren aber interessant.

Peter et al. (siehe [pete92]) berichten über eine mögliche Verbesserung der Beugungseffizienz wenn eine leichte Abwandlung der üblichen Startverteilung (Phase zufällig, Amplitude = gewünschte Objektamplitude) gewählt wird. Die Verbesserungen sind allerdings im Fall des IFT-Algorithmus nicht signifikant. Dagegen sollen sich deutliche Verbesserungen der Simulated-Annealing (SA) und DBS Algorithmen (siehe Kapitel 10) erreichen lassen. Die dort angeführten schlechten Grundergebnisse der SA und DBS-Algorithmen konnten von mir und anderen allerdings nicht bestätigt werden.

Eine weitere interessante Variation des Gerchberg-Saxton-Algorithmus wurde kürzlich von Rabadi, Myler und Weeks zur Verbesserung der Konvergenz und zur Vermeidung von Stagnationen vorgeschlagen ([raba96]). Die Autoren berichten, daß im Falle der Phasenrestauration eine deutliche Verbesserung hinsichtlich Geschwindigkeit und Endqualität der Iteration erreicht werden kann, wenn man verschiedene Auflösungen im Laufe des Algorithmus benutzt. Das Verfahren wird als *Iterative Multiresolution Algorithm* bezeichnet. Ich schildere das Verfahren anhand der Berechnung computergenerierter Hologramme (obwohl es für diesen Zweck bisher nicht verwendet wurde): Zunächst wird die Auflösung (sowohl des Objekts, als auch des Hologramms) künstlich verringert. Das kann in der Praxis zum Beispiel durch Mittelung über jeweils n Pixel erfolgen. Für diese Hologramm/Objekt-Kombination mit geringer Auflösung wird dann der Gerchberg-Saxton-Algorithmus angewendet. Die Auflösung des sich so ergebenden Hologramms wird nun erhöht, indem Zwischenwerte interpoliert werden (die Autoren verwenden dabei eine einfache lineare Interpolation).

Dieses Hologramm wird dann als Startverteilung für einen erneuten Gerchberg-Saxton-Algorithmus mit höherer Auflösung verwendet. Der gesamte Vorgang wird mehrmals wiederholt, bis die volle ursprüngliche Auflösung erreicht ist.

Die erzielten Verbesserungen führen die Autoren darauf zurück, daß durch die geringeren Auflösungen sichergestellt wird, daß der Algorithmus im Einflußgebiet eines lokalen Minimums landet, das sehr nah am globalen Minimum liegt. Erst in diesem Einflußgebiet macht dann eine weitere hochauflösende Optimierung Sinn.

8.1.2 Optische Implementierung

Noel, Khan und Dhadwal (siehe [noel93]) haben eine optische Implementierung eines modifizierten Gerchberg-Saxton-Algorithmus (allerdings nicht zur CGH-Optimierung, sondern

für Anwendungen im Bereich der “Superresolution”²⁾ realisiert.

Fourier-Transformation und Inverse Fourier-Transformation werden dabei über Linsen gebildet. Das ist allerdings nur zulässig, wenn die Eingangswellenfront jeweils punktsymmetrisch in Amplitude und Phase ist.

Die Iterationen werden über Spiegel durchgeführt und hier liegt auch das Problem der an sich eleganten Methoden: Wenn das Verfahren auf komplizierte Objekte³⁾ wie CGHs übertragen werden soll, ist damit zu rechnen, daß minimale Phasenfehler durch die fortlaufende Iteration immer weiter verstärkt werden und so zu stark unsymmetrischen Wellenfronten führen. Dies ist gleichbedeutend mit einem Versagen des Algorithmus und müßte daher verhindert werden.

Die Probleme mit der Forderung nach Symmetrie bei Verwendung von Linsen zur Berechnung der inversen Fourier-Transformation lassen sich durch phasenkonjugierende Methoden vermeiden.

Lo und Indebetouw (siehe [lo92]) haben, ausgehend von einem allgemeinen System, bestehend aus einem Fabry-Perot-Resonator, bei dem einer der Spiegel durch einen phasenkonjugierenden Spiegel ersetzt wurde, ein entsprechendes System aufgebaut. Das Prinzip der Anordnung ist in Abbildung 8.2 dargestellt.

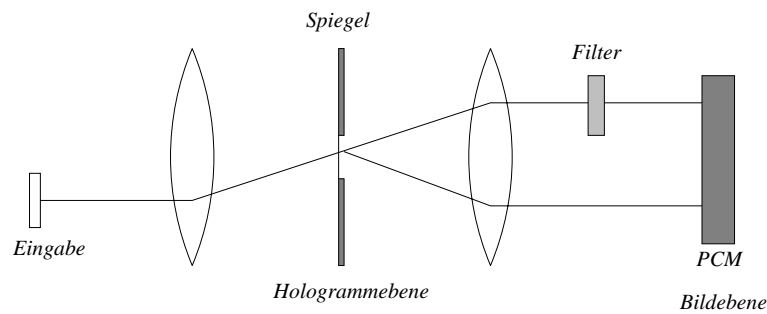


Abbildung 8.2: Optische Implementierung des Gerchberg-Saxton- Algorithmus mittels Phasenkonjugation

Die Anordnung ist ebenfalls zur Auflösungssteigerung gedacht. Der Spiegel entspricht dabei der Hologrammebene. Er blendet (Maske) gleichzeitig hohe Frequenzen aus. Entsprechend wird in der Raumebene (vor dem phasenkonjugierenden Spiegel (PCM)) ein weiteres Filter (in diesem Fall ebenfalls ein einfacher Tiefpaß) angebracht.

Der phasenkonjugierende Spiegel verhindert die Anhäufung der Phasenfehler und sorgt so für ein stabiles Verhalten. Gleichzeitig gleicht er unvermeidbare Intensitäts-Verluste innerhalb der Anordnung aus. Lo und Idebetouw haben die Phasenkonjugation über degenerierte Vierwellen-Mischung in $BaTiO_3$ realisiert.

Obwohl das System für einfache Objekte gute Ergebnisse lieferte, führten den Autoren zufolge die vielen sich zwangsläufig bei der praktischen Realisierung ergebenden Fehler bei komplexen Objekten zu einem unbefriedigenden Systemverhalten.

Die Anordnung müßte für eine Hologrammberechnung dahingehend abgewandelt werden, daß das Filter beim PCM durch die gewünschte Objektmaske ersetzt wird (z.B. über ein

²⁾Auflösungen bis über die beugungsbedingten Auflösungsgrenzen

³⁾Die Autoren haben lediglich einen einfachen Doppelspalt untersucht

LCD). Die Anordnung würde sich dann zur Berechnung von komplexen Hologrammen (Phase und Amplitude werden durch das Hologrammedium beeinflusst) eignen. Diese komplexen Hologramme wollen wir aber vermeiden, da wir zwei LCDs (eines als Phasenmodulator und direkt dahinter ein zweites als Amplitudenmodulator) zu ihrer Realisation benötigen. Abgesehen davon wird die Beugungseffizienz bei komplexen Hologrammen deutlich vermindert⁴ und es ergeben sich in der Praxis eventuell Probleme mit der Justierung der beiden LCDs.

Um ein reines Phasenhologramme zu erhalten, muß in der Hologrammebene entsprechend eine konstante Amplitude hergestellt werden. Dabei darf die Phase nicht beeinflusst werden. Mir ist keine einfache Anordnung bekannt, die selbiges in hoher Geschwindigkeit erledigt⁵

Beide angesprochenen Systeme lassen sich — abgesehen von praktischen Problemen — auf die Berechnung computergenerierter Hologramme übertragen. Ich halte das aber nicht für sinnvoll. Es ist trotz großen technischen Aufwands mit erheblichen Problemen zu rechnen (siehe die wenig beeindruckenden Ergebnisse in den Originalarbeiten der Autoren). Mit geeigneter elektronische Hardware (siehe Abschnitt 15.3.1) läßt sich der Gerchberg-Saxton Algorithmus sehr schnell rein digital (und damit ohne Fehler) durchführen. Die digitale Anordnung ist insgesamt betrachtet billiger, flexibler, weniger empfindlich und benötigt keinerlei Justierung.

Insbesondere beinhaltet die fehlende Flexibilität, daß die optischen Systeme nicht einfach für die Optimierung von 3D-CGHs (s.u.) verwendet werden können.

Ein Ausweichen auf die rein optische Berechnung ist daher nur für Echtzeitanwendungen (z.B. supraauflösende Mikroskope) interessant.

8.1.3 Binarisierung

Auf CGH-Medien können in der Regel nur quantisierte Hologramme dargestellt werden. Im Falle von LCDs als Medien sind in der Regel zwischen 2 und 256 Quantisierungsstufen⁶ verfügbar.

Besonders interessant (preiswerte Displays) ist die Quantisierung auf 2 Graustufen, die sogenannte *Binarisierung*. Man kann mit dem oben beschriebenen Gerchberg-Saxton-Algorithmus zunächst ein CGH berechnen und dann in einem zweiten Schritt dieses Hologramm mit den in Kapitel 11 zu besprechenden Verfahren quantisieren.

Laut Wyrowski (siehe [wyro89a], [broj89]⁷) ergeben sich bessere Ergebnisse, wenn man die

⁴Das wäre allerdings bei den von uns angepeilten Anwendungen durchaus tragbar, da die Rekonstruktionsqualität des komplexen Hologramms wegen des doppelten SBP deutlich verbessert wird.

⁵Eine (komplizierte) Möglichkeit wäre der Einsatz von Smart-Pixeln mit FLC-Schichten zur Phasenmodulation: Auf die Smart-Pixel wird eine zusätzliche Referenzwelle gestrahlt um die Phase des einfallenden Lichts detektieren zu können. Entsprechend wird die FLC-Schicht zur Phasenmodulation angesteuert und kohärent mit konstanter Amplitude bestrahlt. Die Smart-Pixel wären entsprechend dem Zeitverhalten des FLC zu takten. Es ist allerdings mehr als fraglich, ob die praktischen Probleme bei der Justierung (Messung der Phase durch Smart Pixel) gelöst werden können. Eine weitere — leider ebenfalls komplizierte — Möglichkeit wäre eine degenerierte Vierwellenmischung bei der die Signalwelle so stark gewählt wird, daß in jedem Fall ein gesättigtes Beugungsgitter im photorefraktiven Kristall entsteht. Die Referenzwelle (konstante Amplitude) wird dann entsprechend mit konstanter Effizienz gebeugt.

⁶Allerdings ist bei 8 Bit Graustufen = 256 Quantisierungsstufen *effektiv* meist mit einer wesentlich geringeren Quantisierungsanzahl zu rechnen. Die effektive Quantisierung ist aufgrund der relativ großen statistischen Streuung der einzelnen Pixel stark reduziert.

⁷Im entsprechenden Artikel wird eine Abwandlung des Verfahrens zur Quantisierung von Bildern (statt

Quantisierung bereits in den iterativen Algorithmus einbaut. Ein simples *Hard-Clipping*, das heißt Binarisieren nach der Methode

$$C(I) = \begin{cases} 0 & \text{für } I \leq b \\ 1 & \text{für } I > b \end{cases} \quad (8.5)$$

zerstört die Konvergenz des Algorithmus und kann daher nicht verwendet werden. Stattdessen wird eine variable Quantisierung gewählt, die zunächst kaum quantisiert und erst gegen Ende der Optimierung in ein *Hard-Clipping* übergeht. Der Quantisierungs-Operator C hat die Form

$$C_j(I) = \begin{cases} 1 & \text{für } I \geq 1 - g_j \\ I & \text{für } g_j \leq I < 1 - g_j \\ 0 & \text{für } I < g_j \end{cases} \quad (8.6)$$

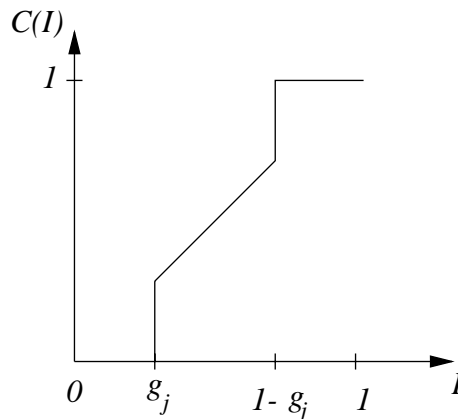


Abbildung 8.3: Quantisierungsoperator im IFT-Algorithmus

Der Parameter g_j wird im Laufe der Optimierung kontinuierlich erhöht.

Die Binarisierung läßt sich insbesondere mit der im folgenden Abschnitt dargestellten Fensterung kombinieren. Damit kann dann das Quantisierungsrauschen, das sich auch hier nicht ganz vermeiden läßt, in den Bereich außerhalb des eigentlichen Bildfensters verschoben werden.

Jennison und Allebach (siehe [jenn89]) verwenden ein ähnliches Verfahren mit einem gegenüber den Arbeiten von Wyrowski leicht abgewandelten Quantisierungs-Operator und nennen den Algorithmus *Projection onto constraint sets (POCS)* (die Bezeichnung “POCS” wird von der Gruppe der Purdue Universität synonym mit unserer Bezeichnung “Iterative Fourier-Transformations-Algorithmen” verwendet). Er ist primär für komplexe Rekonstruktionen (Amplitude und Phase in Bildebene festgelegt) entwickelt und konvergiert für die von uns angestrebte ausschließliche Optimierung der Amplitude nicht immer. Innerhalb des Verfahrens wird die Konvergenz im wesentlichen durch einen freien Parameter geregelt. Die Wahl dieses Parameters muß zeitlich verändert werden und hängt außerdem von der konkreten Anwendung ab.

Eine sehr interessante Variante der Binarisierung nach Wyrowski wurde kürzlich von Zhang et al. (siehe [zhan95]) vorgestellt. In dieser Variante wird die Binarisierung in der Hologrammebene nur für eine bestimmte Anzahl von Hologrammpixeln durchgeführt. Die ent-

Hologrammen) beschrieben.

sprechende Anzahl von Pixel wird dabei im Laufe der Iterationen immer weiter erhöht bis schließlich alle Pixel binarisiert werden.

Welche Pixel binarisiert werden, wird jeweils zufällig gewählt. Das hat den Effekt, daß das Binarisierungsrauschen gleichmäßig über das gesamte Hologramm verteilt wird. Ich habe diesen Algorithmus bisher nicht getestet, die Autoren berichten aber über eine Verringerung des Rekonstruktionsfehlers um ca. 65 Prozent. Allerdings bezieht sich dieser Wert auf simulierte Rekonstruktionen ohne Speckles (und diese machen den Hauptfehler aus), so daß der Gewinn insgesamt in der optischen Rekonstruktion kaum sichtbar wird.

Eine gänzlich andere Methode der Quantisierung innerhalb des IFT-Algorithmus ist die Verwendung der Error-Diffusion (siehe Abschnitt 11.6) innerhalb der Iteration ([weis91]). Außer der Binarisierung soll dieses Verfahren allgemein die Stagnation des Algorithmus vermindern.

8.1.4 Fensterung im IFT-Algorithmus

Um die Qualität der Rekonstruktion von CGHs zu verbessern, müssen Freiheitsgrade genutzt werden. Der Gerchberg-Saxton-Algorithmus nutzt die Objektphase als Freiheitsgrad. Wenn man sich nur für einen bestimmten Bereich des Objekts interessiert (ein sogenanntes *Objektfenster*), erhält man neue Freiheitsgrade, die zur Optimierung genutzt werden können. Die neuen Freiheitsgrade sind die Objektintensitäten außerhalb des Objektfensters.

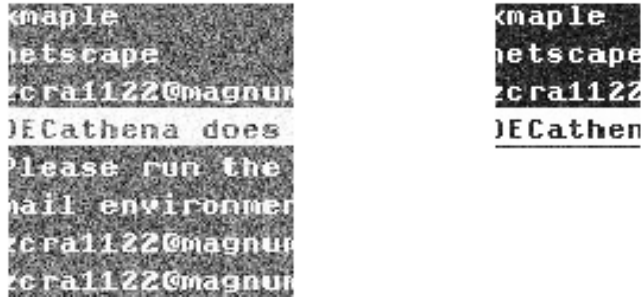
Statt über $N \times N$ Pixel kann man dann nur noch über $N' \times N'$ Pixel mit $N' < N$ in der Rekonstruktion frei verfügen. Dafür wird der Fehler dieser $N' \times N'$ Pixel reduziert sein. Besonders im Zusammenhang mit zeitlich variablen Hologrammedien — also z.B. unseren LCDs — ergeben sich hier interessante Möglichkeiten. Entscheidend für die die Qualitätssteigerung ist das Verhältnis N'/N . Die Position des Fensters im Gesamtobjektfeld ist weitgehend ohne Belang.

Im Sinne einer genauen Vermessung eines Objekts mittels projizierten Mustern können wir das Objekt schrittweise abtasten. Das Objekt wird in Teilbereiche eingeteilt. Für jeden Teilbereich wird ein CGH berechnet. Nacheinander werden alle CGH-Rekonstruktionen auf das Testobjekt projiziert und die entsprechenden Ergebnisbilder aufgenommen. Im Rechner werden die Einzelbilder dann wieder zu einem Gesamtbild zusammengesetzt. Da die Ausgabe der CGHs im Videotakt erfolgen kann, lassen sich so z.B. in 5 Sekunden 100 verschiedene Objektbereiche vermessen. Das SBP des LCDs wird in diesem Fall um den Faktor 100 vergrößert und entsprechend nimmt die Qualität der Vermessung deutlich zu.

Insbesondere für neue Testverfahren von optischen Bauelementen (in erster Linie Asphären) sind mit dieser Technik unter Umständen sehr hohe Genauigkeiten bei größtmöglicher Flexibilität⁸ realisierbar.

Die Modifikation des Gerchberg-Saxton-Algorithmus ist trivial. Beim Ersetzen der Amplitude in der Rekonstruktionsebene durch die gewünschte Objektamplitude erfolgt die Ersetzung nur in dem gewählten $N' \times N'$ Bereich (siehe z.B. [bryn90]). Die Verbesserung der Bildqualität ist — je nach Objekt — teilweise erstaunlich. Abbildung 8.4 zeigt ein Beispiel bei einer Reduzierung der Bereichsgröße um den Faktor 4.

⁸beliebige Elemente lassen sich ohne Umbau des Systems testen, keine Referenzobjekte nötig (CAD-Daten sind ausreichend)



a.) ohne Fensterung

b.) Fensterung (4:1)

Abbildung 8.4: Verbesserung der Bildqualität durch Fensterung

Es sei darauf hingewiesen, daß die bisher besprochenen Freiheitsgrade (Objektphase, Amplituden außerhalb des Fensters) nicht nur zur Optimierung der Rekonstruktionsqualität (quadratischer Fehler), sondern auch zur Erhöhung der Beugungseffizienz (für uns nicht so sehr von Belang) verwendet werden können (siehe [wyo90]).

8.1.5 IFT-Algorithmus mit Bandbreitenkontrolle

Im Abschnitt 7.1 wurde bereits ausführlich auf die Problematik der begrenzten Bandbreite des Hologramms eingegangen. Wenn das aufzuzeichnende Hologramm größer als das Hologrammedium ist, ergeben sich sogenannte Speckles in der Rekonstruktion. Der IFT-Algorithmus läßt sich abwandeln, um die Bandbreite des Hologramms und damit die Speckles zu reduzieren.

Die Modifikation ist einfach, erhöht die Rechenzeit aber um mindestens den Faktor 4⁹. Die Hologrammgröße wird in in jeder Dimension verdoppelt. Wenn das Ausgabemedium also $N \times N$ Pixel aufweist, wird bei der CGH-Berechnung mit $2N \times 2N$ Pixel gerechnet. In der Hologrammebene werden nun einfach die Amplituden in den Randbereichen auf 0 gesetzt. Damit wird das Tiefpaßverhalten des CGHs berücksichtigt.

Variationen dieses Verfahrens sind möglich. Eine ausführlichere Diskussion findet man bei [alle75], [wyo88] und [wyo89b]. Eigene Ergebnisse ergeben eine — im Gegensatz zu den Originalbildern in der Arbeit von Allebach und Liu — deutliche Reduktion der Speckles. Ob dieses Verfahren eingesetzt wird, hängt demnach im wesentlichen von der verfügbaren Rechenkapazität ab.

Selbst wenn das zu rekonstruierende Amplitudenobjekt bandbegrenzt ist, treten Probleme auf. Diese rühren von der Phase des Objekts her. Bereits die vielfach verwendete, zufällige Startphase für die Iterationen ist nicht bandbegrenzt. Anschaulich ist klar, daß die zufälligen Sprünge in der Phase zu hohen Ortsfrequenzen führen, die vom Hologramm nicht gespeichert werden können (begrenzte Apertur des Hologramms) und damit zu Speckles führen. Diese Sprünge können durch den Gerchberg-Saxton-Algorithmus nicht immer wegoptimiert werden. Der Algorithmus stagniert.

Ziel muß es also sein, diese Sprünge zu vermeiden. In einem ersten Schritt kann dazu die

⁹Wegen des erhöhten Speicherbedarfs und der sich daraus ergebenden Verlangsamung ist eher mit noch deutlicheren Geschwindigkeitseinbußen zu rechnen.

Startphase entsprechend ohne Sprünge gewählt werden. Wyrowski (siehe [wyro88]) hat eine Abwandlung des Gerchberg-Saxton-Algorithmus vorgeschlagen, in der solche Sprünge durch eine geeignete Wahl der Startphase vermindert bzw. vermieden werden. Insbesondere bei komplexen Hologrammen (im Gegensatz zu Phasenhologrammen¹⁰) scheint dieses Verfahren von Bedeutung zu sein.

Dazu wird zunächst für jede Bildzeile ein eindimensionaler IFT-Algorithmus angewendet. Die resultierenden Phasen der unterschiedlichen Zeilen werden zu einer Phase für das gesamte (zweidimensionale) Objekt zusammengesetzt und als Startphase für den eigentlichen (zweidimensionalen) IFT-Algorithmus verwendet. Da wir momentan nicht planen, komplexe Hologramme zu benutzen, gehe ich hier nicht weiter auf dieses Verfahren ein.

Eine andere Startphase, die hauptsächlich die Stagnation bei gleichzeitiger Verminderung der Speckles bewirken soll, wurde von Bräuer et al. ([bräu91]) vorgeschlagen.

Unabhängig von der Startphase können sich aber während der Iterationen wieder neue, ungünstige Phasenverteilungen ergeben. Aagedahl (siehe [aage96]) hat jüngst ein Verfahren vorgeschlagen, bei dem in jedem Iterationszyklus das Vorhandensein ungünstiger Phasensprünge geprüft wird. Um solche Sprünge zu eliminieren, wird ein (möglichst kleiner) Bereich in der Umgebung des Sprungs gewählt und mit einer Phasenverteilung ohne Sprünge belegt. Danach wird weiter iteriert (normaler IFT-Algorithmus). Dabei ist allerdings zu beachten, daß bei den folgenden Iterationen der entsprechende Bereich immer wieder mit der besonderen Phasenverteilung belegt wird. Die Phase in diesem Bereich wird also von allen folgenden Iterationen ausgenommen.

Die Wahl der Bereichsgröße um den Phasensprung ergibt sich aus den Arbeiten von Aagedahl ("Speckle-Ordnung"). Die Arbeiten sind noch nicht veröffentlicht und wurden daher von mir bisher nicht verwendet. Sie könnten aber zu einer weiteren Verbesserung der erzielten Rekonstruktionsqualität führen und sollten zusammen mit den Arbeiten von Bräuer et al. nach Möglichkeit auf einen Einsatz geprüft werden.

8.1.6 IFT-Algorithmus mit Amplituden-Phasen-Koppelung

Für unsere Anwendungen bei der Hologrammrekonstruktion werden vor und hinter dem LCD in der Regel Polarisatoren benutzt (siehe ausführliche Diskussion in Kapitel 13).

Damit folgt aber zwangsläufig, daß durch die Anordnung nicht nur die Phase variiert wird, sondern auch die Amplitude eine Variation erfährt. Wie stark die Amplitudenvariation ist, hängt in komplizierter Weise (siehe Kapitel 13) vom LCD selbst und von der Stellung der Eingangs- und Ausgangspolarisatoren ab. In der Praxis findet man zwar Einstellungen, die nur eine sehr geringe Amplitudenmodulation (z.B. $< 10\%$) zur Folge haben, bei diesen Einstellungen ergibt sich oft aber zwangsläufig ein stark eingeschränkter Phasenschieberegion oder eine mangelhafte Gesamtintensität.

Für dieses Problem wird eine einfache Abwandlung des Gerchberg-Saxton-Algorithmus vorgeschlagen: Die Amplitudenvariation in Abhängigkeit von der Phasenschiebung kann in erster Näherung nach Messungen von J. Schweikhardt (siehe [schw96]) als linear angenom-

¹⁰Zumindest konnten die Probleme in der angegebenen Stärke bei unseren Phasenhologrammen nicht nachvollzogen werden.

men werden. Den entsprechenden Zusammenhang schreiben wir als

$$H = a_0 \left[1 - \alpha \frac{\phi}{2\pi} \right] e^{i\phi} \quad (8.7)$$

Die einfache Abwandlung des Gerchberg–Saxton–Algorithmus besteht nun darin, in der Hologrammebene die Amplitude nicht konstant, sondern auf den nach Gleichung (8.7) gegebenen Wert zu setzen. Der zusätzliche Zeitbedarf ist vernachlässigbar.

Da nur die Amplituden verändert werden, wird die Konvergenz des Algorithmus nicht beeinflusst. Eine zusätzliche Beeinflussung der Phasen innerhalb des Algorithmus (um beispielsweise einem begrenzten Phasenschiebebereich Rechnung zu tragen) ist dagegen nicht möglich.

In den Abbildungen 8.6 und 8.8 wird beispielhaft für einen Koppelungsfaktor $\alpha = 1$ die Verbesserung der Rekonstruktionsqualität durch die Modifikation des Algorithmus gezeigt. Zu beachten ist weiterhin, daß die nullte Beugungsordnung stark vermindert ist; ein für die praktische Anwendung unter Umständen gewichtiger Vorteil, der in der Simulation hier nicht sichtbar wird (nullte Beugungsordnung wurde unterdrückt).

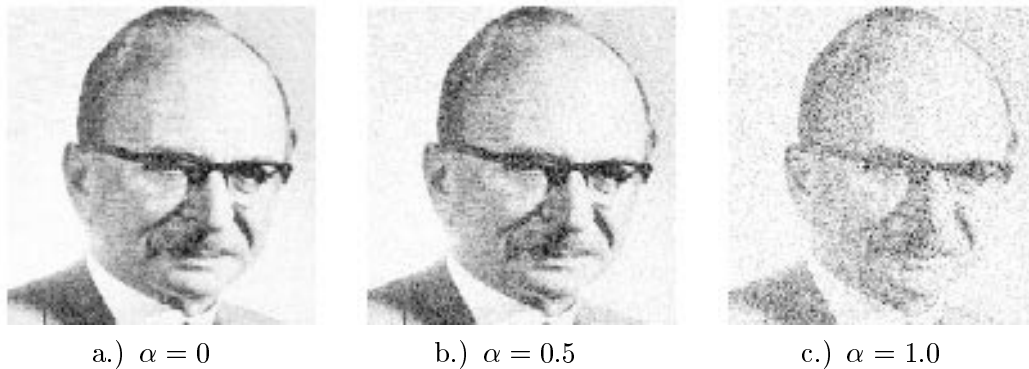


Abbildung 8.5: Gerchberg–Saxton–Algorithmus (GS)

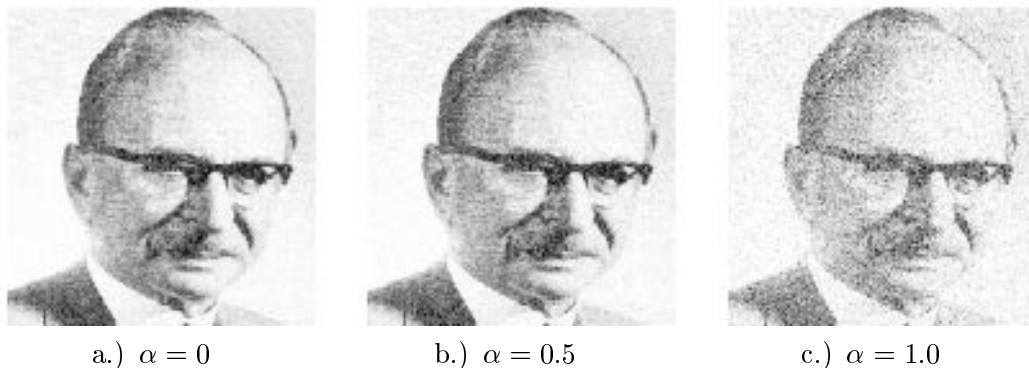


Abbildung 8.6: modifizierter Gerchberg–Saxton–Algorithmus (GS-APC)

8.1.7 Dreidimensionaler IFT–Algorithmus

Die bisher beschriebenen IFT–Algorithmen liefern eine beliebige Lichtverteilung in einer Ausgangsebene. Die Lichtintensität ist nicht in verschiedenen Bildebenen kontrollierbar.

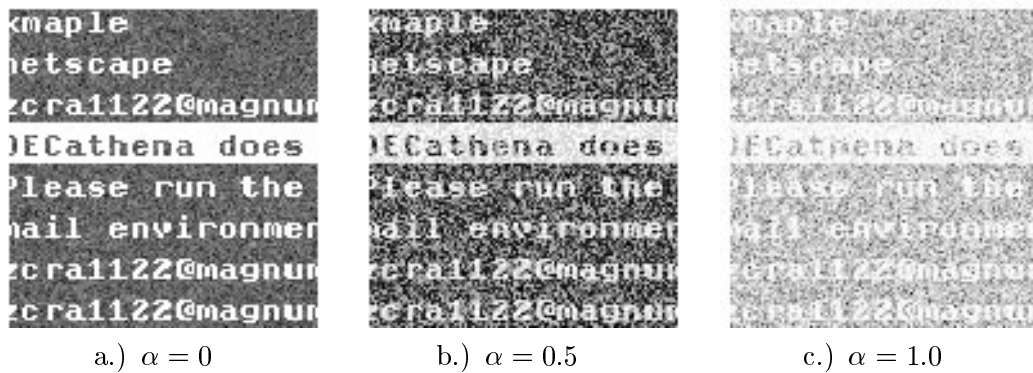


Abbildung 8.7: Gerchberg-Saxton-Algorithmus (GS)

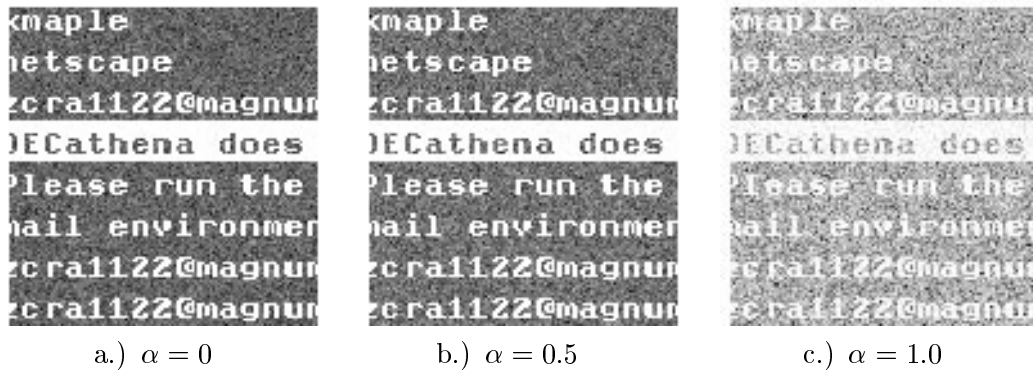


Abbildung 8.8: modifizierter Gerchberg-Saxton-Algorithmus (GS-APC)

Hier wird nun erstmals eine Erweiterung des GS-Algorithmus auf mehrere Ebenen in beliebigen Abständen vorgestellt. Grundidee der Erweiterung ist der im folgenden beschriebene simulierte Einbau von zwei Linsen in die Iterationsschleife.

Man kann das rekonstruierte Bild eines CGHs in die Tiefe verschieben, indem man eine zusätzliche Linse in den Rekonstruktionsaufbau einfügt (siehe Abbildung 8.9).

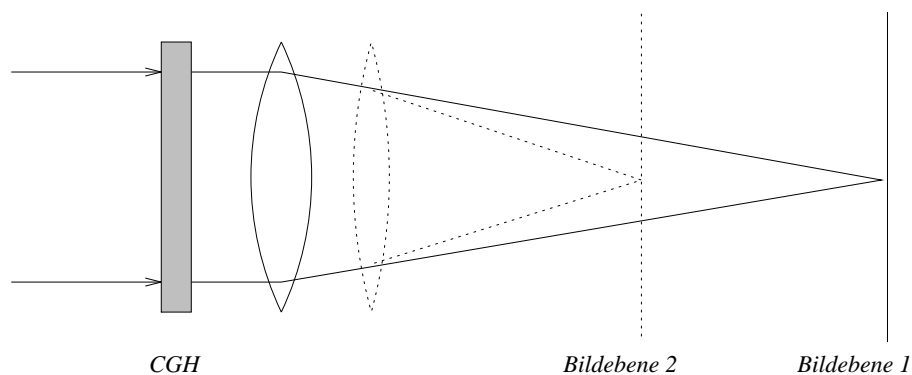


Abbildung 8.9: Prinzip der Rekonstruktion in verschiedenen Ebenen

Entsprechend kann man diese Linse aber mittels (2.17) auch direkt in das Hologramm integrieren. Dazu muß lediglich die Wenn wir gleichzeitig Rekonstruktionen in verschiedenen Bildebenen wollen, dann kann man die entsprechenden Teilhologramme (jedes bestehend aus einem Grundhologramm und einer simulierten Linse) überlagern.

Dieser einfache Ansatz ist aber nicht mit dem Gerchberg-Saxton-Algorithmus vereinbar. Wenn wir mehrere Hologramme getrennt optimieren und dann nach der einfachen, eben geschilderten Vorschrift überlagern, macht das wenig Sinn. Die einzelnen Teilhologramme können nicht optimal zusammenspielen.

Stattdessen wird eine globale Optimierung aller Ebenen gleichzeitig vorgenommen. Zu diesem Zweck wird der einfache Gerchberg-Saxton-Algorithmus entsprechend Abbildung 8.11 abgewandelt. Das DIN-Flußdiagramm ist recht unübersichtlich, weswegen in Abbildung 8.10 eine anschaulichere Darstellung des Algorithmus dargestellt ist. Man sieht deutlich die Grundidee des Einbaus von jeweils zwei Linsen pro Tiefenebene in die Iterationsschleife.

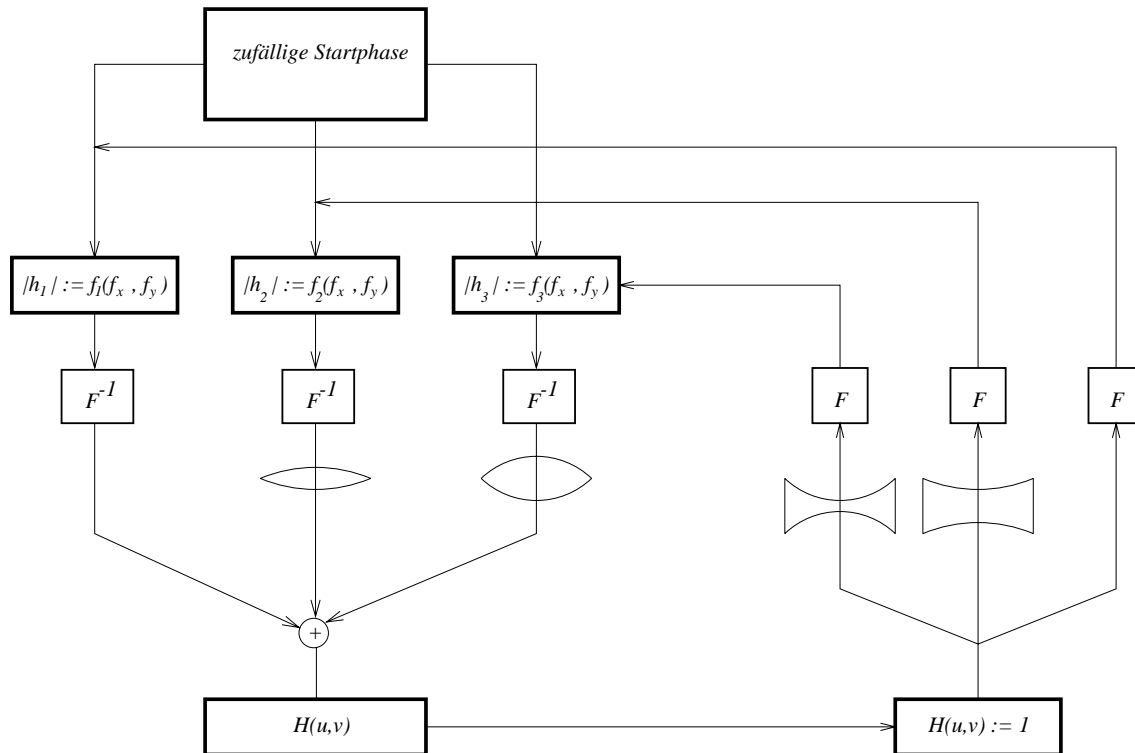


Abbildung 8.10: 3D-Gerchberg-Saxton für (beispielhaft für 3 verschiedene Tiefenebenen)

In den Abbildungen 8.12 und 8.13 sind die optischen Rekonstruktionen bei Verwendung des VGA-LCDs dargestellt. Die Hologramme haben 256×256 Pixel und sind auf 2 Bit quantisiert. Auf die Darstellung der theoretischen Simulation (mit entsprechend besserer Qualität) wird hier verzichtet, da die optischen Rekonstruktionen bereits eine beeindruckend gute Qualität aufweisen.

Die Qualität kann vermutlich weiter verbessert werden, wenn ein gutes und voll vermessenes LCD zur Verfügung steht. Die dargestellten Rekonstruktionen wurden mit dem im Kapitel 14.4.1 beschriebenen LCD durchgeführt, daß uns leider nur sehr kurze Zeit als Leihgabe zur Verfügung stand und deswegen nicht optimal ausgenutzt werden konnte.

In Abbildung 8.12 sieht man die optische Rekonstruktion eines CGHs für 2 Tiefenebenen im 0.1-fachen Abstand der Brennweite. In beiden Ebenen erhält man jeweils ein scharfes Bild. Wie stark die Tiefendifferenz der beiden Bilder ist, erkennt man an der starken Veränderung der nullten Beugungsordnung in b.) gegenüber a.).

Der in Abbildung 8.12 an sich sehr große Tiefenbereich der Abbildung kann noch weiter

vergrößert werden. Beispielhaft ist in Abbildung 8.13 ein Abstand der beiden Ebenen von $0.3 f$ gewählt. Ein Abstand also, der weit über der für sinnvolle Anwendungen nötigen Tiefenschärfe liegt. Kreise und Rechteck sind klar erkennbar. Die nullte Ordnung ist extrem verzerrt. Das CGH ergibt zusätzliche Rekonstruktionen in zwei weiteren Ebenen (insgesamt also 4 Ebenen), die hier nicht dargestellt sind.

In Abbildung 8.13 b.) sieht man innerhalb des Quadrats einen leichten Fehler in Form einer schräg verlaufenden, dunkleren Geraden. Diese Gerade kommt von einer anderen Bildebene, die sich mit der Ebene des Quadrats überlagert. Das muß so sein, denn der Algorithmus berücksichtigt keine Hidden-Line-Effekte. Für die von uns angestrebten Anwendungen der Projektion auf beliebige Oberflächen kann ein entsprechendes Problem nicht auftreten! Zu jedem Pixel in der zweidimensionalen Grund-Bildebene (senkrecht zur optischen Achse) gibt es genau *einen* passenden Tiefenwert.

Die 3D-CGHs wurden in einer leicht abgewandelten Rekonstruktionsanordnung (zusätzliche Objektive und Verlängerung des optischen Weges über Spiegel) direkt ins Auge projiziert. Man “sieht” damit unmittelbar den dreidimensionalen Charakter. Die Größen der Dreiecke aus Abbildung 8.12 sind unterschiedlich und bei leichter Bewegung des Kopfes wird die Parallaxe sehr schön deutlich. Für echte Display-Anwendungen ist die Qualität der CGHs (bedingt durch das Medium LCD) aber meiner Ansicht nach noch zu schlecht.

