

# Kapitel 1

## Einleitung

Bildverarbeitung ist ein noch relativ junges, aber extrem schnell wachsendes Forschungsgebiet, das heute vor allem durch Digitalfotografie und medizinische Bildgebung vorangetrieben wird. Verwandte Felder, die meist im gleichen Rahmen betrachtet werden, sind auch die Bearbeitung von Videos und Computergraphiken, die durch Digitaleffekte in Filmen bzw. durch die Popularität von Spielekonsolen sehr gefragt sind.

In dieser Vorlesung widmen wir uns vor allem den mathematischen Aspekten der Bildverarbeitung, und versuchen einen kleinen Streifzug durch wichtige mathematische Techniken, die hier Anwendung finden, sowie durch einige wichtige Verarbeitungsmodi (Entzaubern, Entzerren, Segmentierung, Restauration, ...) zu bieten. Wir beginnen mit einer recht allgemeinen Diskussion der Grundaufgaben der Bildverarbeitung.

### 1.1 Entstehung digitaler Bilder

Zur Motivation und zum besseren Verständnis für einige Notwendigkeiten in der Bildverarbeitung befassen wir uns zunächst mit der Frage woher digitale Bilder kommen und wie sie typischerweise aussehen. Viele Arten von Bildern und Bildentstehung sind heute bereits alltäglich:



Abbildung 1.1: Zwei Fotografien einer Szenerie mit verschiedenen Fokalfpunkten<sup>1</sup>

- **Scans analoger Bilder und Filme:** Die Ursprünge der Fotografie lagen in einer analogen Technik, die einfallenden Lichtstrahlen wurden direkt auf den fotosensiblen Film abgebildet und erzeugten die bekannten Negative von Fotos. Durch die Entwicklung des Fotofilms wurden die Negative wieder in Positiva umgekehrt. Eine analoge Technik wurde für Filme (Movies, bewegte Bilder) verwendet, die aus einer schneller Abfolge von Bildern erzeugt wurden. Will man nun solch analoge Bilder auf einem Computer bearbeiten (für Zeitungen, Bücher, Präsentationen, Geschichtsforschung, ...), so muss zuerst ein Scan durchgeführt werden, der das analoge in ein digitales Bild übersetzt. Dabei wird das analoge Bild schichtweise beleuchtet und aus der Lichtintensität werden Grau- bzw. Farbwerte des Bildes erzeugt.

Durch Auflösungsbegrenzung des Scanners erhält man zumeist einen erheblichen Qualitätsverlust, zusätzlich auch Artefakte durch die schichtweise Bearbeitung. Vor allem ältere analoge Bilder sind oft verblasst oder verunreinigt. Damit ergeben sich bereits erste wichtige Aufgaben für die Bildverarbeitung, denn man möchte all diese Fehler am Computer möglichst korrigieren.

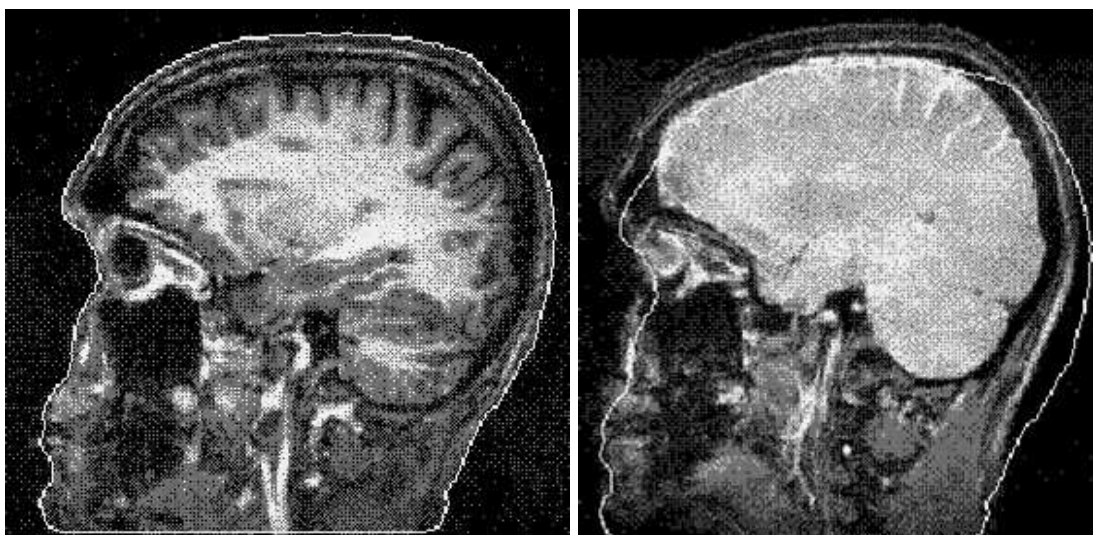


Abbildung 1.2: Nuklearmagnetresonanz (MRI) Bilder mit verschiedener Gewichtung (links T1, rechts Protonendichte)<sup>2</sup>

- **Digitalfotografie:** In den letzten Jahren fand in der Fotografie eine technische Revolution statt, mit einem fast lückenlosen Übergang zu Digitalkameras, die mittels Fotochips direkt ein digitales Abbild betrachteter Szenen erzeugen. Damit ergibt sich kein Fehler wie etwa beim Scannen und auch Verunreinigungen wie bei Filmen treten nicht auf. Durch die Eigenschaften der Fotochips (Detektorrauschen etc.) kann man aber dennoch leicht verrauschte oder mit Artefakten behaftete Bilder erhalten, die einer Verbesserung bedürfen.

<sup>1</sup>Dankenswerterweise zur Verfügung gestellt von Paolo Favaro, Institute for Computer Science, Herriot-Watt University Edinburgh.

<sup>2</sup>Dankenswerterweise zur Verfügung gestellt von Carsten Wolters, Institut für Biomagnetismus und Biosignalanalyse, WWU Münster.

- **Mikroskopie:** In der Mikroskopie versucht man üblicherweise an die Grenzen der Auflösung zu gehen, was durch Ausnützung verschiedener optischer Effekte passiert. Moderne Mikroskope zeichnen Bilder ebenfalls direkt digital auf Chips auf, das Licht wandert vorher aber durch einen oft sehr komplizierten Aufbau (halbdurchlässige Spiegel, Reflektoren, etc.). Dadurch zeichnet man oft nur eine verzerrte Version des Bildes auf, und benötigt eine Korrektur um das eigentlich interessante Bild zu erhalten.

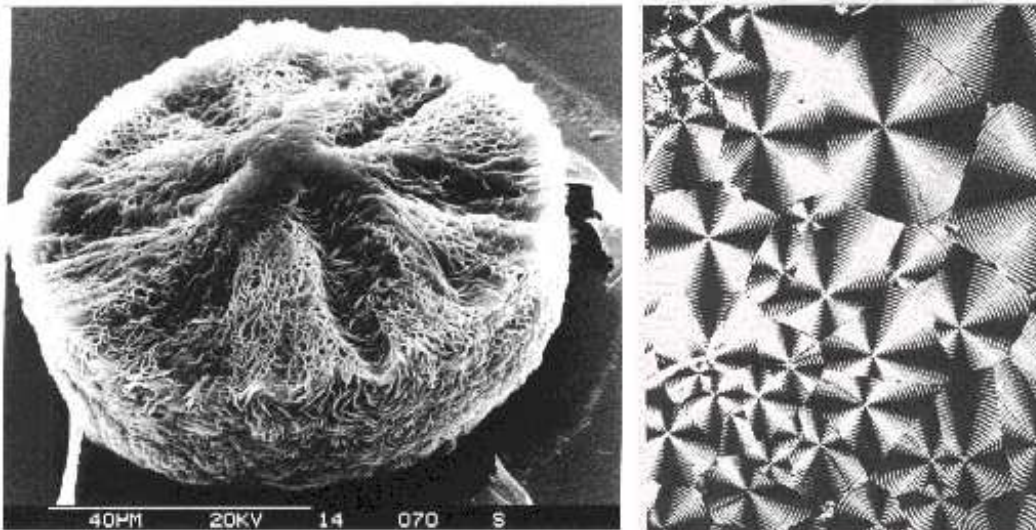


Abbildung 1.3: Mikroskopie eines Polymerkristalls (links) sowie Schnitt durch ein kristallisiertes Material (Polyhydroxybutyrat, rechts)

