

Mathematische Methoden in der Bildverarbeitung

Priv.-Doz. Dr. Sven Ehrich

Vorlesung für Studierende der Mathematik und der Informatik
TU München
Wintersemester 2003/2004

Vorbemerkung

Diese Unterlagen sind kein Ersatz für ein Lehrbuch.

Das Manuskript soll die Nachbereitung der Vorlesung “Mathematische Methoden in der Bildverarbeitung” erleichtern, die im Wintersemester 2003/2004 für Studentinnen und Studenten der Mathematik und der Informatik an der Technischen Universität München gehalten wurde. Für eine sinnvolle Nachbereitung empfiehlt es sich, zusätzlich mindestens ein Standard-Lehrbuch zu verwenden. Entsprechende Literaturhinweise sind am Ende des Manuskripts zu finden.

Soweit dies während der Vorlesung noch nicht geschehen ist, wird außerdem die Bearbeitung der begleitenden praktischen Übungen empfohlen. Die Aufgaben (und Lösungsvorschläge) sind auf der homepage der Vorlesung

www-m12.ma.tum.de/lehre/bv_2003

zu finden.

Teile des Manuskripts wurden bereits in der Vorlesung ”Grundlagen der digitalen Bildverarbeitung” im Sommersemester 1999 präsentiert. Ich danke den Hörerinnen und Hörern dieser Veranstaltung für ihre aktive Teilnahme an der Vorlesung und für die konstruktiven Rückmeldungen. Mein besonderer Dank gilt meiner damaligen Hilfskraft, Herrn Dr. Peter Heinlein, für seine intensive Unterstützung und Zusammenarbeit. Er hat grosse Teile der begleitenden praktischen Übungen konzipiert und mit vielen Vorschlägen zur Überarbeitung des Vorlesungsmanuscript beigetragen. Eventuell verbleibende Fehler bleiben natürlich in meiner Verantwortung.

Sven Ehrich

Oktober 2003

Inhaltsverzeichnis

1	Einleitung	4
1.1	Ausrichtung der Vorlesung	4
1.2	Anwendungen der Bildverarbeitung und bildgebende Verfahren . .	5
1.2.1	Fernerkundung der Erde	5
1.2.2	Bildgebende Verfahren in der Medizin	6
1.2.3	Bildverarbeitung in der Automatisierung	7
1.3	Begriffsklärungen	7
2	Das kontinuierliche und das digitale Bildmodell	8
2.1	Das kontinuierliche Idealbild	8
2.2	Das digitale Grauwertmodell	9
2.3	Diskrete Farbbildmodelle	10
2.3.1	Farbwahrnehmung	10
2.3.2	Farbräume	10
2.4	Elementare Kenngrößen von Bildern	12
3	Codierung von Bildern	15
3.1	Vorbemerkungen zum Informationsgehalt von Bildern	15
3.2	Huffman-Codierung	18
3.3	Binärbilder	18
3.4	Differenzenbildung	19
3.5	Laufängencodierung	20
4	Modifikation der Grauwertverteilung	21
4.1	Lineare Skalierung	22
4.2	Stückweise konstante Skalierung	24
4.3	Grauwertäquilisation	25

5	Geometrische Transformationen und Entzerrung	26
5.1	Vergrößerung	26
5.2	Affine Abbildungen	28
5.3	Interpolation durch Polynome	29
5.4	Methode der kleinsten Quadrate	31
6	Lokale Operatoren und Filter	34
6.1	Gleitender Mittelwert, lineare Filter	34
6.2	Median-Filter	37
6.3	Differenzenoperatoren	38
6.4	Elemente der Mathematischen Morphologie	41
7	Fouriertransformationen	43
7.1	Rückblick auf die eindimensionale kontinuierliche Fouriertransformation und auf Fourierreihen	43
7.2	Zweidimensionale kontinuierliche Fouriertransformation	46
7.3	Eindimensionale diskrete Fouriertransformation (DFT)	48
7.4	Schnelle Fouriertransformation (FFT)	49
7.5	Zweidimensionale diskrete Fouriertransformation	51
7.6	Digitale Filterung mittels Faltung und DFT	53
7.7	Eine Basisdarstellung für die zweidimensionale DFT	54
7.8	Diskrete Cosinustransformation (DCT)	56
7.9	Anwendung der DCT auf Bilddatenkompression (JPEG)	59
8	Wavelet-Transformationen	61
8.1	Elemente der Zeit-Frequenzanalyse	61
8.2	Die kontinuierliche Wavelet-Transformation	63
8.3	Die orthogonale Multiresolutionsanalyse (Multiskalenanalyse)	63
8.4	Mehrdimensionale Multiskalenanalyse	65
8.5	Die diskrete Wavelettransformation	66
9	Klassifikation	70
10	Segmentierung	71

Kapitel 1

Einleitung

1.1 Ausrichtung der Vorlesung

Das Ziel dieser Vorlesung ist es, Methoden zur Verarbeitung von digitalen Bildern zu vermitteln. Dabei werden wir uns, den Umfang der Vorlesung vor Augen, auf die grundlegendsten Methoden zur Informationsverarbeitung und -gewinnung beschränken. Fast immer werden wir dabei davon ausgehen, daß Bilder in digitalisierter Form vorliegen, d.h. codiert als Matrix von Grauwerten bei Schwarz-Weiß-Bildern, oder als Farbbild — allgemeiner als Hyperspektralbild — codiert in drei oder mehr Matrizen.

Die Verarbeitung und Analyse von Bildern ist vor allem eine Frage nach effizienten Algorithmen. Diese Vorlesung ist daher in erster Linie algorithmisch orientiert. Sie haben am Ende der Vorlesung einen Basissatz von Methoden kennengelernt, der einen soliden Ausgangspunkt für eine weitere praktische oder methodische Beschäftigung mit der Bildverarbeitung bildet.

Begleitend zu dieser Vorlesung werden praktische Übungsaufgaben ausgegeben. Ich empfehle, daß Sie diese bearbeiten. Die rechnerische Komplexität vieler der Methoden, die Sie in den folgenden Wochen kennenlernen, werden Sie erst durch die eigene Bearbeitung dieser Aufgaben einschätzen können.

1.2 Anwendungen der Bildverarbeitung und bildgebende Verfahren

Bildverarbeitungsmethoden haben einen sehr breiten Anwendungsbereich. Dazu gehören insbesondere:

- Fernerkundung der Erde
- medizinische Diagnostik
- Automatisierung
- Werkstoffprüfung und Qualitätskontrolle
- Robotersteuerung
- Biologie / Mikroskopie
- (digitales) Fernsehen, (digitale) Bildübertragung

Die ersten drei Anwendungsgebiete werden kurz erläutert.

1.2.1 Fernerkundung der Erde

Die Fernerkundung der Erde ist historisch eine der wichtigsten Anwendungen der Bildverarbeitung. Der Grund ist der offensichtliche militärische Nutzen zur Aufklärung. Die Anwendungen reichen heute aber wesentlich weiter:

- Meteorologie,
- Klimaforschung,
- Ozeanographie,
- Rohstoffsuche,
- Umweltforschung,
- Landnutzungsklassifikation.

(die einzelnen Punkte werden erläutert)

Die bildgebenden Verfahren in der Fernerkundung liefern meist *panchromatische* (=schwarz-weiss) oder multispektrale Bilder. Multispektral bedeutet, daß die Sensoren mehrere Kanäle besitzen, die jeweils in Teilbereichen des elektromagnetischen Spektrums Bilder aufzeichnen.

(Folie elektromagnetisches Spektrum)

Typischerweise ist dies der sichtbare Bereich bis zum Infrarot, bzw. zum thermalen Infrarot, oder der Mikrowellen- und Radarbereich. Außer Radargeräten arbeiten alle Geräte passiv, d.h. sie nehmen die von der Sonne emittierten und auf der Erde reflektierten elektromagnetischen Wellen auf. Radargeräte sind aktive Sensoren: sie senden Wellen aus und empfangen das reflektierte und gestreute Signal. Sie besitzen daher einen wesentlichen Vorteil: sie sind weitgehend unabhängig von der Wetterlage. Sie können damit "operationell" eingesetzt werden.

(Folie Fernerkundungsbilder verschiedener Auflösung mit Erläuterungen)

Welche Sensoren benutzt werden, hängt von dem aufzunehmenden Objekt ab. So weiß man, daß manche Pflanzen sich recht gut durch die Reflektion im "Nahen-Infrarot"-Anteil des Lichts *klassifizieren* lassen.

(Folie Klassifikation anhand der Wellenlängen)

1.2.2 Bildgebende Verfahren in der Medizin

Ich skizziere grob einige wichtige bildgebende Verfahren der Medizin.

- Optische Mikroskopie,
- Ultraschall,
- Röntgendiagnostik,
- Röntgen-Computertomographie,
- MRT (Magnetresonanztomographie, Kernspintomographie).

(Die einzelnen Punkte werden erläutert, zum Teil mit Folien.)

Tomographische Probleme werden im Rahmen dieser Vorlesung nicht behandelt. Für dieses auch mathematisch sehr interessante Gebiet verweise ich auf die einschlägige Literatur, insbesondere auf das Buch von Natterer: *The Mathematics of Computerized Tomography*, Wiley, 1986.

1.2.3 Bildverarbeitung in der Automatisierung

Ein typisches Beispiel ist die visuelle Kontrolle von Druckerzeugnissen in der Fertigung. Ein anderes Beispiel kommt aus der Bankenautomation, bei der Geldscheine bei zum Teil hoher Geschwindigkeit erkannt, geprüft und bewertet werden müssen. Ein drittes Beispiel ist die Automatisierung des Postwesens, bei der u.a. Methoden zur Schrifterkennung (OCR) gefragt sind.

1.3 Begriffsklärungen

Bildverarbeitung und Signalverarbeitung

Ein Bild ist ein spezielles, mindestens zweidimensionales Signal. Insofern hängt Bildverarbeitung eng mit Signalverarbeitung zusammen. Der Unterschied ist, daß (die klassische) Signalverarbeitung in der Regel eindimensionale Signale (Funktionen) betrachtet; die Veränderliche ist hier oft die Zeit. Bei mehrdimensionalen Signalen sind die Koordinaten in der Regel korreliert. Der numerische Aufwand steigt mit der Dimension erheblich.

Bildverarbeitung und Bildanalyse

Bildverarbeitung meint in der Regel rein numerische Operationen, wie Transformationen oder Filtern. Bildanalyse bedeutet dagegen die Interpretation der Daten. Bildanalysemethoden verwenden oft stochastische Modelle.

Im nächsten Abschnitt werden wir uns näher mit dem Begriff "digital" auseinandersetzen.

Kapitel 2

Das kontinuierliche und das digitale Bildmodell

Jede konkrete Verarbeitung von Signalen auf dem Computer ist letztlich digital. Andererseits sind die zu untersuchenden Objekte praktisch immer kontinuierlich. Oft entstehen Bilder durch optische Messungen mit analogen, kontinuierlichen Ausgangssignalen, die mit einem ADC (Analog-Digital-Converter) digitalisiert werden.

Eine natürliche Sichtweise ist also, Bilder als kontinuierliche Größen aufzufassen, die auf Computern in digitalisierter Form vorliegen.

2.1 Das kontinuierliche Idealbild

Definition 2.1 Sei $D \subset \mathbb{R}^2$. Eine Funktion $f : D \rightarrow \mathbb{R}_+$ heißt Idealbild.

Bemerkungen

1. Stetigkeit wird in der Regel nicht vorausgesetzt. Im Gegenteil ! Oft enthalten Unstetigkeiten, z.B. Kanten, die eigentliche Information im Bild. Dagegen setzt man aber in der Regel die Beschränktheit voraus.
2. Die kontinuierliche Sichtweise ist zentral für die Analyse und den Entwurf vieler Algorithmen.

Typische Diskretisierungen sind pixelbasiert (\rightarrow §2.2) oder modellbasiert: Enthält das Bild ausschließlich geometrische Objekte (z.B. genau einen Kreis), so läßt es sich durch Angabe der Parameter der Objekte (Mittelpunkt und Radius) sehr kompakt repräsentieren.

2.2 Das digitale Grauwertmodell

Auf die folgende Definition werden wir in der Vorlesung selten direkt verweisen. Trotzdem ist es sinnvoll am Anfang festzulegen worüber man eigentlich spricht. Auch sollten neue Begriffe immer präzise eingeführt werden.

Definition 2.2 Sei $n \in \mathbb{N}$, $\{(x_1, y_1), \dots, (x_n, y_n)\} \in D^n$; f sei Idealbild.

1. Die Menge $\{f(x_1, y_1), \dots, f(x_n, y_n)\} \in \mathbb{R}_+^n$ heißt *Abtastung von f* .
2. Seien D_1, \dots, D_n offene Teilmengen von D , so daß $D_i \cap D_j = \emptyset$, $i \neq j$, und $\bigcup_{i=1}^n \bar{D}_i = D$, $(x_i, y_i) \in D_i$. Dann heißt $(D_i, f(x_i, y_i))_{i=1}^n$ (*Ortsbereichs-*) *Diskretisierung von f* .
3. Ein *diskretes Bild* ist die Ortsbereichsdiskretisierung eines Idealbildes, bei dem D_1, \dots, D_n jeweils das Innere von Quadraten einer gegebenen Kantenlänge sind, und D ein Rechteck ist.
4. Ist $G = \{G_1, \dots, G_k\} \subset \mathbb{N}_0$, dann heißt $\{D_i, g_i\}_{i=1}^n$, $g_i \in G$, *digitales Bild*. Die Zahl g_i heißt *Grauwert in D_i* . Eine Abbildung, die einem diskreten Bild ein digitales Bild zuordnet, heißt *Quantisierung*.

Bemerkungen

1. Die D_i heißen auch Pixel (aus dem engl. =picture elements).
2. Digitale Bilder lassen sich durch eine Matrix repräsentieren:

$$S = (s(x, y))_{x=0, y=0}^{L-1, R-1}, \quad s(x, y) \in G.$$

Als Konvention im Skript bedeutet die erste Koordinate die Nummer der Zeile und die zweite Zahl die Nummer der Spalte, von links oben aus gezählt. Das Matrixmodell ist grundlegend und wichtig. Wir werden es im folgenden benutzen. Typische Werte sind

$$G = \{0, 1, \dots, 255\}, \quad G = \{0, 1\} \quad (\text{Binärbild})$$

$$L \times R = 512 \times 512, 1024 \times 1280, 960 \times 1280, 480 \times 640, \dots$$

Wenn nicht anders erwähnt, werden im folgenden immer digitale Grauwertmatrizen mit $G = \{0, \dots, 255\}$ betrachtet.

————— *lil*

2.3 Diskrete Farbbildmodelle

2.3.1 Farbwahrnehmung

Die Informationen aus diesem und dem folgenden Abschnitt sind aus dem Buch [1] entnommen, auf das für eine detailliertere Darstellung verwiesen sei. Zur Farbwahrnehmung des Menschen gibt es zwei verschiedene Theorien: die des trichromatischen Farbsehens und die des Farbsehens durch Komplementärfarben.

Trichromatisches Farbsehen. Im menschlichen Auge gibt es drei primäre Klassen von Zapfen mit unterschiedlichen Farbempfindlichkeiten:

- blau bei etwa 440nm
- grün bei etwa 545nm
- rot bei etwa 580nm

(Folie mit Abb. 3.12 aus [1])

Die trichromatische Theorie kann verschiedene Phänomene nicht erklären, wie etwa den Farbantagonismus: verschiedene Farben können nicht in Kombination gesehen werden (z.B. existiert kein "rötliches Grün"). Dies führt zur

Komplementärfarben Theorie. Man nimmt sechs Primärfarben (rot/grün, blau/gelb, schwarz/weiß) an, von denen jeweils zwei komplementär zueinander sind. Bei Mischen der Komplementärfarben erhält man Weiß bzw. Grautöne. Dies entspricht der Tatsache, daß im Auge Farb-Zapfen von Gegenfarb-Umfeldern umgeben sind, mit denen man den Farbantagonismus erklärt (siehe [1, S. 80]).