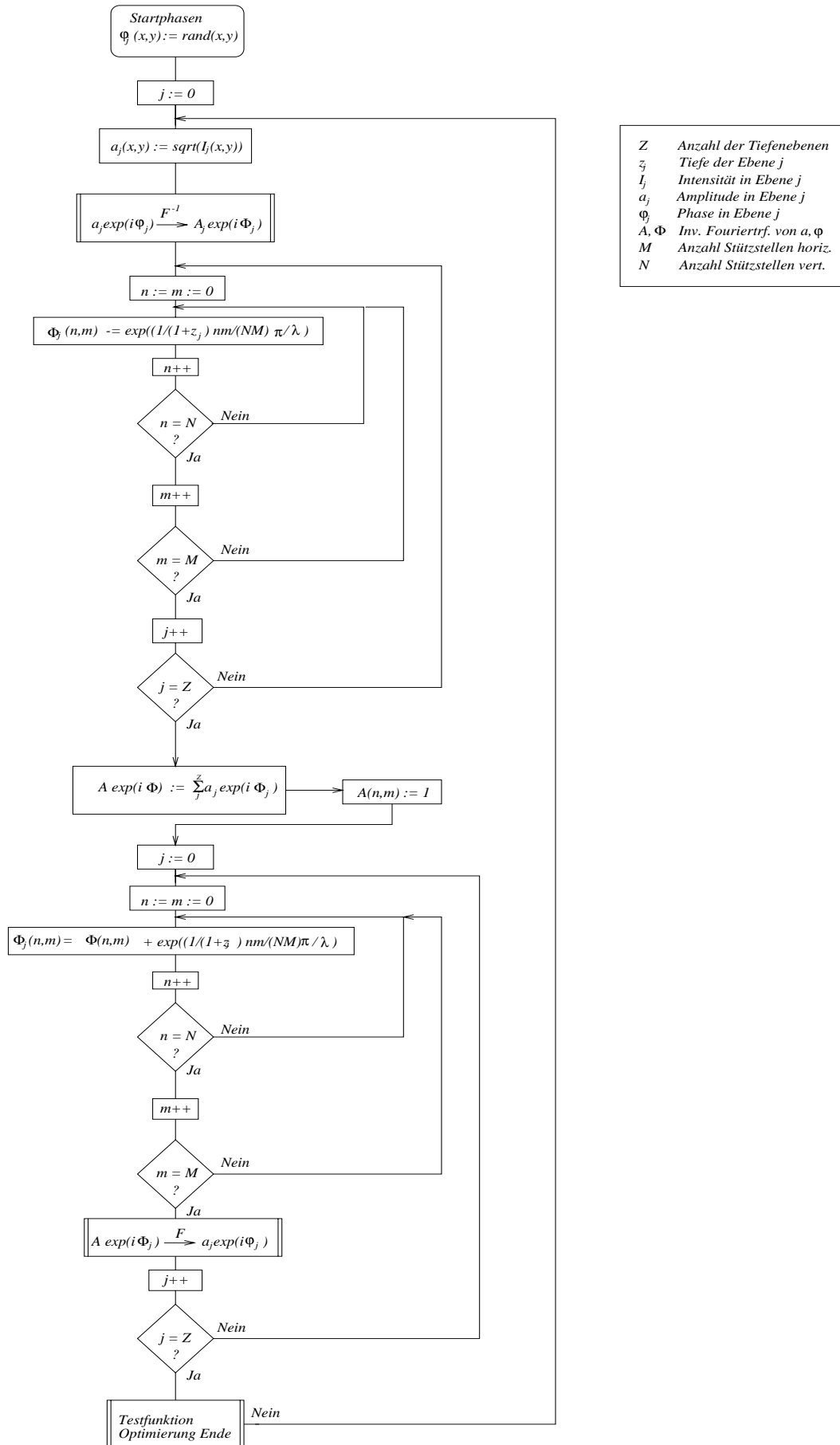
Die Qualität der Rekonstruktion nimmt — wie man es anschaulich erwartet — mit der Anzahl der Tiefenebenen und der Tiefe selbst ab. Eine theoretische Beschreibung der Rekonstruktionsqualität ist schwierig und wurde bisher nicht erbracht.

Der Algorithmus kann in Zukunft weiter verbessert werden, indem die bisher besprochenen Abwandlungen des Gerchberg-Saxton-Algorithmus (insbesondere Fensterung und Amplituden-Phasen-Koppelung beim LCD) zusätzlich berücksichtigt werden. Größere Probleme dürften dabei aus meiner Sicht nicht auftreten. Auch die für die meisten Anwendungen unerhebliche Einschränkung, daß die Objektebenen parallel zur Hologrammebene angenommen werden, kann umgangen werden, wenn man das Verfahren mit Arbeiten von Leseberg und Frère ([lese88]) und Tommasi und Bianco ([tomm91]) kombiniert, die den Effekt der Rotation von Objektebenen bei Verwendung der FFT zur Hologrammberechnung untersucht haben¹¹.

Weiterhin ist zu beachten, daß die Qualität noch deutlich, durch Parkettierung des Hologramms, verbessert werden kann. Eine einfache Parkettierung, wie im Abschnitt 7.1.2 beschrieben, ist dabei allerdings nicht möglich. Es ist anschaulich klar, daß die im CGH verwirklichten Linsenterme nur dann korrekt sind, wenn die gesamte CGH-Fläche mit der Linse ausgefüllt ist. Wenn nachträglich ein 3D-CGH parkettiert wird, dann sind entsprechend mehrere Linsen über das Hologrammedium verteilt, was offensichtlich dem Grundgedanken zuwider läuft. Im Experiment ergibt sich dementsprechend eine deutliche Verschlechterung der Rekonstruktionsqualität.

Eine Umsetzung des Algorithmus hin zu parkettierten 3D-Hologrammen ist möglich, erhöht aber deutlich die Berechnungsdauer.

¹¹Die ursprüngliche Idee für diese Berechnung ist bereits bei Waters ([wate68]) zu finden.



- Z Anzahl der Tiefenebenen
- z_j Tiefe der Ebene j
- I_j Intensität in Ebene j
- a_j Amplitude in Ebene j
- Φ_j Phase in Ebene j
- A, Φ Inv. Fouriertrf. von a, Φ
- M Anzahl Stützstellen horiz.
- N Anzahl Stützstellen vert.

Abbildung 8.11: 3D-Gerchberg-Saxton DIN Flußdiagramm

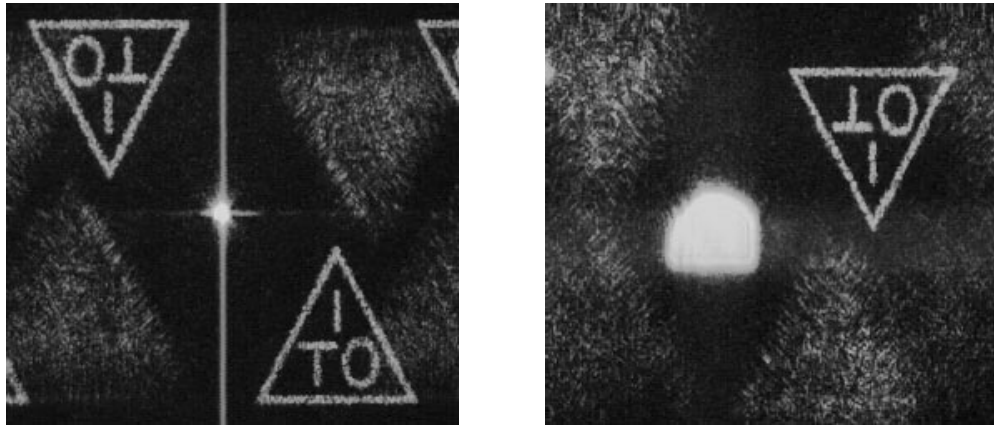


Abbildung 8.12: optische Rekonstruktion eines 3D-CGHs für 2 Tiefenebenen, $0.1 f$

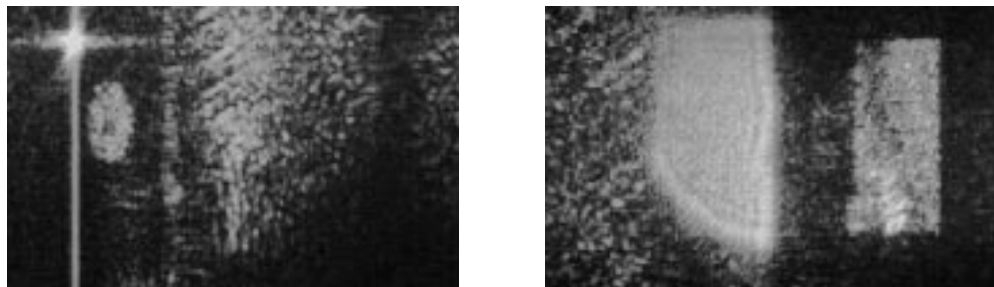


Abbildung 8.13: optische Rekonstruktion eines 3D-CGHs: 2 von 4 Tiefenebenen, $0.3 f$

Kapitel 9

Input–Output–Algorithmus und Global–Iterative–Coding

Die in diesem Kapitel vorgestellten Algorithmen sind eng verwandt mit den im letzten Kapitel besprochenen IFT–Algorithmen. Sie liefern Rekonstruktionen, die hinsichtlich Beugungseffizienz und Rekonstruktionsfehler ungefähr dieselbe Qualität wie IFT–Algorithmen liefern. Auch die typischen Berechnungszeiten unterscheiden sich nicht wesentlich. In der Praxis konnten sich diese Algorithmen (bisher) kaum durchsetzen und auch wir haben lediglich den Einsatz des Global–Iterative–Coding Algorithmus ins Auge gefaßt.

9.1 Input–Output–Algorithmus

Der Input–Output–Algorithmus geht auf Fienup zurück ([fien80]). Während beim Gerchberg–Saxton–Algorithmus direkt die Rekonstruktion modifiziert und dann zu weiteren Iterationen verwendet wird (siehe Abbildung 8.1), wird beim Input–Output–Algorithmus die Rekonstruktion zur Modifikation des ursprünglichen Eingabevektors verwendet. Das Prinzip ist in Abbildung 9.1 veranschaulicht.

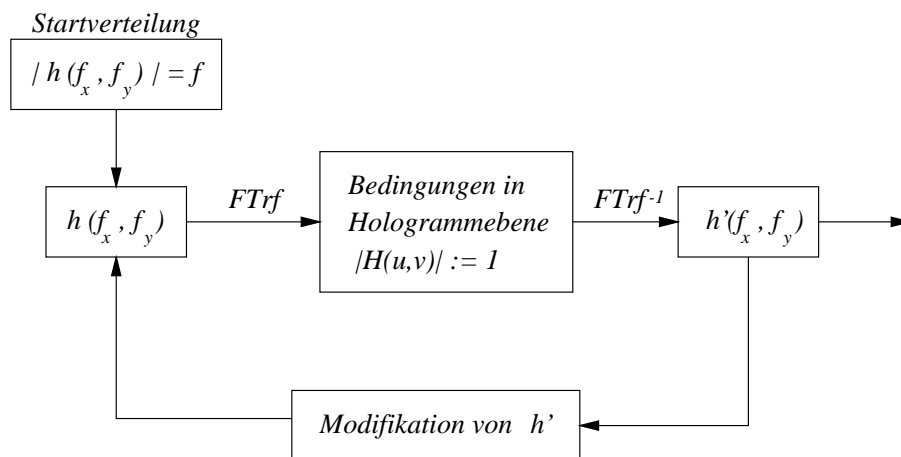


Abbildung 9.1: Input–Output–Algorithmus

Ausgehend von einem Startobjekt mit zufälliger Phase wird eine diskrete Fourier–

Transformation durchgeführt, um das der Objektverteilung entsprechende Hologramm zu erhalten. Dieses Hologramm wird — wie beim GS-Algorithmus — dahingehend modifiziert, daß es dem realen Hologrammedium entspricht. Im Fall eines Phasenhologramms wird also die Amplitude explizit auf einen konstanten Wert gesetzt. Danach wird die inverse, diskrete Fourier-Transformation gebildet. Die erhaltene komplexe Amplitude wird zur Modifikation des Objekts verwendet.

Die entscheidende Frage ist nun, wie die Modifikation des Objekts zu erfolgen hat, um den Fehler der Rekonstruktion zu verringern. Fienup wählt hier folgendes Vorgehen: Die neue Objektamplitude $f'(f_x, f_y)$ ergibt sich aus der bisherigen Objektamplitude $f(f_x, f_y)$ und der Rekonstruktion $h(f_x, f_y)$ zu

$$f'(f_x, f_y) = f(f_x, f_y) + \alpha \Delta f \quad (9.1)$$

Dabei ist Δf der Korrekturterm, der sich aus f , $|f|$ und h mittels

$$\Delta f = \left[|f| \frac{h}{|h|} - g \right] + \left[|f| \frac{g}{|g|} - |f| \frac{f}{|f|} \right] \quad (9.2)$$

ergibt. Der erste Term ist dabei für die Differenz der Intensitäten zuständig während der zweite Term Phasenfehler berücksichtigt und dementsprechend für Anwendungen, bei denen die rekonstruierte Phase ohne Belang ist, weggelassen werden muß.

Eine eher heuristische Begründung für diese Wahl von Δf findet man bei Fienup ([fien80]).

Der Algorithmus kann — genau wie der Gerchberg-Saxton-Algorithmus — nicht nur zur Berechnung von CGHs verwendet werden, sondern ist auch für Phase-Retrieval Probleme geeignet.

Eine Variation dieses Algorithmus, die um ca. 50 Prozent bessere Rekonstruktionsergebnisse für quantisierte Hologramme liefern soll, wurde von Chang und Ersoy ([chan94]) vorgeschlagen.

Dabei wird der Faktor α in (9.1) adaptiv über die Gleichung

$$\alpha = \frac{\alpha_0}{\beta_i} \quad (9.3)$$

gewählt. α_0 ist konstant und β_i ändert sich im Lauf der Iterationen. Bei günstiger Wahl dieser Parameter soll ein besseres Verhalten des Algorithmus erreicht werden. Diese Verbesserung erreichen die Autoren allerdings nur für intensitätsbasierte Rekonstruktionen. Für die verwendeten Hologramme erwies es sich dabei als vorteilhaft, wenn β_i im Laufe der Iterationen von 1.0 bis 1.5 variiert.

Gerade hier liegt aber wieder das Problem dieses Algorithmus: Für viele Anwendungen ist es untragbar, daß erst in längeren Versuchsreihen eine passende Wahl von Parametern gefunden wird, damit der Algorithmus gute Ergebnisse liefert. Der Input-Output-Algorithmus ist in diesem Sinne zu allgemein. Demgegenüber liefert der Gerchberg-Saxton-Algorithmus immer sinnvolle Ergebnisse.

9.2 Global-Iterative-Coding

Zunächst soll an eine einfache Hologramm-Kodierung von Lee erinnert werden (siehe [lee79]). Die Interferenzstreifen eines Hologramms sind gegeben durch die Gleichung

$$2\pi v/T + \varphi(u, v) = 2\pi n \quad (9.4)$$

Dabei wird von einer Objektwelle mit Phase φ in der Hologrammebene und einer Referenzwelle $\exp(-2\pi iv/T)$ ausgegangen (für den von uns normalerweise verwendeten on-axis Fall reduziert sich diese Gleichung auf $\varphi = 2\pi n$). n ist eine ganze Zahl.

Mit einer entsprechenden Kodierung kann die Phase Φ gespeichert werden. Will man zusätzlich die Amplitude A kodieren, dann muß eine Variation von Gleichung (9.4) vorgenommen werden.

Lee bespricht drei verschiedene Methoden, wie dies durchgeführt werden kann. Ich will hier nicht näher auf die Ableitung dieser Methoden eingehen, sondern verweide mit Bergeron et al. lediglich eine dieser Methoden:

Kodiert werden soll $H(u, v) := A(u, v)e^{i\varphi(u, v)}$

$$\cos(2\pi u/T + \varphi(u, v)) = \cos \pi q(u, v) \quad (9.5)$$

mit

$$\pi q(u, v) = \arcsin \frac{A(u, v)}{A_{max}} \quad (9.6)$$

Für binäre Hologramme wird folgendes Clipping gewählt:

$$H(u, v) = \begin{cases} 1 & \text{für } \cos(2\pi n/T + \varphi) - \alpha \cos(\pi q) \geq 0 \\ 0 & \text{für } \cos(2\pi n/T + \varphi) - \alpha \cos(\pi q) < 0 \end{cases} \quad (9.7)$$

Der Parameter α ist im Gegensatz zur Originalarbeit von Lee neu und kontrolliert die Konvergenz des Algorithmus. Bergeron et al. benutzen eine *erweiterte Fensterfunktion*. Korrekturen finden im eigentlichen Rekonstruktionsgebiet und — in abgeschwächter Form — auch in der unmittelbaren Umgebung des Rekonstruktionsgebiets statt.

Die eigentliche Iteration wird mittels

$$H'(u, v) = H(u, v) + H_\epsilon(u, v) \quad (9.8)$$

durchgeführt. Dabei ist H_ϵ das kodierte Hologramm, das der Differenz zwischen der gewünschten und der erhaltenen Rekonstruktion entspricht. Also für komplexe Rekonstruktionen z.B.

$$H_\epsilon = C[f(f_x, f_y) - \lambda h(f_x, f_y)] \quad (9.9)$$

wobei mit C die Kodierung nach Gleichung (9.7) bezeichnet wird. Hier ist anzumerken, daß durchaus andere Kodierungen denkbar sind. Insbesondere für nicht-binäre Hologramme-

dien kommen einfachere Kodierungen in Frage. Bei Phasenhologrammen mit vielen Phasenstufen kann direkt auf eine punktorientierte Kodierung der Phase übergegangen werden ([berg95]).

Der entsprechende Algorithmus ist in Abbildung 9.2 veranschaulicht.

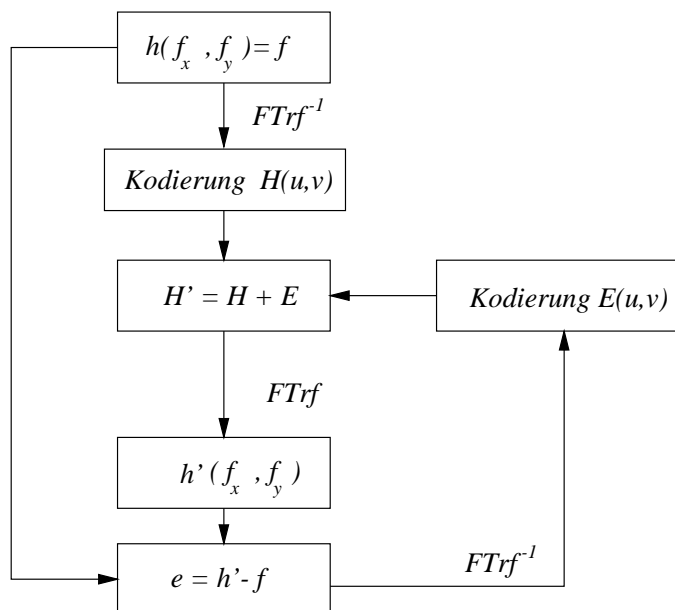


Abbildung 9.2: Global-Iterative-Coding Algorithmus

In der Regel werden für Hologramme mit 128×128 Bildpunkten zwischen 7 und 30 Iterationen benötigt. Gegenüber der einfachen Kodierung nach Lee ergeben sich deutliche Verbesserungen.

Die Konvergenz des Verfahrens konnte von den Autoren nicht bewiesen (und — zumindest für mich — auch nicht plausibel gemacht) werden.

Während mit dem Verfahren für Amplitudenhologramme und binäre Hologramme keine Probleme auftraten, war es unmöglich, die in der neueren Arbeit von Bergeron et al. ([berg95]) besprochene Übertragung des Verfahrens auf Phasenhologramme nachzuvollziehen. Das Verfahren konvergierte hier nur für sehr wenige Fälle. Eine längere Kommunikation mit Herrn Bergeron via EMail ergab schließlich, daß in diesem Fall tatsächlich Probleme bei der Konvergenz auftreten. Inzwischen gäbe es aber Möglichkeiten, die Konvergenz mit einer verbesserten Variante des Algorithmus zu kontrollieren. Leider soll dieser Algorithmus geheimgehalten und für kommerzielle Zwecke eingesetzt werden (bei einer Spin-off Firma der Arbeitsgruppe kann man sich Hologramme berechnen und produzieren lassen).

Abgesehen davon, daß eine solche Berechnung bei einer Firma für unsere Zwecke natürlich völlig uninteressant ist, muß man befürchten, daß die Konvergenz nur bei günstiger Wahl von zusätzlichen Parametern erreicht wird und keine vollautomatische Berechnung, wie sie für unsere Anwendungen benötigt wird, vorgenommen werden kann.

Ein freundlicherweise von Herrn Bergeron für uns mit seinem neuen Algorithmus berechnetes Hologramm zeigte in der Simulation eine leicht verbesserte Qualität gegenüber den IFTA-Hologrammen. Allerdings sind genaue Vergleiche nicht möglich, da wichtige Angaben wie die Anzahl der Iterationen und die optimierte Bereichsgröße fehlten und uns nur die Rekonstruktionen zugesandt wurden.

Kapitel 10

Direkte Suchalgorithmen

Die bisher besprochenen Algorithmen zur CGH Berechnung orientierten sich direkt an dem physikalischen Entstehungsprozeß eines Hologramms. Wir können das Problem der Hologrammberechnung aber auch auf eine Weise betrachten, die im starkem Gegensatz zum bisherigen Vorgehen steht. Wir fassen die Berechnung als ein ganz gewöhnliches Optimierungsproblem auf und wenden verschiedene, in der Informatik bekannte, Optimierungsstrategien an.

Dabei gehen wir direkt vom Hologramm selbst aus. Das Hologramm besteht aus $N \times M$ Pixeln. Jedes Pixel kann Q Werte annehmen. Offensichtlich gibt es genau Q^{MN} mögliche Hologramme. Von diesen Hologrammen suchen wir das, welches die — nach unserer Definition — beste Hologrammrekonstruktion ergibt. Wir haben ein MN -dimensionales Optimierungsproblem.

Methoden, die von diesem Ansatz ausgehen und mit relativ einfachen Verfahren, dieses “beste” Hologramm suchen, werden oft als *direkte Suchalgorithmen* bezeichnet.

Der einfachste Algorithmus wäre sicher, einfach alle möglichen Hologramme zu untersuchen. Der Algorithmus ist extrem einfach und liefert garantiert das günstigste Hologramm. Die Optimierung wird nicht in *lokalen Minima* gefangen. Leider ist dieser grundlegende Algorithmus nicht praktikabel. Q^{MN} ist für sinnvolle CGHs eine astronomisch große Zahl. Eine entsprechende Anzahl von Hologrammen wird sich — auch bei einer Steigerung momentaner Rechenleistungen um viele Größenordnungen — unter keinen Umständen in annähernd sinnvoller Zeit untersuchen lassen.

Die einfache Untersuchung aller möglichen Hologramme muß also dahingehend variiert werden, daß nur eine bestimmte Anzahl von Hologrammen untersucht wird. Entsprechende Probleme treten natürlich nicht nur bei der Berechnung von Hologrammen auf. Hochdimensionale, komplizierte Optimierungsprobleme, und darum handelt es sich hier, sind seit langem bekannt¹ (siehe z.B. [ducc93]). Die entsprechenden Lösungsansätze solcher Probleme basieren auf einer zielgerichteten Wanderung im Lösungsraum und können direkt für die Berechnung von CGHs verwendet werden.

Viele Möglichkeiten zur effektiven Wanderung in diesem Lösungsraum sind bekannt. Besonders beliebt sind Abwandlungen von sogenannten *Gradientenverfahren*, bei denen eine Wanderung jeweils lokal in die Richtung der Verbesserung stattfindet. Die Schrittweiten

¹Ein klassisches Beispiel ist das Problem des Handlungsreisenden.

der Wanderung sind oft variabel und gewisse Zufallselemente werden in die Algorithmen eingebaut, um eine Stagnation in lokalen Minima zu vermeiden.

Lokale Minima sind überhaupt ein Problem aller Optimierungsalgorithmen. Die Wanderung im Lösungsraum führt letztlich in ein Tal im Lösungsraum. Leider ist im Rahmen des Algorithmus jeweils nur eine lokale Kenntnis des Lösungsraums vorhanden. Es kann nicht vorhergesagt werden, ob an anderen Stellen des Lösungsraums eventuell noch viel tiefere Täler, entsprechend besseren Lösungen, existieren. Typische Verfahren findet man in [pres94] (Kapitel 15) und [ducc93].

Für die Optimierung computergenerierter Hologramme werden momentan die folgenden Algorithmen, die im weiteren näher besprochen werden, eingesetzt:

- Direct Binary Search Algorithmus
- Simulated Annealing
- Neuronale Netze
- genetische Algorithmen
- einfache Gradientenverfahren

Viele neuere Optimierungsstrategien (z.B. Sintflut-Algorithmus) wurden bisher nicht hinsichtlich eines Einsatzes zur Berechnung von CGHs untersucht. Zukünftige Verbesserungen sind hier denkbar.

Der Begriff *direkte Suchverfahren* wird nicht einheitlich verwendet. Einige Autoren verstehen unter der Klasse der direkten Suchalgorithmen nur den *Direct-Binary-Search Algorithmus* und den *Simulated Annealing Algorithmus*.

Ein nicht zu unterschätzender Vorteil der direkten Suchmethoden ist es, daß sie ohne Änderung des Algorithmus für beliebige Hologrammgeometrien (z.B. Fresnel-Hologramme) und Hologrammedien (z.B. nichtlineares Verhalten, Pixelfehler usw.) gültig bleiben. Lediglich die Routine zur Berechnung der Rekonstruktion muß entsprechend (z.B. Fresnel-Trf. statt Fourier-Trf.) abgewandelt werden. Der Grundalgorithmus selbst bleibt unverändert. Es lassen sich so auch einfach neue Optimierungsziele setzen (z.B. gleichzeitiges Optimieren von Beugungseffizienz und Rekonstruktionsfehler).

Die direkten Suchmethoden sind in der Regel sehr einfach zu parallelisieren. Daher könnte der enorm große Rechenaufwand, der bisher noch eine Verwendung für große Hologramme ausschließt, sich in der Zukunft als lösbares Problem erweisen.

10.1 Direct-Binary-Search (DBS)

10.1.1 Grundidee

Die Grundidee des DBS-Algorithmus ist erstaunlich einfach und wurde 1987 von Seldowitz, Allebach und Sweeney ([seld87]) für binäre Hologramme vorgeschlagen. Ausgegangen wird von einem zufälligen Hologramm. Für dieses Hologramm werden die Rekonstruktion und in einem zweiten Schritt der entsprechenden Rekonstruktionsfehler berechnet. Jetzt wird ein

zufälliges Pixel des Hogramms umgeklappt, d.h. das Pixel wird auf 1 gesetzt falls es vorher 0 war und umgekehrt. Für dieses neue Hogramm wird wieder der Rekonstruktionsfehler berechnet. Hat sich die Rekonstruktion verschlechtert, dann wird der umgeklappte Pixel wieder auf seinen ursprünglichen Wert zurückgeklappt. Bei einer Verbesserung wird der neue Pixelwert beibehalten.

Das Verfahren ist nochmals in Abbildung 10.1 veranschaulicht.

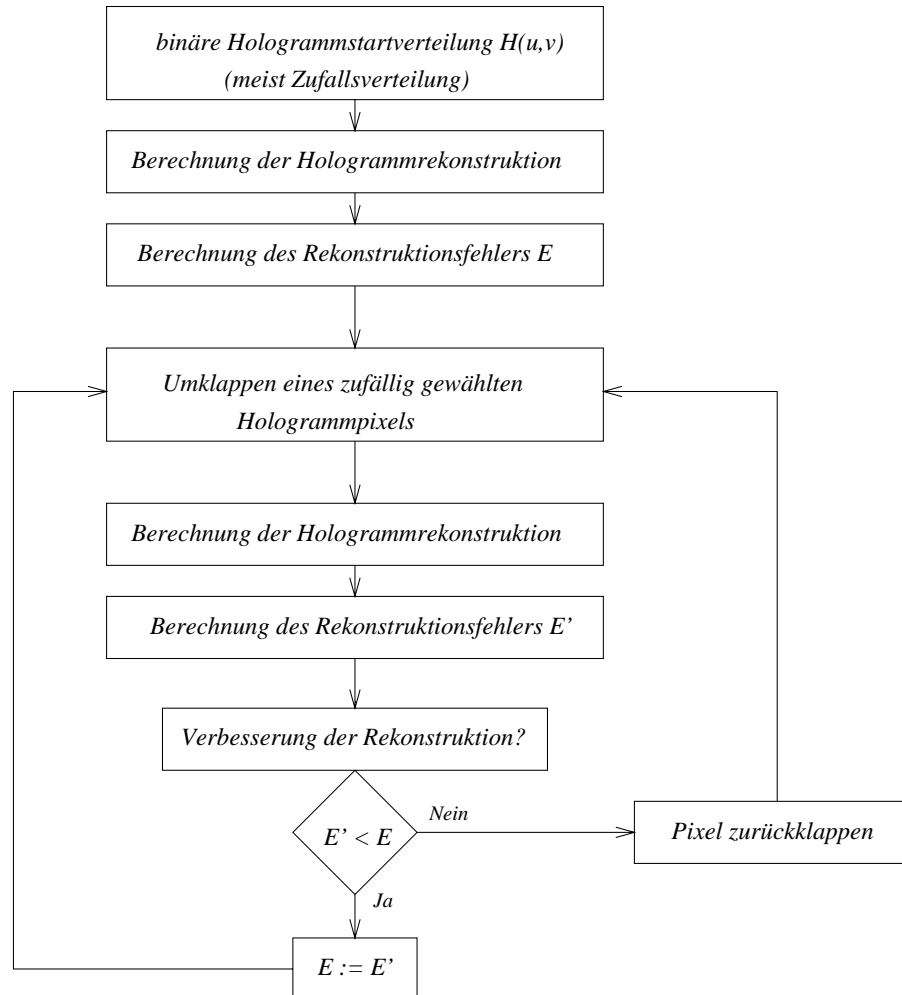


Abbildung 10.1: DBS-Algorithmus

Es ist bemerkenswert, daß dieses einfache Verfahren sehr gute Hogramme liefert. Dies deutet darauf hin, daß der Lösungsraum bei üblichen Hogrammberechnungen sehr viele lokale Minima enthält, die alle annähernd dieselbe Höhe haben².

Der Algorithmus wird nach einer bestimmten Anzahl von Iterationen oder bei Erreichen einer festgelegten Rekonstruktionsqualität abgebrochen. Es ist anschaulich klar, daß bei einer so einfachen Iteration (Änderung eines einzelnen Pixels) sehr viele Iterationen nötig werden. Darin ist auch der Hauptnachteil des DBS-Algorithmus zu sehen: er ist um Größenordnungen langsamer als beispielsweise der IFT-Algorithmus.

Eine gewisse Beschleunigung wird dadurch erreicht, daß es bei einem einzelnen Umklappen zur Berechnung der Rekonstruktion nicht nötig ist, eine komplette zweidimensionale

²Dies legen auch Vergleiche mit Simulated-Annealing-Methoden (siehe [seld87]) nahe.

Fourier-Transformation durchzuführen. Stattdessen wird die bisherige Rekonstruktion abgeändert. Dies erfolgt einfach mittels der Grundformel für die DFT. Demnach ergibt sich die Änderung der bisherigen, rekonstruierten Amplitude g_{nm} zu

$$h'_{nm} = h_{nm} \pm \frac{1}{\sqrt{NM}} \exp \left[2\pi i \left(\frac{km}{M} + \frac{ln}{N} \right) \right] \quad (10.1)$$

Dabei steht das Pluszeichen für ein Umklappen von 0 nach 1 und das Minuszeichen für ein Umklappen von 1 nach 0. Die Koordinaten des umgeklappten Pixels sind (k, l) .

Wenn die Hologrammrekonstruktion nur in einem bestimmten Bereich benötigt wird ("Fensterung"), genügt es, (10.1) auch nur in diesem Bereich auszuwerten. Insbesondere bei im Vergleich zur Hologrammgröße ($N \times M$) kleinen Rekonstruktionsbereichen ($X \times Y$), typischerweise für $X \leq N/4$ und $Y \leq M/4$, ergeben sich mit dem DBS-Algorithmus sehr gute Ergebnisse ([jenn91], [mori95]).

Der Aufwand für (10.1) ist wesentlich geringer als der Aufwand für die komplette 2D-Fourier-Transformation. Allerdings muß man beachten, daß sich Fehler akkumulieren können. Wenn sehr viele Iterationen durchgeführt werden, dann summieren sich die Einzelfehler und die gesamte weitere Optimierung wird unsinnig. Es wird dann zwar weiter optimiert, die Optimierung paßt aber nicht mehr zur realen, physikalischen Hologrammrekonstruktion. Nach einer bestimmten Anzahl von Iterationen ist also eine komplette Neuberechnung der Iteration über die 2D-Fourier-Transformation³ nötig.

10.1.2 Fast-DBS-Algorithmen

Für ein 128×128 Hologramm liegt die typische Berechnungsdauer auf einem i486DX2/66-System (CGHMark = 1.0) in der Größenordnung von 3 Tagen; ein für viele Anwendungen sicher untragbarer Wert. Die Berechnung kann weiter beschleunigt werden, wenn es gelingt, die Berechnung des Rekonstruktionsfehlers direkt aus dem bisherigen Rekonstruktionsfehler und dem geänderten Pixel abzuleiten. Entsprechende Verfahren werden als *Fast DBS-Algorithmen* (*FDBS*) bezeichnet ([jenn89b], [jenn91]).

Dabei ist zu unterscheiden, ob wir uns wie bisher für den Fehler der Intensitäten (entsprechend Gleichung (5.1)) oder den Fehler der komplexen Amplituden der Rekonstruktion interessieren. Während der — für unsere Anwendungen entscheidende — Intensitätsfehler nur relativ kompliziert in eine rekursive Berechnung umgeformt werden kann und dementsprechend auch nur eine geringe Beschleunigung ermöglicht, ergibt sich beim Fehler der komplexen Amplitude eine deutliche Geschwindigkeitssteigerung. Die Optimierung komplexer Amplituden ist für Anwendungen, bei denen in der Bildebene sowohl die Amplitude, als auch die Phase festgelegt sein soll, interessant.

Die Beschleunigung bei Intensitätsfehlern beträgt für ein 64×64 Hologramm lediglich 1.2, während bei komplexen Amplituden eine Beschleunigung um den Faktor 11 erreicht wird.

Bei großen Hologrammen kann sich der erhöhte Speicherbedarf der rekursiven Berechnung eventuell negativ in der Berechnungsdauer niederschlagen.

³Ein oft vergessener Vorteil der FFT ist neben ihrer Geschwindigkeit die geringere Anzahl von Rechenoperationen, die zu einer im Vergleich zur straight-forward Berechnung erhöhten Genauigkeit führt.

FDBS mit komplexen Amplituden

Statt (5.1) verwenden wir als Fehlerformel

$$E = \sum_{x,y \in B} |f(f_x, f_y) - \lambda h(f_x, f_y)|^2 \quad (10.2)$$

Mit der Nomenklatur $\langle f, h \rangle := \sum_{f_x, f_y \in B} f(f_x, f_y) h^*(f_x, f_y)$ erhalten wir durch Einsetzen des Skalierungsfaktors λ^4

$$\lambda = \frac{\langle f, h \rangle}{\langle h, h \rangle} \quad (10.3)$$

für den Fehler

$$E = \langle f, f \rangle - \frac{|\langle f, h \rangle|^2}{\langle h, h \rangle} \quad (10.4)$$

Der erste Term ist konstant und daher für eine Optimierung uninteressant. Die restlichen Terme sollen nun rekursiv berechnet werden. Es gilt mit (10.1)

$$\langle f, h' \rangle = \langle f, h \rangle \pm \frac{1}{\sqrt{NM}} \sum_{m,n \in B} f(m, n) \exp \left[-2\pi i \left(\frac{km}{M} + \frac{ln}{N} \right) \right] \quad (10.5)$$

Den zweiten Term schreiben wir als F_{kl} . Die Matrix F_{kl} ist dabei unabhängig von der aktuellen Iteration und kann einmalig zu Beginn der Optimierung berechnet werden.

Für die Berechnung von E braucht man weiterhin den Term $\langle g, g \rangle$:

$$\langle h', h' \rangle = \langle h, h \rangle + \frac{AB}{MN} \quad (10.6)$$

$$\pm 2 \operatorname{Re} \left[\frac{1}{\sqrt{MN}} \sum_{m,n \in B} h(m, n) \exp \left[-2\pi i \left(\frac{km}{M} + \frac{ln}{N} \right) \right] \right] \quad (10.7)$$

$$= \langle h, h \rangle + \frac{AB}{MN} \pm 2 \operatorname{Re} [G_{kl}] \quad (10.8)$$

Dabei ist G_{kl} die DFT der aktuellen Rekonstruktion im Bereich B, also $G_{kl} = \text{DFT}[rh]$ mit der Fensterfunktion r . Diese Matrix ändert sich offensichtlich bei jedem Iterationsschritt. Da wir aber bei jedem Iterationsschritt nur ein Element der Matrix benötigen, macht es keinen Sinn, jedesmal die gesamte Matrix auf den neusten Stand zu bringen. Statt dessen merkt man sich alle erfolgreichen Umklappungen der letzten S erfolgreichen Iterationen. Diese Umklappungen werden mit a_s bezeichnet. a_s ist 1 für ein Umklappen von 0 nach 1 und -1 für ein Umklappen von 1 nach 0.

Damit kann man dann schreiben

$$h_{nm}^j = h_{nm}^{j-S} + \frac{1}{\sqrt{NM}} \sum_{s=1}^S a_s \exp \left[2\pi i \left(\frac{k_s m}{M} + \frac{l_s n}{N} \right) \right] \quad (10.9)$$

⁴ λ ist gegenüber der Definition in 5.2 abzuwandeln, da hier komplexe Amplituden optimiert werden (s. [zhua94] (6))

Für G_{kl} entsprechend

$$G_{kl}^j = G_{kl}^{j-S} + \frac{1}{\sqrt{NM}} \sum_{s=1}^S a_s R_{k-k_s, l-l_s} \quad (10.10)$$

Damit wird dann

$$\langle h', h' \rangle = \langle h, h \rangle + \frac{AB}{MN} + 2a \operatorname{Re} \left[G_{kl}^{j-S} + \frac{1}{\sqrt{MN}} \sum_{s=1}^S a_s R_{k-k_s, l-l_s} \right] \quad (10.11)$$

Nach S Iterationen wird G über eine FFT komplett neu berechnet. Die optimale Wahl von S ist entscheidend für eine optimale Geschwindigkeit des Algorithmus. Jennison und Allebach (siehe [jenn91]) leiten dieses optimale S anhand von Markov-Ketten ab. Es ergibt sich dabei

$$S = \sqrt{\frac{p C_u}{C_i}} - \frac{C_i}{2} \quad (10.12)$$

Dabei bezeichnet p die Wahrscheinlichkeit für erfolgreiches Umklappen⁵, C_u gibt dabei die Dauer der Berechnung von G_{kl} aus h an. Nach jedem erfolgreichen Umklappen werden die Kosten zur rekursiven Berechnung von E um den Faktor C_i vergrößert.

C_u und C_i ändern sich je nach verwendetem Computer und hängen von der Hologrammgröße ab. Sie werden daher experimentell bestimmt.

Für intensitätsbasierte Fehler (Formel (5.1)) ist das Ableiten einer rekursiven Berechnung des Fehlers leider nicht so einfach. Dementsprechend ist auch der Gewinn an Beschleunigung nur für sehr große Hologramme interessant. Ich vermeide daher, die entsprechenden Formeln abzuleiten. Für Details sei auf [jenn89] verwiesen.

Jennison und Allebach schlagen wegen der geringen Beschleunigung des Algorithmus für intensitätsbasierte Fehler vor, auch für reine Intensitätsobjekte den Algorithmus zur Optimierung von komplexen Amplituden zu verwenden. Entsprechend werden von ihnen zufällige Objektphasen gewählt. Die geringere Rekonstruktionsqualität, die so erreicht wird, wird ihrer Ansicht nach durch die stark beschleunigte Berechnung ausgeglichen. Meiner Meinung nach ist dies aber fragwürdig, da man gerade deshalb die direkten Suchmethoden einsetzen will, um die bestmögliche Qualität in der Rekonstruktion zu erreichen. Dies gilt um so mehr, als daß die Rekonstruktionen bei meinen Versuchen mit diesem Verfahren deutlich schlechter als mit dem intensitätsbasierten Algorithmus waren.

1994 wurde von Zhuang und Ersoy ([zhua94]) eine weitere Beschleunigung erreicht. Dazu werden implizite Symmetrien genutzt. Sie nennen den entsprechenden Algorithmus "*Fast decimation-in-frequency Direct Binary Search (FDFDBS)*".

Der Hauptaufwand für die Berechnung von $\langle h, h \rangle$ ist in (10.11) durch die Summation über die Elemente von R gegeben. Für rechteckförmige Rekonstruktionsbereiche ist $R = DFT[r]$ aber eine zweidimensionale sinc-Funktion und weist Nullstellen in einem regelmäßigen Raster auf. Diese Nullstellen können genutzt werden.

Zhuang und Ersoy erreichen dies, indem sie das Hologramm in Teilblöcke aufspalten. Entsprechend werden auch die Berechnungsformeln aufgespalten und es zeigt sich, daß die Anzahl elementarer Rechenoperationen bei geeigneter Wahl der Blocklagen und Blockgrößen stark vermindert werden kann.

⁵Vorsicht: selbige ändert sich natürlich im Lauf der Iterationen

Insbesondere am Anfang der DBS-Iterationen, wenn die Wahrscheinlichkeit für erfolgreiches Umklappen der Pixel noch groß ist, bringt das Verfahren eine deutliche Beschleunigung. Bei geringer Wahrscheinlichkeit für erfolgreiches Umklappen der Pixel ist das Verfahren dagegen langsamer als die einfache rekursive Beschleunigung. Dementsprechend erreicht ein hybrides Vorgehen, das beide Beschleunigungsvarianten verwendet, die insgesamt höchste Geschwindigkeit.

10.1.3 Variationen

Bisher sind wir davon ausgegangen, daß die Pixel, die umgeklappt werden sollen, *zufällig* gewählt werden. Es zeigt sich allerdings, daß es praktisch keinen Unterschied für das Endergebnis der Optimierung macht, wenn die Pixel fortlaufend *lexikalisch* umgeklappt werden. Auch andere Muster sind denkbar. Einen erwähnenswerten Geschwindigkeitsgewinn wird man allerdings durch diese vereinfachte Wahl der umzuklappenden Pixel nicht erreichen. Für die zufällige Wahl der Pixel genügt ein extrem einfacher und damit schneller Zufallszahlengenerator, so daß der zusätzliche Zeitbedarf praktisch vernachlässigbar ist.

Interessant wären aber *Heuristiken*, die es erlauben Pixel so auszuwählen, daß die Wahrscheinlichkeit einer Verbesserung beim Pixelumklappen erhöht wird. Insbesondere gegen Ende der DBS-Optimierung wird die Wahrscheinlichkeit, daß ein Umklappen zu einer Verbesserung führt, sehr gering. Eine Erhöhung dieser Wahrscheinlichkeit würde den Algorithmus deutlich beschleunigen. Leider sind bisher keine entsprechenden Heuristiken bekannt.

