

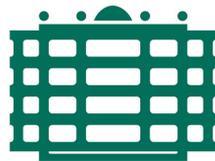
Bildkommunikation

Vorlesungsskript

Basel Fardi

Professur Nachrichtentechnik

Fakultät für Elektrotechnik und Informationstechnik



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Inhaltsverzeichnis

1	Einführung	5
1.1	Anwendungsgebiete der Bildverarbeitung	5
1.2	Schema einer Verarbeitungskette	6
2	Grundlagen	8
2.1	Die visuelle Wahrnehmung des Menschen	8
2.2	Farbe und Farbmodelle	11
2.2.1	Farbmischung	12
2.2.2	Der Rezeptorraum	14
2.2.3	Der psychophysische Raum	15
2.2.4	Farbmodelle	18
3	Bildentstehung	21
3.1	Abbildung	21
3.1.1	Das Lochkameramodell	21
3.1.2	Schärfentiefe	24
3.1.3	Welt- und Kamerakoordinaten	27
3.1.4	Kamerakalibrierung	29
3.2	Das analoge Videosignal	33
3.2.1	Die Auflösung des Bildes	33
3.2.2	Das Bildsignal	34
3.2.3	Das Farbsignal	38
3.2.4	Farbfernsehnormen	42
3.3	Digitalisierung	47

4	Charakterisierung digitalisierter Bilder	50
4.1	Bilder als Funktionen zweier diskreter Variablen	50
4.2	Zufallsvariablen	51
4.3	Zufallsprozesse	56
4.4	Wahrscheinlichkeitsverteilungen	59
5	Punktoperationen	61
5.1	Homogene Punktoperationen	61
5.2	Inhomogene Punktoperationen	64
5.3	Geometrische Transformation	65
6	Bildtransformationen	68
6.1	Grundlagen der Fouriertransformation	68
6.1.1	Definitionen	68
6.1.2	Wichtige Basisfunktionen	70
6.1.3	Stetige Darstellung der Basisfunktionen	72
6.2	Die Fouriertransformation	72
6.3	Diskrete Fouriertransformation	73
6.3.1	Abtasttheorem	73
6.3.2	2D-Abtastung	74
6.3.3	Rekonstruktion	74
6.3.4	DFT	76
6.3.5	2D-DFT	77
6.3.6	Wichtige Eigenschaften der 2D-DFT	77
6.3.7	Anwendungen in der Bildverarbeitung	79
6.3.8	Matrizenschreibweise	79
6.3.9	Periodizitätseigenschaft	80
6.4	Diskrete unitäre Transformationen	81
6.4.1	Kosinus- und Sinustransformation	81
6.4.2	Walshtransformation	83
6.5	FFT	87

7	Nachbarschaftsoperatoren	92
7.1	Lineare verschiebungsinvariante Filter	92
7.1.1	Grundlagen	93
7.1.2	Glättung	97
7.1.3	Kantendetektion	102
7.1.4	Filterdesign	108
7.2	Nichtlineare Filter	110
7.2.1	Rangordnungsfilter	110
7.2.2	Max/Min-Medianfilter	112
8	Morphologische Operatoren	114
8.1	Dilatation und Erosion	114
8.1.1	Dilatation	115
8.1.2	Erosion	115
8.2	Öffnung und Schließung	116
8.2.1	Öffnung	116
8.2.2	Schließung	117
8.3	LOC-Operator	117
9	Segmentierung	120
9.1	Grundlagen	120
9.2	Methodische Ansätze	121
9.2.1	Pixelbasierte Verfahren	121
9.2.2	Kantenbasierte Verfahren	121
9.2.3	Regionenbasierte Verfahren	123

Kapitel 1

Einführung

In dem vorliegenden Vorlesungsskript werden die wichtigsten Grundlagen der Bildverarbeitung behandelt. Der Inhalt orientiert sich hauptsächlich an dem der Vorlesung „Bildkommunikation“ der Professur für Nachrichtentechnik an der Technischen Universität Chemnitz.

Es ist zu beachten, dass die Darstellung einiger Themen unvollständig oder abstrakt ist. Diese werden im Rahmen der Vorlesung und des dazugehörigen Praktikums vervollständigt. Deswegen wird dieses Skript im Zusammenhang mit der oben genannten Vorlesung verwendet und ist nur bedingt zum Selbststudium geeignet.

1.1 Anwendungsgebiete der Bildverarbeitung

Die Bildverarbeitung ist eines der wichtigsten Teilgebiete der Informationstechnik und Datenverarbeitung. Sie beschäftigt sich mit der Manipulation von digitalen Bildern mit dem Ziel, entweder die visuelle Qualität zu verbessern oder bestimmte Eigenschaften zu extrahieren. Es ergeben sich eine Vielzahl von Anwendungsgebiete, die dementsprechend viele Methoden und Verfahren zum Lösen von Problemen mit sich bringen. In der folgenden Übersicht sind einige davon aufgeführt:

- Qualitätskontrolle (Fehlererkennung)
- Robotik (Umgebungserkennung)
- Produktion (Bestückung, Sortierung)
- Medizin, Biologie, Physik (Aufnahmenauswertung)

- Kriminologie (Fingerabdrücke, Portraitauswertung)
- Fernsehtechnik (digitale Bildübertragung, Spezialeffekte)
- Wettervorhersage (Auswertung von Bildfolgen)
- Militärische Anwendungen (Zielerfassung)
- Verkehr (autonome Fahrzeugführung)
- Topografie, Kartografie
- etc

Die Bildverarbeitungsgebiete werden dabei in zwei Kategorien aufgeteilt:

1. Bildverbesserung und Bildtransformation (mit der Anwendung: Bildkodierung)
2. Automatische Bilderkennung (mit der Anwendung: Klassifikation)

1.2 Schema einer Verarbeitungskette

In Abbildung 1.1 wird beispielhaft eine Bildverarbeitungskette schematisch dargestellt. Dabei ist das Ziel die Erkennung eines bestimmten Musters (Objektes) in einer unbekanntem Umgebung. Da das gesuchte Objekt an jeder Stelle des Bildes vorkommen kann, wird das Bild vollständig verarbeitet und ausgewertet. Weiterhin kann diese Verarbeitungskette unabhängig vom Objekttyp eingesetzt werden und stellt somit eine allgemeine Erkennungsstrategie dar.

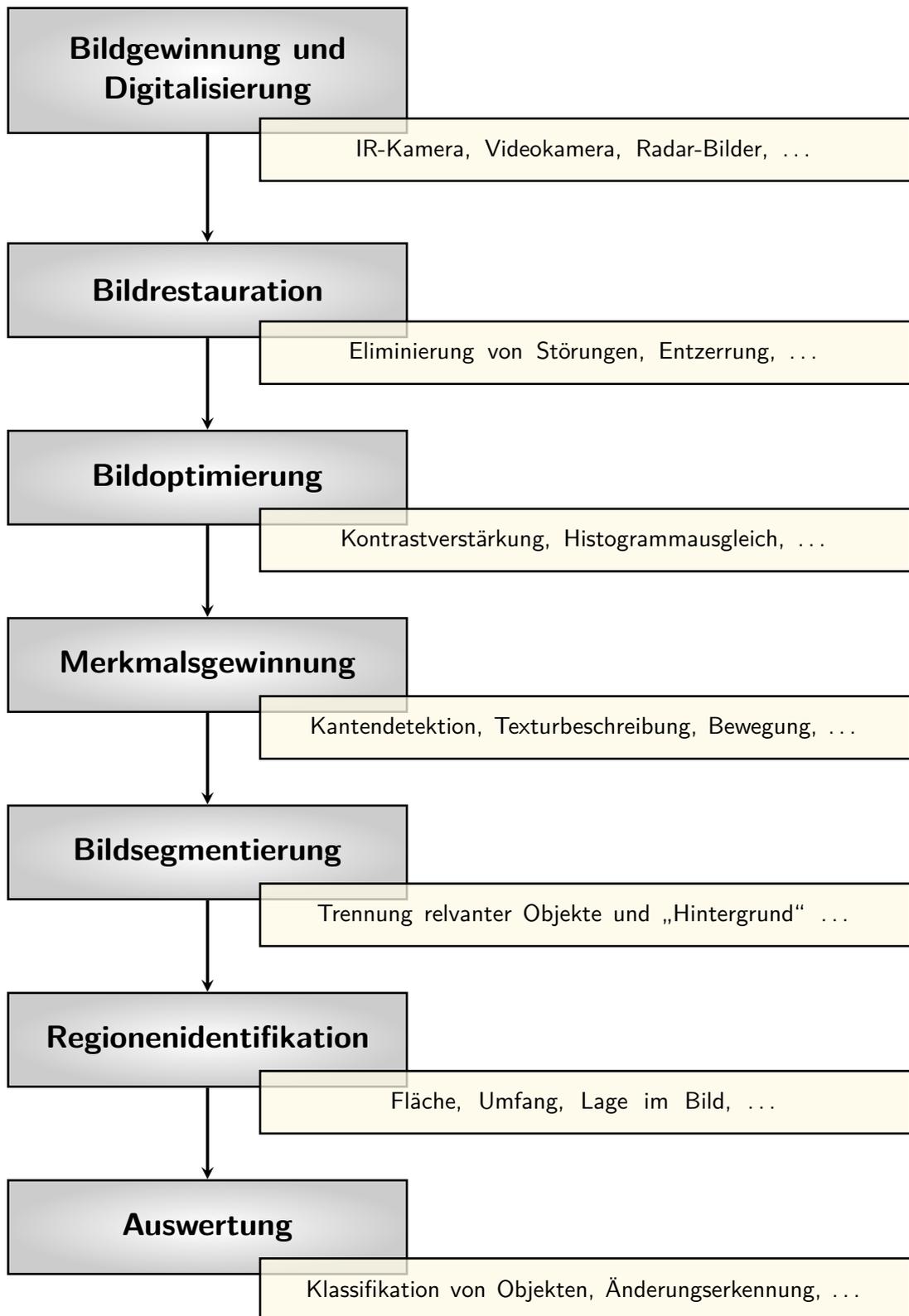


Abbildung 1.1: Schema einer Verarbeitungskette

Kapitel 2

Grundlagen

In diesem Kapitel soll gezeigt werden, wie der Mensch visuelle Wahrnehmungen verarbeitet. Dabei wird erläutert, welche anatomischen Systeme (Zäpfchen, Stäbchen) dafür verantwortlich sind und welche Grenzen (Auflösungsvermögen, Sinnestäuschungen) existieren. In der Farbentheorie gibt es drei theoretische Konzepte für Farbräume. Nach deren Vorstellung werden die praktisch wichtigen Farbmodelle (RGB, HSB, CMY) diskutiert.

2.1 Die visuelle Wahrnehmung des Menschen

Um das Wahrnehmungssystem des Menschen untersuchen zu können, werden folgende Untersuchungsebenen definiert:

1. Anatomie und Optik des Auges sowie Anatomie der Sehbahn In der Netzhaut unterscheidet man zwischen zwei Typen von Rezeptorzellen:

- Stäbchen: Sie können nur Lichtintensitäten unterscheiden, nicht aber Farbe. Im menschlichen Auge befinden sich ca. 120 Millionen Stäbchen.
- Zäpfchen: Sie sind für das Erkennen von Farben verantwortlich. Man vermutet, dass es drei Arten gibt, die sich durch ihre spektrale Empfindlichkeit unterscheiden. Es gibt ca. 6 Millionen im menschlichen Auge. Im Vergleich zu den Stäbchen haben sie lediglich 1/10000 der Helligkeits-Empfindlichkeit.

Diese Rezeptorzellen nehmen das Licht auf, können aber nur einen geringen Teil des Strahlenspektrums erkennen (sichtbares Licht). Abbildung 2.1 zeigt das Strahlenspektrum mit dem darin enthaltenen sichtbaren Licht von violett bis rot.

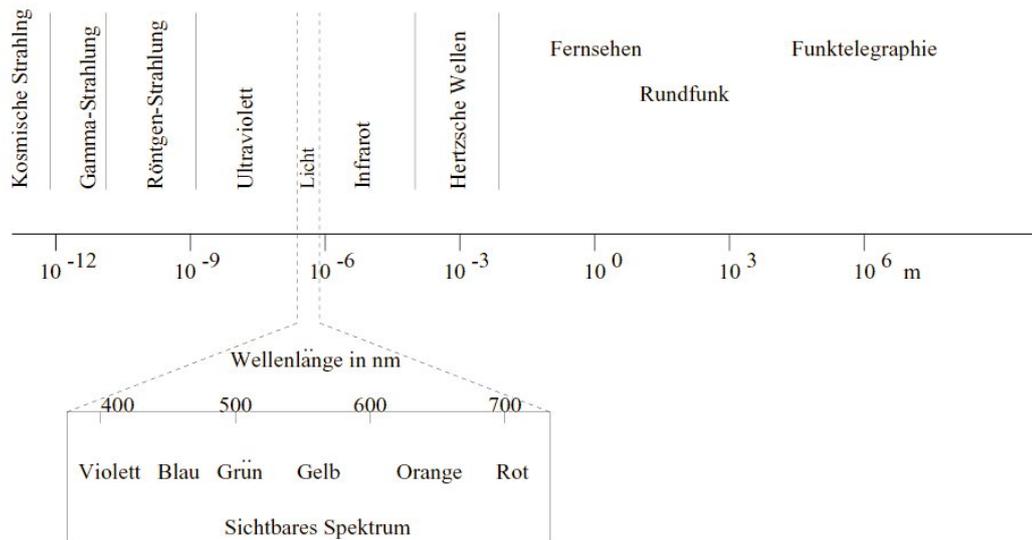


Abbildung 2.1: (aus Rehrmann [8]) Strahlenspektrum

Die Unterschiede zwischen dem visuellen System des Menschen und den heute üblichen Bildverarbeitungssystemen sind:

- Auflösungsvermögen
 - Die Netzhaut des Auges enthält ca. 6 Millionen Zapfenzellen. Ihre größte Dichte besitzen sie im gelben Fleck mit etwa $160.000 \frac{\text{Zäpfchen}}{\text{mm}^2}$. Die Anzahl der Stabzellen liegt bei ca. 120 Millionen.
 - Eine CCD-Kamera hat ein Sensorarray mit nur $756 * 581 \approx 440.000$ Bildpunkte.
- Farbsehen
 - Das menschliche Auge kann weniger als 100 Graustufen und etwa 7 Millionen Farbtöne unterscheiden.
 - Digitale Graustufenbilder (8Bit) enthalten 256 Graustufen, und Farbbilder ($3 * 8\text{Bit}$) bis zu 16,7 Millionen Farbtöne.
- Verarbeitungsleistung
 - Man schätzt, dass sie etwa $10^{11} \frac{\text{Bit}}{\text{s}}$ umfasst. Die Schaltzeiten der Nervenzellen liegen zwar im Millisekundenbereich, aber durch die massiv parallele Informationsverarbeitung kann eine extrem hohe Verarbeitungsleistung erreicht werden.
 - Nur Rechner mit spezieller Hardware können einen Datenstrom von ≈ 20 MByte/s (≈ 25 (Bilder) * 512 (Pixelhöhe) * 512 (Pixelbreite) * 3 Byte (Farben)) verarbeiten, um eine ähnliche Informationsverarbeitung zu realisieren.

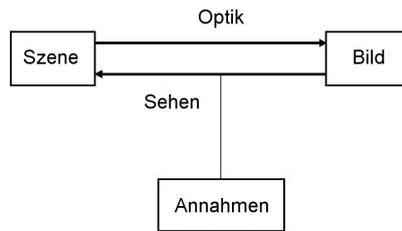


Abbildung 2.2: Prinzip der Wahrnehmung mittels Annahmen.

2. Repräsentation und Algorithmen Wie werden Intensitätsverteilungen repräsentiert und welche neuronalen Operationen gibt es?

Die Realisierung erfolgt entweder elektrisch analog oder elektrisch digital (kodiert).

3. Informationsverarbeitung Was muss überhaupt berechnet werden, um aus Intensitätsverteilungen die gewünschte Information zu erhalten?

Hier stellen sich für Computer und Gehirn die gleichen Probleme: Welche Information kann man aus den Bildern entnehmen?

Beispiele für diese Informationen sind:

- Hell-Dunkel-Unterschiede
- Farbe
- Struktur
- Größe
- Räumliche Tiefe
- Bewegung (zeitliche und örtliche Änderung)
- Regionen (Segmentierung)
- ...

Die Wahrnehmung Das Sehen ist ein Informationsverarbeitungsprozess, der die Beschreibung der Umwelt aufgrund von Bildern rekonstruiert. Diese Inversion ist nur mit Vorwissen oder Annahmen eindeutig möglich (siehe Abbildung 2.2).

Da das Wahrnehmen sehr von der Umgebung abhängig ist, verursachen diese Annahmen Sinnestäuschungen, wie in Abbildung 2.3.

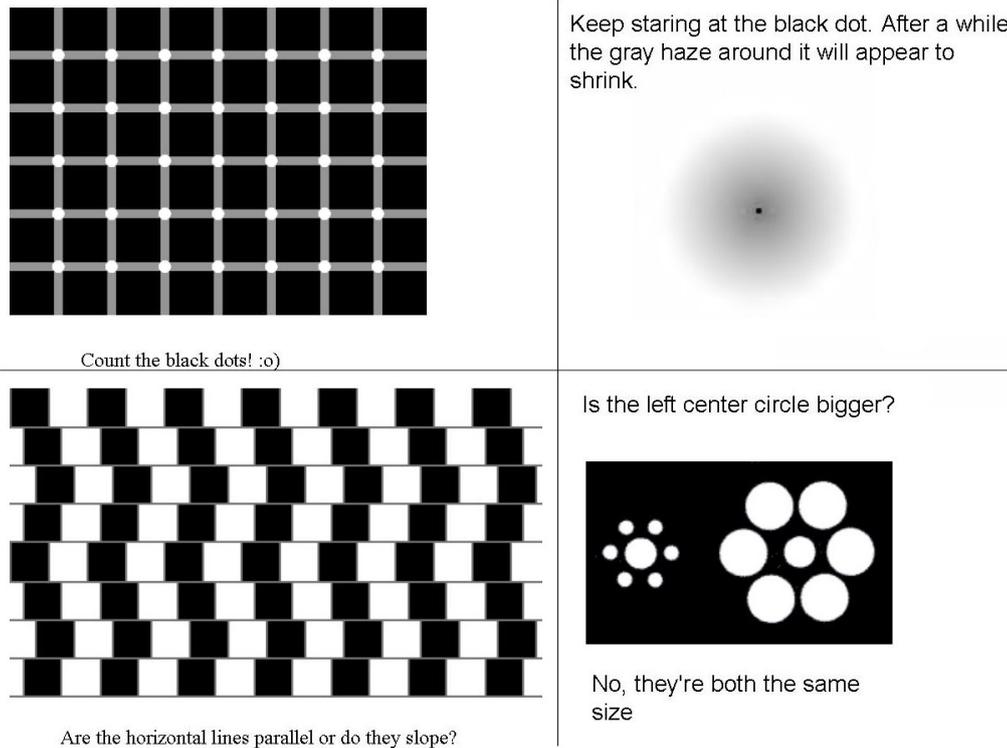


Abbildung 2.3: Illusionen verursacht durch Annahmen beim Sehen.

2.2 Farbe und Farbmodelle

Dieser Abschnitt bezieht sich weitgehend auf das Kapitel 4 des Vorlesungsskriptes von Prof. Mallot [7].

Die Farbe ist keine physikalische, messbare Eigenschaft (wie z.B. Masse oder Volumen), sondern eine Sinnesempfindung und damit eine subjektive Wahrnehmung. Was jedoch messbar ist, das ist der Farbreiz, also das Licht, welches in unser Auge gelangt. Dabei spielt die spektrale Verteilung bei jeder Farbe eine entscheidende Rolle. Man unterscheidet drei wesentliche Begriffe:

1. **Licht und Spektren** bilden die physikalische Grundlage des Farbensehens.
2. **Pigmente oder Farbstoffe** können die spektrale Zusammensetzung des Lichtes verändern.
3. **Farbe** ist eine Wahrnehmung, die auf einem komplizierten physiologischen Prozess beruht.

Für die Theorie des Farbensehens sind drei Konzepte von Farbräumen von Bedeutung:

1. **Physikalisch** betrachtet man Spektren des Lichtes. Die Menge aller Spektren ist ein unendlich-dimensionaler Funktionenraum, da die Intensitäten in allen Wellenlängenbereichen im Prinzip unabhängig voneinander sind. Alle Werte sind nicht negativ, es handelt sich also nicht um einen vollständigen Vektorraum.
2. **Physiologisch** betrachtet man den Raum der möglichen Erregungsmuster in den Rezeptoren („Rezeptorraum“). Entsprechend der Zahl der Rezeptoren ist er dreidimensional. Es handelt sich um einen linearen Unterraum des Spektrenraumes, wobei alle Werte nicht negativ sind.
3. **Psychophysisch** betrachtet man die Mischungsverhältnisse bei der Nachmischung von Farben, sie bilden den eigentlichen Farbenraum.

2.2.1 Farbmischung

Jede Farbwahrnehmung kann aus drei Grundfarben additiv gemischt werden. Dies entspricht physikalisch gesehen einer punktwweisen Addition der Spektren. **Additive** Farbmischung liegt z.B. vor bei:

- Überlagerung von Licht aus unterschiedlichen Projektoren.
- Zeitliche Mischung im Farbkreis: bestimmte Sektoren haben unterschiedliche Farben. Durch Drehen des Kreises entsteht eine Farbmischung.
- Örtliche Mischung durch nahes Aneinanderliegen verschiedenfarbiger kleiner Punkte: z.B. beim Farbbildschirm, Offsetdruck oder beim Weben.

Der additiven Farbmischung steht die subtraktive Farbmischung gegenüber, bei der Pigmente zusammengefügt werden. Die Absorptionsspektren der einzelnen Teilfarben multiplizieren sich dabei. **Subtraktive** Farbmischung liegt z.B. vor bei:

- Mischen von Farben in der Malerei.
(Hinweis: Obwohl Material hinzugegeben (addiert) wird, ist es bezogen auf die Farben eine *subtraktive* Mischung.)
- Übereinanderlegen von Farbfolien vor einem Lichtprojektor.

Abbildung 2.4 zeigt die Grundfarben bei der additiven und subtraktiven Farbmischung, Abbildung 2.5 verdeutlicht noch einmal den Unterschied anhand eines Beispiels. Während bei der additiven Methode die Grundfarben mehr oder weniger frei gewählt

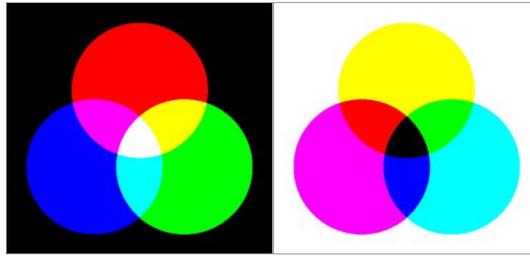


Abbildung 2.4: Links: Additive Farbmischung mit den Grundfarben rot, grün und blau **Rechts:** Subtraktive Farbmischung mit den Grundfarben gelb, cyan und magenta

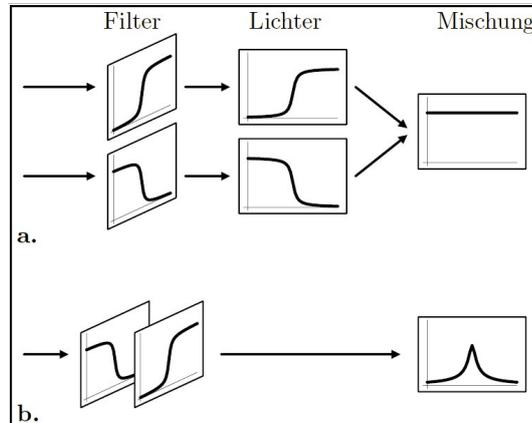


Abbildung 2.5: (aus Mallot [7]) Beispiel für additive und subtraktive Farbmischung. Alle Kurven sind Funktion der Wellenlänge λ . **a.** Additive Farbmischung: Weißes Licht fällt durch zwei Filter mit Hochpass (oben) bzw. Tiefpass (unten). Nach der Absorption erhält man gelbes und blaues Licht. Mittels einer Überlagerung ergibt sich wieder weißes Licht. **b.** Subtraktive Farbmischung: Zwei Filter mit Hochpass- bzw. Tiefpassverhalten werden hintereinandergestellt und weißes Licht fällt drauf. Als Ergebnis erhält man die Multiplikation der Absorptionsspektren: grünes Licht.

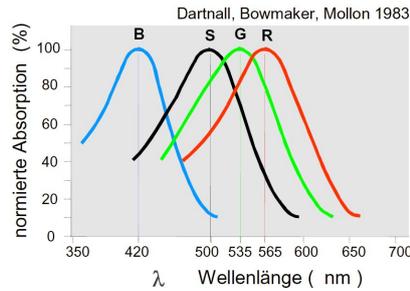


Abbildung 2.6: Absorptionskurven der drei Zapfentypen (rot, blau, grün) und der Stäbchen (schwarz).

werden können, verhält sich dies bei der subtraktiven Methode anders. Dort verdunkelt sich mit der Mischung der einzelnen Farben der Farbton, was zum Beispiel in der Malerei oder beim Drucken nicht erwünscht ist. Deswegen bedient man sich der hellsten, gesättigsten Farben: gelb, cyan und magenta.

Interessant ist, dass Licht mit verschiedener spektralen Zusammensetzung den gleichen Farbeindruck hervorrufen können. Z.B. kann das weiße Licht zusammengesetzt sein aus der gleichmäßigen Verteilung aller Spektralfarben oder der Spektralanteile gelb und blau, das visuelle Ergebnis ist gleich.

2.2.2 Der Rezeptorraum

Die Grundlage für die Farbwahrnehmung ist die Erregung der drei Zapfentypen. Die Erregung eines Rezeptors lässt sich beschreiben mit der Summation der mit der Empfindlichkeit gewichteten spektralen Anteile. Die Erregungen bezeichnet man mit e_k , e_m und e_l für Zapfen mit Absorptionsmaxima im kurzwelligen, mittelwelligen und langwelligen Bereich. Es gilt:

$$\begin{aligned}
 e_k &= \int \varphi_k(\lambda) \cdot I(\lambda) d\lambda \\
 e_m &= \int \varphi_m(\lambda) \cdot I(\lambda) d\lambda \\
 e_l &= \int \varphi_l(\lambda) \cdot I(\lambda) d\lambda
 \end{aligned}$$

Dabei ist $\varphi_i(\lambda)$, $i \in \{k, m, l\}$ die Empfindlichkeit eines Rezeptors, $I(\lambda)$ der Spektralgehalt eines Reizes und λ die Wellenlänge.

Wenn man davon ausgeht, dass die Erregungen der Zapfen im Intervall $[0, 1]$ liegen, so ergibt sich ein würfelförmiger Erregungsraum. Wegen der Überlappung der Absorptionskurven (siehe Abbildung 2.6) ist jedoch nicht jeder Vektor $\vec{e} = (e_k, e_m, e_l)^T$ möglich. Es gibt z.B. keine einzelne Rezeptorerregung der Art $\vec{e} = (1, 0, 0)^T$. Die Menge der realisierbaren Rezeptorerregungen hat ungefähr die Gestalt eines Kegels. Der Querschnitt des Kegels ist dabei annähernd dreieckig. Abbildung 2.7 zeigt den Rezeptorerregungs-

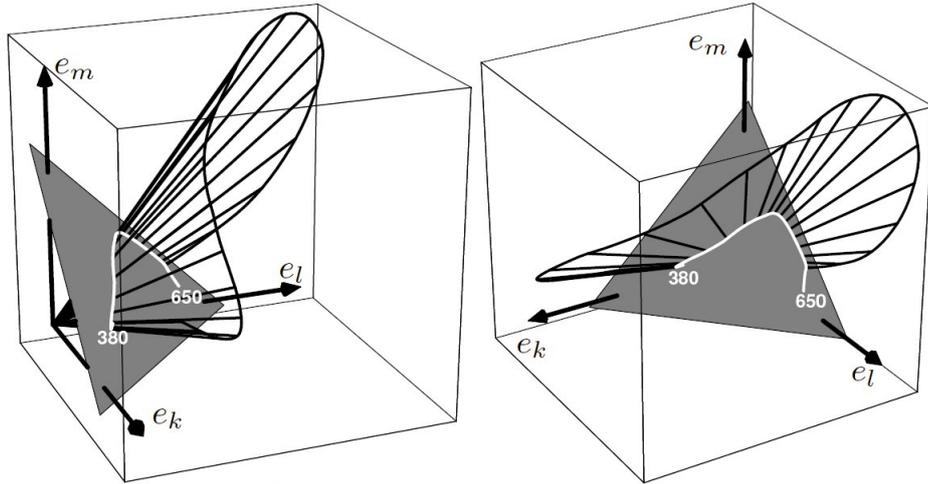


Abbildung 2.7: (aus Mallot [7]) Illustration des Raumes der Rezeptorerregungen. Jede Achse e_i hat das Intervall $[0, 1]$. Entlang einer solchen Achse wird nur die Helligkeit verändert. Die geschwungene Raumkurve zeigt den Spektralfarbenzug von 380 bis 650 nm. Sie entspricht Licht konstanter Irradianz (physikalischer Leistung). **Rechts:** Linkes Bild gedreht, entspricht der Draufsicht auf die Fläche mit einer konstanten Helligkeit.

raum und veranschaulicht die möglichen Vektoren. Das eingezeichnete Dreieck zeigt die konstante Summe der einzelnen Vektoren. Das führt zu der Idee, die Farbdarstellung unabhängig von der Helligkeit zu machen und zweidimensional abzubilden (siehe Abbildung 2.7 rechts).

2.2.3 Der psychophysische Raum

Farbmischung In der Farbmeterik verwendet man nicht den Rezeptorraum, sondern den psychophysischen Raum. Dabei werden drei Grundfarben definiert, die jeweils nicht aus den anderen beiden Farben gemischt werden können (lineare Unabhängigkeit).

Bei einem Wahrnehmungsexperiment wird ein Feld zweigeteilt und auf einen farblosen Hintergrund gebracht. In Abbildung 2.8 gilt es, die linke Seite so additiv nachzumischen, dass die Farbe der Probe auf der rechten Seite gleich ist. Man wählt zuerst drei linear unabhängige Grundfarben mit den Spektren $b_j(\lambda)$, $j \in \{1, 2, 3\}$ aus. Die additive Mischung ist beschreibbar mit $\sum_j a_j \cdot b_j(\lambda)$ mit $0 \leq a_j \leq 1$. Wenn für eine beliebige Probe $f(\lambda)$ mittels der Koeffizienten a_j eine ununterscheidbare Mischung eingestellt wird, so muss die Erregung gleich sein:

$$\int \underbrace{\varphi_i(\lambda) f(\lambda)}_{\text{Probe}} d\lambda = \int \varphi_i(\lambda) \underbrace{\left(\sum_j a_j b_j(\lambda) \right)}_{\text{Nachmischung}} d\lambda \quad (2.1)$$

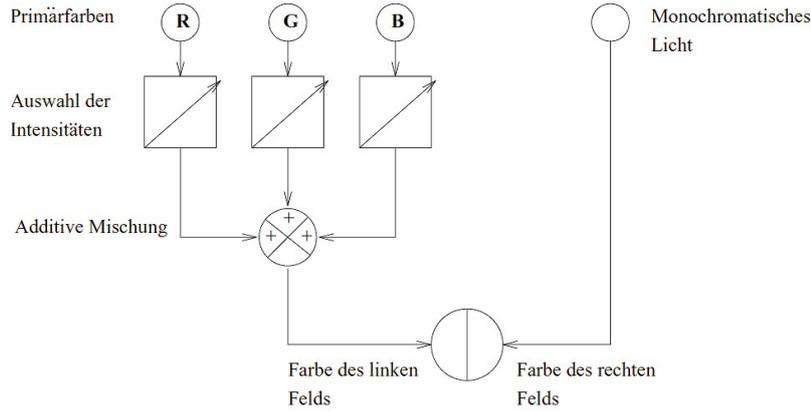


Abbildung 2.8: (aus Rehrmann [8]) Wahrnehmungsexperiment: Die linke Seite des Feldes wird so lange mit den Grundfarben variiert, bis sie der rechten Seite (der Probe) gleicht.

Das Einsetzen von negativen Werten in die Gleichung entspricht praktisch dem Hinzumischen der Farbe zur Probe.

Spektralwertfunktionen In der Farbmeterik spielen die sogenannten Spektralwertfunktionen \bar{b}_j eine große Rolle. Wenn man sich eine reine Spektralfarbe der Wellenlänge λ_0 vorstellt, so sind \bar{b}_j die Farbwerte, mit denen diese Spektralfarbe nachgemischt werden kann. In Gleichung (2.1) werden dabei die allgemeinen Farbwerte a_j durch die Spektralwertfunktionen \bar{b}_j ersetzt. Die Formel ändert sich zu:

$$\int \varphi_i(\lambda) \left(\sum_j \bar{b}_j(\lambda_0) b_j(\lambda) \right) d\lambda = \varphi_i(\lambda_0) \quad (2.2)$$

Wenn man ein beliebiges Spektrum $f(\lambda)$ betrachtet, so erhält man die Rezeptorerregung durch:

$$e_i = \int \varphi_i(\lambda) f(\lambda) d\lambda \quad (2.3)$$

Hier setzt man die linke Seite aus Gleichung (2.2) ein, wobei der Index 0 weggelassen wird und die Integrationsvariable λ durch λ' ersetzt wird.

$$e_i = \int \int \varphi_i(\lambda') \left(\sum_j \bar{b}_j(\lambda) b_j(\lambda') \right) d\lambda' f(\lambda) d\lambda \quad (2.4)$$

Durch Umstellungen ergibt sich:

$$e_i = \int \varphi_i(\lambda') \left(\sum_j b_j(\lambda') \underbrace{\int f(\lambda) \bar{b}_j(\lambda) d\lambda}_{a_j} \right) d\lambda' \quad (2.5)$$

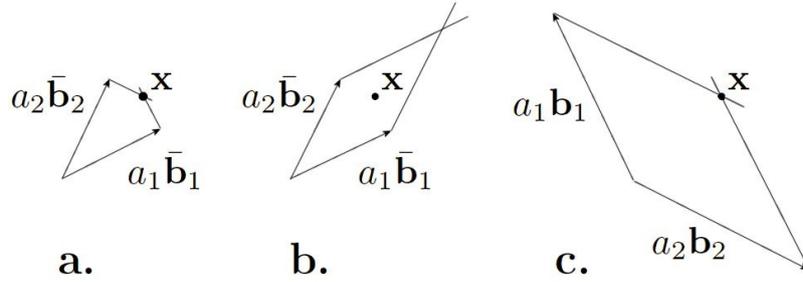


Abbildung 2.9: (aus Mallot [7]) **a.** Durch senkrechte Projektion auf die Achsen \bar{b}_1 und \bar{b}_2 erhält man die „deskriptiven“ Koordinaten des Vektors x mit den Koeffizienten a_1 und a_2 . **b.** Die Addition der Basisvektoren ergibt einen von x verschiedenen Vektor: $x \neq a_1\bar{b}_1 + a_2\bar{b}_2$. **c.** Der Vektor x lässt sich jedoch rekonstruieren, indem man mittels der Gleichung $b_i \sum_j a_j b_j = a_i$ die Werte für b_1 und b_2 findet. Dabei stehen für $j \neq i$ alle b_i senkrecht zu den \bar{b}_j .

Wenn man diese Formel mit der Gleichung (2.1) vergleicht, so sieht man, dass $\int f(\lambda)\bar{b}_j(\lambda)d\lambda$ mit a_j ersetzbar ist. Damit ergeben sich für die Farbwerte a_j :

$$a_j = \int f(\lambda)\bar{b}_j(\lambda)d\lambda \quad (2.6)$$

Die Basis des Farbraums Die Grundfarben und Spektralwertfunktionen bilden zwei Basen, die dual zueinander sind (konstruktive und deskriptive). Bei der Normierung auf 1 gilt:

$$\int b_i(\lambda)\bar{b}_j(\lambda)d\lambda = \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j \end{cases} \quad (2.7)$$

Dabei sind b_i die Grundfarben und \bar{b}_j die Spektralwertfunktionen. Man kann die Gleichung mathematisch interpretieren, so dass für $i \neq j$ alle b_i senkrecht auf alle \bar{b}_j stehen. Folglich haben im Raum der möglichen Spektren, der sich durch $[0, 1]$ ergibt, nicht alle b_i und \bar{b}_j Platz. Wenn man von realisierbaren Grundfarben im Bereich $[0, 1]$ ausgeht, so werden die zugehörigen Spektralwertfunktionen negativ sein. Transformiert man dagegen das Koordinatensystem, so werden die Spektralwertfunktionen positiv, die Grundfarben sind jedoch nicht mehr realisierbar. Somit entscheidet man sich entweder für realisierbare Grundfarben (RGB) wenn man die Methode des Nachmischens verwendet oder für realisierbare Spektralwertfunktionen (XYZ) wenn man Farbfilter konzipieren möchte.

Abbildung 2.9 veranschaulicht die Basen im Koordinatensystem, wobei die Spektralwertfunktionen den Basisvektoren \bar{b}_i entsprechen, die Spektren der Grundfarben den b_i und die Farbwerte den a_i .

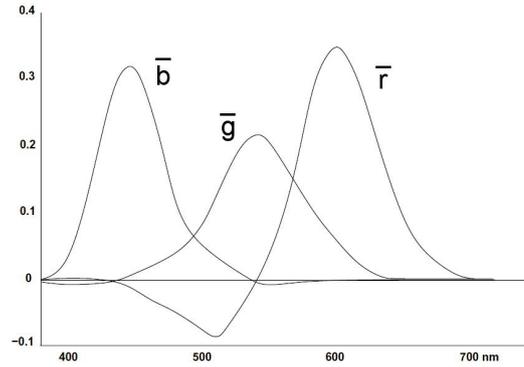


Abbildung 2.10: (aus Rehrmann [8]) Spektrale Empfindlichkeiten von $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ und $\bar{b}(\lambda)$ des RGB-Farbraums.

2.2.4 Farbmodelle

Durch die Wahl von drei Grundfarben legt man ein Koordinatensystem des Farbenraumes fest. Prinzipiell sind diese Farbmodelle gleichwertig, es gibt auch Überschneidungen bei ihren Farbdreiecken. Vor allem die Modelle RGB und XYZ sind von entscheidender Bedeutung.

Das RGB-System Das RGB-System benutzt als Grundfarben reine Spektralfarben mit den Wellenlängen:

$$\lambda_R = 700nm \quad \lambda_G = 546,1nm \quad \lambda_B = 435,8nm$$

Die Spektralwertfunktionen dieses Systems nehmen zum Teil negative Werte an, so dass einige Farben (z.B. $\lambda = 500nm$) nicht nachgemischt werden können (vgl. Abbildung 2.10). Die von der Helligkeit unabhängigen Werte für r , g und b sind durch Normierung entstanden:

$$r = \frac{R}{R + G + B} \quad g = \frac{G}{R + G + B} \quad b = \frac{B}{R + G + B}$$

Das XYZ-System und das CIE-Farbdreieck Das XYZ-System beruht auf dem RGB-System, nur dass hier eine Koordinatentransformation stattgefunden hat, damit alle Werte nicht-negativ werden (vgl. Abbildung 2.11 links). Somit können Filter kreiert werden, die Grundfarben sind aber unreal. Der Koordinatentransformation liegt folgende Matrix zu Grunde:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.173 & 0.21 \\ 0.299 & 0.586 & 0.115 \\ 0 & 0.066 & 1.118 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (2.8)$$

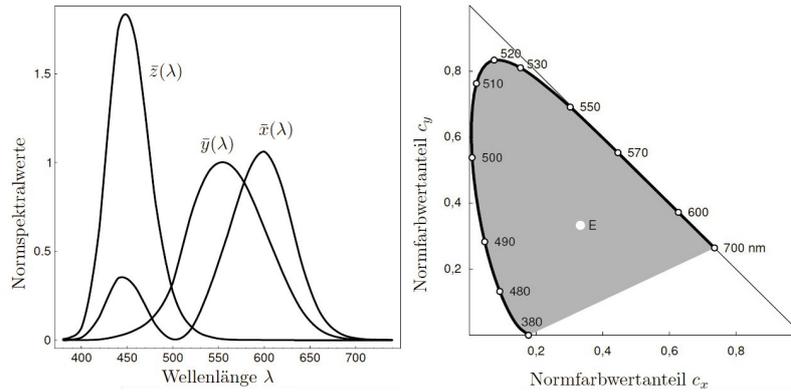


Abbildung 2.11: (aus Mollot [7]) **Links.** Normspektralwertfunktionen für das XYZ-System **Rechts.** Das CIE-Farbdreieck mit dem Spektralfarbenzug (schwarze Linie), dem Weißpunkt (E) und der Purpurlinie (unterer Abschluss des Farbraums).