- **Medizinische Bildgebung:** Die medizinische Bildgebung besteht aus einer Reihe von sehr indirekten Verfahren, bei denen aus gemessenen physikalischen Größen (Schwächung von Röntgenstrahlen durch den Körper, Reflexion von Ultraschallwellen, Zerfallsdaten einer induzierten radioaktiven Substanz, ..) Dichtebilder des Patientenkörpers (oder eines Teils davon) berechnet werden sollen. Die Dichtebilder geben dann Aufschluss über mögliche Erkrankungen und Erkrankungsursachen (Tumore, Verstopfung von Herzgefäßen). Die Verfahren in der medizinischen Bildgebung sind sogenannte aktive Verfahren, d.h. Wellen (z.B. Röntgen- oder Ultraschallstrahlen) werden ausgesandt und die Wirkung des betrachteten Objekts auf diese Bilder wird gemessen).
- **Radar-Bilder:** Vielen Autofahrern bestens bekannt sind Radar-Bilder, die zur Geschwindigkeitsmessung benutzt werden. Bei einer Radar-Messung gewinnt man Bilder in einem sehr bestimmten Wellenbereich durch eine aktive Messung, hier sendet man Strahlen im Radarbereich aus und misst die gestreut zurückkommenden.
- **Satelliten-Bilder:** die Fernerkundung der Erde basiert auf Satelliten-Bildern, die eine vielfältige Anwendung (Meteorologie, Klimaforschung, Ozeanographie, Rohstoffsuche, militärische Aufklärung) finden. Hier verwendet man meist passive Verfahren im sichtbaren Bereich bis zum Infrarot oder Mikrowellen, bei Radarbildern wiederum aktive Verfahren. Eine mathematische Herausforderung entsteht dabei durch Verzerrungseffekte, die in der Erdatmosphäre entstehen können.

- **Künstliche Bilder:** Immer häufiger arbeitet man heute auch mit künstlichen, d.h. am Computer erzeugten Bildern. Diese Bilder sind natürlich weder fehlerbehaftet noch verrauscht, die Herausforderung besteht eher darin die Bilder für das menschliche Auge möglichst realistisch aussehen zu lassen.

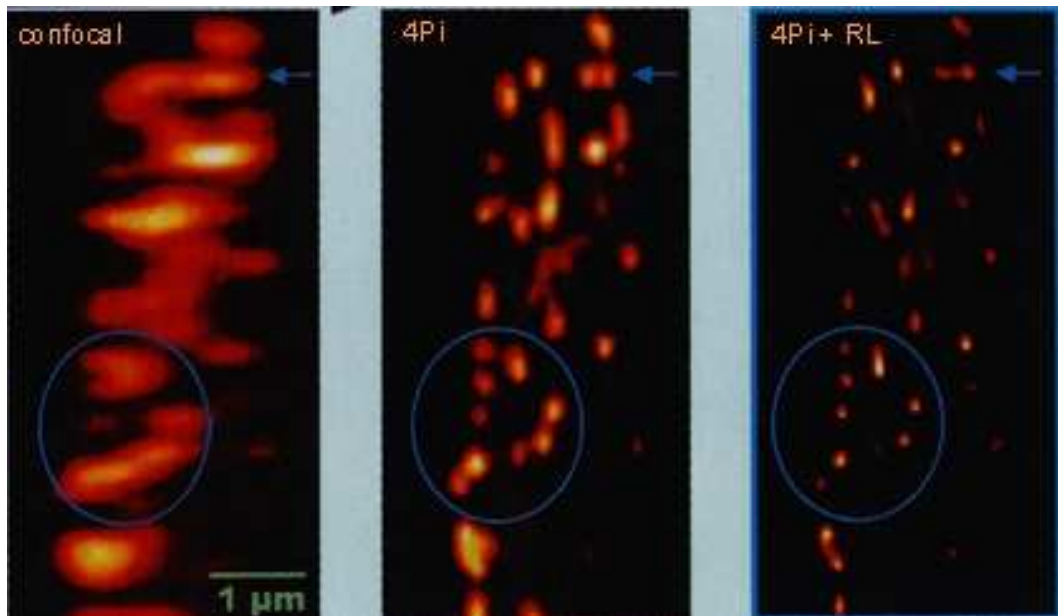


Abbildung 1.4: Mikroskopie in biologischen Anwendungen, klassische konfokale Technik (links), 4-Pi Mikroskopie mit erhöhter Auflösung (mitte) und 4-Pi Mikroskopie mit nicht-linearer Bildrekonstruktionstechnik (rechts)<sup>3</sup>



Abbildung 1.5: Künstliche Bilder: Computergraphik von realen<sup>4</sup>(links) und Phantasiewesen<sup>5</sup>(rechts).

<sup>3</sup>Dankenswerterweise zur Verfügung gestellt von der Abteilung für NanoBioPhotonik, Max-Planck Institut für biophysikalische Chemie, Göttingen.

<sup>4</sup>Quelle: <http://www.fifa07.ea.com>.

<sup>5</sup>Quelle: <http://www.mightandmagic.com>.

## 1.2 Aufgaben der Bildverarbeitung

- **Rekonstruktion (Reconstruction):** In den seltensten Fällen entspricht das aufgenommene Bild tatsächlich den Grau- oder Farbwerten, an denen man interessiert ist. So nimmt man etwa bei einem Computertomographen Mittelwerte der Dichte über Linien durch das Objekt auf, interessiert sich aber für die Dichte selbst, die im Bild dargestellt werden soll. Analoge Probleme treten auch bei allen anderen medizinischen bildgebenden Verfahren auf, sowie bei vielen anderen der oben genannten Anwendungen. Damit benötigt man als Grundlage jeder Weiterverarbeitung zunächst die Bildrekonstruktion, üblicherweise ein schlecht gestelltes Problem bei dessen (approximativer) Lösung einige Information verloren geht.

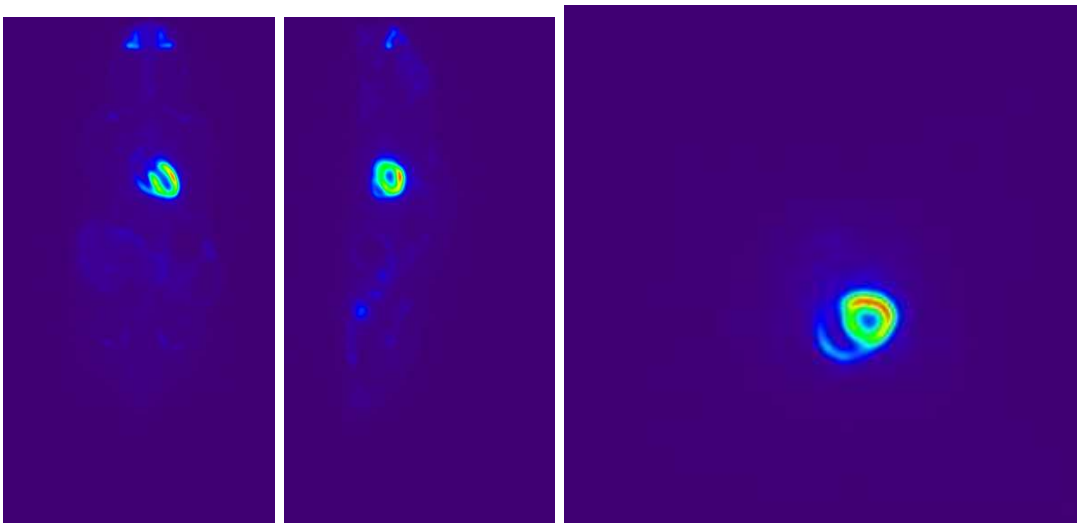


Abbildung 1.6: Bildrekonstruktion aus PET-Bild eines Mäuseherzens, koronaler (links), sagittaler (Mitte) und transverser Schnitt (rechts).<sup>6</sup>

- **Entzerren (Deblurring):** Ein besonders häufig auftretendes Beispiel der Bildrekonstruktion ist die Entzerrung, die mathematisch meist auf die Lösung einer Faltungsgleichung (oder allgemeiner Integralgleichung) führt. Die Verzerrung entsteht, weil man (z.B. bei Fotografietechniken) nicht direkt den Farbwert jedes Punktes, sondern eigentlich eine lokale (oder manchmal auch globale) Mittelung aufnimmt. Passiert die lokale Mittelung auf einer hinreichend kleinen Skala, so wird keine Entzerrung durchgeführt, sondern direkt das verzerrte Bild betrachtet. Den typischen Radius bei der lokalen Mittelung bezeichnet man dann als Auflösungsgrenze des bildgebenden Verfahrens (bzw. der Kamera, des Mikroskops, ..). Die Auflösung liesse sich dann durch eine Entzerrung eventuell noch weiter steigern, allerdings ist in vielen Fällen die Art der lokalen Mittelung (die rückgängig gemacht werden soll) nicht genau bekannt. Mathematisch führt dies auf die Lösung einer Faltungsgleichung mit unbekanntem Kern, und Lösungsansätze für solche Probleme werden als blinde Entzerrung (Blind Deconvolution) bezeichnet.
- **Entrauschen (Denoising):** Aufgrund von Messfehlern, Übertragungsfehlern oder ähnlichem hat man oft mit verrauschten Bildern zu tun, bei denen einzelne Pixelwerte

<sup>6</sup>Dankenswerterweise zur Verfügung vom SFB 656 Molecular Cardiovascular Imaging, WWU Münster.

komplett falsch oder teilweise gestört sind. Entrauschen bezeichnet dann die Aufgabe das Rauschen zu eliminieren, bzw. Bilder ähnlich zum verrauschten zu erzeugen, die aber für das menschliche Auge vernünftiger aussehen.