Eine rechentechnische Beschleunigung kann — je nach verwendetem Computersystem⁶ — durch eine *Tabellierung* der in (10.1) auftretenden trigonometrischen Funktionen erzielt werden.

Peter et al. (siehe [pete92]) berichten über eine deutliche Verbesserung der erreichbaren maximalen Beugungseffizienz durch eine spezielle Wahl der *Hologramm Startverteilung*. Im selben Artikel werden allerdings sehr merkwürdige (theoretisch eigentlich kaum mögliche) Ergebnisse von Rechnungen mit dem Simulated Annealing Algorithmus präsentiert. Falls die direkten Suchmethoden — wider Erwarten — weiter untersucht werden, kann entsprechend die in [pete92] angegebene Startverteilung nochmals überprüft werden.

Üblicherweise wird man als Fehlermaß und damit als Optimierungsziel das Fehlermaß nach (5.1) verwenden. Statt den Skalierungsfaktor λ in dieser Formel jeweils neu zu berechnen, kann er auf einen bestimmten, festen Wert gesetzt werden. Damit ist auf einfache Weise festgelegt, welche Beugungseffizienz angestrebt werden soll und die Beugungseffizienz kann zugunsten der Rekonstruktionsqualität erhöht werden.

Eine Verallgemeinerung von binären Hologrammen hin zu Hologrammen mit weiteren Quantisierungsstufen ist trivial (zusätzlich wird bei jeder Pixeländerung noch per Zufall bestimmt, um “wieviel” der Pixel geändert wird), erhöht die Rechenzeit aber ganz erheblich (exponentiell mit der Zahl der Quantisierungsstufen) und kommt damit für die meisten Anwendungen nicht in Frage.

Eine weitere Beschleunigung kann eventuell für bestimmte einfache Objektklassen mit inhärenter Symmetrie erreicht werden, indem die Fehlerberechnung weiter vereinfacht wird. Das könnte vor allem für Fan-out Elemente oder einfache Strahlformer interessant sein.

⁶Superskalarität, Cache und Objektgröße sind dabei entscheidend

Boutenko und Chevallier ([bout96]) haben kürzlich eine interessante Variante des DBS-Grundalgorithmus vorgeschlagen. Der DBS-Algorithmus wird solange durchgeführt, bis keine Verbesserung mehr durch ein Pixelumklappen erzielt werden kann, also die Optimierung in einem lokalen Minimum feststeckt. Dann werden zwei Pixels gleichzeitig umgeklappt. Es ist durchaus möglich, daß bei einem solchen Doppelumklappen eine weitere Reduktion des Fehlers stattfindet.

Auf den ersten Blick scheint dieses Vorgehen sehr zeitintensiv zu sein, gibt es doch extrem viele mögliche Kombination von zwei Pixeln. Die Autoren zeigen aber analytisch, daß nur ein Bruchteil dieser Kombinationen eine Verbesserung bewirken kann und untersucht werden muß. Die Verbesserungen der Rekonstruktionsqualität, die mit diesem *Second order DBS* erzielt werden, liegen im Bereich von 35 bis 75 % sowohl für komplexe als auch für intensitätsbasierte Optimierungen.

Die Verallgemeinerung der direkten Suchmethoden (auch die im folgenden zu besprechenden Methoden wie Simulated Annealing, Threshold Accepting und genetische Algorithmen) auf drei Dimensionen ist trivial: es werden einfach die Rekonstruktionen in verschiedenen Ebenen berechnet und die entsprechenden Rekonstruktionsfehler der Teilebenen addiert. Nachteil ist dabei, daß der Berechnungsaufwand linear mit der Anzahl der Ebenen steigt, was sich natürlich wegen der an sich schon großen Berechnungskomplexität negativ bemerkbar macht.

Man kann auf diese Verlangsamung verzichten, wenn man Qualitätseinbußen in Kauf nimmt. Dazu wird eine Objektebene gewählt. In dieser Objektebene wird eine komplexe Amplitude festgelegt, die alle Teilebenen korrekt rekonstruiert. Dies ist kein Problem. Dazu werden einfach die komplexen Amplituden der Teilbenen in dieser Objektebene einmalig berechnet. Für die sich ergebende komplexe Amplitude wird dann mit dem komplexen DBS-Algorithmus ein Hologramm berechnet. Nachteil ist, daß die Objektphase nicht optimiert werden kann und als Freiheitsgrad wegfällt.

Alle besprochenen Variationen lassen *nicht* den direkten Einsatz der im letzten Abschnitt besprochenen beschleunigten DBS-Algorithmen zu.

10.1.4 Beurteilung des DBS-Algorithmus

Seldowitz et al. (siehe [seld87]) haben den DBS-Algorithmus ausführlich mit einigen einfachen Detour-Phasen-Hologrammen und einem einfachen Error-Diffusions-Algorithmus für verschiedene Versuchsbedingungen verglichen. Zusammenfassend kann gesagt werden, daß der DBS-Algorithmus sowohl hinsichtlich der Beugungseffizienz als auch der Rekonstruktionsqualität deutlich bessere Ergebnisse liefert, was allerdings angesichts der einfachen, nicht-iterativen (und damit schnellen) Vergleichsverfahren kaum verwundert. Aber auch der Vergleich mit der komplexen Abwandlung der IFT-Algorithmen, dem POCS-Algorithmus (siehe Abschnitt 8.1.3) liefert deutlich bessere Rekonstruktionsergebnisse und höhere Beugungseffizienzen (siehe [jenn89b], [hopt89]). Auch die intensitätsbasierte DBS-Variante liefert für binäre Hologramme hinsichtlich Beugungseffizienz und Rekonstruktionsfehler bessere Ergebnisse als andere Algorithmen, insbesondere als der IFT-Algorithmus mit Binarisierung (siehe [zhan95]).

Die besseren Rekonstruktionsergebnisse dürfen allerdings nicht zu hoch bewertet werden. Die Hologramme des DBS-Algorithmus leiden genauso unter den Speckle-Effekten wie andere Verfahren auch. Dieser Fehler, der aufgrund von Speckles auftritt, ist der domi-

nierende Anteil im Gesamtfehler. Eine Verwendung der DBS-Algorithmen kann dementsprechend für die meisten Anwendungen den optischen Gesamtfehler im Vergleich zu den IFT-Algorithmen nur gering vermindern und es ist dementsprechend fraglich, ob sich der erhöhte Rechenaufwand lohnt. Unsere konkreten Vergleiche optischer Rekonstruktionen bestätigten dies dahingehend, daß kaum ein visueller Unterschied in der Qualität der DBS-Hologramme im Vergleich zu IFT-Hologrammen festgestellt werden konnte.

Der Aufwand für den DBS-Grundalgorithmus ist von der Ordnung $(NM)^4$, kann aber durch die Beschleunigung auf $(NM)^{3/2} \sqrt{\log_2 NM}$ bei komplexen Amplituden vermindert werden (siehe [jenn91]).

Die beschleunigten Varianten des Algorithmus haben den Nachteil, daß eine Modifikation der Rekonstruktionsformel (um z.B. Nichtlinearitäten des Hologrammediums zu berücksichtigen) oder eine Verallgemeinerung auf nicht-binäre Hologrammedien oder andere Fehlerformeln nicht einfach möglich ist. Für jede entsprechende Modifikation muß erneut die beschriebene Ableitung wiederholt werden, um zu einer Beschleunigung zu gelangen (falls dies im Einzelfall überhaupt möglich ist).

Interessant sind alle Verfahren der direkten Suche für einfache Rekonstruktionsobjekte (z.B. ein 4×4 Fan-out Element). In diesem Fall ist der Aufwand für die Berechnung des Rekonstruktionsfehlers gering und die Algorithmen werden praktikabel.

Desweiteren scheint die Rechenzeit für die Berechnung von einzelnen Hologrammen (z.B. für Strahlformungsaufgaben) tragbar zu sein. Für Anwendungen, bei denen sehr viele, eher komplizierte CGHs benötigt werden (und genau deshalb verwenden wir ja LCDs als Hologrammedien), sind dagegen schnellere Verfahren zu bevorzugen.

Momentan wäre bei uns der Einsatz von DBS-Algorithmen nur für Anwendungen zur Rekonstruktion komplexer Amplituden (und das ist eigentlich kurzfristig nicht geplant) interessant. In diesem Fall sollte die aktuellste Variante des DBS-Algorithmus (siehe [zhua94]) verwendet werden.

10.2 Simulated Annealing (SA) und Threshold Accepting (TA)

Bisher sind wir davon ausgegangen, daß nur erfolgreich umgeklappte Pixels auch umgeklappt bleiben. Es ist anschaulich klar, daß ein solches Vorgehen in ein lokales Optimierungsminimum führt. Das globale Minimum wird nicht erreicht. Abwandlungen einfacher Optimierungen, die lokale Minima verlassen können, sind bekannt. Bereits bei der ersten Vorstellung des DBS-Algorithmus ([seld87]) wurde der Einsatz des wohl wichtigsten Vertreters dieser Klasse, dem *Simulated Annealing (SA) Algorithmus*, besprochen.

Der SA-Algorithmus ([kirk83], [otto94], [zell94]) ist eine sehr einfache Abwandlung der DBS Optimierung: wenn ein Pixelumklappen die Rekonstruktion verschlechtert, dann wird nicht wie bisher dieses Umklappen rückgängig gemacht, sondern es wird stattdessen mit einer bestimmten Wahrscheinlichkeit beibehalten. Damit wird es möglich, auch aus lokalen Minima zu entkommen. Die Wahrscheinlichkeit ist dabei dynamisch im Laufe der Iteration abzusinken.

Diese Art der Optimierung wurde direkt von dem physikalischen Modell des Abkühlens von Spingläsern übernommen. Auch dort wird mit abnehmender Temperatur des Glases die

Wahrscheinlichkeit für eine “ungünstige” Umorientierung von Spins immer weiter abnehmen. Besondere Popularität gewannen die Simulated-Annealing-Methoden insbesondere mit dem Einsatz von Neuronalen Netzen, bei denen die Gewichte zwischen den Neuronen (“Synapsen”) über diesen Algorithmus optimiert werden können. Um dem physikalischen Modell näher zu kommen, bestimmt man meist die Wahrscheinlichkeit aus einer künstlichen Temperatur über die Fermi- oder Boltzmann-Verteilung.

Ganz analog zum physikalischen, langsamen Abkühlen ist es entscheidend, wie konkret die Temperatur bzw. die Wahrscheinlichkeit im Laufe der Zeit abgesenkt wird. Senkt man die Temperatur zu schnell ab, wird sich nicht das globale Minimum ergeben. Lediglich bei ausreichend langsamem Abkühlen kann man hoffen, ein globales Minimum zu erreichen.

Ein sehr langsames Abkühlen bedeutet aber gleichzeitig eine starke Verlangsamung des Algorithmus. Da der Grund-DBS-Algorithmus an sich schon sehr langsam ist, wirkt sich dies natürlich besonders negativ aus. Hier liegt ein Hauptproblem, da momentan keine konkreten Aussagen getroffen werden können wie der Temperaturverlauf für verschiedene Optimierungsprobleme konkret gewählt werden soll. Man ist hier bislang auf mehr oder weniger langwieriges Experimentieren mit dem Algorithmus angewiesen.

Seldowitz, Allebach und Sweeney benutzen zum Beispiel folgendes Abkühlungsschema: Die Temperatur T wird gewählt über

$$T = \left(\frac{\alpha}{T_0}\right)^p T_0 \quad (10.13)$$

Dabei ist α ein Parameter, der die Geschwindigkeit des Prozesses regelt und T_0 ist die Starttemperatur. p ist die Iterationsnummer.

Daraus wird dann die Wahrscheinlichkeit W für die Übernahme eines Umklappens, welche eine Verschlechterung um ΔE bewirkt, nach

$$W = \exp(-\Delta E/T) \quad (10.14)$$

berechnet.

Im allgemeinen wird davon ausgegangen, daß der stark vergrößerte Rechenaufwand beim SA-Algorithmus nur zu einer kaum erwähnenswerten Verbesserung der Rekonstruktionsergebnisse führt (siehe z.B. [seld87], [clar93], [bout96], [erso92]). Boutenko gibt eine durchschnittliche Verbesserung von 3.7 % an. Diese geringe Verbesserung legt weitere Arbeiten hinsichtlich der Verwendung beschleunigter SA-Varianten, wie das *Fast-Simulated-Annealing* ([szu87]) oder das *Very-Fast-Simulated-Reannealing* ([ingb89]) nicht nahe.

Teilweise wird zunächst mit einem Simulated Annealing Ansatz ein lokaler Bereich mit guter Beugungseffizienz gesucht, um dann mit einem gewöhnlichen Gradientenverfahren die weitere Optimierung durchzuführen. Damit ist sichergestellt, daß das Verfahren nicht in einem sehr “schlechten” lokalen Minimum konvergiert (z.B. [weib93]).

SA-CGHs zur Generierung von Spot-Arrays (z.B. für digital-optische Informationsverarbeitung) finden sich in [dame91], [vasa92] und [hopt89].

Eine Abwandlung des Simulated Annealing ist der *Threshold Accepting (TA) Algorithmus* (siehe [ducc90], [ducc93]). Während beim SA lediglich Wahrscheinlichkeiten für ein Verwerfen von Umklappvorgängen angegeben werden, gibt es bei TA feste Schwellen. Wenn die Verschlechterung größer als ein bestimmter Schwellwert T , der im Laufe der Zeit im Sinne einer Temperatur abgesenkt wird, ist, dann wird der Umklappvorgang verworfen.

Der TA Algorithmus hat gegenüber dem SA Algorithmus den Nachteil, daß ein Entkommen aus steilen lokalen Minima oft nicht möglich ist. Vorteilhaft ist aber, daß er sich aus vielen weniger steilen lokalen Minima sehr viel schneller als der SA Algorithmus befreien kann. Insgesamt ist der TA-Algorithmus deutlich schneller und liefert für komplizierte Probleme meist gleichwertige Lösungen. Teilweise (z.B. für das Handlungsreisenden Problem) werden bei gleicher Anzahl von Iterationen auch deutlich bessere Ergebnisse als mit SA-Algorithmen erreicht.

Bislang wurde der TA-Algorithmus meines Wissens noch nicht zur Berechnung computergenerierter Hologramme verwendet. Wenn in Zukunft weiter in Richtung direkter Suchverfahren gearbeitet werden sollte, wäre eine Untersuchung dieses Verfahrens aber in jedem Fall interessant.

Kapitel 11

Hologramm–Quantisierung

Viele, insbesondere einfache, CGH–Algorithmen liefern kontinuierliche Amplituden und/oder Phasenverteilungen, die auf dem Hologramm zu kodieren sind. Viele Hologrammedien sind aber nur in der Lage, einige wenige Quantisierungsstufen darzustellen. Typische Beispiele für solche Medien sind über Masken belichtete Schichten, aber auch LCDs. Grundsätzlich kann gesagt werden, daß die Darstellung von Hologrammen um so komplizierter wird, je mehr Quantisierungsstufen verwendet werden sollen.

Vor der Ausgabe des Hologramms muß dementsprechend eine Quantisierung vorgenommen werden, um die Daten dem Hologrammedium anzupassen. In der Praxis besonders wichtig ist die Quantisierung auf zwei diskrete Stufen, die *Binarisierung*.

Einige der bisher besprochenen Algorithmen liefern bereits quantisierte Hologramme. Die direkten Suchverfahren ergeben prinzipbedingt quantisierte (in der Regel binäre) Hologramme. Andere Verfahren, insbesondere die iterativen Fourier–Transformations–Algorithmen, lassen sich dahingehend erweitern, daß innerhalb des Algorithmus selbst eine Quantisierung stattfindet.

In diesem Kapitel werden nun Quantisierungen besprochen, die *nach* der eigentlichen Berechnung des Hologramms angewendet werden. Die Wahl einer günstigen Quantisierung ist um so bedeutender, je weniger Quantisierungsstufen das Hologrammedium zur Verfügung stellt. Für theoretische Abschätzungen hinsichtlich des Quantisierungsfehlers für Phasenhologramme sei auf [pahl82] verwiesen.

Da wir zunächst den Einsatz eines Video–LCDs mit 256 verschiedenen Grau– bzw. Phasenstufen planten, schien die Quantisierung ein eher untergeordnetes Problem zu sein und wurde dementsprechend wenig beachtet. Inzwischen wird allerdings der Einsatz eines VGA–LCDs mit 4 Quantisierungsstufen favorisiert. Die Quantisierung wird dementsprechend wichtiger für uns. Aus Zeitgründen konnten aber leider nicht alle wichtigen Quantisierungsverfahren geprüft werden.

Unbeachtet bleiben in diesem Kapitel ältere, relativ einfache Verfahren zur Quantisierung. Diese Verfahren werden nur noch selten eingesetzt und es existiert eine nicht unerhebliche Anzahl von entsprechenden Varianten.

Bereits im Abschnitt 7.2 wurde die wohl einfachste Variante des Quantisierens, das *Top–Clipping*, kurz mit ihren Vor– und Nachteilen besprochen. Weitere Verfahren findet man bei [lee74]

11.1 Pulse-Width Area-Modulation (PWM) und Pulse-Density Modulation (PDM)

Nehmen wir an, wir wollen das diskrete Amplitudenhologramm $H(k\Delta u)$ kodieren. Für ein kontinuierliches Hologrammedium würde man als Ausgabe $G(u)$ einfach

$$G(u) = H(k\Delta u) * \text{rect}(u, \Delta u) \quad (11.1)$$

wählen. Das Hologramm besteht damit aus gleich großen Zellen der Breite Δu . Die Schwärzung der Zellen ist dabei proportional zu H .

Bei der Binarisierung über eine Pulsbreiten-Modulation (PWM) wählt man stattdessen

$$G(u) = \text{comb}(u, \Delta u) * \text{rect}(u, w(k)) \quad (11.2)$$

Dabei kodiert die Pulsbreite $w(k)$ indirekt die Amplitude und ist zum Beispiel durch

$$w(k) = (H(k)\Delta u, \Delta v) \quad (11.3)$$

gegeben (siehe [bryn90]). Anschaulich wird die Breite einzelner Pulse variiert. Die Schwärzung gibt die Position von Interferenzstreifen, wie sie z.B. in Gleichung (9.4) gegeben war, an. Verschiedene andere Möglichkeiten zur Wahl von $w(k)$ findet man bei Lee ([lee74]).

Nachteil dieses Verfahrens ist offensichtlich, daß eine hohe Auflösung des Hologrammediums zu fordern ist, um die variablen Breiten der kodierten Rechtecke darstellen zu können. Das Verfahren macht dementsprechend für LCDs wenig Sinn.

Statt die Amplitude über die Breite zu kodieren, kann auch die Anzahl der Pulse variiert werden, um denselben Effekt zu erreichen. Dies ist die übliche anschaulich Quantisierung. Das diskrete $H(k)$ wird zunächst quantisiert und dann gemäß (11.1) auf gleichgroße Pixel verteilt. Diese Quantisierung nennen Bryngdahl und Wyrowski ([bryn90]) eine *Pulsdichtenmodulation*. Mir erscheint diese Bezeichnung ungünstig, da der Begriff dahingehend mißverstanden werden kann, daß kleine Pulse mit mehr oder weniger Zwischenraum zur Kodierung der Amplitude verwendet werden könnten (was im Fall computergenerierter Hologramme keinen Sinn macht). Diese Art der Quantisierung wird von mir daher als "direkte" Quantisierung bezeichnet. In der Praxis werden dabei einfach die einzelnen Pixel quantisiert.

11.2 Pulsbreitenmodulation mit variabler Schwelle

Grundidee dieses Verfahrens ist die Verwendung eines Hard-Clippings (nach (8.5)). Allerdings wird die Clipping-Schwelle b variabel gewählt. Die Schwelle b errechnet sich aus dem zu kodierenden Signal nach Just und Bryngdahl ([just87]) zu

$$b(u, v) = \sqrt{1 - H^2(u, v)} \quad (11.4)$$

Die Rekonstruktionseigenschaften lassen sich zum Teil erheblich durch die Wahl von b beeinflussen (siehe [just85]).

Motiviert wird dieser Ansatz durch eine entsprechende Binarisierung von gewöhnlichen Bildern, bei denen die Schwelle periodisch (z.B. dreieckförmige Trägerwelle) gewählt wird (*Trägerfrequenz PWM*). (Diese Art der Binarisierung wurde allerdings von Hauck und Bryngdahl ebenso zur Hologrammquantisierung verwendet ([just85]).

Das Vorgehen ist anschaulich in Abbildung 11.1 dargestellt.

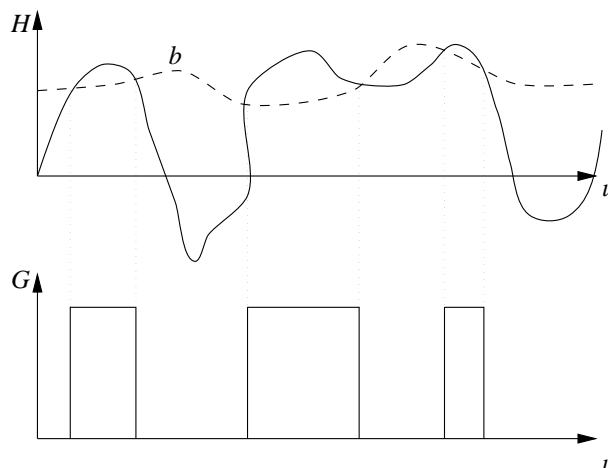


Abbildung 11.1: Quantisierung durch Pulsbreitenmodulation

Weitere Information zu diesem Verfahren finden sich bei [just87].

11.3 Nutzung des Babinet-Effekts

Bei der Beugung an einer Maske erhält man dasselbe Beugungsbild wie bei der Beugung an der entsprechenden inversen Maske, d.h. einer Maske, bei der alle transparenten Stellen undurchlässig gewählt werden und alle undurchlässigen Stellen transparent werden. Die Beugungsbilder unterscheiden sich lediglich in einem für die meisten Anwendungen unerheblichen Phasenfaktor π .

Dieses Phänomen wird *Babinet-Effekt* genannt (siehe z.B. [ditc61] 6.21). Im Falle von Fourier-Hologrammen ist der Effekt trivial. Die Invertierung des binären Hologramms wird durch eine Multiplikation des Hologramms mit $-1 = e^{-\pi}$ erreicht. Dieser konstante Faktor wird von der Fourier-Transformation nicht beeinflusst. Dementsprechend wird auch die Fourier-Transformierte lediglich mit -1 zu multiplizieren sein, was der entsprechenden Phasenschiebung von π entspricht.

Eine allgemeine Begründung ergibt sich direkt aus dem Superpositionsprinzip. Wenn das Beugungsbild ohne Apertur konstante Amplitude besitzt, dann muß nach dem Superpositionsprinzip die beschriebene Beziehung zwischen den beiden Beugungsbildern komplementärer Aperturen bestehen.

Eine wesentliche Änderung ergibt sich allerdings dennoch. Die Energie des gebeugten Lichts wird sich ändern. Es ist unmittelbar klar, daß auf einem entfernten Schirm ein einzelnes Loch in einer Apertur eine wesentlich geringere Intensität wie ein einzelner absorbierender Punkt in einer freien Apertur erreicht¹. Ganz entsprechend wird das invertierte Hologramm

¹Ein Effekt der im übrigen zur Verbesserung von optischen Nahfeld-Mikroskopen eingesetzt werden soll

eine höhere Beugungseffizienz erreichen, wenn bei diesem die transparente Fläche, die undurchlässige Fläche übersteigt. Nach der Berechnung binärer Amplituden-Hologramme sollte also in jedem Fall überprüft werden, ob ein einfaches Invertieren des Hologramms nicht zu einer erhöhten Beugungseffizienz führt.

Just, Hauck und Bryngdahl ([just85]) berichten über einige kleinere Experimente, bei denen nur Teile des Hologramms invertiert werden (*lokaler Babinet Effekt*). In diesem Fall ist das Babinet-Prinzip auch nur für die Teilbereiche anwendbar und das gesamte Beugungsbild wird sich ändern. Mit der entsprechenden Technik konnten verschiedene Beugungsordnungen unterschiedlich beeinflußt werden. Eine systematische Verbesserung wurde damit aber nicht erzielt.

In neuerer Zeit haben Lohmann und Sinzinger ([lohm92]) den Effekt erneut aufgegriffen und auf lokale Bereiche innerhalb von CGHs angewendet. Direkt verwertbare Ergebnisse finden sich in dieser eher theoretischen Arbeit allerdings ebenfalls nicht.

11.4 Interlacing und Iterative Interlacing

Interlacing zur Quantisierung von Hologrammen geht zurück auf Ersoy, Zhuang und Brede ([erso92]).

Grundidee des Interlacing ist es, das gesamte Hologramm in Teilhologramme aufzuspalten. Das 1. Teilhologramm dient als Grundhologramm. Seine Rekonstruktion wird das Objekt annähernd korrekt darstellen. Der noch verbleibende Fehler aufgrund der Quantisierung wird mittels des 2. Teilhologramms ausgeglichen. Auch das gelingt nicht voll, so daß ein weiteres Teilhologramm nötig wird. Insgesamt wird durch den Einsatz von N Teilhologrammen der Gesamtfehler vermindert.

Die Teilhologramme können mit praktisch allen besprochenen CGH-Algorithmen für komplexe Rekonstruktionen berechnet werden. Insbesondere kommen iterative Fourier-Transformations-Algorithmen (POCS), direkte Suchverfahren (DBS) und eine Kombination mit der Error-Diffusion (siehe Abschnitt 11.6) in Frage. Insbesondere die Kombination mit Error-Diffusion ([chan93]) liefert für intensitätsbasierte Hologramme in den Simulationen von Chang und Ersoy sehr gute Ergebnisse.

Die Teilhologramme werden ineinander verschachtelt. Zwei mögliche Geometrien für zwei ineinander verschachtelte Teilhologramme sind in Abbildung 11.2 dargestellt.

Wenn uns eine bestimmte Anzahl von Gesamtpixeln im Hologramm zur Verfügung steht, dann ist klar, daß die Anzahl der Pixel pro Teilhologramm mit der Anzahl der Teilhologramme abnimmt. Das hat anschaulich einen negativen Effekt auf die Qualität der Rekonstruktion, der die Verbesserung aufgrund des Interlacings teilweise wieder aufwiegt. Offensichtlich wird es eine optimale Anzahl von Teilhologrammen geben, um die Gesamtqualität der Rekonstruktion zu optimieren. In den Experimenten von Ersoy et al. ergab sich ein Optimum für 4 Teilhologramme (allerdings sollte man es vermeiden, diesen Wert ohne Prüfung für verschiedene Anwendungen zu übernehmen).

Um eine weitere Verbesserung der Rekonstruktionsqualität zu erzielen, wird zum *Iterative Interlacing* übergegangen ([erso92]). Der nach dem letzten Teilhologramm noch verbleibende Fehler wird zurückgekoppelt und zu einer weiteren Optimierung des 1. Teilhologramms verwendet. Dazu wird einfach das "Zielobjekt" f entsprechend

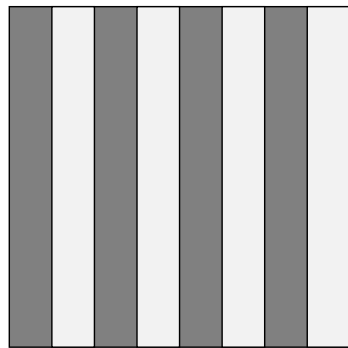
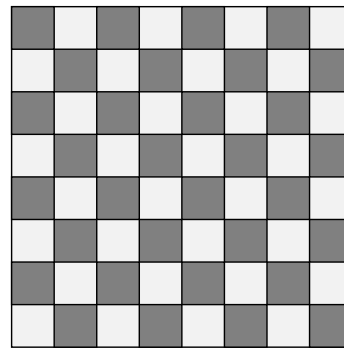
a.) *Streifenmuster*b.) *Schachbrettmuster*

Abbildung 11.2: mögliche Hologramm-Geometrien beim Iterative-Interlacing Algorithmus

$$f'(f_x, f_y) := f(f_x, f_y) - \epsilon(f_x, f_y) \quad (11.5)$$

modifiziert, wobei $\epsilon(f_x, f_y)$ den verbleibenden Fehler des letzten Teilhologramms bezeichnet. Für dieses neue, abgewandelte Objekt wird der komplette Interlace-Algorithmus erneut durchgeführt. Insgesamt sind zwischen 10 und 20 komplette Durchgänge nötig, bis keine Verbesserungen mehr erzielt werden.

Eine leichte Abwandlung dieser Technik (siehe [chan93])) verwendet in (11.5) einen zusätzlichen Skalierungsfaktor², der die Rekonstruktionsqualität weiter verbessern soll.

Es erscheint zunächst unnötig, die Teilhologramme mittels DBS-Methoden zu berechnen. Die DBS-Algorithmen liefern per se binäre Hologramme und machen den Einsatz der Interlacing Technik eigentlich nicht nötig. Da der Aufwand für die Berechnung von DBS-Hologrammen aber sehr stark von der Hologrammgröße abhängt, ist eine Beschleunigung durch die Aufteilung in Teilhologramme möglich. Ersoy et al. geben eine erzielte Beschleunigung um den Faktor 5 an. Erstaunlich ist in diesem Zusammenhang, daß auch die erreichbare Rekonstruktionsqualität trotz der Geschwindigkeitssteigerung (geringfügig) verbessert werden kann. Die Autoren meinen einen gewissen “Simulated Annealing” Effekt zu erzeugen, der das Verlassen lokaler Minima ermöglicht. Eine einleuchtende Begründung dafür fehlt allerdings.

11.5 Neuronale Netze zur Binarisierung

Neuronale Netze werden heute mit mehr oder weniger großem Erfolg für völlig verschiedene Anwendungen eingesetzt (siehe z.B. [zell94]). Dementsprechend wurde 1991 von Just und Ling ([just91]) ein Binarisierungsverfahren für CGHs, basierend auf neuronalen Netzen vorgeschlagen.

Grundidee eines neuronalen Netzes ist eine Menge von N Neuronen n_i über Gewichte, sogenannte *Synapsen*, miteinander zu verbinden und eine einfache Dynamik nach folgender Gleichung anzunehmen:

²der in der Praxis zwischen 1 und 2 gewählt wird

$$n_i = \Theta\left(\sum_{j=1}^N w_{ij}n_j + s_i\right) \quad (11.6)$$

Dabei ist $\Theta(x)$ eine nichtlineare Funktion (z.B. $\tanh(x/\alpha)$) und s_i ein konstanter Schwellwert. Für alle N Neuronen wird (11.6) fortlaufend berechnet³. Die Gewichte w_{ij} bestimmen letztlich die Funktion des Netzes und es existieren viele verschiedene Sonderformen von Netzen, die sich in impliziten Symmetrien der Gewichte unterscheiden. Dementsprechend existieren auch viele verschiedene Lernverfahren, die die Gewichte setzen. Üblicherweise lernt ein neuronales Netz aus einer Menge an Kombination von Ein- und Ausgabevektoren. Bei der Hologrammberechnung fallen die Eingabeneuronen weg. Die innere Dynamik des Netzes liefert die gewünschten Ausgaben.

Ein unter Physikern besonders beliebtes Netz ist das Hopfield-Netz ([hopf82], [hopf86]). Hier wird nicht im eigentlichen Sinn gelernt, sondern es wird analytisch eine Energiefunktion abgeleitet, die konkret in Abhängigkeit von den Gewichten w_{ij} eine Energie E angibt, die vom Netzwerk im Laufe der durch (11.6) beschriebenen Iterationen minimiert wird.

Diese Energie ergibt sich zu (siehe [zell94] (17.3))

$$E = -\frac{1}{2} \sum_{ij} n_i w_{ij} n_j - \sum_i n_i s_i \quad (11.7)$$

Eine Konvergenz der Iterationen wird nur erreicht, wenn zusätzliche Bedingungen, die hier nicht aufgeführt werden sollen (siehe [tutt93]), an die Gewichtsmatrix w_{ij} gestellt werden.

Es genügt nun, ein zu lösendes Problem auf die Form (11.7) zu transformieren und daraus dann die Gewichte w_{ij} abzulesen. Das entsprechende Netzwerk wird dieses Problem dann lösen.

Just und Ling beschäftigen sich mit der CGH-Binarisierung für komplexe Rekonstruktionen. Dementsprechend verwenden sie als Kostenfunktion

$$E = \sum_{ij} \Lambda_{ij} |f_{ij} - h_{ij}| \quad (11.8)$$

Dabei ist Λ eine zusätzliche Gewichtsfunktion, die angibt wie stark die Fehler in verschiedenen Bildbereichen zu berücksichtigen sind (im einfachsten Fall wäre das ein rechteckiger Fensterbereich).

Leider ergibt sich bei dieser eleganten und allgemeinen Vorgehensweise ein Problem. Für die Binarisierung bietet es sich unmittelbar an, jedem Hologrammpixel ein Neuron zuzuordnen. Der Zustand des Neurons n_i gibt dementsprechend den Zustand des Hologrammpixels (Schwärzung bzw. Phase) an. Für in der Praxis auftretende Hologramme wird dann allerdings die Gewichtsmatrix sehr groß und unhandlich. Entsprechend ist natürlich auch der Berechnungsaufwand enorm.

³Verschiedene Möglichkeiten zur Auswahl des jeweils zu berechnenden Neurons n_i stehen zur Verfügung. Die folgenden Formeln gelten für das sogenannte *serielle Update*, bei dem jeweils ein Neuron neu berechnet und geändert wird. Beim *parallelen Update* werden dagegen alle Neuronen *gleichzeitig* geändert.

Glücklicherweise kann man durch eine kleine Näherung das neuronale Netz für die CGH-Binarisierung, so darstellen, daß sich (11.6) über FFTs berechnen läßt und sich die Gewichtsmatrix auf eine zweidimensionale Matrix mit gleichvielen Elementen wie Hologrammpixeln reduziert. Erst durch diese enorme Vereinfachung wird der Ansatz letztlich praktikabel. Die Ableitung der entsprechenden Formeln findet man in [just91]. Es ergibt sich eine erstaunliche Ähnlichkeit des Verfahrens mit den IFT-Algorithmus, auf die in [tutt93] näher eingegangen wird.

Eigene Versuche mit der Methode ergaben nur eine vernachlässigbar kleine Verbesserung gegenüber einer einfachen Quantisierung mittels Top-Clipping. (Die Autoren geben für ein Beispielhologramm eine Verbesserung der Rekonstruktionsqualität gegenüber dem Error Diffusion Algorithmus von ca. 30 % an).

Außerdem kommt das Verfahren in dieser Form nicht direkt für uns in Frage, da — wie beschrieben — die komplexen Amplituden der Rekonstruktion (im Gegensatz zur eigentlich interessierenden Intensität) optimiert werden.

Eine neuere, umfassendere Besprechung (auch für intensitätsbasierte Rekonstruktionen) neuronaler Netze für die Hologrammberechnung findet man bei Li et al. ([li95]). Dort wird eine den DBS-Methoden vergleichbare Rekonstruktionsqualität angegeben.

11.6 Error-Diffusion

Die Binarisierung von Bildern (statt von Hologrammen) ist ein altes und technisch bedeutsames Problem. Die Technik wird in Massenprodukten wie Druckern und Fax-Geräten, aber auch für Zeitungen eingesetzt.

1976 haben Floyd und Steinberg ein neues Verfahren vorgestellt, mit dem Bilder binarisiert werden können. Das Verfahren nennt sich *Error-Diffusion* und ist mittlerweile das meist gebrauchte Binarisierungsverfahren im Computergraphik-Bereich.

Grundidee des Algorithmus ist es, einen bei der Quantisierung eines einzelnen Pixels gemachten Fehler auf die Nachbarpixel zu verteilen und bei der späteren Quantisierung dieser Pixel zu berücksichtigen. Wie dieser Fehler auf die benachbarten Pixel verteilt werden soll, kann im zweidimensionalen Fall⁴ nicht allgemein beantwortet werden. Hier werden verschiedene (auch lokal variierende) Verteilungen eingesetzt. Auf die zugrundeliegende Theorie zur "optimalen" Wahl der Koeffizienten kann hier nicht eingegangen werden. Für Details sei insbesondere auf [fett95], [barn88], [weis92] sowie auf [esch91] für einen Vergleich unterschiedlicher ED-Methoden verwiesen.

Der Algorithmus wurde von Hauck und Bryngdahl zur Quantisierung computergenerierter Hologramme ([hauc84]) verwendet. Im Gegensatz zur Anwendung der ED auf Bilder, liefert die eindimensionale ED im Fall von Hologrammen in der Regel gleich gute oder sogar bessere⁵ Ergebnisse ([casa91]). Der Quantisierungsfehler einer Spalte wird dabei auf die Hologrammpixel derselben Spalte verteilt, was außer für sehr kleine Hologramme ($N < 64$) eine ausreichende Mittelung ergibt.

Es ist erstaunlich, daß die Gesamtstruktur des Hologramms unabhängig von der Wahl der

⁴im eindimensionalen ist der optimale Koeffizient 1 (siehe [casa91])

⁵weil die optimalen Diffusionskoeffizienten im zweidimensionalen Fall nicht exakt bestimmt werden können

Binarisierungsschwelle ist ([hauc84]).

Entscheidend für das Funktionsprinzip des Algorithmus ist, daß das Hologramm aus relativ tieffrequenten Strukturen besteht. Nur so ist gewährleistet, daß genügend Hologrammpixel zur Kodierung über die Error-Diffusion zur Verfügung stehen. Wenn dagegen ein sehr hochfrequentes Hologramm kodiert wird, bei dem zwischen benachbarten Pixeln starke Unterschiede bestehen, dann kann der Algorithmus nicht sinnvoll arbeiten. Dementsprechend erwartet man für solche Hologramme starkes Rauschen. Dieses Rauschen wird sich besonders am Rand der Rekonstruktion bemerkbar machen, da es — wie beschrieben — besonders die hochfrequenten Anteile im Hologramm beeinträchtigt.

Damit ist auch sofort klar, daß die Positionierung des Objektbereichs im Objektraum ein entscheidender Faktor für die Bildqualität der Rekonstruktion ist. Verschiebungen aus dem Zentrum führen zu hochfrequenten Anteilen und dementsprechend zu Störungen.

In mathematischer Formulierung lautet der ED-Algorithmus ([weis89]):

$$H'_{ij} = H_{ij} + \sum_{n=-1}^1 \sum_{m=-1}^1 A_{nm} E_{i-n, j-m} \quad (11.9)$$

$$E_{ij} = H'_{ij} - B_{ij} \quad (11.10)$$

$$B_{ij} = \begin{cases} 1 & \text{für } H'_{ij} > b \\ 0 & \text{für } H'_{ij} \leq b \end{cases} \quad (11.11)$$

Dabei ist b die Quantisierungsschwelle, H das zu kodierende Amplitudenhologramm und A_{nm} die Gewichtsmatrix für die Verteilung der Fehler. Das Ergebnis der Quantisierung ist B_{ij} . Entsprechend kann natürlich auch eine Quantisierung in mehr als zwei Stufen vorgenommen werden. Insbesondere dann ergibt sich für CGHs laut Casasent et al. ([casa91]) im Vergleich zu anderen Verfahren ein günstiges Verhalten.

Hauck und Bryngdahl ([hauc84]) haben eine eindimensionale Error-Diffusion untersucht. Dabei wird einfach für jede Spalte oder Zeile des Hologramms der 1D-ED-Algorithmus angewendet. Es ergaben sich dabei leichte Vorteile hinsichtlich Rauschanteilen, die — laut den Autoren — durch ein Übersprechen zwischen den einzelnen Spalten auftreten sollen.

Durch die Wahl der ED-Koeffizienten läßt sich kontrollieren, in welchen Bereichen der Rekonstruktion das Quantisierungsrauschen überwiegt ([fett95]). Diese Bereiche sind natürlich so zu wählen, daß sie außerhalb des eigentlichen Objektbereichs liegen. Die Rauschbereiche sind insbesondere von der Richtung, in die der ED-Algorithmus fortschreitet, abhängig.

Für die Binarisierung von Bildern wurde der eindimensionale ED-Algorithmus mit Erfolg dahingehend abgewandelt, daß der Fehler nicht auf einer Geraden in eine Richtung propagiert, sondern entlang einer fraktalen Hilbert-Kurve fortschreitet ([mene94]). Diese Kurve deckt das gesamte Bildgebiet ab und soll damit den Fehler besser verteilen und regelmäßige Fehleranhäufungen vermeiden. Bisher wurde diese Variation noch nicht auf computergenerierte Hologramme übertragen.

11.6.1 Komplexe Error-Diffusion

Binäre Amplitudenhologramme können wie in 5.2 ausgeführt natürlich auch als Phasenhologramme verwendet werden. Dem voran ist aber eine Kodierung der Phase (z.B. mittels Burch-Kodierung) gegangen. Für reine Phasenhologramme kann es günstig sein, diesen Schritt zu umgehen. Mit den iterativen Fourier-Transformations-Algorithmen existiert ja beispielsweise ein effektives Mittel, um solche Phasenhologramme zu berechnen. Wenn diese Phasenhologramme mittels des ED-Algorithmus quantisiert werden sollen, dann ist eine Abwandlung der bisherigen Vorgehensweise nötig.

Diese Abwandlung wird durch den Übergang zu komplexen Zahlen erreicht. Weissbach et al. ([weis89]) verwenden dementsprechend folgenden ED-Algorithmus:

$$H'_{ij} = H_{ij} + \sum_{n=-1}^1 \sum_{m=-1}^1 A_{nm} E_{i-n, j-m} \quad (11.12)$$

$$E_{ij} = H'_{ij} - \hat{Q}(h'_{ij}) \quad (11.13)$$

Dabei ist insbesondere der bisherige Hard-Clip-Binarisierer B durch einen Codierungsoperator \hat{Q} ersetzt worden, der die Amplitude auf 1 transformiert.

$$\hat{Q}(H'_{ij}) = \frac{H'_{ij}}{|H'_{ij}|} \quad (11.14)$$

(11.14) würde zu Phasenhologrammen führen, die nicht quantisiert sind. Will man quantisierte Phasenhologramme, dann ist vor der Anwendung von \hat{Q} noch eine Quantisierung bzw. Binarisierung der Phase von H' durchzuführen.