2.3.2 Farbräume

Eine farbige Lichtquelle wird durch ihre Farbanteile im elektromagnetischen Spektrum bestimmt, die durch die Funktion $P : \mathbb{R} \rightarrow \mathbb{R}_0^+$ gegeben seien. Ausgehend vom trichromatischen Farbsehen geht man meist von drei primären Lichtquellen mit zugeordneten spektralen Energieverteilungen $P_k : \mathbb{R} \rightarrow \mathbb{R}_0^+$ aus, $k = 1, 2, 3$, aus denen sich die Lichtquelle zusammensetzt,

$$P = \beta_1 P_1 + \beta_2 P_2 + \beta_3 P_3.$$

Es sei dabei

$$\int_{\mathbb{R}} P_k(\lambda) d\lambda = 1, \quad k = 1, 2, 3.$$

Man ordnet dann jedem der drei Zapfentypen ein Absorptionsspektrum $S_i : \mathbb{R} \rightarrow \mathbb{R}^+$ zu, $i = 1, 2, 3$. Die “spektrale Antwort” der Zapfen auf die farbige Lichtquelle P ist deren Projektion auf diese Absorptionsspektren,

$$\alpha_i(P) = \int_{\mathbb{R}} S_i(\lambda) P(\lambda) d\lambda, \quad i = 1, 2, 3.$$

Bezeichnet man die Antwort des i -ten Zapfentyps auf die k -te Primärquelle P_k mit

$$\alpha_{i,k} := \int_{\mathbb{R}} S_i(\lambda) P_k(\lambda) d\lambda,$$

dann kann $\alpha_i(P)$ geschrieben werden als

$$\alpha_i(P) = \beta_1 \alpha_{i,1} + \beta_2 \alpha_{i,2} + \beta_3 \alpha_{i,3}.$$

Üblicherweise normiert man diese Koeffizienten noch; bezeichnet man mit w_k den Anteil der Primärquelle P_k am weißen Licht, dann betrachtet man statt β_k meist

$$T_k(P) = \frac{\beta_k}{w_k}.$$

Dieses Modell bezeichnet man als “additiven” Farbraum oder RGB-Farbraum (**R**ed**G**reen**B**lue). Additive Farbräume werden etwa für Computermonitore verwendet.

(Erläuterung zum CIE-**R**GB-Farbsystems anhand Folie mit Abb. 3.13 aus [4])

Das CIE-**R**GB-Farbsystem hat teilweise negative Spektralwertkurven. Für Anwendungen sind oft andere Farbräume interessanter, die sich aus dem CIE-**R**GB-System durch lineare Transformationen errechnen lassen (**X**YZ, **N**TSC-**R**GB).

Zusammengefasst werden **R**GB-Farbräume dann verwendet, wenn die Farbe aus dem Mischen mehrerer Strahlungsquellen unterschiedlicher Wellenlängen entsteht. Wenn alle Strahlungsquellen gemeinsam leuchten, entsteht weisses Licht. Umgekehrt entsteht Farbe aber auch dadurch, dass einer weissen Lichtquelle durch Filterung gewisse Wellenlängen entzogen werden. Dieses Farbmodell, was z.B. bei gedruckten Bildern Anwendung findet, nennt man “subtraktiv”. Wichtig sind dabei die **CMY**- (**C**yan**M**agenta**Y**ellow) und **CMYK**-Farbräume (**C**yan**M**agenta**Y**ellow**B**lac**K**). Die Farben spielen hierbei die Rollen von Filtern, alle Filter zusammen ergeben, zumindest theoretisch, Schwarz. Bei Farbdruckern, bei denen die Farben niemals

absolut rein hergestellt werden können, wird das CMYK-Modell verwendet und also schwarz als Primär-“Filter” hinzugenommen. Dies hat auch den Vorteil, dass statt drei Farben nur eine zum Drucken verwendet wird.

Alle Farbräume, ob additiv oder subtraktiv, sind letztlich verschiedene Koordinatensysteme eines 3D-Raumes. Dabei unterscheiden sich die Basisfunktionen leicht, die Räume sind also nicht völlig identisch. Sie werden durch unterschiedliche Anwendungen erforderlich und lassen sich durch lineare Transformationen näherungsweise umrechnen. RGB-Modelle werden z.B. von MS Windows zum Ansteuern der Monitore verwendet. Farbtintenstrahldrucker verwenden typischerweise CMYK. Für die Umsetzung wird ein Programm benötigt, der Druckertreiber.

Andere Anwendungen können z.B. erfordern dass der geometrische Abstand zweier Farben im Farbraum in Relation zum sinnlichen Farbeindruck steht: bei visuell ähnlichen Farben ist die euklidische Distanz der Koordinaten klein, bei unähnlichen Farben eher gross. Dafür dient z.B. der HSV-Farbraum bzw. das HSV-Koordinatensystem (**H**ue**S**aturation**V**alue).

2.4 Elementare Kenngrößen von Bildern

Für die Analyse von Bildern und den Entwurf entsprechender Algorithmen spielen die Konzepte der mathematischen Statistik eine wichtige Rolle. Dafür gibt es viele Gründe; der wichtigste ist wohl dass die Statistik Kenngrößen bereitstellt (z.B. Mittelwert, Varianz), mit deren Hilfe sich wesentliche Bildinhalte sehr komprimieren lassen. Dazu kommt dass meist sowohl die beobachteten Größen als auch die bildgebenden Verfahren stochastischer Natur sind. Oft ist es sinnvoll Pixel als Zufallsvariable zu modellieren. Und nicht zuletzt liefert die Statistik einen umfangreichen Schatz von Modellen, Begriffen und Methoden sowie theoretisch fundierte Aussagen. Den beschränkten Umfang und die Ausrichtung der Vorlesung im Auge, beschränken wir uns hier auf die wichtigsten elementaren Kenngrößen. Eine umfangreichere Darstellung statistischer Konzepte und ihrer Anwendungen ist in [1] und in [4] zu finden. Für eine wesentlich tiefere Behandlung der Thematik siehe [11].

Ein wichtiger Parameter eines gegebenen Bildes ist sein Kontrast. Dieser kann mit Hilfe elementarer Kenngrößen gemessen werden. Man unterscheidet dafür in der Bildverarbeitung zwischen Statistiken erster Ordnung, bei denen Nach-

barschaftsbeziehungen zwischen Pixeln unberücksichtigt bleiben, und Statistiken zweiter Ordnung, für die diese Beziehungen maßgeblich sind.

Definition 2.3

1. $m_S := \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y)$
heißt mittlerer Grauwert.

2. $q_S := \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} (s(x, y) - m_S)^2$
heißt mittlere quadratische Abweichung.

3. Sei

$$H_S(g) := \#\{(x, y) : s(x, y) = g, x = 0, \dots, L - 1, y = 0, \dots, R - 1\}$$

die Anzahl der Bildpunkte (x, y) mit $s(x, y) = g \in G$. Der Graph von H_S heißt Histogramm von S . Die Zahl

$$p_S(g) := \frac{1}{LR} H_S(g)$$

heißt die relative Häufigkeit des Grauwerts g .

Bemerkung. q_S ist ein Maß für den globalen Bildkontrast.

Elementare Folgerungen.

1. $\sum_{g=0}^{255} p_S(g) = \frac{1}{LR} \sum_{g=0}^{255} H_S(g) = 1$

2. $m_S = \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y) = \frac{1}{LR} \sum_{g=0}^{255} g \underbrace{H_S(g)}_{LR p_S(g)} = \sum_{g=0}^{255} g p_S(g)$

3. $q_S = \frac{1}{LR} \sum_{g=0}^{255} (g - m_S)^2 H_S(g) = \sum_{g=0}^{255} (g - m_S)^2 p_S(g)$

Für das Histogramm eines kontrastarmen Bildes gilt

- $p_S(g)$ groß für dicht beieinanderliegende Grauwerte,

- $p_S(g) \approx 0$ sonst.

Falls $p_S(y) \approx \frac{1}{256}$ für alle $g \in G = \{0, 1, \dots, 255\}$, dann liegt ein kontrastreiches Bild vor. In Abschnitt §3.1 wird eine Optimalitätseigenschaft solch gleichverteilter Histogramme gezeigt.

Ein weiteres Hilfsmittel zur Erfassung des lokalen Kontrastes ist die *Grauwert-Übergangsmatrix*.

(x', y') heißt rechter Nachbar von (x, y) , wenn $x' = x$ und $y' = y + 1$.
analog: linker/unterer/oberer Nachbar von (x, y) .

Sei $A(g, g')$ die Anzahl der Paare $(x, y), (x', y')$ mit

- (x', y') ist rechter Nachbar von (x, y) .
- $s(x, y) = g, s(x', y') = g'$

Sei

$$W_{S,S} = (A(g, g'))_{g, g' \in G}.$$

$W_{S,S}$ heißt *Grauwert-Übergangsmatrix*.

analog: linker/unterer/oberer Nachbar von (x, y) ; auch Kombinationen wurden betrachtet.

Beispiel. (rechter Nachbar) $G = \{0, 1, 2, 3\}$

$$S = \begin{bmatrix} 0 & 0 & 1 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 & 3 & 3 \\ 1 & 2 & 2 & 3 & 3 & 3 \\ 2 & 2 & 3 & 3 & 3 & 3 \end{bmatrix}; \quad W_{S,S} = \begin{bmatrix} 4 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

Kontrastarme Bilder haben Grauwert-Übergangsmatrizen, bei denen die Hauptdiagonale dominiert. Bei kontrastreichen Bildern sind auch in den Ecken links unten und rechts oben betragsmäßig große Einträge.

Kapitel 3

Codierung von Bildern

Das Ziel der Bildcodierung ist die Datenreduktion oder die Datenkompression bei digitalen Bildern.

Bilddatenreduktion beinhaltet das Weglassen von *unwichtiger* Information, eine verlustfreie Rekonstruktion ist in der Regel nicht möglich. Dagegen besteht die Bildkompression im Weglassen *redundanter* Information. Hier ist eine Rekonstruktion möglich. Die Begriffe werden in der Literatur aber nicht immer klar auseinandergehalten.

3.1 Vorbemerkungen zum Informationsgehalt von Bildern

Der (syntaktische) Informationsgehalt eines Bildes wird in der minimalen Anzahl bits gemessen, mit der das Bild verlustfrei codiert werden kann. Die Standard-Codierung, die jedem Pixel einen 8-bit-Vektor zur Darstellung seines Grauwertes zuweist, ist meist hochgradig redundant. So ist unmittelbar klar, daß sich ein Grauwert g bereits mit $\lceil 1 + \log_2 g \rceil$ bits binär codieren läßt. Weiterhin ist man bei der Codierung nicht an die Reihenfolge der Grauwerte gebunden. Man weist vielmehr häufig auftretenden Grauwerten kurze Codeworte zu, und selten auftretenden Grauwerten lange. Eine “optimale” Codierung eines Bildes hängt vom Grauwert histogramm ab. Der Informationsgehalt von Bildern ist eine Größe, die aus dem Histogramm bestimmt werden kann. Die Untersuchung und Konstruktion “optimaler” Codierungen ist Gegenstand der Informationstheorie. Ein zentraler Begriff aus diesem Gebiet ist die *Entropie*. Dieser Begriff spielt auch in

der Bildverarbeitung eine Rolle und wird hier oft als “mittlerer (syntaktischen) Informationsgehalt [in bit]” bezeichnet.

Definition 3.1 Sei

$$p_S(g) = \frac{H_s(g)}{LR}$$

die relative Häufigkeit des Grauwerts g im Bild S . Die Größe

$$I_{mit}(S) = \sum_{g \in G} p_S(g) \log_2 \frac{1}{p_S(g)}$$

heißt Entropie oder mittlerer (syntaktischer) Informationsgehalt des Bildes S .

Der Zusammenhang mit der anfangs eingeführten Sprechweise (Informationsgehalt = minimale Anzahl bits zur Codierung des Bildes) ist durch einen wichtigen Satz von C. Shannon gegeben. Sei $l(g)$ die Länge des Codeworts für den Grauwert g , und sei die *mittlere Codewortlänge* l_m des Bildes S gegeben durch

$$l_m(S) = \sum_{g \in G} l(g) p_S(g).$$

Eine Codierung heißt optimal für S , wenn sie $l_m(S)$ minimiert.

Satz 3.1 (Shannon) Jede Codierung erfüllt

$$I_{mit}(S) \leq l_m(S).$$

Für jedes Bild S gibt es eine Codierung, so daß

$$l_m(S) < I_{mit}(S) + 1.$$

Ein konstruktives Verfahren für eine Codierung nach dem Shannon’schen Satz ist die Huffman-Codierung (siehe Abschnitt 3.2).

Mit der oben eingeführten Sprechweise bedeutet Shannons Satz, daß der mittlere Informationsgehalt nach oben durch die optimale mittlere Wortlänge beschränkt ist. Wann wird das Gleichheitszeichen angenommen, d.h., welche Bilder haben einen maximalen mittleren Informationsgehalt bei einer gegebenen mittleren Wortlänge? Der folgende Satz gibt an, wie das Histogramm eines solchen Bildes aussehen muß.

Satz 3.2 Für die Funktion $f : (0, 1)^n \rightarrow \mathbb{R}_+$,

$$f(p_1, p_2, \dots, p_n) := \sum_{i=1}^n p_i \log_2 \frac{1}{p_i},$$

gilt unter der Nebenbedingung

$$\sum_{i=1}^n p_i = 1 \tag{3.1}$$

die Ungleichung

$$f(p_1, p_2, \dots, p_n) \leq \log_2 n.$$

Gleichheit wird genau dann angenommen, wenn

$$p_1 = p_2 = \dots = p_n = \frac{1}{n}.$$

Beweis. Bei dem Beweis verwenden wir die Ungleichung

$$\ln x \leq x - 1. \tag{3.2}$$

Das Gleichheitszeichen ist genau im Fall $x = 1$ gültig. (Beweis der Ungleichung: Übungsaufgabe !)

Sei $D(p_1, \dots, p_n) = f(p_1, \dots, p_n) - \log_2 n$. Mit (3.1) ergibt sich

$$\begin{aligned} D &= \sum_{i=1}^n p_i \left(\log_2 \frac{1}{p_i} - \log_2 n \right) \\ &= \sum_{i=1}^n p_i \log_2 \frac{1}{n p_i}. \end{aligned}$$

Mit $\log_2 n = \log_2 e \ln n$ und Ungleichung (3.2) folgt

$$D = \log_2 e \sum_{i=1}^n p_i \ln \frac{1}{n p_i} \leq \log_2 n \sum_{i=1}^n p_i \left(\frac{1}{n p_i} - 1 \right),$$

wobei das Gleichheitszeichen genau dann angenommen wird, wenn $n p_i = 1$, $i = 1, \dots, n$. QED

Zusammengefaßt bedeutet der Satz, daß alle Codewörter gleich häufig angenommen werden. In Abschnitt §4.3 werden wir eine Transformation für Bilder kennenlernen, bei der das Histogramm des Ergebnisbildes annähernd gleichverteilt ist.

Für weitere Details zur Informationstheorie und zu Shannon's Ergebnis wird auf [6] verwiesen.

3.2 Huffman-Codierung

Schritt 1. Alle vorkommenden Grauwerte g_i werden entsprechend ihrer (relativen) Häufigkeit aufsteigend sortiert.

Schritt 2. Die zwei geringsten Häufigkeiten werden zusammen addiert und bilden eine neue Wahrscheinlichkeit für das so entstandene Zeichenpärchen. Dabei wird das Zeichen mit der größeren Häufigkeit mit einer 0 markiert, das andere mit einer 1. Die beiden aufaddierten Häufigkeiten werden im folgenden wie ein einzelnes Zeichen mit einer neuen Häufigkeit behandelt.

Schritt 3. Wiederhole Schritt 2 bis alle Häufigkeiten aufaddiert sind. Führe dabei die Markierungen nach links fort, so daß eine binäre Zahl entsteht.

Beispiel:

Grauwert	rel. Häufigkeit	Code1	Code 2
g_1	0,4	1	1
g_2	0,2	01	000
g_3	0,2	000	001
g_4	0,15	0010	010
g_5	0,05	0011	011

Das Beispiel zeigt, daß die Bildung des Codes nicht eindeutig ist, wenn mehrere Zeichen bzw. Zeichengruppen gleiche Häufigkeiten aufweisen. Man kann für die so entstandenen Codes weitere Kenngrößen vergleichen, etwa die mittlere Codewortlänge (hier identisch) oder die Varianz des Codes (hier schneidet Code 2 besser ab). Die Varianz kann z.B. eine Rolle spielen, wenn Daten mit einer konstanten Rate über einen Zwischenpuffer übertragen werden, da dieser dann besser ausgenutzt sein wird.

3.3 Binärbilder

Die Erzeugung eines Binärbildes ist eine sehr einfache und effiziente Methode zur Bilddatenreduktion. Binärbilder können außerdem sehr effizient verarbeitet und analysiert werden.

Definition 3.2 Ein Binärbild ist ein digitales Bild mit $G = \{0, 1\}$.

Gegeben sei ein digitales Bild $S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$. Wir geben einen simplen

Algorithmus zur Erzeugung eines Binärbildes.

bestimme Schwellwert c (engl. *threshold*)

für alle Pixel (x, y) setze

$$s'(x, y) = \begin{cases} 1, & s(x, y) \geq c, \\ 0, & \text{sonst.} \end{cases}$$

Eine typische Anwendung ist die Trennung von Vordergrund und Hintergrund. Eine Verfeinerung besteht darin, adaptive Schwellwerte $c(x, y)$ zu verwenden, die von der Umgebung des betrachteten Bildpunktes abhängen.

3.4 Differenzenbildung

Hinter Differenzenbildung steckt die Idee, daß benachbarte Pixel oft ähnliche Grauwerte haben. Differenzen sind häufig kleiner als die Originalgrauwerte und benötigen daher weniger bits zur Codierung.

Das Prinzip der Differenzenbildung ist aus der Numerik bekannt:

- neuer Wert = alter Wert + Korrekturterm.