Wie in Abbildung 2.7 auf Seite 15 gezeigt, lassen sich Farben zweidimensional darstellen wenn man die Helligkeitsunterschiede vernachlässigt. Statt der Farbwerte a_1, a_2, a_3 betrachtet man die normierten Farbwertanteile (chromaticities):

$$c_1 = \frac{a_1}{a_1 + a_2 + a_3}; \quad c_2 = \frac{a_2}{a_1 + a_2 + a_3} \quad (2.9)$$

Eine dritte Angabe für c_3 ist nicht nötig, da sich dieses durch $c_3 = 1 - c_2 - c_1$ errechnet. Abbildung 2.11 zeigt das CIE¹-Farbdreieck in Abhängigkeit der c_i . Die schwarze Linie nennt man *Spektralfarbenzug* und enthält die Farben der höchsten Sättigung. Die Purpurlinie (unterer Abschluss des Farbraums) schließt mit dem Spektralfarbenzug die sogenannte *konvexe Hülle* ein, die alle möglichen Farben beinhaltet. Farben außerhalb dieses Bereichs sind aufgrund des Überlapps der Rezeptorkurven nicht realisierbar.

Das HSB-System (auch HSL, HSV) Bisher haben wir Farben in kartesischen Koordinaten dargestellt. Eine Alternative dazu ist die Darstellung in Polarkoordinaten, die etwas anschaulicher ist. Dabei ist eine Achsengröße die Helligkeit (B...Brightness, L...Luminance, V...Value), eine zweite Achsengröße ist die Sättigung (S...saturation) und der Winkel entspricht dem Farbton (H...hue). Die folgende Tabelle zeigt die Interpretation und Quantisierung der einzelnen Größen.

¹CIE...Commission Internationale de l'Éclairage - Eine Organisation, bestehend aus mehreren Mitgliedsstaaten, die u.a. Standardmodelle für geräteunabhängige Farbraumdarstellung entwickelt haben.

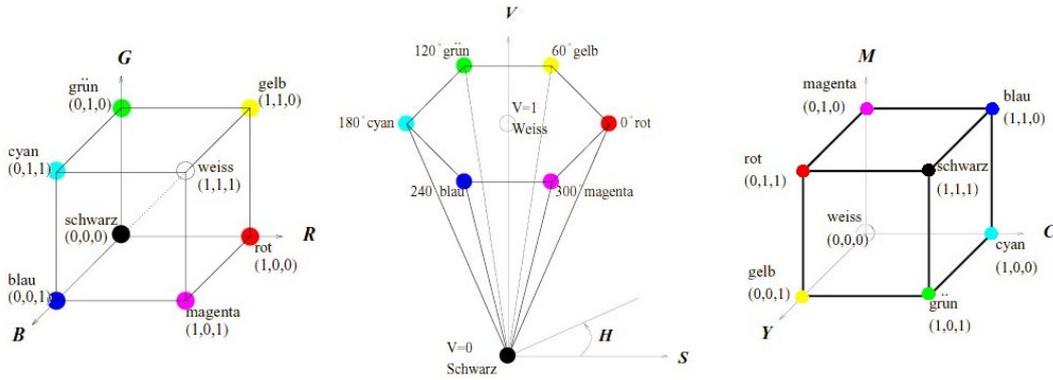


Abbildung 2.12: (aus Rehrmann [8]) Vergleich der Farbmodelle. **Links.** Farbmodell RGB: additive Farbmischung **Mitte.** Farbmodell HSB: Polardarstellung mit Farbton(Hue), Helligkeit(Brightness) und Sättigung(Saturation) **Rechts.** Farbmodell CMY: subtraktive Farbmischung.

Dimension	Interpretation	Wahrnehmbare Unterschiede
Farbton (hue)	Qualität der Farbwahrnehmung (rot, grün, gelb, blau) Wellenlängen der Spektralfarben	ca. 200
Sättigung (saturation)	Grad der Buntheit. Wieviel weiß ist zugemischt?	ca. 20
Helligkeit (brightness)	Lichtintensität gewichtet mit der spektralen Empfindlichkeitsverteilung	ca. 500
Insgesamt		>1.000.000

Das CMY-System Zum Schluss sei noch ein Vertreter der subtraktiven Farbmischung genannt. Dieses Farbmodell wird vorwiegend in der Drucktechnik angewendet, da dort die einzelnen Farben Cyan, Magenta und Yellow auf das Papier gebracht (und gemischt) werden. Manchmal liest man auch vom CMYK-System. Das K steht für Kontrast oder Key und stellt eine zusätzliche Abdunklung dar, da die 3 Sekundärfarben zwar theoretisch gemischt schwarz ergeben, praktisch dies aber nicht ideal erfüllen. Die Umrechnungsmatrix zwischen RGB und CMY lautet:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} C \\ M \\ Y \end{pmatrix} \quad (2.10)$$

Abbildung 2.12 zeigt die grafische Darstellung der Farbmodelle RGB, HSB und CMY im Vergleich.

Kapitel 3

Bildentstehung

In diesem Kapitel sollen die Grundlagen der Bildentstehung erläutert werden. Dazu gehören geometrische Transformationen sowie ein Einblick in die technische Realisierung.

3.1 Abbildung

Die Abbildung ist der geometrische Aspekt der Bildentstehung, bei dem es prinzipiell darum geht, die dreidimensionale Umgebung in eine zweidimensionale Darstellung zu transformieren. Dabei geht immer eine Koordinate verloren und es findet ein wesentlicher Informationsverlust statt. Jedoch ist das menschliche Auge in der Lage, auch aus zweidimensionalen Bildern einen dreidimensionalen Eindruck zu rekonstruieren.

3.1.1 Das Lochkammermodell

Die wesentlichen Abbildungsaspekte sind im Lochkammermodell (vgl. Kapitel 7.3.1 aus Jähne [1]) erläuterbar. Dabei geht man von einer idealen Abbildung aus, da hier eine Lochblende mit infinitesimal kleiner Blendenöffnung genutzt wird. Wie in Abbildung 3.1 gezeigt, kann durch die Blende nur ein Lichtstrahl hindurchdringen und projiziert auf der Bildebene ein Bild im gleichen Seitenverhältnis wie das des Objektes. Der Lichtstrahl des Objektes mit den Koordinaten $(x, y, z)^T$ trifft die Bildebene im Punkt $(x', y', -d')^T$. Die Beziehung zwischen den 3D-Koordinaten der Objektebene und der 2D-Koordinaten der Bildebene lautet:

$$x' = -d' \cdot \frac{x}{z}, \quad y' = -d' \cdot \frac{y}{z}$$

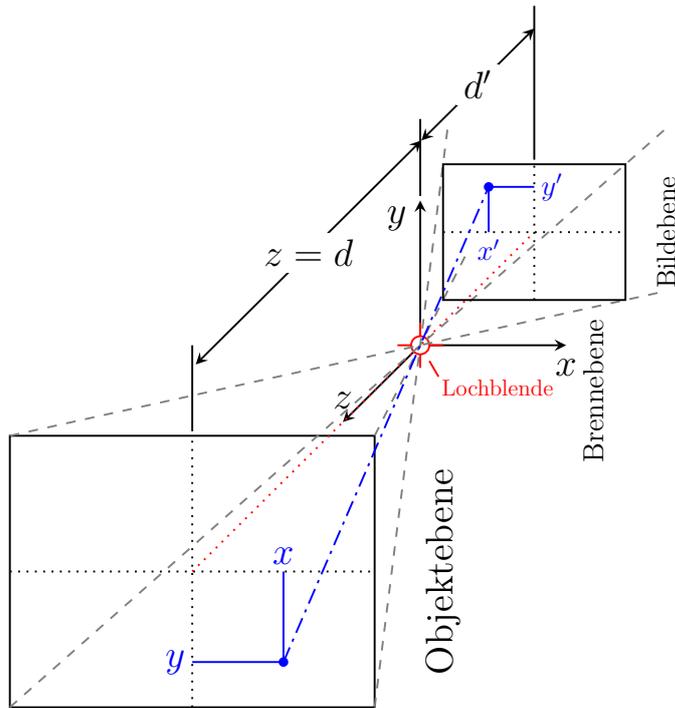


Abbildung 3.1: Bilderzeugung mittels Lochkamera

Mit dem Verlust einer Koordinate enthält die Bildebene nur die *Verhältnisse* des Objektes, jedoch nicht die Größe und die Entfernung.

Wenn Teile der Objektoberfläche verdeckt sind, spricht man von *Okklusion*. Dabei kann es vorkommen, dass die sich überdeckenden Objekte nicht voneinander unterschieden werden können. Diesen Effekt kann man verhindern, indem eine Tiefeninformation hinzugefügt wird. Dabei wird die dritte Dimension durch Aufnahme mehrerer Bilder in Abhängigkeit von z erzeugt, damit die Gesamtheit der Bilder die Informationen eines dreidimensionalen Objektes erhalten. Eine weitere Möglichkeit ist das Betrachten des Objektes aus unterschiedlichen Blickwinkeln, wie z.B. das Stereo-Sehen des Menschen mit zwei Augen.

In der Realität arbeitet man mit lichtempfindlichen Sensoren die erst ab einer bestimmten Beleuchtungsstärke (z.B. CCD ab 2000 Lux bei Blendenzahl 8) arbeiten. Deswegen muss man im Gegensatz zum Lochkameramodell die Öffnung der Blende vergrößern und die einfallenden Strahlen eines Objektpunktes mit Hilfe eines Linsensystems auf einen Bildpunkt fokussieren (Abbildung 3.2).

Dadurch wird die genaue (scharfe) Abbildung auf eine Objektentfernung begrenzt. Abbildung 3.3 zeigt den Strahlenverlauf bei einer dünnen Linse. Dabei gilt die Abbildungsgleichung:

$$\frac{1}{f} = \frac{1}{d} - \frac{1}{d'} \quad (3.1)$$

Das Vorzeichen erklärt sich durch das umgekehrte Bild auf der Bildseite. Das Abbil-

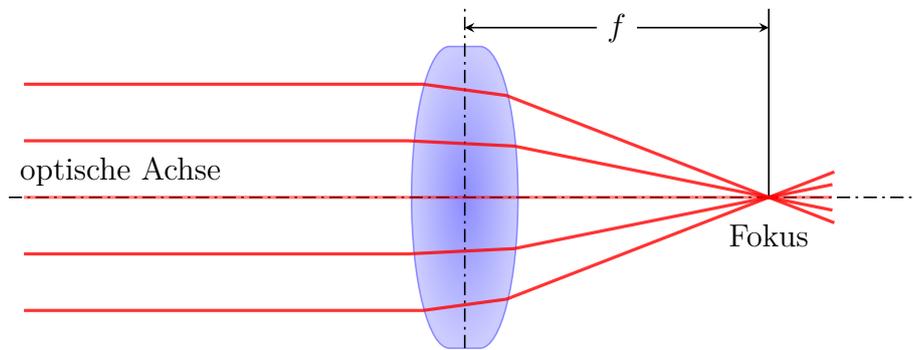


Abbildung 3.2: Fokussierung der Lichtstrahlen durch eine Sammellinse.

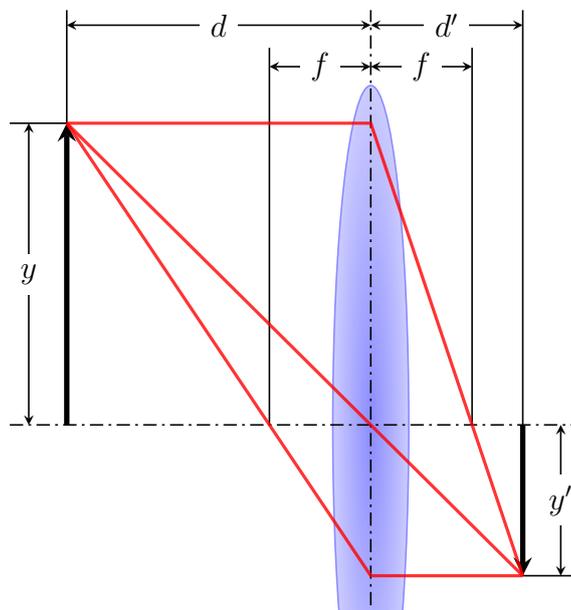


Abbildung 3.3: Abbildung durch eine dünne Linse.

Abbildungsverhältnis ist definiert durch:

$$B = \frac{y'}{y} = \frac{d'}{d}$$

Das Verhältnis von Bild- und Objektgröße nennt man *lateraler Abbildungsmaßstab*. Der *axiale Abbildungsmaßstab* ist dagegen das Verhältnis der Entfernung von Bild und Objekt auf der optischen Achse.

3.1.2 Schärfentiefe

Es wurde schon angedeutet, dass die Objektentfernung für ein scharfes Bild begrenzt ist. Dabei stellt sich die Frage, wie diese sog. *Schärfentiefe* (vgl. Kapitel 7.4.3 aus Jähne [1]) definiert ist.

Das Konzept beruht darauf, dass eine gewisse Unschärfe akzeptiert werden kann, da die bildseitigen Empfängerelemente nicht unendlich klein sind. Für digitale Bilder ist die Größe der Sensorelemente (Pixelgröße) ein natürliches Maß für die akzeptable Unschärfe. Bei der geometrischen Optik wird auf der Bildebene ein Punkt der Objektebene abgebildet. Wenn nun objektseitig die Position verändert wird, so vergrößert sich der Punkt zu einem verschwommenen Kreis. Dieser Kreis muss in die Größe des Sensorelements passen, damit das Bild unverschommen wahrgenommen werden kann.

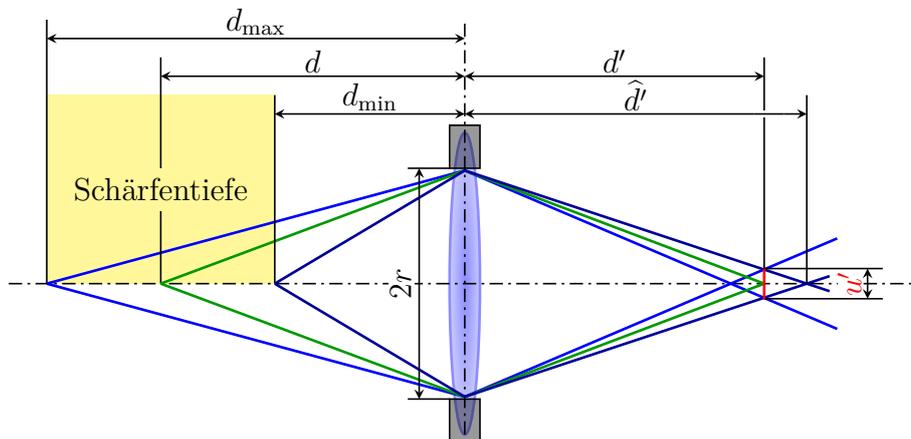


Abbildung 3.4: Illustration der Schärfentiefe eines objektseitigen Punktes auf der optischen Achse. Das Objekt wird innerhalb des Abstandes d_{\min} bis d_{\max} bewegt, die Bildebene bleibt fest.

Abbildung 3.4 zeigt die Schärfentiefe eines Objekts. Dargestellt ist ein auf der optischen Achse verschiebbarer Punkt auf der Objektseite, der in der unverschiebbaren Bildebene einen Kreis mit dem Durchmesser u' erzeugt. Der Wert u' ist dabei der zulässige Unschärfebereich und bestimmt die Schärfentiefe auf der Objektseite. Die Blende kann

mit der Blendenzahl beschrieben werden:

$$n = \frac{f}{2r} \qquad \begin{array}{ll} f & - \text{ Brennweite} \\ 2r & - \text{ Blende} \end{array}$$

Die Schärfentiefe $d_{\max} - d_{\min}$ definiert sich als: Der Bereich vor und hinter der optimalen Gegenstandsweite, in dem die Abbildung eines Gegenstandspunktes nicht größer als u' ist. Die beiden Randgrenzen berechnen sich durch:

$$d_{\min, \max} = \frac{f^2 \cdot d}{f^2 \pm n \cdot u' \cdot (d - f)} \quad (3.2)$$

Herleitung Die Formel (3.2) lässt sich folgendermaßen herleiten:

Abbildungsgleichung (3.1) für fokussierten Objektpunkt nach d' umstellen:

$$d' = \frac{f \cdot d}{f - d} \quad (3.3)$$

Gleichsetzen des Seitenverhältnisses zweier ähnlicher Dreiecke:

$$\begin{aligned} \frac{r}{|\widehat{d}'|} &= \frac{\frac{u'}{2}}{|\widehat{d}'| - |d'|} \\ \frac{r}{-\widehat{d}'} &= \frac{\frac{u'}{2}}{-\widehat{d}' + d'} \end{aligned} \quad (3.4)$$

Auflösen nach \widehat{d}'

$$\begin{aligned} r(d' - \widehat{d}') &= -\frac{u}{2} \cdot \widehat{d}' \\ \widehat{d}' &= \frac{r \cdot d'}{r - \frac{u}{2}} \end{aligned}$$

Gleichung (3.3) einsetzen:

$$\widehat{d}' = \frac{r \cdot f \cdot d}{\left(r - \frac{u}{2}\right) \cdot (f - d)}$$

Durch Verwendung der Blendenzahl wird $r = \frac{f}{2n}$ gesetzt:

$$\widehat{d}' = \frac{f^2 \cdot d}{(f - n \cdot u')(f - d)} \quad (3.5)$$

Jetzt die Abbildungsgleichung (3.1) für den Nahpunkt aufstellen und nach d_{\min} auflösen:

$$d_{\min} = \frac{f \cdot \hat{d}'}{f + \hat{d}'}$$

\hat{d}' ersetzen durch Gleichung (3.5)

$$\begin{aligned} d_{\min} &= \frac{f \cdot \frac{f^2 \cdot d}{(f - n \cdot u')(f - d)}}{f + \frac{f^2 \cdot d}{(f - n \cdot u')(f - d)}} \\ &= \frac{f^2 \cdot d}{\underline{\underline{f^2 + n \cdot u'(d - f)}}} \end{aligned}$$

Analog lässt sich auch die Formel für d_{\max} herleiten, dabei wird das Plus im Nenner durch ein Minus ersetzt.

Beispiel Ein Beispiel soll die Formeln verdeutlichen:

Eine 1.80 m große Person, die in 3.6 m Abstand vor einer Linse steht, soll auf einen Bildwandler abgebildet werden. (Wandlerhöhe: 1 cm, Auflösung: 1000 Pixel, Blendenzahl: 4 und 8)

1. Wie groß ist die Brennweite der Linse?
2. In welchem Bereich kann sich die Person bewegen, ohne den Schärfbereich zu verlassen?

Lösung:

$$y = 1.8 \text{ m} \quad y' = -0.01 \text{ m} \quad d = 3.6 \text{ m} \quad u' = \frac{0.01 \text{ m}}{1000}$$

$$\begin{aligned} \frac{y'}{y} &= \frac{d'}{d} && \Rightarrow && d' = -0.02 \text{ m} \\ \frac{1}{f} &= \frac{1}{d} - \frac{1}{d'} && \Rightarrow && f = 19.9 \text{ mm} \end{aligned}$$

$$d_{\min, \max} = \frac{f^2 \cdot d}{f^2 \pm n \cdot u' \cdot (d - f)} \quad \begin{array}{l} \xrightarrow{n=4} \\ \xrightarrow{n=8} \end{array} \quad \begin{array}{l} d_{\min} = 2.6 \text{ m} \\ d_{\min} = 2.1 \text{ m} \end{array} \quad \begin{array}{l} d_{\max} = 5.6 \text{ m} \\ d_{\max} = 13 \text{ m} \end{array}$$

3.1.3 Welt- und Kamerakoordinaten

Zur Beschreibung von 3D-Objekten im Raum gibt es zwei Möglichkeiten (vgl. Kapitel 7.2.1 aus Jähne [1]). Die Erste ist die Nutzung eines ortsfesten dreidimensionalen orthogonalen Weltkoordinatensystems, welches auf die betrachtete Szene bezogen ist. Die zweite Beschreibungsart beruht auf dem Kamerakoordinatensystem.

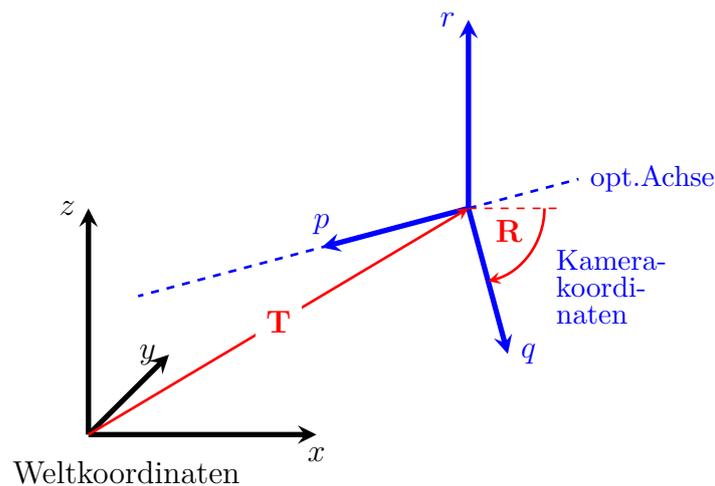


Abbildung 3.5: Überführung des Weltkoordinatensystems in das Kamerakoordinatensystem durch eine Rotation R um den Ursprung und die Translation T .

Beide Koordinatensysteme können durch eine Translation und eine Rotation ineinander umgerechnet werden. Zuerst wird der Ursprung des einen Koordinatensystems mittels Translation in den Ursprung des Anderen bewegt. Dann wird die Richtung mittels Rotation so korrigiert, dass sie mit der des anderen Systems übereinstimmt (siehe Abbildung 3.5).

In vielen Kameraanwendungen wird das Wissen über Lage, Größe und Abstand von Objekten angestrebt. Die eben beschriebene Transformation der Weltkoordinaten in Kamerakoordinaten ist auch mathematisch beschreibbar. Die Kamerakalibrierung sorgt dafür, dass der mathematischen Formulierung Werte zugeordnet werden können. Die

Transformation lässt sich allgemein folgendermaßen darstellen¹:

$$\begin{aligned}\vec{p} &= \mathbf{U} \cdot \vec{x} \\ &= \mathbf{R} \cdot \mathbf{T} \cdot \vec{x}\end{aligned}\tag{3.6}$$

mit $\vec{x} = \begin{pmatrix} x & y & z & 1 \end{pmatrix}^T$... Weltkoordinaten
 $\vec{p} = \begin{pmatrix} p & q & r & 1 \end{pmatrix}^T$... Kamerakoordinaten

Dabei ist

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & -t_x \\ 0 & 1 & 0 & -t_y \\ 0 & 0 & 1 & -t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

die Translationsmatrix und

$$\mathbf{R} = \mathbf{R}_z \cdot \mathbf{R}_y \cdot \mathbf{R}_x$$

mit

$$\mathbf{R}_z(\varphi_z) = \begin{pmatrix} \cos \varphi_z & \sin \varphi_z & 0 & 0 \\ -\sin \varphi_z & \cos \varphi_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Rotation um den Winkel } \varphi_z$$

$$\mathbf{R}_y(\varphi_y) = \begin{pmatrix} \cos \varphi_y & 0 & -\sin \varphi_y & 0 \\ 0 & 1 & 0 & 0 \\ \sin \varphi_y & 0 & \cos \varphi_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Rotation um den Winkel } \varphi_y$$

$$\mathbf{R}_x(\varphi_x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \varphi_x & \sin \varphi_x & 0 \\ 0 & -\sin \varphi_x & \cos \varphi_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Rotation um den Winkel } \varphi_x$$

die Matrizen zur Beschreibung der Rotationen um die einzelnen Koordinatensystemkomponenten.

Um nun die Kamerakoordinaten in Bildkoordinaten zu überführen wird eine zusätzliche Transformation benötigt. Diese setzt sich aus den folgenden zwei Schritten zusammen:

Im ersten Schritt werden die Kamerakoordinaten mit den inneren (intrinsischen) Kame-

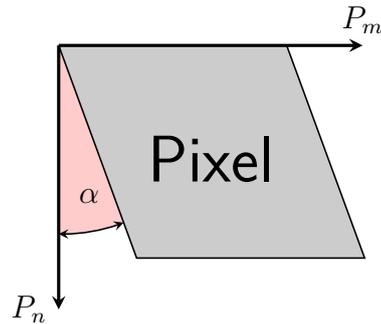
¹Für die Darstellung werden die homogenen Koordinaten verwendet.

raparameter multipliziert. Diese lassen sich mit der folgenden Matrix zusammenfassen:

$$\mathbf{K} = \begin{pmatrix} f_m & S & m_c & 0 \\ 0 & f_n & n_c & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{f}{P_m} & \tan \alpha \cdot \frac{f}{P_n} & m_c & 0 \\ 0 & \frac{f}{P_n} & n_c & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.7)$$

wobei

- f ... Brennweite
- S ... Scherungsfaktor
- P_m ... Pixelbreite
- P_n ... Pixelhöhe
- (m_c, n_c) ... Bildmittelpunkt
- α ... Parameter für die Orthogonalität der Pixelachsen



Die Bildkoordinaten m und n erhält man durch Normierung des Vektors

$$\begin{pmatrix} s \cdot m \\ s \cdot n \\ s \end{pmatrix} = \mathbf{K} \cdot \begin{pmatrix} p \\ q \\ r \end{pmatrix}$$

auf die dritte Komponente:

$$\vec{m} = \begin{pmatrix} m \\ n \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{s \cdot m}{s} \\ \frac{s \cdot n}{s} \\ \frac{s}{s} \end{pmatrix}$$

3.1.4 Kamerakalibrierung

Für die Aufnahme von Bildern mittels Kameras findet eine Digitalisierung statt. Die 3D-Objekte der Szene werden auf die 2D-Fläche des Kamerasensors abgebildet, wobei ein bestimmtes Kameramodell zum Einsatz kommt (z.B. das Lochkameramodell). Wenn man vom Bild aus Rückschlüsse auf die Szene machen möchte, so sind bestimmte Informationen nötig, wie das Bild entstanden ist. Das Kameramodell enthält verschiedene physikalische und geometrische Parameter, die bei der Kamerakalibrierung bestimmt werden sollen. Im Folgenden wird die Methode der direkten Schätzung der Parameter beschrieben.

3.1.4.1 Direkte Schätzung der Parameter

Ausgangspunkt der Kamerakalibrierung ist die Beschreibung der Transformation der Welt- in Bildkoordinaten:

$$s \cdot \vec{m} = \mathbf{A} \cdot \vec{x}$$

mit

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} = \mathbf{K} \cdot \mathbf{U}$$

Um später die Elemente der Matrix A separieren und mittels der Methode der kleinsten Quadrate schätzen zu können (siehe Gleichung 3.11), werden sie auf das Element a_{34} normiert (z.B. $a'_{11} = \frac{a_{11}}{a_{34}}$):

$$s \cdot \begin{pmatrix} m \\ n \\ 1 \end{pmatrix} = \begin{pmatrix} a'_{11} & a'_{12} & a'_{13} & a'_{14} \\ a'_{21} & a'_{22} & a'_{23} & a'_{24} \\ a'_{31} & a'_{32} & a'_{33} & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3.8)$$

Somit kann man für die Bildkoordinaten m, n schreiben:

$$m = \frac{s \cdot m}{s} = \frac{a'_{11} \cdot x + a'_{12} \cdot y + a'_{13} \cdot z + a'_{14}}{a'_{31} \cdot x + a'_{32} \cdot y + a'_{33} \cdot z + 1} \quad (3.9)$$

$$n = \frac{s \cdot n}{s} = \frac{a'_{21} \cdot x + a'_{22} \cdot y + a'_{23} \cdot z + a'_{24}}{a'_{31} \cdot x + a'_{32} \cdot y + a'_{33} \cdot z + 1} \quad (3.10)$$

Wenn man die Parameter a'_{ij} als Vektor separiert, so erhält man für einen Punkt:

$$\underbrace{\begin{pmatrix} m \\ n \end{pmatrix}}_{\vec{b}_1} = \underbrace{\begin{pmatrix} x & y & z & 1 & 0 & 0 & 0 & 0 & -m \cdot x & -m \cdot y & -m \cdot z \\ 0 & 0 & 0 & 0 & x & y & z & 1 & -n \cdot x & -n \cdot y & -n \cdot z \end{pmatrix}}_{\mathbf{M}_1} \cdot \underbrace{\begin{pmatrix} a'_{11} \\ a'_{12} \\ a'_{13} \\ a'_{14} \\ a'_{21} \\ a'_{22} \\ a'_{23} \\ a'_{24} \\ a'_{31} \\ a'_{32} \\ a'_{33} \end{pmatrix}}_{\vec{s}} \quad (3.11)$$

$$\vec{b}_1 = \mathbf{M}_1 \cdot \vec{s}$$

Der Parametervektor \vec{s} ist zu berechnen. Dafür sind 11 Punktkoordinaten (weil 11 Parameter) notwendig, d.h. mindestens 6 Messpunkte. Verwendet man eine größere Anzahl an Messpunkten, wird die Berechnung genauer (Vgl. Statistik, Messen).

$$\vec{b} = \begin{pmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vdots \\ \vec{b}_N \end{pmatrix} = \begin{pmatrix} m_1 \\ n_1 \\ m_2 \\ n_2 \\ \vdots \end{pmatrix} = \mathbf{M} \cdot \vec{s} = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_N \end{pmatrix} \cdot \vec{s} \quad (3.12)$$

Das Gleichungssystem ist dann allerdings überbestimmt, die Lösung wird uneindeutig und kann nach unterschiedlichen Kriterien optimiert werden.

Schätzungsmethode von \vec{s} Der Parametervektor \vec{s} kann mit der Methode der kleinsten Quadrate geschätzt werden. Diese minimiert den Betrag des Fehlervektors \vec{e} :

$$\|\vec{e}\| = \|\vec{b} - \mathbf{M}\vec{s}\| \rightarrow \min$$

Diese Gleichung wird quadriert und das Minimum

$$\min(\|\vec{e}\|^2) = \min((\vec{b} - \mathbf{M}\vec{s})^2)$$

erhält man durch Ableiten:

$$\begin{aligned}\frac{\partial \vec{e}}{\partial \vec{s}} &= 0 \\ \Rightarrow \mathbf{M}^T \vec{b} - \mathbf{M}^T \mathbf{M} \vec{s} &= 0 \\ \vec{s} &= (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \vec{b}\end{aligned}$$

Es ist zu erwähnen, dass die mit \vec{s} parametrisierte 3D-Transformation nur in eine Richtung möglich ist (3D-Welt \rightarrow 2D-Bild). In der anderen Richtung ist nur eine 2D \rightarrow 2D-Abbildung möglich. Wenn man sich beispielsweise einen dreidimensionalen Strahl vorstellt, der eine Ebene schneidet, so ist der Punkt der Ebene eindeutig bestimmt. Geht man jedoch von dem Punkt der Ebene aus, so kann man nicht auf die Richtung des Strahles rückschließen.

2D-Transformation Ein Sonderfall ergibt sich, wenn die 3. Dimension des Weltkoordinatensystems anwendungsspezifisch nicht von Bedeutung ist. So kann $z = 0$ angenommen werden. Gleichung (3.8) vereinfacht sich dann zu:

$$\begin{pmatrix} m' \\ n' \\ s \end{pmatrix} = \begin{pmatrix} a'_{11} & a'_{12} & 0 & a'_{14} \\ a'_{21} & a'_{22} & 0 & a'_{24} \\ a'_{31} & a'_{32} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3.13)$$

Dabei sind $a'_{13} = 0$, $a'_{23} = 0$ und $a'_{33} = 0$, weil $z = 0$. Somit müssen für diesen Sonderfall nur noch 8 Parameter geschätzt werden.

Die Bildkoordinaten m, n lassen sich hier vereinfacht schreiben (vgl. Gleichung (3.9) und (3.10)):

$$\begin{aligned}m &= \frac{p}{r} = \frac{a'_{11} \cdot x + a'_{12} \cdot y + a'_{14}}{a'_{31} \cdot x + a'_{32} \cdot y + 1} \\ n &= \frac{q}{r} = \frac{a'_{21} \cdot x + a'_{22} \cdot y + a'_{24}}{a'_{31} \cdot x + a'_{32} \cdot y + 1}\end{aligned}$$

Man kann die Gleichungen umstellen...

$$\begin{aligned}m \cdot a'_{31} \cdot x + m \cdot a'_{32} \cdot y + m &= a'_{11} \cdot x + a'_{12} \cdot y + a'_{14} \\ n \cdot a'_{31} x + n \cdot a'_{32} \cdot y + n &= a'_{21} \cdot x + a'_{22} \cdot y + a'_{24} \\ (m \cdot a'_{31} - a'_{11}) \cdot x + (m \cdot a'_{32} - a'_{12}) \cdot y &= a'_{14} - m \\ (n \cdot a'_{31} - a'_{21}) \cdot x + (n \cdot a'_{32} - a'_{22}) \cdot y &= a'_{24} - n\end{aligned}$$

...und als Matrix schreiben:

$$\underbrace{\begin{pmatrix} a'_{11} - m \cdot a'_{31} & a'_{12} - m \cdot a'_{32} \\ a'_{21} - n \cdot a'_{31} & a'_{22} - n \cdot a'_{32} \end{pmatrix}}_{\mathbf{G}} \cdot \underbrace{\begin{pmatrix} x \\ y \end{pmatrix}}_{\vec{x}} = \underbrace{\begin{pmatrix} m - a'_{14} \\ n - a'_{24} \end{pmatrix}}_{\vec{d}}$$

$$\vec{x} = \mathbf{G}^{-1} \cdot \vec{d}$$

3.2 Das analoge Videosignal

Ein Signal wird als analog bezeichnet, wenn es die abzubildende physikalische Größe in ihrer ursprünglichen Form wiedergibt (im Gegensatz zum digitalen Signal). Die technische Realisierung beinhaltet jedoch eine Diskretisierung dieses Signals, welche in der Regel eine Beschränkung der Auflösung mit sich bringt.

3.2.1 Die Auflösung des Bildes

Für die Auflösung des Bildes stellt sich die Frage, wie groß der zulässige Unschärfbereich (siehe Abschnitt 3.1.2) sein kann. Dazu betrachtet man die örtliche Abtastfrequenz: $\frac{1}{\Delta x}$ und $\frac{1}{\Delta y}$. Für das Abtasttheorem $f_{x,\max} < \frac{2}{\Delta x}$ und $f_{y,\max} < \frac{2}{\Delta y}$ muss vor der Abbildung optisch gefiltert werden. Die Realisierung ist jedoch sehr aufwendig und wird weggelassen, was Aliasing-Effekte (wie Treppenstrukturen) hervorruft.

Der Kell-Faktor Um die Aliasing-Effekte zu vermindern, muss der Zuschauer einen Abstand zum Wiedergabegerät einnehmen, der größer ist als der erforderliche Abstand, bei dem zwei Zeilen des Bildes gerade nicht mehr unterschieden werden können.

Experimentell wurde der Abstandsfaktor von *Kell* 1934 durch Versuchsreihen an Testpersonen auf 1,5 festgelegt (der Kellfaktor ist der Kehrwert davon : $k \approx 0,67$). Somit können von 100 Zeilen nur noch 67 Zeilen aufgelöst werden.

Die Zeilenanzahl Für die Berechnung der erforderlichen Zeilenanzahl sind zwei Werte von Bedeutung (siehe Abbildung 3.6):

1. Der vertikale Sehwinkel $\alpha = 10^\circ$, mit dem man ein Objekt bequem erfassen kann.
2. Der minimale aufgelöste Sehwinkel $\beta = 1' = \left(\frac{1}{60}\right)^\circ$, der durch die maximale Auflösung des Auges bestimmt wird.

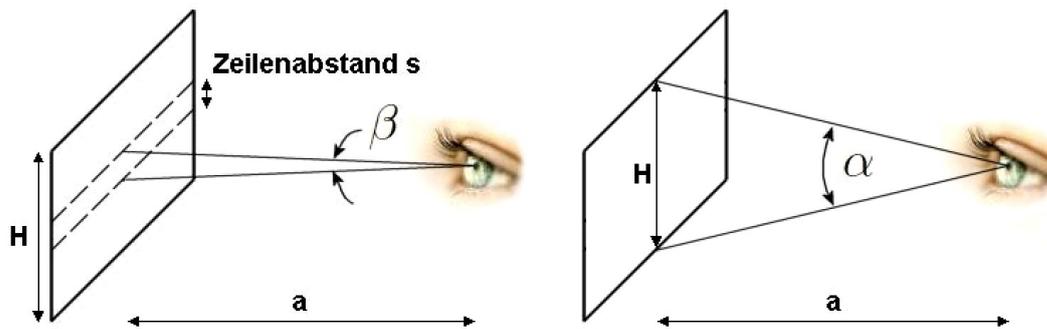


Abbildung 3.6: **Links:** Kleinster Sehwinkel β , bestimmt durch die maximale Auflösung des Auges **Rechts:** Vertikaler Sehwinkel α , mit dem man ein Objekt bequem erfassen kann.

$$s = \frac{H}{Z} \quad \rightarrow \quad Z = \frac{H}{s}$$

$$\tan\left(\frac{\beta}{2}\right) = \frac{s}{2a}$$

$$\tan\left(\frac{\alpha}{2}\right) = \frac{H}{2a}$$

$$Z = \frac{\tan\left(\frac{\alpha}{2}\right)}{\tan\left(\frac{\beta}{2}\right)} = 601 \text{ Zeilen}$$

H	...	Bildhöhe
Z	...	Zeilenanzahl
s	...	Zeilenabstand
a	...	Betrachtungsabstand

In der EU wurden 625 Zeilen als Standard festgelegt.

Die Bildanzahl Es sind ca. $20 \frac{\text{Bilder}}{s}$ nötig, damit die Bewegung im Bild als zusammenhängend empfunden wird (im Kino werden $24 \frac{\text{Bilder}}{s}$ genutzt). Da der Bildwechsel mit Dunkelpausen verbunden ist, würde starkes Flimmern erscheinen, so dass man die Bildwechselfrequenz auf 50 Hz erhöhen wollte. Nachteilig ist jedoch, dass dabei die doppelte Signalbandbreite nötig ist. Deswegen entschied man sich für das *Zeilensprungverfahren* (engl. Interlace).

Das Bild wird dabei in zwei Halbbilder zerlegt, mit je 312,5 Zeilen, die ineinander verkämmt sind. Das erste Halbbild enthält die ungeraden und das zweite Halbbild die geraden Zeilen. Somit hat man die Bildwiederholffrequenz auf $25 \frac{\text{Bilder}}{s}$ beschränkt. Abbildung 3.7 zeigt den Verlauf des Elektronenstrahls bei den zwei Halbbildern des Zeilensprungverfahrens.