Phasenhologramm, die mit dieser komplexen ED quantisiert werden, haben zwar eine geringere Beugungseffizienz als IFTA-Hologramme, rekonstruieren aber die komplexe Amplitude, was für einige Anwendungen interessant sein kann. Desweiteren ist es natürlich wie bei allen ED Algorithmen durch Wahl der Diffusionsmatrix A_{nm} möglich, das Binarisierungsrauschen vom Objektbereich zu separieren. Für Anwendungen, bei denen nicht das gesamte mögliche Rekonstruktionsgebiet benötigt wird (Fensterung), kann das ein erheblicher Vorteil sein. Allerdings muß man auch darauf hinweisen, daß es im Gegenzug ein entscheidender Nachteil ist, wenn wirklich ein möglichst großes Rekonstruktionsgebiet verwendet werden soll.

Nichtzuletzt ist ein entscheidender Vorteil der Error-Diffusion, daß es ein nicht-iteratives und damit sehr schnelles Verfahren ist.

Insgesamt wurden im Rahmen dieser Arbeit nur einige Simulation mit dem ED-Algorithmus durchgeführt. Die Ergebnisse waren allerdings wenig befriedigend und zeigten kaum bessere Ergebnisse als Quantisierungen mit Hard-Clipping. Grundproblem ist dabei sicher, daß unsere Hologramme relativ hochfrequente Anteile enthalten (s.o.). Außerdem ist bei der Verwendung der ED immer das Problem gegeben, daß nur in einem Objektfenster eine rauschfreie Rekonstruktion erreicht werden kann.

Kapitel 12

Weitere Verfahren zur CGH–Berechnung

In diesem Kapitel werden einige weniger gebräuchliche oder für unsere Zwecke weniger in Frage kommenden Verfahren zur Berechnung von CGHs angesprochen. Diese Verfahren wurden im Rahmen dieser Arbeit nicht bzw. nur sehr kurz getestet. Trotzdem sollte man die entsprechenden Algorithmen nicht völlig aus dem Auge verlieren, da natürlich durchaus Anwendungen möglich sind, bei denen sie vorteilhaft eingesetzt werden können.

12.1 Genetische Algorithmen

Die *genetischen Algorithmen* (z.B. [davi91], [blue91], [pric94]) können mit einer gewissen Berechtigung zu den direkten Verfahren gezählt werden. Sie sind ein allgemeines Optimierungsverfahren, das auf beliebige Fehlerfunktionen, also z.B. auf den quadratischen Fehler unserer Rekonstruktion, die Beugungseffizienz bzw. Kombinationen aus beiden, angewendet werden können.

Grundidee ist es, aus einer *Population* von n Hologrammen Kombinationen zu bilden. Zusammen mit den ursprünglichen Hologrammen (*Eltern*) bilden diese Kombinationen eine neue Population¹. Alle Mitglieder dieser Population werden nun hinsichtlich der Fehlerfunktion beurteilt. Im Sinne einer *Selektion* überleben aus dieser Hologrammpopulation nur n Hologramme. Die Überlebenswahrscheinlichkeit hängt vom Fehler des jeweiligen Hologramms ab. Der gesamte Vorgang wird sehr oft wiederholt.

Analog zum natürlichen Vorbild ergibt sich durch die Selektion im Laufe vieler *Generationen* (bzw. Iterationen) eine gut angepasste Population (entsprechend optimierten Hologrammen).

Allgemein haben genetische Algorithmen ein großes zukünftiges Potential. Die Endergebnisse, die mit ihnen bei einer Hologrammoptimierung erzielt werden können, sind sehr gut und hinsichtlich des Rekonstruktionsfehlers besser als die mit IFTA erreichten Ergebnisse. Allerdings liegt der Rechenaufwand momentan noch um mehrere Größenordnungen (Faktor 2000 bei einigen Experimenten von M. Pahlke (ITO Stuttgart)) höher. Beim momentanen Stand sind die Verfahren mit Rechenzeiten im Bereich von Tagen bis Wochen für 16×16

¹einige Algorithmen bilden die neue Population nur aus den Kombinationen der Nachkommen

Hologramme auf herkömmlicher Hardware nur für sehr kleine Hologramme anwendbar.

Die entscheidende Frage wird dementsprechend sein, ob sich die Rechenzeiten deutlich reduzieren lassen. Da genetische Algorithmen sehr junge Optimierungsverfahren sind, darf man hier noch deutliche Verbesserungen erwarten. Folgende grundsätzlichen Möglichkeiten bieten sich für CGHs an:

- *schnelle Hardware*: Mit der im Abschnitt 15.3.1 beschriebenen Hardware läßt sich eine Beschleunigung um den Faktor 100 erwarten. Die Algorithmen sind gut parallelisierbar, was weitere Leistungssteigerungen ermöglicht.
- *Verbesserung der Konvergenz*: Die Anzahl nötiger Generationen hängt stark von der Größe der Population, der Art der Kombination der Hologramme zu neuen Hologrammen und der Berücksichtigung von Zufallselementen (*Mutation*) ab. Momentan fehlen hier noch theoretische Ergebnisse, die helfen, die vielen Parameter, die in den genetischen Algorithmus eingehen, zu optimieren.
- *Beschleunigung der Auswertung*: Der Hauptaufwand innerhalb einer Iteration ist durch die Bewertung der Hologramme gegeben. Dort mußte bisher für jedes Hologramm eine zweidimensionale FFT und eine Summation der Fehler durchgeführt werden. Analog zum Vorgehen bei den DBS Methoden sollte meiner Ansicht nach dieser Aufwand drastisch reduziert werden können, indem nur spezielle Kombinationen der Hologramme zugelassen werden, die eine schnellere Berechnung der Rekonstruktionen aus den Rekonstruktionen der Eltern-Populationen ermöglichen.

Momentan läßt sich nicht mit Bestimmtheit sagen, ob sich die genetischen Algorithmen durchsetzen können. Die bisherigen Ergebnisse sind ermutigend. Momentan ist an einen praktischen Einsatz für die angestrebten Anwendungen allerdings noch nicht zu denken.

12.2 Ping-Pong-Propagation und andere einfache 3D-Verfahren

Die Grundidee der *Ping-Pong-Approximation* ist einfach (siehe [dall80] 6.6.1): das dreidimensionale Objekt wird in diskrete Schichten zerlegt. Ausgehend von der tiefsten Schicht wird die Lichtamplitude über eine Fresnel-Propagation vor der darauf folgenden Schicht berechnet. Die Transmission der folgenden Schicht wird mit dieser Lichtamplitude multipliziert und zur nächsten Schicht propagiert.

Das Verfahren wird bis zur letzten Schicht und der darauf folgenden Propagation zur Hologrammebene fortgesetzt.

Vorteil dieses Verfahrens ist, daß ihm die Simulation der exakten physikalischen Vorgänge zugrunde liegt. Es werden so automatisch Überdeckungseffekte (*"hidden-lines"*) berücksichtigt. Nachteilhaft ist dagegen daß die Objektphasen in den verschiedenen Objektebenen nicht optimiert werden und so die Qualität der Rekonstruktion bei Verwendung von realen Hologrammedien zu wünschen übrig läßt.

Weiterhin ist zu beachten, daß bei Hologrammen mit sehr großen Pixeln (LCDs) Probleme mit kleinen Abständen zwischen den Objektebenen entstehen. Will man die Propagation zwischen eng benachbarten Ebenen berechnen, dann ist die Verwendung der Fresnel-

Beugung nicht mehr zulässig, da die Fresnel-Näherung nicht zutrifft. In diesem Fall muß das Kirchhoffsche Beugungsintegral gelöst werden, was einen erheblichen Rechenaufwand nach sich zieht.

Eine Vereinfachung ergibt sich, wenn man die Interaktion zwischen den verschiedenen Objektebenen vernachlässigt und die Summe der Beugungsbilder aller Ebenen in der Hologrammebene bildet. Das entspricht der Summation der Fourier-Transformationen der Einzelebenen, die dann jeweils noch mit einem entsprechenden quadratischen Phasenfaktor (Linse) multipliziert werden. Dieses Verfahren wurde bereits 1968 von Waters ([wate68]) für dreidimensionale CGHs vorgeschlagen.

Damit umgeht man die Probleme mit der Beugung zwischen eng benachbarter Ebenen, muß allerdings auf eine korrekte Behandlung des hidden-lines Problems verzichten. Da dies für die von uns angestrebten Anwendungen ohne Belang ist, wählten wir diesen Ansatz, der im übrigen von Dallas (siehe [dall80] 6.6.3) *Born-Approximation* (entgegen der üblichen Besetzung dieses Begriffs (siehe z.B. [sale91] 19.1)) genannt wird.

Die Ergebnisse mit diesen einfachen Methoden waren für unsere Zwecke zu schlecht.

12.3 Analytische Verfahren

Bisher wurden die Hologramme numerisch für beliebige Rekonstruktionsmuster berechnet. Die Berechnung erfolgte im wesentlichen über die schnelle Fourier-Transformation. Eine wesentliche Beschleunigung kann erreicht werden, wenn man das gewünschte Objekt aus einfacheren Teilen (Punkte, Linien) zusammensetzt. Für diese einfachen Teile werden analytisch Hologramme berechnet.

Mit dieser Vorgehensweise lassen sich sehr große Geschwindigkeiten zur Berechnung von Hologrammen erreichen. Sie eignet sich dementsprechend insbesondere für Echtzeitanwendungen (z.B. Displays) mit Hologrammedien mit großen Orts-Bandbreiten-Produkten.

Nachteil dieser Verfahren ist, daß die Objektphase nicht für das Hologrammedium optimiert werden kann. Dementsprechend ist die Qualität der Rekonstruktion gegenüber den bisher besprochenen Varianten stark vermindert. Eigene Experimente mit dem Verfahren von Frère et al. führten dementsprechend für die geringe Pixelanzahl unserer Hologramme zu sehr mangelhaften Rekonstruktionsergebnissen.

Frère et al. ([frer86], [lese87]) setzen dreidimensionale Objekte aus Linienelementen zusammen. Dazu nutzen sie, daß Zylinderwellen zu Linien gebündelt werden. Das Hologramm muß lediglich die entsprechenden Zylinderwellen kodieren und es ergeben sich die gewünschten Rekonstruktionen. Auf die passenden Zylinderwellen kommt man durch elementare Substitutionen.

Eine Abwandlung des Verfahrens zur Rekonstruktion beliebiger Kurven im dreidimensionalen wurde von denselben Autoren ebenfalls 1986 vorgestellt ([frer86b]). Die Grundidee ist dabei dieselbe.

12.4 Superpositionen

Um den hohen Berechnungsaufwand, der durch die Fourier-Transformation gegeben ist, zu vermeiden, kann man die Lichtverteilung in der Hologrammebene für einige Grundelemente einmalig berechnen und dann aus diesen Grundelementen das Gesamtobjekt zusammensetzen (siehe z.B. [siem68], [hask74], [schr84] Kap. 4.3.3.3., [nava91]). Die Verschiebung der Grundelemente (z.B. einzelne Punkte oder Linien) bewirkt lediglich eine Phasenmodulation in der Hologrammebene, die direkt analytisch angegeben werden kann.

Üblicherweise werden Punkte und Linien als Grundobjekte verwendet. Möglich sind allerdings auch komplizierte Objekte bis hin zu völlig abstrakten Gebilden (z.B. [wat195]).

Auch für dieses Verfahren kann die Objektphase nicht optimiert werden da bei der Herstellung der Basishologramme nicht das Gesamtobjekt bekannt ist. Ein weiterer Nachteil ist der unter Umständen sehr große Speicherbedarf für die Basishologramme.

12.5 Ray-Tracing

Ray-Tracing-Methoden (siehe z.B. [stei92]) berechnen die Lichtamplitude in der Hologrammebene für jeden Objektpunkt und summieren über alle Objektpunkte statt FFTs zu verwenden. Dieses Verfahren kann deutliche Geschwindigkeitsvorteile ergeben, wenn das Gesamtobjekt aus relativ wenigen Punkten besteht, die zudem in verschiedenen Tiefen liegen. Für komplexe Objekte ist das Verfahren allerdings wesentlich langsamer als der Ansatz über FFTs.

Außerdem besteht bei diesem Verfahren ebenfalls nicht die Möglichkeit die Objektphase passend zum Hologrammedium zu wählen. Entsprechend ist wieder mit einer schlechteren Qualität der Rekonstruktion zu rechnen. Will man dies vermeiden, so wäre der Ray-Tracing Ansatz mit einem zusätzlichen iterativen Verfahren zu verbinden.

12.6 Brechung der Inversions-Symmetrie

Im Standardaufbau zur Rekonstruktion eines Fourier-Hologramms ergibt sich bei Verwendung eines Amplitudenhologramms oder eines binären Phasenhologramms eine Inversions-symmetrie

$$I(f_x, f_y) = I(-f_x, -f_y) \quad (12.1)$$

Diese schränkt ohne Zweifel den nutzbaren Bereich auf die Hälfte ein und führt zu einer Verminderung der Beugungseffizienz (ausgenommen sind die Sonderfälle bei denen die Rekonstruktion genau eine solche Symmetrie zeigen soll). Das Grundproblem, das zur Ausbildung der Symmetrie führt ist, daß das binäre Filter eine rein reelle Transmission hat.

12.6.1 Zusätzliches festes Filter

Man geht von einem zeitlich variablen Hologramm (z.B. LCD) aus. Wenn man aber nun vor dieses Hologramme ein weiteres (zeitlich konstantes) Filter mit der komplexen Transmission $g = a(f_x, f_y)e^{i\theta(f_x, f_y)}$ setzt, läßt sich das Problem der Inversions-symmetrie umgehen. Im

einfachsten Fall wäre dieses Filter ein Phasenfilter mit den Phasen 0 und $\pi/2$. Damit ist anschaulich klar, daß wir nun effektiv in der Hologrammebene über die Werte 0, $\pi/2$, π und $3/2\pi$ verfügen, wenn unser eigentliches Hologramm den Wertebereich $0, \pi$ aufweist. Leider sind aber eben nicht alle Hologrammpixel beliebig zwischen den 4 Werten wählbar. Das führt zu einem Fehler gegenüber einem echten 4-stufigen Hologramm.

Gehen wir zunächst davon aus, daß die beiden Phasenwerte des festen Filters zufällig verteilt sind. Um nun das eigentliche Hologramm zu berechnen muß man die Verteilung des zusätzlichen Filters in die Berechnung einbeziehen. Unsere iterativen Fourier-Transformations-Algorithmen sind hierzu nur bedingt in der Lage. Bei der Variation der Phase in der Hologrammebene kann die Konvergenz des Algorithmus empfindlich gestört werden. Durch ein entsprechendes langsames Anpassen analog der Quantisierung im IFT-Algorithmus kann dieses Problem aber umgangen werden.

Die Algorithmen, die über eine direkte Suche (wichtigster Vertreter DBS siehe Kapitel 10.1) das Hologramm berechnen, können problemlos verwendet werden. Dabei wird einfach bei der Berechnung der Rekonstruktion das zusätzliche Filter einbezogen.

Statt die Verteilung der beiden Phasen des festen Filters zufällig zu wählen, können andere Verteilungen – je nach Anwendung – bessere Ergebnisse liefern. Sie bestimmen insbesondere wie der sich ergebende Fehler im Vergleich zu einem echten 4-stufigen Hologramm in der Rekonstruktionsebene verteilt. Beispiele hierzu finden sich bei [neil93]. Dort wird auch auf den Sonderfall der Verwendung einer Fresnel-Zonenlinse als festes Filter eingegangen.

Davis et al. ([davi88]) teilen ihr Hologramm in zwei zusammenhängende Bereiche und verwenden entsprechend bei der Rekonstruktion eine zusätzliche Phasenschiebung in einem der beiden Bereiche (z.B. über dünne Glasplatte).

12.7 Nutzung von Symmetrien

Was macht die Berechnung computergenerierter Hologramme langsam? Alle Algorithmen müssen zweidimensionale Fourier-Transformationen berechnen. Die Rechenzeit kann drastisch verkürzt werden, wenn es gelingt, über Symmetrien die zweidimensionale Fourier-Transformation durch eine eindimensionale Fourier-Transformation zu ersetzen.

Im allgemeinen ist das nicht möglich. Wenn die Objektdaten aber Symmetrien aufweisen, dann kann wegen der Separierbarkeit der Fourier-Transformation eine Vereinfachung möglich sein.

In diesem Abschnitt werden drei mögliche Vereinfachungen aufgrund von Symmetrien angesprochen. Diese Vereinfachungen werden starken Symmetrien des Objekts nutzen und dementsprechend auch zu starken Vereinfachungen der Berechnung führen. Weniger offensichtliche Symmetrien lassen sich ebenso nutzen. Allerdings ist dann mit einem erhöhten algebraischen Aufwand und einer geringeren Beschleunigung zu rechnen.

Ein allgemeiner Nachteil bei der Verwendung von Symmetrien sollte nicht übersehen werden. Bei der Optimierung von CGHs werden freie Parameter genutzt (z.B. Objektphase) um das Hologramm zu optimieren. Die Symmetrieforderungen, die zur Beschleunigung genutzt werden, gelten auch für diese freien Parameter und reduzieren so die Anzahl der Freiheitsgrade, die zur Verfügung steht, ganz erheblich. Dementsprechend ist jeweils genau zu untersuchen, inwieweit diese Reduktion der Freiheitsgrade zu einer Verminderung der

Optimierungsqualität führt.

Bisher wurde die Beschleunigung aufgrund von Symmetrien nicht in der Praxis benutzt da es das Ziel war, möglichst allgemeine Verfahren zu entwickeln. Insbesondere für einfache Streifenprojektionen sollten sich die Verfahren aber anwenden lassen. Allerdings darf nicht zuviel Symmetrieeinformation verwendet werden da diese Symmetrie zwar die Algorithmen beschleunigt, aber — wie beschrieben — die Optimierungsfreiheitsgrade einschränkt.

12.7.1 Rotationssymmetrische Objekte

Rotationssymmetrische Objekte $f(r)$ enthalten effektiv lediglich eindimensionale Informationen. Mit $r = \sqrt{u^2 + v^2}$, dem Radius in der Bildebene $\rho = \sqrt{x^2 + y^2}$ und $\omega := 2\pi\rho$ läßt sich die Fourier-Transformation nach [schr84] (2.36) umschreiben zu

$$\mathcal{F}[f(u, v)] = \frac{1}{2\pi} \mathcal{HT}[f(r)] = \frac{1}{2\pi} \int_0^\infty r f(r) J_0(\omega r) dr \quad (12.2)$$

\mathcal{HT} ist dabei die *Hankel-Transformation* und durch die rechte Seite von (12.2) definiert. J_0 ist die Bessel-Funktion nullter Ordnung.

Durch einige mathematische Umformungen (siehe [sieg77]) läßt sich die Hankel-Transformation in eine eindimensionale Faltung umschreiben, die numerisch bekanntlich über drei FFTs berechnet werden kann. Das entsprechende Verfahren wurde von Siegman entwickelt ([sieg77]) und nennt sich *Fast-Hankel-Transform*. Dabei ist allerdings zu beachten, daß die Blockgröße verdoppelt werden muß. Unterbleibt dies, so ergeben sich falsche Werte (das Problem ist unter dem Namen *DFT-Alias* insbesondere in der elektrischen Signalanalyse bekannt (z.B. [bach92])). Alternativ kann die Rotationssymmetrie auch einfach über eine eindimensionale Summation über Bessel-Funktionen ausgenutzt werden. In diesem Fall ergibt sich bei N Stützstellen ein Aufwand proportional zu N^2 (für N Positionen jeweils eine Summe aus N Termen berechnen). Im Falle der Fast-Hankel-Transform ist der Aufwand direkt durch die FFTs gegeben und ist daher proportional $3 \cdot 2N \ln 2N$. Nur für sehr große N wird man in der Praxis gegenüber der einfachen Summation Geschwindigkeitsvorteile verbuchen können². Weiterhin ist zu beachten, daß durch die Substitutionen innerhalb des Algorithmus die berechneten Werte nicht auf einem äquidistanten Raster liegen und so praktische Probleme machen können.

Trotzdem wurde die Fast-Hankel-Transform von Khonina et al. zur Berechnung rotationssymmetrischer Hologramme eingesetzt ([khon91]). Die beschleunigte Berechnung ist aber hauptsächlich durch die Reduktion der zwei Dimensionen auf eine Dimension gegeben (s.o.). Eine einfache Summation der Bessel-Basis-Funktionen sollte eine Beschleunigung in der gleichen Größenordnung bewirken und wesentlich weniger praktische Probleme bereiten.

²Man beachte in diesem Zusammenhang, daß der Algorithmus zur Fast-Hankel-Transform wesentlich komplizierter ist und damit einen Teil der potentiellen Geschwindigkeitsvorteile zunichte macht.

12.7.2 Separierung und eindimensionale Objekte (Streifen)

Die Fourier-Transformation einer separierbaren Funktion $h(x, y) = f(x)g(y)$ ist ebenfalls separierbar. D.h.

$$\mathcal{F}[h(x, y)] = \mathcal{F}[f(x)g(y)] = \mathcal{F}[f(x)] \mathcal{F}[g(y)] \quad (12.3)$$

Wenn unser Objekt dementsprechend separierbar ist, läßt sich der Aufwand zur CGH-Berechnung gewaltig reduzieren. Statt einer zweidimensionalen FFT müssen lediglich zwei eindimensionale FFTs ausgeführt werden.

Für einfache, nicht-iterative CGH-Berechnungen ergeben sich keine Probleme. Iterative Verfahren, die die Objektphase optimieren ergeben allerdings dahingehend Probleme, daß die Objektphase entsprechend auch separierbar sein muß. Diese Forderung beinhaltet die bereits angesprochene deutliche Einschränkung der Freiheitsgrade, die uns mit der Objektphase gegeben sind. Abgesehen von den sich daraus ergebenden Qualitätseinbußen, muß bei jedem Iterationsschritt künstlich die Separierbarkeit der zu transformierenden Funktion sichergestellt werden. Unter Umständen kann dadurch sogar die Konvergenz des Algorithmus zerstört werden (insbesondere bei den IFT-Algorithmen ist mit diesem Problem zu rechnen).

In der Regel werden die im SYMAS-Projekt zu projizierenden Muster nicht direkt separierbar sein. Allerdings sind sie meist binär. Diese Information kann genutzt werden. Dazu denkt man sich das (zweidimensionale) $N \times M$ Objekt aus N Zeilen $o_i(x)$ aufgebaut. Entsprechend kann man für das Objekt $f(x, y)$ schreiben:

$$f(x, y) = \sum_{n=1}^N o_n(x)z_n(y) \quad (12.4)$$

Dies bringt noch keine Beschleunigung, weil wegen der Summe entsprechend N eindimensionale FFTs durchgeführt werden müssen. Interessant wird dieser Ansatz wenn — je nach Objekt — entsprechende Vereinfachungen dahingehend durchgeführt werden, daß Glieder der Summe eingespart werden können. Dies ist wegen des binären Charakters des Objekts ($1 \cdot 1 = 1$) einfacher als für koplizierte Objekte³. Im einfachsten Fall fallen z.B. sofort alle Zeilen, in denen kein Objektpunkt gesetzt ist, heraus.

³Diese Vereinfachung wird allerdings wieder durch komplizierte Objektphasen zunichte gemacht!

Kapitel 13

Phasenschiebung in doppelbrechenden Medien

Bisher wurde von einer skalaren Lichttheorie ausgegangen. Sollen Polarisierungseffekte, wie sie bei LCDs üblicherweise auftreten, berücksichtigt werden, muß zu einer vektoriellen Beschreibung (mit 2 Komponenten) des Lichts übergegangen werden¹.

Unser vorrangiges Ziel ist es, LCDs als Phasenmodulatoren einzusetzen. Dazu will ich zunächst den völlig unscheinbar erscheinenden Begriff “Phasenschiebung” aus unterschiedlichen Richtungen näher beleuchten. Dies ist für ein tieferes Verständnis der phasenschiebenden Eigenschaften von LCDs interessant, hat aber außerdem praktische Bedeutung. Ich werde auf dieser Diskussion aufbauend eine – meines Wissens – neue Möglichkeit der Vergrößerung des maximalen Phasenschiebebereichs durch den zusätzlichen Einbau polarisierender Elemente in den Strahlengang, vorschlagen.

Insbesondere die geometrische Phase (“Pancharatnam-Phase”) wird selten in Überlegungen einbezogen (wohl weil man meist intuitiv an linear polarisiertes Licht denkt und die meisten CGH-Medien auch keine anisotropen Eigenschaften aufweisen).

13.1 Die Poincaré-Kugel

Die Poincaré-Kugel ist ein seit langem verwendetes Hilfsmittel zur Beschreibung von Polarisationszuständen (siehe z.B. [born80]). Jeder Punkt auf der Oberfläche dieser Einheits-Kugel gibt eindeutig einen Polarisationszustand an. Ausführliche Einführungen und die Herleitung der Formeln zur Umrechnung der vektoriellen Darstellung der Polarisation auf Stokes-Parameter und damit auf die Poincaré-Kugel findet man in den Standardlehrbüchern der Optik (siehe z.B. [born80] 1.4.). Ich werde hier lediglich die für uns interessanten Ergebnisse angeben.

Für eine elliptisch polarisierte elektromagnetische Welle in z-Richtung, die wir mit $\vec{E} = (a_1 \cos(\omega t), a_2 \cos(\omega t + \delta), 0)$ beschreiben ergeben sich die *Stokes-Parameter* s_i zu

$$s_0 = I = a_1^2 + a_2^2 \tag{13.1}$$

¹Eine voll vektorielle Beschreibung über die Maxwell'schen Gleichungen ist bei der Beugung sichtbaren Lichts an LCDs wegen der Größe der LCD-Pixel Δx mit $\Delta x > 10\lambda$ nicht nötig

$$s_1 = a_1^2 - a_2^2 \quad (13.2)$$

$$s_2 = 2a_1a_2 \cos \delta \quad (13.3)$$

$$s_3 = 2a_1a_2 \sin \delta \quad (13.4)$$

Stokes-Parameter haben gegenüber der kartesischen Beschreibung einige Vorteile. Sie sind direkt einer Messung zugänglich (siehe z.B. [robs74] 1.9., [spie88] S. 3 folgend), lassen sich im Rahmen der Quantenmechanik reinterpreten und ermöglichen teilweise eine anschaulichere Beschreibung.

Eine zweidimensionale Darstellung der Elliptizität ist in Abbildung 13.1 verdeutlicht.

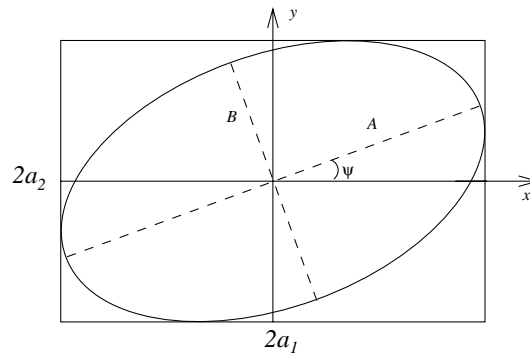


Abbildung 13.1: Polarisationsellipse

Für konstante Intensitäten I , die wir als auf 1 normiert voraussetzen, können die Stokes-Parameter auf der sogenannten *Poincaré-Kugel* mit Radius 1 veranschaulicht werden:

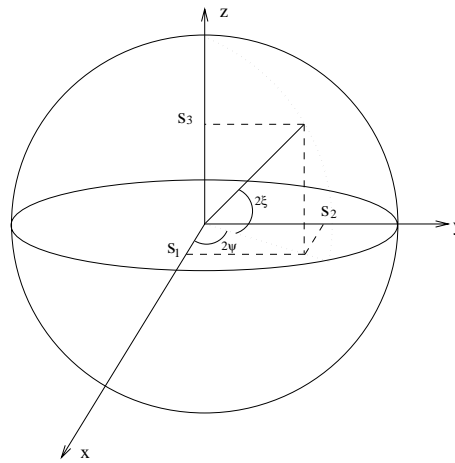


Abbildung 13.2: Poincaré-Kugel

Dabei gilt $\xi = 1/\tan(\pm B/A)$ und

$$s_1 = s_0 \cos 2\xi \cos 2\psi \quad (13.5)$$

$$s_2 = s_0 \cos 2\xi \sin 2\psi \quad (13.6)$$

$$s_3 = s_0 \sin 2\xi \quad (13.7)$$

Durch Einsetzen verifiziert man leicht die folgenden Beziehungen:

- gegenüberliegende Punkte auf der Poincaré-Kugel entsprechen zueinander senkrechten Polarisationen
- die Punkte auf dem Äquator entsprechen linearen Polarisationen
- der Nordpol (Südpol) entspricht rechts-(links-)zirkulärer Polarisation

Das Auftreten des Faktors 2 in den Argumenten der trigonometrischen Funktionen hat seinen tieferen Sinn in der Nichtunterscheidbarkeit von entgegengesetzten (z.B. E_x und $-E_x$) Polarisationen und wird gern als Ausgangspunkt für Analogien zu Fermionen und der Quantenmechanik im allgemeinen (siehe z.B. [berr87], [robs74]) verwendet.

13.2 Jones-Darstellung

Die obige Beschreibung der Polarisation anhand der Poincaré-Kugel hat anschauliche Vorteile, eignet sich jedoch schlecht für numerische Rechnungen. Stattdessen benutzt man dort meist die *Jones-Darstellung*. Dabei werden die x- und y-Komponenten des elektrischen Feldes abzüglich des schnell variablen Zeitfaktors $e^{i\omega t}$ als zweidimensionaler Vektor (“Jones-Vektor”) \vec{J} dargestellt (wir gehen im folgenden immer von in z-Richtung propagierenden Wellen aus):

$$\vec{J} = \begin{pmatrix} |E_x| e^{i\varphi_x} \\ |E_y| e^{i\varphi_y} \end{pmatrix} \quad (13.8)$$

Tritt die elektromagnetische Welle durch ein lineares, zeitinvariantes Element (z.B. einen linearen Polarisator), dann kann die Welle hinter dem entsprechenden Element durch Matrixmultiplikation des Jones-Vektors mit einer das Element beschreibenden Matrix berechnet werden. Eine solche Darstellung mit Matrizen und Vektoren ist oft bequem und übersichtlich und bietet über die Berechnung von Eigenvektoren teilweise elegante Lösungsmöglichkeiten².

Beispielhafte Rechnungen in der Jones-Darstellung und die Angabe der wichtigsten Jones-Matrizen findet man in [fowl75].

Eine weitere Beschreibung mittels Matrizen und Vektoren ist die *Mueller-Darstellung*. Ein *Muellervektor* besteht aus den vier Stokes-Parametern s_i . Ganz entsprechend zur Jones-Darstellung werden optische Elemente durch Matrizen beschrieben. Während bei der Jones-Darstellung die Addition zweier Jones-Vektoren der *kohärenten* Überlagerung zweier Wellen entspricht, ergibt die Addition zweier Muellervektoren eine *inkohärente* Überlagerung. Da wir uns in dieser Arbeit für kohärente Überlagerungen interessieren, werde ich die Mueller-Beschreibung nicht weiter verwenden. Es sei aber darauf hingewiesen, daß sie gegenüber der – für kohärentes Licht einfacheren – Jones-Darstellung einige Vorteile aufweist³. Näheres zur Mueller-Darstellung und eine Umrechnung von Jones-Matrizen in Mueller-Matrizen findet man bei [robs74] 1.8.1.

²Die Beschreibung ist in direkter Analogie zur in der Elektrotechnik gern verwendeten Vierpoltheorie zu sehen.

³1.) Es werden nur Observablen benutzt. 2.) Sie ist allgemeiner als die Jones-Darstellung (beispielsweise läßt sich in der Jones-Darstellung kein vollständiger Depolarisator darstellen (allerdings ist ein solcher auch in der Natur nicht bekannt)) 3.) sie kann auf andere Theorien (z.B. Elementarteilchen) übertragen werden.

13.3 Pancharatnams Phase

Die Phasenschiebung zwischen zwei Lichtfeldern ist auf den ersten Blick trivial. Beispielsweise kann man sofort die Phasenschiebung zwischen zwei in x-Richtung polarisierten Wellen angeben:

Gilt z.B. $E_{1,x} = \cos(\omega t)$ und $E_{2,x} = \cos(\omega t + \phi)$, so ist die Phasenschiebung ϕ . Ganz entsprechend sind auch Phasenschiebungen zwischen anderen Polarisationszuständen jeweils untereinander (also z.B. lineare Polarisation in y-Richtung mit linearer Polarisation in y-Richtung) trivial.

Wie sieht es aber nun mit der Phasenschiebung zwischen einer linear polarisierten Welle und einer beliebig anders polarisierten (elliptische Polarisation) Welle aus? Es muß eine Definition für die Phasenschiebung gefunden werden, die unseren üblichen Erwartungen für alle einfachen Sonderfälle entspricht und in sich konsistent ist.

Die natürlichste Definition geht auf Pancharatnam (siehe z.B. [berr87]) zurück und orientiert sich an der experimentellen Anwendung von Phasenschiebungen. Optische Phasenschiebungen sind vor allem für verschiedene Arten von Interferenzauswertungen wichtig.

Man definiert nun die Phasenschiebung als *die Abweichung von einer Referenzkonfiguration, bei der 2 Wellen in Phase sind. Die zwei Wellen werden dann als in Phase betrachtet, wenn ihre Interferenz eine maximale Intensität ergibt.*

Wir bilden also die Interferenz zwischen zwei *normierten* Wellen A und B mit $|A| = |B| := 1$:

$$I = (A + B)(A + B)^* = |A|^2 + |B|^2 + 2|A||B| \cos(\arg(A - B)) \quad (13.9)$$

$$= 2 + 2 \cos(\arg(A - B)) \quad (13.10)$$

Damit ist die Phase definiert als die Phase des Skalarprodukts zwischen den beiden Wellen.

Zwei Wellen sind damit genau dann *in Phase*, wenn das Skalarprodukt zwischen beiden reell und positiv ist. Dieses “in Phase sein”, also die oben angesprochene Referenzkonfiguration, wird auch als *“Pancharatnam connection”* bezeichnet.

Bei vielen Experimenten wird vor der eigentlichen Interferenz sichergestellt, daß alle Polarisierungen zum Beispiel in x-Richtung liegen. Für diese Art von Experimenten wäre die gerade gegebene Definition der Phasenschiebung unnötig kompliziert (obwohl sie natürlich auch diesen Spezialfall korrekt beschreibt). Bei Experimenten mit LCDs muß aber bei Nichtverwendung von Polarisatoren mit der exakten Phasendefinition gearbeitet werden.

Pancharatnam hat weiterhin gezeigt (siehe für einen einfachen Beweis z.B. [arav92]), daß die Pancharatnam connection nicht-transitiv ist. Das heißt, wenn wir drei Lichtfelder A , B und C haben und A und B sowie B und C in Phase sind, dann sind A und C trotzdem im allgemeinen nicht in Phase. Nur für Sonderfälle (z.B. A, B und C sind linear polarisiert) sind A und C in Phase.

Die Phasenschiebung θ zwischen A und C ergibt sich nach Pancharatnam zu (siehe [arav92])

$$\theta = \arg(AC) = \frac{1}{2} \Omega_{ABC} \quad (13.11)$$

Dabei ist Ω_{ABC} der Raumwinkel der Geodäten (refraktive (“dynamische”) Phasenschiebung

ist entlang dieser Geodäten⁴ Null) zwischen den Polarisationszuständen auf der Poincaré-Kugel bzw. die von den Geodäten auf der Poincaré-Kugel eingeschlossene Fläche.

Diese “geometrische” Phasenschiebung⁵ ist völlig unabhängig von einer eventuell zusätzlich vorhandenen “dynamischen” Phasenschiebung⁶ ! Die Gesamtphasenschiebung ist die Summe aus geometrischer und dynamischer Phasenschiebung (siehe z.B. [mart90]). Man beachte weiterhin, daß die Pancharatnam Phase im Gegensatz zur dynamischen Phase nur von der eingeschlossenen Fläche auf der Poincaré-Kugel, nicht aber von der Wellenlänge abhängig ist.

Die geometrische Phase als Raumwinkel auf der Poincaré-Kugel ist nicht auf 2π -beschränkt (kein modulo 2π).

Verschiedene Autoren ([mart90], [chyb88]) haben die Pancharatnam Phase zur Kontrolle der Phasenschiebung in der Praxis untersucht.

Schmitzer et al. (siehe [schm93b]) haben die geometrische Phase für zwei elliptisch polarisierte Wellen bei Interferenz nach einem Polarisator in Abhängigkeit von der Polarisatorstellung untersucht und erhielten ein stark nichtlineares Verhalten für die Phasenschiebung. Sie kommen zu dem Schluß, daß sich die Pancharatnam Phase deshalb für schnelle Schaltanwendungen eignen könnte.

Es sei darauf hingewiesen, daß die differentialgeometrische Beschreibung der Pancharatnam Connection analog zur in der allgemeinen Relativitätstheorie so wichtigen Parallelverschiebung in der gekrümmten (nicht notwendigerweise kugelförmigen) vierdimensionalen Raumzeit zu sehen ist. Auch dort wird durch den sogenannten “Paralleltransport” die Eigenschaft “Parallelität” (hier in Analogie zur Eigenschaft “in Phase”) für nicht-benachbarte Punkte eingeführt⁷.

Pancharatnams Ergebnis 13.11 hat außerdem erstaunliche direkte Analogien in der klassischen Mechanik (siehe [berr90]) und der modernen Quantenmechanik (“Berry-Phase” in adiabatischen Quanten-Systemen, “Aharonov-Bohm-Effekt” (siehe [berr90], [berr87])), die unter Umständen den tieferen Ursprung des Polarisationsphänomens für Photonen⁸, dessen quantenmechanische Deutung nach wie vor heftig umstritten ist (siehe z.B. [tiwa92], [tiwa93], [gray94], [devi94], [strn86]), näher beleuchten könnten.

Auf dieses aus theoretischer Sicht sehr interessante Gebiet kann hier aber nicht weiter eingegangen werden, da es für unsere praktische Anwendung der Polarisation ohne Belang bleibt.

13.4 Pancharatnams Phase und Phasenmessungen am LCD

Wenn wir eine interferometrische Phasenmessung (z.B. mit einem Mach-Zehnder Interferometer) durchführen, werden wir die gesamte Phasendifferenz $\delta = \delta_{geom} + \delta_{dyna}$ messen.

Interessiert man sich für die Gesamtphasenschiebung eines LCDs, muß bedacht werden,

⁴das ist in diesem Fall die Definition von “Geodäte”

⁵= “topologische” Phasenschiebung = “Pancharatnam Phase” = “Berry Phase”

⁶= “refraktive Phasenschiebung”

⁷entsprechend fallen freie Körper auf Geodäten und die Phasenschiebung bei Transport auf Geodäten ist Null

⁸üblicherweise sieht man die beiden Polarisationszustände des Photons als eine Manifestation zweier Spinzustände des Photons (Boson !) an. Die Wahrscheinlichkeit für den Aufenthalt des Photons im dritten Spinzustand ($m_s = 0$) ist wegen der Transversalität der elektromagnetischen Wellen Null.

daß diese nicht in einfacher Form von der Wellenlänge des verwendeten Lichts abhängt: Während die dynamische Phasenschiebung linear von der Wellenlänge abhängt, ist die Pancharatnam Phase unabhängig von der Wellenlänge. Dies muß bei einer Übertragung der für eine Wellenlänge gewonnenen Meßergebnisse auf eine andere Wellenlänge unbedingt beachtet werden!

Die in [reyn89] 22.3.4 vorgeschlagene Meßmethode ist nur innerhalb einer skalaren Theorie (bzw. gleiche Polarisationen der Teilstrahlen) korrekt und kann daher bei der Messung der Phasenschiebung von LCDs nicht verwendet werden (siehe [wagh95] für eine Diskussion der korrekten Messung der Pancharatnam Phase).

13.5 Vergrößerung des Phasenschieberegions bei LCDs

Für computergenerierte Phasen-Hologramme (und andere Anwendungen) werden LCDs als phasenschiebende Elemente eingesetzt. Dabei ist man bestrebt, möglichst einen Phasenschieberegions von mindestens 2π zu erreichen. Je kleiner der Phasenschieberegions ist, desto größer werden die Fehler in der Rekonstruktion des Hologramms sein.

Preiswerte LCDs sind nur in Massenproduktion herzustellen. Dementsprechend kommen als preiswerte LCDs nur LCDs, die für Display-Anwendungen konstruiert wurden, in Frage. Der Phasenschieberegions ist dabei für den herkömmlichen Display-Betrieb uninteressant. Es nimmt also nicht wunder, daß solche LCD selten den Bereich 0 bis 2π abdecken.

Wie in beiden vorangegangenen Abschnitten dargelegt wurde, ergibt sich die Gesamtphasenschiebung eines LCDs aus einem rein refraktiven Anteil, der durch die Drehung der LC-Moleküle im elektrischen Feld in Verbindung mit den anisotropen Brechungsverhalten zustande kommt und einem geometrischen Anteil, der durch die Änderung des Polarisationszustandes in Verbindung mit externen Polarisatoren hervorgerufen wird.

Der refraktive Anteil ist konstant und kann von uns nicht beeinflußt werden. Den geometrischen Anteil kann man dagegen durch die Wahl der externen, polarisierenden Elemente stark beeinflussen.

In [schw96] wurde die Eingangspolarisation parallel zur Orientierung der LC-Moleküle an der LCD-Oberfläche⁹ gewählt. Diese Vorgehensweise ist üblich und garantiert eine maximale Lichtintensität nach dem LCD. Der lineare Polarisator hinter dem LCD ("Analytator") wird dann entsprechend so gedreht, daß ein maximaler Phasenschieberegions bei ausreichend konstanter Intensität erreicht wird.