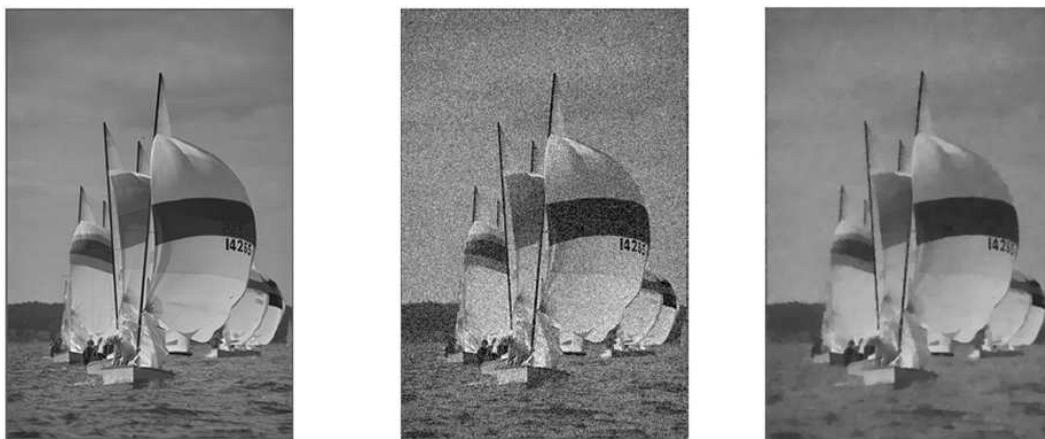


Abbildung 1.7: Entrauschen: sauberes Bild (links), verrauschtes Bild (Mitte), entrauschtes Bild (rechts).<sup>7</sup>

- **Kontrastkorrektur (Contrast Correction):** Aus verschiedenen Gründen (zu wenig Licht, falsche Kameraeinstellungen, zu ähnliche Objekte) erhält man manchmal Bilder mit wenig Kontrast, d.h. alle auftretenden Grau- oder Farbwerte sind in einem sehr ähnlichen Bereich. Um die Unterschiede besser sichtbar zu machen, führt man dann eine Kontrastkorrektur durch, um die Spanne der Farbwerte zu vergrößern. Dafür hat man mehrere Möglichkeiten (lineare oder nichtlineare Transformationen etc.), und die Hauptfragestellung ist dabei geeignete Transformationen zu finden, die genau die interessanten Teile sichtbar machen.
- **Segmentierung (Segmentation):** Als Segmentierung bezeichnet man das automatische Erkennen von Objekten ("objektbasierte Segmentierung") oder von Kanten ("kantenbasierte Segmentierung") in einem Bild. In den meisten Fällen hängen Kanten und Objekte auch direkt zusammen, denn die Objektgrenzen sind ja meist als Kanten in einem Bild sichtbar. Schwierig wird die Segmentierung dadurch, dass man eigentlich nur wichtige Kanten (bzw. solche die auch das menschliche Auge als Kanten klassifizieren würde) erkennen sollte. Die Segmentierung ist deshalb ein Aufgabengebiet der Computer Vision.  
 Mathematisch definiert man eine Kante meist als eine abrupte Änderung des Grau- und Farbwerts im digitalen Modell, bzw. eine Unstetigkeit der Intensitätsfunktion im Kontinuumsmodell.
- **Bildzerlegung (Image Decomposition):** Die Bildzerlegung ist eigentlich die Synthese aus Techniken wie Entrauschen und Segmentierung. Bei der Zerlegung versucht man die verschiedenen Anteile des Bildes, wie etwa Kanten, Texturen oder den Cartoon, zu trennen, um daraus weitere Rückschlüsse ziehen zu können.

<sup>7</sup>Erzeugt mit einem Code von Jinjun Xu, Google Inc., basierend auf einer TV-Methode.

- **Registrierung (Registration):** Vor allem in medizinischen Anwendungen tritt oft das Problem auf, dass man Information aus zwei Bildern zusammensetzen oder vergleichen will, die aus verschiedenen Modalitäten (z.B. Computertomograph und Ultraschall) gewonnen worden. Dabei tritt neben den Eigenheiten jeder Modalität auch das Problem auf, dass Teile des Bilds bewegt wurden (der Patient und seine Körperteile sind eventuell in verschiedener Lage, seine Organe ändern Grösse und Form wie z.B. das schlagende Herz, ...). Registrierung bezeichnet dann die Aufgabe diese Bewegungen und Verformungen zu korrigieren, sodass man zwei Bilder erhält, in denen zwei Punkte in den Bildern mit jeweils gleichen Koordinaten auch dem selben Punkt in der Realität entsprechen. Ein verwandtes Problem ist auch das *Warping*, bei dem man an einem ganzen, möglichst fließenden, Übergang zwischen zwei Bildern. Solche Effekte werden z.B. im Fernsehen gerne verwendet um einen Übergang zwischen zwei Gesichtern (verschiedene Personen oder gleiche Person zu verschiedenen Zeiten) darzustellen.
- **Restauration (Inpainting):** Die Restauration von Bildern ohne Computer ist ein Problem, mit dem viele Museen konfrontiert sind. Dabei werden z.B. beschädigte Bereiche eines Bildes wieder mit Farbe gefüllt, sodass die Restauration ein plausibles Bild ergibt. Ein analoges Problem stellt sich in der digitalen Welt, wenn man z.B. den Schmutz in einem gescannten Bild ausbessern möchte. Man muss das gegebene Bild in eine gewisse (meist nicht allzu grosse) Bildregion fortsetzen (weitermalen, daher der engl. Name Inpainting). Dies automatisch zu tun, damit es für das menschliche Auge vernünftig aussieht, ist eine nicht zu unterschätzende (mathematische) Aufgabe.

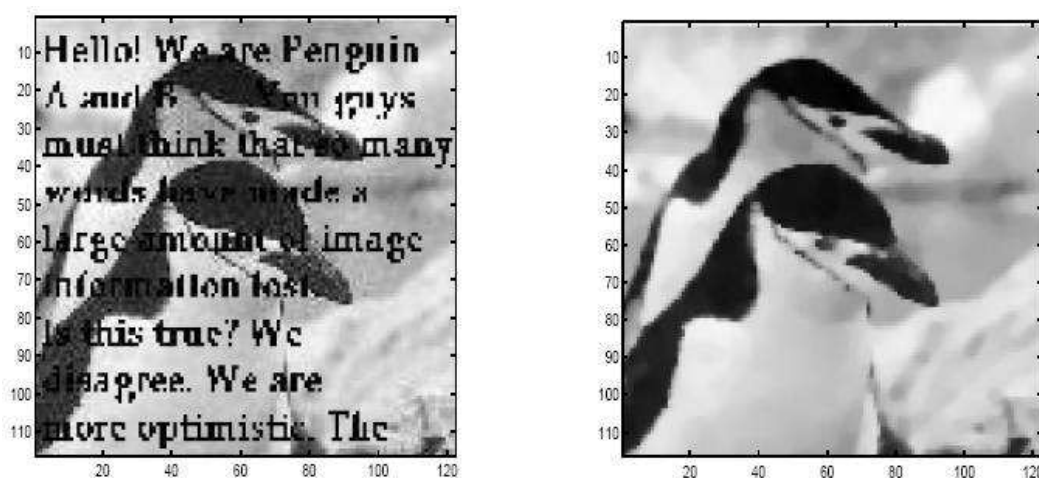


Abbildung 1.8: Inpainting zur Entfernung von Schrift.<sup>8</sup>

- **Bildfusion (Matching):** Bildfusion bezeichnet das Zusammenbringen von Teilen verschiedener Bilder zu einem einzigen Bild. Beispiele dafür sind Einzelbilder in verschiedene Richtungen, die zu einem Panoramabild zusammengesetzt werden sollen, oder mehrere Gruppenfotos, bei denen von jeder Person das beste Bild benutzt werden soll. Schwierigkeiten entstehen dabei durch möglicherweise verschiedenen Kontrast der Bilder sowie beim Verschmelzen der Ränder von einzelnen Teilbilder.