3.2.2 Das Bildsignal

Das zu einem Zeitpunkt komplett vorliegende Bild wird seriell Zeile für Zeile ausgelesen, was mit einer Parallel-Seriell-Wandlung geschieht. Früher beruhte die Darstellung

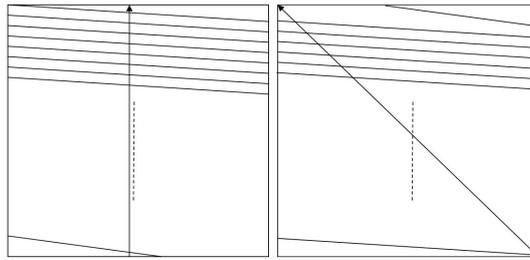


Abbildung 3.7: Zeilensprungverfahren: **Links:** Erstes Halbbild **Rechts:** Zweites Halbbild.

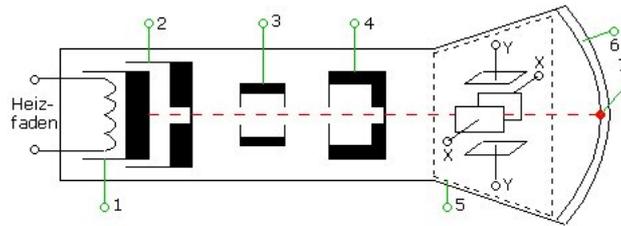


Abbildung 3.8: (aus Elektronik-Kompendium [11]) Braunsche Röhre:
 1.Kathode 2.Wehneltzylinder 3.Elektronen-Optik 4.Anode
 5.Nachbeschleunigungsanode(bis 1,5 kV) 6.Leuchtschicht
 7.Brennpunkt

auf der Braunschen Röhre mittels Elektronenstrahl. Heutzutage verwendet man CCD-Wandler, die ein Feld von Halbleiterbauelementen darstellen, ein analoges Schieberegister.

Braunsche Röhre Die Braunsche Röhre (Elektronenstrahlröhre) dient zur Erzeugung, Bündelung, Ablenkung und Beschleunigung von Elektronen. Diese haben eine geringe Masse und können unter Einfluss von Feldern sehr stark beschleunigt werden. In Abbildung 3.8 haben die einzelnen Bauteile die folgende Bedeutung:

1. **Kathode** An der Kathode liegt eine Spannung von -200 ... -800 Volt. Sie liefert die Elektronen. Durch ein Heizelement wird der Elektronenausritt erhöht.
2. **Wehneltzylinder** Mittels dieses Zylinders lässt sich die Intensität, sprich Helligkeit des Elektronenstrahls beeinflussen.
3. **Elektronen-Optik** Ablenkung der Elektronen im Elektrischen Feld. Einstellung für die Schärfe des darzustellenden Elektronenstrahls.
4. **Anode** Die Anode liegt an einer Spannung von +100 ... +200 Volt und zieht die Elektronen an.
5. **X-Platten** Zur horizontalen Ablenkung.
6. **Y-Platten** Zur vertikalen Ablenkung.

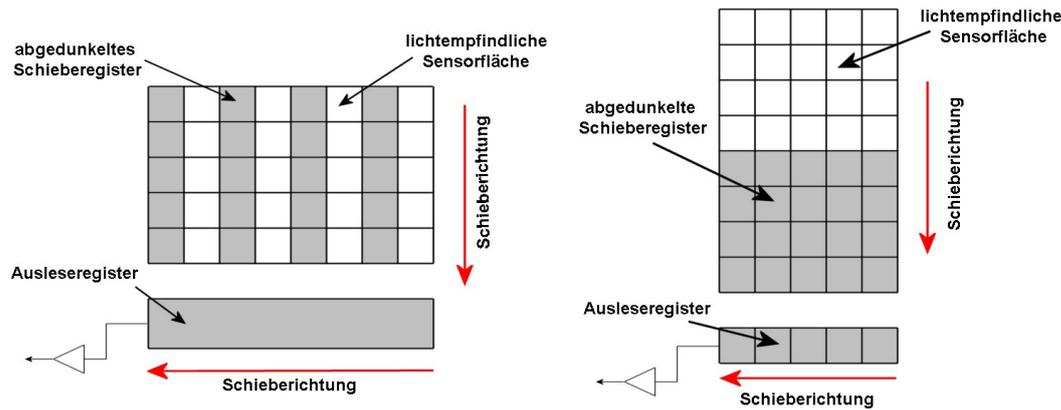


Abbildung 3.9: (Abb. nach www.ccd-sensor.de [12]) CCD-Varianten:
Links: Interline Transfer **Rechts:** Frame Transfer.

CCD-Wandler CCD steht für Charge Coupled Device (ladungsgekoppeltes Bauelement). Die CCD-Sensoren bestehen meistens aus einem ein- oder zweidimensionalen Array von Speicherelementen. Anwendungsgebiete sind vor allem Videokameras, Scanner und Digitalkameras.

Es haben sich hauptsächlich zwei Konzepte durchgesetzt (siehe Abbildung 3.9 und vgl. www.ccd-sensor.de [12]):

1. Interline CCDs
2. Frame-Transfer CCDs

Die *Interline-Sensoren* bestehen aus einem streifenförmigen Muster mit lichtempfindlichen Sensorflächen und lichtundurchlässigen Schieberegistern. Die in den lichtempfindlichen Bereichen aufgenommenen Ladungen werden parallel in die angrenzenden vertikalen Schieberegister übernommen. Daraufhin werden die Ladungen in den vertikalen Schieberegistern zeilenweise in das horizontale Ausleseschieberegister geschoben. Von dort werden sie seriell an einem Verstärker in Spannungssignale umgewandelt und ausgelesen. Die *Frame-Transfer-Sensoren* haben die Sensorflächen und den Speicherbereich blockförmig übereinander angeordnet. Es werden hier die Ladungspakete in der Belichtungsphase komplett übernommen und in den Speicherbereich übergeben.

Die Austastlücken Es gibt zwei Arten von Austastlücken: vertikale und horizontale Austastlücken.

1. **Vertikale Austastlücken** Es werden $1,6 \text{ ms}$ für den Strahlrücksprung nach der Beendigung eines Halbbildes reserviert. Für $25 \frac{\text{Bilder}}{\text{s}}$ und 625 Zeilen entsprechen die $1,6 \text{ ms}$ der Dauer von 25 Zeilen. Deswegen sind im europäischen Fernsehsystem nur 575 Zeilen aktiv.

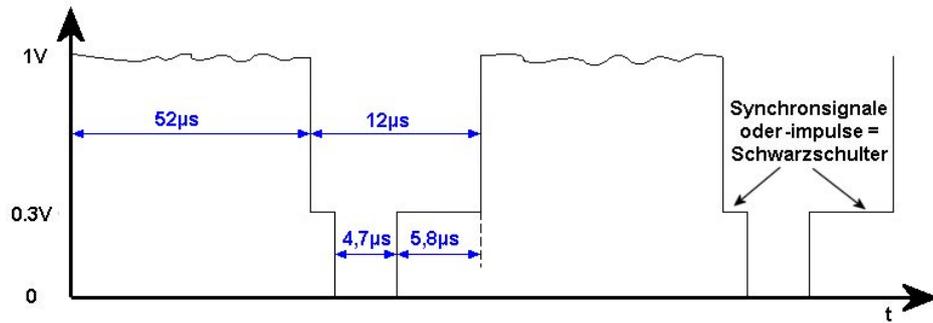


Abbildung 3.10: Horizontale Austastlücken: Für jede Zeile sind $52 \mu s$ nötig, dann erfolgt ein Zeilenrücksprung (horizontale Austastlücke), welcher $12 \mu s$ dauert.

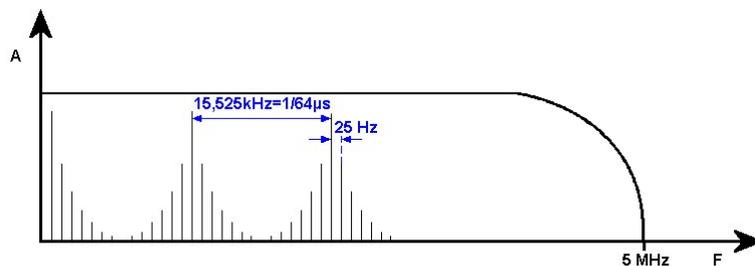


Abbildung 3.11: Das Spektrum eines schwarzen Signals für europäischen Standard.

- Horizontale Austastlücken** Für $40 ms$ Bilddauer und 625 Zeilen hat man eine Zeilendauer von $64 \mu s$, davon sind $12 \mu s$ für den horizontalen Strahlenrücksprung.

Abbildung 3.10 veranschaulicht die horizontale Austastlücke anhand eines Diagramms.

Die Videogrenzfrequenz Die Videobandbreite lässt sich folgendermaßen bestimmen:

- 575 (aktive)Zeilen $\cdot 0.67$ (Kellfaktor) = 385 Zeilen.
- $\frac{385}{2} \approx 193$ (Zeilenpaare) (Mindestens zwei Zeilen sind notwendig, um die kleinste Periode darzustellen).
- $193 \cdot \frac{4}{3} = 257$ (Spaltenpaare) (Seitenverhältnis).
- Die Spaltenpaare müssen in $52 \mu s$ abgelesen werden $\rightarrow \frac{257}{52 \mu s} = \underline{\underline{4,94 MHz}}$

In den europäischen Standardfernsehsystemen wurde die Bandbreite auf $5 MHz$ festgelegt. Das Spektrum eines schwarzen Signals ist in Abbildung 3.11 gezeigt. Andere Standards sind zum Beispiel:

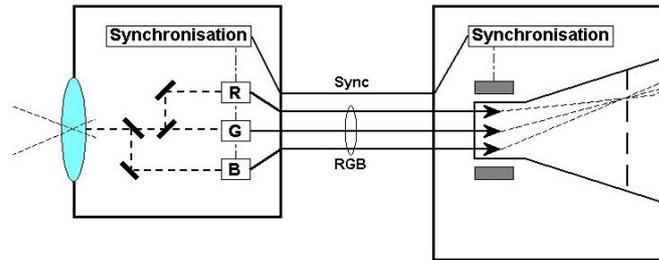


Abbildung 3.12: (Abb. aus dem www [13]) Aufteilung des Lichtes in die RGB-Komponenten und empfängerseitige Nachmischung.

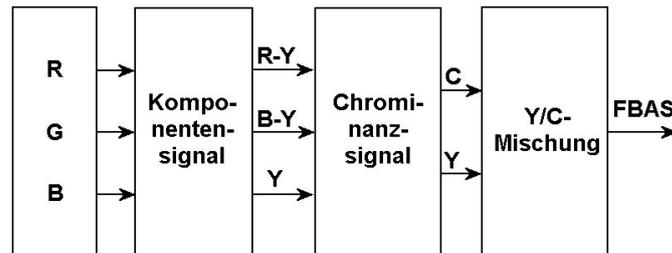


Abbildung 3.13: Verarbeitungskette vom RGB-Signal zum FBAS-Signal.

- USA: 525 Zeilen, 60 Hz Halbbildfrequenz, 4,2 MHz Bandbreite
- Frankreich: ähnlich wie in Deutschland, nur mit 6 MHz Bandbreite

3.2.3 Das Farbsignal

Das Konzept Das Farbsignal beruht auf der Farbmischung der Grundfarben (RGB), welche mittels eines 3-CCD-Wandlers (oder ein Wandler mit 3 Subpixel) gemischt werden. Durch eine Vorfilterung wird das einfallende Licht in drei Grundfarbenlichter zerlegt und den drei Sensoren (Wandlern) zugeführt. Jeder Wandler liefert ein elektrisches Signal, das die volle Bandbreite hat (entspricht der Sensorauflösung). Empfängerseitig findet eine örtliche Farbnachmischung statt (Abbildung 3.12). In der Studioteknik werden die drei Signale ohne große Änderung im Frequenzmultiplex übertragen. Für die Anwendung im Konsumerbereich muss das zu übertragende Farbsignal zwei Forderungen erfüllen:

1. Es muss schwarz/weiß-kompatibel sein (S/W-Empfänger müssen das Signal demodulieren können)
2. Die 5 MHz-Bandbreite des Bildsignals darf nicht überschritten werden.

So muss aus den drei Signalen ein Signal erzeugt werden, das die o.g. Forderungen erfüllt, das FBAS-Signal (Abbildung 3.13).

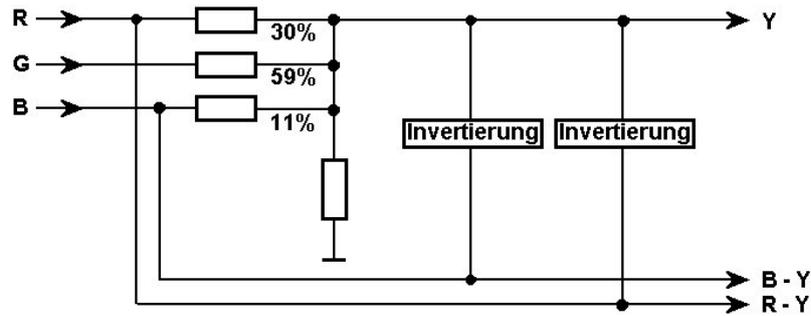


Abbildung 3.14: Entstehung des Komponentensignals aus dem RGB-Signal.

Das Komponentensignal (Y, R-Y, B-Y) Aus den RGB-Werten wird das sog. *Komponentensignal* gebildet. Es enthält die Komponente Y, die nur Helligkeitsinformationen enthält. Sie folgt aus der Transformation $RGB \rightarrow XYZ$ mit $Y = 0,3R + 0,59G + 0,11B$.

R-Y und B-Y sind die *Farbdifferenzsignale* (siehe Abbildung 3.14). Bei der S/W-Übertragung werden die beiden Differenzsignale weggelassen und nur das Luminanzsignal wird übertragen. In der Praxis wird die Bandbreite der Farbdifferenzsignale auf 1,3 bis 3,5 MHz reduziert mit dem Argument, dass das Auflösungsvermögen des menschlichen Auges für feine Farbdetails gering ist. Trotzdem haben diese Signale eine hohe Qualität und werden häufig in der professionellen Studioteknik verwendet, außerdem gelten sie als das Ausgangsformat zur Gewinnung eines hochwertigen Digitalsignals.

Vorteile gegenüber RGB	Nachteile gegenüber FBAS
S/W-kompatibel	getrennte Signalführung erforderlich
kleinere Bandbreite	doppelte Bandbreite

Aus modulationstechnischen Gründen werden die Differenzsignalamplituden vor dem nächsten Schritt mit den Faktoren $C_u = 0,493$ und $C_v = 0,877$ reduziert, so dass $V = 0,877(R - Y)$ und $U = 0,493(B - Y)$ (für die QAM (siehe unten)) entstehen. Daraus folgt für RGB:

$$\begin{aligned}
 R &= \frac{V}{C_v} + Y \\
 B &= \frac{U}{C_u} + Y \\
 G &= Y - \frac{0,3V}{0,59C_v} - \frac{0,11U}{0,59C_u}
 \end{aligned}$$

Das Chrominanzsignal Aus den Signalen U und V wird mit Hilfe der *Quadraturamplitudenmodulation* (QAM) das *Chrominanzsignal* erzeugt (Abbildung 3.15). Bei

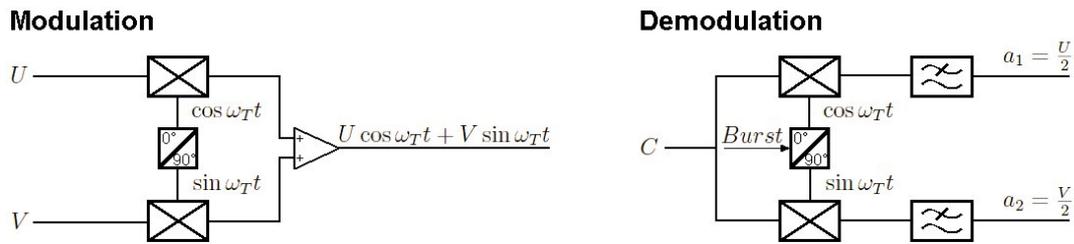


Abbildung 3.15: Modulation und Demodulation des Chrominanzsignals mittels QAM.

der Demodulation errechnet sich a_1 aus dem Chrominanzsignal folgendermaßen:

$$\begin{aligned}
 a_1 &= (U \cdot \cos(\omega_T t) + V \cdot \sin(\omega_T t)) \cdot \cos(\omega_T t) \\
 &= U \cdot \cos^2(\omega_T t) + V \cdot \sin(\omega_T t) \cdot \cos(\omega_T t) \\
 &= \frac{U}{2} \cdot (\cos(0) + \cos(2\omega_T t)) + \frac{V}{2} \cdot (\sin(0) + \sin(2\omega_T t)) \quad (\text{Additionstheoreme}) \\
 &= \frac{U}{2} + \underbrace{\frac{U}{2} \cos(2\omega_T t) + \frac{V}{2} \sin(2\omega_T t)}_{\text{entfällt durch Tiefpass}} \\
 &= \underline{\underline{\frac{U}{2}}}
 \end{aligned}$$

Analog lässt sich a_2 herleiten zu $a_2 = \frac{V}{2}$.

Bei diesem Verfahren wird das Signal mit unterdrücktem Träger moduliert. Deswegen ist für die Demodulation ein Hilfsträger nötig, der mit übertragen werden muss, der sog. *Burst* (siehe unten).

Wegen der momentanen Phasenverschiebung von 90° kann man den Wert des Chrominanzsignals mit $C = \sqrt{U^2 + V^2}$ angeben mit dem dazugehörigen Winkel im (U,V)-Koordinatensystem $M = \arctan \frac{V}{U}$ (Abbildung 3.16). Hierbei sei das sogenannte *Vektorskop* erwähnt. Dies ist ein Gerät, das speziell zur Überprüfung eines Chrominanzsignals geeignet ist und nach dem Prinzip der (U,V)-Koordinatendarstellung funktioniert.

Das FBAS-Signal Die Zusammenführung des Luminanz- und Chrominanzsignals ergibt das FBAS-Signal². Damit das Luminanzsignal möglichst wenig gestört wird, muss die Farbträgerfrequenz relativ hoch sein (feine Strukturen stören weniger als Grobe). Das Chrominanzsignal hat dieselbe vertikal- und horizontalfrequente Rasterung, so kann man das Y-Signal und das Chrominanzsignal mit einer „Frequenzverkämmung“ übertragen. Als Farbträgerfrequenz kann zum Beispiel ein ungeradzahliges Vielfaches der halben Zeilenfrequenz gewählt werden (siehe Abbildung 3.17).

²FBAS...Farb Bild Austast Synchron

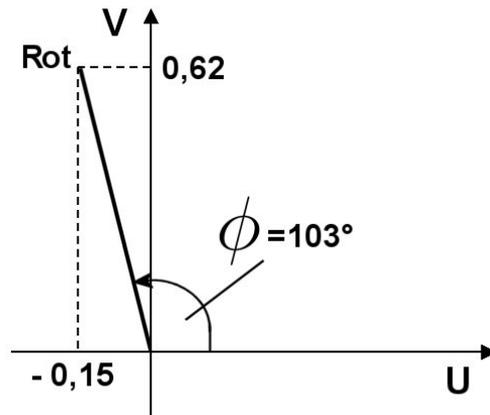


Abbildung 3.16: Darstellung der Chrominanz in Polarkoordinaten, da U und V senkrecht aufeinander stehen. Dargestellt ist Rot mit 100% Farbsättigung und 100% Helligkeit: $R=1$ $B=G=0 \Rightarrow Y=0,3 \Rightarrow U=-0,15$ $V=0,62$ $C=0,63$ und $\phi=103^\circ$.

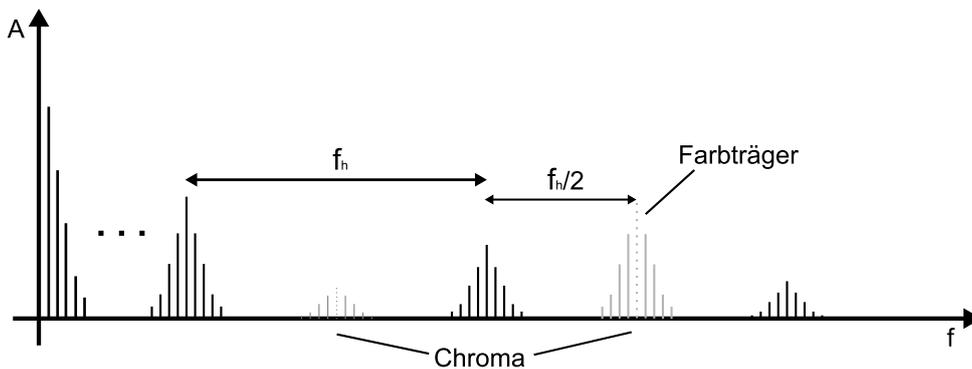


Abbildung 3.17: Darstellung der Farbträgerfrequenz mit der halben Zeilenfrequenz

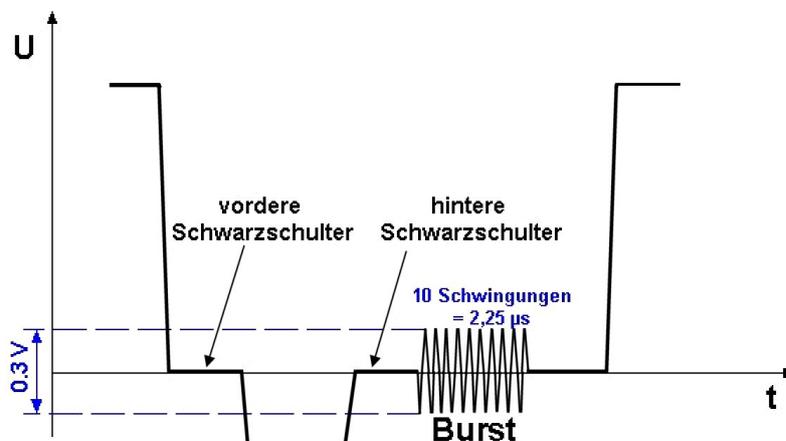


Abbildung 3.18: (Abb. aus dem wvw [13]) Der Burst als Synchronisierungssignal innerhalb der rechten Schwarzschulter einer aktiven Zeile.

Der Burst Zur phasenrichtigen Verkopplung von Sender und Empfänger wird in jeder Zeile im Bereich der hinteren Schwarzscherle ein Synchronisierungssignal, der sogenannte Burst, eingeführt (Abbildung 3.18). Dieser besteht aus 10 Schwingungen der Farbträgerfrequenz mit definierter Phasenlage. Er besitzt bei PAL einen Spannungswert von $0,3V^{SS}$ und ist $2,25\mu s$ lang. Der Burst ermöglicht die Synchronisierung von Phase und Frequenz des Senders mit dem Empfänger. Desweiteren dient der Burst als Referenz für die Regelung der Farbamplitude.

3.2.4 Farbfernsehnormen

PAL, SECAM und NTSC sind die heute gebräuchlichen Farbfernsehnormen. In Deutschland wird das PAL-Format benutzt. Französische Schwarzweiß-Filme sind zwar in SECAM, können aber ohne weiteres auf den hiesigen PAL-Geräten abgespielt werden. Französische Farbfilme können ebenfalls abgespielt werden, jedoch nur in Schwarzweiß-Qualität. Amerikanische NTSC-Filme sind auf PAL-Geräten nicht abspielbar. Dazu ist zusätzliche Technik nötig, die sich in einigen Videoplayern als Pseudo-PAL finden lässt. Bei den europäischen DVD-Playern spielen die verschiedenen Farbfernsehnormen keine Rolle mehr, SECAM wird jedoch nicht mehr verwendet.

NTSC NTSC³ ist eine vor allem in Nord- und Südamerika sowie Japan gebräuchliche analoge Fernsehnorm. Ähnlich wie bei PAL werden für die Farbübertragung die Differenzsignale (Rot-Helligkeit und Blau-Helligkeit) mittels QAM genutzt. Auch bei NTSC wird der Farbträger nicht mit übertragen, so dass in der rechten Schwarzscherle wieder der Burst nötig ist. Folgende Daten kennzeichnen die Farbfernsehnorm:

- 525 Zeilen, davon maximal 486 sichtbar
- Farbträger: 3579,545 Hz
- Bandbreite: 4,2 MHz
- Audioträger: 4,5 MHz
- negative Amplitudenmodulation für das Bild
- Frequenzmodulation für den Ton
- Zeilensprungverfahren(Interlace)
- $29,97 \frac{\text{Vollbilder}}{s}$

³NTSC...National Television Systems Committee

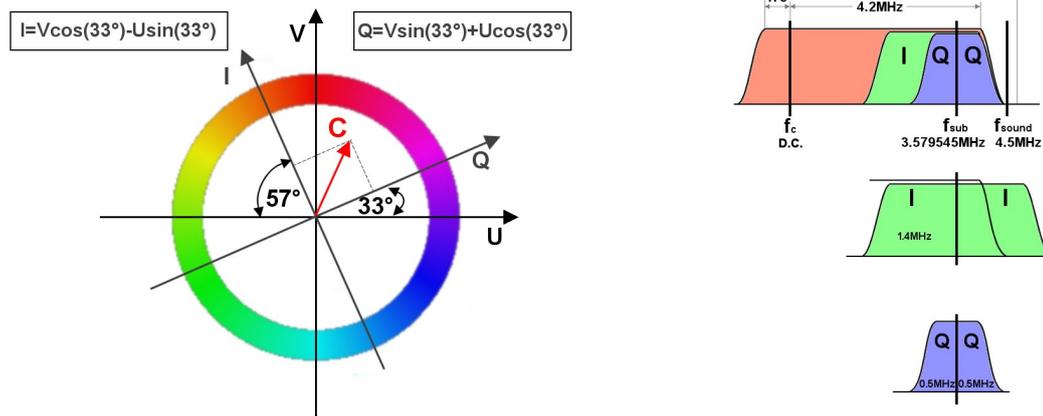


Abbildung 3.19: Links: Darstellung der Farbdifferenzkomponenten U&V und I&Q und deren Konvertierungsformel. **Rechts:** (Abb. aus www [14]) Spektrum der bandbegrenzten Farbdifferenzkomponenten I/Q.

Im digitalen Bereich beschreibt NTSC lediglich die Auflösung 720×480 Bildpunkte und die Bildanzahl von $29,97 \frac{\text{Vollbilder}}{\text{s}}$. Alle anderen oben genannten Eigenschaften sind dafür unbedeutend.

Bei NTSC werden nicht die Signale U und V genutzt, sondern die um 33° geneigten Farbdifferenzkomponenten I und Q (siehe Abbildung 3.19). I und Q liegen in der Richtung der maximalen bzw. minimalen Kontrastempfindlichkeit des menschlichen Auges. Phasenverschiebungen, die in der gesamten Übertragungskette auftreten können, wirken sich als Farbtonfehler aus, da die Farbe mit der Phase beschrieben wird. Somit haben NTSC-Fernseher einen zusätzlichen Regler (engl. „hue control“ oder „tint control“), der eventuelle dauerhafte Phasenfehler ausgleichen soll. Heutzutage geschieht dies oft automatisch, da „intelligente“ Schaltkreise zum Einsatz kommen. Die Farbfehler beim NTSC-Verfahren führten zur scherzhaften Deutung der Abkürzung als „Never The Same Color“.

SECAM SECAM⁴ ist eine vor allem in Frankreich und Osteuropa gebräuchliche analoge Fernsehnorm für die Farbübertragung und wurde von Henri de France entwickelt. Es umgeht das kritische QAM und nutzt für die Farbübertragung die Frequenzmodulation. Dabei werden zeilenweise alternierend die Farbdifferenzsignale Rot-Helligkeit und Blau-Helligkeit übertragen (siehe Abbildung 3.20). Die aktuelle Zeile ergibt sich aus dem Mittelwert der Zeilen (n) und (n-1). Für die Farbträger werden zwei verschiedene Trägerfrequenzen genutzt: für (R-Y) ist $f=4,25\text{MHz}$ und für (B-Y) ist $f=4,4\text{MHz}$.

Die Entwicklung von SECAM in Frankreich war politisch motiviert, um die Industrie von Importen zu schützen. Auch die Verbreitung in den ehemaligen Ostblockstaaten

⁴SECAM... Séquentiel Couleur avec Mémoire

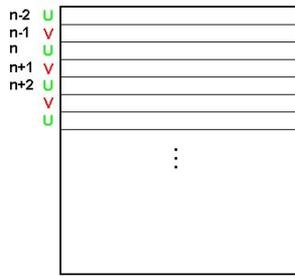


Abbildung 3.20: Zeilenweise alternierende Übertragung der U/V-Komponenten bei SECAM.

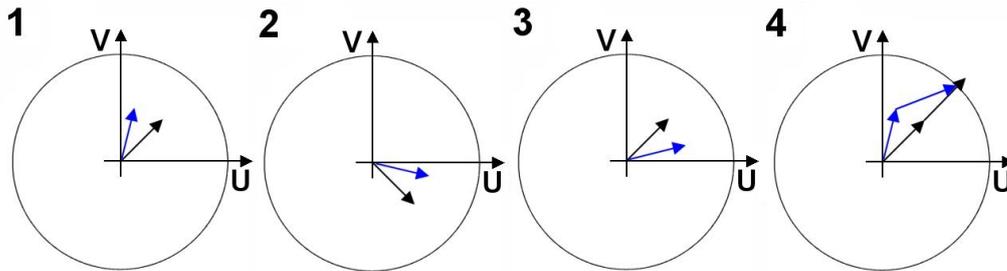


Abbildung 3.21: (Abb. nach wikipedia [15]) Phasenfehler bei PAL: **1** Zeile n , schwarzer Originalzeiger, blauer Zeiger mit Phasenfehler **2** Zeile $n+1$, Phase invertiert **3** Lage der Zeiger nach Rückdrehung **4** vektorielle Addition der schwarzen Originalzeiger bzw. der blauen phasenfehlerbehafteten Zeiger. Es ergibt sich ein geringer Farbsättigungsfehler, der Farbton ist gleich.

war politisch begründet, da sich Frankreich in der Annäherung zu diesen Staaten befand. Viele Oststaaten haben auf PAL umgestellt bzw. befinden sich noch in der Umstellung. SECAM ist allgemein eher rückläufig, in der digitalen Welt der DVD spielt es keine Rolle mehr.

Das Verfahren hat den Nachteil, dass es bei der FM ein starkes Übersprechen im Luminanzkanal geben kann. Folgende Daten kennzeichnen die Farbfernsehnorm:

- Zeilenanzahl: 625
- Farbdifferenzsignalfrequenzen: 4,40625 MHz und 4,250 MHz
- Variationen: SECAM B,G,H ; SECAM D,K,K1,L ; MESECAM
- Bandbreite: 5,0 MHz (bei B,G,H) und 6,0 MHz (bei D,K,K1,L)
- Audioträger: 5,5 MHz (bei B,G,H) und 6,5 MHz (bei D,K,K1,L)

PAL PAL⁵ ist vor allem in Europa gebräuchlich, wird aber auch in Australien und in vielen afrikanischen und asiatischen Ländern verwendet. Es beruht auf dem NTSC-

⁵PAL...Phase Alternating Line

Verfahren, umgeht aber die Empfindlichkeit gegenüber Phasenschwankungen, was mit erheblichem Mehraufwand bei der Schaltungsrealisierung erkauft wird.

Folgende Daten kennzeichnen die Farbfernsehnorm:

- Zeilenanzahl: 625 bzw. 525 bei PAL-M
- Farbsignalträger: 4,43362 MHz
- Variationen: PAL B,G,H ; PAL I ; PAL D ; PAL N ; PAL M
- Bandbreite: 5,0 MHz (bei B,G,H), 5,5 MHz (bei I), 6,0 MHz (bei D), 4,2 MHz (bei M,N)
- Audioträger: 5,5 MHz (bei B,G,H), 6,0 MHz (bei I), 6,6 MHz (bei D), 4,5 MHz (bei M,N)

Auch bei PAL wird die Quadraturamplitudenmodulation für U und V genutzt. Die V-Komponente wird zeilenweise alternierend invertiert (daher auch die Bezeichnung *Phase Alternating Line*). Diese Umschaltung wirkt wie eine zusätzliche Modulation mit halber Zeilenfrequenz. Die PAL-Farbträgerfrequenz wird daher mit einem Viertelzeilen-Offset berechnet (siehe Abbildung 3.22):

$$f_{\text{Träger}} = (n - 1/4) \cdot f_h = 4,43361875 \text{ MHz} \quad (3.14)$$

Im Empfänger werden, im Gegensatz zu NTSC, Farbtonfehler (häufig Phasenfehler) durch Mittelwertbildung über je zwei Zeilen automatisch kompensiert und in einen geringen Farbsättigungsfehler umgewandelt. Das funktioniert aber nur, wenn die Farbe und der Farbtonfehler zwischen beiden Zeilen konstant bleibt. Farbsättigungsfehler fallen dem Menschen weitaus weniger auf als Farbtonfehler und hier steckt auch der entscheidende Vorteil von PAL zu NTSC. Abbildung 3.21 verdeutlicht an einem Beispiel das Prinzip.

Vergleich der Farbfernsehnormen Die Abbildung 3.23 zeigt ein Diagramm, in dem die Luminanz und die Farbdifferenzkomponenten für die einzelnen Farbfernsehnormen in Abhängigkeit zur Zeit dargestellt sind.

Aufgabe

geg.: Mit einem Videosystem wird eine Farbfläche generiert, so dass gesättigtes gelb mit einem Videopegel von 50% entsteht.

ges.: Berechnen Sie die Pegelwerte aller Signalanteile in den folgenden Formen:

- a) RGB b) YUV c) YC (PAL) d) FBAS (PAL)

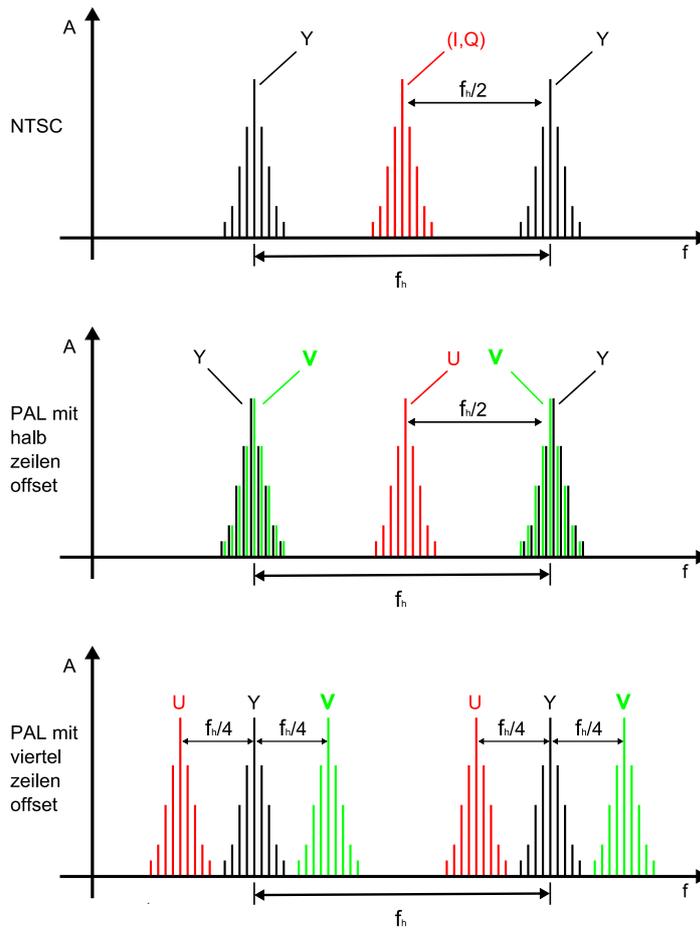


Abbildung 3.22: Vergleich

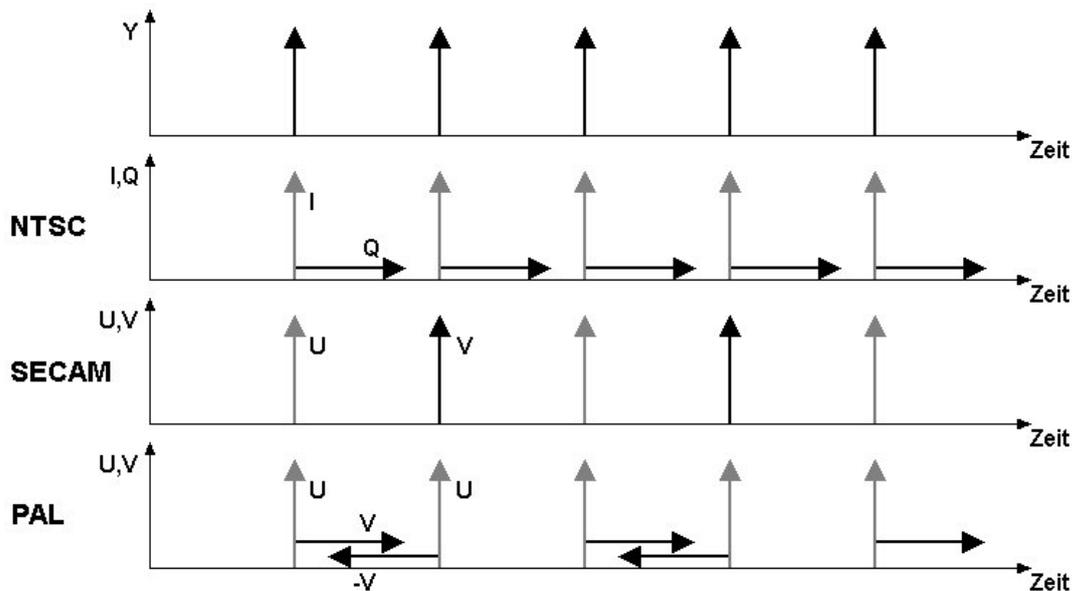


Abbildung 3.23: Vergleich der Farbfernsehnormen mittels Darstellung der Luminanz und der Farbdifferenzkomponenten der einzelnen Farbfernsehnormen.

Lösung:

$$100\% = 0,7V \quad \text{Gelb} = \text{Rot} + \text{Grün}$$

a) $R = 0,5 \cdot 0,7V \quad \underline{\underline{R = 0,35V}}$

$$G = 0,5 \cdot 0,7V \quad \underline{\underline{G = 0,35V}}$$

$$\underline{\underline{B = 0}}$$

b) $Y = 0,3R + 0,59G + 0,11B \quad \underline{\underline{Y = 0,31V}}$

$$R - Y = 0,039 \quad \Rightarrow \quad V = 0,877(R - Y) \quad \underline{\underline{V = 0.034V}}$$

$$B - Y = -0,31 \quad \Rightarrow \quad U = 0,493(B - Y) \quad \underline{\underline{U = -0,154V}}$$

c) $C = \sqrt{U^2 + V^2} \quad \underline{\underline{C = 0,157V}}$

d) $\underline{\underline{FBAS = Y + C = 0.469V}}$

3.3 Digitalisierung

Ein digitales Videoformat in dem Sinne gibt es nicht, hier findet einfach eine Digitalisierung des analogen Videosignals statt. Durch die zeilenweise Abtastung des Bildes liegt eine Diskretisierung in der vertikalen Richtung bereits vor. Die Abtastung der Zeilen fordert laut Abtasttheorem die doppelte maximale Signalfrequenz. Das menschliche Auge kann höchstens 200 Graustufen unterscheiden. Deswegen kann man im einfachsten Fall die Helligkeitswerte eines Bildes mit der Quantisierungsstufenzahl von 256 realisieren, was mit 8 Bit geschieht. So ergibt sich bei einer Bandbreite von 5 MHz eine minimale Datenrate von $80 \frac{\text{MBit}}{\text{s}}$. Je weniger Bits genutzt werden, desto größer ist der Quantisierungsfehler. Für die Bestimmung des Signal-Rauschabstandes ($\frac{\text{Signal}}{\text{Noise}}$) für ein mit $m = 2^n$ digitalisiertes Videosignal gilt:

$$\frac{S}{N} = 20 \log(m\sqrt{12}) \approx (n \cdot 6 + 10) \text{dB} \approx 59 \quad (\text{bei 8Bit})$$

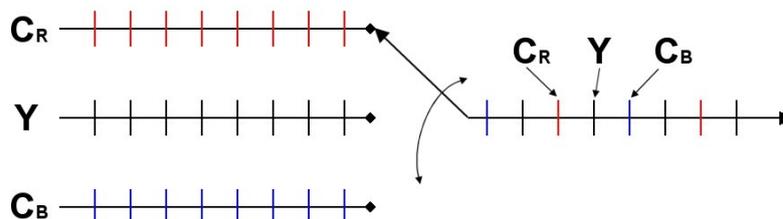


Abbildung 3.24: Bei CCIR 601 werden die Signale C_R, C_B und Y im Zeitmultiplex übertragen. Aufgrund der Häufigkeiten der Signale wird auch die Bezeichnung 4:2:2 dafür verwendet. Dies hat zur Folge, dass nebeneinanderliegende Pixel zwar unterschiedliche Y -Werte aber gemeinsame C_R - bzw. C_B -Werte haben können. Die Farbqualität ist zum 4:4:4-Format kaum unterscheidbar.