Man errechnet zunächst die Grauwertdifferenz zu den rechten Nachbarn:

$$\begin{aligned} S &\rightarrow S' = (s'(x, y)), \\ s'(x, y) &= s(x, y) - s(x, y + 1) + c, \quad \begin{array}{l} x = 0, \dots, L - 2 \\ y = 0, \dots, R - 2 \end{array} \end{aligned}$$

c ist dabei so gewählt, daß $s'(x, y) \geq 0$. Das Abspeichern von $s'(x, y)$ und $s(x, 0)$ ermöglicht eine verlustfreie Rekonstruktion:

$$\begin{aligned} s(x, 1) &= s(x, 0) - s'(x, 0) \\ s(x, 2) &= s(x, 1) - s'(x, 1) \\ &\vdots \end{aligned}$$

Differenzenbildung eignet sich auch zur Kodierung von Bildfolgen (Filmen) und wird beim MPEG-Verfahren (DVD) benutzt. Im Abschnitt über Wavelet-Transformationen kommen wir auf die Idee der Differenzenbildung zurück.

3.5 Lauflängencodierung

Lauflängencodierung besteht im Abspeichern von gleichen Grauwerten in einer Rasterzeile (Grauwert, Beginn, Lauflänge). Die Methode ist besonders effizient bei Binärbildern. Durch Zusammenfassung von “schwankenden” Grauwerten lässt sich oft eine gute Datenreduktion erzielen.

Kapitel 4

Modifikation der Grauwertverteilung

Modifikationen der Grauwertverteilung dienen oft zur Bildverbesserung im Sinne einer Kontrastanreicherung oder zur Normierung des Kontrasts von Bildern. Oft geht eine Kontrastnormierung weiteren Bildverarbeitungsschritten voraus.

Man betrachtet dafür zum einen diskrete Funktionen $f : G \rightarrow G$, die auch *Lookup-tables* (LUT) genannt werden. Sie werden in der Praxis durch einen Speicher realisiert und u.U. auch hardwaremäßig unterstützt.

Sei S' das Bild, das durch Anwendung eines eineindeutigen LUT f auf jedes Pixel eines Bildes S entsteht. Ohne Gefahr der Verwechslung schreiben wir kurz $S' = f(S)$. Dann gilt

$$H_{S'}(g) = H_S(f^{-1}(g)).$$

Ein Beispiel für ein eineindeutiges LUT ist eine Inversion der Grauwerte (Negativbildung). Nicht selten wird die Grauwertmodifikation eine Verringerung der Grauwertstufen bewirken; dann kann die obige Formel natürlich nicht mehr direkt angewandt werden (höchstens wenn man $f^{-1}(g)$ als die Summe aller Urbilder von g auffasst). Oft werden durch Grauwertmodifikationen einzelne Grauwertbereiche gestreckt, andere zusammengefasst. Um die Auswirkung einer solchen Operation auf das Histogramm des Bildes zu erfassen, ist es sinnvoll das Bild kurzzeitig als ein kontinuierliches Idealbild zu auffassen. Dann können wir die Häufigkeit von Grauwerten allerdings nicht mehr zählen.

Übungsaufgabe: wie kann man das Histogramm eines kontinuierlichen Idealbildes definieren ?

Wir betrachten nun monoton steigende und differenzierbare Funktionen $f : (0, 255) \rightarrow (0, 255)$, die auch *Gradationskurven* genannt werden. Ist f eine Gradationskurve, dann geht das Histogramm des gradierten Bildes $S' = f(S)$ für $g \in [0, 255]$ hervor aus

$$\tilde{H}_{S'}(g) = \frac{H_S(f^{-1}(g))}{f'(f^{-1}(g))}, \quad f' = \frac{df}{dg} \quad (4.1)$$

Das heißt, wenn die Gradationskurve steiler ist, dann wird der Grauwertbereich aufgespreizt. Ist sie flacher, dann werden viele Grauwerte zusammengefasst. Für ein digitales Bild kann eine LUT als Abtastung einer Gradationskurve verstanden werden. Dann kann man die obige Formel näherungsweise anwenden, meist mit ganzzahliger Rundung.

Wir diskutieren hier die lineare Skalierung, die stückweise konstante Skalierung und die Grauwertäquilisation (Tonwertangleichung).

4.1 Lineare Skalierung

$$S \rightarrow S' = (s'(x, y)), \quad s'(x, y) = (s(x, y) + c_1) c_2$$

- $c_1 > 0, c_2 = 1$: Bild wird heller
- $c_1 < 0, c_2 = 1$: Bild wird dunkler
- $c_1 = 0, c_2 > 1$: Bild wird kontrastreicher
- $c_1 = 0, c_2 < 1$: Bild wird kontrastärmer

Für den Mittelwert des linear skalierten Bildes gilt

$$\begin{aligned} m_{S'} &= \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s'(x, y) \\ &= \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} (s(x, y) + c_1) c_2 \\ &= c_2 m_S + c_1 c_2. \end{aligned}$$

Für die mittlere quadratische Abweichung gilt

$$\begin{aligned} q_{S'} &= \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} (s'(x, y) - m_{S'})^2 \\ &= \frac{c_2^2}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} (s(x, y) - m_S)^2 = c_2^2 q_S \end{aligned}$$

Bemerkungen zur praktischen Implementierung.

1. Bei der praktischen Implementierung ist eine Rundung erforderlich.
2. Da $s'(x, y) \in G = \{0, 1, \dots, 255\}$, ist die praktische Skalierung stückweise linear,

$$s'(x, y) = \begin{cases} 0, & \text{falls } (s(x, y) + c_1)c_2 \leq 0, \\ 255, & \text{falls } (s(x, y) + c_1)c_2 \geq 255, \\ (s(x, y) + c_1)c_2 & \text{sonst.} \end{cases}$$

Dies bewirkt i.A. einen Informationsverlust (bei den extremen Grauwerten).

Methoden zur Bestimmung von c_1 und c_2 .

1. Seien g_{\min}, g_{\max} die minimalen/maximalen Grauwerte von S . Diese Werte können aus dem Histogramm bestimmt werden. Dann setze

$$c_1 = -g_{\min}, \quad c_2 = \frac{255}{g_{\max} - g_{\min}}.$$

Es gilt

$$s(x, y) \in [g_{\min}, g_{\max}] \Rightarrow s'(x, y) \in [0, 255].$$

Diese Skalierung bewirkt also eine Kontrastanreicherung, (im wesentlichen) ohne Verlust von Bildinformationen.

2. Ein Spezialfall dieser Skalierung ist das *clipping*. Clipping führt zum Schwärzen und Weißen gewisser Regionen. Wähle dafür $\varepsilon > 0$ und g', g'' so, daß

$$g_{\min} < g' < g_{\min} + \varepsilon, \quad g_{\max} - \varepsilon < g'' < g_{\max},$$

$$c_1 = -g', \quad c_2 = \frac{255}{g'' - g'}.$$

Eine lineare Skalierung ergibt

$$s(x, y) \in [g_{\min}, g_{\max}]$$

$$\Rightarrow s'(x, y) \in \left[\underbrace{(g_{\min} - g')}_{<0} \frac{255}{g'' - g'}, \underbrace{(g_{\max} - g')}_{>g'' - g'} \frac{255}{g'' - g'} \right]$$

$$\Rightarrow s'(x, y) = \begin{cases} 0 & s(x, y) \leq g' \\ 255 & s(x, y) > g'' \\ (s(x, y) + c_1)c_2 & \text{sonst.} \end{cases}$$

Typischerweise werden 1-5% der Bildpunkte "geclipt".

4.2 Stückweise konstante Skalierung

Gegeben seien $g_0 = g'_0 = 0$ sowie die K Grauwertpaare

$$(g_1, g'_1), \dots, (g_{K-1}, g'_{K-1}), (g_K, 255).$$

Das stückweise konstant skalierte Bild erhält man durch

$$s'(x, y) = \begin{cases} 0 & \text{falls } 0 \leq s(x, y) < g_1 \\ g'_k & \text{falls } g_k \leq s(x, y) < g_{k+1} \\ 255 & \text{falls } g_K \leq s(x, y). \end{cases}$$

(Skizze)

Stückweise konstante Skalierung tritt z.B. bei der Verringerung der Grauwertstufen auf, insbesondere beim Erzeugen eines Binärbildes.

4.3 Grauwertäquilisation

Die Grauwertäquilisation dient der Normierung der Grauwertverteilung von Bildern. Eine Normierung vereinfacht die weitere Verarbeitung. In der Regel faßt man dabei mehrere Grauwerte zusammen, so daß das grauwertäquilierte Bild weniger Grauwerte als das Originalbild hat. Das menschliche Auge kann etwa 30 Graustufen auseinanderhalten, bei den üblichen 256 Graustufen eines digitalen Bildes läßt sich eine Zusammenfassung der Grauwerte für eine Grauwertäquilisation also rechtfertigen. Man kann die Formel (4.1) benutzen, um bei einem gegebenen Histogramm die zu einer Grauwertäquilisation führende Gradationskurve zu berechnen.

→ siehe Übungsblatt

Kapitel 5

Geometrische Transformationen und Entzerrung

Oft ist es notwendig, ein Bild $S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$ in andere Ortskoordinaten zu überführen. Wichtige Beispiele sind

- Erzeugung von Deckungsgleichheit bei Bildensembeln
- Geometrische Verzerrung bei Flugzeugscanneraufnahmen
- Bildverzerrung durch Aufnahmetechnik

Die Aufgabe ist, aus dem “verzerrten” Bild $S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$ ein “entzerrtes” Bild $S' = (s'(x', y'))_{x=0, y=0}^{L-1, R-1}$,

$$x' = f_1(x, y), \quad y' = f_2(x, y)$$

zu berechnen. Das “entzerrte” Bild entspricht in der Praxis oft der orthogonalen Projektion auf die Kameraebene.

5.1 Vergrößerung

$S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$ seien die gegebenen Bildpunkte. Eine Bildvergrößerung erreicht man mittels

$$x' = ax, \quad y' = ay, \quad a > 1.$$

Eine Bildverkleinerung erzielt man analog mit $0 < a < 1$. Das Hauptproblem bei der Bildvergrößerung ist, daß “Lücken” zu füllen sind. Zusätzlich zu der obigen

linearen Abbildung ist also ein Interpolationsproblem zu lösen.

Beispiel. Bei $a = 2$, d.h.

$$s'(2x, 2y) = s(x, y),$$

sind nicht alle Einträge von S' erklärt. Zur Festsetzung der fehlenden Grauwerte zieht man die Nachbarpixel hinzu. Sei die Gauß-Funktion (engl. floor function) $\lfloor \cdot \rfloor$ definiert durch

$$\lfloor \xi \rfloor := \max\{x \in \mathbb{Z}; \quad x \leq \xi\}$$

(für positive ξ ist dies der ganzzahlige Anteil von ξ). Zu den Standardmethoden für eine Bildvergrößerung gehören die durch die folgenden Vorschriften definierten Interpolationsmethoden:

1.

$$s'(x', y') = s\left(\left\lfloor \frac{x'}{a} \right\rfloor, \left\lfloor \frac{y'}{a} \right\rfloor\right)$$

2. Arithmetische Mittelbildung:

$$\begin{aligned} s'(x', y') = & \frac{1}{4} \left[s\left(\left\lfloor \frac{x'}{a} \right\rfloor, \left\lfloor \frac{y'}{a} \right\rfloor\right) + s\left(\left\lfloor \frac{x'}{a} \right\rfloor, \left\lfloor \frac{y'}{a} \right\rfloor + 1\right) \right. \\ & \left. + s\left(\left\lfloor \frac{x'}{a} \right\rfloor + 1, \left\lfloor \frac{y'}{a} \right\rfloor\right) + s\left(\left\lfloor \frac{x'}{a} \right\rfloor + 1, \left\lfloor \frac{y'}{a} \right\rfloor + 1\right) \right] \end{aligned}$$

3. Gewichtete Mittelbildung über die Nachbarpixel:

$$\begin{aligned} s'(x', y') = & a_{00}s\left(\left\lfloor \frac{x'}{a} \right\rfloor, \left\lfloor \frac{y'}{a} \right\rfloor\right) + a_{01}s\left(\left\lfloor \frac{x'}{a} \right\rfloor, \left\lfloor \frac{y'}{a} \right\rfloor + 1\right) \\ & + a_{10}s\left(\left\lfloor \frac{x'}{a} \right\rfloor + 1, \left\lfloor \frac{y'}{a} \right\rfloor\right) + a_{11}s\left(\left\lfloor \frac{x'}{a} \right\rfloor + 1, \left\lfloor \frac{y'}{a} \right\rfloor + 1\right) \end{aligned}$$

Durch Festlegung von

$$a_{00} + a_{01} + a_{10} + a_{11} = 1$$

erreicht man, daß konstante Bilder konstant interpoliert werden. Die Gewichte a_{mn} können von der Lage von $\left(\frac{x'}{a}\right), \left(\frac{y'}{a}\right)$ im Quadrat

$\left[\left\lfloor \frac{x'}{a} \right\rfloor, \left\lfloor \frac{x'}{a} \right\rfloor + 1\right] \times \left[\left\lfloor \frac{y'}{a} \right\rfloor, \left\lfloor \frac{y'}{a} \right\rfloor + 1\right]$ abhängen.

Noch allgemeiner können für eine Mittelbildung auch noch weiter entfernt liegende Nachbarn hinzugezogen werden. Dies bedeutet einen höheren Rechenaufwand, aber oft werden dadurch bessere Effekte erzielt (größere "Verschmierung").

5.2 Affine Abbildungen

Eine Funktion $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ heißt affine Abbildung, wenn für alle $\alpha_j \in \mathbb{R}$ ($j = 1, \dots, n$) mit $\sum_{j=1}^n \alpha_j = 1$ und alle $\begin{pmatrix} x_j \\ y_j \end{pmatrix} \in \mathbb{R}^2$ gilt

$$\phi \left(\sum_{j=1}^n \alpha_j \begin{pmatrix} x_j \\ y_j \end{pmatrix} \right) = \sum_{j=1}^n \alpha_j \phi \begin{pmatrix} x_j \\ y_j \end{pmatrix}.$$

Ist $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ eine affine Abbildung, so gilt

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \phi \begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{R}^{2 \times 2},$$

denn mit Definition von

1. $\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} := \phi \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
2. $\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} := \phi \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}; \quad \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} := \phi \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$
3. $A := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

folgt

1. $\phi \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$
2. $\phi \begin{pmatrix} 0 \\ 1 \end{pmatrix} = A \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$
3. und für beliebiges $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2,$

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (1 - x - y) \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ \phi \begin{pmatrix} x \\ y \end{pmatrix} &= x \phi \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \phi \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (1 - x - y) \phi \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &= A \begin{pmatrix} x \\ 0 \end{pmatrix} + A \begin{pmatrix} 0 \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \\ &= A \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \end{aligned}$$

Jede affine Abbildung setzt sich aus den folgenden Abbildungen zusammen:

1. Verschiebung (Translation)

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad x' = x + b_1, \quad y' = y + b_2.$$

2. Drehung um Winkel α um den Ursprung

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix}$$
$$\Rightarrow A = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}.$$

Mit

$$A^T = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}$$

folgt

$$\begin{aligned} A^T A &= \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

A^T beschreibt die zu A inverse Abbildung (Drehung um $-\alpha$). Es gilt

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{pmatrix} x \\ y \end{pmatrix} = A^T \begin{pmatrix} x' \\ y' \end{pmatrix},$$

Verschiebung und Drehung sind winkeltreue Abbildungen.

3. Vergrößerung bzw. Verkleinerung (s.o.)

5.3 Interpolation durch Polynome

$$\begin{aligned} x' &= f_1(x, y) = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 + \dots \\ y' &= f_2(x, y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + \dots \end{aligned}$$

Praktisch haben die oft Polynome den Gesamtgrad 1,2 oder 3. Die Koeffizienten a_i, b_i werden bestimmt durch

- Systemkorrekturen (falls die Position der Aufnahmekamera bekannt ist)

- durch die Paßpunktmethod (z.B. bei Luftbildern)

Im folgenden wird die Paßpunktmethod beschrieben. Sei $S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$ sei das verzerrte Originalbild. $R = (r(x, y))_{x=0, y=0}^{L-1, R-1}$ sei das "Referenzbild", an das das Bild S angeglichen werden soll. Gesucht ist das entzerrte Bild $S' = (s'(x, y))$, das mit (Teilen von) R deckungsgleich sein soll.

Beispiel. Interpolation durch lineare Polynome

$$x' = a_0 + a_1x + a_2y, \quad y' = b_0 + b_1x + b_2y.$$

6 Koeffizienten \Rightarrow 3 Paßpunkte werden benötigt

(Skizze mit Paßpunkt (u_j, v_j) , Originalpunkt (x_j, y_j))

$$\begin{aligned} u_j &= a_0 + a_1x_j + a_2y_j, & j &= 0, 1, 2 \\ v_j &= b_0 + b_1x_j + b_2y_j, & j &= 0, 1, 2 \end{aligned}$$

Setze

$$\mathbf{u} := \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad \mathbf{v} := \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad \mathbf{a} := \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \quad \mathbf{b} := \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

$$P := \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{pmatrix}.$$

Die Paßpunktbedingungen sind dann

$$P\mathbf{a} = \mathbf{u}, \quad P\mathbf{b} = \mathbf{v},$$

und die Koeffizientenvektoren \mathbf{a}, \mathbf{b} lassen sich durch Lösung des linearen Gleichungssystems bestimmen.