Man kann aber vermuten, daß die geometrische Phase unter Umständen vergrößert werden kann, wenn die Richtung der Eingangspolarisation variiert wird. In der Tat ist dies der Fall. Auf der Poincaré-Kugel wird die eingeschlossene Fläche auf der Kugeloberfläche künstlich vergrößert.

Um die optimale Eingangspolarisation zu finden wurde eine Simulation der entsprechenden Vorgänge durchgeführt. Als Jones-Matrizen dienten dabei theoretisch berechnete Jones-Matrizen für TN-LCDs (siehe [sale90] (das Ergebnis von Saleh wurde mit eigenen Simulationen verglichen und bestätigt)).

Das Simulationsprogramm berechnet für beliebige Stellungen des Eingangs- und des Aus-

⁹Vorzugsrichtung der LC-Moleküle wird meist durch Riefen in der Glasoberfläche festgelegt

gangspolarisators und Spannungen U am LC von 1.0 V bis 2.0 V die Ausgangs-Intensitäten und die Phasenschiebungen bezüglich eines Referenzstrahls (mit $U = 1.0$ V).

In Abbildung 13.3 und 13.4 sind Phasenschiebung und Intensität der üblichen Eingangspolarisator-LCD-Analysator Kombination für die 0/90 Stellung (Eingangspolarisator in 0 Grad Stellung, Analysator in 90 Grad Stellung) dargestellt. Man erkennt eine starke Variation der Intensität in Verbindung mit einem relativ geringen Phasenschiebebereich von ca. 0.6π .

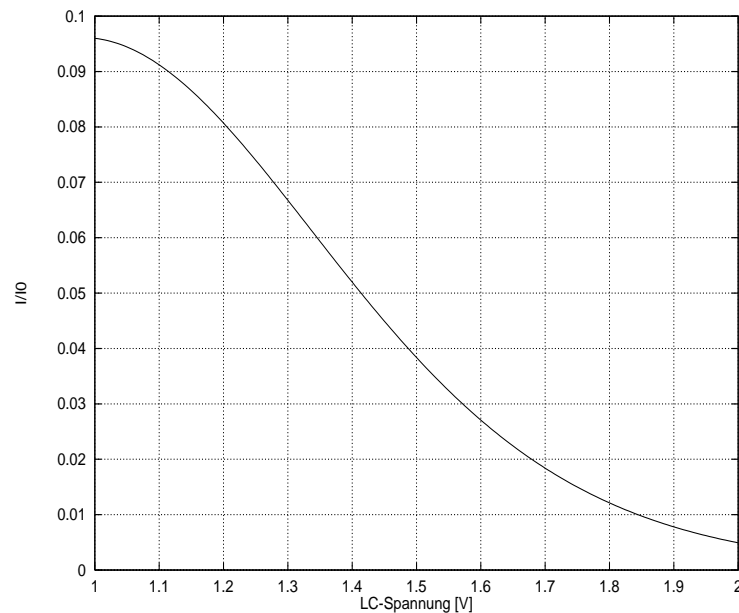


Abbildung 13.3: 0/90-Stellung: Intensität in Abhängigkeit von LC-Spannung

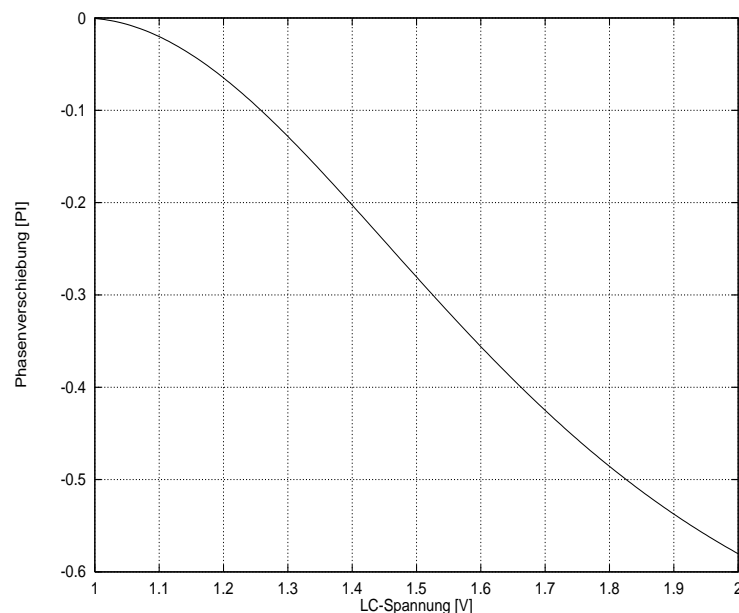


Abbildung 13.4: 0/90-Stellung: Phasenschiebung in Abhängigkeit von LC-Spannung

In einem weiteren Schritt wird die Simulation für alle möglichen Stellungen (1 Grad Schritte) der Polarisatoren durchgeführt und der jeweils maximal erreichbare Phasenschiebebereich

gespeichert. Dabei werden nur solche Stellungen berücksichtigt, bei denen die Ausgangsintensität einen minimalen Wert nicht unterschreitet und bei dem die Schwankung der Ausgangsintensität (in Abhängigkeit von der angelegten LC-Spannung) nicht zu groß wird. Die entsprechenden Parameter sind – je nach Anwendung – vom Benutzer wählbar.

Das Ergebnis einer solchen Rechnung (max. Kontrastvariation $K = |I_1 - I_2| / |I_1 + I_2| < 0.1$ und $I_{min} > 0.1I_0$) ist in Abbildung 13.5 dargestellt.

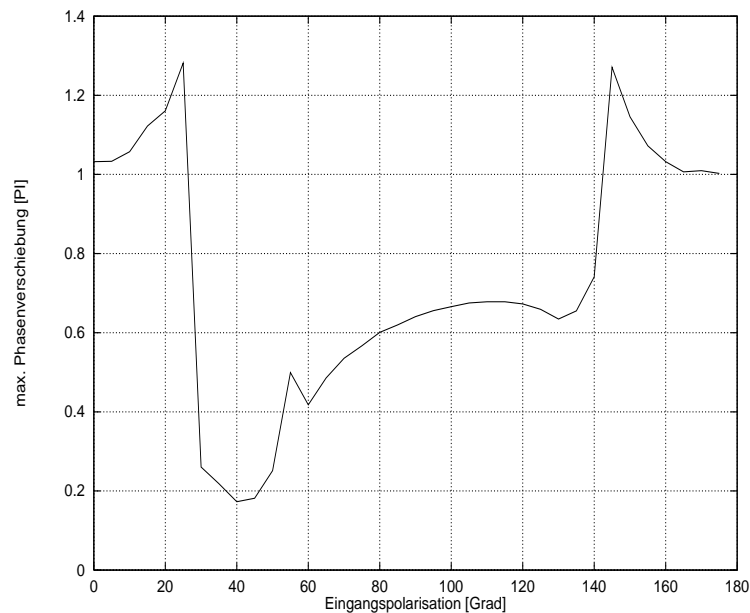


Abbildung 13.5: maximal erreichbare Phasenschiebung in Abhängigkeit von Eingangspolarisation

Man erkennt deutlich, daß eine optimale Phasenschiebung für das LCD bei einer Eingangspolarisation von 25 bzw. 145 Grad (der Analysator muß entsprechend auf 124 bzw. 64 Grad gestellt werden) und nicht bei 0 Grad erreicht wird.

Im Beispiel kann damit also der maximale Phasenschiebebereich von 1.03π (0/110-Stellung) auf 1.28π (25/124-Stellung) – auf Kosten der maximalen Intensität – gesteigert werden.

Es sei darauf hingewiesen, daß die Kurve in Abbildung 13.5 stark mit den je nach Anwendung zu wählenden Toleranzen für die Intensitäten variiert.

In Abbildung 13.6 13.7 sind beispielhaft die Intensitäts- und Phasenverläufe für die “optimale” 25/125-Stellung gezeigt.

Der Gewinn von rund 25 % an Phasenschiebung wäre an sich schon interessant, läßt sich aber verbessern, wenn zusätzlich noch weitere polarisierende Elemente in den Strahlengang eingebracht werden.

Beispielhaft wurde in der Simulation eine zusätzliche $\lambda/4$ -Platte zwischen LCD und Analysator berücksichtigt. Je nach Stellung der $\lambda/4$ -Platte wurde wieder – wie oben beschrieben – der maximale Phasenschiebebereich (beliebige Stellung des Eingangs- und Ausgangspolarisators) ermittelt. Das Ergebnis ist in Abbildung 13.8 dargestellt.

Man erkennt die starke Zunahme der geometrische Phase resultierend in einer Gesamtphasenschiebung von bis zu 3.6π .

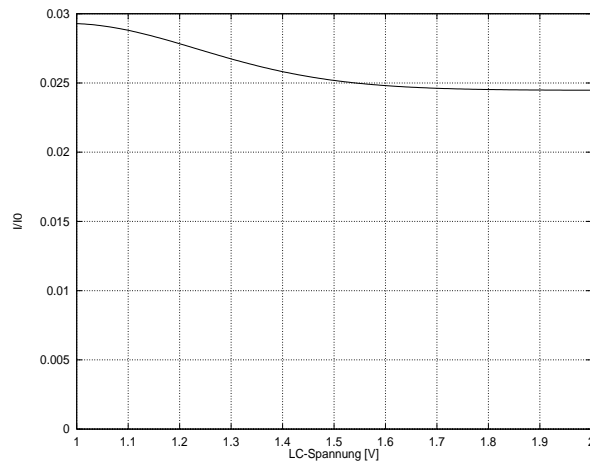


Abbildung 13.6: 0/90-Stellung: Intensität in Abhängigkeit von LC-Spannung

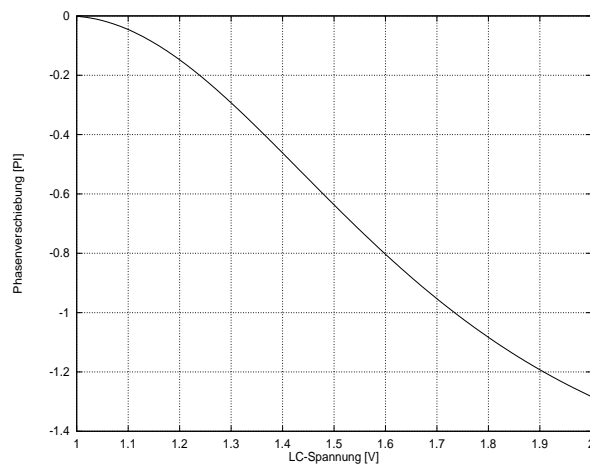


Abbildung 13.7: 0/90-Stellung: Phasenschiebung in Abhängigkeit von LC-Spannung

Der Einsatz von $\lambda/2$ -Platten bringt — wie erwartet — keine Verbesserungen.

Eine $\lambda/4$ Platte ist ein sehr einfaches Element. Unter Umständen können aber kompliziertere Elemente eine weitere Verbesserung bewirken. Eine allgemeine Beschreibung geht von zwei beliebigen linearen Elementen, repräsentiert durch die Jones-Matrizen J_1 und J_2 , vor bzw. hinter dem LCD aus. Der Zusammenhang zwischen dem Eingangslichtfeld \vec{v}_i und dem Ausgangslichtfeld \vec{v}_o wird dann gegeben durch

$$\vec{v}_o = L(\alpha) J_1 LCD(U) J_2 L(\beta) \vec{v}_i \quad (13.12)$$

Die linearen Polarisatoren (Eingangspolarisation und Analysatorstellung) sind dabei durch L und die spannungsabhängige Jones-Matrix des LCDs durch LCD repräsentiert.

Für beliebige α und β sind nun die $16+16=32$ freien Parameter der beiden Jones-Matrizen so zu bestimmen, daß ein "optimales" Systemverhalten erreicht wird (als "optimal" für unsere Anwendung gilt ein Kompromiß zwischen großem Phasenschiebebereich und ausreichend hoher und konstanter Amplitude).

Sind die entsprechenden Jones-Matrizen bestimmt, dann kann untersucht werden, wie sie

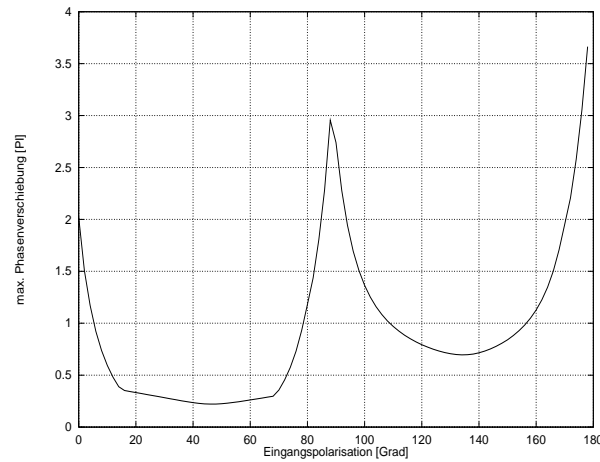


Abbildung 13.8: maximal erreichbare Phasenschiebung in Abhängigkeit von Eingangspolarisation

(annähernd) durch reale optische Elemente realisiert werden können.

Die beschriebene Optimierungsaufgabe, mit den vom Benutzer wählbaren Nebenbedingungen, ist kaum analytisch lösbar. Ein numerisches Vorgehen scheint daher besser geeignet zu sein.

Im Rahmen meiner Diplomarbeit kann ich aus Zeitgründen diese interessante Idee nicht weiter verfolgen. Die in der Simulation erzielten theoretischen Ergebnisse legen es meiner Ansicht nach aber nahe, in näherer Zukunft die beschriebene Optimierung durchzuführen und die Ergebnisse experimentell zu verifizieren. Unter Umständen könnten dadurch preiswerte LCDs mit geringem Phasenschieberegion entsprechend teurere SLMs – nicht nur im Bereich computergenerierter Hologramme – für viele Anwendungen ersetzen.

Kapitel 14

TN–LCDs als phasenschiebende, optische Elemente

In diesem Kapitel wird die Verwendung industriell verfügbarer *Liquid Crystal Displays* (LCDs) als phasenschiebende Elemente besprochen. Dabei werde ich mich insbesondere auf den Einsatz zur Erzeugung computergenerierter Hologramme beschränken. Die Ergebnisse lassen sich aber direkt auch auf andere optische Anwendungen (Korrelatoren, Interferometrie, optische Informationsverarbeitung) übertragen.

Eine allgemeine Darstellung der Physik und Technik von LCDs kann im Rahmen dieser Diplomarbeit nicht erbracht werden. Dieses Wissen wird als bekannt vorausgesetzt. Bei Bedarf sei auf die Fachliteratur, in der eine Vielzahl von hervorragenden Einführungen (siehe .z.B. [good73]) vorliegen, verwiesen.

Zunächst werde ich die für CGH–Anwendungen relevanten Eigenschaften und Grundlagen anhand von *Twisted–Nematic–LCDs* (TN–LCD) schildern. Der Grund für die intensive Betrachtung der TN–LCDs ist durch ihre Marktdominanz begründet. TN–LCDs sind vergleichsweise preiswert auch in optischer Qualität und in der für optische Zwecke nötigen geringen Stückzahl und Größe erhältlich.

14.1 Typischer technischer Aufbau eines TN–LCDs

Der typische Aufbau eines TN–LCDs ist in Abbildung 14.1 dargestellt.

Eine Änderung der optischen Eigenschaften (s.u.) wird durch ein über transparente, meist lithographisch belichtete Elektroden¹ angelegtes Feld hervorgerufen. Außer dieser Beeinflussung der optischen Eigenschaften über elektrooptische Effekte, haben auch andere physikalische Parameter (Magnetfeld, Temperatur, akustische Energie, elektromagnetische Strahlung) Einfluß auf das Verhalten der Flüssigkristall–Schicht, die meist aus organischen, stabförmigen Molekülen besteht. Für Display–Anwendungen wird wegen der unkomplizierten und schnellen Ansteuerung in der Regel immer eine elektrische Regelung verwendet. Es ist aber zu beachten, daß Umgebungsparameter (insbesondere die Temperatur) eine zusätzliche (und für unsere Anwendungen unerwünschte) Änderung der optischen Eigenschaften bewirken können (insbesondere bei Spezialanwendungen mit weitem Bereich der

¹in der Regel Indium–Oxid (ITO) und/oder (Antimon–dotiertes) Zinn–Oxid

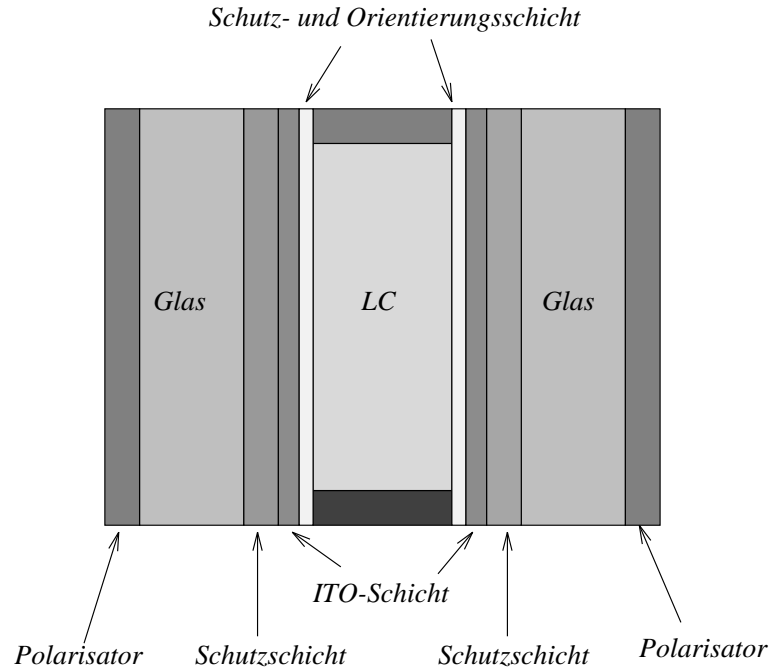


Abbildung 14.1: Prinzipieller -Aufbau eines TN-LCD

Umgebungstemperatur).

Als Spacer werden meist inerte, organische Materialien (z.B. Mylar oder Teflon) verwendet. Die Flüssigkristallschichtdicke liegt üblicherweise im Bereich von 10 bis 20 μm , kann aber auch 5 oder 50 μm betragen. Zwischen dem Glas und den Elektroden findet man oft eine zusätzliche diffusionshemmende Schicht (z.B. SiO_2 -Verbindungen), die das Eindringen von Alkali-Ionen des Glases in die Elektroden bei den für die Bearbeitung relativ hohen Temperaturen (400 Grad Celsius) verhindert.

Die Displays sind hermetisch versiegelt², um ein Eindringen von Fremdstoffen (z.B. Wasser) zu verhindern. Dabei verbleiben zunächst mehrere Öffnungen im Abstandsrahmen, in die der Flüssigkristall eingefüllt werden kann (Vakuumbefüllung).

Im feldfreien Fall wird linear polarisiertes Licht entlang der Verdrillung um 90 Grad gedreht. Man spricht von “waveguiding”. Wenn die angelegte Spannung die sogenannte *Schwellwertspannung* übersteigt, geht der Flüssigkristall in den homeotropen Zustand über. Der waveguiding Effekt ist aufgehoben und die Polarisationssebene des einfallenden Lichts wird nicht gedreht. Der Haupteffekt ist also – im Gegensatz zu LCDs die nach dem Prinzip der feldinduzierten Doppelbrechung arbeiten – eine Drehung der Polarisationssebene mit einem nur geringen Anteil an Doppelbrechung.

Als Schwellwertspannung ergibt sich laut [good73] (3)

$$V_s^2 = \frac{1}{\Delta\epsilon\epsilon_0} \left[K_{11} \frac{\pi^2}{4} + (K_{33} - 2K_{22})\varphi_0^2 \right] \quad (14.1)$$

Dabei sind die K_{ii} elastische Konstanten des Flüssigkristall. φ_0 ist der Verdrehungswinkel (hier 90 Grad), ϵ ist die Dielektrizitätskonstante des Vakuums und $\Delta\epsilon$ ist die Änderung der

²Material zur Versiegelung muß zum Flüssigkristall passen

Dielektrizitätszahl.

Die Schwellwertspannung kann — je nach verwendetem Flüssigkristall — relativ klein werden (0.9 V).

14.2 LCD-Adressierungstechniken

14.2.1 Elektrische Adressierung

Im folgenden wird kurz auf die Ansteuerung von LCDs mit einer großen Anzahl von Einzel-elementen eingegangen. Die Kenntnis der Ansteuerung ist insbesondere dann von Belang, wenn man die Ansteuersignale am LCD konkret zur Qualitätsbeurteilung (s.u.) messen will.

Bei der $V:V/2$ Ansteuerung wird an die Spaltenleitung des anzusteuernenden Pixels die Spannung $-V/2$ angelegt, Die Zeilenleitung wird auf die Spannung $+V/2$ gelegt. Damit ergibt sich eine Spannung von V an der gewünschten Zelle. An allen anderen Zellen liegt "nur" eine Spannung mit dem Betrag $|V/2|$. Im einfachsten Fall wählt man für V die Schwellspannung V_s .

Wenn man eine kurze Ansprechzeit der Zellen (also eine schnelle Reaktion auf eine Bildänderung) will, dann ist es erforderlich, eine größere Spannung als V_s zur Verfügung zu haben (das ist unmittelbar ersichtlich, weil der Einschwingvorgang bis zu V_s damit schneller erfolgt). Damit handelt man sich aber das Problem ein, daß auch falsche Zellen geschaltet werden, wenn die Spannung größer $2V_s$ gewählt wird.

Um das Problem zu umgehen, kann zu einer $V:V/3$ Ansteuerung übergegangen werden. Dabei werden die Zeilen mit 0 oder $-2/3V$ und die Spalten mit $\pm V/3$ angesteuert. Damit ergibt sich an der gewünschten Zelle die Spannung von V , an den anderen Zellen aber lediglich $V/3$. Der Störabstand ist also erhöht.

Je größer die Anzahl der anzusteuernenden Elemente ist, desto größer werden auch die Probleme einer adäquat schnellen Ansteuerung. Bei gleicher Refreshrate muß zwangsläufig die Zeit, zur Ansteuerung eines Pixels abnehmen, wenn die Anzahl der Pixel erhöht wird.

Jedes Pixel kann also nur relativ kurz mit einem Signal versorgt werden. Man geht zu sogenannten "line-at-a-time" Displays über. Dabei werden immer alle Pixel einer kompletten Zeile gleichzeitig beschrieben. Alle Spalten werden also mit den entsprechenden Pixelwerten geladen. Erst dann wird ein Impuls an die zu beschreibende Zeile angelegt. Damit wird die Ansteuerzeit von $10^{-7}s$ auf $60 \mu s$ bei einem typischen TV-Bild verlängert.

Das Zeitverhalten konventioneller Flüssigkristalle liegt in der Größenordnung von Millisekunden und ist damit offensichtlich für eine solche Ansteuerung zu langsam. Um dieses Problem zu umgehen, speichert jedes ansteuerbares Pixel lokal die auszugebende Information. Dazu wird ein zusätzlicher Kondensator verwendet³. Es muß nun aber verhindert werden, daß die dort gespeicherte Energie zu benachbarten Zellen abfließt. Dies wird durch den zusätzlichen Einbau einer Diode oder eines Transistors als Isolationselement – wie in Abbildung 14.2 zu sehen ist – erreicht.

Das nichtlineare Element (Transistor, Diode) hat dabei die zusätzliche Wirkung der Vers-

³bis ca. 100.000 Pixel genügt unter Umständen auch die Kapazität der Flüssigkristall-Schicht.

teilung der Ansteuer-Kennlinie. Damit wird der Kontrast erhöht. Andererseits bringt die Nichtlinearität das Problem mit sich, daß möglichst die Kennlinien aller Pixel exakt dieselbe Nichtlinearität aufweisen sollten, um sichtbare Fehler im Bild zu vermeiden. Die Nichtlinearität selbst (z.B. Diodenkennlinie) ist aber stark temperaturabhängig und es ist schwierig, alle Dioden exakt gleich herzustellen.

Bei Verwendung von Dünnschicht Feldeffekt-Transistoren (*Thin-Film-Transistor (TFT)*) entfallen diese Probleme, da das Element direkt als Schalter wirkt und die Nichtlinearität damit so stark wird, daß – unabhängig von Temperatur- und Einzelelement-Streuungen – der Transistor in jedem Fall in die Sättigung geht oder ungeschaltet bleibt (siehe Abbildung 14.2). Dünne TFT-Schichten aus CdS oder CdSe sind in der Regel für sichtbares Licht ausreichend transparent. Dieser Vorteil entfällt bei der technologisch besser eingeführten Si-Technologie, die den Vorteil hat, daß ohne Probleme gleich ein erheblicher Teil der Ansterelektronik auf dem Display selbst untergebracht werden kann (daraus ergibt sich natürlich eine Reduktion der Anschlüsse und damit verbunden auch der Kosten).

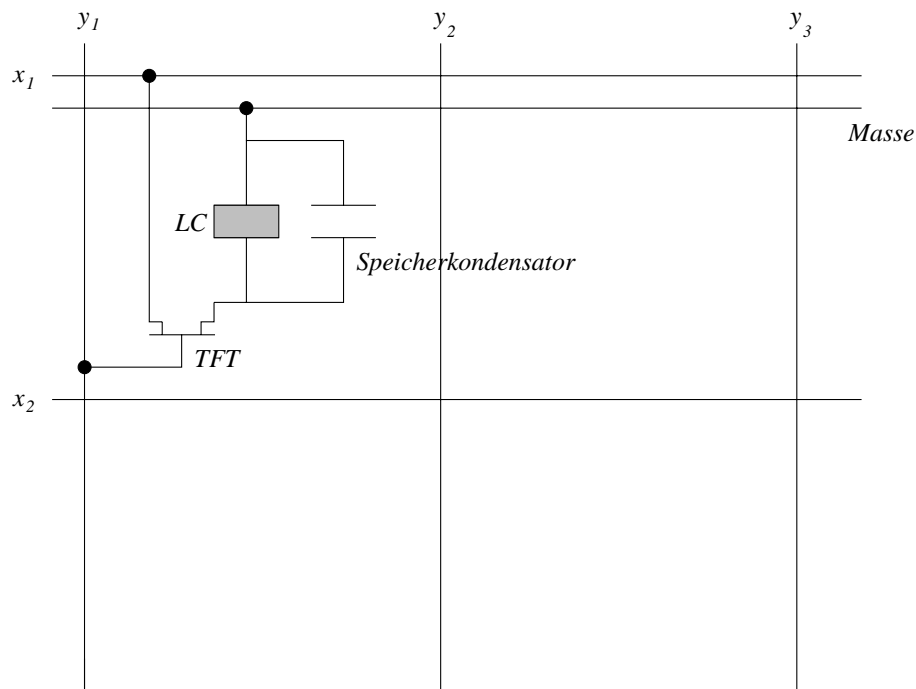


Abbildung 14.2: Pixelaufbau mit TFT und Speicherkondensator

14.2.2 Optische Adressierung

Eine optische Ansteuerung kann entscheidende Vorteile haben: in der Flüssigkristallschicht werden keine Anschlüsse und Steuerelemente benötigt. Es entfällt also eine Substruktur, die erstens die Intensität und Beugungseffizienz vermindert, zweitens die Kosten erhöht und drittens die Miniaturisierung beschränkt. Kirk et al. (siehe [kirk94]) benutzten eine Anordnung bei der ein elektrisch adressiertes LCD verkleinert auf einen optisch adressierten SLM abgebildet wird. Dabei erzielten sie Pixelabstände von $12 \mu\text{m}$. Maximal lassen sich momentan Auflösungen von bis zu 300 Zeilen/mm erreichen.

Für einen Einsatz im Bereich computergenerierter Hologramme ergibt sich bei einer optischen Ansteuerung allerdings das Problem, daß noch ein weiteres Ausgabemedium (digital

→ optisch), z.B. ein zweites LCD, verwendet werden muß. Die Folge davon ist ein erhöhter technischer wie finanzieller Aufwand. Momentan wollen wir diesen Aufwand vermeiden, weshalb hier nicht näher auf optische Ansteuerungen von LCDs eingegangen wird.

14.3 Konkurrierende Lichtmodulatoren

Welche Echtzeit-Medien sich für CGHs in Zukunft durchsetzen werden, kann nicht mit Bestimmtheit vorausgesagt werden. Insbesondere Japan sieht die Zukunft im LCD-Bereich (das wundert angesichts der Markt- und Forschungsdominanz Japans in diesem Bereich allerdings nicht). Es werden Entwicklungen hin zu holographischem Video⁴ langfristig als möglich erachtet (auch wenn die Bildqualität bisheriger Ergebnisse weit von diesem Ziel entfernt ist). Sehr viele Artikel zu diesem Thema werden (leider oft nur auf japanisch) veröffentlicht (siehe z.B. [hash92], [sato92]).

Alternativ können statt LCDs auch andere elektrisch adressierbare, echtzeitfähige SLMs eingesetzt werden. Die in dieser Arbeit dargestellten Algorithmen sind weitgehend unabhängig von der Wahl der Display-Technologie und lassen sich direkt für diese SLMs einsetzen. Im folgenden werden kurz die entsprechenden Technologien mit ihren Stärken und Schwächen in Hinblick auf den Einsatz als CGH-Medien aufgeführt:

Digital-Mirror-Devices (DMDs) (siehe [pape83], [greg88]): miniaturisierte Spiegel werden über verschiedene Effekte⁵ bewegt.

- bis zu 2048×1152 Pixel (Pixelgröße: $17 \times 17 \mu m$) sind kommerziell erhältlich
- maximale, theoretische Bildfrequenz: 100kHz
- Kontrast: bis zu 400:1
- hohe Füllfaktoren (Verhältnis aktive Pixelfläche zu gesamter Pixelfläche): 85 - 90%
- *bisher* sind nur Amplitudenhologramme realisierbar⁶
- langfristig wahrscheinlich sehr günstiges Preis-Leistungsverhältnis erreichbar

Multiple Quantentrog (SEED) Modulatoren (siehe [lent93], [mill87], [worc96]): extrem dünne Schichten (z.B. 10 nm) aus verschiedenen Halbleitern (*Multiple Quantum well modulators*) zeigen eine ausgeprägte Abhängigkeit der Absorption in Abhängigkeit von angelegten elektrischen Feldern (*Quantum-confined Stark Effect*). Man spricht auch von "SEED-Modulatoren" (*Self-Electroopic Effect Devices*).

- extreme Schaltgeschwindigkeiten (prinzipiell bis zu 100 ps, SLMs mit z.B. 128×128 Pixeln arbeiten momentan mit bis zu 100.000 Bildern/s).
- sehr energiesparend (Im Bereich von $fJ/\mu m^2$)
- reflektierende und transmittierende Anordnungen sind möglich
- Kontrastverhältnisse bis zu 100:1
- Erweiterbar hin zu Smart Pixel Technologie

⁴evtl. nur mit horizontaler Parallaxe um die technischen Anforderungen zu reduzieren

⁵am gebräuchlichsten ist die von Texas Instruments verwendete "Cantilever Beam"-Technik

⁶Versuche, phasenschiebende DMDs herzustellen sind erfolgreich abgeschlossen. Allerdings sind noch keine hochauflösende Elemente erhältlich ([cohn96]).

- noch nicht ausgereift, nicht kommerziell erhältlich

Magnetooptische SLMs (siehe [ross83]): ausgenutzt wird der Faraday-Effekt in Verbindung mit externen Polarisatoren.

- Kein Refresh nötig
- Pixelgrößen von ca. $10\mu m$ bis $500\mu m$
- Schaltzeiten von $50\mu s$ bis $1ms$
- Graustufen-Displays sind schwierig herzustellen

Akustooptische Modulatoren in Verbindung mit zusätzlichem Scannen (siehe z.B. [luce92])

- hohe eindimensionale Auflösung für CGHs mit horizontaler Parallaxe (z.B. Display-Anwendungen)
- hohe Geschwindigkeiten
- großer finanzieller und technischer Aufwand für Gesamtsysteme (AOM, Ansteuerung, Scannen)

Pockels read-out optical Modulator (PROM): (siehe z.B. [iwas92], [nise72]) Photo-refraktive Kristalle (z.B. BSO) werden über den elektrooptischen Effekt zur Speicherung von Bildern und zur Modulation von einfallendem Licht (andere Wellenlänge) verwendet.

- Auflösungen bis zu 300 Zeilen/mm
- Kontrastverhältnis: 1000:1
- Schreib-Lösch-Zyklus: bis zu 1 kHz
- maximale Speicherzeit: ca. 2 Stunden
- bei häufigem Auslesen ist Refresh nötig
- elektrische Spannungen im kV-Bereich nötig

Zusammenfassend bieten aus meiner Sicht neben den LCDs kurzfristig DMDs die größten Erfolgsaussichten als echtzeitfähige CGH-Medien. Mittel- bis langfristig können die SEED-Modulatoren sehr interessant werden.

14.4 Experimentelle Ergebnisse

Schweikhardt ([schw96]) hat im Rahmen seiner Diplomarbeit die Vermessung eines LCDs durchgeführt. Dort findet sich ein exakter Arbeitsplan zur Messung der Jones-Matrix und der dynamischen Phasenschiebung eines LCDs. Die genaue Kenntnis der Jones-Matrix ist wichtig, um das Verhalten des LCDs hinsichtlich Phasen- und Amplitudenbeeinflussung in Abhängigkeit von externen Polarisatoren zu beschreiben. Wenn man sich lediglich für die Phasenschiebung bei einer bestimmten Polarisatorstellung interessiert, dann sind die entsprechenden Meßmethoden auch bei dieser Polarisatorstellung durchzuführen. Nur so ist gewährleistet, daß man die geometrischen und dynamischen Phasenanteile mißt.

Wenn man die komplette Messung der Jones-Matrix vermeiden will, bietet es sich an, zunächst in einem einfachen Mach-Zehnder-Aufbau eine günstige Stellung der Polarisatoren zu ermitteln. Dabei achtet man auf folgende Punkte

- Gesamtintensität sollte nicht zu gering werden
- Phasenschiebung sollte möglichst groß werden
- Amplitudenvariation (mit Phase) sollte möglichst klein sein

Für diese Stellung der Polarisatoren wird dann im Mach-Zehnder Interferometer eine grob-orientierende Phasenmessung durchgeführt. Weiterhin kann auf einfache Weise die Unebenheit der LCD-Oberfläche untersucht werden. Dazu werden einige Streifen im Interferometer eingestellt und eine Messung des Streifenbildes ohne LCD und mit LCD durchgeführt. Die sich ergebenden Bilder werden automatisch ausgewertet.

Eine exakte Phasenmessung kann dann über verschiedene Meßmethoden erfolgen. Wir verwenden die meist gebräuchlichste Methode (z.B. [reyn89] 22.3.2⁷), eine Doppelspaltanordnung bei der die beiden Spalte unterschiedliche Bereiche des LCDs abdecken. Damit erhält ein Spalt die Referenzphase (Grauwert 0) und der andere Spalt die zu messende Phase (beliebiger Grauwert). Im Beugungsbild ergibt sich ein Streifenmuster, das mit zunehmender Phasenschiebung wandert. Eine genaue Analyse findet man bei Schweikhardt ([schw96]). Diese Methode wird häufig verwendet und hat den Vorteil, daß keine geeichte Intensitätsmessung (mit CCDs ist das aufwendig) nötig ist.

Weitere Arbeiten, die sich mit der Messung der Phasenschiebung von LCDs beschäftigen, findet man in [mccl96], [dou96].

In diesem Kapitel werde ich einige praktische Ergebnisse mit einem neuen LCD beschreiben. Leider stand uns das LCD nur sehr kurze Zeit als Leihgabe des Herstellers zur Verfügung. Auf eine genaue Vermessung mußte daher aus Zeitgründen verzichtet werden. Trotz einer sicherlich nicht optimalen Einstellung des LCDs, konnten sehr gute Rekonstruktionen der Hologramme erzielt werden. Der Vergleich mit den am videoadressierten LCD gewonnenen Erfahrungen ermöglicht es uns, einige allgemeine Empfehlungen zur Verwendung von LCDs als Hologrammedien in einem abschließenden Abschnitt dieses Kapitels zu geben.

14.4.1 Eigenschaften des VGA1-LCDS von CRL

Das verwendete LCD "VGA1" ist ein neues Produkt der Firma CRL Smectic Technology⁸. Es handelt sich um ein VGA-adressiertes TFT twisted nematic LCD mit 4 Graustufen und 640×480 Pixeln. Der Hersteller kann bisher keine Angaben über optische Eigenschaften des Displays machen. Geplante Anwendungen sind optische Korrelatoren, Head-mounted-Displays, allgemeine Displays und Markierungssysteme.

Die technischen Daten ergeben sich nach dem Datenblatt und unseren ergänzenden Messungen zu:

- 640×480 Pixel
- aktive LCD-Fläche: 26.9×20.2 mm
- Pixelgröße: $24 \times 31 \mu\text{m}$ (Pixelpitch: $42 \times 42 \mu\text{m}$)
- maximale Bildwechselfrequenz: 25 Hz

⁷leider mit einem Rechenfehler

⁸CRL, Dawley Road, Hayes, Middlesex, UB3 1HH, UK, (smectic@thorn-emi-crl.co.uk)

- Kontrast $> 30:1$
- Quantisierung in 4 Graustufen
- großflächige Phasenvariation über das gesamte Display: ca. 2π
- maximale Phasenschiebung⁹: 0.8π

Besonderer Vorteil des LCDs ist die Ansteuerung über ein VGA-Signal (auf die entscheidende Wichtigkeit dieses Punktes werden wir später noch näher eingehen).

Das LCD ist zu einem Preis von ca. 10.000 DM über die Firma Döhner Elektrooptik in Deutschland erhältlich.

Alle Messungen wurden mit einem ITI FG 100 Framegrabber und einer Pulnix TM 765 CCD durchgeführt.

Mikroskopische und elektronische Messungen

Um die mikroskopische Struktur des LCDs beurteilen zu können, wurde ein einfaches Mikroskop aufgebaut, mit dem das LCD im arbeitenden Zustand betrachtet werden konnte. Damit wurde es möglich, die Schärfe des Bildes, die Homogenität der Pixel und mikroskopische Defekte beurteilen zu können.

Das LCD wurde zwischen gekreuzten Polarisatoren (entsprechend einem hohen Kontrast) in kollimiertem Laser-Licht (HeNe 633 nm) verwendet. Die vom Hersteller verwendeten Polarisationsfolien wurden durch optische Polarisationsfilter ersetzt, um Fehler aufgrund von minderwertigen Polarisationsfolien zu vermeiden. Zur Abbildung wurde ein vorhandenes Mikroskop-Objektiv (170- 3.2/0.12) verwendet.

Der Anteil der defekten Pixel (Totalausfall des Pixels) liegt im Bereich von 0.5 .. 1 %. Die Homogenität der Pixel (Änderung der Helligkeit einzelner Pixel) wurde rein visuell überprüft, war aber sehr zufriedenstellend (Abweichungen $< 10 \%$).

Auf der Ansterelektronik des LCDs ist ein Potentiometer zur Justierung der Pixel-Dot-Clock eingestellt. Mit dieser Einstellung wird erreicht, daß die Pixel im Computer exakt auf die physikalischen Pixel des LCDs geschrieben werden. Die vertikale Synchronisation der Bilder ist unproblematisch weil vertikale Synchronisationssignale in allen gebräuchlichen Videoformaten vorhanden sind. Eine horizontale Synchronisation innerhalb einer Zeile fehlt aber weil die nötigen Bandbreiten für diese Synchronisation zu hoch sind. Hier liegt das Hauptproblem von *videoadressierten LCDs*. Die Diskretisierung im Computer ist nicht identisch mit der späteren Diskretisierung in LCD Pixel. Es ergeben sich starke Störungen. Wenn man beispielsweise einen einzelnen Pixel setzen will, dann werden auf dem LCD in der Regel mehrere nebeneinanderliegende Pixel angesteuert. Die Ansteuerung erfolgt dabei nicht mit der gewünschten, sondern einer abgeschwächten Spannung. Während dies für Display-Anwendungen oft noch tragbar ist, ergeben sich bei computergenerierten Hologrammen starke Störungen.

Insbesondere hochfrequente Anteile im Hologramm, also einzelne Pixelmuster, werden nicht richtig dargestellt. Dementsprechend wird eine Rekonstruktion am Rand stark abgeschwächt sein. Außerdem treten zusätzliche Geisterbilder auf. Wir haben lange Zeit

⁹bei sinnvollen Intensitätsmodulationen (s.o.)

versucht, diese Fehler mit verschiedenen Methoden bei unserem videoadressierten LCD zu vermindern. Insbesondere wurde versucht, eine Grafikkarte im Lowlevel Bereich so zu programmieren, daß der Verschmierungseffekt vermindert ist und wirklich einzelne Pixel angesteuert werden können. Weiterhin gab es Versuche, das Problem bereits bei der Hologrammberechnung zu berücksichtigen. Da letztlich aber eben keine hochfrequenten Anteile korrekt dargestellt werden können, führte auch dieser Ansatz nur zu geringen Verbesserungen.

Eine wirkliche Verbesserung kann nur durch eine verbesserte Ansteuerelektronik erreicht werden. Hier bieten sich zwei Verfahren an: bei einer voll digitalen Ansteuerung fallen die Probleme natürlich weg. Ideal wäre demgemäß eine spezielle Grafikkarte mit rein digitalem Ausgang. Die entsprechende Elektronik wäre dafür allerdings extrem aufwendig, was man leicht an der Komplexität herkömmlicher Grafikkarten abschätzen kann. Ein Selbstbau ist für uns kaum sinnvoll bzw. durchführbar.