Standards für die Digitalisierung Es existieren zwei wichtige Standards für die Digitalisierung des Videosignals:

- **Das digitale FBAS-Signal**

- $f_{\text{abast}} = 4 \cdot f_{\text{farb}} = 17,7\text{MHz} = 141,6 \frac{\text{MBit}}{\text{s}}$ Datenrate bei 8 Bit
- findet in bestimmten Aufzeichnungsgeräten (D2 und D3) Anwendung
- Abtastwerte pro Zeile: $0,000052\text{s} \cdot 17700000\text{Hz} = 920 \frac{1}{\text{Zeile}}$

- **Das digitale Komponentensignal CCIR 601**

- Abtastrate des Luminanzsignals: $13,5\text{MHz} = 2,25\text{MHz} \cdot 6$
- Abtastrate der Komponentensignale ist die Hälfte: $6,75\text{MHz}$
- serielle Übertragung im Zeitmultiplex (vgl. Abbildung 3.24) \Rightarrow für 8 Bit ergibt sich die Datenrate von: $8\text{Bit}(13,5 + 2 \cdot 6,75)\text{MHz} = 216 \frac{\text{MBit}}{\text{s}}$

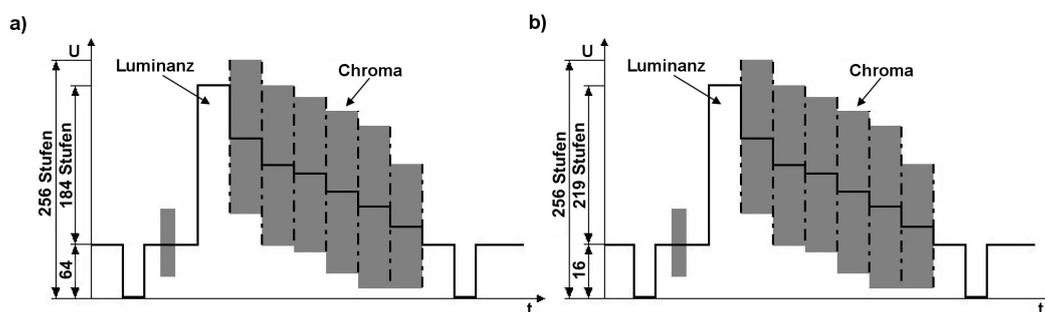


Abbildung 3.25: Quantisierungsstufen bei a) FBAS und b) CCIR.

Die Quantisierungsstufen unterscheiden sich bei den zwei Standards. Abbildung 3.25 veranschaulicht dies.

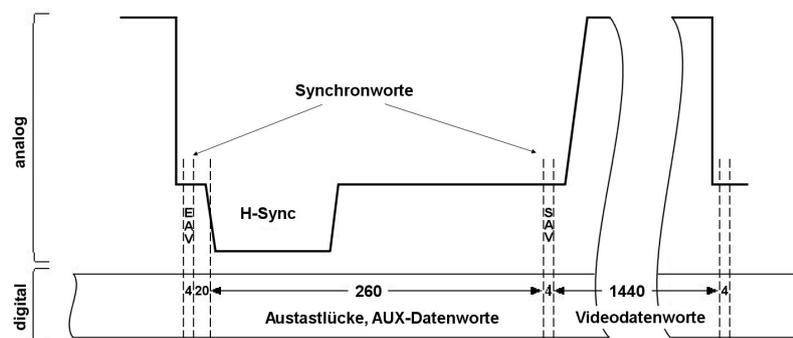


Abbildung 3.26: Zuordnung von analogen und digitalen Synchronsignalen.

Synchronsignale Auch bei dem digitalisierten Videosignal muss es Synchronsignale geben, um Sender und Empfänger zu synchronisieren. Dabei werden wie in Abbildung 3.26 4 Byte für das Signal SAV⁶ und ebenfalls 4 Byte für EAV⁷ genutzt. Die Synchronsignale haben den folgenden Aufbau:

SAV/EAV: FF 00 00 XY

Die acht Bit, welche sich hinter XY verbergen haben folgende Bedeutung:

	X				Y			
Bit Nr.	0	1	2	3	4	5	6	7
Bedeutung	Parity	Fehlerschutz			0=SAV 1=EAV	1 bei V-Lücke	0=1.Halbbild 1=2.Halbbild	1

⁶SAV...Start of Active Video

⁷EAV...End of Active Video

Kapitel 4

Charakterisierung digitalisierter Bilder

In diesem Kapitel sollen verschiedene Eigenschaften von Bildern vorgestellt werden, die vor allem statistisch begründet sind. Es werden Begriffe und deren Bedeutung für die Bildverarbeitung eingeführt, wie z.B. *Wahrscheinlichkeitsdichtefunktion*, *Histogramm*, *Mittelwert*, *Varianz* etc. Die Betrachtung des Bildes als statistische Größe ermöglicht eine statistische Aussagen über dessen Struktur zu machen. Dabei spielen auch die *Normalverteilung* und deren diskretes Analogon, die *Binomialverteilung*, eine Rolle .

4.1 Bilder als Funktionen zweier diskreter Variablen

Bilder werden als digitale Zahlenfelder (Vgl. Matrizen) mit $M \cdot N$ Bildpunkten repräsentiert. Ein Bildpunkt wird als *Pixel* (picture element) bezeichnet. Dabei wird für die zwei Dimensionen die folgenden Bezeichnungen festgelegt:

$$\begin{array}{ll} m = 0, 1, \dots, M - 1 & m \dots \text{Spaltenindex} \\ n = 0, 1, \dots, N - 1 & n \dots \text{Zeilenindex} \end{array}$$

Gerade für die Programmierung ist es wichtig, dass diese Festlegungen nicht geändert werden, damit bei der Entwicklung von Algorithmen keine Vertauschungen auftreten. Mathematisch gesehen können digitale Bilder als Funktionen zweier diskreter örtlicher

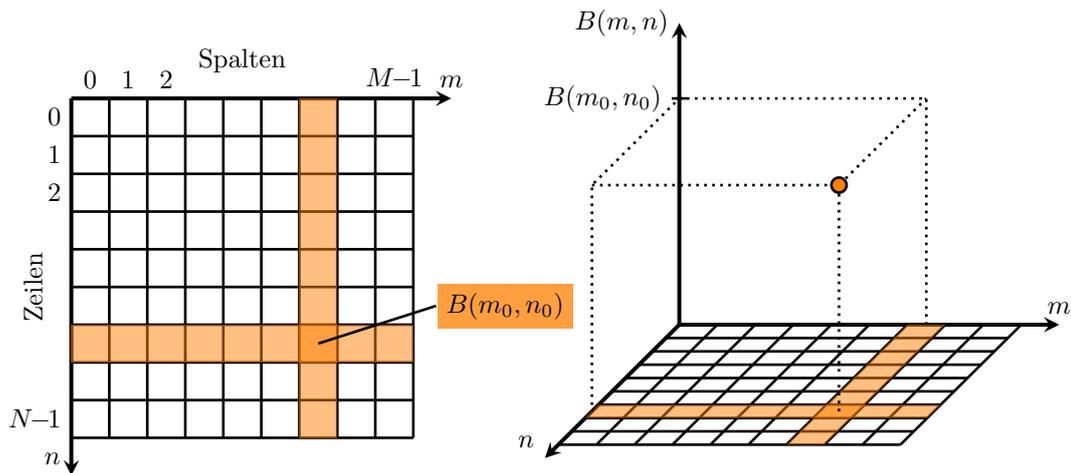


Abbildung 4.1: Links: Bilddarstellung als Funktion der Variablen $m \in \{0, M - 1\}$ und $n \in \{0, N - 1\}$. Rechts: Koordinatensystem zur pseudo-3D-grafischen Darstellung von Grauwerten eines Bildes.

Variablen (x_1, x_2) betrachtet werden. Wenn man den Grauwert mit g und das Bild mit B bezeichnet, so gilt:

$$g_{\min} \leq B(x_1, x_2) \leq g_{\max}$$

oder entsprechend des Bildkoordinatensystems:

$$g_{\min} \leq B(m, n) \leq g_{\max}$$

Bei einer 8 Bit Digitalisierung ist $[g_{\min}, g_{\max}] = [0, 255]$.

K -kanalige Bilder, wie z.B. Farbbilder mit $K = 3$, können als K -dimensionale Vektoren der K diskreten Funktionen $B(m, n)$ beschrieben werden:

$$\vec{B} = \begin{pmatrix} B_0(m, n) \\ B_1(m, n) \\ \vdots \\ B_{K-1}(m, n) \end{pmatrix}$$

4.2 Zufallsvariablen

Wahrscheinlichkeitsdichtefunktion Das Bild wird als eine Reihe von Messungen aufgefasst, wobei jedes Pixel eine Messung ist. Wegen der stochastischen Natur des Prozesses werden die gemessenen Größen (die Grauwerte) durch eine Zufallsvariable g beschrieben. So kann der beobachtete Prozess mit Hilfe der *Wahrscheinlichkeitsdichtefunktion* $f(g)$ charakterisiert werden. Für einen *homogenen*¹ Prozess lässt sich $f(g)$

¹Hierbei wird angenommen, dass die einzelnen Messungen voneinander unabhängig sind und dass die Auftrittswahrscheinlichkeit eines Grauwertes von der Position im Bild unabhängig ist.

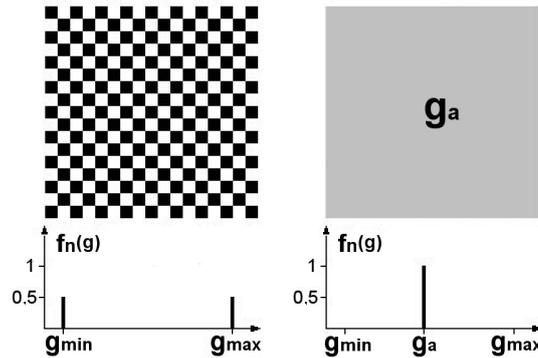


Abbildung 4.2: Links: Schwarz-weißes Schachbrettmuster enthält die Grauwerte schwarz und weiß gleich häufig: $f_n(g_{\min}) = f_n(g_{\max}) = 0,5$. Für alle Grauwerte $g \neq g_{\min}$ und $g \neq g_{\max}$ gilt: $f_n(g) = 0$. Rechts: Homogenes Bild enthält nur den Grauwert g_a . Es gilt: $f_n(g_a) = 1$ und $f_n(g) = 0$ für alle anderen Grauwerte.

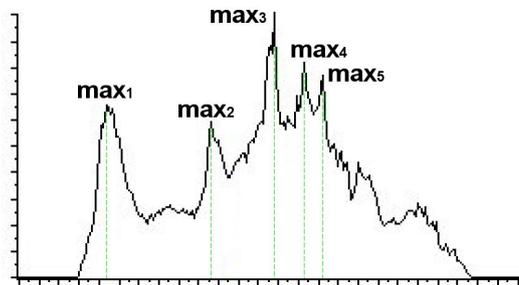


Abbildung 4.3: Anwendung des Histogramms: Bei der Segmentierung werden z.B. die Maxima der Grauwerte als Kriterium für die Segmentierung verwendet.

durch die Häufigkeitsverteilung, das sog. *Histogramm*, einfach abschätzen.

Histogramm Das Histogramm ist ein Vektor, der für jede Quantisierungsstufe ein Element aufweist. Jedes Element enthält die Anzahl der Pixel, deren Grauwert mit dem Index des Elementes übereinstimmt und es kann in tabellarischer Form (Liste) oder grafisch (Balkendiagramm) wiedergegeben werden. Das Histogramm gibt keine Auskunft über die räumliche Verteilung der Grauwerte. Durch die Normierung über die Anzahl A der Bildpunkte erhält man die *relative Häufigkeitsverteilung*

$$f_n(g) = \frac{f(g)}{A}$$

und es gilt:

$$\sum_{g_{\min}}^{g_{\max}} f_n(g) = 1 \quad \text{und} \quad 0 \leq f_n(g) \leq 1$$

Die Abbildung 4.2 zeigt das Histogramm für ein homogenes Bild und ein Schachbrettmuster.

Anwendungen:

- Segmentierung, siehe Abbildung 4.3
- Adaptive Einstellung von Filterparameter
- Bildverbesserung, Helligkeit- und Kontrastoptimierung (Punktoperation)

Mittelwert und Varianz Der Mittelwert (oder Erwartungswert) gibt Auskunft über die allgemeine Helligkeit des Bildes. Er lässt sich berechnen durch:

$$\mu = \frac{1}{MN} \sum_{m=0}^{N-1} \sum_{n=0}^{M-1} B(m, n)$$

Die Varianz (oder mittlere quadratische Abweichung) ist ein Maß für die Streuung der Messwerte (Grauwerte). Sie wird berechnet mit:

$$\begin{aligned} \sigma^2 &= \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (B(m, n) - \mu)^2 \\ &= \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (B^2(m, n) - 2\mu \cdot B(m, n) + \mu^2) \\ &= \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} B^2(m, n) - 2\mu \underbrace{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} B(m, n)}_{\mu} + \underbrace{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mu^2}_{\mu^2} \end{aligned}$$

Die Formel vereinfacht sich zu:

$$\sigma^2 = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} B^2(m, n) - \mu^2$$

Diese Werte können auch aus dem Histogramm leicht berechnet werden:

$$\mu = \sum_{g_{\min}}^{g_{\max}} g \cdot f_n(g)$$

$$\sigma^2 = \sum_{g_{\min}}^{g_{\max}} (g - \mu)^2 \cdot f_n(g)$$

Beispiele:

	Homogenes Bild mit g_a	Schachbrettmuster S/W
Mittelwert μ	$\mu = g_a$	$\mu = g_{\text{mittel}} = 127,5$ bei 256 Stufen
Varianz σ^2	$\sigma^2 = 0$	$\sigma^2 = \sigma_{\text{max}}^2 = 127,5^2 = 16.256,25$

Die zentralen Momente Zur genaueren Charakterisierung des Histogramms werden die zentralen Momente verwendet. Deren allgemeine Formulierung lautet:

$$M_n = \sum_{g_{\min}}^{g_{\max}} (g - \mu)^n \cdot f_n(g)$$

Die ersten Momente haben dabei die folgende Bedeutung (vgl. auch Abbildung 4.4):

- Das nullte Moment M_0 ist immer eins.
- Das erste Moment M_1 ist immer null.
- Das zweite Moment M_2 entspricht der Varianz.
- Das dritte Moment M_3 , die Schiefe des Histogramms, ist ein Maß für die Asymmetrie um den Mittelwert. Für eine Funktion mit gerader Symmetrie ist es null.
- Das vierte Moment M_4 kann als Maß für die Abweichung von der Normalform benutzt werden. Für die Normalform ist $\frac{M_4}{(M_2)^2} = 3$ und so kann man die Maßzahl $M_z = \frac{M_4}{(M_2)^2} - 3$ definieren.

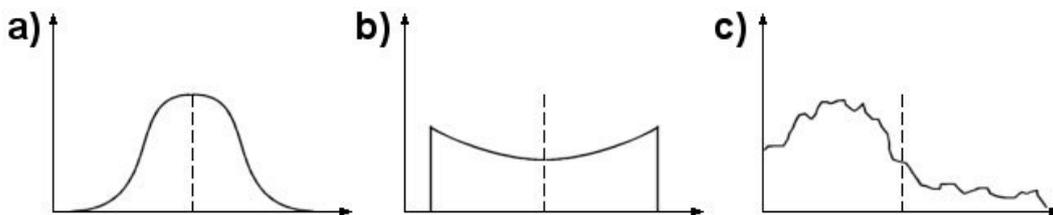


Abbildung 4.4: Beispiele für Histogramme und deren M_3 - und M_z -Momente.
a) $M_3 = 0$, $M_z = 0$ **b)** $M_3 = 0$, $M_z > 0$ **c)** $M_3 < 0$

Entropie Die Entropie ist ein Maß für den mittleren Informationsgehalt oder die Anzahl der notwendigen Bits zur Speicherung eines Bildes. Sie ist definiert durch:

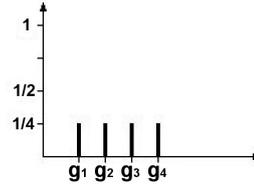
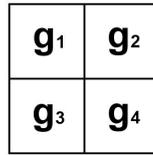
$$H = - \sum_{g_{\min}}^{g_{\max}} (f_n(g) \cdot \log_2 f_n(g))$$

Beispiele:

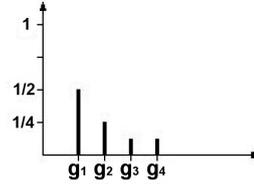
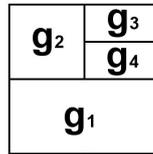
- Homogenes Bild: $H = 0$
- Binärbild, mit zwei gleichhäufigen Werten: $H = 1$

Aufgabe: Berechnen Sie die Entropie für folgende Bilder:

a) $P_1 = P_2 = P_3 = P_4 = \frac{1}{4}$



b) $P_1 = \frac{1}{2} P_2 = \frac{1}{4} P_3 = P_4 = \frac{1}{8}$



Lösung:

a) $\underline{H = 2}$

b) $H = -[-2 \cdot \frac{1}{4} - 1 \cdot \frac{1}{2} - 3 \cdot 2 \cdot \frac{1}{8}] = \underline{\underline{\frac{14}{8} \approx 1,75}}$

Grauwertübergangsmatrix Die Grauwertübergangsmatrix (cooccurrence matrix) wird zur Beschreibung von bestimmten Bildeigenschaften verwendet. Sie wird durch eine festgelegte Relation ρ zwischen Bildpunkten definiert, die z.B. sein kann:

rechter Nachbar

$$g_1 \rightarrow g_2$$

g_2

↑

in vierer Umgebung

$$g_2 \leftarrow g_1 \rightarrow g_2$$

↓

g_2

g_2

↗

in der Diagonale

g_1

↙

g_2

Wenn man mit B das Bild $B(m, n)$ und a als die Häufigkeit der Grauwertkombination bezüglich der Relation ρ bezeichnet, so lässt sich die Grauwertübergangsmatrix definieren als:

$$CO_{B,\rho}(g_1, g_2) = [a_{g_1, g_2}] \quad (4.1)$$

Ein Beispiel soll den Sachverhalt verdeutlichen: Gegeben sei ein Bild mit den Grauwerten $G = \{0, 1, 2, 3\}$. Es hat folgende Grauwertmatrix:

$$B = \begin{pmatrix} 0 & 0 & 1 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 & 3 & 3 \\ 0 & 1 & 1 & 2 & 3 & 3 \\ 1 & 2 & 2 & 3 & 3 & 3 \\ 2 & 2 & 3 & 3 & 3 & 3 \end{pmatrix}$$

Die Grauwertübergangsmatrizen für ρ =rechter Nachbar und ρ =Viererumgebung lauten:

rechter Nachbar		Viererumgebung									
$g_1 \setminus g_2$	0	1	2	3	$g_1 \setminus g_2$	0	1	2	3		
$CO_{B,\rho}(g_1, g_2) =$	0	4	4	0	0	$CO_{B,\rho}(g_1, g_2) =$	0	18	8	0	0
	1	0	2	5	0		1	8	8	9	0
	2	0	0	2	6		2	0	9	10	9
	3	0	0	0	7		3	0	0	9	32

Wenn man z.B. für ρ Viererumgebung den Wert für $g_1 = 1, g_2 = 2$ ermitteln möchte, so geht man zu jeder 1 im Bild und inkrementiert das Ergebnis (beginnend bei 0) bei jeder 2, die sich in den Pixeln der Nachbarschaft befinden.

Anwendung findet die Grauwertübergangsmatrix in der Texturanalyse.

4.3 Zufallsprozesse

Um eine statistische Aussage über die Bildstruktur machen zu können, wird das ganze Bild als eine statistische Größe betrachtet, ein stochastisches Feld. Jedem Pixel wird eine Zufallsvariable zugeordnet und so besitzt das stochastische Feld $M \cdot N$ Zufallsvariablen. Mittelwert und Varianz können für jede Zufallsvariable über eine Bildfolge im Zeitraum T berechnet werden.

$$\mu(m, n) = \frac{1}{T} \sum_{t=0}^{T-1} B_t(m, n)$$

$$\sigma^2(m, n) = \frac{1}{T} \sum_{t=0}^{T-1} (B_t(m, n) - \mu(m, n))^2$$

Anwendung findet dies in der Untersuchung des Pixelverhaltens auf Homogenität (unter bestimmten Bedingungen, wie z.B. einer konstanten, homogenen Beleuchtung).

Außerdem stellt die statistische Betrachtungsweise die Grundlage für die grauwertbasierte Klassifikation dar. Jedes Pixel eines Bildes stellt ein Merkmal dar, weshalb der Merkmalsvektor $M \cdot N$ Zeilen aufweist. Mit den über die Zeit t erhobenen Stichproben

kann für jedes Pixel eine Wahrscheinlichkeitsbeschreibung (-Funktion) gewonnen werden. Diese $M \cdot N$ -dimensionale Wahrscheinlichkeitsfunktion $p(i, j, m, n, m', n')$ gibt an, wie hoch die Wahrscheinlichkeit ist, dass am Punkt (m, n) der Grauwert g_i vorliegt und gleichzeitig an einem anderen Punkt (m', n') der Grauwert g'_i .

Die folgenden Abbildungen (4.5, 4.6) zeigen beispielhaft die Wahrscheinlichkeitsdichtefunktion zweier Merkmale (Pixel) und die zugehörige "kombinierte" Wahrscheinlichkeitsdichte, die bei unkorrelierten Merkmalen durch Multiplikation gewonnen werden kann.

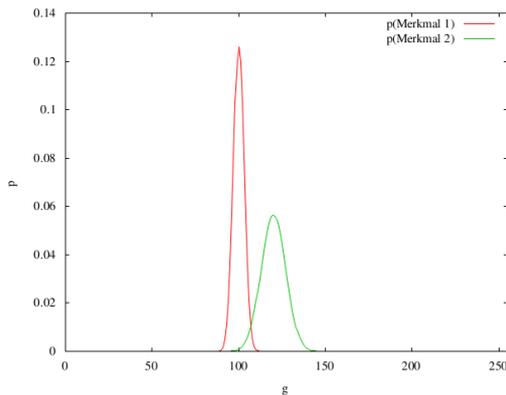


Abbildung 4.5:
Wahrscheinlichkeitsverteilung der beiden Merkmale 1 und 2

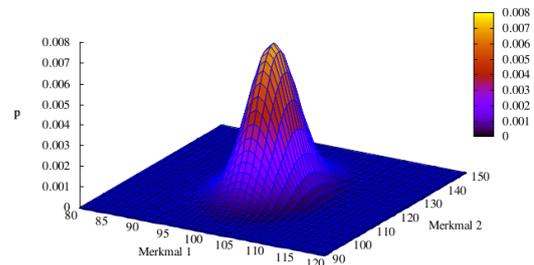


Abbildung 4.6: 3d-
Wahrscheinlichkeitsverteilung der unkorrelierten Merkmale 1 und 2

Korrelation Die Autokorrelationsfunktion verknüpft die Grauwerte der einzelnen Pixel miteinander und wird durch den Mittelwert des Produktes der Zufallsvariablen an zwei Punkten (m, n) und (m', n') beschrieben. Sie ist definiert durch:

$$\begin{aligned} R_{BB}(m, n, m', n') &= \frac{1}{T} \sum_{t=0}^{T-1} B_t(m, n) \cdot B_t(m', n') \\ &= \sum_{i=g_{\min}}^{g_{\max}} \sum_{j=g_{\min}}^{g_{\max}} g_i g_j \cdot p(i, j, m, n, m', n') \end{aligned}$$

Im Fall eines homogenen statistischen Prozesses, in dem alle Zufallsvariablen die gleiche statistische Charakteristik aufweisen, können statistische Parameter, wie Mittelwert und Autokorrelationsfunktion, aus einer Messung durch örtliche Mittelung abgeschätzt werden. Die Autokorrelationsfunktion ist definiert durch:

$$R_{BB}(m, n) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} B(i, j) \cdot B(i + m, j + n)$$

Analog dazu die Kreuzkorrelationsfunktion, welche die Ähnlichkeit zwischen zwei Bildern B und C beschreibt (vgl. Abbildung 4.7):

$$R_{CB}(m, n) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} C(i, j) \cdot B(i + m, j + n)$$

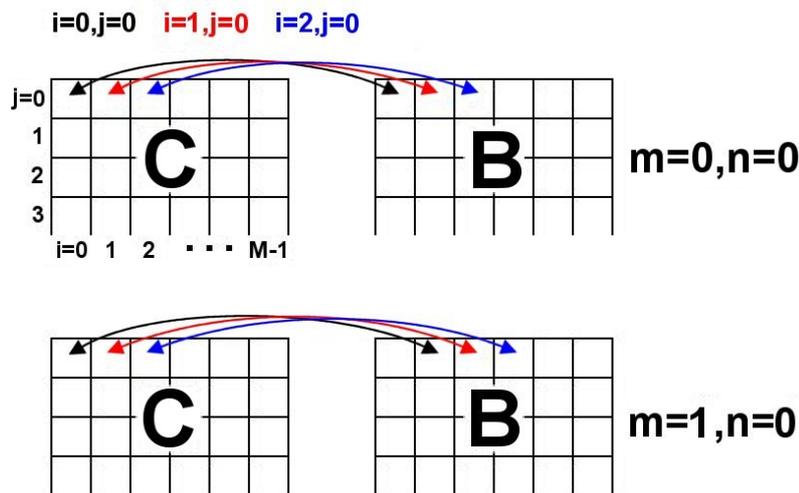


Abbildung 4.7: Kreuzkorrelation zweier Bilder. Dargestellt ist die Berechnung von $R_{CB}(m, n) = R_{CB}(0, 0)$ (oben) und $R_{CB}(m, n) = R_{CB}(1, 0)$ (unten).

Anwendungen sind in der Mustererkennung (Voraussetzung ist eine zufällige Verteilung des Musters), Texturanalyse und Bewegungsdetektion (Korrelationsdetektor) zu finden.

Kovarianz Die Kovarianzfunktion wird verwendet, um festzustellen, ob das Rauschen an Bildpunkten miteinander korreliert ist. Auch hier wird unterschieden zwischen der Autokovarianzfunktion. . .

$$C_{BB}(m, n) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (B(i, j) - \mu_B(i, j)) \cdot (B(i + m, j + n) - \mu_B(i + m, j + n))$$

. . . und der Kreuzkovarianzfunktion:

$$C_{CB}(m, n) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (C(i, j) - \mu_C(i, j)) \cdot (B(i + m, j + n) - \mu_B(i + m, j + n))$$

4.4 Wahrscheinlichkeitsverteilungen

Normalverteilung Die Normalverteilung ist durch den Mittelwert und die Varianz komplett beschrieben:

$$f(g) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(g-\mu)^2}{2\sigma^2}}$$

Für optimale Helligkeitsbedingungen kann diese Verteilung benutzt werden, um das Rauschen eines Bildsensorelements (allgemeiner auch das Systemrauschen) zu modellieren.

Binomialverteilung Die Binomialverteilung ist das diskrete Analogon zur Normalverteilung. Sie ist definiert durch:

$$f(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

mit

$$\begin{aligned} \mu &= n \cdot p & p &\in [0, 1] \\ \sigma^2 &= n \cdot p \cdot (1-p) & n &\geq 0 \\ & & k &\in \{0, \dots, n\} \end{aligned}$$

Beispiel: geg.: $p = \frac{1}{2}, n = 256$
 ges.: μ, σ^2
 Lsg.: $\mu = n \cdot p = 128$
 $\sigma^2 = n \cdot p(1-p) = 64$

Für die Berechnung der Binomialverteilung gilt: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, mit den Sonderfällen $\binom{0}{0} = 1$ und $\binom{n}{n} = \binom{n}{0} = 1$. Für die Berechnung kann auch das Pascalsche Dreieck verwendet werden:

n							$\frac{\sigma^2}{p=\frac{1}{2}}$					
0	1						0					
1	1		1				1/4					
2	1			2		1	1/2					
3	1				3		3	3/4				
4	1					4	6	4	1	1		
5	1						5	10	10	5	1	5/4
⋮	$\binom{5}{0}$	$\binom{5}{1}$	$\binom{5}{2}$	$\binom{5}{3}$	$\binom{5}{4}$	$\binom{5}{5}$	⋮					

Spezialfall: symmetrische Binomialverteilung

$$\begin{aligned} p = \frac{1}{2} \quad \Rightarrow \quad p^k(1-p)^{n-k} &= \left(\frac{1}{2}\right)^k \cdot \left(1 - \frac{1}{2}\right)^{n-k} \\ &= \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{n-k} \\ &= \frac{1}{2^n} \end{aligned}$$

Übergang zur Normalverteilung Für $n \rightarrow \infty$ konvergiert die Binomialverteilung gegen eine Normalverteilung, oder umgekehrt:

für hinreichend viele Stichproben n (Faustregel: $\sigma^2 \geq 9$) kann die Binomialverteilung durch die Normalverteilung approximiert werden:

$$f(k) \approx \frac{1}{\sqrt{2\pi \cdot np \cdot (1-p)}} \cdot \exp\left(-\frac{(k-np)^2}{2 \cdot np \cdot (1-p)}\right)$$

Kapitel 5

Punktoperationen

Dieses Kapitel befasst sich mit Punktoperationen, bei denen der Grauwert und/oder die Position eines Bildpunktes unabhängig von seiner Umgebung modifiziert wird.

$$B'(y_1, y_2) = f_{(x_1, x_2)}(B(x_1, x_2)) \quad (5.1)$$

Anwendungen sind z.B. Helligkeits- und Kontrastoptimierung und die geometrische Bildkorrektur. Man unterscheidet zwischen *homogener* und *inhomogener* Punktoperation. Die geometrische Transformation, bestehend aus *Translation*, *Skalierung* und *Rotation*, wird im Zusammenhang mit der *Interpolation* im dritten Teil dieses Kapitels behandelt.

5.1 Homogene Punktoperationen

Homogene Punktoperationen sind von der Position des Bildpunktes unabhängig.

$$B'(m_1, n_1) = f(B(m_2, n_2)) \quad (5.2)$$

Die Berechnung für ein ganzes Bild B kann sehr aufwendig sein. Beispiel: Für ein Bild mit $575 \cdot 767 = 440830$ Pixel und $f(B) = (a \cdot \log(B(m, n)) + b)$ ergibt das mehr als eine Million Operationen. Nützlich für diesen Rechenaufwand sind deshalb LUTs (Look Up Tables). Diese sind Speicher, die alle Werte von $f(g)$ beinhalten. Für $g \in G = [0, 1, 2, \dots, 255]$ (8-Bit-Felder) hat die LUT 256 Elemente. Die Verarbeitung reduziert sich auf eine Ersetzung der Grauwerte durch das entsprechende Tabellenelement, das über den jeweiligen Grauwert indiziert wird (Abbildung 5.1). Diese Tabellen können auch in die Hardware (z.B. in Framegrabbern zwischen A/D-Wandler und Speicher) integriert werden.

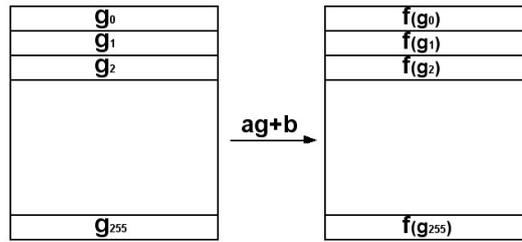


Abbildung 5.1: Prinzip einer Look Up Table. Dem Grauwertvektor (links) wird ein Funktionsvektor (rechts) zugewiesen.

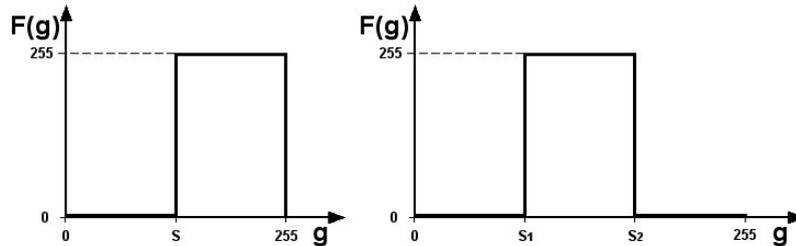


Abbildung 5.2: Links: Schwellwertbildung mit Untergrenze S . Rechts: Schwellwertbildung mit Unter- und Obergrenze S_1 und S_2 .

Schwellwertbildung Die Schwellwertbildung dient zur Erzeugung von Binärbildern. Dabei gilt für die Schwellwertfunktion (Abbildung 5.2 links):

$$F(g) = \begin{cases} 0 & \text{für } g < S \\ g_{\max} & \text{für } g \geq S \end{cases} \quad \begin{array}{l} g_{\max} \text{ entspricht} \\ \text{dem maximalen} \\ \text{Wert der gegebenen} \\ \text{Bittiefe (255} \\ \text{bei 8 Bit)} \end{array}$$

Eine Erweiterung ist gegeben, wenn auch eine Untergrenze festgelegt wird (Abbildung 5.2 rechts):

$$F(g) = \begin{cases} 0 & \text{für } g < S_1 \text{ und } g > S_2 \\ g_{\max} & \text{für } S_1 \leq g \leq S_2 \end{cases}$$

Lineare homogene Punktoperation Lineare homogene Punktoperationen mit der Abbildungsfunktion:

$$f(g) = a \cdot g + b$$

haben mit $a = 1$ und $b \neq 0$ eine reine Helligkeitsänderung zur Folge. Für $a > 1$ findet eine Kontrastverstärkung und für $0 < a < 1$ eine Kontrastverringernng statt. Abbildung 5.3 veranschaulicht die Abhängigkeit der Funktion von den Parametern a und b anhand von Funktionenscharen.

Als Beispiel soll der Bereich $[50, 190]$ durch eine Kontrastverstärkung auf den gesamten

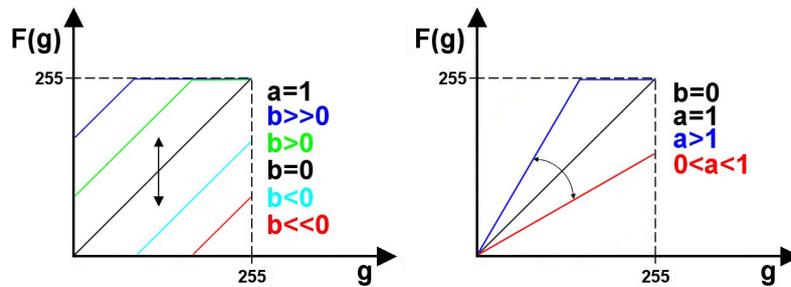
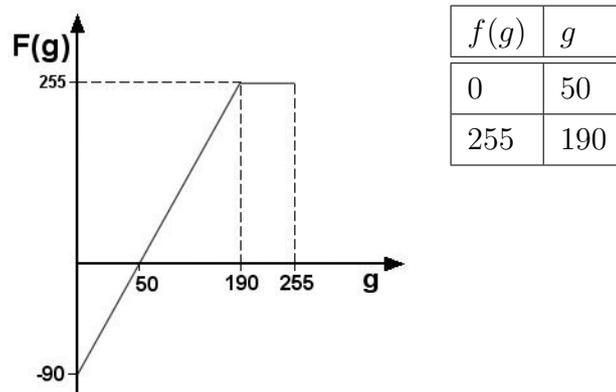


Abbildung 5.3: Auswirkungen bei Parameteränderung von a und b in der Abbildungsgleichung $F(g) = a \cdot g + b$.

Links: Helligkeitsänderung durch Variierung von b . Es folgt eine Parallelverschiebung der Kurve mit oberer/unterer Grauwertgrenze (255/0).

Rechts: Kontraständerung durch Anstiegsvarierung (a). Auch hier sind die Grauwertgrenzen 255 (oben) und 0 (unten).

Bereich $[0, 255]$ abgebildet werden. Dabei ergibt sich die lineare Funktion $f(g) = \frac{255-0}{190-50} \cdot g - 50 \cdot \frac{255}{190-50} = 1.82 \cdot g - 91$ mit folgendem Graphen:



Nichtlineare homogene Punktoperation Bei der nichtlinearen homogenen Punktoperation werden die Grauwerte durch eine nichtlineare Funktion abgebildet, wie z.B.:

$$f(g) = g_{\max} \cdot \left(\frac{g}{g_{\max}} \right)^{\frac{1}{\gamma}} \quad \text{mit } \gamma > 1$$

Viele Videosensoren nehmen die Lichtintensität I nicht linear, sondern exponentiell auf: $U = U_{\max} \cdot \left(\frac{I}{I_{\max}} \right)^{\gamma}$. Dies hat den Zweck, den Kontrastumfang zu steigern, da das menschliche Auge ein logarithmisches und relatives Auflösungsvermögen hat. Wenn man z.B. den Gammawert auf 1 einstellt, so liegt der Kontrastumfang bei 25; ist der Gammawert 0,4, so liegt der Kontrastumfang bei 316 (mehr als zwölfmal höher).

Wird jedoch eine lineare Funktion benötigt, so wendet man die obige nichtlineare Funktion auf die Funktion der Kamera an, siehe Abbildung 5.4.

Äquidensiten Äquidensiten sind Linien oder Flächen, die in einem Bild Punkte gleichen Grauwertes oder Farbdichte verbinden. Dabei wird jeweils einem Grauwertbereich

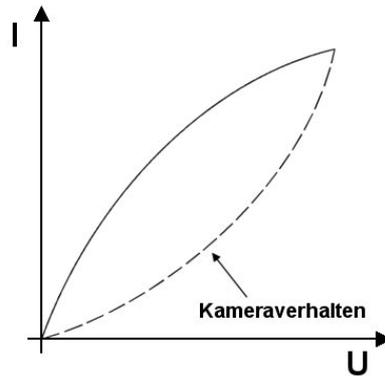


Abbildung 5.4: Nichtlineare homogene Punktoperation. Die gestrichelte gammakorrigierte Funktion wird durch ihre Umkehrfunktion linearisiert.