<sup>8</sup>Erzeugt mit einem Code von Lin He, RICAM Linz, basierend auf einer TV-Methode.

- **Bildkompression (Compression):** Ein trotz ständig wachsender Speichermedien noch immer aktuelles Problem ist die Bildkompression (und noch mehr die Videokompression). Klassische Formate wie Bitmap (bmp), die wirklich Grau- oder Farbwerte für jedes Pixel abspeichern, benötigen einen relativ hohen Speicheraufwand, auch bereits für typische natürliche Digitalbilder. Deshalb versucht man nur die wichtigen Teile der Bildinformation abzuspeichern. Dies ist vor allem möglich für alternative Repräsentationen (mathematisch gesehen andere Basen des Bildraums, wie Wavelets), die viele vernachlässigbare Anteile enthalten (mathematisch kleine Koeffizienten von Basiselementen, die durch Null gut approximiert werden). Dies passiert etwa, wenn wir ein Bitmap-Bild in das JPEG-Format umwandeln.

Neben den klassischen Aufgaben der Bildverarbeitung betrachtet man heute auch viele Aufgaben für Videos und Graphiken mit ähnlichen Techniken:

- **Bewegungserkennung (Motion Detection):** Bei der Bewegungserkennung versucht man aus einer Bildsequenz die sich bewegenden Objekte und deren Geschwindigkeiten zu erkennen. Dies ist kein einfaches Problem und auch kein eindeutig lösbares, wenn man keine Informationen über Vorder- und Hintergrund hat. Ein Beispiel ist eine weiße Scheibe, die sich vor schwarzem Hintergrund bewegt. Man kann dabei nicht unterscheiden, ob sich der weiße oder schwarze Anteil bewegt, denn es könnte sich ja auch eine schwarze Schablone vor ruhigem weißen Hintergrund bewegen.
- **Videokompression (Movie Compression):** Analog zur Bildkompression, mit noch viel grösseren Datensätzen (2+1 D oder 3+1 D). Neben der Kompression einzelner Bilder in der Sequenz wird dabei oft auch versucht, nur Teile der Bildsequenz abzuspeichern und Zwischenbilder durch Interpolation zu rekonstruieren.
- **Oberflächenglättung (Surface Smoothing):** Zur Visualisierung oder Restaurierung von Oberflächen ist Oberflächenglättung nötig, die kleine Oszillationen entfernt oder die Oberfläche glatt interpoliert. Die mathematischen Aufgaben dabei sind analog zum Entrauschen oder zur Restauration von Bildern.

### 1.3 Bildmodelle

Die klassische Beschreibung eines Bildes passiert durch seine Grauwerte, denen man eine gewisse Intensität in einem Intervall  $[a, b]$  zuordnet, wobei  $a$  weiss und  $b$  schwarz beschreibt. Dies ist ein sehr naheliegendes Vorgehen, da jede Graustufe ja einer Überlagerung von weissen und schwarzen Anteilen entspricht, d.h. mathematisch einer konvexen Kombination von  $a$  und  $b$ . Da sich der Grauwert natürlich lokal ändert, was ja erst ein interessantes Bild ausmacht, stellt sich die Frage nach der Repräsentation der Lokalität. Es gibt dazu zwei Standardansätze:

- **Diskret:** Bei einer diskreten Darstellung geht man von einem regelmässigen Gitter (mit  $N_1 \times N_2$  Punkten) über dem Bildgebiet aus, und ordnet jedem Gitterpunkt einen Grauwert zu. Damit ist das Bild mathematisch als eine Matrix  $U \in [a, b]^{N_1 \times N_2}$  definiert. Analog kann man natürlich auch 1D Bilder (Signale), 3D Bilder, oder auch Videos (mit zusätzlicher Zeitdimension) repräsentieren. Man bezeichnet die Grauwerte in den Gitterpunkten auch als Pixel. Für die einfachere Schreibweise werden wir im Verlauf des Skriptums meist  $N_1 = N_2 = N$  wählen.



- **Kontinuierlich:** Bei einer kontinuierlichen Darstellung geht man von einer Funktion definiert auf einem Bildgebiet  $\Omega \subset \mathbb{R}^d$  (bei der klassischen Definition eines Bildes  $d = 2$ ) aus, d.h.  $u : \Omega \rightarrow [a, b]$ . Die kontinuierliche Darstellung ist eigentlich eine Idealisierung, die auf keinem Computer realisiert werden kann. Andererseits liefert die kontinuierliche Darstellung eine sehr gute Basis für die Analysis und auch die Konstruktion von Verfahren, wie wir später sehen werden. Denkt man z.B. an mehrere Versionen eines Bildes mit verschiedener Auflösung, ist es naheliegend sich den Grenzwert für infinitesimale Auflösung (unendliche Anzahl an Gitterpunkten) vorzustellen, der eine Funktion liefern sollte. Die verschiedenen Bildern lassen sich dann als Samples dieser Funktion interpretieren.

Die Verbindung zwischen den beiden Darstellungen liefert die übliche Interpretation von Pixeln als lokal konstanter Grauwert in einer Zelle um den jeweiligen Punkt. In dieser Weise konstruiert man zur diskreten auch eine kontinuierliche Darstellung durch stückweise konstante Interpolation. Zoomt man in ein Digitalbild im Pixel-Format (z.B. bmp) so erkennt man tatsächlich diese Art der Darstellung.

Analog zu Graustufenbildern definiert man Farbbilder. Dabei wird statt einem Grauwert ein Tripel von Farbwerten (z.B. für Rot, Grün, und Blau) verwendet.

### 1.3.1 Diskretes Modell: Random Fields

Eine klassische Sichtweise für das diskrete Modell sind sogenannte *Gibbs<sup>9</sup> random fields*. Wie der Name schon deutlich macht, handelt es sich dabei um eine stochastische Beschreibung der Bilder, die auf einer a-priori Annahme über den Bildraum und die Wahrscheinlichkeit des Auftretens einzelner Bilder beruhen. Dabei wird das Bild  $U$  als Zufallsvariable  $\mathcal{U}$  über einem geeigneten Wahrscheinlichkeitsraum betrachtet. Die a-priori Wahrscheinlichkeit für die Realisierung  $U$  wird (in Analogie zur statistischen Mechanik) über ein Energiefunktional  $E : [a, b]^{N \times N} \rightarrow \mathbb{R}$  definiert, nämlich als

$$\mathbb{P}(\{\mathcal{U} \in \Theta\}) = c_E \int_{\Theta} e^{-\beta E(U)} dU, \quad (1.1)$$

wobei  $c_E$  eine geeignete Normierungskonstante ist, sodass

$$\mathbb{P}(\{\mathcal{U} \in [a, b]^{N \times N}\}) = c_E \int_{[a, b]^{N \times N}} e^{-\beta E(U)} dU = 1.$$

Stochastische Modelle werden nicht nur für Bilder, sondern auch fast immer für das Rauschen in Bildern verwendet. Ein klassisches Modell für das Rauschen ist identisch verteiltes Gauss<sup>10</sup>-sches Rauschen in jedem Punkt, d.h.

$$f_{ij} = u_{ij}^0 + n_{ij}^\sigma, \quad (1.2)$$

---

<sup>9</sup>Josiah Willard Gibbs (1839-1903), US-amerikanischer Physiker und Mathematiker, Professor an der Yale University, New Haven. Gilt als Begründer der physikalischen Chemie und brachte fundamentale Beiträge zur statistischen Mechanik, elektromagnetischen Theorie des Lichts und zur Vektoranalysis. Nach ihm wurden auch die Gibb'sche Phasenregel der Thermodynamik, das Gibb'sche Paradoxon der statistischen Mechanik und das Gibb'sche Phänomen (siehe unten) benannt.