Bemerkungen.

1. Die Paßpunkte sind in der Regel mit Fehlern behaftet.
2. Bei mehr als drei Paßpunkten kann eine stückweise lineare Interpolation nach vorheriger Triangulation erfolgen.
3. Verallgemeinerung auf höhere Polynomgrade:
Gesamtgrad 2 \rightarrow 12 unbekannte Koeffizienten \rightarrow 6 Paßpunkte
Gesamtgrad 3 \rightarrow 20 unbekannte Koeffizienten \rightarrow 10 Paßpunkte

5.4 Methode der kleinsten Quadrate

Oft ist die Veränderung der Geometrie eines Bildes unerwünscht, die bei Polynomen höheren Grades (als 1) auftritt. Dann werden bevorzugt lineare Polynome verwendet. Wenn mehr als drei Paßpunkte vorliegen, dann kann die Methode der kleinsten Quadrate verwendet werden, um die Summe der Quadratfehler der Abweichungen an den Paßpunkten zu minimieren.

Aufgabe: Bestimme $\mathbf{a} := \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, $\mathbf{b} := \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ so, daß

$$\begin{aligned} \sum_{j=1}^p [u_j - (a_0 + a_1 x_j + a_2 y_j)]^2 &\rightarrow \min(\mathbf{a}) \\ \sum_{j=1}^p [v_j - (b_0 + b_1 x_j + b_2 y_j)]^2 &\rightarrow \min(\mathbf{b}) \end{aligned}$$

Die zu minimierenden Funktionen

$$\begin{aligned} F(a_0, a_1, a_2) &:= (\mathbf{u} - P\mathbf{a})^T (\mathbf{u} - P\mathbf{a}), \\ G(b_0, b_1, b_2) &:= (\mathbf{v} - P\mathbf{b})^T (\mathbf{v} - P\mathbf{b}) \end{aligned}$$

sind "glatt", d.h. differenzierbar. Eine notwendige Bedingung für das Minimum ist, daß dort die partiellen Ableitungen verschwinden,

$$\frac{\partial F}{\partial a_i}(a_0, a_1, a_2) = 0, \quad \frac{\partial G}{\partial b_i}(b_0, b_1, b_2) = 0, \quad i = 0, 1, 2.$$

Es ergeben sich die Normalgleichungen

$$P^T P\mathbf{a} = P^T \mathbf{u}, \quad P^T P\mathbf{b} = P^T \mathbf{v}.$$

Die Lösungen \mathbf{a}_{\min} , \mathbf{b}_{\min} erhält man durch Lösen des linearen Gleichungssystems.

Beurteilung der Güte der geometrischen Entzerrung:

1. mittlerer quadratischer Fehler:

$$\frac{1}{p} F(\mathbf{a}_{\min}), \quad \frac{1}{p} G(\mathbf{b}_{\min})$$

2. Vektorfeld der "Fehlervektoren":

$$\xi_j = \text{Paßpunkt} \begin{pmatrix} u_j \\ v_j \end{pmatrix} - \begin{pmatrix} u'_j \\ v'_j \end{pmatrix}$$

mit

$$\begin{aligned} u'_j &= a_0^{\min} + a_1^{\min} x_j + a_2^{\min} y_j \\ v'_j &= b_0^{\min} + b_1^{\min} x_j + b_2^{\min} y_j \end{aligned}$$

Die Bildentzerrung ist "gut", wenn die Richtungen der Fehlervektoren gleichverteilt sind und die Längen der Fehlervektoren klein.

3. zusätzliche Paßpunkte zur Kontrolle

Bestimmung der Paßpunkte.

1. von Hand

2. halbautomatisch

1. Schritt: ungefähre Bestimmung der (x_j, y_j) von Hand (oder aus Vorabinformationen)

2. Schritt: Flächenkorrelation. Bilde $m \times m$ -Umgebung ($m = 3, 5, 7, \dots$) von (u_j, v_j) in R und $(x_j, y_j) \in S$.

(Skizze)

Die $m \times m$ -Umgebung in S kann entsprechend der Vorabinformation über die Verzerrung an die Geometrie angepaßt sein. Setze

$$\begin{aligned} a(k, l) &= r \left(u_j - \frac{m-1}{2} + k, v_j - \frac{m-1}{2} + l \right), \\ b(k, l) &= s \left(x_j - \frac{m-1}{2} + k, y_j - \frac{m-1}{2} + l \right), \\ &k, l = 0, \dots, m-1. \end{aligned}$$

Ein Maß für den Vergleich der Bildausschnitte ist

$$\frac{1}{m^2} \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} (b(k+p, l+q) - a(k, l))^2.$$

Bestimme p, q so, daß das Maß am kleinsten wird. Ausquadrieren ergibt

$$\frac{1}{m^2} \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} (b(k+p, l+q)^2 + a(k, l)^2) - \underbrace{\frac{2}{m^2} \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} b(k+p, l+q)a(k, l)}_{\text{}}$$

Die unterklammerte Größe nennt man die Korrelation von $(a(k, l))_{k,l}$ mit $(b(k+p, l+q))_{k,l}$. Nimmt man an, daß $\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} (b(k+p, l+q)^2 + a(k, l)^2)$ konstant bleibt, dann ist das Maß ist am kleinsten, wenn die Korrelation am größten ist. Bestimme (p, q) so, daß die Korrelation am größten ist, und ordne $(x_j + p, y_j + q)$ dem Paßpunkt (u_j, v_j) zu.

3. automatisch: aufgrund von Vorabinformationen über die Lage der Kamera und evtl. über die Topographie des aufgenommenen Objekts.

Kapitel 6

Lokale Operatoren und Filter

Lokale Operatoren berücksichtigen im Gegensatz zu Punktoperatoren auch Grauwerte in der Umgebung eines Pixels, im allgemeinen aber nicht alle Bildpunkte (\rightarrow globale Operatoren). Lokale Operatoren haben gegenüber globalen den Vorteil, daß sie schneller berechnet werden können. Wir reden in diesem und in dem folgenden Abschnitt statt von lokalen Operatoren meist von *Filtern*. Dies ist eine Bezeichnung aus der Praxis für solche Operatoren, die

- zur Rauschunterdrückung,
- zur Erkennung von fehlerhaften Pixeln, gestörten Bildzeilen und anderen Bildstörungen,
- zur Beseitigung von Unschärfen und anderer Bildverbesserung,
- zur Unterdrückung von unnötigen feinen Details,
- zur Detektion von Kanten

dienen. Die Detektion von Kanten nimmt dabei eine besonders wichtige Stellung ein, da Kanten oft die wichtigsten (semantischen) Informationsträger eines Bildes sind.

6.1 Gleitender Mittelwert, lineare Filter

Beispiele für lokale Operatoren sind gleitende Mittelwertoperatoren, die zur Glättung von Grauwerten dienen.

Definition 6.1 Seien m_1, m_2 natürliche Zahlen, $H = (h(u, v))$ eine reelle $m_1 \times m_2$ -Matrix. Der Operator $S \rightarrow S' = (s'(x, y))$,

$$s'(x, y) = \sum_{u=0}^{m_1-1} \sum_{v=0}^{m_2-1} s(x+k-u, y+k-v)h(u, v), \quad M = \sum_{u=0}^{m_1-1} \sum_{v=0}^{m_2-1} h(u, v) = 1, \quad (6.1)$$

heißt (gewichteter) gleitender Mittelwert.

Gleitende Mittel sind durch ihre charakteristische Matrix (Maske)

$$H = (h(u, v))_{u,v=0}^{m_1-1, m_2-1}$$

charakterisiert. Für den klassischen moving average lautet die Maske

$$\begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix}.$$

Gebräuchlich sind auch Modifikationen des gleitenden Mittelwertes, bei denen diese Maske anders aussieht (siehe Beispiel unten).

Gleitende Mittelwerte und ihre Modifikationen sind *lineare* Operatoren. Genauer gesagt sind sie diskrete *Faltungsoperatoren*.

Zur vollständigen Definition des transformierten Bildes muß noch der Rand definiert werden. Hierfür gibt es mehrere Standardansätze,

- das Weglassen des Randes (dann wird das Bild kleiner),
- durch Extrapolation des Bildes am Rand,
- eine andere Mittelwertbildung für die Randpunkte, z.B. Mittelbildung über alle Pixel, die noch im Bild enthalten sind,
- durch periodische Fortsetzung von S oder Spiegelung von S am Rand und “randübergreifender” Mittelwertbildung.

Mittelwert und Varianz des Ergebnisbildes sind abhängig von der Behandlung des Randes. Im allgemeinen gilt

$$m_S \approx m_{S'}, \quad q_S \geq q_{S'}.$$

Beispiel. Eine der am häufigsten verwendeten gleitenden Mittel ist der moving average bezgl. der 8-Umgebung von (x, y) .

(Skizze)

Die Maske lautet bei der 8-er Umgebung

$$H = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Andere über 3×3 -Matrizen definierte Filter sind

- das gleitende Mittel bezgl. der 4-er Umgebung:

$$M = 5, \quad H = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

- der Gaußfilter:

$$M = 16, \quad H = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

- die Binomialfilter: diese sind Verallgemeinerungen des Gaußfilters. Der zugrundeliegende eindimensionale Zahlenvektor beim Gaußfilter ist $(1, 2, 1)$, beim Binomialfilter werden die Zeilen höherer Ordnung aus dem Pascal'schen Dreieck genommen (die Binomialkoeffizienten).

Alle obigen Filter sind *Glättungsfilter* oder *Tiefpaßfilter*, die niedrige Frequenzen im wesentlichen unberührt lassen und hochfrequente Anteile aus dem Bild eliminieren. Damit führen sie auch zu einer Rauschverminderung. Im Abschnitt 7.5 werden wir hierauf noch tiefer eingehen. Hier werden wir die Wirkung der Filter am praktischen Beispiel studieren.

- Wirkung des Filters auf eine vertikale Kante

$$\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{array} \rightarrow \begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ 0 & \frac{1}{3} & \frac{2}{3} & 1 \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Scharfe Kanten werden verschmiert.

- Vergleich von gleitendem Mittelwert und Gaußfilter
gleitender Mittelwert:

$$\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ -1 & 1 & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{array} \rightarrow \begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Gaußfilter:

$$\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ -1 & 1 & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{array} \rightarrow \begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

- Tiefpaßfilterung mit linearen Filtern geht immer auf Kosten der Schärfe und der Detailauflösung.

6.2 Median-Filter

Der Median-Filter ist ein nichtlinearer Filter, der Ausreißer (Impulsrauschen) im Bild unterdrückt.

Definition 6.2 *Es seien n reelle Zahlen $x_1 \leq x_2 \leq \dots \leq x_n$ gegeben. Der Median von x_1, \dots, x_n ist definiert durch*

$$\text{median}(x_1, \dots, x_n) := \begin{cases} x_{\frac{n+1}{2}}, & n \text{ ungerade,} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n+2}{2}}), & n \text{ gerade.} \end{cases}$$

Der (gleitende) Median spielt eine wichtige Rolle in der *robusten Statistik*, da er weniger empfindlich gegen Ausreißer ist als etwa der (gleitende) Mittelwert. Er ist für ungerade n diejenige Zahl, die die Summe der Abstände zu x_1, \dots, x_n minimiert. Für gerade n leistet dies jede Zahl aus dem Intervall $(x_{\frac{n}{2}}, x_{\frac{n+2}{2}})$, die obige Definition ist die natürlichste.

Wie beim gleitenden Mittelwert wird beim Medianfilter eine Umgebungsmatrix (Beispiel: 3×3 oder 5×5) zum Bildpunkt (x, y) definiert. Dann werden gemäß Definition 6.2 die Grauwerte in der Umgebungsmatrix der Größe nach geordnet, und der Medianwert ist der neue Grauwert. Auf diese Weise werden einzelne (in Abhängigkeit der Größe der Umgebungsmatrix auch mehrere) fehlerhafte Pixel vom Medianfilter eliminiert. Ein Nachteil der Medianfilter ist, daß auch fehlerfreie Grauwerte verändert werden. Diesem Umstand kann durch Modifikationen der Methode begegnet werden.

6.3 Differenzenoperatoren

Wir definieren in diesem Abschnitt *Differenzenfilter*. Wie in Abschnitt §6.1 sind dies Faltungsoperatoren, aber sie unterscheiden sich durch eine wesentliche Eigenschaft von den Filtern in (??): Die Summe M aller Einträge verschwindet.

Definition 6.3 Seien m_1, m_2 natürliche Zahlen, $H = (h(u, v))$ eine reelle $m_1 \times m_2$ -Matrix. Der Operator $S \rightarrow S' = (s'(x, y))$,

$$s'(x, y) = \sum_{u=0}^{m_1-1} \sum_{v=0}^{m_2-1} s(x+k-u, y+k-v)h(u, v), \quad M = \sum_{u=0}^{m_1-1} \sum_{v=0}^{m_2-1} h(u, v) = 0, \quad (6.2)$$

heißt *Differenzenfilter*.

Beispiele.

- $H = [-1, 1]$ liefert eine Vorwärts-Differenz und damit eine Approximation an die Ableitung in positive x -Richtung
- $H = [-1, 0, 1]$ liefert eine zentrierte Differenz in x -Richtung
- $H = \frac{1}{6} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$ liefert eine zentrierte Differenz in x -Richtung und eine Mittelwertbildung in y -Richtung. Dieser Operator heißt *Prewitt-Operator* (in x -Richtung).
- $H = \frac{1}{6} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$ liefert eine zentrierte Differenz in x -Richtung und eine modifizierte Mittelwertbildung in y -Richtung (gemäß Gauß-Mittel). Dieser Operator heißt *Sobel-Operator* (in x -Richtung).
- Die transponierten Matrizen der obigen Operatoren liefern die entsprechenden Differenzen in y -Richtung.

Differenzenfilter sind *Hochpaßfilter*, die niedrige Frequenzen unterdrücken (siehe Abschnitt 7.5).

Jeder Differenzenfilter ist eine Approximation an die partiellen Ableitungen $s_x(x, y)$ und $s_y(x, y)$ des kontinuierlichen Idealbildes, entsprechende Differenzierbarkeit vorausgesetzt. Aus dem Gradienten

$$\text{grad } s(x, y) = \begin{bmatrix} s_x(x, y) \\ s_y(x, y) \end{bmatrix}$$

läßt sich für jeden Richtungsvektor

$$\mathbf{u} = \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix}$$

die Ableitung in Richtung \mathbf{u} durch Multiplikation errechnen,

$$D_{\mathbf{u}} = \mathbf{u} \cdot \text{grad } s(x, y) = \cos \varphi \cdot s_x(x, y) + \sin \varphi s_y(x, y).$$

Richtungsableitungen sind im Zusammenhang mit der *Konturpunktfindung* wichtig. Sie dienen damit zur *Kantenfindung* (allgemeiner zur Detektion von Singularitäten) und außerdem zur Erzeugung von *pseudoplastischen Bildern*.

Neben den Differenzenfiltern, die erste Ableitungen approximieren, werden in der Bildverarbeitung auch diskrete Ableitungen zweiter Ordnung benutzt. Im kontinuierlichen Fall ist der *Laplace-Operator* definiert durch

$$\Delta s(x, y) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) s(x, y).$$

Die diskreten Laplace-Filter zu den 4-er und 8-er Nachbarschaften sind definiert durch

$$H_{L1} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \quad H_{L2} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

Auch die Maske

$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

wird verwendet und als diskreter Laplace-Operator bezeichnet. Mit einem Laplace-Filter werden strukturreiche Bereiche hervorgehoben und weniger strukturreiche Bereiche unterdrückt: (Beispiele für H_{L1})

- Ein Einzelpunkt wird vergrößert und verstärkt.

$$\begin{array}{ccccccc}
 & & \vdots & & & & \vdots \\
 & & 0 & 0 & 0 & & 0 & -1 & 0 \\
 \dots & & 0 & 1 & 0 & \dots & \rightarrow & \dots & -1 & 4 & -1 & \dots \\
 & & 0 & 0 & 0 & & & & 0 & -1 & 0 \\
 & & \vdots & & & & & & \vdots & & &
 \end{array}$$

- Eine Linie wird vergrößert und verstärkt.

$$\begin{array}{ccccccccccc}
 & & & & \vdots & & & & & & \vdots \\
 & & & & 0 & 0 & 1 & 0 & 0 & & 0 & -1 & 2 & -1 & 0 \\
 \dots & & & & 0 & 0 & 1 & 0 & 0 & \dots & \rightarrow & \dots & 0 & -1 & 2 & -1 & 0 & \dots \\
 & & & & 0 & 0 & 1 & 0 & 0 & & & & 0 & -1 & 2 & -1 & 0 \\
 & & & & \vdots & & & & & & & & \vdots & & & & &
 \end{array}$$

- Das Ende einer Strecke wird verstärkt.

$$\begin{array}{ccccccccccc}
 & & & & \vdots & & & & & & \vdots \\
 & & & & 0 & 0 & 1 & 0 & 0 & & 0 & -1 & 2 & -1 & 0 \\
 \dots & & & & 0 & 0 & 1 & 0 & 0 & \dots & \rightarrow & \dots & 0 & -1 & 3 & -1 & 0 & \dots \\
 & & & & 0 & 0 & 0 & 0 & 0 & & & & 0 & 0 & -1 & 0 & 0 \\
 & & & & \vdots & & & & & & & & \vdots & & & & &
 \end{array}$$

- Eine Kante wird eine verbreiterte Linie.

$$\begin{array}{ccccccccccc}
 & & & & \vdots & & & & & & \vdots \\
 & & & & 0 & 0 & 1 & 1 & & & 0 & -1 & 1 & 0 \\
 \dots & & & & 0 & 0 & 1 & 1 & \dots & \rightarrow & \dots & 0 & -1 & 1 & 0 & \dots \\
 & & & & 0 & 0 & 1 & 1 & & & & 0 & -1 & 1 & 0 \\
 & & & & \vdots & & & & & & & \vdots & & & & & &
 \end{array}$$

- Eine Linie mit Ecke wird vergrößert und verstärkt.

$$\begin{array}{ccccccc}
& & \vdots & & & \vdots & \\
& & 0 & 0 & 1 & 0 & \\
& & 0 & 0 & 1 & 0 & \\
\cdots & 1 & 1 & 1 & 0 & \cdots & \rightarrow & \cdots & 2 & 2 & 2 & -1 & \cdots \\
& & 0 & 0 & 0 & 0 & & & -1 & -1 & -1 & 0 & \\
& & \vdots & & & \vdots & & & & & & &
\end{array}$$

Der Hauptzweck von Laplace-Filtern ist die Kontrastverstärkung. Ist S ein Bild und S' das aus S durch Laplace-Filterung entstandene Bild, dann liefert $S + S'$ eine Verstärkung der Kanten in S .