Die zweite Möglichkeit ist, ein quantisiertes Signal für die Ansteuerung zu verwenden. Das Display kann sich anhand der Quantisierungsstufen selbst korrekt synchronisieren. Als Signalnorm bietet sich dabei VGA an, da die überwiegende Mehrzahl der heute verwendeten Computer diese Norm unterstützt. Mit dieser Art der Ansteuerung können wir herkömmliche Grafikkarten verwenden. Dementsprechend wird der Computer nicht zur laufenden Ausgabe des Hologramms benötigt, sondern kann sich anderen Aufgaben zuwenden (z.B. Einlesen von CCD-Bildern und Auswertung von gemessenen Daten). Weiterhin können Standardroutinen zur Darstellung von Grafiken verwendet werden und wir sind nicht darauf angewiesen, eigene Graphik-Grundfunktionen zu programmieren.

Die Entwicklung eigener Ansteuerelektroniken ist aus meiner Sicht daher ähnlich unsinnig wie die Entwicklung eines eigenen Computers. Der Aufwand, um die Qualität industrieller Lösungen zu erreichen, ist zu hoch und lohnt sich nur für Spezialanwendungen.

Diese Zusammenhänge brachten uns zu der Überzeugung, daß mit VGA-LCDs eine deutlich verbesserte Rekonstruktionsqualität der Hologramme erreicht werden sollte, was uns letztlich zum Test des hier beschriebenen LCDs führte.

Das VGA-LCD synchronisiert sich automatisch auf die digitalen VGA-Signale des Computers. Diese Synchronisationsschwelle ist korrekt einzustellen um ein scharfes und helles Bild zu erhalten. Die Einstellung wurde im Mikroskopaufbau untersucht. Wie erwartet, ergeben sich bei korrekter Einstellung nicht die vom videoadressierten LCD her bekannten Verschmierungseffekte. Dies ist auch der Hauptgrund für die deutlich bessere Rekonstruktionsqualität, die wir mit dem neuen LCD erreichen.

Die korrekte Ansteuerung wurde nicht nur rein optisch, sondern auch elektronisch mit einem Oszilloskop untersucht. Wie erwartet, ergaben sich saubere Signale, die den von uns eingeschriebenen Pixelmustern entsprachen. Bei dieser Untersuchung läßt sich auch gut die Variation der Ansteuerspannung mit unterschiedlichen Einstellungen der übrigen Regler (Helligkeit, Kontrast, Nichtlinearität der einzelnen Stufen) beobachten. Die Konstanz der Ansteuerspannungen einzelner Pixel ist besser als 5 % und es finden im Gegensatz zum videoadressierten LCD keine globalen oder lokalen Änderungen der Ansteuerspannungen in Abhängigkeit vom Bildinhalt statt.

Bei dem videoadressierten LCD, das für reine Display-Anwendungen gedacht war, regelt die Ansteuerelektronik automatisch je nach Bildinhalt die Ansteuerspannungen nach. Das ist vergleichbar mit einer automatischen Belichtungsregelung bei Foto-Kameras. Für uns ist eine solche Regelung natürlich sehr störend, da wir bei der Ausgabe eines bestimmten

Pixelwerts eine entsprechende Spannung erwarten, die unabhängig von den anderen Pixeln des Hologramms sein soll.

Auch hier ist das neue VGA-LCD klar dem bisher verwendeten LCD vorzuziehen.

Eine Untersuchung der elektrischen Eigenschaften (z.B. Konstanz der Ansteuerspannung) über längere Zeiträume konnte nicht durchgeführt werden, da uns das Display nicht lange genug zur Verfügung stand.

Großflächige Phasenvariation und Phasenschiebung

Zum Zeitpunkt der Messung stand leider kein 4-Phasen-Interferometer zur Verfügung, mit dem wir eine genaue Vermessung der Phasenvariation über die LCD Fläche hätten durchführen können. Deshalb wurde ein einfacheres, aber für die von uns benötigten Zwecke ausreichendes, Meßverfahren verwendet.

In einem Mach-Zehnder-Interferometer werden ohne LCD ca. 20 waagrechte Interferenz-Streifen eingestellt. Das entsprechende Bild wird aufgenommen. Ein weiteres Bild wird nach dem Einbringen des LCDs (ohne Spannungen) aufgenommen.

Die beiden Bilder werden dann mit einem einfachen Mittelwertfilter (*pnmconvol*) von groben Störungen befreit. Anhand der beiden Bilder wird über ein Programm die Phasenschiebung berechnet. Ausgehend von den Zeilenanfängen werden die Interferenzstreifen in beiden Bildern abgefahren. Dabei wird jeweils das Maximum des Interferenzstreifens als Koordinate gespeichert. Die Differenz der Koordinaten zwischen den beiden Bildern liefert zusammen mit der lokalen Periode der Interferenzstreifen direkt die Phasenschiebung und damit die Phasenstörungen über die LCD-Fläche.

Bei der Ermittlung des lokalen Maximums der Interferenzstreifen wird die ungefähr vorausgesagte Position dieses Maximums berücksichtigt. Außerdem ist ein minimaler Kontrast gegenüber den Dunkelstreifen zu fordern. Das Programm verfolgt die Interferenzstreifen korrekt bis zu einer maximalen Steigung der Streifen von 45 Grad¹⁰. Die Interferenzstreifen werden in diskrete Bereiche eingeteilt, für die jeweils eine Phasenschiebung berechnet wird. Das Ergebnis der automatischen Auswertung ist in Abbildung 14.3 dargestellt.

Die Phasenstörung ist annähernd sphärisch. Der Mittelpunkt dieser sphärischen Störung liegt in einer der Ecken des LCDs.

Nachteil dieses einfachen Verfahrens ist die geringe räumliche Auflösung (1:20 bei 20 Streifen), die für den großflächigen Fehler aber ohne Belang ist.

Die Genauigkeit der Phasenbestimmung selbst wird folgendermaßen abgeschätzt: Durch die Mittelung (jeweils ca. 30 Pixel) kann von einer Bestimmung des Maximums auf mindestens 2 Pixel genau ausgegangen werden. Bei einer Periode der Interferenzstreifen ist die Genauigkeit damit $2/30 * 2\pi = 0.13\pi$.

Die Messung der maximalen Phasenschiebung wurde mittels des Mach-Zehnder Interferometers und der in [schw96] beschriebenen Methode (Doppelspalt) durchgeführt. Bei beiden Verfahren ergab sich ein Phasenschieberegion von lediglich 0.8π .

Mit dem VGA-LCD können 4 Quantisierungsstufen erreicht werden. Die entsprechenden,

¹⁰ anders orientierte Streifen lassen sich einfach dadurch untersuchen, daß die entsprechenden Bilder vor der Auswertung gedreht werden

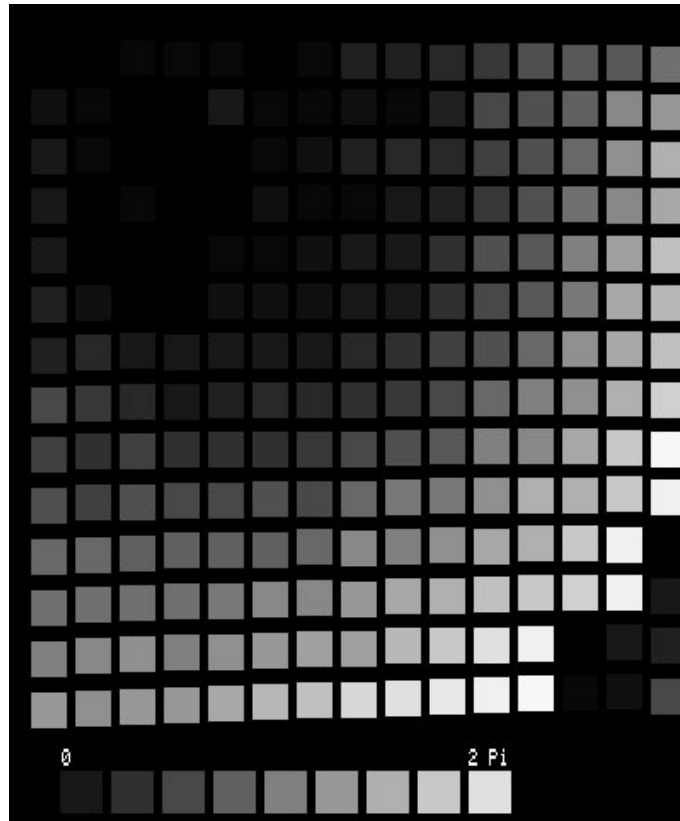


Abbildung 14.3: Phasenfehler des CRL VGA-LCDs

am LC anliegenden, Spannungen lassen sich über zwei Potentiometer an der Ansteuerelektronik regeln. Leider ist die Regelung sehr empfindlich und kaum praktikabel durchführbar. Da nur zwei Regler vorhanden sind und beide Regler alle vier Spannungen beeinflussen, ist ein iteratives Vorgehen zur Einstellung nötig. Während das bei Display-Anwendungen, bei denen direkt der Kontrast eingestellt wird, recht unproblematisch ist, ergeben sich bei der verhältnismäßig komplizierten Messung der Phasenschiebung hier erhebliche Probleme.

In der Praxis war es uns nicht möglich, vier annähernd äquidistante Phasenwerte einzustellen. Die Einstellung von drei Phasenwerten war bereits recht schwierig zu erreichen. Hier kann man dem Hersteller nur empfehlen, entsprechend vier Potentiometer (für jede Ansteuerungsspannung einen getrennten Regler) vorzusehen, wenn das Element zur Phasenschiebung eingesetzt werden soll.

Phasenschiebung in einer Reflexionsanordnung

Das LCD wurde ebenfalls in einer reflektiven Anordnung untersucht. Reflexionsanordnungen haben den Vorteil, daß idealerweise eine doppelte (dynamische) Phasenmodulation erreicht werden kann. Im praktischen Aufbau muß bei einer solchen Anordnung sichergestellt werden, daß das Licht nach der Reflexion wieder durch exakt denselben Pixel geworfen wird. Dies macht im Falle eines einfachen Spiegels wegen der geringen Pixelgrößen eine extrem genaue Justierung nötig. Alternativ kann der Spiegel direkt auf das LCD aufgedampft werden.

Barnes et al. haben ein afokales Reflexionssystem verwendet, das in [barn89] genauer be-

geschrieben wird. Diese Anordnung ist wahrscheinlich die eleganteste Lösung. Es bleibt aber festzustellen, daß der Aufwand gegenüber der transmittiven Anordnung deutlich erhöht ist und natürlich auch eine geringere Gesamtintensität bei doppeltem Durchlauf durch das LCD erreicht wird. Aus diesen Gründen fassen wir die Anordnung in Reflexion momentan nicht ins Auge.

Die Messungen am VGA-LCD zeigten zudem, daß in dieser Anordnung starke räumliche Inhomogenitäten der Phasenschiebung bei diesem LCD auftreten, die einen praktischen Einsatz ebenfalls nicht nahe legen.

14.4.2 Optische Rekonstruktion computergenerierter Hologramme mit dem VGA-LCD

Wie die in den letzten Abschnitten beschriebenen Messungen erwarten lassen, liefert das VGA-LCD eine gegenüber unserem videoadressierten LCD deutlich verbesserte Rekonstruktionsqualität. Trotz des geringen Phasenschiebereichs von 0.8π und der geringen Anzahl von Quantisierungsstufen ergeben sich erfreulich gute Rekonstruktionen.

Im folgenden werden einige optische Rekonstruktionen für Hologramme unterschiedlicher Größe und Berechnungsalgorithmen gezeigt. Alle CGHs wurden über den Gerchberg-Saxton-Algorithmus berechnet und mit einem einfachen Hard-Clipping quantisiert, so daß eventuell noch weitere Verbesserungen in der Zukunft möglich sind. Die gesamten Messungen standen unter starkem Zeitdruck (LCD als Leihgabe), so daß das LCD sicher noch nicht optimal verwendet wurde (Anpassung der Hologrammberechnung an LCD).

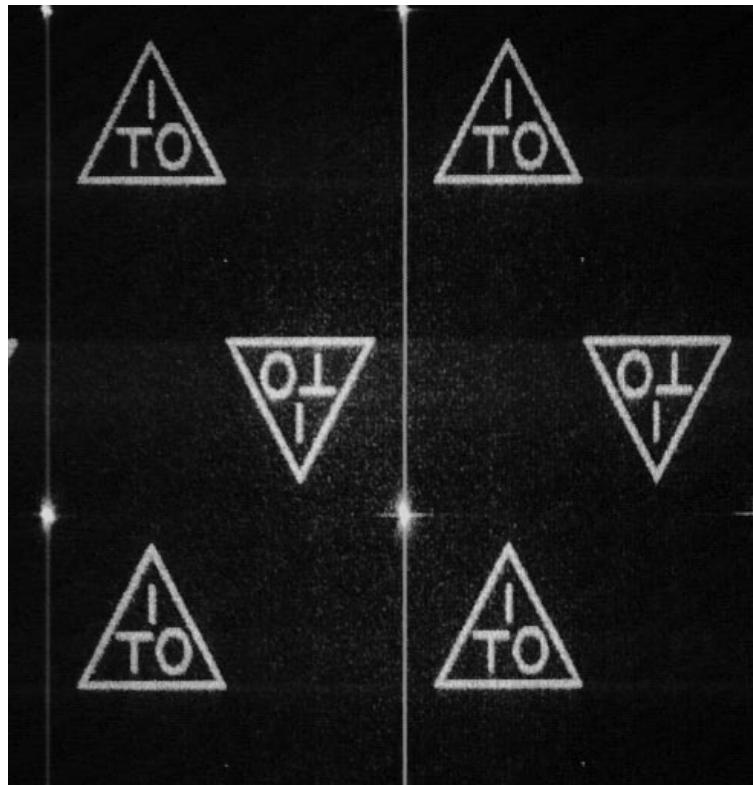


Abbildung 14.4: Optische Rekonstruktion: 256×256 Hologramm 4-fach parkettiert

Bei einer Verstellung der Pixel-Dot-Clock, resultierend in einer Desynchronisation der Pixel

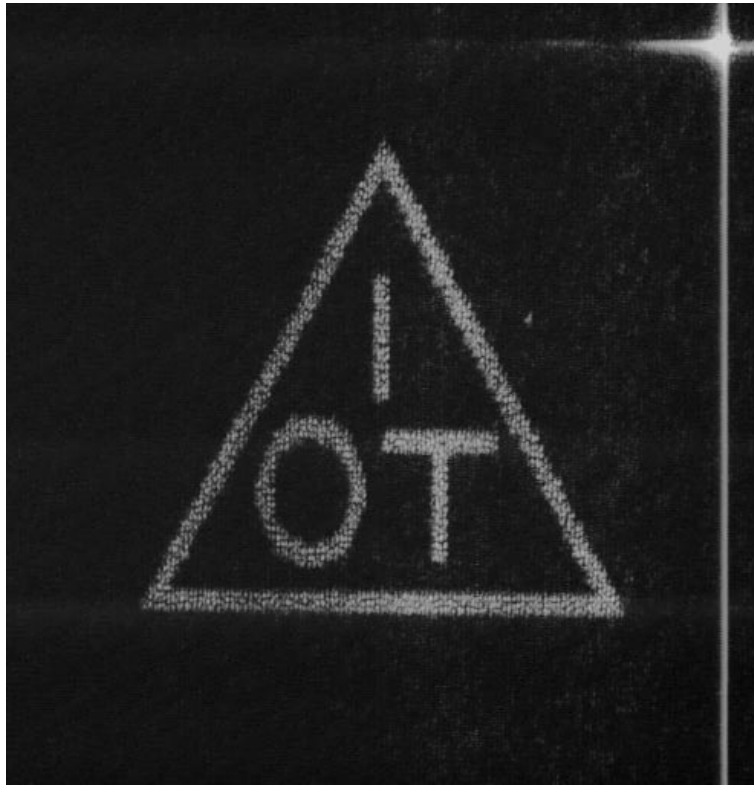


Abbildung 14.5: Optische Rekonstruktion: 256×256 Hologramm 4-fach parkettiert, vergrößerte Darstellung

ergibt sich der in Abbildung 14.8 gezeigte und vom videoadressierten LCD bekannte Effekt zusätzlicher Geisterbilder und der Abschwächung in den Randbereichen (siehe Diskussion im Abschnitt 14.4.1).

Die Rekonstruktion eines Grauwertbildes (siehe Abbildung 14.7) ist — wie erwartet — weniger gut.

14.4.3 Auswahl von LCDs als Hologrammedien

In diesem Abschnitt sollen die wesentlichen Merkmale, auf die beim Kauf von LCDs für computergenerierte Hologramme geachtet werden muß, nochmals tabellarisch zusammengestellt werden.

Auf folgende Display-Parameter sollte beim Kauf geachtet werden:

- Optische Ebenheit der Glaskörper
- Polarisationsfolien sollten sich entfernen lassen
- Grund-Geometrie (Apertur, Pixelanzahl, Mikrostruktur)
- möglichst gute Homogenität der Pixel
- maximaler Phasenschiebebereich (möglichst groß)
- Anzahl effektiver Phasenstufen (möglichst groß)

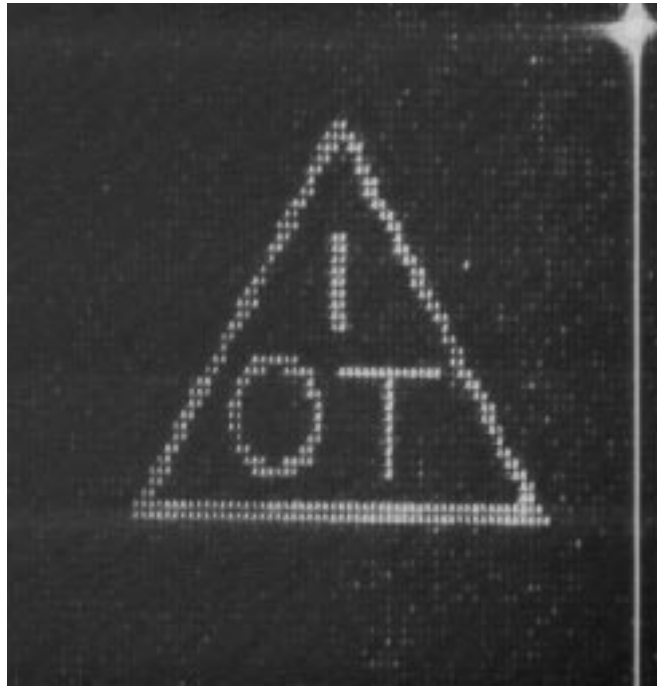


Abbildung 14.6: Optische Rekonstruktion: 128×128 Hologramm 16-fach parkettiert, vergrößerte Darstellung

- digitale Ansteuerung¹¹!
- bei Farb-LCDs auf Möglichkeit zum Farbabgleich der einzelnen Farbkanäle (RGB) achten
- keine “intelligente” Helligkeits- oder Kontrastregelung des Displays

¹¹falls dennoch eine analoge Ansteuerung verwendet wird, muß darauf geachtet werden, daß sich folgende Parameter nicht ändern: Videokabel, Framegrabber-Spannung, Helligkeitsregelung und Stabilität der Ansteuerung



Abbildung 14.7: Optische Rekonstruktion eines Graustufenbildes

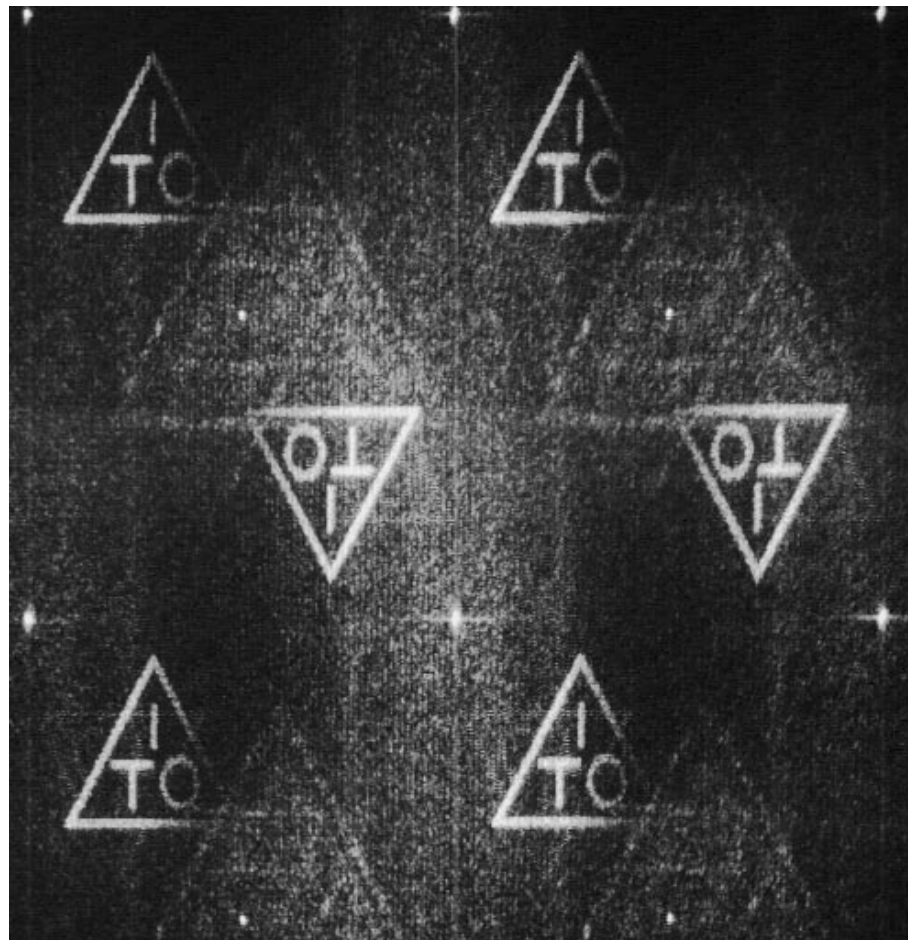


Abbildung 14.8: Optische Rekonstruktion bei falsch eingestellter Horizontalsynchronisation

Kapitel 15

Computer, Rechenzeiten, Benchmarks

In diesem Kapitel wird auf die praktischen Probleme und Möglichkeiten zur Berechnung computergenerierter Hologramme mittels digital-elektronischer und optischer Methoden eingegangen. Dabei will ich in erster Linie praktische Hilfen bei der konkreten Auswahl von Algorithmen und Hardware geben.

Die bisherige Beschreibung der Algorithmen zur Hologrammberechnung wird jetzt dahingehend erweitert, daß die praktische Umsetzung der Algorithmen in konkreter Hardware besprochen wird. Die entsprechende Hardware wird in der Regel ein herkömmlicher general-purpose Computer sein, aber auch spezielle Beschleuniger können heute zu vertretbaren Kosten eingesetzt werden. Bei der Umsetzung müssen wir uns vorrangig um die effiziente Berechnung der diskreten Fourier-Transformation kümmern, da diese letztlich die Hauptrechenzeit in Anspruch nimmt.

15.1 Grundsätzliche Beurteilung

Der Computer ist das zentrale Hilfsmittel zur Berechnung der CGHs. In der Praxis ist die Entscheidung zu treffen, welche Algorithmen auf welchen Computern zur Lösung eines bestimmten Berechnungsproblems zu verwenden sind. Allgemeine Entscheidungen der Form: "Grundsätzlich wird Algorithmus X auf dem Rechner Y verwendet" machen hier wenig Sinn. Die Wahl muß sich an der konkreten Anwendung orientieren.

Während zur Berechnung eines Laser-Strahlformungselements eine Optimierungsdauer von Tagen bis Wochen durchaus akzeptabel sein kann, ist für Display-Anwendungen eine möglichst schnelle Berechnung nötig. Nichtzuletzt hängen Rechenzeit und Optimierungsqualität aber auch vom SBP des Hologramms, also der Anzahl ansteuerbaren Pixel, ab.

Die Berechnungs-Algorithmen können in drei Klassen eingeteilt werden:

- schnelle Verfahren
- mittelschnelle Verfahren
- langsame Verfahren

Dabei beziehen sich die Begriffe “schnell” und “langsam” auf die Anzahl von elementaren Rechenschritten, die zur CGH-Berechnung beim jeweiligen Algorithmus nötig sind. Sie sind zunächst losgelöst von der tatsächlichen Rechenzeit auf einem realen Computer zu sehen.

Die Qualität der CGH-Rekonstruktion nimmt mit der Anzahl der Rechenschritte des Algorithmus zu. Für die meisten Anwendungen werden heute mittelschnelle Verfahren eingesetzt. Die Qualitätssteigerung beim Übergang hin zu langsamen Verfahren wird oft als nicht ausreichend signifikant angesehen, um den zusätzlichen Rechenaufwand zu rechtfertigen.

15.2 Benchmarks

Um die Eignung verschiedener Computer und ihrer Konfigurationen in Zukunft besser vergleichen zu können und Fehlananschaffungen zu vermeiden, wurde ein Testprogramm (“cgh-bench.c”) geschrieben. Bei der Anschaffung bzw. Ausstattung neuer Rechner kann so bereits im Vorfeld auf einfache Weise abgeschätzt werden, ob der entsprechende Rechner für eine bestimmte Anwendung ausreichend schnell ist.

Das Programm berechnet die für verschiedene Algorithmen nötigen Grundalgorithmen mit unterschiedlichen Feldgrößen und protokolliert die entsprechenden Ausführungszeiten. Die verschiedenen Teilergebnisse werden zu einem Gesamt-Leistungsindex (“CGHMark”) gewichtet. Dieser CGHMark gibt in etwa die “Leistungsfähigkeit des Systems zur Berechnung von CGHs im allgemeinen” an. Er ist für ein i486 DX2/66 (256kB 2nd level Cache) System zu 1.0 normiert. Eine Tabelle der entsprechenden Leistungsdaten verschiedener Computer findet sich in der folgenden Aufstellung.

System	CGHMark
UltraSparc 170 Solaris 2.5 cc -fast lmopt	7.46
UltraSparc 170 Solaris 2.5 cc -fast	6.68
UltraSparc 1 SunOs 5.5 cc	5.07
P100 (ASync-Cache, Intel FX) gcc -O3, Linux 1.2.8	3.59
P90 (PB Cache, Intel FX) gcc -O3, Linux 1.3.98	3.58
P90 (PB Cache, UMC) gcc -O3, Linux 2.0	2.92
DEC AlphaStation 3000 gcc -O3	2.90
DEC AlphaStation 3000 gcc -O3 (load 0.3)	2.48
Sun SS10 Solaris 2.4 gcc -O3 -lmopt	2.30
AMD K5-P90-SSA (PB Cache, UMC) gcc -O3, Linux 2.0	2.10
Sun SS10 Solaris 2.4 gcc -O3	1.79
Sun Sparcstation LX gcc -O3 -lmopt	1.21
RS6000 (viele andere User) load 0.5	1.12
i486DX2/66 256kB 2nd Level Cache, Linux 1.2.8	1.00
Sun Sparcstation LX cc -fast -lmopt -xlibmil	0.79
Sun Sparcstation LX gcc -O3	0.74

(Einige der Tests wurden freundlicherweise am Rechenzentrum der Universität Ulm von Herrn Ebach durchgeführt)

Es muß darauf hingewiesen werden, daß die Ergebnisse nicht nur vom Rechner selbst,

sondern wesentlich vom verwendeten Betriebssystem und Compiler (und den Compiler-Einstellungen) abhängen.

Im Bereich der Forschung wäre es wenig vernünftig, jeden neuen Algorithmus zunächst in Bezug auf die Rechenzeit zu optimieren, um dann später festzustellen, daß er nicht die erwarteten Ergebnisse liefert. Eine Optimierung ist zeitaufwendig und lohnt nur für Algorithmen, die dann auch längere Zeit eingesetzt werden. Daher wurden im Testprogramm keine (hoch)optimierten Routinen verwendet. Dies kommt der täglichen Arbeit am Computer am nächsten und ist daher die sinnvollste Grundlage für Geschwindigkeitsvergleiche.

Erstaunlich ist, daß unabhängig vom verwendeten Prozessor das Mainboard einen starken Einfluß auf die Ergebnisse hat (vgl. P90 UMC mit P90 Intel FX).

Wenn bereits feststeht, welcher Algorithmus mit welchen CGH-SBPs Verwendung findet, kann auf die entsprechende Einzelposition der Testprogramm-Ausgabe zurückgegriffen werden. Diese streuen nämlich zum Teil erheblich. Der direkte Pentium-Ersatz von AMD (AMD K5-P90-SSA) erreicht zwar lediglich ein CGHMark von 2.3 (im Vergleich zu 2.9 für den P90), kann aber trotzdem für unsere Anwendungen interessant sein. Während er FFTs bis 128×128 um bis zu 40 % langsamer als der P90 berechnet, ist er für größere Felddimensionen, wie sie bei uns und natürlich ganz allgemein in der optischen Meßtechnik oft vorkommen, deutlich schneller (z.B. 35 % bei 256×256). Dies ist zunächst ein eher unerwartetes Ergebnis, das aber deutlich zeigt, daß man genau untersuchen muß, für welche CGH-Anwendung welcher Rechner geeignet ist.

15.3 Fast-Fourier-Transformation

Die Berechnung der diskreten Fourier-Transformation nach (3.3) bedeutet im eindimensionalen einen Berechnungsaufwand der Ordnung N^2 (im zweidimensionalen Ordnung N^4). Seit 1965 ist ein schneller Algorithmus zur Berechnung von (3.3) bekannt: die *schnelle Fourier-Transformation (FFT)*.

Mit ihr reduziert sich der Aufwand auf die Ordnung $N \ln N$ (bzw. im zweidimensionalen $2N^2 \ln N$). Insbesondere für große N bedeutet das eine erhebliche Beschleunigung. Viele Algorithmen der Signalverarbeitung (und auch die meisten Algorithmen zur Berechnung computergenerierter Hologramme) wären ohne die FFT in der Praxis überhaupt nicht anwendbar.

In diesem Abschnitt wird vorrangig dargestellt, was bei der Auswahl und Benutzung von FFT-Routinen zu beachten ist. Dabei geht es uns insbesondere um eine kurze Darstellung der üblichen Begriffe und Bezeichnungen, um dem Leser die Auswahl eines passenden FFT-Algorithmus zu ermöglichen. Die mathematischen Grundlagen und fortgeschrittene FFT-Algorithmen können hier nicht dargestellt werden.

Das Grundprinzip der FFT ist die fortgesetzte Aufteilung des Problems (die FFT gehört zur Klasse der "Divide-and-Conquer"-Algorithmen). Wenn man statt eine DFT mit N Punkten ausführen zu müssen, zwei DFTs mit je $N/2$ Punkten und eine geeignete Zusammenfassung der beiden Teilergebnisse verwenden könnte, dann wäre der Gesamtaufwand – wegen der quadratischen Abhängigkeit von N – reduziert. Zunächst wird also das Problem in drei Teilprobleme (2 DFTs und 1 Zusammenfassen der Teilergebnisse) zerlegt.

Wir schreiben für die DFT (siehe 3.3)

$$a(f) = \sum_0^N A(n\Delta x)e^{-2\pi if2n\Delta x} \quad (15.1)$$

$$= \sum_0^{N/2} A(2n\Delta x)e^{-2\pi if2n\Delta x} + e^{-2\pi ifN\Delta x} \sum_0^{N/2} A((2n+1)\Delta x)e^{-2\pi if2n\Delta x} \quad (15.2)$$

Beide Summen sind selbst diskrete Fourier-Transformationen mit $N' = N/2$. Jede dieser beiden DFTs kann weiter aufgespalten werden. Die fortgesetzte Teilung des Problems wird solange durchgeführt, bis das Problem trivial ist ($N' = 1$). Dies ist genau dann möglich, wenn N eine Zweierpotenz ist. Damit ergibt sich dann (s. z.B. [stea84] Kap. 6) ein Aufwand proportional $N \ln N$.

Wenn N keine Zweierpotenz ist, werden die Daten nicht in zwei, sondern in eine andere Anzahl von Teil-DFTs zerlegt.

Teilweise werden *Tabellen* zur Berechnung der Sinus- und Cosinus-Terme verwendet. Bei den heutigen superskalaren Rechnerarchitekturen ist es allerdings fraglich, ob der erhöhte Speicherbedarf (und die damit verbundene Gefahr eines Zugriffs außerhalb des Cache-Speichers) und eine eventuell nicht optimale Auslastung der CPU den potentiellen Vorteil in einen Nachteil verwandelt. Um große Tabellen zu vermeiden, können auch *teiltabellierte Funktionen*, aus denen dann in einigen wenigen Iterationen der wirkliche Funktionswert berechnet wird, verwendet werden (siehe z.B. [dobb95]).

Ganz ohne Tabellierung kommt die *rekursive Berechnung* aus. Dabei werden Sinus- und Cosinus-Werte aus vorher bereits benötigten Werten ohne großen Aufwand berechnet. Nachteilig ist allerdings die Kumulation von Rundungsfehlern.

Sollen nur reelle oder rein imaginäre Signale transformiert werden, kann der Aufwand nochmals halbiert werden. Entweder werden gleichzeitig zwei Funktionen transformiert (*Simultan-FFT*), oder – falls nur eine Funktion zu transformieren ist – eine Funktion wird in zwei Hälften aufgeteilt, transformiert und dann “entflechtet” (*Reelle FFT*) (siehe.B. [bach92] Kapitel 8.3).

Man spricht von *Radix- n Algorithmen* (siehe.B. Grant Kap. 12), wenn die fortgesetzte Teilung des Problems in n Teilprobleme erfolgt. Das oben dargestellte Grundprinzip war also ein Radix-2 Algorithmus. Der Vorteil von einem größeren n als 2 ist, daß sehr viele Zugriffe auf den Speicher gespart werden können und damit der Algorithmus an Geschwindigkeit gewinnt. Ein beliebter Sonderfall ist der *Radix-4*-Algorithmus. Er benötigt wie auch der Radix-2 Algorithmus letztlich keine echte Multiplikation als Basisoperation, sondern nur die “Multiplikation” mit ± 1 oder $\pm i$. Eine Implementierung (auch in Hardware) wird damit besonders effizient. Algorithmen mit $n > 4$ werden im Vergleich hierzu verlangsamt.

Eine *Mixed-Radix-Implementation* liegt vor, wenn n im Laufe der Berechnung variiert. Es wird z.B. solange Radix-4 transformiert wie dies möglich ist. Dann wird zu Radix-2 übergegangen. Die restlichen Daten ($N' \neq 2^k$) werden über Radix-3, Radix-5 usw. (Primzahlen) bearbeitet. Mixed-Radix-Implementierungen sind damit insbesondere bei Daten-Größen, die keine 4-er Potenz (oder 2-er Potenz) sind, interessant.

Weiterhin wird zwischen *Frequenz-Dezimierung* und *Zeit-Dezimierung*¹ unterschieden. Das oben dargestellte Grundprinzip war eine Zeit-Dezimierung (oder “Zeitzerlegung”). Die

¹decimation in frequency, decimation in time

Eingangsdaten (“Zeitbereich”) wurden in zwei Hälften aufgeteilt. Genauso kann man den Algorithmus auf einer Zerlegung des Ergebnisses (“Frequenz”) aufbauen (siehe z.B. [stea84] Kap. 6.6).

Wegen der enormen Bedeutung der FFT existieren sehr viele Varianten des Algorithmus. Für eine optimale Implementierung muß der Algorithmus so geändert werden, daß er für eine ganz bestimmte Anwendung (z.B. Symmetrien in den Eingangs- oder Ausgangsdaten) auf einem ganz bestimmten Rechner (Speichergrößen, Cache-Organisation, Prozessor-Organisation usw.) optimal arbeitet. Eine eigene Implementierung ist sehr aufwendig und lohnt sich nur für Spezialanwendungen. Normalerweise sollte einer der frei zugänglichen Algorithmen verwendet werden. Zu beachten ist dabei, daß einige Algorithmen implizit eine Normierung vornehmen (das kann je nach Anwendung nützlich oder aber auch sinnlos sein). Meist ist die Normierung nur in der Hin- oder nur in der Rücktransformation vorhanden, was zu unterschiedlichen Ausführungszeiten von Hin- und Rücktransformation führt.

Weiterhin sollte man auf die für die Anwendung nötige Rechengenauigkeit achten. Obwohl viele Rechner in doppelter Genauigkeit genauso schnell (oder sogar schneller) arbeiten, ist aufgrund des auch doppelt so großen Speicherbedarfs und der damit verbundenen Probleme (Cache !) eventuell eine erhebliche Beschleunigung bei Beschränkung auf einfache Genauigkeit möglich. Eine Ersetzung der Fließkommaarithmetik durch Integer- oder Festkommaarithmetik ist bei modernen Prozessoren (Superskalarität) selten sinnvoll.

15.3.1 Beschleunigung der FFT

Die Verwendung der FFT beschleunigt die Berechnung der diskreten Fourier-Transformation ganz erheblich. In der Regel wird die FFT über einen general-purpose-Computer berechnet. Weitere Beschleunigungen sind aber über den Einsatz von spezieller optischer und elektronischer Hardware möglich.

Digitale Signal- und Vektorprozessoren

Eine drastische Beschleunigung kann durch den Einsatz spezieller FFT-Hardware erreicht werden (siehe [gran89] Kap. 12 für eine Einführung). Im einfachsten Fall werden dazu digitale Signalprozessoren (DSPs) eingesetzt. Die meisten DSPs haben spezielle Instruktionen, die bei der Berechnung von FFTs wichtig sind. Moderne CPUs holen allerdings gegenüber diesen DSPs immer mehr auf. Als Beispiel für einen modernen DSP sei der CAP von Motorola angeführt (momentan noch in der Entwicklung), für den eine komplexe (20 Bit) 1024-1D-FFT mit 131 Mikrosekunden angegeben wird. Ein weiterer (und momentan sehr beliebter) DSP ist der 320C40 von Texas Instruments ([ti96]) (50 MFLOPs skalar). Er liefert dasselbe Ergebnis in 1550 Mikrosekunden.

Eine weitere Beschleunigung kann durch den Einsatz von allgemeinen Vektorprozessoren oder speziellen FFT-Prozessoren (FFTIC) erreicht werden. Von der Firma Catalina Research Inc, Colorado² liegt uns ein Angebot über eine PC-Steckkarte (ISA-BUS) mit einem LH9124 Vektorprozessor (Sharp) vor (500 MOPS vektoruell).

Die Karte berechnet eine zweidimensionale (24 Bit) 512x512 FFT in knapp 40 ms (inklusive

²Catalina Research Inc., 1321 Aeroplaza Dr., Colorado Springs, Colorado, USA

Transfer zwischen Karte und Hostcomputer) bzw. eine eindimensionale 1024 FFT in 80 μ s. Das entspricht in etwa einem Geschwindigkeitsvorteil gegenüber unserem Pentium 90 System vom Faktor 100.

Durch den Einsatz einer solchen Karte könnten zeitintensive Hologrammberechnungen (insbesondere genetische Algorithmen und Direkte Suchmethoden) auch für größere LCDs durchgeführt werden. Die Rechenzeit würde von Wochen auf Stunden schrumpfen.

Nachteil dieser Karte ist der (noch) recht hohe Preis von ca. 25.000 US-Dollar. Eine Karte mit ähnlichen Leistungsmerkmalen wird von Sundance (SMT 311) zum Preis von 10.000 Brit. Pfund angeboten³. Beide Karten verwenden denselben FFTIC LH9124.

Für Spezialanwendungen und vor allem im Hinblick auf den üblicherweise extremen Preisverfall sollte man den möglichen Einsatz solcher Karten aber in jedem Fall im Auge behalten.

Andere mögliche Hochleistungsprozessoren, die für einen Einsatz in Frage kommen, sind der DSP BDS0124 (1D 1024 komplex: 64 Mikrosek.) und der Vektorprozessor TM-66 swiFFT: (680 MFLOPS vektoruell, 1D 1024 komplex: 80 Mikrosek.).

Eine weitere deutliche Geschwindigkeitssteigerung kann nur durch Parallelisierung⁴ oder eine optische Generierung der Fourier-Transformation erfolgen.

Analog-optische Berechnung

Problematisch an der rein optischen FT-Generierung für computergenerierte Hologramme sind mehrere Punkte. Zunächst muß bedacht werden, daß die FT-Generierung selbst zwar sehr schnell ist, der optische Prozessor aber mit den Eingangsdaten (z.B. über ein LCD) versorgt und das Ergebnis ebenfalls wieder digitalisiert (z.B. CCD) werden muß. Hier entsteht ein Flaschenhals, der eine deutliche Geschwindigkeitssteigerung gegenüber elektronischen FFTICs nur mit relativ großem technischen und finanziellen Aufwand ermöglicht. Um dieses Problem zu umgehen, muß nach Möglichkeit das Digitalisieren vermieden werden und der Algorithmus zur Hologrammberechnung weitestgehend optisch realisiert werden. Das ist sicher nur für ausgewählte Algorithmen sinnvoll bzw. möglich.

Die optische Rekonstruktion hat einen gewaltigen Vorteil: Fehler in der Rekonstruktion müssen nicht erst digital simuliert werden (was unter Umständen gar nicht möglich ist, weil der Fehlerursprung nicht bekannt ist) sondern werden automatisch wegoptimiert. Solche Fehler können z.B. defekte Pixel, großflächige LCD-Fehler oder eine mangelhafte Justierung sein.

Welche Algorithmen eignen sich nun zum Einsatz optischer Fourier-Transformationen? Im Abschnitt 8.1.2 wurden bereits zwei mögliche optische Implementierungen der IFT-Algorithmen besprochen. Für sehr einfache, nicht-iterative Algorithmen lohnt sich der Aufwand der optischen Lösung nicht. Wegen des enormen Rechenaufwands sind die direkten Suchverfahren die interessantesten Algorithmen bezüglich einer optischen Realisation. Insbesondere genetische Algorithmen könnten unter Umständen stark hiervon profitieren.

Die DBS-Algorithmen sind meiner Ansicht nach keine geeigneten Kandidaten. Beim DBS

³<http://www.sundance.com/s311.htm>

⁴insbesondere die oben angesprochenen LH9124 lassen sich kaskadieren wobei sich die Leistung bei FFT-Berechnungen annähernd linear mit der Prozessorzahl erhöht

Verfahren wurde jeweils 1 Hologrammpixel geändert, die Rekonstruktion und der mittlere Fehler berechnet. Bei den uns interessierenden Displays mit bis zu 10^6 Pixeln ändert ein einzelner Pixel statistisch auch nur 10^{-6} der Intensität der Rekonstruktion. Die Verbesserung oder Verschlechterung kann also nicht mehr mittels üblicher Kameras gemessen werden. Die Genauigkeit des optischen Fourier-Prozessors ist für diesen Algorithmus bei weitem nicht ausreichend.