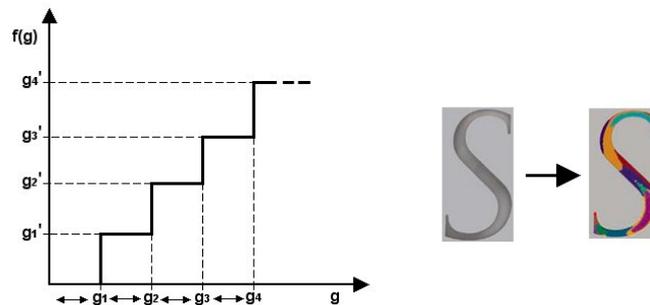


Abbildung 5.5: Segmentierungstechnik Äquidensiten. **Links:** Einem bestimmten Grauwertbereich wird mit der Funktion ein Grauwert zugewiesen. **Rechts:** (Abb. aus [www \[16\]](#)) Farbliche Hervorhebung der Segmente.

ein bestimmter Grauwert zugewiesen. Um sich gering unterscheidende Grauwerte darzustellen kann man statt Grauwerten auch unterschiedliche Farben zuweisen. Diese einfache Segmentierungstechnik wird in [Abbildung 5.5](#) veranschaulicht.

5.2 Inhomogene Punktoperationen

Bei inhomogenen Punktoperationen hängt die Abbildungsfunktion vom Ort des Bildes ab:

$$B'(m, n) = f_{(m,n)}(B(m, n))$$

Sie wird für die Korrektur von Aufnahme Fehlern verwendet, die z.B. bei inhomogener Beleuchtung oder unterschiedlicher Empfindlichkeit der CCD-Sensoren entstehen können. Die einfachste Form (mit B_1 als Originalbild und B_2 als Korrekturbild) lautet:

$$B'(m, n) = B_1(m, n) - B_2(m, n)$$

Man kann den Aufnahme fehler als additiv oder multiplikativ auffassen und entsprechend die Korrektur definieren:

- Additive Korrektur

$$B'(m, n) = B_1(m, n) - B_2(m, n) + \mu_{B_2}$$

μ_{B_2} : Mittelwert von B_2

- Multiplikative Korrektur

$$B'(m, n) = B_1(m, n) \cdot \frac{\mu_{B_2}}{B_2(m, n)}$$

Das Korrekturbild B_2 lässt sich durch die Aufnahme einer Vorlage mit konstanter Reflektivität erzeugen.

5.3 Geometrische Transformation

Geometrische Transformationen definieren die Beziehung zwischen den Punkten zweier Bilder. Wenn die Abbildungsfunktion mit M bezeichnet wird, so gilt die Beziehung zwischen dem Ausgangsbild B' und dem Eingangsbild B :

$$B' = M(B)$$

Die Grundtypen der linearen Koordinatentransformation (affine Abbildung) sind:

- **T** Translation
- **S** Skalierung
- **R** Rotation

Sie können durch Vektoraddition und Matrixmultiplikation ausgedrückt werden:

$$\begin{pmatrix} m' \\ n' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix} + \begin{pmatrix} t_m \\ t_n \end{pmatrix}$$

Für homogene Koordinaten kann die affine Abbildung mit einer einzigen Matrixmultiplikation beschrieben werden:

$$\begin{pmatrix} m' \\ n' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_m \\ a_{21} & a_{22} & t_n \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m \\ n \\ 1 \end{pmatrix}$$

Die Grundtypen lauten in Matrixform:

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & -t_m \\ 0 & 1 & -t_n \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} S_m & 0 & 0 \\ 0 & S_n & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

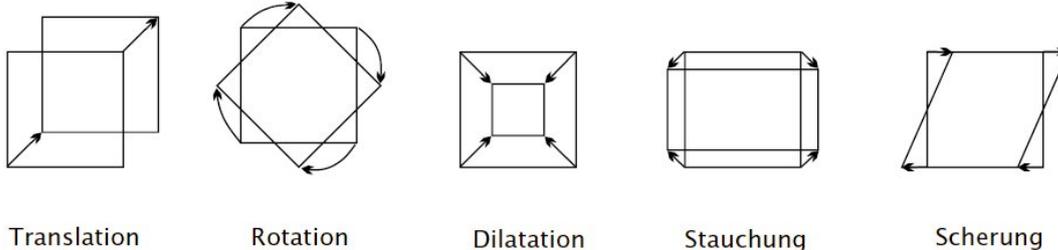


Abbildung 5.6: (aus Jähne [1]) Elementar geometrische Transformation eines planaren Oberflächenelements.

Alle möglichen geometrischen Transformationen in Abbildung 5.6 lassen sich durch Kombination der drei Grundtypen beschreiben.

Interpolation Bei der geometrischen Transformation werden die Bildpunkte i.a. auf Zwischengitterpositionen transformiert. Der klassische Ansatz hierfür ist die lineare Interpolation. Zur Vereinfachung der Berechnungen werden normierte Variablen genutzt: $\hat{x} = \frac{x}{\Delta x}$. Abbildung 5.7 zeigt ein Beispiel für die lineare Interpolation, bei dem aus Symmetriegründen die zwei Punkte $\frac{1}{2}$ und $-\frac{1}{2}$ gewählt wurden. Daraus folgt die Interpolationsgleichung:

$$g(\hat{x}) = \frac{g_{\frac{1}{2}} + g_{-\frac{1}{2}}}{2} + (g_{\frac{1}{2}} - g_{-\frac{1}{2}}) \cdot \hat{x} \quad \text{wobei } \hat{x} \in \left[-\frac{1}{2}, \frac{1}{2}\right]$$

im Bsp: $\hat{x} = 0$.

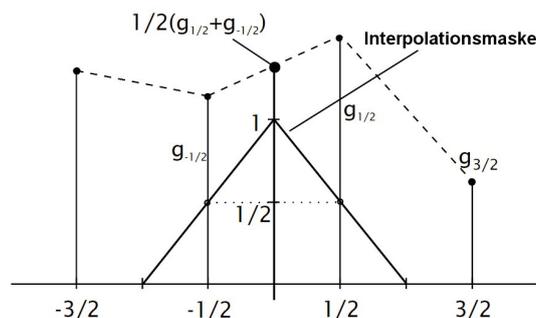


Abbildung 5.7: (aus Jähne [1]) Veranschaulichung der linearen Interpolation mit den Punkten $\frac{1}{2}$ und $-\frac{1}{2}$.

Die kontinuierliche Interpolationsmaske für die lineare Interpolation ist in der Abbildung 5.7 ebenfalls zu erkennen. Sie lässt sich beschreiben mit:

$$h_1(\hat{x}) = \begin{cases} 1 - |\hat{x}| & |\hat{x}| \leq 1 \\ 0 & \textit{sonst} \end{cases}$$

Kapitel 6

Bildtransformationen

Bildtransformationen sind für die Bildverarbeitung von entscheidender Bedeutung, da sie verschiedene Eigenschaften hervorheben oder komplizierte Rechenwege erheblich vereinfachen können. Dazu wollen wir die analoge und diskrete *Fouriertransformation* kennenlernen, aber auch die *Sinus-* und *Kosinustransformation*, sowie die *Walshtransformation*. Eine schnelle Berechnung der Fouriertransformation in Form der *Fast Fouriertransformation (FFT)* soll zum Schluss vorgestellt werden.

6.1 Grundlagen der Fouriertransformation

„Bei der mathematischen Beschreibung ist es zweckmäßig, Signale als Punkte oder Vektoren in einem Funktionenraum – dem Signalraum, Signaltransformationen als Abbildungen in diesem Raum und Eigenschaften der Signale als Eigenschaften des Raumes zu betrachten.“

6.1.1 Definitionen

Wir werden das Bild als Element eines normierten linearen metrischen Raumes (des Funktionenraumes) beschreiben. Der Funktionenraum ist ein linearer Raum. Es gilt:

1. $\vec{a}_1, \vec{a}_2 \in \mathbb{R}^n$, $\vec{a}_1 + \vec{a}_2 = \vec{a}_3 \in \mathbb{R}^n$
2. $0 \in \mathbb{R}^n$, $\vec{a} \in \mathbb{R}^n$, $\vec{a} + 0 = \vec{a}$
3. α ist eine beliebige Zahl, $\vec{a} \in \mathbb{R}^n$: $\alpha \cdot \vec{a} = b \in \mathbb{R}^n$

In dem Funktionenraum wird das folgende Skalarprodukt definiert:

$$\underbrace{(\vec{a}_1, \vec{a}_2)}_{\text{Skalarprodukt}} = \underbrace{\int_T a_1(t) \cdot a_2^*(t) dt}_{\text{stetig}} = \underbrace{\sum_{n=0}^{N-1} a_{1,n} a_{2,n}^*}_{\text{diskret}}$$

wobei \vec{a}_1 und \vec{a}_2 im allgemeinen komplexwertig sind und \vec{a}^* komplex konjugiert zu \vec{a} ist.

Der Raum heißt metrisch, wenn in ihm zwischen den Elementen ein Abstand (eine Metrik) definiert ist. Für diskrete Signale ist die euklidische Metrik folgendermaßen definiert:

$$l_N^2 = d(\vec{a}_1, \vec{a}_2) = \sqrt{\sum_{n=0}^{N-1} (a_{1,n} - a_{2,n})^2}$$

Für stetige Signale gilt dagegen:

$$l_T^2 = d(\vec{a}_1, \vec{a}_2) = \sqrt{\int_T |a_1(t) - a_2(t)|^2 dt}$$

Ein Raum mit einem Skalarprodukt kann zu einem normierten metrischen Raum gemacht werden. Hier sei auf die *euklidische Norm* hingewiesen, die folgendermaßen definiert ist:

$$\|\vec{a}\| = \sqrt{(\vec{a}, \vec{a})} \quad \text{für } \vec{a} = \vec{a}_1 - \vec{a}_2$$

Dieser Raum heißt bei einer endlichen Dimension N *euklidischer Raum*.

Beispiel:

Sei n_k eine unabhängige Zufallsgröße und

$$a_{2,k} = a_{1,k} + n_k \quad \text{mit } k = 0, 1, 2, \dots, (N-1)$$

Alle Abweichungen der Signale a_1 und a_2 voneinander sind im Signal $n = \{n_k\}$ enthalten. Dieses kann statistisch vollständig durch seine mehrdimensionale Wahrscheinlichkeitsdichte ausgedrückt werden:

$$p(n) = p(n_0) \cdot p(n_1) \cdot \dots \cdot p(n_{N-1})$$

mit

$$p(n_k) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left(-\frac{n_k^2}{2\sigma^2}\right)$$

$$\Rightarrow p(n) = \frac{1}{\sqrt{(2\pi)^N \cdot \sigma^N}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (a_{2k} - a_{1k})^2\right)$$

Sie ist durch den euklidischen Abstand vollständig beschrieben:

$$d(\vec{a}_1, \vec{a}_2) = \sqrt{\sum_{k=0}^{N-1} (a_{2,k} - a_{1,k})^2}$$

Im euklidischen Raum gilt:

- \vec{a}_1 und \vec{a}_2 sind orthogonal, wenn $(\vec{a}_1, \vec{a}_2) = 0$ gilt.
Orthogonale Vektoren sind linear unabhängig und können deswegen als Basis linearer Räume genutzt werden.
- $\{\vec{\varphi}_k\}$ mit $k = 0, \dots, N-1$ ist eine Basis von der der Raum A^N aufgespannt wird.
- Jeder Vektor kann in diesem Raum durch $\vec{a} = \sum_{k=0}^{N-1} \alpha_k \cdot \vec{\varphi}_k$ repräsentiert werden oder im Sinne des Funktionenraumes:

$$a(t) = \sum_{k=0}^{N-1} \alpha_k \cdot \varphi_k(t) \quad (6.1)$$

- $\{\vec{\psi}_l\}$ seien Vektoren, die zu $\{\vec{\varphi}_k\}$ paarweise orthogonal und normiert (dual) sind.
 $(\vec{\varphi}_k, \vec{\psi}_l) = \delta(k, l) = \begin{cases} 1 & \text{für } k = l \\ 0 & \text{für } k \neq l \end{cases}$
 \Rightarrow
 $\alpha_l = (\vec{a}, \vec{\psi}_l) \quad (6.2)$
- Wenn die Basis $\{\vec{\varphi}_k\}$ normierte, paarweise orthogonale Vektoren enthält, d.h. dual zu sich selbst ist, dann nennt man sie orthonormal:
 $(\vec{\varphi}_k, \vec{\varphi}_l) = \delta(k, l)$

Es sei noch der *Hammingabstand* (-metrik) erwähnt, welcher sich auf binäre Signale bezieht. Er ist mit dem Modulo-2-Operator (\oplus) folgendermaßen definiert:

$$d(\vec{a}_1, \vec{a}_2) = \sum_{k=0}^{N-1} (a_{1k} \oplus a_{2k})$$

6.1.2 Wichtige Basisfunktionen

Impulsfunktionen

$$\varphi_k(t) = \text{rect} \left(\frac{t - kT}{T} \right) = \text{rect } x = \begin{cases} 1 & \text{für } 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases}$$

Die duale Basis für dieses Funktionensystem bildet die Funktion:

$$\psi_k(t) = \frac{1}{T} \operatorname{rect} \left(\frac{t - kT}{T} \right)$$

Für den zweidimensionalen Fall:

$$\varphi_{k,l}(t_1, t_2) = \operatorname{rect} \left(\frac{t_1 - kT_1}{T_1} \right) \cdot \operatorname{rect} \left(\frac{t_2 - lT_2}{T_2} \right)$$

Mit dem dualen Funktionensystem:

$$\psi_{k,l}(t_1, t_2) = \frac{1}{T_1 T_2} \cdot \operatorname{rect} \left(\frac{t_1 - kT_1}{T_1} \right) \cdot \operatorname{rect} \left(\frac{t_2 - lT_2}{T_2} \right)$$

Harmonische Funktionen

$$\varphi_k^S(t) = \sin \left(\frac{\pi kt}{T} \right)$$

$$\varphi_k^C(t) = \cos \left(\frac{\pi kt}{T} \right)$$

$$\varphi_k^e(t) = e^{\left(\frac{\pi i kt}{T} \right)}$$

Die duale Basis:

$$\psi_k(t) = \frac{1}{T} \cdot e^{\left(-\frac{\pi i kt}{T} \right)}$$

Für den zweidimensionalen Fall:

$$\varphi_{k,l}(t_1, t_2) = e^{\left[2\pi i \left(\frac{kt_1}{T_1} + \frac{lt_2}{T_2} \right) \right]}$$

Abtastfunktionen

$$\varphi_k(t) = \frac{\sin \left[2\pi F \left(t - \frac{k}{2F} \right) \right]}{2\pi F \left(t - \frac{k}{2F} \right)} = \operatorname{sinc} \left[2\pi F \left(t - \frac{k}{2F} \right) \right]$$

$$\int_{-\infty}^{+\infty} \operatorname{sinc} \left[2\pi F \left(t - \frac{k}{2F} \right) \right] \cdot \operatorname{sinc} \left[2\pi F \left(t - \frac{l}{2F} \right) \right] \cdot dt$$

$$= \frac{1}{2F} \cdot \operatorname{sinc}[2\pi(k-l)] = \begin{cases} \frac{1}{2F} & \text{für } k = l \\ 0 & \text{für } k \neq l \end{cases}$$

Die duale Basis:

$$\psi_k(t) = 2F \cdot \operatorname{sinc} \left[2\pi F \left(t - \frac{k}{2F} \right) \right]$$

Für den Fall, dass das Argument der *sinc*-Funktion 0 ist, gilt:

$$\frac{\sin(0)}{0} = \frac{0}{0} \text{ ist undefiniert}$$

⇒ Anwendung der L'Hospital'schen Regel mit $\lim_{k \rightarrow 0} \rightarrow 0$

$$\frac{\cos(0)}{2F} = \frac{1}{2F}$$

Walshfunktionen Die Walshfunktionen sollen hier nur vollständigerhalber genannt werden. Sie werden später behandelt.

6.1.3 Stetige Darstellung der Basisfunktionen

Ersetzt man den Index k der Basisfunktionen durch eine kontinuierliche Veränderliche $f \in F$, so kann man die Darstellung in (6.1) als Grenzfall der kontinuierlichen Darstellung betrachten.

$$a(t) = \int_F \alpha(f) \varphi(t, f) df$$

Analog dazu aus (6.2):

$$\alpha(f) = \int_T a(t) \psi^*(f, t) dt$$

Die Funktion $\alpha(f)$ nennt man *Integraltransformierte* des Signals $a(t)$ oder sein Spektrum nach der stetigen Basis $\varphi(t, f)$.

6.2 Die Fouriertransformation

Die Darstellung von Signalen nach den Exponentialfunktionen heißt *Fourierentwicklung* oder *Fouriertransformation*:

$$\alpha(f) = \frac{1}{T} \int_T a(t) \cdot e^{-2\pi i f t} dt$$

Die Umkehrfunktion lautet:

$$a(t) = \int_F \alpha(f) \cdot e^{+2\pi i f t} df$$

Wichtige Eigenschaften und Transformationspaare sind:

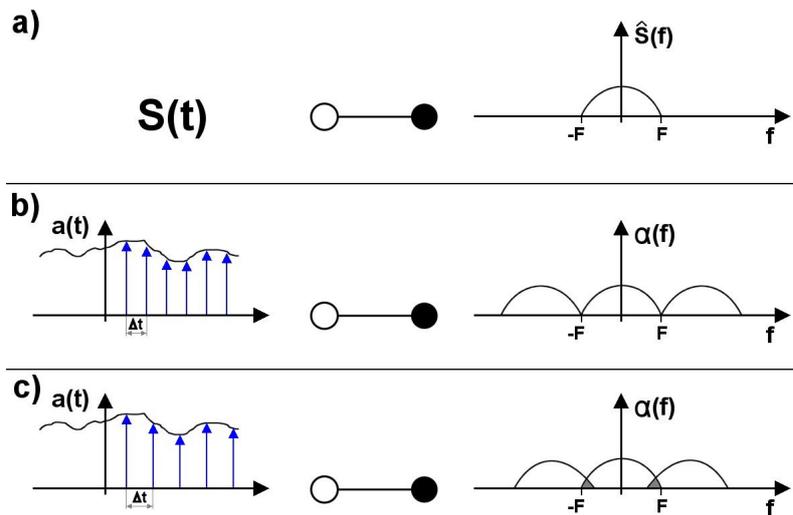


Abbildung 6.1: a) Das Signal wird als Funktion mit begrenztem Spektrum betrachtet. b) $\Delta t = \frac{1}{2F}$ Abtastung mit δ -Impulskamm. Die Spektren wiederholen sich. c) $\Delta t > \frac{1}{2F}$ Die wiederholten Spektren überlappen sich.

	Zeitfunktion	Fouriertransformierte
Faltungstheorem	$\int_{-\infty}^{+\infty} a(\tau) \cdot b(t - \tau) d\tau$	$\circ \text{---} \bullet \quad \alpha(f) \cdot \beta(f)$
Ableitung	$\frac{d^n a(t)}{dt^n}$	$\circ \text{---} \bullet \quad (-2\pi i f)^n \cdot \alpha(f)$
Rechteckfunktion	$rect\left(\frac{t + \frac{T}{2}}{T}\right)$	$\circ \text{---} \bullet \quad T \cdot sinc(\pi T f)$
Spaltenfunktion	$2F \cdot sinc(2\pi F t)$	$\circ \text{---} \bullet \quad rect\left(\frac{f + F}{2F}\right)$
Diskretes Signal	$\frac{1}{2F} \sum_{k=-\infty}^{+\infty} a\left(\frac{k}{2F}\right) \cdot \delta\left(t - \frac{k}{2F}\right)$	$\circ \text{---} \bullet \quad \sum_{l=-\infty}^{+\infty} \alpha(f - 2Fl)$
Periodisches Signal	$\sum_{k=-\infty}^{+\infty} a(t - kT)$	$\circ \text{---} \bullet \quad \frac{1}{T} \sum_{s=-\infty}^{+\infty} \alpha\left(\frac{s}{T}\right) \cdot \delta\left(f - \frac{s}{T}\right)$
	$e^{-2\pi i p t}$	$\circ \text{---} \bullet \quad \delta(f - p)$
	$\cos(2\pi p t)$	$\circ \text{---} \bullet \quad \frac{\delta(f-p) + \delta(f+p)}{2}$

6.3 Diskrete Fouriertransformation

„Die Methoden der Beschreibung von Operatoren sind durch die Methoden der diskreten Darstellung der verarbeiteten Signale völlig festgelegt.“

6.3.1 Abtasttheorem

Wenn man die Signale als Funktionen mit begrenztem Fourierspektrum $(-F, F)$ ansieht (siehe Abbildung 6.1 a)), so muss man, damit eine fehlerfreie Rekonstruktion des kontinuierlichen, originalen Signals erfolgt, die Abtastwerte mit einer Schrittweite von $\Delta t = \frac{1}{2F}$ ermitteln (siehe Abbildung 6.1 b) und c)).

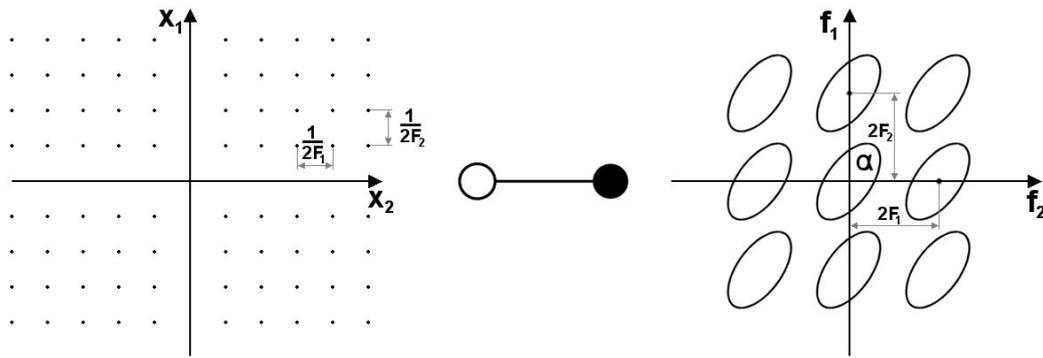


Abbildung 6.2: Abtastung eines zweidimensionalen Bildes.

6.3.2 2D-Abtastung

Die zweidimensionale Abtastung soll Abbildung 6.2 verdeutlichen. Auch hier kommt es zu sich wiederholenden Spektren.

6.3.3 Rekonstruktion

Wenn man ein kontinuierliches Signal rekonstruieren möchte, so ist eine Interpolation der Abtastpunkte nötig:

$$a(t) = \sum_{k=-\infty}^{+\infty} a(k \cdot \Delta t) \cdot h(t - k \cdot \Delta t)$$

Dabei ist $h(t - k \cdot \Delta t)$ eine Interpolationsfunktion, die die Abtastwerte wichtet. Sie wird folgendermaßen bestimmt:

Um das originale Signal wiedergewinnen zu können, müssen die verschobenen Spektren eliminiert werden. Dazu multipliziert man das Spektrum $\sum_{l=-\infty}^{+\infty} \alpha(f - 2Fl)$ mit der Funktion $rect\left(\frac{f+F}{2F}\right)$ (siehe Abbildung 6.3). Die Multiplikation der Spektren im Frequenzbereich...

$$\alpha(f) = rect\left(\frac{f+F}{2F}\right) \cdot \sum_{l=-\infty}^{+\infty} \alpha(f - 2Fl)$$

...ergibt eine Faltung der Signale im Zeitbereich:

$$a(t) = \sum_{k=-\infty}^{+\infty} a(k \cdot \Delta t) \cdot sinc\left[2\pi F \left(t - \frac{k}{2F}\right)\right]$$

Zur Herleitung:

$$\frac{1}{2F} \sum_{k=-\infty}^{+\infty} \alpha(k \cdot \Delta t) \cdot \delta\left(t - \frac{k}{2F}\right) \circ \bullet \sum_{l=-\infty}^{+\infty} \alpha(f - 2Fl)$$

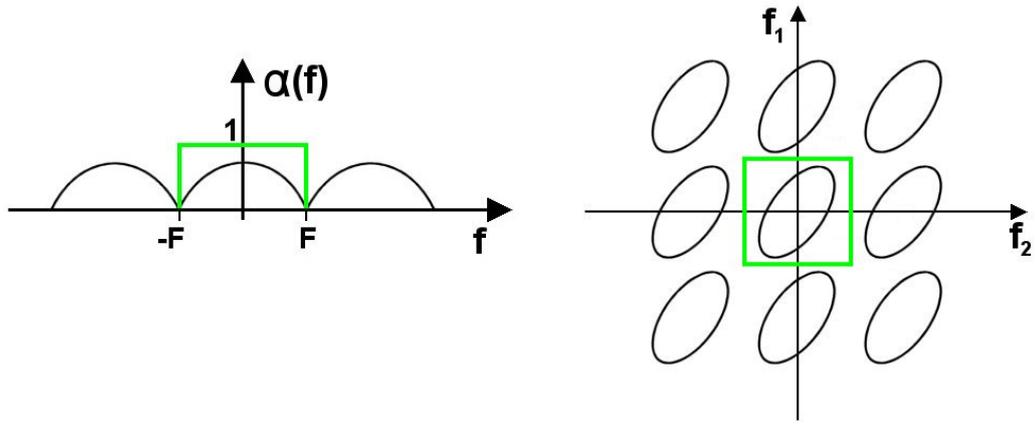


Abbildung 6.3: Multiplikation des Spektrums mit der *rect*-Funktion für 1D (links) und 2D (rechts).

mit

$$\int_{-\infty}^{+\infty} 2F \cdot \text{sinc}(f\pi Ft') \cdot b(t-t') dt' \quad \circ \bullet \quad \text{rect}\left(\frac{f+F}{2F}\right) \cdot \beta(f)$$

und

$$\begin{aligned} \beta(f) &= \sum_{l=-\infty}^{+\infty} \alpha(F-2Fl) \\ b(t-t') &= \frac{1}{2F} \sum_{k=-\infty}^{+\infty} a(k \cdot \Delta t) \cdot \delta\left(t - \frac{k}{2F} - t'\right) \\ \Rightarrow a(t) &= \int_{-\infty}^{+\infty} 2F \cdot \text{sinc}(2\pi Ft') \cdot \frac{1}{2F} \sum_{k=-\infty}^{+\infty} a(k \cdot \Delta t) \cdot \delta\left(t - \frac{k}{2F} - t'\right) dt' \\ &= \sum_{k=-\infty}^{+\infty} a(k \cdot \Delta t) \cdot \int_{-\infty}^{+\infty} \text{sinc}(2\pi Ft') \cdot \delta\left(t - \frac{k}{2F} - t'\right) dt' \\ \Rightarrow a(t) &= \sum_{k=-\infty}^{+\infty} a(k \cdot \Delta t) \cdot \text{sinc}\left[2\pi F \left(t - \frac{k}{2F}\right)\right] \end{aligned}$$

Für die Multiplikation eines zweidimensionalen Spektrums mit der *rect*-Funktion ergibt sich analog dazu:

$$a(t_1, t_2) = \sum_{k_1=-\infty}^{+\infty} \sum_{k_2=-\infty}^{+\infty} a\left(\frac{k_1}{F_1}, \frac{k_2}{F_2}\right) \cdot \text{sinc}\left[2\pi F_1 \left(t_1 - \frac{k_1}{2F_1}\right)\right] \cdot \text{sinc}\left[2\pi F_2 \left(t_2 - \frac{k_2}{2F_2}\right)\right]$$

Dies bedeutet, dass das Signal $a(t)$ nach dem Basissystem der Abtastfunktionen $\text{sinc}\left[2\pi F \left(t - \frac{k}{2F}\right)\right]$ dargestellt wird. Und es gilt:

$$a(k \cdot \Delta t) = 2F \int_{-\infty}^{+\infty} a(t) \cdot \text{sinc}\left(2\pi F \left[t - \frac{k}{2F}\right]\right) dt$$

$$= \int_{-\infty}^{+\infty} a(t) \cdot \delta\left(t - \frac{k}{2F}\right) dt$$

6.3.4 DFT

Bei der DFT handelt es sich um die diskrete Darstellung der Fouriertransformation nach der Basis:

$$\varphi_k(t) = 2F \cdot \text{sinc} \left[2\pi F \left(t - \frac{k}{2F} \right) \right]$$

Damit ergibt sich für die Fouriertransformierte:

$$\begin{aligned} \alpha(f) &= \int_{-\infty}^{+\infty} a(t) \cdot e^{-2\pi i f t} dt \\ &= \sum_{k=-\infty}^{+\infty} a\left(\frac{k}{2F}\right) \cdot \int_{-\infty}^{+\infty} \text{sinc} \left[2\pi F \left(t - \frac{k}{2F} \right) \right] \cdot e^{-2\pi i f t} dt \\ \alpha(f) &= \frac{1}{2F} \sum_{k=-\infty}^{+\infty} a\left(\frac{k}{2F}\right) \cdot e^{-2\pi i f \frac{k}{2F}} \end{aligned} \quad (6.3)$$

Betrachtet wird das periodische Signal:

$$a(t) = \sum_{k=-\infty}^{+\infty} a(t - kT)$$

Sein Spektrum ist:

$$\begin{aligned} \alpha(f) &= \frac{1}{T} \sum_{s=-\infty}^{+\infty} \alpha\left(\frac{s}{T}\right) \cdot \delta\left(f - \frac{s}{T}\right) \\ \alpha\left(\frac{s}{T}\right) &= \int_{-\infty}^{+\infty} \alpha(f) \cdot \delta\left(f - \frac{s}{T}\right) df \end{aligned} \quad (6.4)$$

Setzt man Gleichung 6.3 in 6.4 ein, erhält man:

$$\begin{aligned} \alpha\left(\frac{s}{T}\right) &= \frac{1}{2F} \cdot \sum_k a\left(\frac{k}{2F}\right) \cdot \underbrace{\int_{-\infty}^{+\infty} \delta\left(f - \frac{s}{T}\right) \cdot e^{-2\pi i f \frac{k}{2F}} \cdot df}_{e^{-2\pi i \frac{k}{2F} \frac{s}{T}}} \\ \alpha\left(\frac{s}{T}\right) &= \frac{1}{2F} \cdot \sum_k a\left(\frac{k}{2F}\right) \cdot e^{-2\pi i \frac{k}{2F} \frac{s}{T}} \end{aligned}$$

Dabei ist $k < TF$. Wir nennen $N = 2TF = \frac{T}{\Delta t}$ und $\alpha_s = \frac{1}{T} \alpha\left(\frac{s}{T}\right)$.

Für die DFT gilt dann:

$$\alpha_s = \frac{1}{N} \sum_{k=0}^{N-1} a_k \cdot e^{-2\pi i \frac{ks}{N}} \quad (6.5)$$

Die inverse Transformation IDFT lautet:

$$a_k = \sum_{s=0}^{N-1} \alpha_s \cdot e^{2\pi i \frac{ks}{N}}$$

6.3.5 2D-DFT

Analog zu der eindimensionalen DFT folgt für die 2D-DFT die Hintransformation:

$$\alpha_{r,s} = \frac{1}{N_1 N_2} \sum_{k=0}^{N_1-1} \sum_{l=0}^{N_2-1} a_{k,l} \cdot e^{-2\pi i \left(\frac{ks}{N_1} + \frac{lr}{N_2} \right)}$$

Dementsprechend ist die Rücktransformation gegeben durch:

$$a_{k,l} = \frac{1}{N_1 N_2} \sum_{s=0}^{N_1-1} \sum_{r=0}^{N_2-1} \alpha_{r,s} \cdot e^{2\pi i \left(\frac{ks}{N_1} + \frac{lr}{N_2} \right)}$$

Um einen Eindruck der Fouriertransformation zu bekommen, sind in Abbildung 6.4 einige Beispiele von Bildtransformationen aufgeführt.

6.3.6 Wichtige Eigenschaften der 2D-DFT

Separierbarkeit Operationen sind dimensionsweise nacheinander ausführbar. Dies sieht man in der folgenden Formel, wobei (1) und (2) vertauschbar sind:

$$\alpha_{r,s} = \frac{1}{N_1 N_2} \sum_{k=0}^{N_1-1} \underbrace{\left(\sum_{l=0}^{N_2-1} a_{k,l} \cdot e^{-2\pi i \frac{lr}{N_2}} \right)}_{(1)} \underbrace{e^{-2\pi i \frac{ks}{N_1}}}_{(2)}$$

Rotationssatz Der Rotation des Bildes um den Winkel φ steht eine Rotation des Spektrums um den selben Winkel gegenüber:

$$k' = k \cdot \cos \varphi + l \cdot \sin \varphi \quad ; \quad l' = l \cdot \cos \varphi - k \cdot \sin \varphi$$

$$r' = r \cdot \cos \varphi + s \cdot \sin \varphi \quad ; \quad s' = s \cdot \cos \varphi - r \cdot \sin \varphi$$

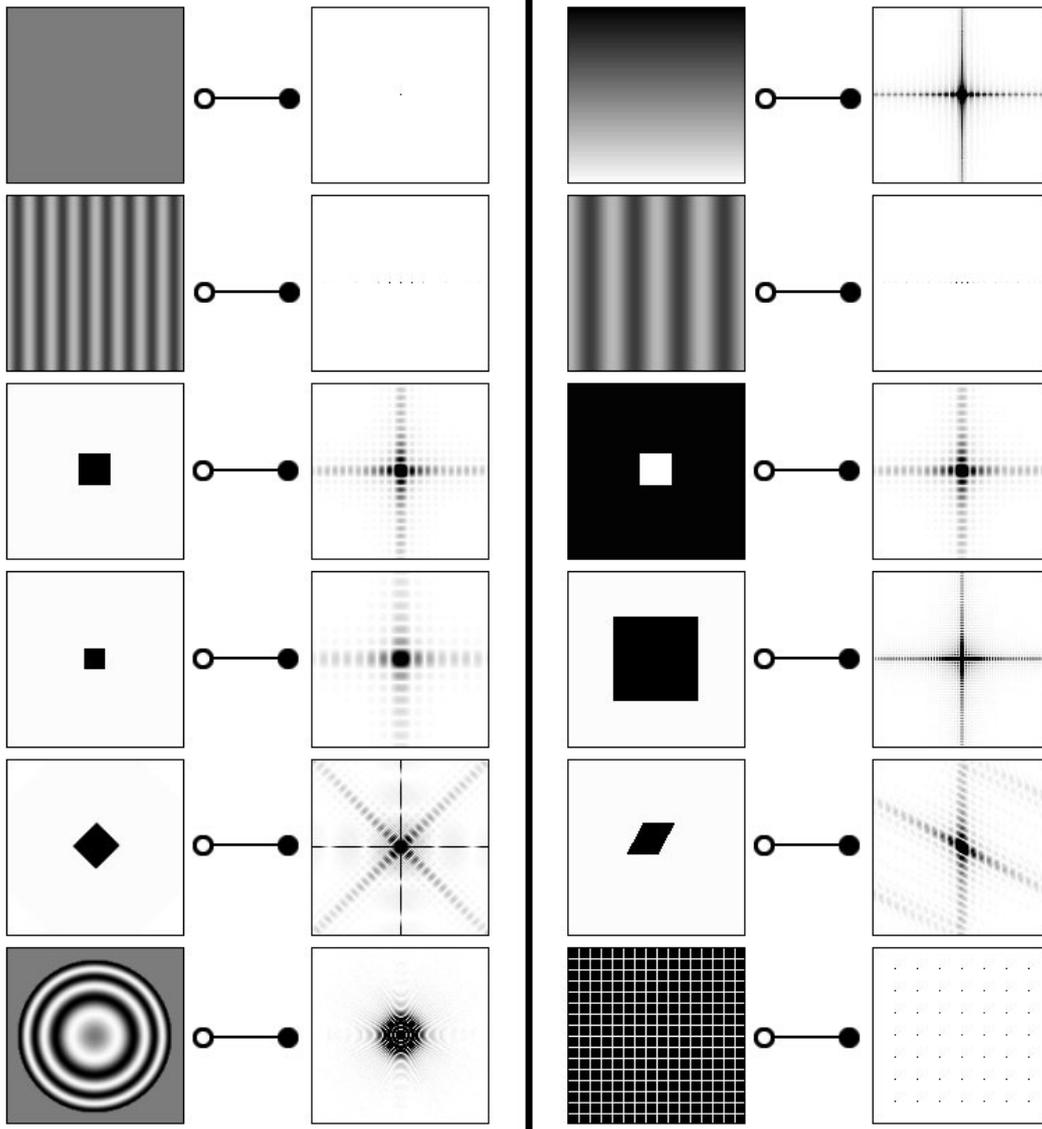


Abbildung 6.4: Beispiele für die Transformation von Bildern in den Fourierraum.

Verschiebungssatz Es gilt bei der 2D-DFT das Gleiche wie bei der eindimensionalen Transformation: Eine Verschiebung im Ortsbereich bewirkt eine lineare Phasendrehung im Ortsfrequenzbereich und hat keinen Einfluss auf die Amplituden im Spektrum. (→ Abb. 6.5)

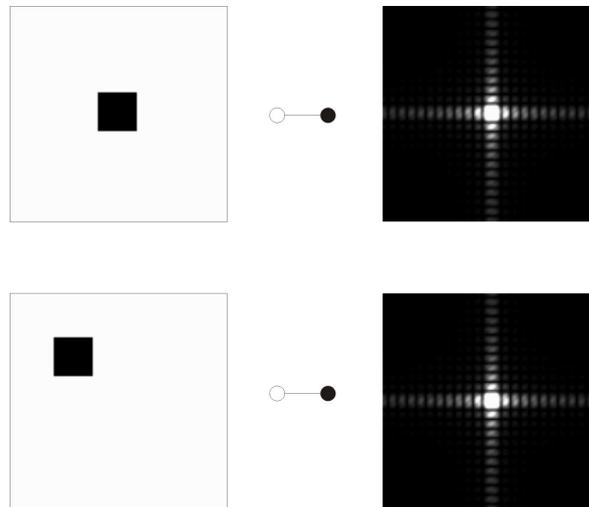


Abbildung 6.5: Verschiebungssatz

6.3.7 Anwendungen in der Bildverarbeitung

- Filterdesign
- Texturanalyse (lokale Frequenzen)
- Bildrestauration und -optimierung
- Bewegungsanalyse
- Gewinnung von Invarianten zur Mustererkennung (Fourierdeskriptoren)

6.3.8 Matrizenschreibweise

- Jede quadratische Matrix besitzt genau eine inverse Matrix und es gilt: $UU^{-1} = I$, wobei I die Einheitsmatrix ist.
- U ist orthogonal, wenn gilt: $U^{-1} = U^T$, wobei U^T die transponierte Matrix von U ist.
- U ist unitär, wenn gilt: $U^{-1} = U^{*T}$, wobei U^{*T} die adjungierte (transponierte und konjugiert-komplexe) Matrix von U ist.

Die Matrix der DFT gehört zur Klasse der unitären Matrizen. Für die Hintransformation gilt:

$$\vec{\alpha} = \frac{1}{\sqrt{N}} \cdot T_N^{ks} \cdot \vec{a}$$

Einer unitären Matrix entsprechend gilt für die Rücktransformation:

$$\vec{a} = \frac{1}{\sqrt{N}} \cdot (T_N^{ks})^{*T} \cdot \vec{\alpha} \quad (6.6)$$

mit $T_N^{ks} = e^{-2\pi i \frac{ks}{N}}$

6.3.9 Periodizitätseigenschaft

Die DFT zeigt eine charakteristische Periodizität:

$$e^{-2\pi i \frac{k}{N}} = e^{-2\pi i \frac{k+nN}{N}}, \quad n \in \mathbb{Z}$$

$$T_N^k = T_N^{(k+nN)}$$

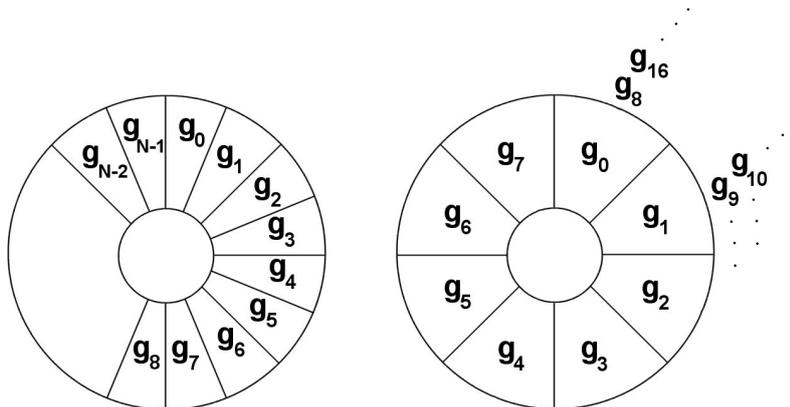


Abbildung 6.6: Links: Geometrische Darstellung der Periodizität mit dem Fourierreing. Rechts: Beispiel für $N = 8$.