6.4 Elemente der Mathematischen Morphologie

Die Morphologie ist die Lehre von der Form. Die mathematische Morphologie beschäftigt sich mit Methoden, die es ermöglichen, Formen aus Bildern zu extrahieren. Ausführliche Informationen zur mathematischen Morphologie sind im Buch [1] zu finden, in dem auch auf weiterführende Literatur verwiesen wird.

Wir betrachten hier Operationen auf Binärbildern. Die zulässigen Grauwerte sind also 0 und 1. Zwei wesentliche Operationen sind die *Dilatation* (Expansion) und die *Erosion* (Kontraktion). Die Dilatation hat das Ziel, die Flächen mit dem Grauwert 1 (weiß) zu vergrößern. Dazu betrachtet man für jeden Pixel (x, y) seine 8-er Umgebung (allgemeiner: ein *strukturierendes Element*). Für das dilatierte Bild $S' = (s'(x, y))$ gilt $s'(x, y) = 1$ genau dann, wenn in der Umgebung von (x, y) mindestens ein Pixel mit Grauwert 1 liegt, und es gilt $s'(x, y) = 0$ sonst.
(Skizze)

Die Erosion hat das gegenteilige Ziel. Für das erodierte Bild gilt $s'(x, y) = 1$, falls in der Umgebung von (x, y) alle Pixel den Grauwert 1 haben, und sonst $s'(x, y) = 0$.
(Skizze)

Eine Anwendung von Erosion und anschließender Dilatation heißt *Öffnung* (*Opening*). Umgekehrt heißt eine Anwendung von Dilatation und anschließender Erosion *Schließen* (*Closing*). Durch Öffnen verschwinden kleine Objekte vollständig. Ebenso werden konvexe Randstrukturen abgetragen. Beim Schließen wachsen eng aneinanderliegende Objekte zusammen, Löcher in Objekten werden geschlossen und konkave Strukturen an den Objektträgern werden aufgefüllt. Zusätzlich ent-

stehen Brücken zwischen eng zusammenliegenden Objekten, die einzelne Objekte stark verändern.

Die mathematische Morphologie liefert auch eine wichtige Methode, um einzelne Bildelemente aus dem (Binär-) Bildzusammenhang zu nehmen und zu isolieren. Diese Methode heißt *grassfire* (Steppenbrand). Wir betrachten die Grassfire-Operation zum Zählen von Objekten.

Nehmen wir an, daß der Hintergrund im Bild den Grauwert 0 und die Objekte den Grauwert 1 haben. Zunächst müssen verschiedene Objekte im Bild auch einwandfrei getrennt sein, d.h. die Aussenkonturen der Objekte dürfen sich nicht berühren. Hierfür eignet sich z.B. eine Öffnungsoperation. Dann muss ein Punkt ermittelt werden, der sich innerhalb eines Objekts befindet, z.B. mittels einer Suche nach einem Pixel mit Grauwert 1. Dieser Punkt heißt dann Zündpunkt. Von dem Zündpunkt ausgehend werden die 4-er oder 8-er Nachbarschaften abgearbeitet, indem sie auf den Grauwert 0 gesetzt (*verbrannt*) werden. Nach Abbrennen eines Objekts wird der Objektzähler um eins erhöht und der Vorgang wird beginnend mit der Suche nach einem Zündpunkt wiederholt. Die Suche bricht ab wenn kein solcher Punkt mehr gefunden wird.

Die mathematische Morphologie stellt noch weitere Methoden zur Objektzählung zur Verfügung, siehe [1, §4.6] und die dort erwähnte Literatur.

Kapitel 7

Fouriertransformationen

Anders als im vorhergehenden Kapitel transformieren die in diesem und im folgenden Abschnitt vorgestellten Methoden das Bild in einen anderen Raum, den *Frequenzraum*. Dort wird das Bild verarbeitet und dann mit der inversen Transformation wieder in den *Ortsraum* zurück abgebildet. Wir wiederholen kurz die für diese Vorlesung relevanten Eigenschaften der eindimensionalen Fouriertransformation und der Fourierreihen, gehen auf die zweidimensionale kontinuierliche Transformation ein und behandeln dann die wichtige diskrete Fourier-Transformation. Abschliessend gehen wir auf die diskrete Cosinus-Transformation und ihre Anwendung zur Bilddatenkompression ein.

7.1 Rückblick auf die eindimensionale kontinuierliche Fouriertransformation und auf Fourierreihen

Die Fouriertransformation einer absolutintegablen Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ ist gegeben durch

$$\mathcal{F}(s, x) = \int_{\mathbb{R}} s(t) e^{-i2\pi xt} dt,$$

und s kann aus $\mathcal{F}(s)$ durch Rücktransformation gewonnen werden,

$$s(t) = \int_{\mathbb{R}} \mathcal{F}(s, x) e^{i2\pi xt} dx.$$

Die für die Signal- und Bildverarbeitung wichtigsten Eigenschaften sind:

- Linearität

- das Faltungstheorem
- das Ähnlichkeitstheorem

Fourierreihen sind eine andere Möglichkeit, Funktionen in ihre Frequenzanteile zu zerlegen. Bei Fourierreihen geht man von periodischen Funktionen aus, als Periode nimmt man meist 2π an. Um die Grundperiode $M > 0$ zu erreichen, multipliziert man das Argument mit $\frac{2\pi}{M}$. Periodische Funktionen sind bei der kontinuierlichen Fouriertransformation ausgeschlossen, da keine periodische Funktion (außer der Nullfunktion) absolutintegabel über der reellen Achse ist. Für Bilder paßt zunächst keines der beiden Konzepte, da Bilder nur für eine Teilmenge des \mathbb{R}^2 (oder \mathbb{C}) definiert sind – vgl. Definition 2.1. Sie können aber für beide Konzepte passend gemacht werden. Für die kontinuierliche Fouriertransformation setzt man dafür den Funktionswert ausserhalb des Definitionsbereiches des Bildes zu null. Für Fourierreihen setzt man das Bild in beide Richtungen periodisch fort.

Als *Fourierreihe* einer Funktion $s : (0, M) \rightarrow \mathbb{R}$ bezeichnet man die Funktion

$$s(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{2\pi kx}{M}\right) + b_k \sin\left(\frac{2\pi kx}{M}\right) \right), \quad (7.1)$$

wobei für $M = 2\pi$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} s(x) \cos(kx) dx, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} s(x) \sin(kx) dx$$

und für andere M müssen diese Formeln entsprechend transformiert werden. Der konstante Anteil in der Fourierreihe ist mit dem Faktor $\frac{1}{2}$ versehen, diese Vereinbarung macht spätere Fallunterscheidungen unnötig. Man schreibt \sim , solange nicht geklärt ist ob die Reihe auf der rechten Seite konvergiert und damit die Funktion überhaupt existiert, und ob dann die rechte Seite mit der Funktion auf der linken Seite punktweise identisch ist. Dies ist unter genügend starken Bedingungen an s der Fall, siehe die Vorlesungen über Analysis und vertiefend über Approximationstheorie. Mittels einer Fourierreihe wird die Funktion also auf die Zahlenfolgen (a_k) , (b_k) abgebildet, und die Frage der Inversion, also die Rekonstruktion der Funktion aus den Koeffizienten, kann unter den oben erwähnten Voraussetzungen an s mittels der Darstellung (7.1) garantiert werden.

Neben diesen *reellen* Fourierreihen werden häufig auch *komplexe Fourierreihen* betrachtet. Dies dient im wesentlichen nur der bequemeren Schreibweise. Man

setzt für den geraden und den ungeraden Teil der Fourierreihe

$$g(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi kx}{M}\right), \quad u(x) = \sum_{k=1}^{\infty} b_k \sin\left(\frac{2\pi kx}{M}\right),$$

sowie $\beta_k = -ib_k$. Dann folgt

$$s(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{2\pi kx}{M}\right) + i\beta_k \sin\left(\frac{2\pi kx}{M}\right) \right), \quad (7.2)$$

und mit

$$\frac{a_k + \beta_k}{2} e^{\frac{2\pi i}{M} kx} + \frac{a_k - \beta_k}{2} e^{-\frac{2\pi i}{M} kx} = \frac{a_k + \beta_k}{2} \omega^{kx} + \frac{a_k - \beta_k}{2} \omega^{-kx},$$

$\omega = e^{\frac{2\pi i}{M}}$, schließlich

$$s(x) \sim \sum_{k=-\infty}^{\infty} c_k \omega^{kx}, \quad c_0 = \frac{a_0}{2}, \quad c_{\pm k} = \frac{a_k \mp i b_k}{2}.$$

Die Zahl ω ist die primitive M -te Einheitswurzel, also $\omega^M = 1$. Die Funktion s ist immer noch reellwertig, die Notation ist nur etwas kompakter geworden. Der Nachteil ist, daß jetzt aus der Reihe eine ‘‘Doppelreihe’’ (Laurent-Reihe) geworden ist.

Mit dieser Notation kann eine Funktion s , unter den oben erwähnten Bedingungen, durch ihr *Spektrum*

$$\dots, c_{-2}, c_{-1}, c_0, c_1, c_2, \dots$$

repräsentiert werden. Neben der oben erwähnten Repräsentierbarkeit ist auch die Frage der Approximationsgenauigkeit der Teilsummen

$$\frac{a_0}{2} + \sum_{k=1}^N a_k \cos(kx) + b_k \sin(kx)$$

bzw.

$$\sum_{k=-N}^N c_k \omega^{kx}$$

von großer Bedeutung.

7.2 Zweidimensionale kontinuierliche Fouriertransformation

Die zweidimensionale Fouriertransformation eines kontinuierlichen Idealbildes S ist definiert als

$$\mathcal{F}(s; t_1, t_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} s(x, y) e^{2\pi i(t_1 x + t_2 y)} dx dy.$$

Für Bilder mit kompaktem Träger definiert man dabei $s(x, y) = 0$ für alle (x, y) , die nicht im Bereich D des Bildes liegen. Eine hinreichende Bedingung für die Existenz des Integrals ist die Absolutintegrierbarkeit der Bildfunktion $s(x, y)$,

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |s(x, y)| dx dy < \infty.$$

Die Rücktransformation des Fourierbildes in den *Ortsraum* heißt *inverse zweidimensionale Fouriertransformation*,

$$s(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{F}(s; t_1, t_2) e^{2\pi i(t_1 x + t_2 y)} dt_1 dt_2.$$

Das Fourierbild ist eine komplexe Funktion, die durch Real- und Imaginärteil,

$$\mathcal{F}(s; t_1, t_2) = \Re\{\mathcal{F}(s; t_1, t_2)\} + i\Im\{\mathcal{F}(s; t_1, t_2)\}$$

oder durch

$$\mathcal{F}(s; t_1, t_2) = |\mathcal{F}(s; t_1, t_2)| e^{i\phi(t_1, t_2)}$$

mit dem *Betrags-* oder *Amplitudenspektrum*

$$|\mathcal{F}(s; t_1, t_2)| = \sqrt{\Re\{\mathcal{F}(s; t_1, t_2)\}^2 + \Im\{\mathcal{F}(s; t_1, t_2)\}^2}$$

und dem *Phasenspektrum*

$$\phi(t_1, t_2) = \arctan \left[\frac{\Im\{\mathcal{F}(s; t_1, t_2)\}}{\Re\{\mathcal{F}(s; t_1, t_2)\}} \right]$$

dargestellt werden kann. Das Quadrat des Amplitudenspektrums bezeichnet man als *Leistungs-* oder *Powerspektrum*

$$P(s; t_1, t_2) = |\mathcal{F}(s; t_1, t_2)|^2.$$

Analog dem eindimensionalen Fall gibt es auch hier ein Ähnlichkeitstheorem: wenn

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

eine invertierbare Matrix ist und $\mathbf{x} = (x, y)$, $\mathbf{t} = (t_1, t_2)$, dann ist die Fouriertransformierte von $s(A\mathbf{x})$ gegeben durch

$$\frac{1}{\det A} \cdot \mathcal{F}(s; B \cdot \mathbf{t}),$$

wobei $B = (A^{-1})^T$. Ein wichtiger Sonderfall ist die Rotation um einen Winkel α . Ist

$$A = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix},$$

dann gilt

$$B = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} = A,$$

und die Fouriertransformierte ist also im Frequenzbereich um den gleichen Winkel α gedreht wie das Bild im Ortsbereich.

Das Fourierbild ist für alle $x, y \in \mathbb{R}$ definiert. In der Nachrichtentechnik geht man meist von bandbeschränkten Funktionen aus, bei denen also nicht beliebig große Frequenzen auftreten. Bezeichnet man mit F_x und F_y die Beträge der betragsmäßig größten auftretenden Frequenzen in x - und y -Richtung, dann heißen $f_{Sx} = 2F_x$ die Abtastfrequenz des Bildes S in x -Richtung, und $f_{Sy} = 2F_y$ die Abtastfrequenz des Bildes S in y -Richtung. Dieses sind nach dem Abtastsatz der Nachrichtentechnik die Frequenzen, mit der das kontinuierliche Bild diskretisiert werden muß, um das kontinuierliche Bild aus den Bildpunktwerten eindeutig rekonstruieren zu können. In der Nachrichtentechnik ist die Annahme der Bandbeschränktheit dadurch gerechtfertigt, daß das Signal (analog) über einen Kanal übertragen wird.

Bei digitalen Bildern kann man eigentlich nicht von bandbeschränkten Funktionen ausgehen, eine Unstetigkeit (Kante) im Bild führt zu einer Fouriertransformierten von nicht mehr kompaktem Träger. Die Fouriertransformierten fallen allerdings ab. Praktisch betrachtet man in diesem Fall oft einen endlichen Bandbereich, in dem die größten Amplituden liegen.

7.3 Eindimensionale diskrete Fouriertransformation (DFT)

Bei der diskreten Fouriertransformation ordnet man einem Vektor $(s(0), s(1), \dots, s(M-1)) \in \mathbb{C}^m$ im "Ortsraum" oder "Zeitraum" einen Vektor $(\hat{s}(0), \hat{s}(1), \dots, \hat{s}(M-1)) \in \mathbb{C}^m$ im "Frequenzraum" zu. Man benutzt dafür die *Teilsommen der Fourierreihen*. Man betrachtet die Zahlen $s(0), s(1), \dots, s(M-1)$ als Werte einer "abgetasteten" Funktion $s : \mathbb{C} \rightarrow \mathbb{C}$ und interpoliert sie durch die Teilsumme einer Fourierreihe. Anschließend ordnet man dieser Teilsumme den diskreten Vektor ihrer Koeffizienten zu.