<sup>10</sup>Johann Carl Friedrich Gauss (1777-1855), deutscher Mathematiker. Namensgeber einer Vielzahl von Sätzen und Methoden, sowie einer physikalischen Einheit. Bedarf hier wohl keiner weiteren Erläuterung, cf. [12].

wobei  $n_{ij}^\sigma$  unabhängige normalverteilte Zufallsvariable mit Varianz  $\sigma^2$  sind, und  $u$  bzw.  $u^0$  das verrauschte bzw. saubere Bild bezeichnen. Damit können sich natürlich verrauschte Grauwerte punktweise relativ stark verändern, wegen des schnellen Abklingens der Normalverteilung ist dies jedoch sehr unwahrscheinlich, vor allem für kleines  $\sigma$ . Wie häufig in der Stochastik erhält man eine kleinere Varianz aus dem Gesetz der grossen Zahl. Dazu betrachten wir das Quadrat des normierten  $\ell^2$ -Fehlers

$$\frac{1}{N_1 N_2} \sum_{ij} (f_{ij} - u_{ij}^0)^2 = \frac{1}{N_1 N_2} (n_{ij}^\sigma)^2 = \frac{\sigma^2}{N_1 N_2} \sum_{ij} \mathbb{E}(\sum_{ij} X_{ij}^2),$$

mit der standardnormalverteilten Zufallsvariable  $X_{ij} = \frac{n_{ij}^\sigma}{\sigma}$ . Aus den Eigenschaften der Normalverteilung

$$\mathbb{E}(n_{ij}^\sigma) = 0, \quad \mathbb{V}(n_{ij}^\sigma) = \sigma^2,$$

erhalten wir

$$\frac{1}{N_1 N_2} \sum_{ij} \mathbb{E}((f_{ij} - u_{ij}^0)^2) = \frac{1}{N_1 N_2} \sum_{ij} \mathbb{E}(n_{ij}^2) = \frac{1}{N_1 N_2} \sum_{ij} \mathbb{V}(n_{ij}^\sigma) = \sigma^2.$$

Dies ist noch kein Gewinn zur lokalen Varianz, da wir wieder die Grösse  $\sigma^2$  erhalten. Nun können wir aber auch die Varianz des  $\ell^2$ -Fehlers betrachten. Die Summe aus den  $N_1 N_2$  Quadraten von unabhängigen normalverteilten Zufallsvariablen  $X_i$  ist  $\chi^2$ -verteilt mit  $N_1 N_2$  Freiheitsgraden. Die Varianz einer solchen  $\chi^2$ -verteilten Zufallsvariable ist  $2N_1 N_2$  und damit folgt

$$\mathbb{V}\left(\frac{1}{N_1 N_2} \sum_{ij} (f_{ij} - u_{ij}^0)^2\right) = \frac{\sigma^2}{N_1 N_2} \mathbb{V}(\sum_{ij} X_{ij}^2) = 2\sigma^2.$$

Damit ist der  $\ell^2$ -Fehler mit äusserst grosser Wahrscheinlichkeit durch ein (kleines) Vielfaches von  $\sigma$  beschränkt.

Auch andere stochastische Rausch-Modelle sind in der Bildverarbeitung weit verbreitet, am bekanntesten ist *salt-and-pepper noise* bei dem Grauwerte von Pixeln entsprechend den Werten eines Poisson<sup>11</sup>-Prozesses geändert werden.

Eine weiteres wichtiges stochastisches Konzept sind Bayes<sup>12</sup>-sche Modelle. Dabei wird jedem Bild eine a-priori Wahrscheinlichkeit oder Wert einer Wahrscheinlichkeitsdichte (wie bei der Gibb'schen Formel oben) zugeordnet. Zusätzlich benötigt man noch eine Wahrscheinlichkeit für das Auftreten einer bestimmten Messung unter einem gegebenen Bild. Das Bayes'sche Modell basiert dann auf der sogenannte *a-posteriori Wahrscheinlichkeit* (a-posteriori probability, auch Bayes'sche Formel genannt)

$$\mathbb{P}(U | F) = \frac{\mathbb{P}(F | U)\mathbb{P}(U)}{\mathbb{P}(F)}. \quad (1.3)$$

Hier ist  $U$  das gesuchte Bild und  $F$  die Beobachtung. Die Formel für  $\mathbb{P}(U | F)$  beschreibt die Wahrscheinlichkeit, dass die Beobachtung  $F$  von einem Bild  $U$  erzeugt wurde. Dabei ist

<sup>11</sup>Simeon-Denis Poisson (1781-25. April 1840), französischer Physiker und Mathematiker, Professor an der Ecole Polytechnique in Paris. Arbeitete auf einer Vielzahl von Gebieten (Akustik, Elastizität, Wärmetheorie, Differentialgeometrie, Infinitesimalrechnung, Wahrscheinlichkeitstheorie). Namensgeber für mehrere physikalische und mathematische Begriffe.

<sup>12</sup>Thomas Bayes (1702-1761), englischer Mathematiker und Pfarrer. Wurde vor allem durch eine zwei Jahre nach seinem Tod veröffentlichte Arbeit zur Wahrscheinlichkeit bekannt, cf. [12]

$\mathbb{P}(F | U)$  die Wahrscheinlichkeit, dass bei gegebenem Bild  $U$  die Beobachtung  $F$  eintritt, und  $\mathbb{P}(U)$ ,  $\mathbb{P}(F)$  sind a-priori Wahrscheinlichkeiten für das Bild und die Beobachtung. Durch Maximierung der Wahrscheinlichkeit kann man auch einen Schätzer konstruieren, den sogenannten *maximum a-posteriori probability (MAP)* Schätzer

$$\hat{U} = \arg \max_U \mathbb{P}(U | F). \quad (1.4)$$

Man sieht aus (1.3) und (1.4), dass nur die a-priori Wahrscheinlichkeit für  $U$  wichtig ist, da  $F$  ohnehin fix gehalten wird. Der Nenner  $\mathbb{P}(F)$  ist dann nur eine Normierungskonstante. Wir verdeutlichen dieses Modell anhand eines Gauss'schen Modells für das Rauschen und eines Gibbs-Modells für die a-priori Wahrscheinlichkeit. In diesem Fall haben wir kontinuierliche Zufallsvariable und verwenden daher die Dichten

$$\rho_{\mathbb{P}}(U | F) = \rho_{\mathbb{P}}(F | U)\rho_{\mathbb{P}}(U),$$

und den MAP-Schätzer

$$\hat{U} = \arg \max_U \rho_{\mathbb{P}}(U | F). \quad (1.5)$$

Da beim Gauss'schen Modell für das Rauschen  $F$  durch Addition normalverteilter Zufallsvariable zu  $U$  entsteht, ist das Ereignis der Beobachtung von  $F$  unter  $U$  äquivalent zu  $n_{ij}^{\delta} = F_{ij} - U_{ij}$ . Da wir unabhängige normalverteilte Zufallsvariable haben, kann die Wahrscheinlichkeitsdichte als

$$\rho_{\mathbb{P}}(F | U) = \prod_{i,j} \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{(F_{ij}-U_{ij})^2}{2\sigma^2}}$$

berechnet werden. Zusammen mit dem Gibbs-Modell erhalten wir die a-posteriori Wahrscheinlichkeitsdichte als

$$\rho_{\mathbb{P}}(U | F) = C e^{-\beta E(u)} \prod_{i,j} e^{-\frac{(F_{ij}-U_{ij})^2}{2\sigma^2}}.$$

Den MAP-Schätzer erhalten wir analog durch Minimierung des negativen Logarithmus der Wahrscheinlichkeitsdichte, d.h.

$$\hat{U} = \arg \min_U \left( \frac{1}{2} \sum_{ij} (F_{ij} - U_{ij})^2 + \sigma^2 \beta E(u) \right). \quad (1.6)$$

### 1.3.2 Kontinuierliches Modell: Funktionenräume, Variationsmethoden

Das Minimierungsproblem in (1.6) liefert einen natürlichen Übergang zu Variationsmethoden, die sehr häufig in der Bildverarbeitung verwendet werden. Die Grundlage für die Entwicklung und das Verständnis von Variationsmethoden ist meist das kontinuierliche Bildmodell und die Betrachtung in Funktionenräumen.