Die Periodizität ist auch geometrisch darstellbar, indem man die Elemente kreisförmig in Segmente anordnet, wobei ein Kreis einer Periode entspricht. Abbildung 6.6 zeigt den sog. *Fourierreing* für die eindimensionale DFT.

Als Beispiel soll für $N = 8$ der Fourierreing und die entsprechende Matrixschreibweise

dargestellt werden (vgl. Abbildung 6.6 rechts):

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_7 \end{pmatrix} = \frac{1}{\sqrt{8}} \begin{pmatrix} T^0 & T^0 & T^0 & \dots & T^0 \\ T^0 & T^1 & T^2 & \dots & T^7 \\ T^0 & T^2 & T^4 & T^6 & \dots & T^6 \\ T^0 & T^3 & T^6 & T^1 & \dots & T^5 \\ \vdots & \vdots & & & \vdots & \\ T^0 & T^7 & & \dots & T^1 & \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_7 \end{pmatrix}$$

mit $T^{14} = T^6$, $T^{21} = T^5$, $T^{10} = T^2$, usw.

Für 2D gilt:

$$\alpha_{r,s} = \frac{1}{\sqrt{N_1 N_2}} \cdot T_{N_2} \cdot a_{k,l} \cdot T_{N_1}$$

mit $T_{N_1} = e^{-2\pi i \frac{ks}{N_1}}$, $T_{N_2} = e^{-2\pi i \frac{rl}{N_2}}$

Die Bildmatrix $B_{m,n}$ kann dabei a sein. Auf jeden Fall sind a , α Matrizen.

6.4 Diskrete unitäre Transformationen

6.4.1 Kosinus- und Sinustransformation

Nach der Eulerschen Formel kann man T_N^{ks} aus Gleichung (6.6) auflösen zu:

$$T_N^{ks} = \cos\left(2\pi \frac{ks}{N}\right) - i \sin\left(2\pi \frac{ks}{N}\right)$$

Die Kosinusfunktionen sind orthogonal, bilden aber kein vollständiges Basissystem in dem Funktionenraum (man kann keine ungeraden Funktionen damit abbilden). Ausweg: Man macht sich das Bild künstlich gerade (oder ungerade für die Sinusentwicklung) durch eine Halbierung der Wellenzahl oder Spiegelung des Signals im negativen Bereich (siehe Abbildung 6.7). Für die Kosinus- bzw. Sinustransformation kann ebenso die Matrixschreibweise verwendet werden:

$$C_N^{ks} = \sqrt{\frac{2}{N}} \cdot \cos\left(\frac{\pi ks}{N}\right)$$

$$S_N^{ks} = \sqrt{\frac{2}{N+1}} \cdot \sin\left(\frac{\pi(k+1)(s+1)}{N+1}\right)$$

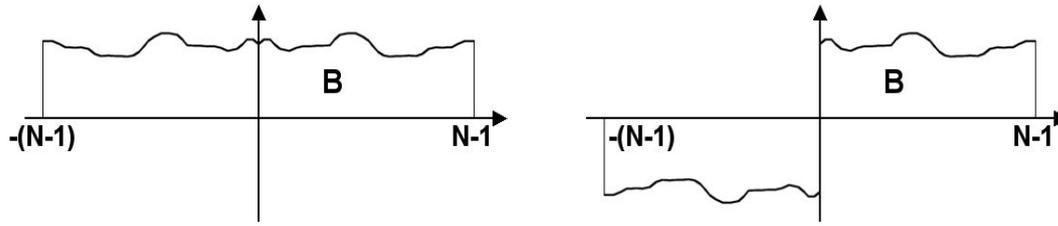


Abbildung 6.7: **Links:** Achsensymmetrische Spiegelung des Signals für die Kosinustwicklung. **Rechts:** Nullpunktsymmetrische Spiegelung des Signals für die Sinustwicklung.

Dann ist die zweidimensionale, diskrete Kosinustransformation (2D-DCT) definiert durch:

$$\alpha_{r,s} = \sum_{k=0}^{N_1-1} \sum_{l=0}^{N_2-1} a_{k,l} \cdot C_{N_1}^{ks} \cdot C_{N_2}^{lr}$$

Die Rücktransformation lautet:

$$a_{k,l} = \sum_{s=0}^{N_1-1} \sum_{r=0}^{N_2-1} \alpha_{r,s} \cdot C_{N_1}^{ks} \cdot C_{N_2}^{lr}$$

Analog folgend lautet die Sinustransformation:

$$\alpha_{r,s} = \sum_{k=0}^{N_1-1} \sum_{l=0}^{N_2-1} a_{k,l} \cdot S_{N_1}^{ks} \cdot S_{N_2}^{lr}$$

Die Rücktransformation ist dann für den Sinus:

$$a_{k,l} = \sum_{s=0}^{N_1-1} \sum_{r=0}^{N_2-1} \alpha_{r,s} \cdot S_{N_1}^{ks} \cdot S_{N_2}^{lr}$$

Im Gegensatz zur DFT rechnet die DCT nicht mit komplexen, sondern nur mit reellen Werten. Sie transformiert Bilddaten in eine Form, die gut komprimiert werden kann. Die Umsetzung in Hard- und Software kann effizient implementiert werden, weswegen sie Anwendung in Kodierungsverfahren, wie *JPEG* und *MPEG*, findet.

Die diskrete Kosinustransformation nutzt sog. *Basisbilder*, um die Informationen eines Bildes abzuspeichern. Dabei werden gleichgroße Blöcke einer Größe von maximal 8x8 Pixel gebildet, da sonst die Rechenzeiten zu hoch wären. Um die Informationen eines 8x8-Blockes abzuspeichern, sind insgesamt 64 Basisbilder nötig. Diese Basisbilder werden durch den Inhalt des zu transformierenden Bildes unterschiedlich gewichtet. Abbildung 6.8 veranschaulicht die Basisbilder der DCT.

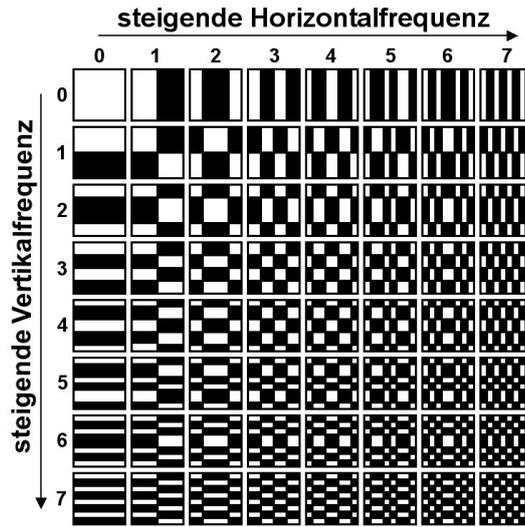


Abbildung 6.8: (nach: www [17]) Die 64 Basisbilder eines 8x8-Blockes. Mit der unterschiedlichen Wichtung der einzelnen Basisbilder ist jede Möglichkeit des Blockes realisierbar.

6.4.2 Walshtransformation

Die Rademacherfunktionen Die Rademacherfunktionen sind definiert durch:

$$rad_s(t) = \text{sign} \left[\sin \left(2^s \pi \frac{t}{T} \right) \right]$$

Die Funktionen mit dem Parameter s sind orthogonal aber nicht vollständig. So gibt es gerade Funktionen, die man damit nicht abbilden kann (z.B. $\text{sign} \left[\cos \left(2^s \pi \frac{t}{T} \right) \right]$). Abbildung 6.9 zeigt die ersten vier Rademacherfunktionen.

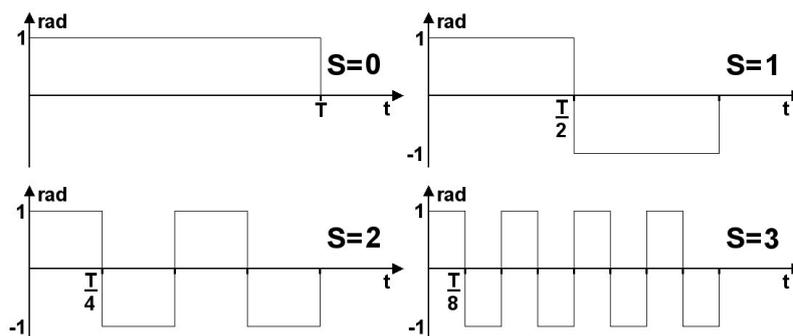


Abbildung 6.9: Die Rademacherfunktionen für $s = \{0, 1, 2, 3\}$.

Die Walshfunktionen Die Walshfunktionen sind die Erweiterung der Rademacherfunktionen zu einem vollständigen System. Sie sind definiert durch:

$$wal_s(t) = \frac{1}{\sqrt{T}} \prod_{m=0}^{\infty} [rad_{m+1}(t)]^{s_m^G}$$

man definiere: s_m^G : die m-te Stelle des Gray-Codes der Nummer s
 $s_m^G = s_m \oplus s_{m+1}$: \oplus bezeichnet die Addition Modulo 2
 m : ist die Nummer der Binärstelle (von rechts nach links)
 s_m : ist die m-te Dualstelle der Dualdarstellung der Zahl
 $s = \sum_{m=0}^{\infty} s_m \cdot 2^m$

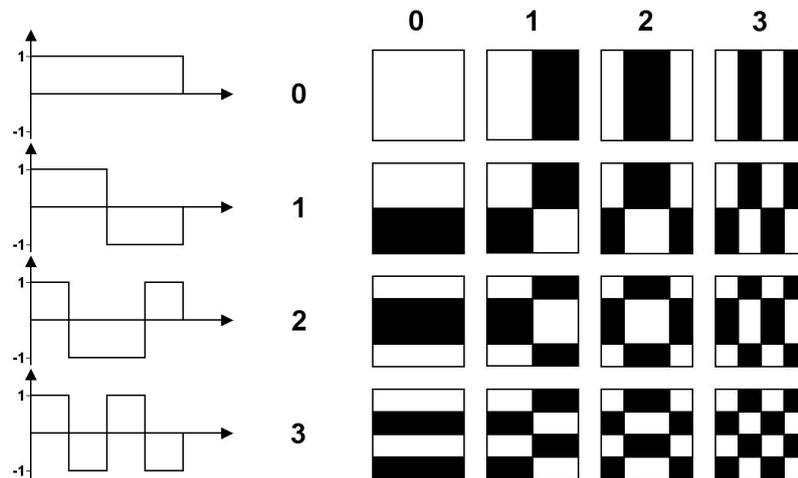


Abbildung 6.10: Die ersten 4 Walshfunktionen. **Links:** Eindimensional. **Rechts:** Zweidimensional.

Abbildung 6.10 zeigt die ersten 4 Walshfunktionen für den eindimensionalen und zweidimensionalen Fall.

Beispiel Man gehe von den BCD-Zahlen (linke Tabelle) aus und rechne sie um in die s_m^G -Zahlen (rechte Tabelle) mittels \oplus :

s_3	s_2	s_1	s_0	s
0	0	0	0	0
0	0	0	1	1
0	0	1	0	2
0	0	1	1	3
0	1	0	0	4
0	1	0	1	5

 \Rightarrow

s_3^G	s_2^G	s_1^G	s_0^G	s^G
0	0	0	0	0
0	0	0	1	1
0	0	1	1	2
0	0	1	0	3
0	1	1	0	4
0	1	1	1	5

Am Beispiel der Zahl 4 ergibt sich die folgende Berechnung:

$s = 4$	s_m^G
$s_0 = 0$	$s_0^G = 0$
$s_1 = 0$	$s_0^G = 0$
$s_1 = 0$	$s_1^G = 1$
$s_2 = 1$	$s_1^G = 1$
$s_2 = 1$	$s_2^G = 1$
$s_3 = 0$	$s_2^G = 1$

Für die Rademacher-Funktionen ergibt sich folgende Tabelle:

																m
$rad_m(t) :$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0
	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	1
	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	2
	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	3
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	4

Die Walshfunktionen haben dagegen folgende Werte:

																s
$wal_s(t) :$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0
	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	1
	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	2
	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	3
	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+	4
	+	+	-	-	-	-	+	+	-	-	+	+	+	+	-	5

Bequeme Darstellung Für $rad_{m+1}(t) = (-1)^{\tau_{m+1}}$, $\tau = \frac{t}{T}$ kann man die Walshfunktion folgendermaßen darstellen:

$$wal_s(\tau) = \frac{1}{\sqrt{T}} \prod_{m=0}^{\infty} (-1)^{\tau_{m+1} s_m^G} = \frac{1}{\sqrt{T}} (-1)^{\sum_{m=0}^{\infty} \tau_{m+1} s_m^G}$$

mit $\tau = \sum_{m=0}^{\infty} \tau_{m+1} \cdot 2^{-(m+1)}$

Ein Beispiel für $T = 8$ und mit $\tau = \frac{t}{T}$:

τ	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	$\frac{7}{8}$
2^0	0	0	0	0	0	0	0	0
2^{-1}	0	0	0	0	1	1	1	1
2^{-2}	0	0	1	1	0	0	1	1
2^{-3}	0	1	0	1	0	1	0	1

Man kann die Basisfunktionen (den Kern) auch in der Matrixschreibweise ausdrücken.

$$W = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}$$

Eine sehr interessante Eigenschaft der Walshfunktionen ist:

$$wal_s(t) \cdot wal_r(t) = wal_{s \oplus r}(t)$$

Zum Beispiel gilt $wal_0(t) \cdot wal_1(t) = wal_1(t)$ oder $wal_2(t) \cdot wal_3(t) = wal_1(t)$.

Die zweidimensionale Walshfunktionen in rechtwinkligen Koordinaten:

$$wal_{s,r}(t_1, t_2) = wal_s(t_1) \cdot wal_r(t_2)$$

Die diskrete Walsh-Transformation Die diskrete Walsh-Transformation ist definiert durch:

$$\alpha_s = \sum_{k=0}^{N-1} a_k \cdot wal_s(k) \quad , k = 0, 1, 2, \dots, 2^n - 1$$

$$\alpha_s = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} a_k \cdot (-1)^{\sum_{j=0}^{N-1} s_{n-1-j}^G \cdot k_j}$$

Die Rücktransformation wird beschrieben mit:

$$a_k = \frac{1}{\sqrt{2^n}} \sum_{s=0}^{2^n-1} \alpha_s \cdot (-1)^{\sum_{j=0}^{N-1} s_{n-1-j}^G \cdot k_j}$$

Wenn man die Formeln für die zweite Dimension ausweitet, so ergibt sich für die zweidimensionale Walsh-Transformation:

$$\alpha_{s,r} = \frac{1}{\sqrt{NM}} \sum_{k=0}^{2^n-1} \sum_{l=0}^{2^m-1} a_{k,l} \cdot (-1)^{\sum_{i=0}^{M-1} r_{n-1-i}^G \cdot l_i} \cdot (-1)^{\sum_{j=0}^{N-1} s_{n-1-j}^G \cdot k_j}$$

6.5 FFT

Die Fast Fourier Transformation (kurz FFT) ist ein Algorithmus, der die Berechnung der DFT erheblich vereinfacht und damit den Rechenaufwand verringert. Geht man von einer Größe von N Werten für die DFT aus, so ist der Rechenaufwand (Anzahl der Multiplikationen):

	Rechenaufwand für N Werte	Beispiel für $N = 2^{10}$
DFT	N^2	1048576 Operationen
FFT	$N \cdot \log_2(N)$	10240 Operationen

Diese Formel für den Rechenaufwand entsteht folgendermaßen: Wenn man den Ausgangsvektor mit der Länge N und einer Komplexität von N^2 in zwei Teilvektoren zerlegt, so ist die Komplexität nach der Zerlegung $2 \left(\frac{N}{2}\right)^2$. Dieses Prinzip heißt *divide and conquer* und wird sukzessive $\log_2(N)$ -mal durchgeführt bis schließlich nur noch Vektoren der Länge 1 übrig bleiben. Die Komplexität ist dann wie in der Tabelle genannt $N \cdot \log_2(N)$.

Ausgangspunkt des Algorithmus ist die Formel für die 1D-DFT aus Gleichung (6.5):

$$\alpha_s = \frac{1}{N} \sum_{k=0}^{N-1} a_k \cdot e^{-2\pi i \frac{ks}{N}} \quad (6.7)$$

Außerdem wir aus Gleichung (6.6) die Matrixschreibweise

$$T_N^{ks} = e^{-2\pi i \frac{ks}{N}} \quad , T = e^{-2\pi i}$$

zur Vereinfachung gebraucht.

Erste Zerlegung Die Eingangsfolge für alle a_k wird in zwei Teilfolgen zerlegt (siehe Abbildung 6.11). Dabei enthält die eine Folge alle geradzahligen und die Andere alle ungeradzahligen Indizes. Hinweis: im Folgenden wird der Vorfaktor $\frac{1}{N}$ weggelassen, da er für die Funktionsweise der FFT eine untergeordnete Rolle spielt. Aus (6.7) wird:

$$\alpha_s = \sum_{k=0}^{\frac{N}{2}-1} a_{2k} \cdot \underbrace{e^{-2\pi i \frac{2ks}{N}}}_{e^{-2\pi i \frac{ks}{N}}} + \sum_{k=0}^{\frac{N}{2}-1} a_{(2k+1)} \cdot \underbrace{e^{-2\pi i \frac{(2k+1)s}{N}}}_{e^{-2\pi i \frac{ks}{N}} \cdot e^{-2\pi i \frac{s}{N}}}$$

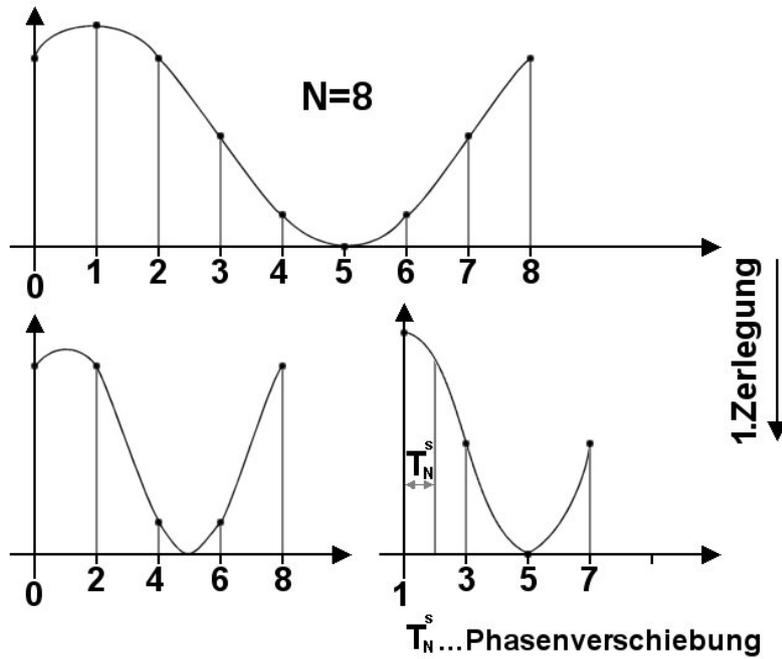


Abbildung 6.11: 1. Zerlegung der Eingangsfolge (oben) in zwei Teilfolgen (unten) bei der FFT. Die zweite Teilfolge hat eine Phasenverschiebung von T_N^s .

$$\alpha_s = \sum_{k=0}^{\frac{N}{2}-1} a_{2k} \cdot e^{-2\pi i \frac{ks}{N}} + \sum_{k=0}^{\frac{N}{2}-1} a_{(2k+1)} \cdot e^{-2\pi i \frac{ks}{N}} \cdot e^{-2\pi i \frac{s}{N}}$$

Verkürzt lässt sich in Matrixschreibweise schreiben:

$$\alpha_s = a_{2k} \cdot T_{\frac{N}{2}}^{ks} + a_{(2k+1)} \cdot T_{\frac{N}{2}}^{ks} \cdot T_N^s$$

Für $\alpha_g = a_{2k} \cdot T_{\frac{N}{2}}^{ks}$ und $\alpha_u = a_{(2k+1)} \cdot T_{\frac{N}{2}}^{ks}$ kann man vereinfachen:

$$\alpha_s = \alpha_g + \alpha_u \cdot \underbrace{T_n^s}_{\text{Phasenverschiebung}}$$

Die Teiltransformationen liefern einen Vektor mit $\frac{N}{2}$ Werte. Für die Transformation wird die doppelte Anzahl benötigt.

Die Symmetrieeigenschaft Aufgrund der Periodizität gilt:

$$\begin{aligned} e^{-2\pi i \left(\frac{s}{N} + \frac{N}{2}\right)} &= -e^{-2\pi i \frac{s}{N}} \\ T_N^{s+\frac{N}{2}} &= -T_N^s \\ \Rightarrow \alpha_s &= \alpha_g + \alpha_u \cdot T_N^s \\ \alpha_{s+\frac{N}{2}} &= \alpha_g - \alpha_u \cdot T_N^s \end{aligned}$$

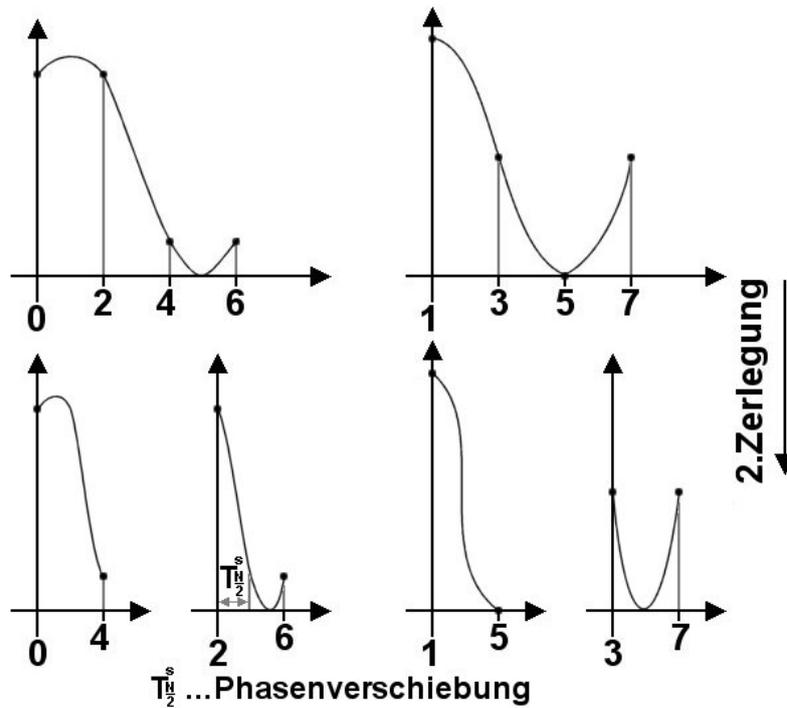


Abbildung 6.12: 2. Zerlegung der zwei Einzelfolgen (oben) in vier Teilfolgen (unten) bei der FFT. Auch hier tritt eine Phasenverschiebung auf: $T_{\frac{N}{2}}^s$.

Zweite Zerlegung Die sukzessive Halbierung der Vektorlänge wird $(\log_2(N))$ -mal durchgeführt bis ein Vektor mit der Länge 1 entsteht. Das Verfahren setzt voraus, dass N eine Zweierpotenz ist. Die zweite Zerlegung errechnet sich mit:

$$\alpha_s = \sum_{k=0}^{\frac{N}{4}-1} a_{4k} \cdot e^{-2\pi i \frac{4ks}{N}} + \sum_{k=0}^{\frac{N}{4}-1} a_{(4k+2)} \cdot e^{-2\pi i \frac{4ks}{N}} \cdot \underbrace{e^{-2\pi i \frac{2s}{N}}}_{T_{\frac{N}{2}}^s} + \sum_{k=0}^{\frac{N}{4}-1} a_{(4k+1)} \cdot e^{-2\pi i \frac{4ks}{N}} \cdot \underbrace{e^{-2\pi i \frac{s}{N}}}_{T_N^s} + \sum_{k=0}^{\frac{N}{4}-1} a_{(4k+3)} \cdot e^{-2\pi i \frac{4ks}{N}} \cdot T_{\frac{N}{2}}^s \cdot T_N^s$$

Das Flussdiagramm der FFT ist in Abbildung 6.13 dargestellt.

Triviale Multiplikationen Am Beispiel mit $N = 8$ sollen die trivialen Multiplikationen erläutert werden. Wenn man $N = 8$ in die Gleichung für T_N^S einsetzt, so ergibt sich:

$$T_N^S = e^{-2\pi i \frac{S}{8}} = e^{-\pi i \frac{S}{4}} = \cos\left(\frac{\pi}{4}S\right) - i \sin\left(\frac{\pi}{4}S\right)$$

Für $S = \{0, 1, 2, 3\}$ erhält man:

$$T_N^0 = 1 - 0 \quad ; \quad T_N^1 = e^{-\pi i \frac{1}{4}} \quad ; \quad T_N^2 = 0 - i \quad ; \quad T_N^3 = e^{-\pi i \frac{3}{4}}$$

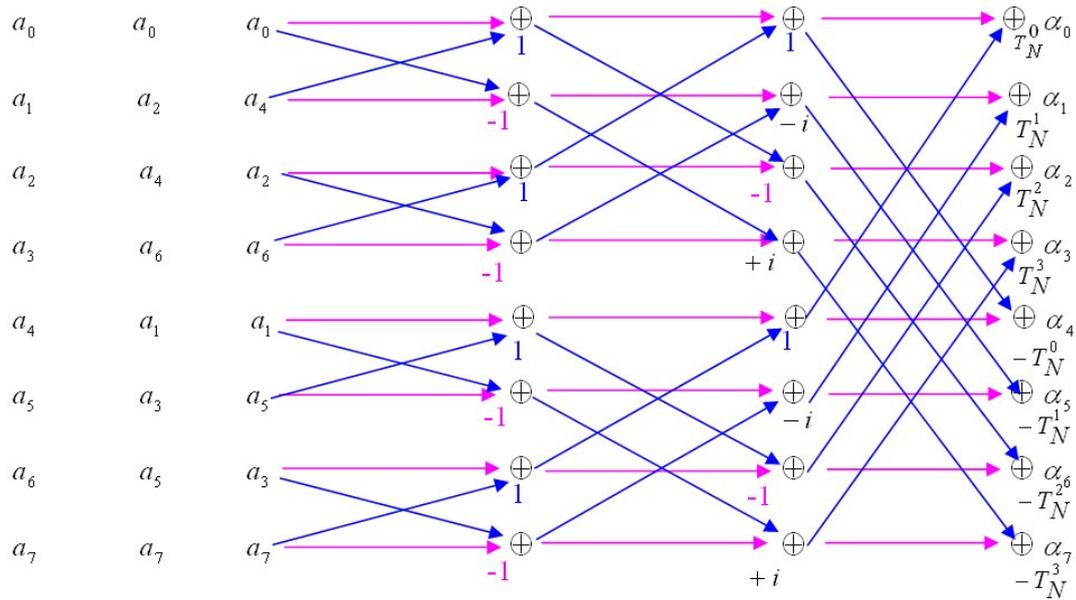


Abbildung 6.13: Flussdiagramm der FFT.

$T_{\frac{N}{2}}^S$ ergibt sich mit dem Einsetzen von $N = 8$ zu:

$$T_{\frac{N}{2}}^S = e^{-\pi i \frac{S}{2}} = \cos\left(\frac{\pi}{2}S\right) - i \sin\left(\frac{\pi}{2}S\right)$$

Für $S = \{0, 1\}$ erhält man:

$$T_{\frac{N}{2}}^0 = 1 - 0 \quad ; \quad T_{\frac{N}{2}}^1 = 0 - i$$

In den ersten beiden Schritten des Zusammensetzens werden nur triviale Multiplikationen mit 1 oder i durchgeführt. In den weiteren Schritten verringern sich die trivialen Multiplikationen jeweils um die Hälfte. Der Grund liegt im Benutzen von Zwischenergebnissen. Der Aufwand sinkt auf $\frac{N}{2} \log_2(N)$ komplexe Multiplikationen. Zusätzlich hat man in jeder Stufe N Additionen und Subtraktionen, was insgesamt einen Aufwand von $N \log_2(N)$ ergibt. Wenn alle trivialen Multiplikationen vermeidbar wären, so hätte man $\frac{N}{2}(\log_2(N) - 3)$ komplexe Multiplikationen bzw. $\frac{1}{2}(\log_2(N) - 3)$ pro Pixel.

Beispiel: Für $N = 512$ braucht die DFT $512 \cdot 512$ komplexe Multiplikationen. Die FFT kommt dagegen mit 3 komplexen Multiplikationen und 8 komplexen Additionen pro Element aus. Dies entspricht 12 reellen Multiplikationen und 22 reellen Additionen.

2D-FFT Die Formel der zweidimensionalen FFT lautet:

$$\alpha_{r,s} = \frac{1}{N_1 N_2} \sum_{l=0}^{N_2-1} \sum_{k=0}^{N_1-1} a_{l,k} \cdot e^{-2\pi i \frac{ks}{N_1}} \cdot e^{-2\pi i \frac{lr}{N_2}}$$

Aufgrund der Separierbarkeitseigenschaft des DFT-Kerns kann man die 2D-FFT in zwei eindimensionale Transformationen zerlegen:

$$\text{Zeilentransformation: } \alpha_{l,s} = \frac{1}{N_1} \sum_{k=0}^{N_1-1} a_{l,k} \cdot e^{-2\pi i \frac{ks}{N_1}}$$

$$\text{Spaltentransformation: } \alpha_{r,s} = \frac{1}{N_2} \sum_{l=0}^{N_2-1} \alpha_{l,s} \cdot e^{-2\pi i \frac{lr}{N_2}}$$

Für jede Teiltransformation kann der 1D-FFT-Algorithmus eingesetzt werden. Für $N_1 = N_2 = M$ hat man $2 \cdot M \cdot \frac{M}{2} \cdot \log_2(M)$ statt M^4 Operationen. Für $M = 512$ braucht man 2 359 296 komplexe Multiplikationen statt 68 719 476 736.

Kapitel 7

Nachbarschaftsoperatoren

Bei den *linearen Nachbarschaftsoperatoren* wird in diesem Kapitel die Faltung und die Transferfunktion vorgestellt. Außerdem werden die wichtigsten Operatortypen und ihre Eigenschaften erklärt. Danach sollen uns die *nichtlinearen Nachbarschaftsoperatoren* zum Vergleich interessieren.

Ein Nachbarschaftsoperator ist ein Operator, der die Grau- bzw. Farbwerte in der Umgebung eines Pixels erfasst, diese geeignet miteinander verknüpft und das Ergebnis auf das Pixel zurückschreibt (siehe Abbildung 7.1). Er gehört zu den lokalen Operatoren; zum Vergleich, es gibt noch Punktoperatoren (ändern die Bildpunkte bzw. deren Inhalt unabhängig von der Nachbarschaft) und globale Operatoren (abhängig vom gesamten Bild). Wenn man einen Nachbarschaftsoperator auf ein Bild anwendet, so entsteht wiederum ein Bild, jedoch mit veränderten Eigenschaften. Es können somit z.B. folgende Ziele erreicht werden:

- Tiefpassfilterung und Glättung
- Hochpassfilterung und Kantendetektion

7.1 Lineare verschiebungsinvariante Filter

Bei einem linearen Filter werden die Werte der Nachbarbildpunkte in einer *Linear-kombination* miteinander verknüpft und auf das aktuelle Zielpixel geschrieben. Da sie unabhängig von der Position im Bild wirken (verschiebungsinvariant), nennt man sie auch *linear homogen* oder LSI: *Linear Shift Invariant*. Sie besitzen begrenzte Träger, weshalb sie auch als *FIR¹-Filter* bezeichnet werden.

¹FIR...finite impulse response

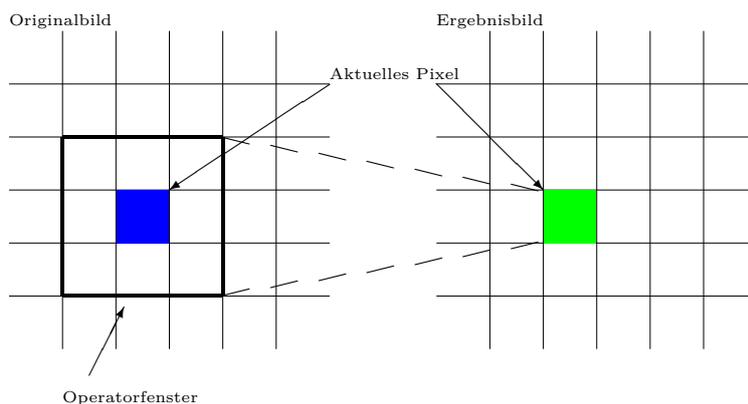


Abbildung 7.1: Prinzip lokaler Operatoren: Die zu einem aktuellen Pixel benachbarten Grauwerte werden beliebig miteinander verknüpft und das Ergebnis an die gleiche Stelle im Zielbild geschrieben.

7.1.1 Grundlagen

Die Faltung Die wichtigste mathematische Operation bei den Nachbarschaftsoperatoren ist die Faltung. Für den kontinuierlichen Fall ist sie definiert durch:

$$(f * g)(\vec{y}) := \int f(\vec{x}) \cdot g(\vec{y} - \vec{x}) \cdot (dx)^n$$

Für den diskreten Fall lautet die Definition:

$$(f * g)(\vec{m}) := \sum_{x \in X} f(\vec{x}) \cdot g(\vec{m} - \vec{x})$$

Die Eigenschaften der Faltung sind in der folgenden Tabelle zusammengefasst:

Linearität	$f * (\alpha \cdot g + \beta \cdot h) = \alpha \cdot f * g + \beta \cdot f * h$
Additivität	$f * g + h * g = (f + h) * g$
Kommutativität	$f * g = g * f$
Assoziativität	$f * (g * h) = (f * g) * h$

Faltung mit Filtermasken Die Anwendung eines linearen verschiebungsinvarianten Operators auf ein Bild lässt sich durch die Faltung des Bildes mit einer Faltungsmaske realisieren, deren Koeffizienten der Impulsantwort des Operators entsprechen. Für den eindimensionalen Fall gilt:

$$b'_m = (b * h)_m = \sum_k^{M-1} b_k \cdot h_{m-k} = \sum_k^{M-1} b_{m-k} \cdot h_k$$

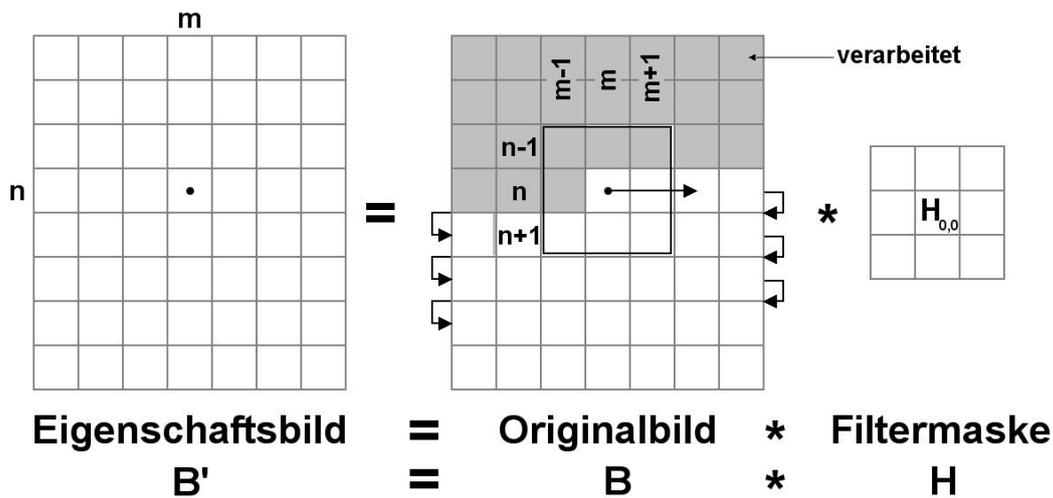


Abbildung 7.2: Anwendung eines 3x3-Operators auf ein Bild B . Die Filtermaske wird von links oben nach rechts unten über das Bild geschoben.

Der zweidimensionale Fall trifft für die Anwendung eines Operators auf ein Bild zu:

$$\begin{aligned}
 B'_{m,n} &= (B * H)_{m,n} = \sum_{k=-K}^K \sum_{l=-K}^K B_{k,l} \cdot H_{m-k,n-l} \\
 &= \sum_{k=-K}^K \sum_{l=-K}^K B_{m-k,n-l} \cdot H_{k,l}
 \end{aligned} \tag{7.1}$$

Wobei K die „halbe“ Höhe/Breite einer quadratischen Filter-Maske (Kernel) ist.

Im folgenden Beispiel ist eine 3x3-Maske gegeben mit $K = 1 \Rightarrow k = \{-1, 0, 1\}$.

$B_{m-1,n-1}$	$B_{m,n-1}$	$B_{m+1,n-1}$
$H_{1,1}$	$H_{0,1}$	$H_{-1,1}$
$B_{m-1,n}$	$B_{m,n}$	$B_{m+1,n}$
$H_{1,0}$	$H_{0,0}$	$H_{-1,0}$
$B_{m-1,n+1}$	$B_{m,n+1}$	$B_{m+1,n+1}$
$H_{1,-1}$	$H_{0,-1}$	$H_{-1,-1}$

In der Abbildung 7.2 ist verdeutlicht, wie die 3x3-Maske auf ein Bild B angewendet wird.

Symmetrien Fast alle Filter, die in der Bildverarbeitung verwendet werden, besitzen gerade oder ungerade Symmetrie bezüglich des Mittelpunktes.

für 1-dimensionale Signale/Filter	
Gerade mit $(2K+1)$ Koeffizienten	$h(k) = [h_k, \dots, h_1, h_0, h_1, \dots, h_k]$ $b'_m = h_0 \cdot b_m + \sum_{k=1}^K h_k \cdot (b_{m-k} + b_{m+k})$
Gerade mit $(2K)$ Koeffizienten	$h(k) = [h_k, \dots, h_1, h_1, \dots, h_k]$ $b'_m = \sum_{k=1}^K h_k \cdot (b_{m-k} + b_{m+k})$
Ungerade mit $(2K+1)$ Koeffizienten	$h(k) = [h_k, \dots, h_1, h_0, -h_1, \dots, -h_k]$ $b'_m = \sum_{k=1}^K h_k \cdot (b_{m-k} - b_{m+k})$
Ungerade mit $(2K)$ Koeffizienten	$h(k) = [h_k, \dots, h_1, -h_1, \dots, -h_k]$ $b'_m = \sum_{k=1}^K h_k \cdot (b_{m-k} - b_{m+k})$

Bei den letzten zwei Reihen der Tabelle ist b'_m deswegen identisch, da h_0 zu 0 gewählt wird. Für 2-dimensionale Filter ergibt sich z.B. für gerade Symmetrie:

$$\begin{aligned}
B'_{m,n} = & H_{0,0} B_{m,n} + \sum_{k=1}^K H_{k,0} \cdot (B_{m-k,n} + B_{m+k,n}) + \sum_{l=1}^K H_{0,l} \cdot (B_{m,n-l} + B_{m,n+l}) \\
& + \sum_{k=1}^K \sum_{l=1}^K H_{k,l} \cdot (B_{m-k,n-l} + B_{m-k,n+l} + B_{m+k,n-l} + B_{m+k,n+l})
\end{aligned}$$

Dagegen erhält man für horizontal ungerade und vertikal gerade Symmetrie:

$$\begin{aligned}
B'_{m,n} = & \sum_{k=1}^K H_{k,0} \cdot (B_{m-k,n} - B_{m+k,n}) \\
& + \sum_{k=1}^K \sum_{l=1}^K H_{k,l} \cdot (B_{m-k,n-l} - B_{m+k,n-l} + B_{m-k,n+l} - B_{m+k,n+l})
\end{aligned}$$

Formalismus: Die Anwendung eines Operators auf ein Bild liefert ein Eigenschaftsbild B' . Sie wird wie folgt ausgedrückt:

$$B' = OB = O\{\delta\} * B$$

In den folgenden Abschnitten wird zur Vereinfachung $O\{\delta\}$ durch O ersetzt:

$$B' = O * B$$

Die Transferfunktion Laut dem Faltungstheorem lässt sich die Faltung des Bildes mit einer Faltungsmaske durch die Multiplikation deren Fouriertransformierten ersetzen. Damit kann der Einfluss der Faltungsmaske auf die Bildstruktur analytisch untersuchen und näher charakterisiert werden.