Weible und Herzig (siehe [weib93]) haben versucht, das Problem durch eine leichte Abwandlung des DBS-Algorithmus zu vermindern. Dazu werden zunächst jeweils N Pixel zu einem "Superpixel" zusammengefaßt. Dieser Pixel muß so groß gewählt werden, daß ein Umklappen dieses Superpixels zu einer meßbaren Änderung der Rekonstruktionsqualität führt. Nach der Optimierung der Superpixel werden dann die Superpixel einzeln weiter optimiert. Die Verwendung der Superpixel dient also lediglich zum Herstellen einer günstigen Startkonfiguration.

Wieso können nun die Einzelpixel trotzdem noch weiter optimiert werden, obwohl die Genauigkeit des Fourier-Prozessors – wie oben beschrieben – dafür nicht ausreichend ist? In Wirklichkeit wird nicht der einfache DBS-Algorithmus, sondern eine Form des Simulated-Annealing Algorithmus ausgeführt. Es werden auch Pixeländerungen zugelassen, die nicht eine direkte Verbesserung der Bildqualität zur Folge haben. Nur in einem rein statistischen Sinn werden die Pixeländerungen, die zu einer Verbesserung führen, bevorzugt.

Allerdings ist die Wahrscheinlichkeit, daß auch inkorrektes Umklappen nicht verworfen wird, konstant und durch die Anzahl der LCD-Pixel und die Meßanordnung der Rekonstruktion gegeben. Diese Wahrscheinlichkeit kann nicht beliebig verringert werden. Dies steht im Gegensatz zum originalen SA-Algorithmus bei dem diese Wahrscheinlichkeit (die "Temperatur") zeitlich abnimmt. Dementsprechend ist mit einer schlechteren Konvergenz zu rechnen. Wieviel schlechter, kann im allgemeinen nicht vorhergesagt werden (siehe Diskussion zum SA-Algorithmus in Abschnitt 10.2), hängt aber im wesentlichen von der Anzahl der LCD-Pixel ab.

Das Zusammenfassen einzelner Pixel zu Superpixeln kann in diesem Sinn als eine zusätzliche künstliche Temperaturänderung im SA-Algorithmus interpretiert werden. Ob allerdings wirklich dieselbe Qualität erreicht wird, ist fraglich. Aus diesem Grund verwenden Weible und Herzig die so gewonnenen Hologramme als Ausgangsmaterial für eine weitere, rein digitale Optimierung nach dem Gradientenverfahren ("steepest-descent", "downhill"). Bei einem solchen Vorgehen fällt allerdings der Vorteil weg, optische Fehler der Anordnung (z.B. defekte Pixel des LCDs, Dejustierungen usw.) über die Optimierung eliminieren zu können.

Überhaupt ist es sehr fraglich, ob für die von uns angepeilten hochauflösenden LCDs (Weible und Herzig experimentierten mit 32×32 Pixeln) das Verfahren überhaupt eine Beschleunigung bringen kann. Die an sich schon recht schlechte Konvergenz des SA-Algorithmus wird zusätzlich weiter verschlechtert und die Anzahl der nötigen Iterationen erhöht sich damit enorm.

Für DBS- und SA-Algorithmen in Verbindung mit hochauflösenden LCDs ist meiner Ansicht nach deshalb nur eine elektronische Beschleunigung sinnvoll.

Hybride Systeme

Die analog-optische Berechnung leidet — wie oben beschrieben — für viele Anwendungen im Bereich der CGH-Berechnung unter dem Problem der begrenzten Auflösung. Rein digitale, optische Computer umgehen dieses Problem, sind momentan aber für Fouriertransformationen der geforderten Arraydimensionen noch nicht technisch realisierbar (siehe z.B. [kane96]).

Die Probleme von rein digitalen optischen Systemen lassen sich durch den Einsatz von Smart-Pixels⁵ vermindern. Ein Teil der informationsverarbeitenden Funktion des Gesamtsystems wird durch mikroelektronische Elemente erledigt. Dabei können Taktraten von bis 100 MHz erreicht werden. Man nimmt die Geschwindigkeitseinbußen gegenüber einer rein optischen Verarbeitung in Kauf, erhält sich aber gegenüber der rein elektronischen Verarbeitung einige Vorteile. Insbesondere können beliebige optische Verbindungen realisiert werden. Diese Verbindungen sind extrem schnell, zeigen kein Übersprechen, sind unempfindlich gegen externe Störungen und führen zu einer elektrischen Isolation der einzelnen Systemteile (siehe[lent93]).

Ozaktas und Miller haben ein interessantes digital-optisches System vorgeschlagen, das auf dieser Idee aufbaut (siehe [ozak96]).

Der Grundaufbau für ein dreistufiges System (beliebige Anzahl von Stufen sind denkbar) ist in Abbildung 15.1 dargestellt.

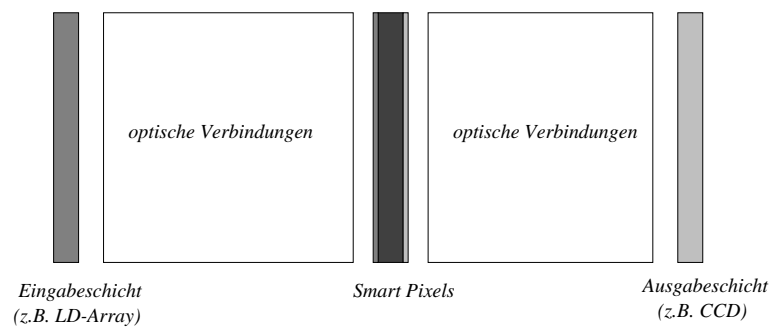


Abbildung 15.1: Digital-optischer Computer (2 Stufen)

Zwischen den Eingangs-, Ausgangs- und Zwischenebenen werden durch optische Elemente Verbindungen hergestellt. Die optischen Verbindungen können — je nach — Anwendung durch einfache optische Elemente wie Linsen oder — für komplizierte Verbindungen — über CGHs realisiert werden.

Unabhängig vom Einsatz eines solchen optischen Prozessors zur CGH-Berechnung ergibt sich eine sehr interessante Anwendungsmöglichkeit für unsere LCD-CGHs: wenn wir die optischen Verbindungen zwischen den Ebenen mittels LCD-CGHs realisieren, erhalten wir einen sehr flexiblen digital-optischen Computer. Bisherige optische Computer sind weit vom elektronischen General-Purpose Computer, wie er heute üblicherweise verwendet wird, entfernt. Die optischen Systeme werden für eine und nur eine ganz spezielle Aufgabe gebaut. Ein nachträgliches Abändern des Computers hin zu einer neuen Anwendung ist schwierig bis unmöglich.

Durch den Einsatz von LCDs-CGHs kann dieser Umbau aber auf einfachste Weise und

⁵zweidimensionale Arrays aus kleinen Elementarprozessoren mit optischem Input und Output

innerhalb von Sekundenbruchteilen erfolgen. Ein herkömmlicher Personal Computer steuert die Eingabe und die LCD-CGHs und erfaßt die Ausgaben des optischen Computers. Die “Programmierung” dieses optischen Computers würde dementsprechend in der Festlegung der Verbindungen zwischen den Schichten liegen (der PC errechnet passende CGHs für diese Verbindungen und gibt sie auf die LCDs).

Ein Problem der vorgeschlagenen Anordnung ist allerdings, daß — aus praktischen Gründen — Inkohärenz zwischen den einzelnen Smart-Pixeln anzunehmen bzw. zu fordern ist. Dementsprechend muß entweder für jedes Smart-Pixel ein eigenes, kleines Hologramm zur Verfügung stehen oder man muß ein Hologramm so optimieren daß es für alle Pixel die jeweils gewünschten Beugungsbilder ergibt (das erfordert eine Abwandlung der üblichen Hologrammberechnungsalgorithmen, die aber grundsätzlich möglich sein sollte).

Die Auflösung von LCDs begrenzt die mögliche Anzahl der Smart-Pixel. Auf welchen Wert kann nicht auf einfache Weise abgeschätzt werden, sondern erfordert eine detailliertere Untersuchung. Die Smart-Pixels müssen Laserdioden (inkohärentes Laserdioden-Array) verwenden (siehe z.B. [chen93] als Beispiele von Smart-Pixels in Verbindung mit oberflächenemittierenden Laserdioden mit Vertikalresonatoren⁶).

Nach diesem kleinen Ausflug in eine mögliche neue Anwendung von CGHs komme ich zurück auf die einfache Anordnung (ohne CGHs) zur Berechnung der Fourier-Transformation. Eine genaue Beschreibung der Implementierung eines 16x16 Butterflies findet man bei [ozak96]. Die Abschätzung der Geschwindigkeit des Systems zeigt allerdings, daß gegenüber den rein elektronischen Spezialprozessoren (s.o.) kaum Vorteile erzielt werden können. Da der Aufwand (Smart-Pixels) in jedem Fall den der rein elektronischen Lösung übersteigt, scheint mir auch dieses Verfahren zur FFT Berechnung (zumindest beim gegenwärtigen Stand der Technik) nicht sinnvoll.

⁶= “Vertical cavity surface emitting lasers (VCSELs)”

Kapitel 16

Beurteilung der verschiedenen Verfahren

Bereits im Abschnitt 5.3 wurde ausführlich dargelegt, daß es den “besten” Algorithmus zur Hologrammberechnung nicht gibt und daß kaum allgemeingültige Aussagen getroffen werden können. Trotzdem will ich hier versuchen wenigstens einige kleine Hilfen zur schnellen Auswahl eines geeigneten Algorithmus zu geben.

Am Ende dieses Kapitels werde ich Beugungseffizienzen und Rekonstruktionsfehler für einige ausgewählte Verfahren tabellarisch angeben. Dabei ist zu beachten, daß diese Ergebnisse lediglich den in Abschnitt 5.3.3 beschriebenen Versuchsbedingungen entsprechen *und nicht ohne weiteres auf andere Bedingungen und Testbilder übertragen werden können*. Weiterhin sind verschiedene Parameter, die die Ergebnisse entscheidend beeinflussen können, bei den Algorithmen änderbar.

Für die Praxis entscheidender sind dagegen die qualitativen Entscheidungshilfen, die ich im folgenden geben will. Dazu teile ich die Algorithmen entsprechend ihrer Geschwindigkeit in drei Grundklassen ein:

- *nicht-iterative Verfahren*: einfache Diffusoren, Analytische Berechnungen, Raytracing, Error-Diffusion, einfache Quantisierungen
- *iterative Verfahren mit geringer Iterationsanzahl*: IFT-Algorithmen (Gerchberg-Saxton, POCS), Hopfield Netze, Input-Output-Algorithmus
- *iterative Verfahren mit erhöhter Iterationsanzahl*: Direct-Binary-Search, Simulated-Annealing, Threshold-Accepting, genetische Algorithmen

Die nicht-iterativen Verfahren bieten sich dann an, wenn eine sehr schnelle Berechnung der Hologramme (z.B. Display-Anwendungen) nötig ist. Die erreichbare Qualität ist in der Regel deutlich schlechter als mit iterativen Verfahren. Da wegen des geringen Orts-Bandbreiten-Produkts unserer Hologrammedien die Qualität der Rekonstruktionen bereits deutlich eingeschränkt ist, scheinen die nicht-iterativen Verfahren zum alleinigen Einsatz kaum geeignet. Unter Umständen sind sie in Kombination mit iterativen Verfahren einsetzbar (z.B. Error-Diffusion in Verbindung mit IFT-Algorithmen).

Die *iterativen Verfahren mit geringer Iterationsanzahl* sind momentan bei einer Berücksichtigung von Rechenzeit und Rekonstruktionsqualität die günstigsten Verfahren.

Verfahren	Rekonstruktion	Rechenaufwand	Vielseitigkeit	Stabilität	Speicherbedarf	nur binär
Einfache Diffusoren	–	+	–	++	+	Nein
IFT	+	0	0	++	0	Nein
POCS (komplex)	+	0	0	+	0	Nein
GIC	+	0	–	–	0	Nein
Input-Output Alg.	+	0/+	–	–	0	Nein
DBS (Intens.)	++	--	++	+	+	Ja
DBS (Ampl.)	++	–	0	+	+	Ja
SA, TA	++	---	++	0	+	Ja
Genetische Alg.	++	---	+	0	+	Nein
Neuronale Netze	+	0	0	0	0	Nein
Analytische Berechnung	–	++	–	++	++	Nein
Superpositionen	–	++	–	++	--	Nein
Raytracing	–	+/-	–	++	0	Nein

Abbildung 16.1: Beurteilung verschiedener Berechnungsalgorithmen für Phasenhologramme

Die erreichbare Qualität ist sehr gut. Für viele Anwendungen scheinen diese Verfahren die geeignete Wahl zu sein.

Iterative Verfahren mit großer Iterationsanzahl, also Algorithmen der direkten Suche, verbessern zwar die Rekonstruktionsqualität minimal, erfordern aber einen um Größenordnungen höheren Berechnungsaufwand. Ihr Einsatz kommt momentan nur bei CGHs in Frage, bei denen die Berechnungsdauer nur eine untergeordnete Rolle spielt und eine optimale Rekonstruktionsqualität erreicht werden muß (z.B. Strahlformung für Hochleistungslaser). Ein weiteres Einsatzgebiet ist die Rekonstruktion komplexer Lichtamplituden, insbesondere bei einfachen Objekten aus möglichst wenig Punkten.

In Abbildung 16.1 wird der Versuch unternommen, die unterschiedlichen Verfahren zur Berechnung von Phasenhologrammen nach ihren Vor- und Nachteilen zu beurteilen. Aus den oben genannten Gründen ist die Aufstellung lediglich als qualitatives Hilfsmittel zu sehen. Bei anderen Autoren würde eine entsprechende Aufstellung teilweise sicher andere Ergebnisse liefern.

Mit “Vielseitigkeit” bezeichne ich die Variations- und Kombinationsmöglichkeiten, die der entsprechende Algorithmus bietet. “Rekonstruktion” bezeichnet die Rekonstruktionsqualität und beinhaltet sowohl die Beugungseffizienz als auch den Rekonstruktionsfehler. Der Parameter “Stabilität” dient zur Beurteilung, inwieweit der Benutzer den Algorithmus steuern muß. Ideal sind dabei Algorithmen, die für beliebige Objekte eine stabile Konvergenz ohne äußere Eingriffe erreichen.

Im Zusammenhang mit der Beurteilung der Rekonstruktionsfehler muß darauf hingewiesen werden, daß für viele Anwendungen die Speckle-Effekte zu deutlichen Störungen führen und den Hauptfehler darstellen. In diesem Fall muß die Reduktion der Speckles gegenüber der Optimierung der Grundqualität, auf die ich mich in der Abbildung beziehe, im Vordergrund stehen.

Die verschiedenen Quantisierungsverfahren (insbesondere die Error Diffusion) sind mit verschiedenen der oben angeführten Grundalgorithmen kombinierbar. Die Bewertung dieser Quantisierungen hängt stark von dem entsprechenden Grundverfahren ab. Allgemeine An-

gaben sind hier wenig hilfreich.

In Abbildung 16 sind beispielhaft die Beugungseffizienzen und Rekonstruktionsfehler für einige unterschiedliche Hologramme angegeben. Es handelt sich dabei jeweils um Phasenhologramme mit QS Quantisierungsstufen und einer Größe von 128×128 Pixeln. Die Fehlerdefinitionen entsprechen den in Abschnitt 5.3 getroffenen Vereinbarungen.

Die Ergebnisse sind in sich schlüssig, sollen aber vor allem zeigen, wie stark die Ergebnisse bei unterschiedlichen Versuchsbedingungen (verwendetes Objekt, Anzahl Quantisierungsstufen) variieren.

Verfahren	Testbild	QS	$e \cdot 10^{-4}$	η [%]
Direkte Berechnung, konstante Phase	ITO	256	8.63	51
Direkte Berechnung, zufällige Phase	ITO	256	5.00	80
Direkte Berechnung, zufällige Phase	ITO	2	9.05	33
Direkte Berechnung, zufällige Phase	ITO	3	7.44	55
Direkte Berechnung, zufällige Phase	ITO	4	6.54	65
Direkte Berechnung, zufällige Phase	ITO	6	6.28	73
Gerchberg-Saxton, 5 Iterationen	ITO	256	4.49	84
Gerchberg-Saxton, 10 Iterationen	ITO	256	4.27	85
Gerchberg-Saxton, 50 Iterationen	ITO	256	3.95	87
Gerchberg-Saxton, 100 Iterationen	ITO	256	3.86	88
Gerchberg-Saxton, 500 Iterationen	ITO	256	3.79	89
Gerchberg-Saxton, 5 Iterationen	ITO	4	6.23	68
Gerchberg-Saxton, 10 Iterationen	ITO	4	6.14	69
Gerchberg-Saxton, 50 Iterationen	ITO	4	5.92	71
Gerchberg-Saxton, 100 Iterationen	ITO	4	5.90	71
Gerchberg-Saxton, 500 Iterationen	ITO	4	5.85	72
Direkte Berechnung, konstante Phase	TEXT	256	59.00	52
Direkte Berechnung, zufällige Phase	TEXT	256	24.30	95
Gerchberg-Saxton, 5 Iterationen	TEXT	256	11.60	99
Gerchberg-Saxton, 10 Iterationen	TEXT	256	9.02	99
Gerchberg-Saxton, 50 Iterationen	TEXT	256	5.01	100
Gerchberg-Saxton, 100 Iterationen	TEXT	256	4.15	100
Gerchberg-Saxton, 5 Iterationen	TEXT	4	24.90	95
Gerchberg-Saxton, 10 Iterationen	TEXT	4	24.20	95
Gerchberg-Saxton, 50 Iterationen	TEXT	4	23.53	95
Gerchberg-Saxton, 100 Iterationen	TEXT	4	23.30	95
Direkte Berechnung, konstante Phase	GABOR	256	24.80	100
Direkte Berechnung, zufällige Phase	GABOR	256	10.80	100
Gerchberg-Saxton, 5 Iterationen	GABOR	256	5.25	100
Gerchberg-Saxton, 10 Iterationen	GABOR	256	4.00	100
Gerchberg-Saxton, 50 Iterationen	GABOR	256	2.05	100
Gerchberg-Saxton, 100 Iterationen	GABOR	256	1.43	100
Gerchberg-Saxton, 500 Iterationen	GABOR	256	0.73	100

Abbildung 16.2: Beispiele verschiedener Beugungseffizienzen und Rekonstruktionsfehler

Kapitel 17

Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Überblick vorhandener Verfahren zur Berechnung computergenerierter Hologramme gegeben. Dabei wurden insbesondere Algorithmen untersucht, die für einen Einsatz unter Verwendung von LCDs als Hologrammedien in Frage kommen. Insbesondere die Iterativen Fourier-Transformations-Algorithmen, die trotz guter Qualität nur vergleichsweise moderate Rechenzeiten benötigten, versprechen momentan für die von uns angepeilten Anwendungen den größten Erfolg und wurden ausführlich dargestellt. Der Grundalgorithmus dieser Klasse von Algorithmen, der Gerchberg-Saxton-Algorithmus, wurde auf drei Dimensionen verallgemeinert. Die direkten Suchverfahren lassen sich direkt auf drei Dimensionen verallgemeinern.

Es wurde ein neuer Ansatz zur Vergrößerung des maximalen Phasenschiebebereichs von LCDs mittels der Nutzung zusätzlicher polarisierender Elemente über geometrische Phasen vorgeschlagen und in der Simulation verifiziert.

Praktische Probleme bei der Verwendung der LCDs wurden zusammen mit Schweikhardt untersucht. Die gewonnenen Ergebnisse erlaubten uns, die für den Einsatz der LCDs als Hologrammedien wesentlichen Eigenschaften genauer angeben zu können und führten zur Kaufempfehlung eines VGA-LCDs.

Das VGA-LCD stand als Leihgabe zur Verfügung, wurde vermessen und zur Wiedergabe von Phasenhologrammen eingesetzt. Die positiven Ergebnisse der Hologrammrekonstruktionen bestätigten unsere Erwartungen.

Computergenerierte Hologramme sind ein sehr junger Zweig der Optik. Die Bedeutung dieses Zweiges kann man an den zu Beginn dieser Arbeit genannten Anwendungen und an der immensen weltweiten Forschungsaktivität abschätzen. Die Aktivität zeigt sich unter anderem in der hier dargestellten Vielfalt unterschiedlichster Verfahren zur Berechnung der Hologramme.

Neben technologischen Verbesserungen im Bereich der Lichtmodulatoren und der Erschließung neuer Anwendungsgebiete sind auch bei den Berechnungsalgorithmen — trotz der bereits heute nur noch schwer überschaubaren Anzahl an möglichen Verfahren — noch viele Verbesserungen möglich.

Abschließend will ich daher die aus meiner Sicht erfolgversprechendsten Ansätze zur weiteren Verbesserung der *Berechnung* computergenerierter Hologramme nochmals zusammenfassen:

- Einsatz spezieller Hardware zur FFT-Berechnung.
- Beschleunigung genetischer Algorithmen
- Untersuchung neuerer mathematischer Optimierungsalgorithmen hinsichtlich eines Einsatzes zur CGH-Berechnung
- Untersuchung der Speckle-Reduktion innerhalb der IFT-Algorithmen
- Untersuchung des Iterative-Multiresolution-Algorithmus (siehe Abschnitt 8.1.1) zur CGH-Berechnung

Anhang A

Grundlagen der Beugungstheorie

Die Ausführungen dieses Abschnittes sind als Erinnerungen für den physikalisch bereits geschulten Leser gedacht. Es können und sollen hier weder tiefere Grundlagen der Elektrodynamik noch der Optik im speziellen vermittelt werden. Für eine ausführlichere Besprechung der hier ausgeführten Schritte sei auf die Standardlehrbücher der theoretischen Optik und der Elektrodynamik verwiesen (z.B. [born80], [ditc61], [fowl75], [jack83]) hingewiesen.

Ausgehend von den Maxwell-Gleichungen werde ich die Fresnel-Kirchhoff Beugungsformel ableiten. Dabei wird die skalare Wellengleichung und damit dann die skalare Beugungstheorie abgeleitet. Für sehr kleine Beugungsstrukturen, ist dieser Ansatz nicht mehr sinnvoll. Peter et al. ([pete93]) geben als praktische Faustformel eine minimale Strukturweite $\Delta x > 10\lambda$ an, bei der die skalare Theorie noch ohne Bedenken verwendet werden kann. Für LCDs und sichtbares Licht ist diese Bedingung in jedem Fall erfüllt.

Die kovariante Formulierung¹ hat den Vorteil einer einfacheren und übersichtlicheren Rechnung bedingt aber zwangsweise elementare Kenntnisse der Tensormathematik und soll dem Leser hier nicht zugemutet werden.

A.1 Wellengleichung

Wie gehen von den *Maxwell-Gleichungen* aus:

Inhomogene Feldgleichungen

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{j} \quad (\text{A.1})$$

$$\nabla \cdot \vec{E} = \rho \quad (\text{A.2})$$

Homogene Feldgleichungen

$$\nabla \cdot \vec{B} = 0 \quad (\text{A.3})$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (\text{A.4})$$

Bildet man die Rotation von (A.4) ergibt sich wegen $\text{rot rot} = -\Delta + \text{grad div}$ (läßt sich leicht mittels Tensoranalysis nachprüfen) bei Abwesenheit von frei beweglichen Ladungen ($\rho = \vec{j} = 0$)

¹siehe z.B. Schröder: "Spezielle Relativitätstheorie", Verlag Harri Deutsch (1987)

$$\nabla \times \nabla \times E = -\Delta E + \nabla(\nabla \vec{E}) \quad (\text{A.5})$$

Wegen (A.1) ergibt sich dann mit $\vec{D} = \epsilon_0 \epsilon_r \vec{E}$ und $\vec{B} = \mu_0 \mu \vec{H}$

$$\nabla \times \nabla \times \vec{E} = -\frac{\mu \epsilon}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} \quad (\text{A.6})$$

Also wegen (A.2)

$$\Delta \vec{E} - \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} := \square \vec{E} = 0 \quad (\text{A.7})$$

Analog leitet man leicht $\square \vec{H} = 0$ ab.

A.2 Kirchhoff–Integral–Theorem

Mit (A.7) wird nun ein Zusammenhang zwischen dem elektromagnetischen Feld in einem beliebigen Punkt innerhalb einer geschlossenen Oberfläche und dem Feld auf der Oberfläche abgeleitet.

Dabei wird von der vektoriellen Beschreibung des elektromagnetischen Feldes zu einer skalaren Theorie übergegangen. Dazu denke man sich ohne Einschränkung der Allgemeinheit² eine elektromagnetische Welle, die in Richtung der z-Achse propagiert. Bekanntlich sind elektromagnetische Wellen transversale Wellen. Das heißt sowohl das E- als auch das B-Feld haben keine Komponente in z-Richtung. Weiterhin stehen E- und B-Feld senkrecht aufeinander. Das B-Feld kann vollständig aus dem E-Feld berechnet werden (ebenso umgekehrt). Es genügt also, wenn wir uns auf ein Feld (wir wählen das E-Feld) beschränken. Für eine vollständige Theorie muß dementsprechend E_x und E_y in Betracht gezogen werden. Bei unpolarisiertem oder vollständig linear polarisiertem Licht ist das aber nicht nötig. Es genügt, eine Komponente, z.B. E_x für weitere Betrachtungen auszuwählen³. Man spricht von einer *skalaren Theorie*, da eben lediglich noch ein Skalar (z.B. E_x) zur vollständigen Beschreibung nötig ist. Aus dem skalaren Feld lassen sich dann alle anderen interessierenden Größen ermitteln.

Wir bezeichnen das skalare Feld E_x mit U und verwenden V als eine Art Hilfsfeld. Zu beachten ist, daß wir kein beliebiges Hilfsfeld V benutzen dürfen. V muß die Wellengleichung $\square V = 0$ erfüllen. Wir wählen eine einfache Kugelwelle für V :

$$V := V_0 \frac{e^{i(\vec{k}\vec{r}-wt)}}{r} \quad (\text{A.8})$$

Zusammen mit dem Satz von Green

$$\int_{\partial V} (V \text{grad}_n U - U \text{grad}_n V) dA = \int_V (V \Delta U - U \Delta V) dV \quad (\text{A.9})$$

²es muß ja lediglich das Koordinatensystem entsprechend gelegt werden

³Man denke sich am einfachsten linear polarisiertes Licht in x-Richtung (d.h. $E_y(\vec{r}, t) := 0$). Wenn nun keine polarisierende Elemente in der optischen Anordnung Verwendung finden bleibt wirklich überall $E_y = 0$ und muß nicht weiter beachtet werden.

der für alle skalaren Funktionen U und V gilt, die die üblichen Integrabilitätsbedingungen erfüllen, kann man mit (A.7) leicht nachweisen, daß

$$\int_{\partial V} \left[\frac{e^{i\vec{k}\vec{r}}}{r} \text{grad}_n U - U \text{grad}_n \frac{e^{i\vec{k}\vec{r}}}{r} \right] dA = 0 \quad (\text{A.10})$$

gilt, wenn U und V eine Zeitabhängigkeit der Form $e^{i\omega t}$ haben. In (A.9) und (A.10) haben wir mit n die Flächennormale zur Integrationsoberfläche bezeichnet. Der konstante Faktor $V_0 e^{i\omega t}$ läßt sich kürzen und fällt damit weg.

Um nun zu dem Wert von U an einer beliebigen Stelle im Integrationsgebiet zu gelangen, ist die eben getroffene Definition von V von Bedeutung. Wir sehen unmittelbar anhand der letzten Gleichung, daß für r gegen Null Divergenzen auftreten. Dieses Problem wird wie üblich durch Aussparung eines kleinen kreisförmigen Gebietes um $r = 0$ bei der Volumenintegration (man bedenke, daß das hier betrachtete Oberflächenintegral nach dem Satz von Gauß ja auch als Volumenintegral über das eingeschlossene Gebiet betrachtet werden kann) berücksichtigt. Das Integrationsgebiet ist in Abbildung A.1 veranschaulicht.

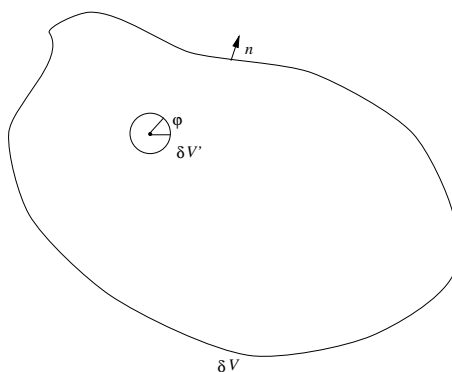


Abbildung A.1: Integrationsgebiet

Dementsprechend ist dann aber auch bei der Berechnung des Oberflächenintegrals, die gesamte Oberfläche zu berücksichtigen. Es wird also ein Term für die innere Oberfläche abgezogen:

$$\begin{aligned} & \int_{\partial V} \left[\frac{e^{i\vec{k}\vec{r}}}{r} \text{grad}_n U - U \text{grad}_n \frac{e^{i\vec{k}\vec{r}}}{r} \right] dA \\ & - \int_{\partial V'} \left[\frac{e^{i\vec{k}\vec{r}}}{r} \frac{\partial U}{\partial r} - U \frac{\partial}{\partial r} \frac{e^{i\vec{k}\vec{r}}}{r} r^2 \right] d\Omega = 0 \end{aligned} \quad (\text{A.11})$$

Dabei haben wir $d\Omega = 4\pi d\varphi$ verwendet.

Wir betrachten nun für den Grenzfall $r \rightarrow 0$ das zweite Integral in (A.11). Führt man die Ableitungen aus, dann ergibt sich sofort für dieses Integral

$$\int_{\partial V'} U(0) d\Omega = 4\pi U(0) \quad (\text{A.12})$$

Dies eingesetzt in (A.11) ergibt das *Kirchhoff-Integral-Theorem*

$$U = -\frac{1}{4\pi} \int_{\partial V} \left[U \text{grad}_n \frac{e^{i\vec{k}\vec{r}}}{r} - \frac{e^{i\vec{k}\vec{r}}}{r} \text{grad}_n U \right] dA \quad (\text{A.13})$$

mit dessen Hilfe das Feld im Innern einer – realen oder rein imaginären – Oberfläche aufgrund der Feldstärke auf der Oberfläche berechnet werden kann.

A.3 Fresnel–Kirchhoff-Formel

Das im letzten Abschnitt abgeleitete Kirchhoff–Integral–Theorem wird nun auf ein möglichst allgemeines Beugungsproblem angewandt. Es wird sich dann eine Verallgemeinerung des bekannten Huygensschen Prinzips ergeben.

Dabei gehen wir von einer Punktlichtquelle in L aus. Diese wirft Licht auf eine beliebige Apertur A . Die Apertur beugt Licht nach P . Wir interessieren uns für die Lichtamplitude in P .

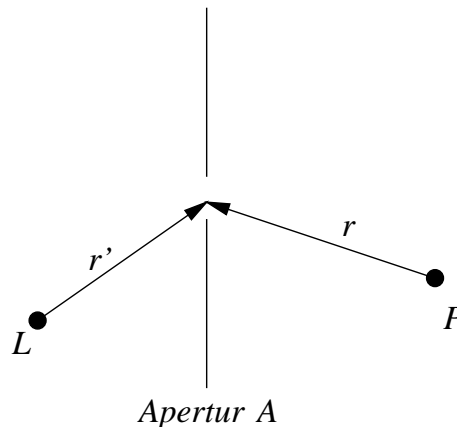


Abbildung A.2: Zur Fresnel–Kirchhoff–Formel

Der Abstand zwischen der Punktlichtquelle L und einem Punkt auf der Apertur sei r' . Die Amplitude in diesem Punkt r' ist – da es sich bei L um eine Punktlichtquelle handelt – durch die Kugelwellenamplitude

$$U = U_0 \frac{e^{i(kr' - \omega t)}}{r'} \quad (\text{A.14})$$

gegeben.

Nun kann das Kirchhoff–Integral–Theorem (A.13) angewendet werden. Um die Amplitude in P zu bekommen müssen wir das Integral in (A.13) lediglich über die Fläche der Apertur ausführen da der Rest der gedachten Oberfläche die Amplitude 0 aufweist (man denke sich einen Integrationsweg im Unendlichen).

Dabei haben wir allerdings implizit die Annahme verwendet, daß die Lichtamplitude bei der Apertur unabhängig davon ist, ob die Apertur vorhanden ist oder nicht. Diese Annahme wird als *St. Venants Hypothese* (siehe [ditc61] 6.13) bezeichnet und kann lediglich im Sinne einer Näherung korrekt sein. Eine vollständig korrekte Berechnung müßte mittels der Maxwell–Gleichungen und den durch die Apertur aufgeprägten Randbedingungen erfolgen (und zwar für jedes konkrete Beugungsproblem erneut). Dieser Aufwand ist in der Praxis nicht zu leisten.

Es zeigt sich aber experimentell, daß im allgemeinen St.Venants Hypothese Ergebnisse liefert, die mit den experimentellen Ergebnissen sehr gut übereinstimmen.

Wir können jetzt direkt (A.13) anwenden:

$$U_p = \frac{U_0 e^{-i\omega t}}{4\pi} \int_A \left[\frac{e^{ikr}}{r} \operatorname{grad}_n \frac{e^{ikr'}}{r'} - \frac{e^{ikr'}}{r'} \operatorname{grad}_n \frac{e^{ikr}}{r} \right] dA \quad (\text{A.15})$$

Mit

$$\operatorname{grad}_n \frac{e^{ikr}}{r} = \cos(\vec{n}, \vec{r}) \frac{\partial}{\partial r} \frac{e^{ikr}}{r} = \cos(\vec{n}, \vec{r}) \left(\frac{ike^{ikr}}{r} - \frac{e^{ikr}}{r^2} \right) \quad (\text{A.16})$$

bzw.

$$\operatorname{grad}_n \frac{e^{ikr'}}{r'} = \cos(\vec{n}, \vec{r}') \left(\frac{ike^{ikr'}}{r'} - \frac{e^{ikr'}}{r'^2} \right) \quad (\text{A.17})$$

Dabei bezeichnet (\vec{n}, \vec{r}) den Winkel zwischen der Flächennormalen \vec{n} der Apertur und \vec{r} .

Die Terme e^{ikr}/r^2 und $e^{ikr'}/r'^2$ sind – außer für extrem kleine Abstände r bzw. r' – vernachlässigbar.

Damit ergibt sich die *Kirchhoff-Integral-Formel*, die eine exakte Formulierung des bekannten *Huygens-Prinzips* darstellt

$$U_p = \frac{ikU_0 e^{-i\omega t}}{4\pi} \int_A \frac{e^{ik(r+r')}}{rr'} \left[\cos(\vec{n}, \vec{r}) - \cos(\vec{n}, \vec{r}') \right] dA \quad (\text{A.18})$$

Der Term $[\cos(\vec{n}, \vec{r}) - \cos(\vec{n}, \vec{r}')]$ wird als *Obliquity factor* bezeichnet und kann in vielen praktischen Fällen (große Abstände) als 1 für die Vorwärtsstreuung angenommen werden. (Für die Rückwärtsstreuung⁴ wird er dann 0.)

Ein weiterer Unterschied – abgesehen vom Obliquity Factor – zum einfachen Huygens-Prinzip ist das Auftreten des Faktors i . Er bedeutet, daß die gebeugten Wellen um 90 Grad gegenüber den einfallenden Wellen phasenverschoben sind.

Zu beachten ist, daß obwohl für die Herleitung von (A.18) einige Näherungen gemacht werden mußten, die Genauigkeit einer Beugungsberechnung mittels (A.18) für praktisch alle relevanten Probleme mehr als ausreichend ist. Für die Anwendung bei computergenerierten Hologrammen werden üblicherweise die im Kapitel 2 besprochenen drastischeren Näherungen, die zur Fresnel- und Fraunhofer-Transformation führen, verwendet.

⁴In der üblichen Huygensschen Darstellung fehlt eine Begründung für diese Berücksichtigung der Ausbreitungsrichtung

Anhang B

Fourier–Transformation

Im folgenden werden die wichtigsten Regeln, für die praktische Arbeit mit Fourier–Transformationen angeben. Dabei wird die Fourier–Transformation von Großbuchstaben mit Kleinbuchstaben gekennzeichnet, also f ist die Fourier–Transformation von F : $f = \mathcal{F}[F]$. Für Beweise und eine ausführliche Einführung sei auf [bach92] und [meyb91] verwiesen.

$$a(f) = \mathcal{F}[A(t)] := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A(t) e^{-2\pi i f t} dt \quad (\text{B.1})$$

$$F[\delta(t)] = 1 \quad (\text{B.2})$$

$$\mathcal{F}[\text{rect}(t/a)] = a \text{ sinc}(a f) := a \frac{\sin(\pi a f)}{a \pi f} \quad (\text{B.3})$$

$$\mathcal{F}[\text{sinc}(t/a)] = a \text{ rect}(a f) \quad (\text{B.4})$$

$$\mathcal{F}[\cos(2\pi a t)] = \frac{1}{2}(\delta(f + a) + \delta(f - a)) \quad (\text{B.5})$$

Dabei ist rect definiert als 1 im Bereich $-1/2$ bis $1/2$ und 0 außerhalb dieses Bereichs.

$$\mathcal{F}[\text{comb}(t, T)] = \frac{1}{T} \text{comb}(f, 1/T) \quad (\text{B.6})$$

Mit

$$\text{comb}(t, T) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \quad (\text{B.7})$$

Faltungssatz:

$$\mathcal{F}[A(t) * B(t)] = a(f) b(f) \quad (\text{B.8})$$

mit der Faltungsdefinition¹

$$A(t) * B(t) := \int_{-\infty}^{\infty} A(t)B(t - \tau)d\tau \quad (\text{B.9})$$

Verschiebung:

$$\mathcal{F}[A(t + \tau)] = e^{2\pi if\tau} a(f) \quad (\text{B.10})$$

$$\mathcal{F}^{-1}[a(f + f_0)] = e^{-2\pi itf_0} A(t) \quad (\text{B.11})$$

Dehnung:

$$\mathcal{F}[B(at)] = \frac{1}{|a|} b(f/a) \quad (\text{B.12})$$

Konjugation:

$$\mathcal{F}[A^*(t)] = a^*(-f) \quad (\text{B.13})$$

Parseval-Theorem ("Energiesatz")

$$\int_{-\infty}^{\infty} \|A(t)\|^2 dt = \int_{-\infty}^{\infty} \|a(f)\|^2 df \quad (\text{B.14})$$

Für *reelle Signale* $A(t)$ gilt:

$$a(-f) = a^*(f) \quad (\text{B.15})$$

Fouriertransformation werden in der Praxis über die hier aufgeführten Regeln und Standard-Fourier-Transformations-Paare² oder numerisch (siehe Abschnitt 3.2) berechnet.

¹Die Faltung wird üblicherweise konsequenterweise anders definiert und aus dieser Definition dann der angegebene Ausdruck abgeleitet.