Dazu spezifiziert man meist einen Funktionenraum  $\mathcal{X}$  der die gewünschten Bilder enthält und einen zweiten Funktionenraum  $\mathcal{Y}$  für die Beobachtung. So kann man z.B. beim Entrauschen  $\mathcal{X} \leftrightarrow \mathcal{Y}$  wählen und das verrauschte Bild wieder in der Form

$$f(x) = u^0(x) + n^{\sigma}(x) \quad (1.7)$$

modellieren, üblicherweise mit einer lokal normalverteilten Variable  $n^\sigma(x)$  mit Varianz  $\sigma^2$ . Tastet man den Fehler wieder auf einem geeigneten Gitter ab, so gilt

$$\frac{1}{N_1 N_2} \sum_{ij} (f(x_{ij}) - u(x_{ij}))^2 \rightarrow \int_{\Omega} (f(x) - u(x))^2 dx,$$

für  $N_1, N_2 \rightarrow \infty$ . Deshalb ist es naheliegend einen Fehler der Form

$$\int_{\Omega} (f(x) - u(x))^2 dx \approx \sigma^2$$

anzunehmen, d.h. man erhält eine Abschätzung für die Norm von  $f - u$  im Hilbertraum

$$L^2(\Omega) = \{v : \Omega \rightarrow \mathbb{R} \mid v \text{ messbar, } \int_{\Omega} v^2 < \infty\}. \quad (1.8)$$

Die Norm in  $L^2(\Omega)$  ist definiert durch

$$\|v\|_{L^2(\Omega)} := \sqrt{\int_{\Omega} v(x)^2 dx}. \quad (1.9)$$

Die hier benutzten Integrale sind natürlich alle als Lebesgue<sup>13</sup>-Integrale zu verstehen.

Wegen der Fehlerschranke in  $L^2(\Omega)$  ist es naheliegend, den Raum für die Beobachtung als  $\mathcal{Y} = L^2(\Omega)$  zu wählen. Dieser Raum enthält aber sehr stark oszillierende Funktionen, unter anderem eben auch das Rauschen. Deshalb versucht man normalerweise durch die Wahl eines kleineren Teilraums  $\mathcal{X}$  glattere Bilder zu erzwingen, wobei man Glattheit meist über den Gradienten definiert. Bei vielen Bildern findet man einige Regionen mit einem sehr ähnlichen Grauwert, d.h. einem kleinen Gradienten. Nur in wenigen Regionen, vor allem bei Kanten, kommt es zu starken Änderungen des Gradienten. Damit sollte es möglich sein die typische Glattheit von Bildern über Integrale des Gradienten zu definieren (eine punktweise Betrachtung des Gradienten scheidet wegen der starken lokalen Änderung aus). Eine einfache Möglichkeit wäre der Sobolevraum<sup>14</sup>

$$H^1(\Omega) = \{v : \in L^2(\Omega) \mid \nabla v \in L^2(\Omega; \mathbb{R}^d)\}, \quad (1.10)$$

mit der Norm

$$\|v\|_{H^1(\Omega)} := \sqrt{\int_{\Omega} (v(x)^2 + |\nabla v(x)|^2) dx}. \quad (1.11)$$

Es stellt sich jedoch heraus, dass diese Wahl zu einschränkend für viele typische Bilder ist. Unter anderem liegt eine unstetige stückweise konstante Funktion nie in  $H^1(\Omega)$ , dies würde z.B. alle nur aus schwarz und weiss bestehenden Bilder, also eine sehr einfache Klasse, eliminieren.

Betrachtet man die Sobolevräume

$$W^{1,p}(\Omega) = \{v : \in L^p(\Omega) \mid \nabla v \in L^p(\Omega; \mathbb{R}^d)\}, \quad (1.12)$$

<sup>13</sup>Henri Leon Lebesgue (1875-1941), französischer Mathematiker, Professor in Nancy. Erweiterte den Integralbegriff und begründete damit die Maßtheorie. Nach ihm benannt sind auch das Lebesgue-Maß und der Satz von Lebesgue benannt, auch der Satz über die dominierte Konvergenz von Integralen stammt von ihm.

<sup>14</sup>Sergei Lwowitzsch Sobolew (1908-1989), russischer Mathematiker, Professor in Moskau und Novosibirsk. Brachte grundlegende Beiträge zur funktionalanalytischen Behandlung partieller Differentialgleichungen und zu Distributionen, später auch zur Numerik, cf. [12].

mit der Norm

$$\|v\|_{W^{1,p}(\Omega)} := \left( \int_{\Omega} (|v(x)|^p + |\nabla v(x)|^p) dx \right)^{1/p} \quad (1.13)$$

für  $1 \leq p < 2$ , so erhält man grössere und interessantere zulässige Klassen von Bildern, leider aber immer noch nicht die stückweise konstanten Grauwertfunktionen. Dies sieht man daraus, dass die distributionelle Ableitung einer stückweise konstanten Funktion nie eine  $L^p$ -Funktion, sondern nur ein Maß auf  $\Omega$  wird (der Gradient ist dann eine vektorwertiges Maß). Dies sieht man einfach am eindimensionalen Beispiel der Heaviside<sup>15</sup>-Funktion  $H$ , wobei  $H(x) = 0$  für  $x < 0$  bzw.  $H(x) = 1$  für  $x > 0$  gilt. Die distributionelle Ableitung der Heaviside-Funktion ist die Dirac<sup>16</sup>  $\delta$ -Distribution, die man als beschränktes Maß (auch Radon<sup>17</sup>-Maß genannt) auffassen kann. Mit dieser Verallgemeinerung erhält man den Raum von Funktionen beschränkter Variation

$$BV(\Omega) = \{v \in L^1(\Omega) \mid \partial_{x_i} v \text{ ist Radon-Maß, } i = 1, \dots, d\}, \quad (1.14)$$

mit der Norm

$$\|v\|_{BV(\Omega)} = \int_{\Omega} |v(x)| dx + |u|_{BV}, \quad (1.15)$$

mit der Halbnorm (oder *totalen Variation*)

$$|u|_{BV} = \sqrt{\sum_{i=1}^d \left( \int_{\Omega} d|\partial_{x_i} v| \right)^2}. \quad (1.16)$$

Wie man aus dem Beispiel der Heaviside-Funktion sofort sieht, ist der Raum  $BV(\Omega)$  tatsächlich gross genug um auch stückweise konstante Funktionen (und auch viele andere Bilder mit Kanten) zu enthalten. Daher ist  $BV(\Omega)$  ein recht angenehmer Raum zur Beschreibung von Bildern, der uns (ebenso wie die totale Variation) im Laufe dieser Vorlesung noch häufig begegnen wird.

Variationsmethoden in Funktionenräumen können als natürliche Erweiterung von (1.4) bzw. (1.6) gesehen werden. Dabei minimiert man meist Funktionale der Form

$$R(u, f) + \alpha E(u) \rightarrow \min_{u \in \mathcal{X}}, \quad (1.17)$$

mit einem Fitting-Funktional  $R : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{\infty\}$ , das die Abweichung eines Bilds  $u$  von gegebenen Daten oder einer gegebenen Aufgabenstellung misst, sowie einer Regularisierungsenergie  $E : \mathcal{X} \rightarrow \mathbb{R} \cup \{\infty\}$ . Die Regularisierungsenergie spielt die selbe Rolle wie  $E$  in der

<sup>15</sup>Oliver Heaviside (1850-1925), britischer Mathematiker und Physiker. Erwarb sich autodidaktisch umfassende mathematische Kenntnisse und begründete die moderne vektoranalytische Form der Maxwell'schen Theorie durch formale Anwendung der Operatorrechnung. Die Heaviside-Funktion benutzte er zur Modellierung des Stromflusses in einem elektrischen Schaltkreis, cf. [12]

<sup>16</sup>Paul Adrien Maurice Dirac (1902-1984), britischer Physiker. Mitbegründer der Quantenphysik, entwickelte eine Fassung der Quantenmechanik (Transformationstheorie), die die Matrizenmechanik Heisenbergs und die Wellenmechanik Schrödingers als Spezialfälle enthält. Nach ihm wurden auch die Dirac-Gleichung (eine relativistische Wellengleichung für Elektronen) und die Fermi-Dirac-Statistik (zur Beschreibung von Fermionen) benannt, cf. [12].

<sup>17</sup>Johann Radon (1887-1956), österreichischer Mathematiker, Professor in Hamburg, Greifswald, Erlangen, und Wien. Nach ihm wurden auch die Radon-Transformation (in der Computertomographie von Bedeutung) und der Satz von Radon-Nikodym in der Maßtheorie benannt, cf. [12].

a-priori Wahrscheinlichkeit bei (1.4), es führt also a-priori Wissen über zu erwartende Bilder ein. Der Regularisierungsparameter  $\alpha > 0$  wird dann zur Gewichtung der beiden Terme benutzt. Im Falle des Entrauschens bei Gauss'schem Rauschen kann man (1.6) direkt zu

$$\frac{1}{2} \int_{\Omega} (u - f)^2 dx + \alpha E(u) \rightarrow \min_{u \in \mathcal{X}} \quad (1.18)$$

verallgemeinern. Die Frage der geeigneten Wahl der Energie  $E$  und des Raumes  $\mathcal{X}$  werden wir im nächsten Kapitel genauer diskutieren.