Satz 7.1 Zu gegebenen reellen Zahlen $s(0), \dots, s(M-1)$ sei

$$\hat{s}(x) = \frac{1}{M} \sum_{m=0}^{M-1} s(m) \omega^{-mx}, \quad \omega = e^{\frac{2\pi i}{M}}.$$

Dann gilt

$$s(m) = \sum_{k=0}^{M-1} \hat{s}(k) \omega^{km}, \quad m = 0, 1, \dots, M-1.$$

Beweis. Der Beweis erfolgt durch Einsetzen:

$$\begin{aligned} s(m) &= \sum_{k=0}^{M-1} \hat{s}(k) \omega^{km}, \quad m = 0, 1, \dots, M-1 \\ &= \sum_{k=0}^{M-1} \left(\frac{1}{M} \sum_{r=0}^{M-1} s(r) \omega^{-rk} \right) \omega^{km} \\ &= \frac{1}{M} \sum_{r=0}^{M-1} s(r) \sum_{k=0}^{M-1} \omega^{k(m-r)}. \end{aligned}$$

Für $r = m$ hat die innere Summe den Wert $\sum_{k=0}^{M-1} \omega^0 = M$, für $r \neq m$ gilt nach der Formel für die Teilsummen der geometrische Reihe

$$1 + \omega^{m-r} + (\omega^{m-r})^2 + \dots + (\omega^{m-r})^{M-1} = \frac{(\omega^{m-r})^M - 1}{\omega^{m-r} - 1} = \frac{(\omega^M)^{m-r} - 1}{\omega^{m-r} - 1}.$$

Da ω M -te Einheitswurzel ist, ist der Zähler $= 0$ und der Nenner $\neq 0$. Damit folgt das Ergebnis. QED

Satz 7.1 ist das diskrete Analogon zur kontinuierlichen Fouriertransformation und wird als Diskrete Fourier-Transformation bezeichnet (DFT, genauer

DFT(M), da ein Vektor der Länge M transformiert wird). Wie man leicht sieht, läßt sich die DFT als Matrix-Vektor-Multiplikation realisieren,

$$\begin{pmatrix} \hat{s}(0) \\ \hat{s}(1) \\ \hat{s}(2) \\ \vdots \\ \hat{s}(M-1) \end{pmatrix} = \frac{1}{M} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \cdots & \omega^{-M+1} \\ 1 & \omega^{-2} & \omega^{-4} & \cdots & \omega^{-2M+2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{-M+1} & \omega^{-2M+2} & \cdots & \omega^{-(M-1)^2} \end{pmatrix} \begin{pmatrix} s(0) \\ s(1) \\ s(2) \\ \vdots \\ s(M-1) \end{pmatrix}. \quad (7.3)$$

Bei naiver Berechnung erfordert die Matrix-Vektor-Multiplikation M^2 Multiplikationen und $M(M-1)$ Additionen.

7.4 Schnelle Fouriertransformation (FFT)

Die spezielle Struktur der Matrix in (7.3) ermöglicht eine wesentlich schnellere Berechnung der DFT. Sei M gerade. Durch Zusammenfassung der Summanden mit geraden und ungeraden Indizes gelangt man zu

$$\begin{aligned} \hat{s}(x) &= \frac{1}{M} \sum_{m=0}^{\frac{M}{2}-1} s(2m) \omega^{-2mx} + \frac{1}{M} \sum_{m=0}^{\frac{M}{2}-1} s(2m+1) \omega^{-(2m+1)x} \\ &= \frac{1}{2} \frac{1}{M/2} \sum_{m=0}^{\frac{M}{2}-1} s(2m) \tilde{\omega}^{-mx} + \frac{\omega^{-x}}{2} \frac{1}{M/2} \sum_{m=0}^{\frac{M}{2}-1} s(2m+1) \tilde{\omega}^{-mx}, \end{aligned} \quad (7.4)$$

wobei $\tilde{\omega} = e^{\frac{2\pi i}{M/2}}$ einen doppelt so großen Winkel wie ω beschreibt. Sei

$$s' = (s(0), s(2), \dots, s(M-2)), \quad s'' = (s(1), s(3), \dots, s(M-1)).$$

Dann bedeutet (7.4)

$$\begin{aligned} 2\hat{s}(j) &= \hat{s}'(j) + \omega^{-j} \hat{s}''(j), & j = 0, \dots, \frac{M}{2} - 1, \\ 2\hat{s}(j + \frac{M}{2}) &= \hat{s}'(j) - \omega^{-j} \hat{s}''(j), & j = 0, \dots, \frac{M}{2} - 1. \end{aligned} \quad (7.5)$$

Die erste Zeile ist klar, und für die zweite Zeile muß zusätzlich nur

$$\omega^{-\frac{M}{2}-j} = \omega^{-\frac{M}{2}} \omega^{-j} = e^{-\pi i} \omega^{-j} = -\omega^{-j}$$

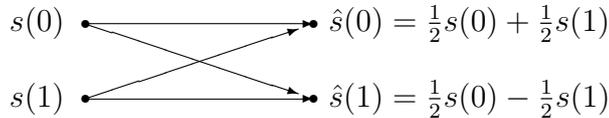
beachtet werden. Berechnet man die DFT über (7.5), dann benötigt die Transformation von M Zahlen die doppelte Anzahl Multiplikationen wie die Transformation von $\frac{M}{2}$ Zahlen, zuzüglich $\frac{M}{2}$ Multiplikationen in (7.5). Für $M = 2^\mu$ benötigt man damit bei rekursiver Anwendung $\frac{1}{2}M\mu$ Multiplikationen. Dies beweist man per vollständiger Induktion über μ . Für $\mu = 0$ ist die Aussage trivialerweise richtig, und für $\mu + 1$ ist die Anzahl der Multiplikationen nach der obigen Regel

$$2\left(\frac{1}{2}2^\mu\mu\right) + 2^\mu = \frac{1}{2}2^{\mu+1}\mu + \frac{1}{2}2^{\mu+1} = \frac{1}{2}2^{\mu+1}(\mu + 1).$$

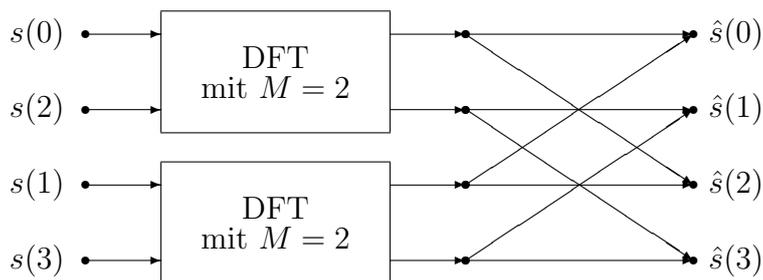
Diese Berechnung der DFT ist also schneller als $M^2 = 2^{2\mu-2}$. Sie heißt *Fast Fourier Transform* (FFT). Die FFT besteht aus drei Grundschritten: Aufteilen von s in s' und s'' , durchführen von zwei DFT($\frac{M}{2}$) und Berechnung von $\hat{s}(0), \dots, \hat{s}(M-1)$ aus (7.5).

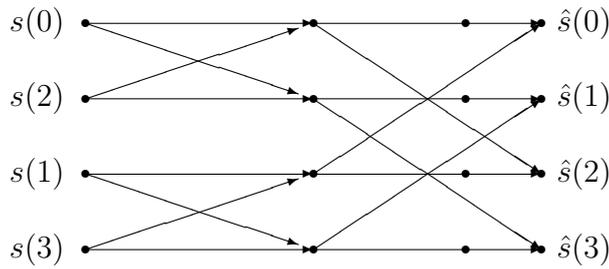
Für die Implementierung ist es nützlich, sich die Verknüpfung der Zahlen in einem Graphen zu veranschaulichen:

$M = 2$



$M = 4$





$M = 8$ Übungsaufgabe

7.5 Zweidimensionale diskrete Fouriertransformation

Für die zweidimensionale DFT kann zunächst in x -Richtung und dann in y -Richtung transformiert werden, oder umgekehrt. Damit ergeben sich die Transformationsregeln unmittelbar aus dem eindimensionalen Fall, es muß nur beachtet werden, daß die Bildbreite L und die Bildhöhe R unterschiedlich groß sein können,

$$\hat{s}(u, v) = \frac{1}{LR} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y) \omega_L^{-xu} \omega_R^{-vy}$$

mit $\omega_L = e^{\frac{2\pi i}{L}}$ und $\omega_R = e^{\frac{2\pi i}{R}}$, sowie

$$s(x, y) = \sum_{u=0}^{L-1} \sum_{v=0}^{R-1} \hat{s}(u, v) \omega_L^{ux} \omega_R^{vy}.$$

Man bezeichnet diese Transformationen auch mit $\text{DFT}(L \times R)$ bzw. inverser $\text{DFT}(L \times R)$, und man schreibt

$$S \stackrel{\text{DFT}}{\longleftrightarrow} \hat{S}.$$

Für die Matrix-Darstellung der DFT definieren wir

$$F_L = (\omega_L^{-xu})_{u,x=0}^{L-1}, \quad F_R = (\omega_R^{-yv})_{y,v=0}^{R-1}.$$

Die DFT ist damit gegeben durch

$$\hat{S} = \frac{1}{LR} F_L S F_R. \quad (7.6)$$

Es gilt $F_L^{-1} = (\omega_L^{xu})_{u,x=0}^{L-1}$, $F_R^{-1} = (\omega_R^{yv})_{v,y=0}^{R-1}$ (Beweis: Übungsaufgabe !). Damit folgt die Darstellung für die inverse DFT,

$$S = (LR)F_L^{-1}\hat{S}F_R^{-1}.$$

Eine naive Berechnung der DFT($L \times R$) per Matrixmultiplikation erfordert L^2R^2 Multiplikationen. Die DFT läßt sich aber darstellen als

$$\hat{s}(u, v) = \frac{1}{L} \sum_{x=0}^{L-1} \tilde{s}(x, v) \omega_L^{-xu}, \quad \tilde{s}(x, v) = \frac{1}{R} \sum_{y=0}^{R-1} s(x, y) \omega_R^{-yv}.$$

Eine zweidimensionale DFT läßt sich also über L eindimensionale DFT(R) und anschliessend eine eindimensionale DFT(L) berechnen (Zeilen-Spalten-Verfahren). Das Zeilen-Spalten-Verfahren läßt sich mit der FFT beschleunigen, der Aufwand ist dann $\mathcal{O}(N^2 \log N)$, wenn $L = R = N$ eine Zweierpotenz ist.

Im folgenden geben wir einige Eigenschaften der zweidimensionalen DFT an.

- **Linearität.** $S + T \xleftrightarrow{\text{DFT}} \hat{S} + \hat{T}$, $\alpha S \xleftrightarrow{\text{DFT}} \alpha \hat{S}$ ($\alpha \in \mathbb{C}$)

- **Symmetrie.**

$$(s(-x, -y))_{x=0, y=0}^{L-1, R-1} \xleftrightarrow{\text{DFT}} (\hat{s}(-u, -v))_{u=0, v=0}^{L-1, R-1},$$

wobei

$$s(-x, -y) = \begin{cases} s(0, 0), & x = 0, y = 0, \\ s(0, R - y), & x = 0, y \in \{1, \dots, R - 1\}, \\ s(L - x, 0), & x \in \{1, \dots, L - 1\}, y = 0, \\ s(L - x, R - y), & x \neq 0 \text{ und } y \neq 0. \end{cases}$$

Beachte, daß S als doppelt periodisch vorausgesetzt wurde, $s(x+L, y+R) = s(x, y)$.

- **Verschiebung.** Für feste $j, k \in \mathbb{Z}$ gilt

$$(s(x - j, y - k))_{x=0, y=0}^{L-1, R-1} \xleftrightarrow{\text{DFT}} (\hat{s}(u, v) \omega_L^{-ju} \omega_R^{-kv})_{u=0, v=0}^{L-1, R-1}.$$

Beweis als Übungsaufgabe !

- **Zyklische Faltung.** Für zwei $L \times R$ -Matrizen $S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$ und $T = (t(x, y))_{x=0, y=0}^{L-1, R-1}$ definieren wir die zyklische Faltung

$$Q := S * T = (q(x, y))_{x=0, y=0}^{L-1, R-1}, \quad q(x, y) = \sum_{j=0}^{L-1} \sum_{k=0}^{R-1} s(x - j, y - k) t(j, k).$$

Für negative Indizes ist dabei die obige Definition zu nehmen. Bei dieser direkten Berechnung der Faltung werden $LR \cdot LR$ Multiplikationen und $LR \cdot (LR - 1)$ Additionen benötigt. Beispiele für Faltungen wurden im letzten Kapitel gegeben, insbesondere der Gauß-Filter mit der Maske

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} t(-1, -1) & t(-1, 0) & t(-1, 1) \\ t(0, -1) & t(0, 0) & t(0, 1) \\ t(1, -1) & t(1, 0) & t(1, 1) \end{bmatrix}.$$

Mit der Matrix

$$T = \begin{bmatrix} 4 & 2 & 0 & \cdots & 0 & 2 \\ 2 & 1 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 2 & 1 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

bedeutet Gauß-Filterung die Faltung $S * T$. In der Maske des Gauß-Filters sind nur die Elemente $\neq 0$ in der Umgebung von (x, y) aufgeführt.

Die Faltungseigenschaft der DFT ist

$$S * T \xrightarrow{\text{DFT}} (LR)(\hat{S}\hat{T}) = (LR \hat{s}(u, v) \hat{t}(u, v))_{u=0, v=0}^{L-1, R-1}.$$

Mit $\hat{S}\hat{T}$ ist also hier die *elementweise Multiplikation* gemeint. Beweis der Faltungseigenschaft ist Übungsaufgabe !

Wir definieren auch im diskreten Fall ein (diskretes) *Leistungsspektrum* oder *Powerspektrum*

$$P = (p(u, v))_{u, v=0}^{L-1, R-1}, \quad p(u, v) := |\hat{s}(u, v)|^2 = (\Re \hat{s}(u, v))^2 + (\Im \hat{s}(u, v))^2.$$

Das Leistungsspektrum ist verschiebungsunabhängig (Beweis: Übungsaufgabe !).

7.6 Digitale Filterung mittels Faltung und DFT

Eine Anwendung der DFT ist die digitale Filterung mittels Faltungsoperatoren, wie im letzten Kapitel behandelt. Ist die Matrix T voll besetzt, dann erfordert eine

direkte Berechnung der Faltung, wie oben angegeben, $(LR)^2$ Multiplikationen und $LR(LR-1)$ Additionen. Mit der Darstellung (7.6) lautet die Faltungseigenschaft

$$\frac{1}{LR}F_L(S * T)F_R = LR \left[\left(\frac{1}{LR}F_L S F_R \right) \cdot \left(\frac{1}{LR}F_L T F_R \right) \right],$$

und daraus folgt

$$S * T = (LR)^2 F_L^{-1} \left[\left(\frac{1}{LR}F_L S F_R \right) \cdot \left(\frac{1}{LR}F_L T F_R \right) \right] F_R^{-1}.$$

In den eckigen Klammern werden die Matrizen wieder elementweise multipliziert. Es sind insgesamt drei DFTen und zusätzlich LR Multiplikationen durchzuführen. Dies ist insbesondere bei großen Bildern ein deutlich geringerer numerischer Aufwand als bei einer direkten Durchführung der Faltung, wenn die Filtermatrix voll besetzt ist.

7.7 Eine Basisdarstellung für die zweidimensionale DFT

Bei der eindimensionalen DFT sind die Basisfunktionen als Frequenzen interpretierbar. In diesem Abschnitt wird die zweidimensionale DFT als Basistransformation aufgefasst, und die zugehörigen Basismatrizen werden ermittelt.

Sei $\mathbb{C}^{L \times R}$ die Menge aller (L, R) -Matrizen mit komplexen Einträgen. Für zwei solche Matrizen $S = (s(x, y))_{x=0, y=0}^{L-1, R-1}$ und $T = (t(x, y))_{x=0, y=0}^{L-1, R-1}$ definieren wir das Skalarprodukt

$$\langle S, T \rangle = \frac{1}{M} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y) \overline{t(x, y)}, \quad M = L \cdot R.$$

Die Größe $\langle S, S \rangle$ bezeichnet man als Energie der Matrix S . Wir betrachten linear unabhängige Matrizen $B_{u,v} = (b_{u,v}(x, y))$, $u = 0, \dots, L-1$, $v = 0, \dots, R-1$, die eine Basis von $\mathbb{C}^{L \times R}$ bilden. Diese Basis heißt orthonormal, falls

$$\langle B_{u,v}, B_{u',v'} \rangle = \begin{cases} 1, & \text{falls } u = u', v = v', \\ 0, & \text{sonst.} \end{cases}$$

Satz 7.2 *Es sei $B_{u,v}$, $u = 0, \dots, L-1$, $v = 0, \dots, R-1$ eine orthogonale Basis von $\mathbb{C}^{L \times R}$. Für jede komplexe $L \times R$ -Matrix $S = (s(x, y))$ gilt*

$$S = \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} \alpha_{x,y} B_{x,y}, \quad \alpha_{x,y} = \langle S, B_{x,y} \rangle.$$

Beweis. Zu zeigen ist die Darstellung der $\alpha_{x,y}$. Sei u, v fest. Es gilt

$$\langle S, B_{u,v} \rangle = \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} \alpha_{x,y} \langle B_{x,y}, B_{u,v} \rangle = \alpha_{u,v}$$

wegen der Orthonormalität. QED.