²G.A. Campbell, R.N. Foster: "Fourier Integrals for Practical Applications", McGraw-Hill 1948

Anhang C

Abkürzungen und Symbole

AH	Amplitudenhologramm
AOM	Akustooptischer Modulator
BSO	$Bi_{12}SiO_{20}$
CCD	Charged-Coupled Devive (Halbleiter Kamera)
CGH	Computergeneriertes Hologramm
CPU	Central Processing Unit
DBS	Direct Binary Search
DFT	Diskrete Fourier-Transformation
DMD	Digital Mirror Device
DOE	Diffraktives Optisches Element
DSP	Digital Signal Processor
ED	Error Diffusion
FFT	Fast Fourier Transform (schnelle Fourier-Transformation)
FFTIC	FFT-Integrated Circuit
FLCD	Ferroelektrisches LCD
GIC	Global-Iterative-Coding
HOE	Holographisches Optisches Element
IFTA	Iterative Fourier-Transformations Algorithmen
ITO	Institut für Technische Optik, Uni-Stuttgart
LCD	Liquid Crystal Display (Flüssigkristallanzeige)
MFLOPS	Millions of Floating Point Operations per Second
PCM	Phase-Conjugate Mirror (Phasen-konjugierender Spiegel)
PH	Phasenhologramm
POCS	Projections onto Constraints
PROM	Pockels Read-out Modulator
RGB	Rot-Grün-Blau
SA	Simulated Annealing
SBP	Space-Bandwidth-Product (Orts-Bandbreiten-Produkt)
SEED	Self Electrooptic Effect Devices
SLM	Spatial Light Modulator
SNR	Signal-to-Noise-Ratio (Signal-Rausch-Verhältnis)
TA	Threshold Accepting
TFT	Thin-Film-Transistor
TN-LCD	Twisted-Nematic LCD
VCSEL	Vertical Cavity Surface Emitting Lasers

B	Bereich der Rekonstruktion
b	Clipping-Schwelle
E	mittlerer quadratischer Fehler
e	normierter quadratischer Fehler
η	Beugungseffizienz
f_x, f_y	(reduzierte) Koordinaten in der Objektebene/Rekonstruktionsebene
$f(f_x, f_y)$	komplexe Objektamplitude
f	Brennweite der Rekonstruktionslinse
$h(f_x, f_y)$	rekonstruierte komplexe Amplitude
,	Fresnelpropagator
$H(u, v)$	komplexes Hologramm
L	Hologrammlänge
λ	Wellenlänge, Skalierungsparameter für E
μ, ν	reduzierte Koordinaten in der Hologrammebene
N, M	Anzahl Hologrammpixel in x- bzw. y-Richtung
n_i	Zustand des Neurons i
ν_{max}	maximale Auflösung (cut-off frequency)
Q	Anzahl der Quantisierungsstufen
s_i	Schwellwert des Neurons i
s_i	Stokes Parameter
T	Temperatur im SA-Algorithmus
θ	Phase in der Hologrammebene
u, v	Koordinaten in der Hologrammebene
w_{ij}	Gewichtsmatrix für neuronales Netz
X, Y	Bereichsgröße der Rekonstruktion
z	Abstand Hologramm zur Rekonstruktionsebene
ϕ	Phase in der Objektebene

Literaturverzeichnis

- [aage96] H. Aagedahl: “Theorie der Speckles in der diffraktiven Optik mit Anwendung in der Strahlformung”, 97. Jahrestagung der DGaO, Neuchatel, 29.5.96, Vortrag und persönl. Kommunikation
- [akah73] H. Akahori: “Comparison of Deterministic Phase Coding with Random Phase Coding in Terms of Dynamic Range”, Applied Optics 12, S. 2336 (1973)
- [alle75] J.P. Allebach, B. Liu: “Minimax spectrum shaping with a bandwidth constraint”, Applied Optics 14, S. 3062 (1975)
- [alle76] J.P. Allebach, N.C. Gallagher, B. Liu: “Aliasing error in digital holography”, Applied Optics 15, S. 2183 (1976)
- [alle81] J.P. Allebach: “Representation-related errors in binary digital holograms: a unified analysis”, Applied Optics 20, S. 290 (1981)
- [arav92] P.K. Aravind: “A simple proof of Pancharatnam’s theorem”, Optics Communications 94, S. 191 (1992)
- [bach92] W. Bachmann: “Signalanalyse — Grundlagen und mathematische Verfahren”, Vieweg (1992)
- [barn88] E. Barnard: “Optimal error diffusion for computer-generated holograms”, JOSA A 5, S. 1803 (1988)
- [barn89] T.H. Barnes, T. Eiju, K. Matusda, N. Ooyama: “Phase-only modulation using a twisted nematic liquid crystal television”, Applied Optics 28, S. 4845 (1989)
- [baue93] H.J. Bauer: “Elektronische Digital- und Analogmeßtechnik”, Vorlesungsskript Univ. Stuttgart (1993)
- [berg93] A. Bergeron, H.H. Arsenault, J. Gauvin, D.J. Gingras: “Computer-generated holograms improved by a global iterative coding”, Optical Engineering 32, S. 2216 (1993)
- [berg95] A. Bergeron, J. Gauvin, F. Gagnon, D. Gingras, H.H. Arsenault, M. Doucet: “Phase calibration and applications of a liquid-crystal spatial light modulator”, Applied Optics 34, S. 5133 (1995)
- [berr87] M.V. Berry: “The adiabatic phase and Pancharatnam’s phase for polarized light”, Journal of Modern Optics, 34, S. 1401 (1987)
- [berr90] M.V. Berry: “Anticipations of the Geometric Phase”, Physics Today 12/90, S. 34 (1990)
- [bonn93] L. Bonnel, P. Gravey: “Application of Photothermoplastic Holograms to Single Mode Fibre Optical Interconnects”, 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993
- [blue91] T. Blümecke: “Optimierung mit Evolutionsstrategien und genetischen Algorithmen”, CT 12/91, S. 228 (1991)

- [born80] M. Born, E. Wolf: "Principles of Optics", Pergamon Press (1980)
- [bout96] V. Boutenko, R. Chevallier: "Second-order direct binary search algorithm for the synthesis of computer-generated holograms", *Optics Communications* 125, S. 43 (1996)
- [broj89] M. broja, F. Wyrowski, O. Bryngdahl: "Digital Halftoning by iterative Procedure", *Optics Communications* 69, S. 205 (1989)
- [bräu91] R. Bräuer, F. Wyrowski, O. Bryngdahl: "Diffusers in digital holography", *JOSA A* 8, S. 572 (1991)
- [bron89] I.N. Bronstein, K.A. Semendjajew: "Taschenbuch der Mathematik", Harri Deutsch Verlag (1989)
- [brow66] B.R. Brown, A.W. Lohmann: "Complex Spatial Filtering with Binary Masks", *Applied Optics* 5, S. 967 (1966)
- [brow91] D.M. Brown, A.D. Kathman: "Off-axis spherical element telescope with binary optic corrector", in I. Cindrich, S.H. Lee: "Computer and Optically generated Holographic Optics", *Proceedings of the SPIE Vol. 1555*, S. 114 (1991)
- [bryn74] O. Bryngdahl: "Optical Map Transformations", *Optics Communications* 10, S. 164 (1974)
- [bryn90] O. Bryngdahl, F. Wyrowski: "Digital Holography: Computer-generated Holograms", in E. Wolf: *Progress in Optics Vol. 28*, S. 3 (1990)
- [buck79] J. Bucklew, N.C. Gallagher, Jr.: "Comprehensive error models and a comparative study of some detour-phase holograms", *Applied Optics* 18, S. 2861 (1979)
- [burc70] C.B. Burckhardt: "Use of a Random Phase Mask for the Recording of Fourier Transform Holograms of Data Masks", *Applied Optics* 9, S. 695 (1970)
- [care91] N. Carender, D. Casasent, F. Coetzee, D. Yu: "Hough Transform Computer Generated Holograms: New Output Format", in I. Cindrich, S.H. Lee: "Computer and Optically generated Holographic Optics", *Proceedings of the SPIE Vol. 1555*, S. 182 (1991)
- [casa91] D. Casasent, F. Coetzee, S. Natarajan: "Accuracy of CGH Encoding Schemes for optical data processing", *Proceedings of the SPIE Vol. 1555: "Computer and optically generated holographic optics 4"*, S. 23 (1991)
- [caul79] H.J. Caulfield: "Handbook of Optical Holography", Academic Press New York (1979)
- [chan93] M.P. Chang, O.K. Ersoy: "Iterative interlacing error diffusion for synthesis of computer-generated holograms", *Applied Optics* 32, S. 3122 (1993)
- [chan94] M.P. Chang, O.K. Ersoy: "The modified input-output algorithm for the synthesis of computer-generated holograms", *Optik* 95, S. 155 (1994)
- [chen93] J. Cheng, P. Zhou, S.Z. Sun, S. Hersee, D.R. Myers J. Zolper, G.A. Vawter: "Surface-emitting laser-based smart pixels for two-dimensional optical logic and reconfigurable optical interconnections", *IEEE Journal of Quantum Electronics* 29, S.741 (1993)
- [chu72] D.C. Chu, J.W. Goodman: "Spectrum Shaping with Parity Sequences", *Applied Optics* 11, S. 1716 (1972)
- [chu72b] D.C. Chu: "Polyphase Codes with good periodic Correlation Properties", *IEEE Transactions on Information Theory* 7/72, S. 531 (1972)
- [chyb88] T.H. Chyba, L.J. Wang, L. Mandel: "Measurement of the Pancharatnam Phase for a light beam", *Optics Letters* 13, S. 562 (1988)

- [cohn94] R.W. Cohn, M. Liang: "Approximating fully complex spatial modulation with pseudorandom phase-only modulation", *Applied Optics* 33, S. 4406 (1994)
- [cohn96] R.W. Cohn, M. Liang: "Pseudorandom phase-only encoding of real-time spatial light modulators", *Applied Optics* 35, S. 2488 (1996)
- [czic93] R.H. Czichy, M. Wittig: "Diffractive optics for advanced free space laser communication terminals", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993
- [clar93] M.S. Clark: "A Direct Search Method for the Computer Design of Holograms", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993
- [dall80] W.J. Dallas: "Computer-Generated Holograms", in B.R. Frieden: "The computer in optical research", Springer, S. 291 (1980)
- [dame91] M.P. Dames, R.J. Dowling, P. McKnee, D. Wood: "Efficient optical elements to generate intensity weighted spot arrays: design and fabrication", *Applied Optics* 30, S. 3685 (1991)
- [davi88] J.A. Davis, G.W. Bach, D.M. Cottrell, R.A. Lilly: "Suppression of selected diffraction orders with programmable masks written on spatial light modulators", *Applied Optics* 27, S. 2949 (1988)
- [davi91] L. Davis: "Handbook of genetic algorithms", Van Nostrand Reinhold (1991)
- [dein95] M. Deininger, L. Wang, K. Gerstner, T. Tschudi: "Optical phase step method for absolute ranging interferometry using computer-generated holograms", *Applied Optics* 34, S. 5620 (1995)
- [ditch61] R.W. Ditchburn: "Light", Dover 1991
- [devi94] E. De Vito, A. Levrero: "Pancharatnam's Phase for polarized light", *Journal of Modern Optics* 11, S. 2233 (1994)
- [dobb95] J.G.G. Dobb: "Faster FFTs", *Dr. Dobb's Journal* 2/95, S. 125 (1995)
- [dou96] R. Dou, M.K. Giles: "Simple technique for measuring the phase property of a twisted nematic liquid crystal television", *Optical Engineering* 35, S. 808 (1996)
- [duec90] G. Dueck, T. Scheuer: "Threshold Accepting: A general purpose optimization superior to simulated Annealing", *Journal of Computational Physics* 90, S. 161 (1990)
- [duec93] G. Dueck, T. Scheuer, H.M. Wallmeier: "Toleranzschwelle und Sintflut: neue Ideen zur Optimierung", *Spektrum der Wissenschaft* 3/93, S. 42 (1993)
- [enge87] G. Engelen-Müllers, F. Reutter: "Numerische Mathematik für Ingenieure", BI Wissenschaftsverlag (1987)
- [erso92] O.K. Ersoy, J.Y. Zhuang, J. Brede: "Iterative Interlacing approach for synthesis of computer-generated holograms", *Applied Optics* 31, S. 6894 (1992)
- [esch91] R. Eschbach: "Comparison of error diffusion methods for computer-generated holograms", *Applied Optics* 30, S. 3702 (1991)
- [farh87] H. Farhoosh, M.R. Feldman, S.H. Lee, C.C. Guest, Y. Fainman, R. Eschbach: "Comparison of binary encoding schemes for electron-beam fabrication of computer generated holograms", *Applied Optics* 26, S. 4361 (1987)
- [fett95] F. Fetthauer, S. Weissbach, O. Bryngdahl: "Computer-generated Fresnel holograms: quantization with the error diffusion algorithm", *Optics Communications* 114, S. 230 (1995)

- [fien80] J.R. Fienup: "Iterative method applied to image reconstruction and to computer-generated holograms", *Optical Engineering* 19, S. 297 (1980)
- [fowl75] G.R. Fowls: "Introduction to modern optics", Dover (1989)
- [frer86] C. Frère, D. Leseberg, O. Bryngdahl: "Computer-generated holograms of three-dimensional objects composed of line segments", *JOSA A* 3, S. 726 (1986)
- [frer86b] C. Frère, O. Bryngdahl: "Computer-generated holograms: reconstruction of curves in 3-D", *Optics Communications* 60, S. 369 (1986)
- [gabo70] D. Gabor: "Laser Speckle and its elimination", *IBM Journal of Research and Development* 9/70, S. 509 (1970)
- [gale91] M. Gale: "Diffractive microstructure for security applications", 3rd international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 342, S. 205 (1991)
- [gall73] N.C. Gallagher, B. Liu: "Method for Computing Kinoforms that Reduce Image Reconstruction Error", *Applied Optics* 12, S. 2328 (1973)
- [gerc72] B.W. Gerchberg, W.O. Saxton: "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures", *Optik* 35, S. 237 (1972)
- [good68] J.W. Goodman: "Introduction to Fourier Optics", McGraw-Hill (1968)
- [good73] L.A. Goodman: "Liquid Crystal Displays", *Journal of Vacuum Science Technology* Vol. 5, S. 804 (1974)
- [gran89] P.M. Grant, C.F.N. Cowan, B. Mulgrew, J.H. Dripps: "Analogue and Digital Signal Processing and Coding", Chartwell-Bratt Studentlitteratur 1989
- [gray94] T.P. Grayson, J.R. Torgerson, G.A. Barbosa: "Observation of a nonlocal Pancharatnam phase shift in the process of induced coherence without induced emission", *Physical Review A* 49, S. 626 (1994)
- [greg88] D.A. Gregory, R.D. Juday, J. Sampsel, R. Gale, R.W. Cohn, S.E. Monroe, Jr: "Optical characteristics of a deformable-mirror spatial light modulator", *Optics Letters* 13, S. 10 (1988)
- [háje92] M. Hájek: "Conservative iterative fourier transform algorithms in digital holography", in F. Wyrowski: "Workshop on Digital Holography", *Proceedings of the SPIE* Vol. 1718, S. 63 (1992)
- [hari84] P. Hariharan: "Optical Holography", Cambridge University Press (1984)
- [hash92] N. Hashimoto, K. Hoshino, S. Morokawa: "Improved Real-Time Holography System with LCDs", *SPIE* Vol. 1667 "Practical Holography 6", Ed. S.A. Benton, S. 2 (1992)
- [hask74] R.E. Haskell, P. Tamura: "Another look at Computer Generated Holograms", *Proceedings of the SPIE* Vol. 52, "Coherent optical Processing", S. 43 (1974)
- [hauc84] R. Hauck, O. Bryngdahl: "Computer-generated holograms with pulse-density modulation", *JOSA A* 1, S. 5 (1984)
- [haup92] C. Haupt, M. Pahlke, M. Jäger, H.J. Tiziani: "Design of diffractive optical elements for CO₂-laser material processing" in F. Wyrowski: "Workshop on Digital Holography", *Proceedings of the SPIE* Vol. 1718, S. 175 (1992)
- [hemb94] C. Hembd, H.J. Tiziani: "Diffractive optical elements for CO/sub/2 laser beam diagnostics", *Optics and Laser Technology* 26, S.9 (1994)
- [hopf82] J.J. Hopfield: "neural networks and physical systems with emergent collective computational abilities", *Proceedings of the National Academy of Science* 79, S. 2554 (1982)

- [hopf86] J.J. Hopfield, D.W. Tank: "Computing with Neural Circuits: A Model", *Science* 233, S. 625 (1986)
- [hopt89] R.G. Hoptroff, P.W. McOwan, T.J. Hall, W.J. Hossack, R.E. Burge: "Two optimization approaches to COHOE Design", *Optics Communications* 73, S. 188 (1989)
- [ingb89] L. Ingber: "Very fast simulated re-annealing", *Mathematical and Computer Modelling* 12, S. 967 (1989)
- [iwai96] T. Iwai, T. Asakura: "Speckle Reduction in Coherent Information Processing", *Proceedings of the IEEE* Vol. 84/5, S. 765 (1996)
- [iwas92] S. Iwasa, E.D. Hartfield: "High Speed Laser Recording on the ITEK PROM", *Proceedings of the SPIE* Vol. 52, "Coherent optical Processing", S. 27 (1974)
- [iwat92] S. Iwata, S. Hasegawa, S. Maeda, S. Kayashima, F. Yamagishi: "Holographic straight line scanner using a holoplate", *SPIE* Vol. 1667 "Practical Holography 6", Ed. S.A. Benton, S. 284 (1992)
- [jack83] J.D. Jackson: "Klassische Elektrodynamik", de Gruyter (1983)
- [jenn89] B.K. Jennison, J.P. Allebach: "Iterative approaches to computer-generated holography", *Optical Engineering* 28, S. 629 (1989)
- [jenn89b] B.K. Jennison, J.P. Allebach, D.W. Sweeney: "Direct binary search computer-generated holograms: an accelerated technique of wavefront quality", *Proceedings of the SPIE* Vol. 1052: "Holographic Optics: Optically and Computer generated" (1989)
- [jenn91] B.K. Jennison, J.P. Allebach: "Efficient design of direct-binary-search computer-generated holograms", *JOSA A* 8, S. 652 (1991)
- [just85] D. Just, R. Hauck, O. Bryngdahl: "Computer-generated holograms: structure manipulation", *JOSA A* 2, S. 644 (1985)
- [just87] D. Just, O. Bryngdahl: "Analysis of carrier procedures for discrete halftoning", *Optics Communications* 64, S. 23 (1987)
- [just91] D. Just, D.T. Ling: "Neural networks for binarizing computer-generated holograms", *Optics Communications* 81, S. 1 (1991)
- [kane96] J.S. Kane, T.H. Grycewicz, T.G. Kincaid: "Realizing optical logic with a smart-pixel spatial light modulator", *Applied Optics* 35, S. 1249 (1996)
- [kato91] M. Kato, S. Kadowaki, Y. Komma, Y. Hori: "Recent advances in optical head with holographic optical elements", *Proceedings of the SPIE* Vol. 1507: "Holographic Optics 3: Principles and Applications", S. 36, (1991)
- [kell93] P.E. Keller, A.F. Gmitro: "Computer-generated holograms for optical neural networks: on-axis versus off-axis geometry", *Applied Optics*, 32 S. 1304 (1993)
- [khon91] V.V. Kotlyar, V.A. Soifer: "Fast Hankel transform for focusator synthesis", *Optik* 88, S. 183 (1991)
- [kirk94] A. Kirk, T. Tabata, M. Ishikawa, H. Toyoda: "Reconfigurable computer generated holograms", *Optics Communications* 105, S. 302 (1994)
- [kotl91] V.V. Kotlyar, V. Nikolskii: "Iterative Calculation of an optical element focused on a given region", *Opt. Spektrosk. (USSR)* 71, S. 113 (1991)
- [kotl94] V.V. Kotlyar, P.G. Serafimovich: "Adaptive Iterative Method of Calculating Kinofoms", *Optics and Spectroscopy* 77, S. 606 (1994)

- [kirk83] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi: "Optimization by Simualted Annealing", Science 220, S. 671 (1983)
- [lee74] W.H. Lee: "Binary Synthetic Holograms", Applied Optics 13, S. 1677 (1974)
- [lee78] W.H. Lee: "Making Computer generated holograms" in E. Wolf: "Progress in Optics" Vol. 16, S. 121-229 (1978)
- [lee79] W.H. Lee: "Binary computer-generated holograms", Applied Optics 18, S. 3661 (1979)
- [leit64] E.N. Leith, J. Upatnieks: "Wavefront Reconstruction with diffused Illumination on Three-Dimensional Objects", JOSA 54, S. 1295 (1964)
- [lent93] A.L. Lentine, D.A.B. Miller: "Evolution of the SEED Technology: Bistabile Logic Gates to Optoelectronic Smart Pixels", IEEE Journal of Quantum Electronics Vol. 29, S. 655 (1993)
- [lese87] D. Leseberg: "Computer generated holograms: cylindrical, conical and helical waves", Applied Optics 26, S. 4385 (1987)
- [lese88] D. Leseberg, C. Frère: "Computer-generated holograms of 3-D objects compose of tilted planar segments", Applied Optics 27, S. 3020 (1988)
- [li95] C.F. Li, R. Eschbach, R. Easton: "Neural network for optimization of binary hologram with printing model", SPIE Vol. 2406 "Practical Holography 9", S. 304 (1995)
- [lo92] K.P. Lo, G. Indebetouw: "Iterative image processing using a cavity with a phase-conjugate mirror", Applied Optics 31, S. 1745 (1992)
- [lohm67] A.W. Lohmann, D.P. Paris: "Binary Fraunhofer Holograms, Generated by Computer", Applied Optics 6, S. 1739 (1967)
- [lohm71] A.W. Lohmann: "How to make computer holograms", Proceeding of the SPIE Vol. 25: "Develkopments in Holography", S. 43 (1971)
- [lohm92] A.W. Lohmann, S. Sinzinger: "Local Babinet effect", JOSA A 9, S. 1154 (1992)
- [luce92] M. Lucente: "Optimization of Hologram Computation for Real-Time Display", SPIE Vol. 1667 "Practical Holography 6", Ed. S.A. Benton, S. 32 (1992)
- [mart90] M. Martinelli, P. Vavassori: "A geometric (Pancharatnam) phase approach to the polarization and phase control in the coherent optics circuits", Optics Communications 80, S. 166 (1990)
- [mcc196] J.L. McClain, P.S. Erbach, D.A. Gregory, F.T.S. Yu: "Spatial light modulator phase depth determination from optical diffraction information", Optical Engineering 35, S. 951 (1996)
- [mene94] B. Menees: "Dithern von Bildern mit fraktalen Hilbert-Kurven", MC 6/94, S. 92 (1994)
- [meyb91] K. Meyberg, P. Vachenaer: "Höhere Mathematik 2", Springer (1991)
- [mill87] D.A.B. Miller: "Quantum wells for optical information processing", Optical Engineering 26, S. 368 (1987)
- [mori95] K. Mori, R. Ohba: "Optical Hough-transform processor with a two-dimensional array of computer-generated holograms", Applied Optics 33, S. 3111 (1995)
- [nava91] M.T. Navarro, J.J. Egozcue, A. Fimia: "Computer generated holograms of linear segments", Proceedings of the SPIE Vol. 1507: "Holographic Optics 3: Principles and Applications", S. 142, (1991)
- [neif95] M.A. Neifeld: "Improvements inb the capacity of computer-generated holographic storage using the Lee method with sparse multivalued reconstructions", Applied Optics 34, S. 1396 (1995)

- [noel93] E. Noel, R.R. Khan, H.S. Dhadwai: "Optical implementation of a regularized Gerchberg iterative algorithm for super-resolution", *Optical Engineering* 32, S. 2866 (1993)
- [obri93] D.C. O'Brien, T.D. Wilkinson, R.J. Mears: "Programmable CGHs with Large Space Bandwidth Product", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993
- [owec93] Y. Owechko: "Cascaded-grating holography for artificial neural networks", *Applied Optics* 32, S. 1381 (1993)
- [ozak96] H.M. Ozaktas, D.A.B. Miller: "Digital Fourier Optics", *Applied Optics* 35, S. 1212 (1996)
- [pahl82] M. Pahlke: "Optimierung von Beugungsstrukturen computergenerierter Hologramme für hohen Beugungswirkungsgrad", Diplomarbeit Institut für Technische Optik, Universität Stuttgart (1992)
- [pape83] D.R. Pape, L.J. Hornbeck: "Characteristics of the deformable mirror device for optical information processing", *Optical Engineering* 22, S. 675 (1983)
- [pete92] T. Peter, F. Wyrowski, O. Bryngdahl: "Comparison of iterative methods to calculate quantized digital holograms" in F. Wyrowski: "Workshop on Digital Holography", *Proceedings of the SPIE* Vol. 1718, S. 55 (1992)
- [pete93] T. Peter, F. Wyrowski, O. Bryngdahl: "Relation between angles of diffraction and diffraction efficiency of diffractive elements", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993, S. 80 (1993)
- [pres94] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery: "Numerical Recipes in C", Cambridge University Press (1994)
- [pric94] K.V. Price: "Genetic Annealing", *Dr. Dobb's Journal*, 10/94, S. 127 (1994)
- [neil93] M.A.A. Neil, E.G.S. Paige: "Breaking of Inversion Symmetry in 2-Level, Binary Fourier Holograms", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993
- [nise72] P. Nisenson, S. Iwasa: "Real Time Optical Processing with $Bi_{12}SiO_{20}$ PROM", *Applied Optics* 11, S. 2760 (1972)
- [otto94] T. Otto: "Travelling Salesman - eine neue Strategie für eine alte Aufgabe?", *CT* 11/94, S. 188 (1994)
- [raba96] W.A. Rabadi, H.R. Myler, A.R. Weeks: "Iterative multiresolution algorithm for image reconstruction from the magnitude of its Fourier transform", *Optical Engineering* 35, S. 1015 (1996)
- [reyn89] G. O. Reynolds, J. B. DeVelis, G. B. Parrent, Jr., B. J. Thompson: "The New Physical Optics Notebook: Tutorials in Fourier Optics" SPIE Press (1989)
- [robs74] B.A. Robson: "The Theory of Polarization Phenomena", Clarendon Press, Oxford (1974)
- [ross83] W.E. Ross, D. Psaltis, R.H. Anderson: "Two-dimensional magneto-optic spatial light modulator for signal processing", *Optical Engineering* 22, S. 485 (1983)
- [sale90] B.E.A. Saleh: "Theorie and Design of the liquid crystal TV as an optical spatial phase modulator", *Optical Engineering* 29, S. 240 (1990)
- [sale91] B.E.A. Saleh, M.C. Teich: "Fundamentals of Photonics", John Wiley & Sons (1991)
- [sato92] K. Sato, K. Higuchi, H. Katsuma: "Holographic Television by Liquid Crystal Device", *SPIE* Vol. 1667 "Practical Holography 6", Ed. S.A. Benton, S. 19 (1992)

- [schm93a] H. Schmitzer, J. Höpfeld, W. Dultz: "An Application of the Topological Optical Phases in Holography", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993
- [schm93b] H. Schmitzer, S. Klein, W. Dultz: "Nonlinearity of Pancharatnam's Topological Phase", Physical Review Letters 71, S. 1530 (1993)
- [schr84] D. Schreier: "Synthetische Holographie", Physik Verlag Weinheim (1984)
- [spie88] Proceedings of the SPIE, Vol. 88: "Polarized Light", San Diego 1976
- [schw96] Jens Schweikhardt: "Digitale Holographie mit LCD", Diplomarbeit ITO Stuttgart (1996)
- [seld87] M.A. Seldowitz, J.P. Allebach, D.W. Sweeney: "Synthesis of digital holograms by direct binary search", Applied Optics 26, S. 2788 (1987)
- [sieg77] A.E. Siegman: "Quasi fast Hankel transform", Optics Letters 1, S. 13 (1977)
- [siem68] W. J. Siemens-Wapniarski, M.P. Givens: "The Experimental Production of Synthetic Holograms", Applied Optics 7, S. 535 (1968)
- [silv93] R. Silvennoinen, J. Räsänen, K. Peiponen, C. Gu: "Computer-generated hologram in sensing metal surface quality", Optics Communications 95, S. 231 (1993)
- [soar91] O.D.D. Soares, L.M. Bernardo, F.P. Morais, M.I. Pinto, L.A. Lage: "Holotags for security applications", 3rd international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 342, S. 213 (1991)
- [spau91] K.E. Spaulding, G.M. Morris: "Achromatization of optical waveguide components", Proceedings of the SPIE Vol. 1507: "Holographic Optics 3: Principles and Applications", S. 45 (1991)
- [stea84] S.D. Stearns: "Digitale Verarbeitung analoger Signale", Oldenbourg Verlag (1984)
- [stei92] A.D. Stein, Z. Wang, J.S. Leigh, Jr.: "Computer-generated holograms: A simplified ray-tracing approach", Computers in Physics 6, S. 389 (1992)
- [strn86] J. Strnad: "Photons in introductory quantum physics", Am. J. Phys 54, S. 650 (1986)
- [szu87] H. Szu, R. Hartley: "Fast Simulated Annealing", Physics Letters A, S. 157 (1987)
- [ti96] Texas Instruments, <http://www.ti.com/sc/docs/dsps/prodinfo/overview.htm>
- [tiwa92] S.C. Tiwari: "Geometric phase in optics: Quantal or classical?", Journal of Modern Optics 39, S. 1097 (1992)
- [tiwa93] S.C. Tiwari: "Pancharatnam phase for photon", Optik 93, S. 32 (1993)
- [tizi80] H.J. Tiziani: "Prospects of testing aspheric surfaces with computer-generated holograms", Proceedings of the SPIE Vol. 235: "Aspheric Optics: Design, Manufacture, Testing", S. 72 (1980)
- [tizi88] H.J. Tiziani: "Optische Informationsverarbeitung", Vorlesungsskript Institut für Technische Optik, Univ. Stuttgart, 3. Auflage (1988)
- [tomm91] T. Tommasi, B. Bianco: "Spatial frequency analysis for the Computer Generated Holography of 3-D objects", Proceedings of the SPIE Vol. 1507: "Holographic Optics 3: Principles and Applications", S. 136, (1991)
- [tori78] Y. Torii: "Synthesis of deterministic phase codes for phase shifter in holography", Optics Communications 24, S. 175 (1978)
- [tutt93] T. Tuttaß, O. Bryngdahl: "Image Halftoning: Fouier transform/neural net iteration", Optics Communications 99, S. 25 (1993)

- [urqu91] K.S. Urquhart, S.H. Lee: "Phase-only encoding method for complex wavefronts", Proceedings of the SPIE, Vol. 1555: "Computer and Optically Generated Holographic Optics 4", San Diego (1991)
- [urqu93] K.S. Urquhart, R. Stein, S.H. Lee: "Computer-generated holograms fabricated by direct write of positive electron-beam resist", Optics Letters 18, S. 308 (1993)
- [vasa92] A. Vasara, M.R. Taghizadeh, J. Turunen, J. Westerholm, E. Noponen, H. Ichikawa, M. Miller, T. Kaakkola, S. Kuisma: "Binary surface-relief gratings for array illumination in digital optics", Applied Optics 31, S. 3320 (1992)
- [veld92] W.B. Veldkamp, T.J. McHugh: "Binäre Optik", Spektrum der Wissenschaft 7/92, S. 44 (1992)
- [wagh95] A.G. Wagh, V.C. Rakhecha: "On measuring the Pancharatnam phase Part 1: Interferometry", Physics Letters A 197, S. 107 (1995)
- [wagn77] M. Wagner: "Elemente der Theoretischen Physik, Band 2", Vieweg Verlag 1977
- [wate68] J.P. Waters: "Three-Dimensional Fourier-Transform Method for Synthesizing Binary Holograms", JOSA 58, S. 1284 (1968)
- [watl95] J.A. Watlington, M. Lucente, C. Sparell, V.M. Bove: "A hardware architecture for rapid generation of electro-holography fringe patterns", SPIE Vol. 2406 "Practical Holography 9", S. 172 (1995)
- [weib93] K.J. Weible, H.P. Herzig: "Optical Optimization of Binary Phase Fan-Out Elements", 4th international Conference on Holographic Systems, Components and Applications, IEE Proceedings No. 379, Neuchatel 1993, S. 129 (1993)
- [weis89] S. Weissbach, F. Wyrowski, O. Bryngdahl: "Digital Phase Holograms: Coding and Quantization with an error diffusion concept", Optics Communications 72, S. 37 (1989)
- [weis91] S. Weissbach, F. Wyrowski: "Application of error diffusion in diffractive optics", Proceedings of the SPIE Vol. 1507: "Holographic Optics 3: Principles and Applications", S. 149 (1991)
- [weis92] S. Weissbach, F. Wyrowski: "Error diffusion procedure: theory and applications in optical signal processing", Applied Optics 31, S. 2516 (1992)
- [wind95] R. Windecker, H.J. Tiziani: "Topometry of technical and biological objects by fringe projection", Applied Optics 34, S. 3644 (1995)
- [worc96] T.L. Worchesky, K.J. Ritter, B. Lane: "Large Arrays of spatial light modulators hybridized to silicon integrated circuits", Applied Optics 35, S. 1180 (1996)
- [wyro88] F. Wyrowski, O. Bryngdahl: "Iterative Fourier-transform algorithm applied to computer holography", JOSA A 5, S. 1058 (1988)
- [wyro89a] F. Wyrowski: "Iterative quantization of digital amplitude holograms", Applied Optics 28, S. 3864 (1989)
- [wyro89b] F. Wyrowski, O. Bryngdahl: "Speckle-free reconstruction in digital holography", JOSA A 6, S. 1171 (1989)
- [wyro90] F. Wyrowski: "Diffraction efficiency of analog and quantized digital amplitude holograms: analysis and manipulation", JOSA A 7, S. 383 (1990)
- [wyro90b] F. Wyrowski: "Diffractive optical elements: iterative calculation of quantized, blazed phase structures", JOSA 7, S. 961 (1990)
- [wyro91] F. Wyrowski: "Digital phase-encoded inverse filter for optical pattern recognition", Applied Optics 30, S. 4650 (1991)

- [yin95] S. Yin, G. Lu, J. Zhang, F.T.S. Yu, J. N. Mait: “Kinoform–based Nipkow disk for a confocal microscope”, *Applied Optics* 34, S. 5695 (1995)
- [yone76] S. Yonezawa: “A Deterministic Shifter for Holographic Memory Devices”, *Optics Communications* 19, S. 370 (1976)
- [zell94] A. Zell: “Simulation Neuronaler Netze”, Addison–Wesley (1994)
- [zhan95] E. Zhang, S. Noethe, C.H. Dietrich, R. Männer: “Gradual and random binarization of gray–scale holograms”, *Applied Optics* 34, S. 5987 (1995)
- [zhua94] J.Y. Zhuang, O.K. Ersoy: “Fast decimation–in–frequency direct binary search algorithms for synthesis of computer–generated holograms”, *JOSA A* 11, S. 135 (1994)

Index

- Abbildungsmaßstab, 15
- Abkürzungen, 146
- Adressierung, 111
- Akustooptischer Modulator, 114
- Aliasing, 16, 44, 97
- Amplitude-information Hologramm, 20
- Amplituden-Phasen-Kopplung, 61
- Amplitudenhologramm, 20, 30, 31, 39
- analog-optische Berechnung (FTrf.), 129
- Analytische Verfahren, 94
- Ansprechzeit, 111
- Anwendungen computergenerierter Hologramme, 8, 131
- Apertur, 28, 47
- Array-Generator, 45
- Asphärische Elemente, 9, 59
- Auflösung, 21, 29
- Autokorrelation, 48, 50

- Babinet-Effekt, 85
- Bandbreite, 43, 44, 60
- Bandpaßabtastung, 18
- Benchmarks, 125
- Bessel-Funktionen, 97
- Beugung, 11
- Beugungseffizienz, 31, 32, 34
- Bias-Term, 22
- Bildebenen-Hologramm, 21
- Bildverarbeitung, 9
- binäre Hologramme, 31
- binäre Optik, 31
- Binarisierung, 57
- Blaze-Gitter, 35
- Boltzmann-Verteilung, 81
- Born-Approximation, 94
- Breitenstörterm, 35
- Burch-Kodierung, 39

- CGHMark, 125

- DC-Anteil, 31
- Dehnung, 145
- Detour-Phasen-Hologramme, 40

- DFT-Alias, 97
- diffraktive Optik, 11
- Diffusor, 24, 47
- Diffusoren, 43
- Digital Mirror Device, 113
- Digitaler Signalprozessor, 128
- Direct Binary Search, 47, 55, 73, 87, 129
- Direkte Suchalgorithmen, 72
- Diskrete Fourier-Transformation, 16
- Diskretisierung, 116
- Displays, 9, 64
- Doppelspalt, 118
- dreidimensionale CGHs, 9, 62, 79, 93, 94
- dreidimensionaler DBS-Algorithmus, 79
- dreidimensionaler IFT-Algorithmus, 62
- Dynamik, 28, 48

- Eltern, 92
- Energiesatz, 145
- Error Diffusion, 59, 86, 89

- Faltung, 97, 145
- Faraday-Effekt, 113
- Fast Hankel-Transform, 97
- Fast-DBS-Algorithmus, 75
- Fensterung, 59, 75
- Fermi-Verteilung, 81
- FFT, 126
- Formelsymbole, 146
- Fourier-Hologramme, 13, 22
- Fourier-Transformation, 14, 144
- Fraunhofer-Beugung, 13
- Fraunhofer-Hologramm, 20
- Freiheitsgrade, 37, 59
- Fresnel-Beugung, 11
- Fresnel-Hologramm, 19, 20
- Fresnel-Kirchhoff-Formel, 142

- gap-and-overlap, 41
- general-purpose diffusor, 43
- genetische Algorithmen, 92, 129
- Gerchberg-Saxton-Algorithmus, 52

- Gerchberg–Saxton–Algorithmus, optische Implementierung, 55
- Global–Iterative–Coding, 70
- globales Extremum, 81
- Gradientenverfahren, 73, 81, 130
- Höhenstörterm, 35
- Hankel–Transformation, 97
- Hard–Clipping, 57
- Haskell–Kodierung, 42
- Heuristiken, 78
- hidden–lines, 64
- Holographie, 19
- Holographische Displays, 9
- Homogenität, 116
- Hopfield–Netz, 88
- Huang–Prasada–Kodierung, 40
- Huygens–Prinzip, 143
- hybride Systeme, 130
- image–specific diffusor, 43
- Impulskompression, 50
- Indium–Oxid, 109
- Inkohärentes Hologramm, 21
- Input–Output–Algorithmus, 68
- Iteratives Interlacing, 86
- Interlacing, 86
- Inversions–Symmetrie, 95
- Iterative Fourier–Transformations–Algorithmen, 52
- Iterative Multiresolution Algorithm, 55
- Jones–Matrix (Messung), 114
- Kardinalreihe, 18
- Kinoform, 20, 30
- Kirchhoff–Integral–Theorem, 140
- Kodierung, 39
- Komplexes Hologramm, 30
- konfokales Mikroskop, 9
- Konjugation, 145
- konstante Phase, 49
- Kontrast, 34, 119
- Konvergenz, 53, 61
- Korrektur optischer Systeme, 9
- LCD, 34, 103, 109
- Leakage, 18
- Lee–Kodierung, 70
- Lee–Kodierung, 41
- Leistungsspektrum, 48
- line–at–a–time, 111
- Linsen–Diffusor, 51
- Linsenloses Fourier–Hologramm, 22
- Lohmann–Diffusor, 49
- Lohmann–Kodierung, 41
- lokaler Babinet–Effekt, 86
- Mach–Zehnder–Interferometer, 118
- Magnetooptische Modulatoren, 113
- Maxwell–Gleichungen, 139
- Mikromarkierung, 10
- Mikrooptik, 9
- Mikrostruktur, 35
- Mittelwertfilter, 118
- Mixed–Radix Algorithmen, 127
- Modulation, 30
- Multiple Quantentrog Modulatoren, 113
- Mutation, 93
- Nebenbedingungen, 37
- Neuronale Netze, 81, 87
- Normierung, 128
- Nullte Ordnung, 30
- Nyquist–Frequenz, 18
- Oberflächenuntersuchung, 9
- Objektphase, 37, 60
- Obliquity–Faktor, 143
- off–axis, 30
- off–axis, 20
- on–axis, 20, 30
- Optimierung, 37
- Optimierung von Algorithmen, 126
- Optimierungsziel, 78
- optisch adressiertes LCD, 45, 112
- optische Computer, 9
- optische Verbindungen, 9
- Orts–Bandbreiten–Produkt, 21, 22, 24, 46, 59
- Ortsfrequenz, 15
- Parkettierung, 45, 46, 65
- Parseval–Theorem, 145
- periodische Wiederholung, 45
- Phase–information Hologramm, 20
- Phase–Retrieval, 52
- Phasenhologramm, 20, 30
- Phasenkonjugation, 20, 56
- Phasenmessung, 115
- Phasenschiebung, 99
- Phasenstörung, 118
- Pinp–Pong–Propagation, 93

- Pixel-Dot-Clock, 116
- Polarisation, 110
- Polarisatoren, 61
- Population, 92
- Projection onto Constraints, 58
- Projektionen, 34
- PROM – Pockels Read-out Modulator, 114
- pseudoskopisches Bild, 20
- Puls-Density Modulation, 84
- Puls-Width-Modulation, 84
- Pulsbreiten-Modulation, 84
- Pulsdichtenmodulation, 84
- punktorientierte Hologramme, 31
- punktorientiertes Hologramm, 39
- Punktsymmetrisches Bild, 23

- Quantisierung, 30, 36, 37, 57, 78, 83

- Radar-Codes, 50
- Radix-n Algorithmen, 127
- random diffusor, 43
- Random-Diffusor, 48
- Ray-Tracing, 95
- Rechengenauigkeit, 128
- Reelle FFT, 127
- reelle Signale, 145
- Referenzwelle, 20
- Reflexionsanordnung, 119
- Rekonstruktionsfehler, 32, 33
- Rekonstruktionsformel, 18
- Rekonstruktionslinse, 15, 24
- Rekonstruktionsqualität, 32
- rekursive Berechnung (FFT), 127
- Rotationssymmetrie, 20, 97

- Sampling-Theorem, 18
- Satz von Green, 140
- Scanning, 9
- Schnelle Fourier-Transformation, 126
- Schröder-Code, 51
- Schröder-Diffusor, 51
- Schwellwertspannung, 110
- Second-order-DBS, 79
- SEED, 113
- Selektion, 92
- Separierung, 97
- Sicherheitstechnik, 9
- Signal-Rausch-Verhältnis, 33
- Simulated Annealing, 80, 130
- Simulation, Speckles, 47
- Simultan-FFT, 127
- Sintflut-Algorithmus, 73
- SLM – Spatial Light Modulator, 113
- Smart-Pixel, 57, 130
- Spacer, 109
- Speckle-Ordnung, 61
- Speckles, 16, 44, 60
- spectrum levelling, 43
- Speicherbedarf, 128
- Speicherkapazität, 24
- Spektrale Diffusion, 18
- Spot-Arrays, 81
- Störabstand, 111
- Stagnation, 55
- Startphase, 60, 78
- Startverteilung, 55, 78
- Statistische Mittelung, 47
- statistische Streuung, 57
- stochastische Diffusoren, 43
- Strahldiagnose, 9
- Strahlformung, 9
- Streifenprojektion, 9
- Superpixel, 129
- Superpositionen, 95
- Superresolution, 56
- SYMAS, 9
- Symmetrie, 78, 95
- Symmetrien, 96
- Synchronisation, 116

- Tabellen, 127
- Temperatur, 81
- Testbilder, 35
- Thin Film Transistor, 112, 115
- Threshold Accepting, 80
- Tiefenschärfe, 9, 64
- Tiefpaß, 44, 49
- Top-Clipping, 48
- Trägerfrequenz, 22
- Transformationen, 9
- Translation, 30
- Triangulation, 9
- Trigonometrische Interpolation, 46
- twisted nematic, 110
- twisted nematic LCD, 104, 109

- Vektorprozessor, 128
- Verschiebung, 145
- Vertical Cavity Surface Emitting Lasers, 132
- VGA-Ansteuerung, 117

virtuelle Rekonstruktion, 22

waveguiding, 110

Weißlicht-Hologramm, 21

Wellengleichung, 140

Wellenlänge, 15

Wiener-Khintchine-Theorem, 48

Windowing, 37

zellenorientierte Hologramme, 31

zero-padding, 47

zusätzliche Phasenfilter, 95

Zwangsbedingungen, 37

Zylinderwellen, 94