### 1.3.3 Morphologische Darstellung

Bei einer morphologischen Darstellung betrachtet man Bilder in einer gegenüber Kontraständerungen invarianten Weise, nämlich über die Level Sets

$$\mathcal{M}_{\alpha}(u) = \{x \mid u(x) < \alpha\}.$$

Ändert man den Kontrast des Bildes, d.h. führt man eine monotone Transformation  $u(x) \mapsto G(u(x))$  in jedem Punkt  $x$  durch, so bleiben die Level Sets insgesamt die selben, es wird nur  $\mathcal{M}_{\alpha}$  zu  $\mathcal{M}_{H(\alpha)}$  geändert.

Will man kontrastinvariante Methoden zur Bildverarbeitung erzeugen, so bedient man sich der morphologischen Darstellung und arbeitet damit eigentlich mit geometrischen Objekten, nämlich den Level Sets  $\mathcal{M}_{\alpha}$  und ihren Rändern. Hat man nun vernünftige Methoden z.B. zur Glättung oder Registrierung der Level Sets, dann kann man diese auch wieder zu einer morphologischen Methode für die Gesamtfunktion zusammensetzen. Die Funktion  $u$  kann aus ihren Level Sets mit der Formel

$$u(x) = \inf\{\alpha \in \mathbb{R} \mid x \in \mathcal{M}_{\alpha}\} \quad (1.19)$$

rekonstruiert werden.

Bei Transformationen von Gebieten spricht man auch von *morphologischen Operationen*. Wichtige Beispiele sind die *Erosion* und *Dilatation* einer Menge  $A$ . Dabei geht man von einem *Strukturelement*  $S \subset \mathbb{R}^d$ , mit  $0 \in S$  aus (meist ist  $S$  eine kleine Umgebung von Null), und bildet die Dilatation und Erosion als

$$D_S(A) = \{x \in \mathbb{R}^d \mid x + S \cap A \neq \emptyset\} \supset A \quad (1.20)$$

$$E_S(A) = \{x \in \mathbb{R}^d \mid x + S \subset A\} \subset A \quad (1.21)$$

Eine analoge morphologische Darstellung ist natürlich völlig analog im Fall eines diskreten Bildmodells möglich.

### 1.3.4 Darstellung im Frequenzraum

Eine in der Signalverarbeitung ( $d = 1$ ) und auch in der mehrdimensionalen Bildverarbeitung beliebte Darstellung ist jene im Frequenzraum, d.h. man betrachtet die Fourier<sup>18</sup>-Transformation  $\hat{u}$  anstatt des ursprünglichen Signals  $u$ . Die Fourier-Transformation kann man

<sup>18</sup>Jean Baptiste Joseph Fourier (1768-1830), französischer Mathematiker und Physiker, Professor an der Ecole Polytechnique in Paris. Beschäftigte sich mit der Wärmeleitung in Festkörpern, nach ihm wurde auch das Fourier'sche Abkühlungsgesetz benannt.

sowohl für ein diskretes Signal als

$$\hat{u}_k = \sum_{j=1}^N u_j e^{-2\pi \frac{kj}{N}} \quad (1.22)$$

oder für ein kontinuierliches Signal als

$$\hat{u}(\omega) = \mathcal{F}u(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u(x) e^{-i\omega x} dx \quad (1.23)$$

definieren. Für Signale auf beschränkten Intervallen genügt es einige Samples der Fourier-Transformation zu betrachten (wobei die Funktion ausserhalb des Intervalls durch Null fortgesetzt wird), nämlich die Koeffizienten der Fourier-Reihe. Für Signale auf dem Intervall  $[-\pi, \pi]$  sind diese gegeben durch

$$\hat{f}_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} u(x) e^{-ikx} dx. \quad (1.24)$$

Bei allen Varianten der Fourier-Transformation wird das Signal in seine Schwingungsteile zerlegt, wachsendes  $|\omega|$  entspricht höheren Frequenzen. Für  $\omega = 0$  erhält man den niedrigsten Frequenzanteil, nämlich genau den Mittelwert. Dabei geht auch keine Information verloren, da die Fourier-Transformierte eine stetige Inverse besitzt. Im Fall der stetigen Fouriertransformation ist die Inverse gegeben durch

$$u(x) = \mathcal{F}^{-1}\hat{u}(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{u}(\omega) e^{i\omega x} d\omega. \quad (1.25)$$

Da man Rauschen oft mit den hochfrequenten Anteilen identifiziert, versucht man Ent-rauschen auch durch Verkleinern hochfrequenter Anteile zu erreichen. Filter basierend auf Fourier-Transformationen definiert man dazu in der Form

$$v = \mathcal{F}^{-1}(\psi \mathcal{F}u) \quad (1.26)$$

mit einer geeigneten Filterfunktion  $\psi$ . Ein einfaches Beispiel ist der low-pass Filter definiert durch  $\psi(\Omega) = H(\omega_0 - |\omega|)$ , der alle Frequenzen über  $\omega_0$  eliminiert. Solche Filter sind sehr erfolgreich beim Ent-rauschen glatter Signale, allerdings treten wieder Probleme bei Kanten auf. Der Grund dafür ist das langsame Abklingverhalten der Fourier-Transformation un-stetiger Funktionen, die Filter eliminieren in diesem Fall nicht nur das Rauschen, sondern auch einen wichtigen Teil des Signals. Lässt man die hohen Frequenzen bei einer un-stetigen Funk-tion weg, so tritt das sogenannte Gibbs-Phänomen auf, d.h. es kommt zu einer Glättung der Unstetigkeit mit Über- bzw. Unterschwingen.

Auch für mehrdimensionale Bilder ist die Fouriertransformation ein wichtiges Hilfsmittel, dann definiert durch

$$\hat{u}(\omega) = \mathcal{F}u(\omega) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} u(x) e^{-i\omega \cdot x} dx. \quad (1.27)$$

Besonders häufig wird die Fouriertransformation bei Bildern im Zuge des Entzerrrens einge-setzt. Dies liegt daran, dass Verzerrung meist zu einer Faltung des Bilds mit einem Kern  $k$

führt und die Fouriertransformation eine Faltung stark vereinfacht. Genauer wird die Beobachtung einer Verzerrung durch die Faltung der Bildintensität  $u$  mit einem Kern  $k$  (oft ein Gauss-Kern) modelliert, d.h.

$$f(x) = (k * u)(x) = \int_{\mathbb{R}^d} k(x-y)u(y) dy. \quad (1.28)$$

Unter geeigneten Voraussetzungen an die Funktionen  $u$  und  $k$  (cf. [3]) gilt der Fourier'sche Faltungssatz

$$\mathcal{F}(k * u) = (2\pi)^{d/2}(\mathcal{F}k)(\mathcal{F}u), \quad (1.29)$$

d.h. die Faltung wird im Frequenzraum zur einfachen Multiplikation zweier Funktionen.

Ein Nachteil der Fourier-Transformation ist die fehlende räumliche Lokalisierung. Die Frequenzanteile werden nur für das ganze Bild berechnet und es kann nicht mehr einfach zwischen Regionen unterschiedlicher Intensität unterschieden werden. Deshalb gab es in den achtziger und neunziger Jahren des letzten Jahrhunderts verstärkte Bemühungen lokalisierte Versionen der Fouriertransformation zu entwickeln. Ein erster Ansatz dazu ist eine direkte Lokalisierung der zur Transformation benutzen Wellen  $e^{-ik \cdot x}$  zu einer Welle der Form  $\psi_{y,\omega} = w(x-y)e^{-ix \cdot \omega}$ , wobei  $w$  eine Fensterfunktion mit lokalem Träger um den Punkt  $x = 0$  ist. Man erhält daraus die sogenannte gefensterte Fouriertransformation (windowed Fouriertransform) oder auch Gabortransformation<sup>19</sup>

$$U(\omega, y) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} u(x)w(x-y)e^{-i\omega \cdot x} dx. \quad (1.30)$$

Man nennt die lokalisierende Welle  $\psi_{y,\omega}$  auch Gabor-Wavelet. Skaliert man die Fensterfunktion so, dass

$$\int_{\mathbb{R}^d} w(x)^2 dx = \int_{\mathbb{R}^d} w(x-y)^2 dx = 1$$

gilt, dann kann  $u$  mittels

$$u(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} U(\omega, y)w(x-y)e^{ix \cdot \omega} d\omega dy \quad (1.31)$$

aus der Gabor-Transformation rekonstruiert werden. Die Gabor-Transformation weist nun aber den Nachteil auf, dass man eine  $2d$ -dimensionale Funktion erhält und damit sehr viel redundante Information (sowie eventuell ein Speicherproblem). Darüberhinaus ist die lokale Skala auf der man Wellen betrachtet durch die spezielle Wahl der Fensterfunktion (bzw. deren Träger) bereits festgelegt, also kann man wieder nur zwischen sehr bestimmten Bildregionen unterscheiden. Eine Version mit Variablen Skalen kann man mit der Wavelet-Transformation

$$U(\omega, y, \sigma) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} u(x)w\left(\frac{x-y}{\sigma}\right)e^{-i\omega \cdot x} dx. \quad (1.32)$$

Die Redundanz wird natürlich durch die dritte Variable  $\sigma$  noch grösser, was sich im allgemeinen auch in numerischer Stabilität in der Algorithmen niederschlägt. Dieser Nachteile kann durch die Verwendung von Wavelets basierend auf Multiresolutions-Analysis behoben werden, die wir im nächsten Abschnitt vorstellen werden.