Die Fouriertransformierte eines Filters heißt *Transferfunktion*. Für einen gegebenen

eindimensionalen Filter

$$h(k) = [h_0, h_1, \dots, h_k] = \sum_{k=0}^K h_k \cdot \delta(k - k_0)$$

ergibt sich die Fouriertransformierte $h(k) \circ \bullet \hat{h}(\tilde{s})$ für $(K+1)$ Koeffizienten zu:

$$\hat{h}(s) = \sum_{k=0}^K h_k \cdot e^{-i2\pi \frac{(k-k_0)s}{N}} \quad (7.2)$$

Mit $\tilde{s} = \frac{2s}{N}$ ergeben sich für die Symmetrien folgende Vereinfachungen:

Gerade mit $(2K+1)$ Koeffizienten	
$h(k) = [h_k, \dots, h_1, h_0, h_1, \dots, h_k]$	$\circ \bullet \hat{h}(\tilde{s}) = h_0 + 2 \sum_{k=1}^K h_k \cdot \cos(\pi k \tilde{s})$
Gerade mit $(2K)$ Koeffizienten	
$h(k) = [h_k, \dots, h_1, h_1, \dots, h_k]$	$\circ \bullet \hat{h}(\tilde{s}) = 2 \sum_{k=1}^K h_k \cdot \cos \left[\pi \left(k - \frac{1}{2} \right) \tilde{s} \right]$
Ungerade mit $(2K+1)$ Koeffizienten	
$h(k) = [h_k, \dots, h_1, 0, -h_1, \dots, -h_k]$	$\circ \bullet \hat{h}(\tilde{s}) = 2i \sum_{k=1}^K h_k \cdot \sin(\pi k \tilde{s})$
Ungerade mit $(2K)$ Koeffizienten	
$h(k) = [h_k, \dots, h_1, -h_1, \dots, -h_k]$	$\circ \bullet \hat{h}(\tilde{s}) = 2i \sum_{k=1}^K h_k \cdot \sin \left[\pi \left(k - \frac{1}{2} \right) \tilde{s} \right]$

Bei einer geraden Koeffizientenanzahl $(2K)$ wird das Ergebnis des Filters wegen der Symmetrie auf einen Zwischengitterpunkt abgebildet, d. h. um $\frac{(2k-1)}{2}$ verschoben.

Auch für den zweidimensionalen Fall kann eine Transferfunktion aufgestellt werden.

Die Formel für das Beispiel einer Maske mit gerader Symmetrie $(2K+1)(2K+1)$ lautet:

$$\begin{aligned} \hat{H}_{\tilde{r}, \tilde{s}} = & H_{0,0} + 2 \sum_{k=1}^K H_{k,0} \cdot \cos(\pi k \tilde{s}) + 2 \sum_{l=1}^K H_{0,l} \cdot \cos(\pi l \tilde{r}) \\ & + 4 \sum_{k=1}^K \sum_{l=1}^K H_{k,l} \cdot \cos(\pi l \tilde{r}) \cdot \cos(\pi k \tilde{s}) \end{aligned}$$

Die Eigenfunktionen oder Eigenbilder Spezielle Typen von Bildern für die die Aussage

$$O * B = \lambda * B$$

gilt, nennt man *Eigenbilder*. Die Anwendung eines Operators auf ein Bild ist die Multiplikation mit dem Faktor λ , welcher *Eigenwert* des Operators heißt. Die Basisfunktionen der DFT sind die Eigenfunktionen für alle LSI-Operatoren und die komplexen Fourierkoeffizienten sind ihre Eigenwerte.

$$B_{m,n}^{r,s} = e^{-i2\pi \frac{mr}{N}} \cdot e^{-i2\pi \frac{ns}{M}}$$

Der Beweis erfolgt über das Faltungstheorem:

$$B' = OB = O * B$$

$$O * B^{kr} \quad \circ \text{---} \bullet \quad \hat{O}^{kr} \hat{B}^{kr} \quad \bullet \text{---} \circ \quad \hat{O}^{kr} \cdot B^{kr}$$

Beispielsweise sei die Ableitung $\frac{\partial}{\partial t} e^{i\omega_0 t}$ gegeben. Dabei ergeben sich folgende Fouriertransformationen:

- $h = \frac{\partial}{\partial t} \quad \circ \text{---} \bullet \quad \hat{h} = i\omega$
- $g(t) = e^{i\omega_0 t} \quad \circ \text{---} \bullet \quad \hat{g} = \delta(\omega - \omega_0)$

Die Multiplikation der Fouriertransformierten lautet:

$$\hat{h} \hat{g} = \hat{h}_{\omega_0} \hat{g} = i\omega_0 \cdot \delta(\omega - \omega_0)$$

Mit der Rücktransformation und unter Beachtung der Ausblendeigenschaft des Dirac-Impulses erhält man:

$$i\omega_0 \cdot \delta(\omega - \omega_0) \quad \bullet \text{---} \circ \quad i\omega_0 \cdot e^{i\omega_0 t} = \frac{\partial e^{i\omega_0 t}}{\partial t} = \hat{h} g(t)$$

7.1.2 Glättung

Glättungsfilter haben die Aufgabe, Strukturen mit hohen Frequenzen abzuschwächen. Dabei ist ein Sonderfall die Glättung mit normierter Mittelung, bei der die Summe der Filterkoeffizienten gleich eins ist.

Rechteckfilter Die einfachste Filter zur Glättung von Bildern sind die Rechteckfilter. Dazu gehört beispielsweise folgendes 1x3-Filter:

$${}^3R_k = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

Der Skalierungswert $\frac{1}{3}$ wird gewählt, damit der Grauwert in einer homogenen Umgebung erhalten bleibt. Ein Beispiel für die Glättung mit dem obigen Filter ist:

$$\begin{bmatrix} \cdot & \vdots & \vdots & \vdots & \vdots & \cdot \\ \dots & 0 & 0 & 1 & 1 & \dots \\ \dots & 0 & 0 & 1 & 1 & \dots \\ \dots & 0 & 0 & 1 & 1 & \dots \\ \cdot & \vdots & \vdots & \vdots & \vdots & \cdot \end{bmatrix} * {}^3R_k = \begin{bmatrix} \cdot & \vdots & \vdots & \vdots & \vdots & \cdot \\ \dots & 0 & \frac{1}{3} & \frac{2}{3} & 1 & \dots \\ \dots & 0 & \frac{1}{3} & \frac{2}{3} & 1 & \dots \\ \dots & 0 & \frac{1}{3} & \frac{2}{3} & 1 & \dots \\ \cdot & \vdots & \vdots & \vdots & \vdots & \cdot \end{bmatrix}$$

Wie man sieht, werden scharfe Kanten in rampenartige Übergänge umgewandelt. Die Transferfunktion des obigen Filters mit $(2K+1)$ Elementen und gerader Symmetrie lautet mit Gleichung (7.2):

$${}^3R_k = \left[\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right] \circ \bullet \quad {}^3\hat{R}_{n\tilde{s}} = \frac{1}{3} + \frac{2}{3} \cos(\pi n\tilde{s})$$

mit $n\tilde{s} = \frac{\tilde{s}}{\tilde{s}_{max}}$ als normierte Frequenz mit $n\tilde{s} = [0, \dots, 1]$

Für $n\tilde{s} = \frac{2}{3}$ gilt ${}^3\hat{R} = \frac{1}{3} + \frac{2}{3} \cdot \left(-\frac{1}{2}\right) = 0$. Die Transferfunktion hat also bei dieser Frequenz einen Nulldurchgang und ist im Intervall $\frac{2}{3} < n\tilde{s} \leq 1$ negativ, was eine Phasendrehung von 180° bedeutet.

Das Rechteckfilter mit 5 Elementen: ${}^5R_k = \left[\frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \right]$. Mit Gleichung (7.2) erhält man wieder die Transferfunktion:

$${}^5R_k = \left[\frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \right] \circ \bullet \quad {}^5R_{n\tilde{s}} = \frac{1}{5} + \frac{2}{5} [\cos(\pi n\tilde{s}) + \cos(2\pi n\tilde{s})]$$

Bei dieser Transferfunktion liegen die Nullstellen bei 0,4 und 0,8. Beide Transferfunktionen sind zum Vergleich in Abbildung 7.3 zu betrachten.

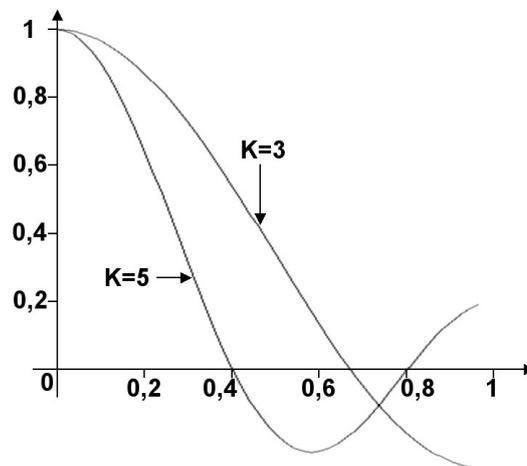


Abbildung 7.3: Dargestellt sind die Transferfunktionen ${}^3\hat{R}_{n\tilde{s}}$ für $K = 3$ und ${}^5\hat{R}_{n\tilde{s}}$ für $K = 5$ in Abhängigkeit von $n\tilde{s}$. Durch die Nulldurchgänge ergeben sich Phasenverschiebungen von 180° .

Ein Beispiel für ein zweidimensionales Rechteckfilter ist:

$${}^3R_{k,l} = {}^3R_k * {}^3R_l = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{3} [1 \quad 1 \quad 1] * \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Die Transferfunktion dieses Filter errechnet sich zu:

$${}^3R_{k,l} \circ \bullet \quad {}^3\hat{R}_{n\tilde{s}, n\tilde{r}} = \left[\frac{1}{3} + \frac{2}{3} \cos(\pi n\tilde{s}) \right] \left[\frac{1}{3} + \frac{2}{3} \cos(\pi n\tilde{r}) \right]$$

Vorteil der Rechteckoperatoren ist die schnelle Berechnung. Sie können durch rekursive Berechnung realisiert werden, so dass unabhängig von der Filtergröße immer nur drei Rechenoperationen benötigt werden. Man könnte die folgende Gleichung deuten als „nächstes Pixel addieren und letztes Pixel subtrahieren“.

$$b'_n = b'_{n-1} + \frac{1}{2k+1}(b_{n+k} - b_{n-k-1})$$

Nachteilig wirkt sich jedoch das Oszillieren der Amplitudendämpfung im Spektrum aus.

Binomialfilter Filter mit einem besseren Dämpfungsverhalten sind die Binomialfilter. Deren Amplitudendämpfung nimmt monoton mit der Frequenz zu. Die einfachste Glättungsmaske ist gegeben durch: $G^1 = \frac{1}{2}[1 \ 1]$. Größere Glättungsmasken können mit dem Pascalschen Dreieck (Abschnitt 4.4) berechnet werden oder durch das $(P-1)$ -mal sukzessive Anwenden des G^1 -Filters ($\cong G^P$ -Filter).

$$G^P = \underbrace{G^1 G^1 \dots G^1}_P = \frac{1}{2^P} \underbrace{\left[\begin{array}{cc} 1 & 1 \end{array} \right] * \left[\begin{array}{cc} 1 & 1 \end{array} \right] * \dots * \left[\begin{array}{cc} 1 & 1 \end{array} \right]}_P$$

Beispiele für gerade Binomialfilter mit $(2K+1)$ Elementen sind:

- $G^2 = \frac{1}{4} \left[\begin{array}{ccc} 1 & 2 & 1 \end{array} \right]$
- $G^4 = \frac{1}{16} \left[\begin{array}{ccccc} 1 & 4 & 6 & 4 & 1 \end{array} \right]$

Für die Transferfunktion wird ein Binomialfilter mit $(2K)$ Elementen und gerader Symmetrie betrachtet. Die Transferfunktion von G^1 lautet:

$$G^1 \circ \bullet \quad \hat{G}^1 = \cos\left(\frac{\pi}{2} n\tilde{s}\right)$$

Da $0 \leq n\tilde{s} \leq 1$, liegt das Argument des Kosinus im Intervall $I = [0; \frac{\pi}{2}]$, so dass der Kosinus Werte zwischen 0 und 1 ergibt. Wenn man größere Binomialfilter bildet, so

lautet die Transferfunktion:

$$G^P = \underbrace{G^1 G^1 \dots G^1}_P$$

$$\circ \bullet \hat{G}^P = \underbrace{\cos\left(\frac{\pi}{2} n \tilde{s}\right) \cos\left(\frac{\pi}{2} n \tilde{s}\right) \dots \cos\left(\frac{\pi}{2} n \tilde{s}\right)}_P = \cos^P\left(\frac{\pi}{2} n \tilde{s}\right)$$

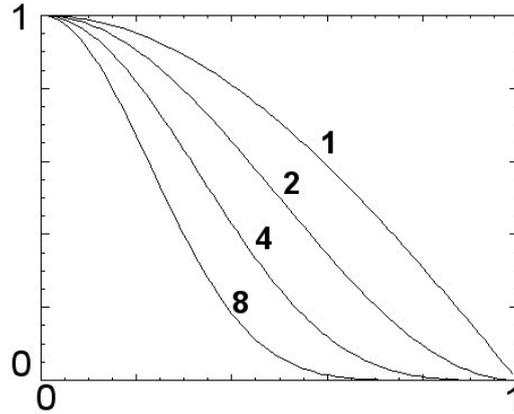


Abbildung 7.4: Transferfunktionen der Binomialfilter G^1, G^2, G^4, G^8 .

Beispielsweise gilt für G^2 :

$$G^2 = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \quad \circ \bullet \hat{G}^2 = \cos^2\left(\frac{\pi}{2} n \tilde{s}\right) = \frac{1}{2} + \frac{1}{2} \cos(\pi n \tilde{s})$$

Den Einfluss des Parameters P kann man in Abbildung 7.4 betrachten.

Die zweidimensionalen Binomialfilter entstehen durch Faltung eines horizontalen mit einem vertikalen Binomialfilter:

$$G_{k,l}^P = G_k^P * G_l^P = \frac{1}{2^P} [\dots] * \frac{1}{2^P} \begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

Ein Beispiel dazu:

$$G_{k,l}^2 = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} * \frac{1}{4} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Die Transferfunktion der zweidimensionalen Maske $G_{k,l}^2$ lautet:

$$\hat{G}_{n \tilde{s}, n \tilde{r}}^2 = \cos^2\left(\frac{\pi}{2} n \tilde{s}\right) \cos^2\left(\frac{\pi}{2} n \tilde{r}\right) = \left[\frac{1}{2} + \frac{1}{2} \cos(\pi n \tilde{s})\right] \left[\frac{1}{2} + \frac{1}{2} \cos(\pi n \tilde{r})\right]$$

Abbildung 7.5 zeigt zum Vergleich die zweidimensionalen Transferfunktionen des obigen Binomialfilters mit einem Rechteckfilter. Man vergleiche die Darstellung auch mit

den eindimensionalen Transferfunktionen aus Abbildung 7.4 und Abbildung 7.3.

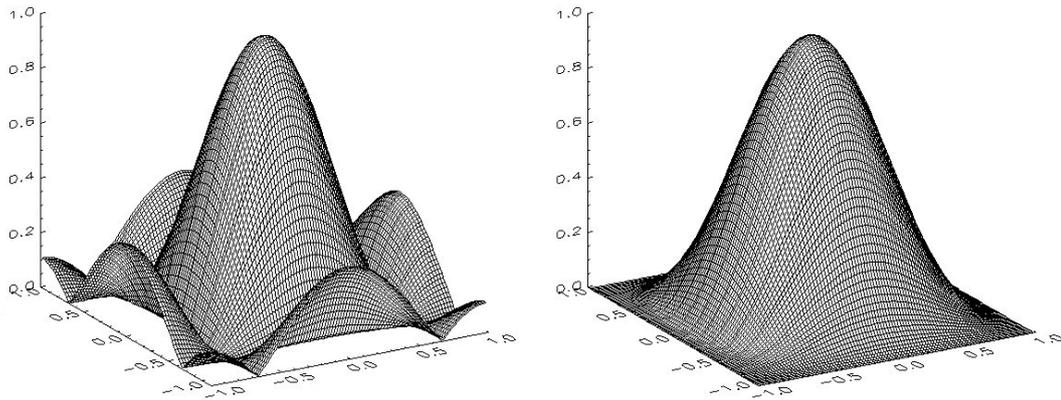


Abbildung 7.5: Dargestellt sind jeweils die Beträge der Transferfunktion in Abhängigkeit von \tilde{r} und \tilde{s} . **Links** TF vom Rechteckfilter $\frac{1}{9}([1 \ 1 \ 1] * [1 \ 1 \ 1]^T)$. **Rechts** TF vom Binomialfilter $\frac{1}{16}([1 \ 2 \ 1] * [1 \ 2 \ 1]^T)$, der bessere Eigenschaften besitzt, da seine TF mit zunehmender Wellenzahl monoton fällt.

Die Binomialglättung lässt sich schnell berechnen, wenn man Zwischenergebnisse benutzt. Abbildung 7.6 zeigt die Idee für eine G^4 -Maske.

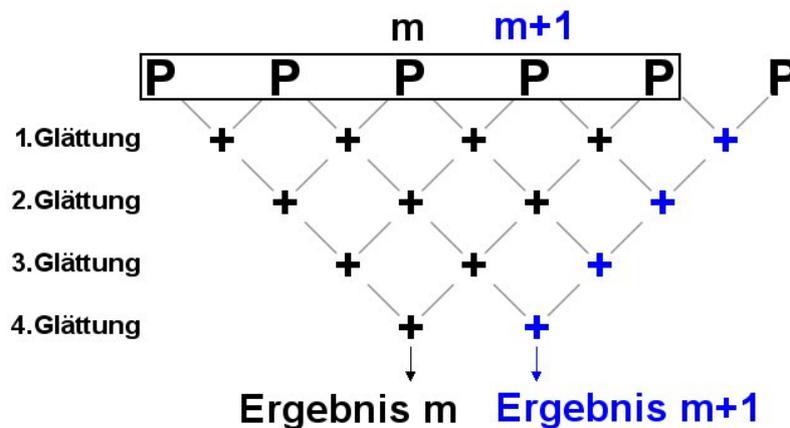


Abbildung 7.6: Schnelle Berechnung der Binomialglättung. Die Anwendung einer G^4 -Maske wird aufgeteilt in eine vierfache Anwendung der elementaren G^1 -Maske. Bei dem Berechnen des Ergebnisses von $m + 1$ wird die linke Diagonale abgezogen und die rechte Diagonale addiert, so dass viele Zwischenergebnisse genutzt werden.

Um die unterschiedliche Wirkung von Rechteckfiltern und Binomialfiltern zu zeigen, ist in Abbildung 7.7 die Anwendung von 3×3 - und 7×7 -Rechteckoperatoren bzw. Binomialoperatoren auf ein Bild dargestellt. Da in diesen Bildern sich der Grauwert in konzentrischen Kreisen von der Bildmitte zum Bildrand sinusförmig ändert, wobei sich die Wellenzahl vergrößert, kann man die Phasenverschiebung bei den Rechteckoperatoren gut erkennen.

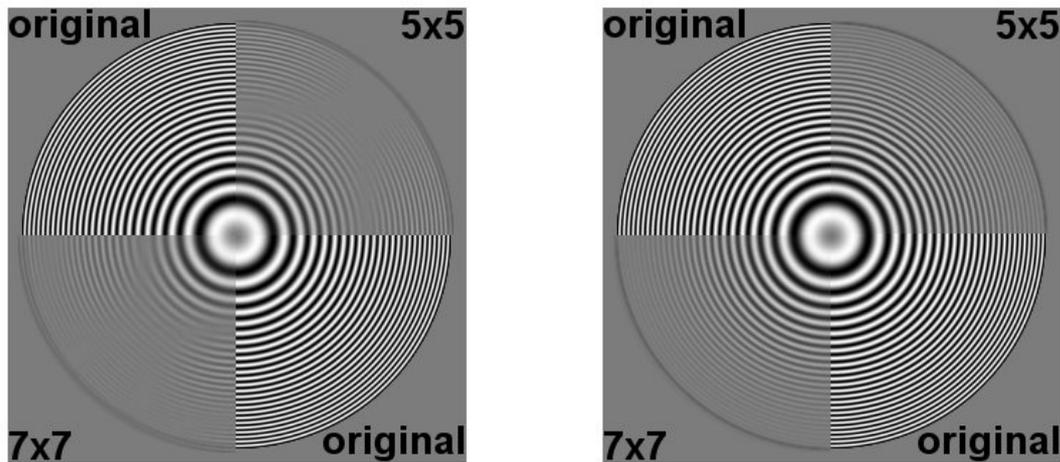


Abbildung 7.7: Vergleich der Glättungswirkung bei Rechteckfiltern und Binomialfiltern. Die Wellenzahl nimmt zum Bildrand hin zu, so dass die Operatorwirkung gut erkennbar ist. Im jeweiligen Bild wurde rechts oben eine 5x5-Glättungsmaske und links unten ein 7x7-Filter angewendet. **Links** Rechteckfilter - es sind Phasenverschiebungen von 180° zu erkennen. **Rechts** Binomialfilter - es sind aufgrund besserer Eigenschaften keine Phasenverschiebungen vorhanden.

7.1.3 Kantendetektion

Ableitungen Kanten können durch Erkennung von Grauwertänderungen gefunden werden. Man macht sich die Ableitung zu Nutze, welche die Grauwertanstiege im Bild liefert. Maxima und Minima zeigen die Kanten an. Man unterscheidet bei (2K) Maskenelementen zwischen Rückwärtsdifferenz:

$$-D_m = \frac{\partial B_{m,n}}{\partial m} \approx \frac{B_{m,n} - B_{m-\Delta m,n}}{\Delta m}$$

und Vorwärtsdifferenz:

$$+D_m = \frac{\partial B_{m,n}}{\partial m} \approx \frac{B_{m+\Delta m,n} - B_{m,n}}{\Delta m}$$

Die zugehörige (2K) Maske lautet:

$$-D_k = \left[\begin{array}{cc} 1 \bullet & -1 \end{array} \right]$$

$$+D_k = \left[\begin{array}{cc} 1 & -1 \bullet \end{array} \right]$$

• bezeichnet den Punkt, an den das Ergebnis zurückgeschrieben wird. Für (2K+1) Maskenelemente gibt es die symmetrische Differenz:

$$sD_m = \frac{B_{m+\Delta m,n} - B_{m-\Delta m,n}}{2\Delta m}$$

Die zugehörigen $(2K+1)$ Masken lauten für vertikale Kanten:

$${}^sD_k = \frac{1}{2} \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$$

...und für horizontale Kanten:

$${}^sD_l = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

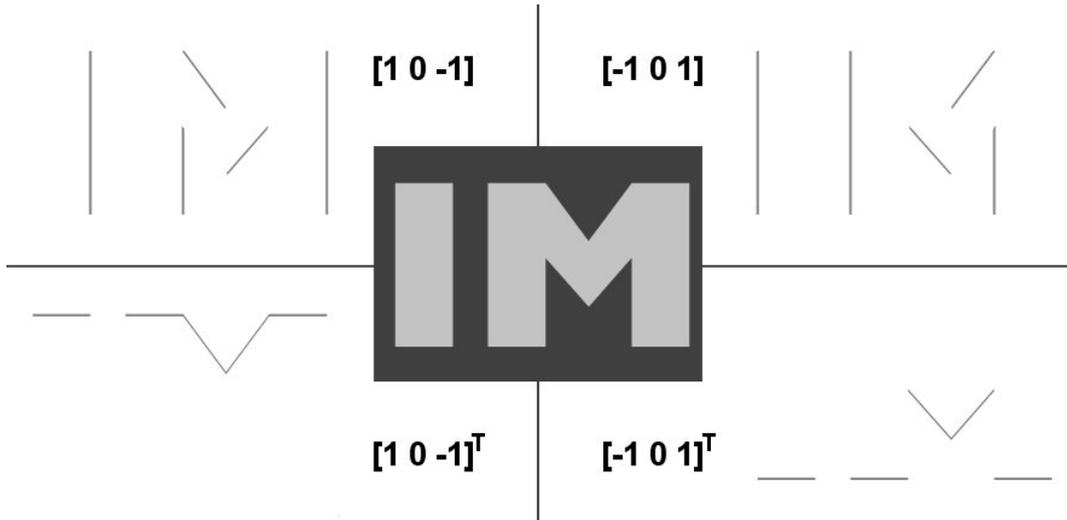


Abbildung 7.8: Das Bild in der Mitte wurde mit den angegebenen Filtern der symmetrischen Differenz gefiltert. Die gefundenen Kanten sind graue Linien. Hinweis: Die gefilterten Bilder sind hier als Negativ dargestellt.

Abbildung 7.8 zeigt anhand eines Beispiels, wie Kanten im Bild mit der symmetrischen Differenz gefunden werden.

Die Ableitung zweiter Ordnung ermöglicht die Detektion von Kanten mittels Nulldurchgänge. Sie ist gegeben durch:

$$\begin{aligned} \frac{\partial^2 B_{m,n}}{\partial m^2} &= \frac{\partial}{\partial m} \frac{\partial B_{m,n}}{\partial m} = \frac{B_{m+\Delta m,n} - B_{m,n} - [B_{m,n} - B_{m-\Delta m,n}]}{\Delta m \Delta m} \\ &= \frac{B_{m+\Delta m,n} - 2B_{m,n} + B_{m-\Delta m,n}}{\Delta m^2} \end{aligned}$$

Die Masken für die zweite Ableitung lauten:

$$D_k^2 = {}^{-}D_k \cdot {}^{-}D_k = \begin{bmatrix} 1 & -1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$$

Die Transferfunktionen der Ableitungsoperatoren lauten:

$$\begin{aligned} \frac{\partial}{\partial x} &\text{---} \bullet \quad \omega \quad \text{oder} \quad D \quad \text{---} \bullet \quad i \cdot n \tilde{\xi} \\ \frac{\partial^2}{\partial x^2} &\text{---} \bullet \quad -\omega^2 \quad \text{oder} \quad D^2 \quad \text{---} \bullet \quad -n^2 \tilde{\xi}^2 \end{aligned}$$

Für die Transferfunktionen kann auch eine Taylorreihenentwicklung durchgeführt werden:

$${}^s\hat{D}_{\tilde{s}} = \iota \sin(\pi {}^n\tilde{s}) \approx \iota \pi {}^n\tilde{s} - \underbrace{\frac{\iota(\pi {}^n\tilde{s})^3}{6}}_{\text{Fehler}} + O(\tilde{s}^5)$$

$$\hat{D}_{\tilde{s}}^2 = -2 + 2 \cos(\pi {}^n\tilde{s}) \approx -(\pi {}^n\tilde{s})^2 + \underbrace{\frac{1}{12}(\pi {}^n\tilde{s})^4}_{\text{Fehler}} + O(\tilde{s}^6)$$

Der Gradientenoperator Wenn man die lokale Kantenrichtung bestimmen möchte, macht man sich den Gradientenoperator zu Nutze. Er ist definiert durch:

$$\vec{\nabla} = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{pmatrix}$$

Betrag und Richtung des Gradientenoperators lauten:

$$|\vec{\nabla} \cdot g| = \sqrt{\left(\frac{\partial g}{\partial x_1}\right)^2 + \left(\frac{\partial g}{\partial x_2}\right)^2} \quad \varphi = \arctan\left(\frac{\frac{\partial g}{\partial x_2}}{\frac{\partial g}{\partial x_1}}\right)$$

Im Sinne der Bildfilterung lassen sich die partiellen Ableitungen wie folgt schreiben:

$$B_{m,n}^m = \frac{1}{2} \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} * B_{m,n} \quad \text{und} \quad B_{m,n}^n = \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * B_{m,n}$$

Der Betrag vom Gradientenoperator errechnet sich zu:

$$B_{m,n}^D = \sqrt{B^m \bullet B^m + B^n \bullet B^n} \quad \bullet : \text{punktweise Multiplikation}$$

Schließlich lautet die Richtung:

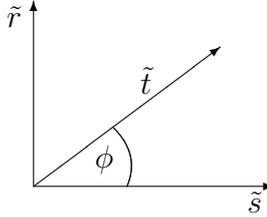
$$B_{m,n}^\varphi = \arctan\left(\frac{B^n}{B^m}\right)$$

Regularisierung durch Querglättung Es gibt das grundsätzliche Problem, dass Ableitungsoperatoren auf einem diskreten Gitter nur approximiert werden können. Je höher die Auflösung ist, desto genauer kann die Approximation sein. Dies kann bei einer vorgegebenen Diskretisierungsschrittweite durch Erweiterung der Maskengröße erreicht werden. Es treten zwei Arten von Fehlern auf.

1. Anisotropie: Die Berechnung des Betrags des Gradienten hängt von der Kantenrichtung ab.

2. Richtungsfehler: Die berechnete Kantenrichtung weicht von der Tatsächlichen ab.

Der Ableitungsoperator für zwei Dimensionen ist gegeben durch:

$$\begin{bmatrix} n_{\tilde{s}} \\ n_{\tilde{r}} \end{bmatrix} \xrightarrow{\text{Polarkoordinaten}} \begin{bmatrix} \tilde{t} \\ \phi \end{bmatrix}$$


Schreibt man die Ableitungen einzeln, so erhält man:

$${}^s\hat{D}_{\tilde{s}} = \iota \sin(\pi n_{\tilde{s}}) = \iota \sin(\pi \tilde{t} \cos \phi)$$

$${}^s\hat{D}_{\tilde{r}} = \iota \sin(\pi n_{\tilde{r}}) = \iota \sin(\pi \tilde{t} \sin \phi)$$

Der Betrag des Ableitungsoperators ergibt sich zu:

$$|\hat{D}| = \sqrt{\sin^2(\pi \tilde{t} \cos \phi) + \sin^2(\pi \tilde{t} \sin \phi)},$$

der Winkel zu:

$$\varphi = \arctan \frac{\sin(\pi \tilde{t} \sin \phi)}{\sin(\pi \tilde{t} \cos \phi)}$$

Über eine Taylorentwicklung mit Fehlerglieder für kleine \tilde{t} kann man Betrags- und Winkelfehler näherungsweise ausdrücken mit:

$$\text{Betragsfehler} \approx \frac{(\pi \tilde{t})^3}{12} \sin^2(2\phi) + O(\tilde{t}^5) \quad (7.3)$$

$$\text{Winkelfehler} \approx \frac{(\pi \tilde{t})^2}{24} \sin(4\phi) + O(\tilde{t}^4) \quad (7.4)$$

Für $\phi = 0^\circ, 45^\circ, 90^\circ, \dots$ ist der Richtungsfehler gleich Null.

Eine Verbesserung des Fehlers in Betrag sowie in der Richtung erhält man, wenn in die orthogonale Richtung der Ableitung eine Glättung durchgeführt wird. Ein bekannter Operator, der diese Regularisierung beinhaltet, ist der Sobel-Kantendetektor:

$$D_k = {}^sD_k G_l^2 = \frac{1}{2} \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} * \frac{1}{4} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

$$D_l = {}^sD_l G_k^2 = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} * \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

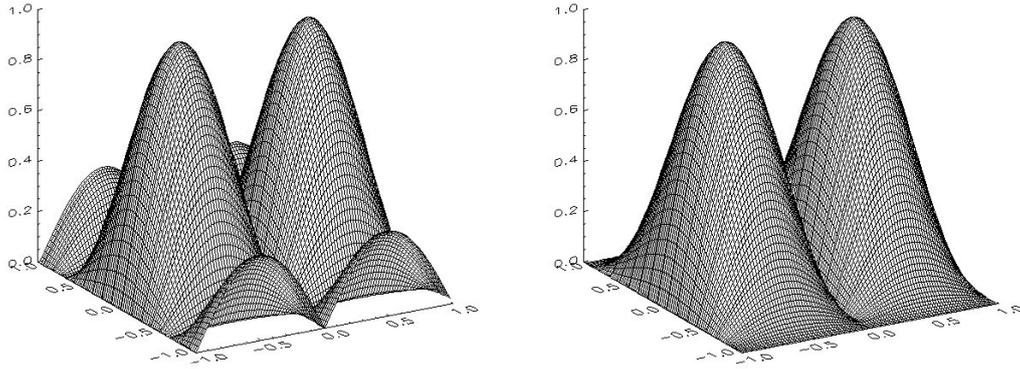


Abbildung 7.9: Dargestellt sind jeweils die Beträge der Transferfunktion in Abhängigkeit von \tilde{r} und \tilde{s} . **Links** TF vom Ableitungsoperator $\frac{1}{6}([1 \ 0 \ -1] * [1 \ 1 \ 1]^T)$. **Rechts** TF vom Sobeloperator $\frac{1}{8}([1 \ 0 \ -1] * [1 \ 2 \ 1]^T)$, der bessere Eigenschaften aufgrund der Querglättung mit einer Binomialmaske besitzt.

Nach einer Taylorreihenentwicklung von der Transferfunktion des Betrages und der Richtung erhält man geringere Fehler als in Gleichung (7.3) und (7.4):

$$\text{Betragsfehler} \approx -\frac{(\pi\tilde{t})^3}{24} \sin^2(2\phi) + O(\tilde{t}^5)$$

$$\text{Winkelfehler} \approx -\frac{(\pi\tilde{t})^2}{48} \sin(4\phi) + O(\tilde{t}^4)$$

Abbildung 7.9 zeigt mittels der Transferfunktionen den Unterschied zwischen dem Sobel-Filter und einem mit einer einfachen Glättung Regularisierten Operator. Weitere Informationen können in Jähne [1] und in einer Veröffentlichung von Scharf, Körköl und Jähne [10] nachgelesen werden.

Laplace Operator Die Kantendetektion mittels Nulldurchgänge, also zweiter Ableitung, ermöglicht der Laplace Operator. Er ist richtungsunabhängig und hebt somit Kanten beliebiger Orientierung hervor.

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$$

Seine Maske wird folgendermaßen gebildet:

$$L = D_k^2 + D_l^2 = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Die Transferfunktion des Laplace-Operators lautet:

$$\hat{L} = -4 + 2 \cos(\pi\tilde{s}) + 2 \cos(\pi\tilde{r})$$

Mit Hilfe einer Taylorreihe kann man als Näherung schreiben:

$$\hat{L} \approx -(\pi\tilde{t})^2 + \frac{(\pi\tilde{t})^4}{12} + \frac{(\pi\tilde{t})^4}{48} \cos(4\varphi) + O(\tilde{t}^6) \quad (7.5)$$

mit: $\tilde{s} = \tilde{t} \cos \varphi$
 $\tilde{r} = \tilde{t} \sin \varphi$
 $\tilde{t} = \sqrt{\tilde{r}^2 + \tilde{s}^2}$

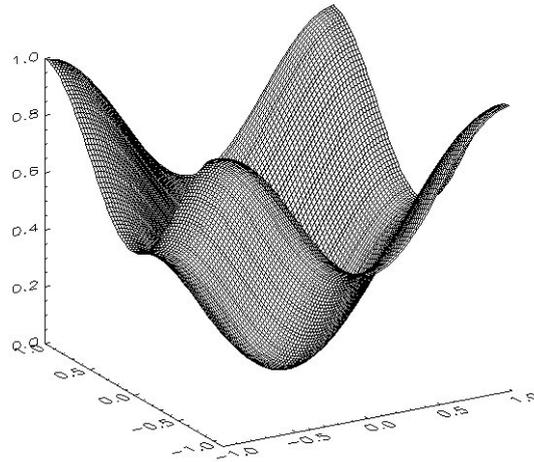


Abbildung 7.10: Dargestellt ist der Betrag der Transferfunktion des Laplaceoperators $\frac{1}{8}([1 \ -2 \ 1] + [1 \ -2 \ 1]^T)$ in Abhängigkeit von \tilde{r} und \tilde{s} .

Abbildung 7.10 zeigt den Betrag der Transferfunktion des Laplaceoperators. Wie anhand dieser Darstellung zu erkennen ist, weist der approximierte Laplace-Operator ein gewisses Isotropieverhalten auf. Dies ist mit dem dritten Term der Taylorreihe 7.5 quantifizierbar.