Ein naheliegendes Beispiel ist die Indikatormatrix $B_{u,v} = (b_{u,v}(x, y))$,

$$b_{u,v}(x, y) = \begin{cases} 1, & u = x \text{ und } v = y, \\ 0, & \text{sonst.} \end{cases}$$

Satz 7.3 Für die DFT gilt $\hat{s}(u, v) = \langle S, B_{u,v} \rangle$,

$$B_{u,v} = (b_{u,v}(x, y)), \quad b_{u,v}(x, y) = e^{2\pi i(xu/L + yv/R)},$$

also

$$B_{u,v} = \begin{bmatrix} 1 \\ \omega_L^u \\ \vdots \\ \omega_L^{(L-1)u} \end{bmatrix} \begin{bmatrix} 1 & \omega_R^v & \cdots & \omega_R^{(R-1)v} \end{bmatrix}.$$

Diese Matrizen bilden eine orthonormale Basis.

Beweis. Es gilt

$$\begin{aligned} \langle S, B_{u,v} \rangle &= \frac{1}{M} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y) \overline{B_{u,v}(x, y)} \\ &= \frac{1}{M} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y) e^{-2\pi i \frac{xu}{L}} \cdot e^{-2\pi i \frac{yv}{R}} \\ &= \frac{1}{M} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} s(x, y) \omega_L^{-xu} \cdot \omega_R^{-yv} \\ &= \hat{s}(u, v) \end{aligned} \tag{7.7}$$

Zum Beweis der Orthonormalität zeigen wir zunächst

$$\sum_{x=0}^{L-1} e^{2\pi i x u/L} = \begin{cases} L, & L|u, \\ 0, & L \nmid u. \end{cases}$$

Es gelte $L|u$, dann folgt $g := xu/L \in \mathbb{Z}$, $e^{2\pi i g} = 1$ und $\sum_{x=0}^{L-1} 1 = L$. Für $L \nmid u$ folgt $z := e^{2\pi i u/L} \neq 1$, also $\sum_{x=0}^{L-1} z^x = 1 + z + z^2 + \cdots + z^{L-1} = \frac{z^L - 1}{z - 1}$, und $z^L = (e^{2\pi i u/L})^L = e^{2\pi i u} = 1$. Somit gilt $\sum_{x=0}^{L-1} z^x = \frac{1-1}{z-1} = 0$.

Im zweiten Schritt berechnen wir

$$\begin{aligned}
& \langle B_{u,v}, B_{u',v'} \rangle \\
&= \frac{1}{M} \sum_{x=0}^{L-1} \sum_{y=0}^{R-1} e^{2\pi i x(u-u')/L} \cdot e^{2\pi i y(v-v')/R} \\
&= \frac{1}{M} \left(\sum_{x=0}^{L-1} e^{2\pi i x(u-u')/L} \right) \cdot \left(\sum_{y=0}^{R-1} e^{2\pi i y(v-v')/R} \right) \\
&= \begin{cases} \frac{1}{M} \cdot L \cdot R = 1, & \text{falls } L|(u-u') \iff u-u'=0 \\ & \text{und } R|(v-v') \iff v-v'=0, \\ 0, & \text{sonst.} \end{cases}
\end{aligned}$$

QED

Die Basismatrizen ergeben in diesem Fall also tatsächlich die DFT. Die Matrizen $B_{u,v}$ sind "Basisbilder" und spielen die Rolle von Frequenzen. Insbesondere ist

$$B_{0,0} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix},$$

und

$$\hat{s}(0,0) = m_S$$

der mittlere Grauwert der Matrix S .

7.8 Diskrete Cosinustransformation (DCT)

Die DFT basierte auf der Interpolation der Abtastwerte mit den Basisfunktionen ω^{-k} . Neben der DFT gibt es noch weitere Transformationen, die auf anderen Interpolationsmethoden basieren. Wir behandeln die Diskrete Cosinus-Transformation (DCT), die von großer praktischer Bedeutung ist. Auf der DCT basieren der JPEG (Joint Picture Expert Group) Standard für die Standbildkompression und der MPEG (Motion Picture Expert Group) Standard für Bewegtbildkompression (MPEG2 wird bei DVD's verwendet).

Satz 7.4 *Zu gegebenen reellen Zahlen $s(0), \dots, s(N-1)$ sei*

$$\hat{s}(u) = \sqrt{\frac{2}{N}} \varepsilon(u) \sum_{x=0}^{N-1} s(x) \cos\left(\frac{(2x+1)u\pi}{2N}\right), \quad \varepsilon(u) = \begin{cases} \frac{1}{\sqrt{2}}, & u=0, \\ 1, & u=1, \dots, N-1. \end{cases}$$

Dann gilt

$$s(x) = \sqrt{\frac{2}{N}} \sum_{u=0}^{N-1} \varepsilon(u) \hat{s}(u) \cos\left(\frac{(2x+1)u\pi}{2N}\right), \quad x = 0, 1, \dots, N-1.$$

Die Transformation in Satz 7.4 heißt *Diskrete Cosinus-Transformation* (*DCT*, genauer *DCT(N)*). Definiert man die *DCT(N)*-Matrix C_N durch

$$C_N = \sqrt{\frac{2}{N}} \left[\varepsilon(u) \cos \frac{(2x+1)u\pi}{2N} \right]_{u,x=0}^{N-1},$$

dann ist die *DCT(N)*-Transformation gegeben durch

$$\hat{s} = C_N s,$$

wobei $\hat{s} = (\hat{s}(0), \dots, \hat{s}(N-1))$ und $s = (s(0), \dots, s(N-1))$. Die Rücktransformation ist dann gegeben durch

$$s = C_N^{-1} \hat{s}.$$

Satz 7.4 ist also bewiesen, wenn wir den folgenden Satz bewiesen haben.

Satz 7.5 C_N ist eine orthogonale $N \times N$ -Matrix, d.h.

$$C_N^{-1} = C_N^T.$$

Beweis. Sei $(e(x, y))_{x,y=0}^{N-1} = C_N^T C_N$. Zu zeigen ist

$$e(x, y) = \begin{cases} 1, & x = y, \\ 0, & x \neq y. \end{cases}$$

Ausmultiplizieren ergibt

$$e(x, y) = \frac{2}{N} \left(\frac{1}{2} + \sum_{u=1}^{N-1} \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)u\pi}{2N} \right), \quad x, y = 0, \dots, N-1.$$

Wegen

$$2 \cos \alpha \cos \beta = \cos(\alpha + \beta) + \cos(\alpha - \beta)$$

folgt

$$e(x, y) = \frac{1}{N} \left(1 + \sum_{u=1}^{N-1} \cos \frac{(x+y+1)u\pi}{N} + \sum_{u=1}^{N-1} \cos \frac{(x-y)u\pi}{N} \right).$$

Nun gilt aber

$$\sum_{u=1}^{N-1} \cos \frac{xu\pi}{N} = \begin{cases} N-1, & x=0, \\ 0, & x=1, 3, \dots, 2N-1 \\ -1, & x=2, 4, \dots, 2N-2. \end{cases} \quad (7.8)$$

Damit folgt für $x = y$

$$e(x, y) = \frac{1}{N} \left(1 + \underbrace{\sum_{u=1}^{N-1} \cos \frac{(2x+1)u\pi}{N}}_{=0 \text{ nach (7.8)}} + \sum_{u=1}^{N-1} 1 \right) = \frac{1}{N} (1 + N - 1) = 1.$$

Für $x \neq y$ gilt

$$e(x, y) = \frac{1}{N} (1 + 0 - 1) = 0.$$

Zum Beweis von (7.8) betrachte man $\omega = \omega_{2N} = e^{-2\pi i/2N} = \cos \frac{\pi}{N} - i \sin \frac{\pi}{N}$. Dann gilt

$$\begin{aligned} \omega^{2N} &= 1, & \omega^x &\neq 1, & x &= 1, \dots, 2N-1, \\ \omega^{ux} &= \cos \frac{ux\pi}{N} - i \sin \frac{ux\pi}{N}, \\ \omega^{-ux} &= \cos \frac{ux\pi}{N} + i \sin \frac{ux\pi}{N}, \end{aligned}$$

und somit

$$\omega^{ux} + \omega^{-ux} = 2 \cos \frac{ux\pi}{N}, \quad x, u = 0, 1, \dots$$

Wir zeigen

$$\sum_{u=0}^{2N-1} \omega^{xu} = \begin{cases} 2N, & x=0 \\ 0, & x=1, \dots, 2N-1. \end{cases} \quad (7.9)$$

Für $x = 0$ ist (7.9) klar. Für $x = 1, \dots, 2N-1$ gilt

$$\sum_{u=0}^{2N-1} (\omega^x)^u = \frac{(\omega^x)^{2N} - 1}{\underbrace{\omega^x - 1}_{\neq 0}} = \frac{1 - 1}{\omega^x - 1} = 0.$$

Umformen von (7.9) liefert (7.8),

$$\sum_{u=0}^{2N-1} \omega^{xu} = 1 + \sum_{u=1}^{N-1} \omega^{xu} + \underbrace{(\omega^N)^x}_{=(-1)^x} + \sum_{u=1}^{N-1} \underbrace{\omega^{(2N-u)x}}_{(\omega^{2N})^x \omega^{-ux}}$$

$$\begin{aligned}
&= 1 + (-1)^x + \sum_{u=1}^{N-1} (\omega^{xu} + \omega^{-xu}) \\
&= 1 + (-1)^x + 2 \sum_{u=1}^{N-1} \cos \frac{ux\pi}{N} \\
&= \begin{cases} 2N, & x = 0, \\ 0, & x = 1, \dots, 2N - 1, \end{cases}
\end{aligned}$$

woraus die Behauptung folgt. QED

Die Definition der DCT zeigt, welche Koeffizienten zu hochfrequenten und welche Koeffizienten zu niedrigfrequenten Schwingungen gehören. Durch Weglassen der hochfrequenten Anteile läßt sich mit der DCT eine Datenreduktion erreichen. Oft ist dies ohne wesentlichen Informationsverlust möglich, etwa wenn der hochfrequente Anteil zum großen Teil vom Rauschen verursacht wird.

Es existieren schnelle Algorithmen, die eine $DCT(N)$ mit $\mathcal{O}(N \log N)$ Rechenoperationen durchführen (wenn N eine Zweierpotenz ist).

7.9 Anwendung der DCT auf Bilddatenkompression (JPEG)

Eine zweidimensionale DCT ist wie bei der DFT durch Zeilen-Spalten-Separation möglich. In der Praxis wird das Bild in *Blöcke* zerlegt (JPEG: 8×8) und die Transformation wird blockweise durchgeführt. Die Blöcke können parallel bearbeitet werden. Sei $S = (s(x, y))_{x, y=0}^7$ ein 8×8 Teilbild mit ganzzahligen Einträgen,

$$s(x, y) \in \{0, \dots, 2^p - 1\} \quad (\text{z.B. } p = 8).$$

Eine $DCT(8 \times 8)$ liefert $\hat{S} = (\hat{s}(u, v))_{u, v=0}^7$ mit

$$\hat{s}(u, v) = \frac{1}{4} \varepsilon(u) \varepsilon(v) \sum_{x=0}^7 \sum_{y=0}^7 s(x, y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}.$$

Die inverse $DCT(8 \times 8)$ lautet

$$s(x, y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 \varepsilon(u) \varepsilon(v) \hat{s}(u, v) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}.$$

Jedes Wort $\hat{s}(x, y)$ trägt Information über den gesamten 8×8 -Block. Der Koeffizient $\hat{s}(0, 0)$ ist bis auf einen Faktor das arithmetische Mittel aller $s(x, y)$, also der

mittlere Grauwert in dem 8×8 -Block. Der Informationsgehalt der Größen $\hat{s}(u, v)$ nimmt von links oben nach rechts unten ab. In dieser Richtung werden die Werte im allgemeinen kleiner. Mit der inversen *DCT* lassen sich die Größen $s(x, y)$ wiedergewinnen. Kleine Änderungen von $\hat{s}(u, v)$ bewirken kleine Änderungen von $s(x, y)$ (die *DCT* ist eine *Isometrie*).

Zu einer Kompression kommt man durch Quantisierung im transformierten Bild \hat{S} , bei der kleine oder unwichtige Größen $\hat{s}(u, v)$ zu 0 gesetzt werden,

$$\tilde{s}(u, v) = \begin{cases} 0, & \text{falls } |\hat{s}(u, v)| \leq \tau, \\ \hat{s}(u, v), & \text{falls } |\hat{s}(u, v)| > \tau. \end{cases}$$

Die Konstante τ heißt Schwellwert.

Bei zu großer Kompression ist die Blockstruktur im Bild deutlich sichtbar, dies ist ein Nachteil des Verfahrens. Ein Vorteil ist dagegen, dass sich häufig ein hoher Kompressionsgrad erzielen läßt. Der Kompressionsgrad ist definiert als

$$\text{Kompressionsgrad} = \frac{\text{ursprüngliche Datenmenge}}{\text{komprimierte Datenmenge}}.$$

Kapitel 8

Wavelet-Transformationen

Ein Nachteil der kontinuierlichen und der diskreten Fouriertransformationen ist, daß für die Berechnung der Transformierten das gesamte Signal herangezogen wird. Es liegt nach der Transformation eine exakte Frequenzinformation vor, aber es ist keine Aussage mehr möglich über den Ort (die Zeit), an dem (zu der) diese Frequenz aufgetreten ist. Dies ist nicht immer ein Nachteil, z.B. ist das Rauschen in der Regel gleichmäßig über ein Signal bzw. über ein Bild verteilt. Für viele Anwendungen ist aber die Zeit- oder Ortslokalisierung der Frequenzanteile entscheidend. Dazu gehört die Kompression. Wie im vorangegangenen Abschnitt gesehen, basiert der Erfolg des JPEG-Algorithmus für die Kompression auf einer Einteilung des Bildes in Unterblöcke, die getrennt transformiert werden. Für die so berechneten Frequenzen ist also ihr Ort bekannt. Dies motiviert die weitere, systematische Beschäftigung mit dem Thema Zeit-Frequenz-Lokalisation und führt uns in diesem Kapitel zu orthogonalen Multiresolutionsanalysen (Multiskalenanalysen) und schließlich auf die Wavelets.

8.1 Elemente der Zeit-Frequenzanalyse

Wir gehen in diesem Abschnitt von der eindimensionalen kontinuierlichen Fouriertransformation aus (siehe Abschnitt 7.1),

$$\mathcal{F}(s, x) = \int_{\mathbb{R}} s(t) e^{-i2\pi xt} dt, \quad x \in \mathbb{R},$$

mit der Rücktransformation

$$s(t) = \int_{\mathbb{R}} \mathcal{F}(s, x) e^{i2\pi xt} dx.$$

Für $s_1, s_2 \in L_2(\mathbb{R})$ führen wir das innere Produkt

$$\langle s_1, s_2 \rangle = \int_{\mathbb{R}} s_1(t) \overline{s_2(t)} dt$$

ein. Ein wichtiges Mittel zur Zeit-Frequenz-Lokalisation ist die *gefensterte Fourier-Transformation*

$$\mathcal{F}^{Fenster}(s; x, \tau) = \int_{\mathbb{R}} s(t) g(t - \tau) e^{-i2\pi xt} dt.$$

Die Funktion g heißt dabei Fensterfunktion, sie ist typischerweise um den Nullpunkt herum betragsmäßig groß und fällt für betragsmäßig große Argumente stark ab. Diese Erweiterung der kontinuierlichen Fouriertransformation kann als Messung der Amplituden der Frequenzkomponenten von s mit der Frequenz x in der Umgebung des Zeitpunkts oder Ortes τ interpretiert werden. Häufig verwendete Fensterfunktionen sind

- $g_\alpha(x) = \begin{cases} 1 + \cos(\pi\alpha x) & \text{für } -\frac{1}{\alpha} \leq x \leq \frac{1}{\alpha}, \\ 0 & \text{sonst.} \end{cases}$
- $g_\alpha(x) = \frac{1}{2\sqrt{\pi\alpha}} e^{-\frac{x^2}{4\alpha}}, x \in \mathbb{R},$

mit verschiedenen $\alpha > 0$. Das letzte Beispiel führt zur *Gabor-Transformation*.

Sei $g_{x,\tau}(t) = g(t - \tau) e^{i2\pi xt}$, und sei $G_{x,\tau}$ die Fouriertransformierte von $g_{x,\tau}$. Mit der *Parsevalschen Gleichung*

$$\langle s_1, s_2 \rangle = \langle \mathcal{F}(s_1), \mathcal{F}(s_2) \rangle$$

folgt

$$\mathcal{F}^{Fenster}(s; x_0, \tau_0) = \int_{\mathbb{R}} s(t) \overline{g_{x_0, \tau_0}(t)} dt = \int_{\mathbb{R}} \mathcal{F}_s(x) \overline{G_{x_0, \tau_0}(x)} dx. \quad (8.1)$$

Die Fouriertransformierte G_{x_0, τ_0} ist dabei von der Form $G_{x_0, \tau_0}(x) = G(x - x_0) e^{i2\pi\tau_0(x - x_0)}$, wobei G die Fouriertransformierte der Funktion g ist. Daraus folgt, daß die Berechnung der Fenster-Fouriertransformierten auch als Filterung mit Bandpässen verschiedener Mittenfrequenzen, aber gleicher Hüllkurve $G(x)$ angesehen werden kann. Dies bedeutet insbesondere, daß die Fensterung für alle Frequenzbereiche mit derselben Breite durchgeführt wird.

8.2 Die kontinuierliche Wavelet-Transformation

Die eindimensionale kontinuierliche Wavelettransformation zerlegt ein Signal $s \in L^2(\mathbb{R})$ in aufeinanderfolgende Frequenzbänder. Sie verwendet zur Transformation keine Sinus- oder Cosinus-Schwingungen wie die Fouriertransformation, sondern eine Familie von orthogonalen Basen, die durch Translation um b und Skalierung durch a einer einzigen Funktion $\psi \in L^2(\mathbb{R})$, dem *mother wavelet* oder schlicht *wavelet*, generiert wird,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right).$$

Eine oft benutzte Funktion ist die *Mexican-Hat-Funktion*

$$\psi(t) = (1 - t^2) e^{-\frac{t^2}{2}}.$$

Die kontinuierliche Wavelet-Transformation ist definiert durch

$$\mathcal{F}^W(a, b) = |a|^{-1/2} \int_{\mathbb{R}} s(t) \psi\left(\frac{t-b}{a}\right) dt, \quad a, b \in \mathbb{R}, \quad a \neq 0.$$

Für eine bestimmte Position b beschreibt $\mathcal{F}^W(a, b)$ die Details von s mit der Auflösungsstufe a .

Die Eigenschaften der Wavelet-Transformierten hängen vom verwendeten Wavelet ψ ab. Eine gängige Bedingung an die Fouriertransformierte Ψ von ψ ist

$$0 < C_\psi = \int_0^\infty \frac{|\Psi(x)|^2}{x} dx < \infty.$$

Dies ist auch eine notwendige Bedingung für die Existenz der inversen Wavelet-Transformation. Wenn diese existiert, dann lautet sie

$$\frac{1}{C_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{F}^W(a, b) \frac{1}{a^2} \psi_{a,b}(t) da db.$$

Diese Funktion ist unter bestimmten Voraussetzungen an die Funktion s punktweise mit s identisch (siehe [3]).

8.3 Die orthogonale Multiresolutionsanalyse (Multiskalenanalyse)

Die kontinuierliche Wavelettransformation einer Funktion s ist eine redundante Darstellung von s . Analog den Fourierreihen betrachtet man Reihenentwicklungen in Wavelets, die unter bestimmten Bedingungen die Funktion s darstellen. Solche Bedingungen werden hier nicht diskutiert, siehe dazu [3].

Es werden zwei verschiedene Basissysteme benötigt: ein Basissystem für die Projektion in die Auflösungsstufen und ein Basissystem für die Detailinformationen, d.h. für die Differenzen zwischen Projektionen von benachbarten Auflösungen. Wir betrachten hier nur solche Basisfunktionen, die paarweise orthogonal zueinander sind bezgl. des Standardproduktes in $L^2(\mathbb{R})$. Die Projektionen in die Auflösungsstufe bezeichnen wir mit P_m , wobei $m = 0$ eine gegebene Ausgangsauflösung bezeichnet. Dann bezeichnet P_{-1} die nächstgrößere Auflösung, P_1 die nächstfeinere. Die Differenzinformation (Detailinformation) bezeichnen wir mit Q_m , so daß

$$\begin{aligned} P_1 s &= P_0 s + Q_0 s \\ P_2 s &= P_1 s + Q_1 s \\ P_3 s &= P_2 s + Q_2 s \quad \text{usw.} \end{aligned}$$

Diese Idee ist mit der Differenzenbildung verwandt, die wir in Abschnitt 3.4 kennengelernt haben. Sie taucht in vielen Teilgebieten der angewandten Mathematik und Informatik auf, z.B. auch in der Computergraphik.

Definition 8.1 *Eine orthogonale Multiresolutionsanalyse (Multiskalenanalyse) ist eine aufsteigende Folge von abgeschlossenen Unterräumen $V_m \in L^2(\mathbb{R})$ und Projektionen $P_m : L^2(\mathbb{R}) \rightarrow V_m$, die die folgenden Bedingungen erfüllen:*

1. $\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots$

- 2.

$$\begin{aligned} \lim_{m \rightarrow -\infty} \|P_m s\|_2 &= 0, & \bigcap_{m \in \mathbb{Z}} V_m &= \{0\}, \\ \lim_{m \rightarrow \infty} \|P_m s - s\|_2 &= 0, & \overline{\bigcup_{m \in \mathbb{Z}} V_m} &= L^2(\mathbb{R}) \end{aligned}$$

3. *Skalierungseigenschaft:*

$$f \in V_m \iff f(\cdot/2) \in V_{m-1}$$

4. *Translationseigenschaft:*

$$f \in V_0 \Rightarrow f(\cdot - n) \in V_0, \quad \forall n \in \mathbb{Z}$$

5. Es existiert eine Funktion $\phi \in V_0$ mit $\int_{\mathbb{R}} \phi(x) dx \neq 0$, so daß die Menge $\phi_{0,n} := \{\phi(x-n) | n \in \mathbb{Z}\}$ eine orthonormale Basis von V_0 bilden. Die Funktion ϕ heißt (orthogonale) Skalierungsfunktion oder auch Generator der orthogonalen Multiresolutionsanalyse.

Bemerkungen.

1. Der Unterraum $W_m = V_{m+1} \setminus V_m$ enthält die Detailinformation zwischen den Räumen V_m und V_{m+1} . Die Abbildungen $Q_m = P_{m+1} - P_m$ bilden von $L^2(\mathbb{R})$ in W_m ab.
2. Es existiert eine Funktion $\psi \in W_0$, so daß die Menge der Translate $\{\psi_{0,n} := \psi(x-n) | n \in \mathbb{Z}\}$ eine orthonormale Basis für W_0 bildet. Diese Funktion heißt (mother-) wavelet.

In vielen Anwendungen wird keine explizite Form der Skalierungsfunktionen und Wavelets benötigt. Da $\phi \in V_0 \subset V_1$, existiert eine Reihe der Form

$$\phi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \phi(2x - k), \quad h_k = \langle \phi, \phi_{1,k} \rangle,$$

mit Koeffizienten $(h_k)_{k \in \mathbb{Z}} \in l^2$ (dem Raum der quadratisch summierbaren biinfinite Folgen), wobei $\phi_{m,n}(x) = 2^{-m/2} \phi(2^{-m}x - n)$. Diese fundamentale Gleichung wird *Zwei-Skalen-Gleichung*, *Verfeinerungsgleichung* oder *Dilatationsgleichung* genannt. Mit der üblichen Normierung

$$\int_{\mathbb{R}} \phi(x) dx = 1$$

ist eine Skalierungsfunktion durch die Koeffizienten h_k eindeutig festgelegt. Für die Wavelets ergeben sich die Koeffizienten g_k in der Darstellung

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \phi(2x - k), \quad g_k = \langle \psi, \phi_{1,k} \rangle,$$

durch den *alternating flip*,

$$g_k = (-1)^k h_{-k+1}.$$

8.4 Mehrdimensionale Multiskalenanalyse

Die Verallgemeinerung der eindimensionalen Multiskalenanalyse ins Höherdimensionale liegt nahe. Eine Multiskalenanalyse des $L^2(\mathbb{R}^n)$ besteht aus einer aufstei-

genden Folge abgeschlossener Unterräume $\{V_m\}_{m \in \mathbb{Z}}$ des $L^2(\mathbb{R}^n)$, so daß

$$\overline{\bigcup_{m \in \mathbb{Z}} V_m} = L^2(\mathbb{R}^n) \quad \text{und} \quad \bigcap_{m \in \mathbb{Z}} V_m = \{0\},$$

und die Projektionen sind analog definiert. Der Unterschied zur eindimensionalen Multiskalenanalyse besteht darin, daß Translationen um $k \in \mathbb{Z}^n$ zugelassen sind und der Übergang von V_m nach V_{m+1} mit Hilfe einer regulären Matrix A , der Dilationsmatrix, beschrieben wird,

$$f(\cdot) \in V_m \iff f(A \cdot) \in V_{m+1}.$$

Der Grundraum V_0 wird durch eine Skalierungsfunktion $\phi \in L^2(\mathbb{R}^n)$ erzeugt. Somit erzeugt

$$\{\phi_{m,k} = |\det A|^{-m/2} \phi(A^m \cdot -k) \mid k \in \mathbb{Z}^n\}$$

den Raum V_m . Die Wahl der Dilatationsmatrix A gibt eine große Freiheit bei der Konstruktion der Wavelets, aber standardmäßig werden zwei Einschränkungen gemacht. Die Dilatationsmatrix A soll in jede Richtung strecken, d.h. die Eigenwerte von A müssen betragsmäßig größer als 1 sein. Darüber hinaus soll A oft nur ganzzahlige Einträge enthalten, denn dann folgt $A\mathbb{Z}^n \subset \mathbb{Z}^n$.

8.5 Die diskrete Wavelettransformation

Wie bei der DFT gehen wir in diesem Abschnitt davon aus, daß keine kontinuierliche Funktion vorliegt, sondern nur *Abtastwerte* – hier allerdings zunächst unendlich viele, $\dots, s(-1), s(0), s(1), \dots$

Wie bei der kontinuierlichen Wavelettransformation werden verschiedene Auflösungsstufen betrachtet. Bei der ersten (“feinsten”) Auflösungsstufe betrachtet man jeden Punkt einzeln, bei den jeweils folgenden werden mehrere Punkte zusammengefasst und gemittelt. Die diskrete Wavelet-Transformation berechnet diese Transformation und auch die Unterschiede zwischen den einzelnen Auflösungen.

Die Grundlage dafür bildet die *schnelle Wavelet-Transformation* (auch *Pyramidalalgorithmus* genannt). Für diesen Algorithmus gehen wir zunächst davon aus, daß nicht Abtastwerte

$$\dots, s(-1), s(0), s(1), \dots$$

vorliegen, sondern die Koeffizienten

$$\dots, c_{-1}^0, c_0^0, c_1^0, \dots$$

einer Funktion $s \in V_0$,

$$s(x) = \sum_{k \in \mathbb{Z}} c_k^0 \phi(x - k).$$

Wie bisher bezeichne ψ das zu ϕ gehörende orthogonale Wavelet. Wir führen die Bezeichnungen

$$\begin{aligned} c_k^m &= \langle s, \phi_{m,k} \rangle, & c^m &= \{c_k^m \mid k \in \mathbb{Z}\} \in l^2, \\ d_k^m &= \langle s, \psi_{m,k} \rangle, & d^m &= \{d_k^m \mid k \in \mathbb{Z}\} \in l^2 \end{aligned}$$

ein. Mit Hilfe der Skalierungsgleichungen folgt

$$\begin{aligned} c_k^m &= \langle s, \phi_{m,k} \rangle = \sum_{l \in \mathbb{Z}} h_l \langle s, \phi_{m-1,2k+l} \rangle = \sum_{l \in \mathbb{Z}} h_{l-2k} c_l^{m-1}, \\ d_k^m &= \langle s, \psi_{m,k} \rangle = \sum_{l \in \mathbb{Z}} g_l \langle s, \phi_{m-1,2k+l} \rangle = \sum_{l \in \mathbb{Z}} g_{l-2k} c_l^{m-1}. \end{aligned}$$

Ausgehend von der Folge c^0 können die Folgen c^1, c^2, \dots berechnet werden, die die Projektionen der Funktion s auf die Wavelets und Skalierungsfunktionen der nächstgrößeren Skalen darstellen.

Zur kürzeren Schreibweise führen wir die Zerlegungsoperatoren H und G ein,

$$\begin{aligned} (Hc)_k &= \sum_{l \in \mathbb{Z}} h_{l-2k} c_l, \\ (Gc)_k &= \sum_{l \in \mathbb{Z}} g_{l-2k} c_l. \end{aligned}$$

Die schnelle Wavelet-Transformation ist dann durch das folgende Schema aus [5, S. 127] gegeben.

Schnelle Wavelet-Transformation

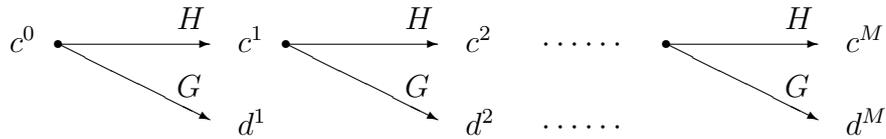
Eingabe: $c^0 = \{c_k | k \in \mathbb{Z}\}$, Zerlegungstiefe M

berechne für $m = 1, \dots, M$

$$c^m = Hc^{m-1}$$

$$d^m = Gc^{m-1}$$

Ausgabe: $c^M; d^m, m = 1, \dots, M$



Aus den Werten $\{c^M, d^m | m = 1, \dots, M\}$ läßt sich die Ausgangsfolge c^0 wieder rekonstruieren. Wir rekonstruieren zunächst c^0 aus c^1 und d^1 . Mit den Skalierungsgleichungen folgt

$$\begin{aligned} \sum_{k \in \mathbb{Z}} c_k^0 \phi_{0,k} &= \sum_{j \in \mathbb{Z}} c_j^1 \phi_{-1,j} + \sum_{j \in \mathbb{Z}} d_j^1 \psi_{-1,j} \\ &= \sum_{j \in \mathbb{Z}} c_j^1 \sum_{l \in \mathbb{Z}} h_l \phi_{0,2j-l} + \sum_{j \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} g_l \phi_{0,2j-l}. \end{aligned}$$

Wir sortieren auf der rechten Seite um und machen einen Koeffizientenvergleich. Es folgt

$$c_k^0 = \sum_{l \in \mathbb{Z}} c_l^1 h_{2l-k} + \sum_{l \in \mathbb{Z}} d_l^1 g_{2l-k}.$$

Analog erhält man c^{M-1} aus d^M und c^M . Definiert man die zu H und G adjungierten¹ Operatoren H^* und G^* durch

$$(H^*c)_k = \sum_{l \in \mathbb{Z}} h_{k-2l} c_l,$$

$$(G^*c)_k = \sum_{l \in \mathbb{Z}} g_{k-2l} c_l,$$

¹Ein Operator H^* heißt zu H adjungiert, wenn $\langle Hc, b \rangle = \langle c, H^*b \rangle$. Hier sind b, c Folgen, und das innere Produkt ist das diskrete Standardprodukt in l^2 . Die adjungierte Abbildung erhält man hier durch die oben erwähnte Umsortierung der Summen.

dann lautet der Rekonstruktionsschritt

$$c^{m-1} = H^* c^m + G^* d^m.$$

Schnelle Wavelet-Rekonstruktion

Eingabe: $M; c^M, d^m, m = 1, \dots, M$

berechne für $m = M, \dots, 1$

$$c^{m-1} = H^* c^m + G^* d^m$$

Ausgabe: c^0

Der Aufwand der schnellen Wavelet-Transformation (und der Rücktransformation) ist $\mathcal{O}(n)$, siehe [5].

Wir gingen anfangs davon aus, daß nicht die Abtastwerte, sondern die Koeffizienten der Waveletentwicklung auf der feinsten Skala gegeben sind. Was ändert sich, wenn tatsächlich “nur” Abtastwerte vorliegen? Dies ist eine Standardsituation in der Praxis. Zwei Alternativen sind denkbar und werden praktisch verwendet.

- Berechnung der Koeffizienten aus den Abtastwerten mittels Quadraturformeln. Dieser Schritt heißt auch *Präfiltern*.
- Direktes Einsetzen der Abtastwerte in die Schnelle Wavelet-Transformation. Bei näherem Hinsehen ist dies eine ungenaue Variante der ersten Alternative. Sie ist einfach zu realisieren und führt oft zu geglätteten Bildern. Strang [9] bezeichnet dieses Vorgehen als *wavelet crime*.

Kapitel 9

Klassifikation

Kapitel 10

Segmentierung

Literaturverzeichnis

- [1] W. Abmayr: *Einführung in die digitale Bildverarbeitung*. B.G. Teubner, Stuttgart, 1994.
- [2] K.R. Castleman: *Digital Image Processing*. Prentice Hall, 1995.
- [3] I. Daubechies: *Ten lectures on wavelets*. SIAM Publ., Philadelphia, 1992.
- [4] Th. Lehmann, W. Oberschelp, E. Pelikan, R. Repges: *Bildverarbeitung für die Medizin*. Springer, 1997.
- [5] A. K. Louis, P. Maaß, A. Rieder: *Wavelets*. B. G. Teubner, 1994.
- [6] O. Mildenerger: *Informationstheorie und Codierung*, Vieweg Verlag, 1992.
- [7] F. Natterer: *The Mathematics of Computerized Tomography*, Wiley, 1986.
- [8] W.K. Pratt: *Digital image processing*. John Wiley and Sons, New York, 1978.
- [9] G. Strang, T. Nguyen: *Wavelets and Filter Banks*. Wellesley Cambridge Press, 1996.
- [10] F. Wahl: *Digitale Bildsignalverarbeitung*. Springer Verlag, Berlin, 1984.
- [11] G. Winkler: *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer, Applications of Mathematics 27, 1991.
- [12] P. Zamperoni: *Methoden der digitalen Bildsignalverarbeitung*. Vieweg, Braunschweig, 1989.