Der Laplace-Operator lässt sich auch mit Glättungsoperatoren konstruieren. Eindimensional lautet die zweite Ableitung:

$$\frac{\partial^2}{\partial x^2} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} = 4 \left[\frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} - \underbrace{\frac{1}{4} \begin{bmatrix} 0 & 4 & 0 \end{bmatrix}}_{I\text{-Identitätsoperator}} \right]$$

Die Anwendung von I auf das Bild ergibt das Bild wieder. Man kann den Laplaceoperator beschreiben mit:

$$L = 4(G^2 - I)$$

Zweidimensional ergibt sich für den Operator:

$$L = 4 \left[\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} - \frac{1}{16} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right] = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Diese Approximation weist besseres Isotropieverhalten auf und lautet näherungsweise für kleine Wellenzahlen (vgl. Gleichung 7.5):

$$\hat{L} \approx -(\pi\tilde{t})^2 + \frac{3}{32}(\pi\tilde{t})^4 - \frac{(\pi\tilde{t})^4}{96} \cos(4\varphi) + O(\tilde{t}^6)$$

Mit \bar{B} als Mittelwertbildung und $\Delta B \approx \bar{B} - B$ lässt sich der Laplace-Operator folgendermaßen approximieren:

$$L = \Delta B \approx \left[\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - \frac{1}{9} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

7.1.4 Filterdesign

Die Konsistenz eines Operators:

„Ein Operator ist dann konsistent diskretisiert, wenn er bei einem Grenzübergang der Schrittweite $h \rightarrow 0$ in den kontinuierlichen Operator übergeht.“

Der Differentialoperator D soll auf einem Intervall $[a, b]$ diskret dargestellt werden:

$$a = x_0 < x_1 < x_2 < \dots < x_{n-1} = b$$

$$x_i = a + i \cdot h \quad y_i = y(x_i) \quad \text{für } i = 0, 1, 2, \dots, n-1$$

mit: $h := \frac{b-a}{n-1}$

Die allgemeine Form des kontinuierlichen Differentialoperators D :

$$D = \sum_{j=0}^J \alpha_j \cdot \frac{\partial^j}{\partial x^j}$$

Die diskrete Darstellung:

$$D_h = \left[\beta_{-m} \quad \dots \quad \dots \quad \beta_m \right]$$

Taylorreihe an dem Punkt x_i :

$$y(x_i \pm lh) = y_i \pm lhy'_i \pm \frac{(lh)^2}{2!}y''_i \pm \frac{(lh)^3}{3!}y'''_i \pm \dots = \sum_{j=0}^J \frac{(lh)^j}{j!}y^{(j)}$$

Das Faltungsergebnis an dem Punkt x_i :

$$(D_h y)_i = \left(\sum_{j=0}^J C_j (\beta_{-m}, \dots, \dots, \beta_m, h) y_i^j \right) + O(h^r)$$

$r > 0$ gibt die Konsistenzordnung an

D_h ist eine konsistente Diskretisierung von D , wenn:

$$\begin{aligned} C_j (\beta_{-m}, \dots, \dots, \beta_m, h) &= \alpha_j \quad \text{für } j = 0, 1, \dots, J \\ &= \sum_{l=-m}^{+m} \frac{(lh)^j}{j!} \beta_l \end{aligned}$$

Beispiele:

1. Erste Ableitung mit einer 1x3-Maske. Dazu gehören drei Gleichungen:

$$\alpha_0 = 0 = \beta_{-1} + \beta_0 + \beta_1$$

$$\alpha_1 = 1 = (-\beta_{-1} + \beta_1)h$$

$$\alpha_2 = 0 = \beta_{-1} + \beta_1$$

Lösung:

$$\beta_{-1} = -\frac{1}{2h}$$

$$\beta_0 = 0$$

$$\beta_1 = \frac{1}{2h}$$

Die Maske lautet schließlich:

$$\left[\begin{array}{ccc} -1 & 0 & 1 \end{array} \right] \frac{1}{2h}$$

2. Erste Ableitung mit einer 1x5-Maske. Dazu gehören fünf Gleichungen:

$$\alpha_0 = 0 = \beta_{-2} + \beta_{-1} + \beta_0 + \beta_1 + \beta_2$$

$$\alpha_1 = 1 = (-2\beta_{-2} - \beta_{-1} + \beta_1 + 2\beta_2)h$$

$$\alpha_2 = 0 = 4\beta_{-2} + \beta_{-1} + \beta_1 + 4\beta_2$$

$$\alpha_3 = 0 = -8\beta_{-2} - \beta_{-1} + \beta_1 + 8\beta_2$$

$$\alpha_4 = 0 = 16\beta_{-2} + \beta_{-1} + \beta_1 + 16\beta_2$$

Lösung:

$$\begin{aligned}\beta_{-2} &= \frac{1}{12h} \\ \beta_{-1} &= \frac{-2}{3h} \\ \beta_0 &= 0 \\ \beta_1 &= \frac{2}{3h} \\ \beta_2 &= \frac{-1}{12h}\end{aligned}$$

Die Maske lautet schließlich:

$$\left[\begin{array}{ccccc} 1 & -8 & 0 & 8 & -1 \end{array} \right] \frac{1}{12h}$$

Für die Berechnung der Filtermaske der zweiten Ableitung muss man $\alpha_2 = 1$ wählen und alle anderen α_i zu null setzen. Am Beispiel der 1x3- bzw. 1x5-Masken ergeben sich die folgenden Filter:

$$\begin{aligned}1x3: & \frac{1}{h^2} \left[\begin{array}{ccc} 1 & -2 & 1 \end{array} \right] \\ 1x5: & \frac{1}{12h^2} \left[\begin{array}{ccccc} -1 & 16 & -30 & 16 & -1 \end{array} \right]\end{aligned}$$

In der Bildverarbeitung wird meistens $h = 1$ gewählt.

7.2 Nichtlineare Filter

Bei Nichtlinearen Filtern gilt das Distributivgesetz im Gegensatz zu den Linearen Filtern nicht mehr:

$$O(B_1 + B_2) \neq OB_1 + OB_2$$

7.2.1 Rangordnungsfilter

Die Rangordnungsfilter sind charakterisiert durch Vergleichen und Selektieren. Sie sortieren alle Grauwerte der Pixel, die innerhalb einer Filtermaske liegen, in absteigender Reihenfolge. Der Grauwert vom Zentralpixel wird je nach Kriterium durch einen Anderen ersetzt. Ein Beispiel ist der Medianfilter, der den mittleren Grauwert in der Filtermaske, den sog. Median, zurückschreibt (siehe Abbildung 7.11). Vorteil ist, dass „Ausreißer“ (z.B. sog. Salz- und Pfefferrauschen als weiße und schwarze Fehlerpixel) eliminiert werden (vgl. Abbildung 7.12), ohne dass scharfe Kanten geglättet werden (vgl. Abbildung 7.13 links).

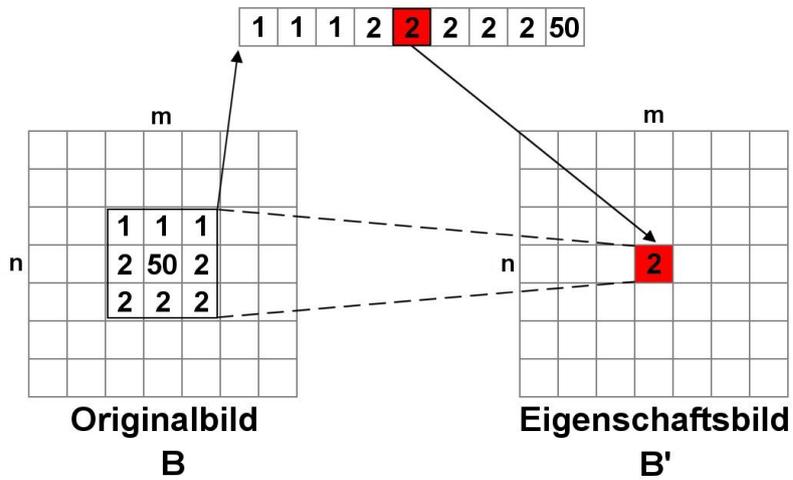


Abbildung 7.11: Wirkungsweise des Medianfilters. Die Filterelemente werden der Größe nach sortiert und das mittlere Element auf das Zentralpixel zurückgeschrieben. Somit werden abweichende lokale Extremwerte eliminiert.

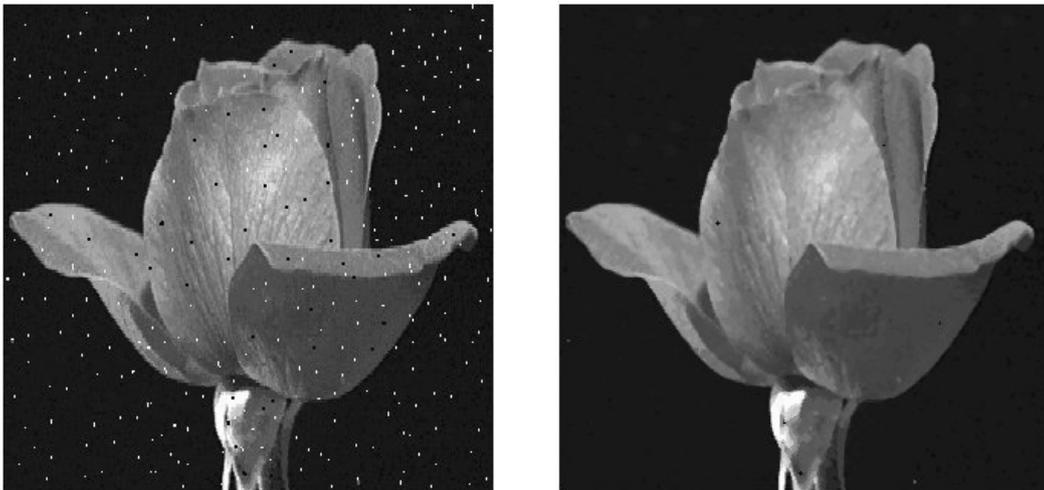
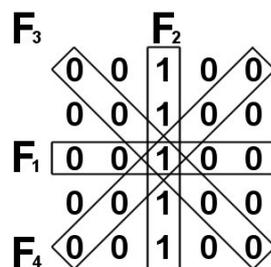


Abbildung 7.12: Eliminierung von Salz- und Pfefferrauschen durch den Medianfilter. Es bleiben Fehlerstellen übrig, die jedoch durch eine erneute Anwendung entfernt werden können. Dies hat wiederum ein größeres Weichzeichnen des Bildes zur Folge.

7.2.2 Max/Min-Medianfilter

Nachteilig beim Medianfilter ist, dass Linien mit einer Breite von einem Pixel eliminiert werden (vgl. Abbildung 7.13 rechts). Abhilfe schafft der *Max/Min-Medianfilter*, der mit eindimensionalen Fenstern arbeitet. Diese Fenster sind für eine $(2K+1) \times (2K+1)$ -Maske über die Diagonalen, Waagerechte und Senkrechte gelegt:

$$\begin{aligned}
 F_1 &= \{B_{(m-K,n)} & , \dots & , B_{(m,n)} & , \dots & , B_{(m+K,n)}\} \\
 F_2 &= \{B_{(m,n-K)} & , \dots & , B_{(m,n)} & , \dots & , B_{(m,n+K)}\} \\
 F_3 &= \{B_{(m-K,n-K)} & , \dots & , B_{(m,n)} & , \dots & , B_{(m+K,n+K)}\} \\
 F_4 &= \{B_{(m-K,n+K)} & , \dots & , B_{(m,n)} & , \dots & , B_{(m+K,n-K)}\}
 \end{aligned}$$


Die Masken F_1 bis F_4 werden nun mit folgenden Funktionen verknüpft:

$$\begin{aligned}
 m_0 &= \text{Median}\{F_1, F_2, F_3, F_4\} \\
 m_1 &= \text{Median}\{F_1\} \\
 m_2 &= \text{Median}\{F_2\} \\
 m_3 &= \text{Median}\{F_3\} \\
 m_4 &= \text{Median}\{F_4\} \\
 m_a &= \text{Max}\{m_1, m_2, m_3, m_4\} \\
 m_i &= \text{Min}\{m_1, m_2, m_3, m_4\}
 \end{aligned}$$

Das Ergebnis des Max/Min-Median-Operators ist letztendlich durch die folgende Entscheidungsregel gegeben:

$$B'_{m,n} = \begin{cases} m_a & \text{if } |m_a - m_0| \geq |m_0 - m_i| \\ m_i & \text{if } |m_a - m_0| < |m_0 - m_i| \end{cases} \quad (7.6)$$

Für die obige senkrechte 1-Pixel-Linie erhält man beispielsweise für die m_i :

$$\left. \begin{aligned} m_0 &= 0 \\ m_1 &= 0 \\ m_2 &= 1 \\ m_3 &= 0 \\ m_4 &= 0 \\ m_a &= 1 \\ m_i &= 0 \end{aligned} \right\} \underline{\underline{B'_{m,n} = 1 = m_a}} \quad (7.7)$$

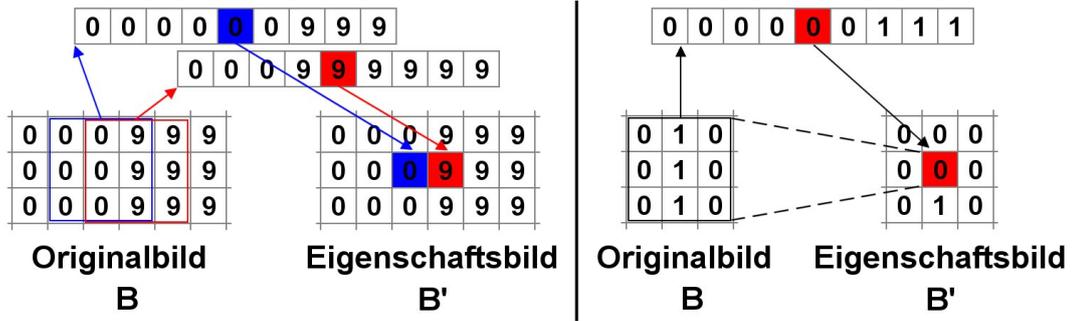


Abbildung 7.13: Links: Vorteil des Medianfilters: Scharfe Kanten werden beim Medianfilter nicht geglättet. Rechts: Nachteil des Medianfilters: 1-Pixel-Kanten verschwinden gänzlich.

Ein weiteres Beispiel:

$$\left. \begin{array}{l}
 3 \ 3 \ 1 \ 3 \ 3 \\
 3 \ 3 \ 1 \ 3 \ 3 \\
 3 \ 3 \ 1 \ 3 \ 3 \\
 3 \ 3 \ 1 \ 3 \ 3 \\
 3 \ 3 \ 1 \ 3 \ 3 \\
 m_0 = 3 \\
 m_1 = 3 \\
 m_2 = 1 \\
 m_3 = 3 \\
 m_4 = 3 \\
 m_a = 3 \\
 m_i = 1
 \end{array} \right\} \begin{array}{l}
 |m_a - m_0| = 0 \\
 |m_0 - m_i| = 2
 \end{array} \Rightarrow \underline{\underline{B'_{m,n} = 1 = m_i}} \quad (7.8)$$

Kapitel 8

Morphologische Operatoren

Die Morphologie ist die *Gestaltlehre* bzw. die *Formenlehre*. Die morphologischen Operatoren bilden eine Alternative zu den klassischen Faltungsoperatoren aus dem vorherigen Kapitel. Sie verändern mit ihrer Anwendung die Gestalt eines Objektes im Bild. Es werden in diesem Kapitel die wichtigsten Operatoren *Erosion* und *Dilatation* sowie die zusammengesetzten Operatoren *Öffnung* und *Schließung* vorgestellt.

Das sogenannte *Strukturelement* bestimmt den Einflussbereich der morphologischen Maske, also die lokale Umgebung, in der die Operationen durchgeführt werden. Abbildung 8.1 zeigt ein paar Beispiele für Strukturelemente.

8.1 Dilatation und Erosion

Neben den Strukturelementen beeinflussen auch die Rangordnungsoperationen die Wirkung eines morphologischen Operators. Für Binärbilder können diese Operationen durch einfache Mengenoperationen ersetzt werden. Man unterscheidet zwischen der Verkleinerung (Erosion¹) und der Vergrößerung (Dilatation²) von Strukturen.

¹lat., erodieren=aus-, wegnehmen

²Dilatation=Ausdehnung

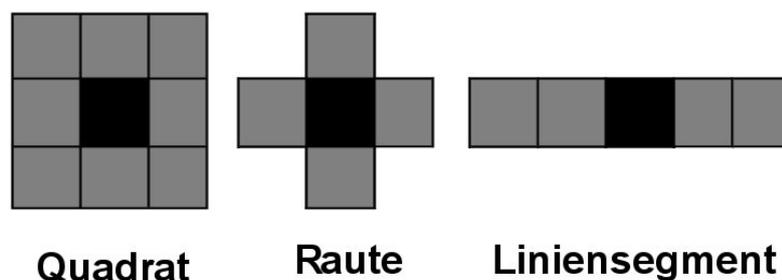


Abbildung 8.1: (Abb. nach Soille) Strukturelemente bei morphologischen Operatoren zur Festlegung der lokalen Umgebung, in der die Operation durchgeführt wird.

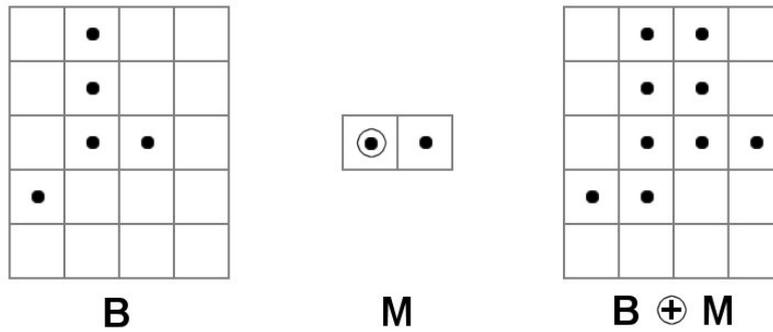


Abbildung 8.2: (Abb. nach Haralick/Shapiro) Dilatation am Beispiel eines Binärbildes mit einer 2x1-Maske als Strukturelement. Die Dilatation entspricht einer logischen ODER-Operation.

8.1.1 Dilatation

Für Grauwertbilder ist die Dilatation des Bildes B mit dem Strukturelement M durch

$$B'_{m,n} = B \oplus M = \{P : M_P \cap B \neq \emptyset\} \quad (8.1)$$

definiert. Dabei ist B' die Menge aller Pixel P , für die die Schnittmenge von B und M_P nicht die leere Menge ist. Man bezeichnet diese Operation auch als *Minkowski-Addition*.

Für Binärbilder ergibt sich ein Sonderfall. Die Dilatation kann man sich dann wie eine logische ODER-Operation vorstellen. Betrachtet werden alle Elemente innerhalb des Strukturelements, die beispielsweise nur die Werte eins und null annehmen können. Wenn mindestens ein Element eine Eins enthält, so wird auch eine Eins zurückgeschrieben, ansonsten eine Null. Im Vergleich zu den schon bekannten Rangordnungsoperatoren, hat der Maximumoperator den gleichen Effekt und stellt somit eine Alternative dar.

Abbildung 8.2 verdeutlicht die Dilatation anhand eines Beispiels.

8.1.2 Erosion

Die Erosion ist das Pendant zur Dilatation und lässt sich für ein Grauwertbild B mit dem Strukturelement M mit

$$B'_{m,n} = B \ominus M = \{P : M_P \subseteq B\} \quad (8.2)$$

beschreiben. Dabei ist B' die Menge aller Pixel P , für die M_P vollständig in B enthalten ist. Man bezeichnet diese Operation auch als *Minkowski-Subtraktion*.

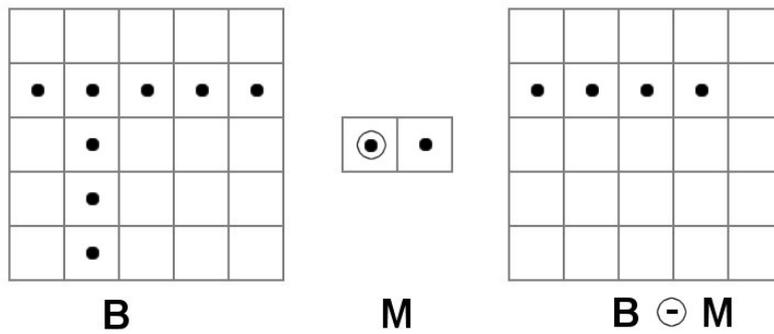


Abbildung 8.3: (Abb. nach Haralick/Shapiro) Erosion am Beispiel eines Binärbildes mit einer 2x1-Maske als Strukturelement. Die Dilatation entspricht einer logischen UND-Operation.

Für Binärbilder ergibt sich wieder ein Sonderfall. Diesmal ist die Operation mit einer UND-Operation zu erklären. Die Elemente innerhalb des Strukturelementes werden wieder betrachtet und nehmen die Werte null und eins an. Wenn mindestens ein Element den Wert null darstellt, so wird auch eine Null zurückgeschrieben, andernfalls eine Eins. Das Gegenstück zu den Rangordnungsoperatoren ist hier der Minimumoperator. Abbildung 8.3 verdeutlicht die Erosion anhand eines Beispiels.

8.2 Öffnung und Schließung

Durch Hintereinanderanwenden der elementaren Operationen Dilatation und Erosion ergeben sich komplexe morphologische Operationen, wie z.B. die Öffnung (engl. opening) und Schließung (engl. closing).

8.2.1 Öffnung

Durch die Erosion verschwinden nicht nur die Strukturen, die das Strukturelement nicht vollständig füllen, es erfolgt auch ein Schrumpfen des Objektes. Um diese Verkleinerung wieder rückgängig zu machen, bedient man sich der Öffnung. Sie entsteht durch eine Erosion, gefolgt von einer Dilatation:

$$B'_{m,n} = (B \ominus M) \oplus M \quad (8.3)$$

Kleine Strukturen verschwinden bei der Öffnung vollständig, deswegen bezieht sich die Wiederherstellung der Größe nur auf einen Teil der Strukturen. Ein Beispiel ist in Abbildung 8.4 zu betrachten.

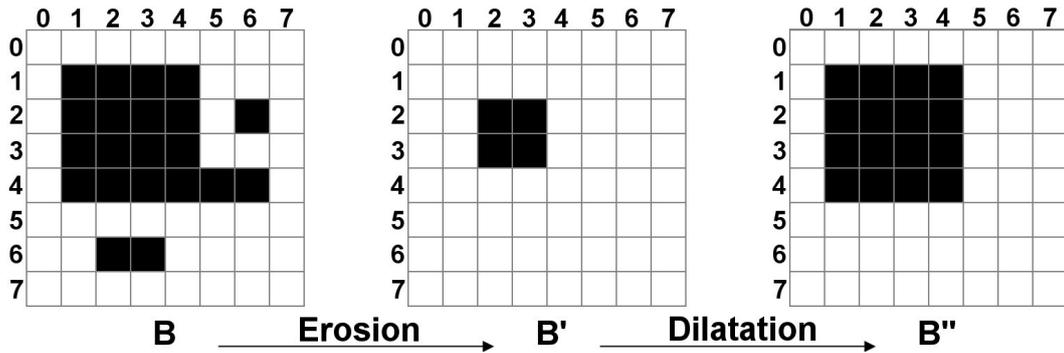


Abbildung 8.4: (Abb. nach Köhn [9]) Öffnung am Beispiel eines Binärbildes mit einer 3x3-Maske des Strukturelements.

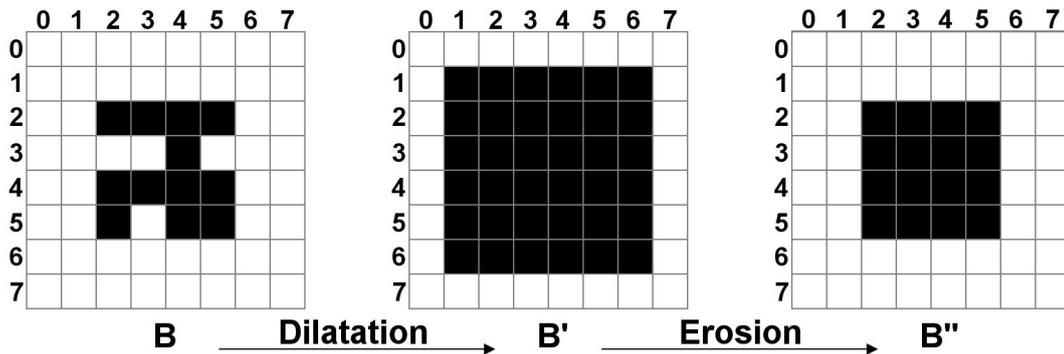


Abbildung 8.5: (Abb. nach Köhn [9]) Schließung am Beispiel eines Binärbildes mit einer 3x3-Maske des Strukturelements.

8.2.2 Schließung

Die Schließung ist das Gegenstück zur Öffnung. Es wird erst eine Dilatation und dann eine Erosion durchgeführt:

$$B'_{m,n} = (B \oplus M) \ominus M \quad (8.4)$$

Dabei vergrößert die Dilatation die Strukturen und füllt gleichzeitig kleine Löcher und Risse aus. Mit der anschließenden Erosion wird der Vergrößerung entgegengewirkt. Ein Beispiel ist in Abbildung 8.5 zu betrachten.

8.3 LOC-Operator

Der LOC-Operator steht für *Lokale Orientierungskodierung*. Er wird für die Erkennung von geometrischen Strukturen verwendet. Bei der Anwendung des Operators werden alle Pixel des Fensters mit dem Zentralpixel verglichen. Je nachdem, ob der Grauwert größer, gleich oder kleiner als der des Zentralpixels ist, wird ihnen der Wert null oder eins zugeordnet. Im Anschluss werden alle Einsen addiert und in das Zentralpixel

zurückgeschrieben:

$$f_{(k,l)} = \begin{cases} 1 & B_{(m,n)} > B_{(m+k,n+l)} \\ 0 & \text{andernfalls} \end{cases} \quad \forall (k,l) \in \text{Maske} \quad (8.5)$$

$$B'_{(m,n)} = \sum f_{(k,l)}$$

Die Vorteile des LOC-Operators sind geringer Rechenaufwand und die Unabhängigkeit der Resultate von den absoluten Grauwerten und dem Kontrast.

Für die Objekte *Punkt*, *Linie*, *Kante* und *Ecke* erhält man nach obiger Kodierungsregel die folgenden Bilder:

	0 0 0 0 0		0 0 0 0 0
	0 0 0 0 0		0 0 0 0 0
Punkt:	0 0 20 0 0	\xrightarrow{M}	0 0 8 0 0
	0 0 0 0 0		0 0 0 0 0
	0 0 0 0 0		0 0 0 0 0
	0 0 10 0 0		0 0 6 0 0
	0 0 10 0 0		0 0 6 0 0
Linie:	0 0 10 0 0	\xrightarrow{M}	0 0 6 0 0
	0 0 10 0 0		0 0 6 0 0
	0 0 10 0 0		0 0 6 0 0
	0 0 10 10 10		0 0 3 0 0
	0 0 10 10 10		0 0 3 0 0
Kante:	0 0 10 10 10	\xrightarrow{M}	0 0 3 0 0
	0 0 10 10 10		0 0 3 0 0
	0 0 10 10 10		0 0 3 0 0
	0 0 0 0 0		0 0 0 0 0
	0 0 0 0 0		0 0 0 0 0
Ecke:	0 0 10 10 10	\xrightarrow{M}	0 0 5 3 3
	0 0 10 10 10		0 0 3 0 0
	0 0 10 10 10		0 0 3 0 0

Diese Resultate sind jedoch nicht eindeutig, da sie keine Richtungskodierung des Objektes beinhalten. Eine 4-Bit-Kodierung bei einer 3x3-Maske könnte folgendermaßen aussehen:

$$\begin{matrix} 0 & 1 & 0 \\ 2 & R & 4 \\ 0 & 8 & 0 \end{matrix}$$

Damit verändern sich obigen Kodierungen für Punkt, Linie, Kante und Ecke zu:

	0	0	0	0	0		0	0	0	0	0
Punkt:	0	0	20	0	0	\xrightarrow{M}	0	0	15	0	0
	0	0	0	0	0		0	0	0	0	0
	0	0	0	0	0		0	0	0	0	0
	0	0	10	0	0		0	0	6	0	0
	0	0	10	0	0		0	0	6	0	0
Linie:	0	0	10	0	0	\xrightarrow{M}	0	0	6	0	0
	0	0	10	0	0		0	0	6	0	0
	0	0	10	0	0		0	0	6	0	0
	0	0	10	10	10		0	0	2	0	0
	0	0	10	10	10		0	0	2	0	0
Kante:	0	0	10	10	10	\xrightarrow{M}	0	0	2	0	0
	0	0	10	10	10		0	0	2	0	0
	0	0	10	10	10		0	0	2	0	0
	0	0	0	0	0		0	0	0	0	0
	0	0	0	0	0		0	0	0	0	0
Ecke:	0	0	10	10	10	\xrightarrow{M}	0	0	3	1	1
	0	0	10	10	10		0	0	2	0	0
	0	0	10	10	10		0	0	2	0	0

Anwendungsgebiete der LOC-Operatoren:

- Modellbasierte Segmentierung (Modellanpassung)
- Stereovision (Korrespondenzanalyse)
- Bewegungserkennung (Bestimmung von Verschiebungsvektoren)

Kapitel 9

Segmentierung

In diesem Kapitel sollen die drei wichtigsten methodischen Ansätze zur Bildsegmentierung vorgestellt werden. Dazu gehören die *pixelbasierten*, *kantenbasierten* und *regionenbasierten* Verfahren. Zu Letzterem wird das *Regionenwachstumsverfahren* und das *Split & Merge*-Verfahren vorgestellt.

Das Ziel der Segmentierung ist die Unterteilung des Bildes in bedeutungsvolle Teilbereiche mit homogenen Eigenschaften, wie z.B. Grauwert, Farbe oder Textur. Somit stellt die Segmentierung den ersten Schritt in Richtung Mustererkennung dar. Nach der Segmentierung kann man die Form der Segmente (Objekte) analysieren und Formmerkmale für die Erkennung von Mustern verwenden.

9.1 Grundlagen

Definition (Segmentierung) Unter der Segmentierung eines diskreten Bildsignals $B_{m,n}$ versteht man die Unterteilung von B in disjunkte ($B_i \cap B_j = 0$), nichtleere Teilmengen B_1, B_2, \dots, B_p , so dass mit einem zu definierenden Einheitlichkeitskriterium E gilt:

- $\bigcup_{i=1}^P B_i = B$
- B_i ist zusammenhängend \forall_i mit $i = 1, \dots, P$.
- $\forall B_i$ ist das Einheitlichkeitskriterium $E(B_i)$ erfüllt.
- Für jede Vereinigungsmenge zweier benachbarter B_i, B_j ist $E(B_i \cup B_j)$ nicht erfüllt.

Definition (Segmente) Segmente sind Klassen, bei denen die Bildpunkte, die zu einer Klasse gehören, im Sinne einer vorgegebenen Nachbarschaftsrelation zusammenhängen.

9.2 Methodische Ansätze

Bei der Segmentierung gibt es grundlegend verschiedene Konzepte:

1. Pixelbasierte Verfahren: verwenden nur die Grauwerte der einzelnen Pixel, „beachten die lokale Nachbarschaft überhaupt nicht“.
2. Regionenbasierte Verfahren: untersuchen die Grauwerte in zusammenhängenden Regionen, „Analysieren homogener Bereiche“.
3. Kantenbasierte Verfahren: erkennen Kanten und versuchen ihnen zu folgen, „achten nur auf Diskontinuitäten“.

Im Folgenden wollen wir uns die einzelnen Ansätze näher anschauen.

9.2.1 Pixelbasierte Verfahren

Das pixelbasierte Verfahren ist der einfachste Ansatz zur Segmentierung und sollte deswegen immer als Erstes gewählt werden, bevor man sich den komplizierteren Methoden zuwendet. Es werden die Grauwerte der einzelnen Pixel betrachtet und in dem Histogramm dargestellt. Bei einer sog. *bimodalen* Verteilung mit zwei Maxima muss ein geeigneter Schwellwert festgelegt werden (*Schwellwertverfahren*), der die Qualität der Segmentierung maßgeblich bestimmt. Eine Region ist durch einen unteren und einen oberen Schwellwert definiert:

$$R = \{P \mid S_u \leq g(P) \leq S_o\} \quad (9.1)$$

Abbildung 9.1 zeigt den Einfluss des Schwellwertes auf das Segmentierungsergebnis. Das pixelbasierte Verfahren eignet sich jedoch nur für homogene Hintergründe und Objekte mit konstantem Grauwert. Sobald diese zwei Bedingungen nicht mehr erfüllt sind kann es passieren, dass Objekte als Hintergrund bzw. Hintergründe als Objekte detektiert werden.

9.2.2 Kantenbasierte Verfahren

Die kantenbasierte Verfahren gehen davon aus, dass Objekte immer durch Kanten begrenzt sind. Findet und verfolgt man diese Kanten, so hat man das Objekt erkannt. Wie man Kanten erkennt, ist uns schon aus dem Kapitel der Nachbarschaftsoperatoren bekannt. Entweder man bedient sich der ersten Ableitung und sucht die Maxima oder

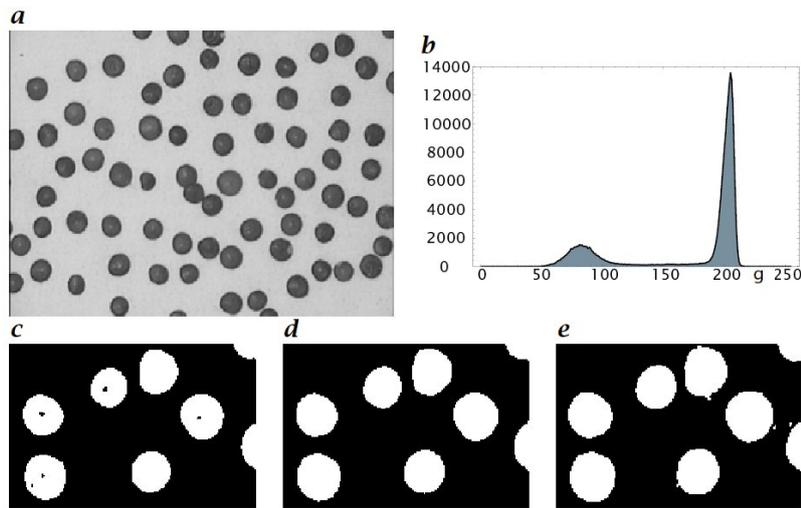


Abbildung 9.1: (Abb. aus Jähne [1]) Segmentierung mit einem globalen Schwellwert: **a** Originalbild; **b** Histogramm; **c** bis **e** oberer rechter Sektor von **a**, segmentiert mit globalen Schwellwerten 110, 147 bzw. 185.

man findet die Nulldurchgänge der zweiten Ableitung. Nachdem eine Kante, zum Beispiel mit dem Grauwertgradienten, gefunden wurde, muss sie nur noch verfolgt werden, um das Objekt zu begrenzen. Vor- und Nachteile des kantenbasierten Verfahren sind:

- V: Lineare Variierung der Hintergrundhelligkeit beeinflusst nicht die Lage der Kante.
- V: Unabhängigkeit von der Intensität der Kante.
- N: Je verschmierter die Kanten, desto fehlerbehafteter ist die detektierte Kantenposition.

Ein wichtiger Begriff, der in dieser Thematik auftauchen kann, ist die *Nonmaxima-Unterdrückung*. Sie ist lediglich eine Kantenverdünnung zu einem Pixel Kantenbreite, da die ermittelten Kanten breiter sein können. Zweck ist das Bilden von Kantensegmenten.

Während der Kantenverfolgung stellt sich die Frage, wie eine solche Kante gespeichert werden kann. Dazu hat sich der *Kettencode* durchgesetzt. Von einem Pixel aus betrachtet, kann sich die Kante nur in acht verschiedene Richtungen fortsetzen. Folglich braucht man nur diese Richtung zu speichern, was mit drei Bit möglich ist.

Weitere Informationen zu Skelettierung und Kettencode können im Vorlesungsskript von Dr.Steinmüller [6] nachgelesen werden.

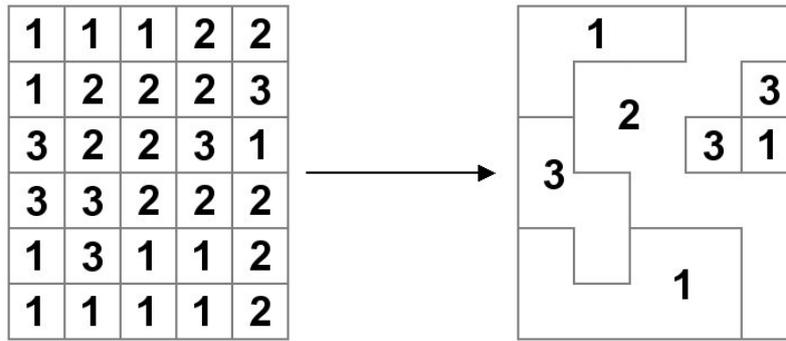


Abbildung 9.2: (Abb. nach Steinmüller [6]) Einfaches Beispiel für die regionenbasierte Segmentierung. **Links** Eingangsbild. **Rechts** Detektierte Regionen aufgrund ihrer Grauwerte.

9.2.3 Regionenbasierte Verfahren

Die regionenbasierten Verfahren ordnen einen Bildpunkt aufgrund seines Grauwertes oder einer Textureigenschaft einem Objekt zu. Das kann zur Folge haben, dass isolierte Punkte oder kleine Punktgruppen entstehen, die das eigentliche Objekt nicht berühren. Abbildung 9.2 zeigt ein einfaches Beispiel. Einer Region muss nicht zwangsläufig ein bestimmter Grauwert zugewiesen sein, es kann auch ein Grauwertbereich sein. Die Standardverfahren für die Verschmelzung der einzelnen Regionen sind das sog. *Regionenwachstumsverfahren* und *Split & Merge*.

Regionenwachstumsverfahren Das Regionenwachstumsverfahren (oder Flächenwachstumsverfahren, engl.: region crowing) unterliegt laut Rehrmann [8] folgendem Algorithmus:

1. Unterteile das Bild in initiale Zellen (1x1, 2x2 oder 4x4 Pixel).
2. Berechne ein statistisches Maß der Intensitäten über alle Zellen. Beginne mit der ersten Zelle in der linken, oberen Ecke.
3. Vergeiche die Statistiken einer Zelle mit jeder der benachbarten Zellen, um zu entscheiden, ob sie ähnlich sind. Falls der Vergleich erfolgreich ist, verschmelze die Zellen zu einem Segment und aktualisiere das statistische Maß für das Segment.
4. Lasse das Segment durch den Vergleich mit all seinen Nachbarn weiter wachsen, bis keine Nachbarn mehr hinzugenommen werden können. Markiere das Segment als fertige Region.
5. Nimm die nächste unbearbeitete Zelle und wiederhole die Schritte 3-4, bis alle Zellen bearbeitet sind.

Split & Merge Im Gegensatz zum Regionenwachstumsverfahren, dass von kleinen initialen Zellen ausgeht, fängt der Split & Merge-Algorithmus beim Bild an. Zuerst wird im Split-Teil das Bild rekursiv zergliedert, bis das Homogenitätskriterium erfüllt ist. Ausgehend von diesen homogenen Teilgebieten werden im anschließendem Merge-Teil benachbarte Teilgebiete, die das Homogenitätskriterium erfüllen, zusammengefasst. Der Algorithmus lautet nach Rehrmann [8]:

1. Teile das Bild in 2^n gleichgroße, quadratische Regionen auf.
2. (Split) Falls für eine Region R in dieser Struktur $H(R) = \text{FALSE}$ gilt, teile die Region in vier Teilregionen auf.
3. (Merge) Falls für vier Regionen eines Quadranten R_{k1}, \dots, R_{k4} $H(R_{k1} \cup R_{k2} \cup R_{k3} \cup R_{k4}) = \text{TRUE}$ gilt, verschmelze sie zu einer Region.
4. (Grouping) Falls es irgendwelche benachbarte Regionen R_i und R_j gibt (möglicherweise unterschiedlicher Größe), für die $H(R_i \cup R_j) = \text{TRUE}$ gilt, verschmelze diese Regionen.

Literaturverzeichnis

- [1] Bernd Jähne: *Digitale Bildverarbeitung*, 5. Auflage, Springer Verlag Berlin Heidelberg, 2002
- [2] Peter Haberäcker: *Digitale Bildverarbeitung*, Grundlagen und Anwendungen, 5.Auflage, Hanser Verlag München Wien 1991
- [3] F.M. Wahl: *Digitale Bildsignalverarbeitung*, Springer Verlag Berlin Heidelberg, 1989
- [4] Ulrich Schmidt: *Digitale Videotechnik*, Franzis Verlag, 1996
- [5] Pierre Soille: *Morphologische Bildverarbeitung*, Springer Verlag Berlin Heidelberg, 1996
- [6] Dr. Johannes Steinmüller: *Bildverarbeitung*, TU-Chemnitz, Sommersemester 04
- [7] Hanspeter A. Mallot: *Kognitive Neurobiologie*, Universität Tübingen, Wintersemester 04/05
- [8] Volker Rehrmann: *Vorlesung Digitale Bildverarbeitung*, Universität Koblenz-Landau, Wintersemester 99/00
- [9] Carsten Köhn: *Skript EPA Draft*, Fachhochschule Bochum, 10.Juli 2003
- [10] Hanno Scharr, Stefan Körkel und Bernd Jähne *Numerische Isotropieoptimierung von FIR-Filtern mittels Querglättung* Universität Heidelberg
- [11] Elektronik-Kompendium:
<http://www.elektronik-kompendium.de/sites/mes/0307081.htm>
- [12] Funktionsprinzipien von CCD-Sensoren:
<http://www.ccd-sensor.de/index.html>
- [13] Fernsehtechnik:
http://info.electronicwerkstatt.de/bereiche/fernsehtechnik/tvsignale/tv_5.html

- [14] NTSC Informationen:
<http://www.ntsc-tv.com/ntsc-main-04.htm>
- [15] PAL Informationen:
http://de.wikipedia.org/wiki/Phase_Alternating_Line
- [16] Äquidensiten Informationen:
<http://www.abc26.de/buch.htm>
- [17] Informationen zu Basisbildern der DCT:
<http://www.flaskmpeg.info/board/thread.php?threadid=3102&sid=>