<sup>19</sup>Denes Gabor (1900-1979), ein ungarischer Ingenieur bei Siemens, später Professor am Imperial College in London. Entwickelte u.a. das Prinzip der Holographie. Nobelpreisträger für Physik 1971, cf. [12]



### 1.3.5 Wavelet Darstellung

Der Grundstein für den Erfolg von Wavelets war der Zugang über Multiresolutions-Analysis (MRA), der 1989 von Meyer<sup>20</sup> und Mallat<sup>21</sup> vorgeschlagen wurde. Der Einfachheit halber beschränken wir uns bei der Darstellung auf den eindimensionalen Fall, mehrdimensionale Erweiterungen sind jedoch möglich.

**Erinnerung: Orthonormalbasen (ONB).** Eine Grundlage für die Darstellung mit Wavelets sind *orthonormale Basen* eines Hilbertraums. Sei  $\mathcal{X}$  ein separabler Hilbertraum. Eine linear unabhängige Teilmenge  $\{\phi_k\}_{k \in \mathbb{Z}}$  von  $\mathcal{X}$  heisst orthonormale Basis (ONB), falls für jedes  $u \in \mathcal{X}$  die Reihe

$$\|u\|_{\mathcal{X}}^2 = \sum_{k \in \mathbb{Z}} \langle u, \phi_k \rangle^2$$

konvergiert und es gilt

$$u = \sum_{k \in \mathbb{Z}} \langle u, \phi_k \rangle \phi_k.$$

Eine *orthonormale MRA* auf dem Raum  $L^2(\mathbb{R})$  ist eine geordnete Kette geschlossener Teilräume

$$\dots \subset \mathcal{V}_{-1} \subset \mathcal{V}_0 \subset \mathcal{V}_1 \subset \dots, \quad (1.33)$$

die die folgenden Bedingungen erfüllt:

1. **Vollständigkeit:**  $\overline{\lim_{j \rightarrow \infty} \mathcal{V}_j} = L^2(\mathbb{R})$  und  $\lim_{j \rightarrow \infty} \mathcal{V}_j = \{0\}$ .
2. **Dyadische Ähnlichkeit:**  $u(x) \in \mathcal{V}_j \Leftrightarrow u(2x) \in \mathcal{V}_{j+1}$ .
3. **Translationseigenschaft:** Es existiert eine Skalierungsfunktion  $\phi \in \mathcal{V}_0$ , sodass  $\{\phi(x - k)\}_{k \in \mathbb{Z}}$  ist orthonormale Basis von  $\mathcal{V}_0$ .

Basierend auf einer MRA kann man den ganzen Raum  $L^2(\mathbb{R})$  durch die Skalierungsfunktion darstellen. Dazu verwendet man die Familie

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad j, k \in \mathbb{Z}. \quad (1.34)$$

Aus der Vollständigkeit sieht man zunächst, dass die Menge  $\{\phi_{0,k}\}_{k \in \mathbb{Z}}$  eine Basis von  $L^2(\mathbb{R})$  ist. Aus der dyadischen Ähnlichkeit folgern wir nun, dass  $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$  eine orthonormale Basis von  $\mathcal{V}_j$  ist:

**Lemma 1.1.** *Für  $j$  in  $\mathbb{Z}$  ist  $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$  eine orthonormale Basis von  $\mathcal{V}_j$ .*

**Beweis.** Zunächst sieht man aus der dyadischen Ähnlichkeit, dass für  $u \in \mathcal{V}_j$  die Funktion  $v$ , definiert durch  $v(x) = u(2^{-j}x)$  in  $\mathcal{V}_0$  liegt. Da  $\{\phi(\cdot - k)\}_{k \in \mathbb{Z}}$  eine ONB von  $\mathcal{V}_0$  ist, gilt

$$v(y) = \sum_{k \in \mathbb{Z}} \gamma_k \phi(y - k), \quad \gamma_j = \int_{\mathbb{R}} v(y) \phi(y - k) dy.$$

<sup>20</sup>Yves Meyer (\*1989), französischer Mathematiker., Professor an der Ecole Normale Supérieure de Cachan, zahlreiche Arbeiten auf dem Gebiet der nichtlinearen Analysis und bei der Entwicklung der Wavelet-Theorie.

<sup>21</sup>Stephane Mallat (\*?), französischer Mathematiker. Professor in New York und an der Ecole Polytechnique in Paris, zahlreiche Arbeiten auf dem Gebiet der Signal- und Bildverarbeitung, vor allem zur Theorie und Implementierung von Wavelet-Methoden.

Mit der Transformation  $x = 2^{-j}y$  folgt

$$u(x) = v(2^j x) \sum_{k \in \mathbb{Z}} \gamma_k \phi(2^j x - k) = \sum_{k \in \mathbb{Z}} \gamma_k 2^{-j/2} \phi_{j,k}(x)$$

und es gilt

$$\begin{aligned} \gamma_k 2^{-j/2} &= 2^{-j/2} \int_{\mathbb{R}} v(y) \phi(y - k) dy = 2^{-j/2} \int_{\mathbb{R}} u(x) \phi(2^j x - k) 2^j dx \\ &= \int_{\mathbb{R}} u(x) \phi_{j,k}(x) dx = \langle u, \phi_{j,k} \rangle. \end{aligned}$$

Damit folgt, dass  $u = \sum_{k \in \mathbb{Z}} \langle u, \phi_{j,k} \rangle \phi_{j,k}$  gilt, also ist  $\phi_{j,k}$  eine ONB von  $\mathcal{V}_j$ .  $\square$

Wir sehen also, dass wir durch  $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$  jede Funktion in  $\mathcal{V}_j$  darstellbar ist. Da die  $V_j$  gegen  $L^2(\Omega)$  konvergieren, erhält man die Konvergenz der Projektionen

$$P_j u = \sum_{k \in \mathbb{Z}} \langle u, \phi_{j,k} \rangle \phi_{j,k} \quad (1.35)$$

gegen die ursprüngliche Funktion  $u$  für  $j$  gegen unendlich. Für grösser werdendes  $j$  erhalten wir also immer bessere Approximationen von  $u$  auf kleiner werdenden Skalen.

Eine offene Frage ist aber noch die Existenz einer MRA, da wir ja eine Funktion  $\phi$  finden müssen, sodass die obigen drei Bedingungen erfüllt sind. Nun soll  $\phi \in \mathcal{V}_0$  gelten, und da  $\mathcal{V}_0 \subset \mathcal{V}_\infty$  gilt, folgt auch  $\phi \in \mathcal{V}_1$ . Der Raum  $\mathcal{V}_1$  hat die ONB  $\{\phi_{1,k}\}_{k \in \mathbb{Z}}$ , also muss es auch eine Darstellung der Funktion  $\phi$  bezüglich dieser ONB geben. Dies bedeutet

$$\phi(x) = \sum_{k \in \mathbb{Z}} \gamma_k \phi_{1,k}(x) = \sum_{k \in \mathbb{Z}} \sqrt{2} \gamma_k \phi(2x - k) \quad (1.36)$$

mit den Koeffizienten

$$\gamma_k = \int_{\Omega} \phi(x) \phi(2x - k) dx.$$

Die Gleichung (1.36) ist eine Funktionalgleichung für die Skalierungsfunktion. Es lässt sich zeigen, dass durch ein  $\phi$ , das (1.36) erfüllt, auch wirklich eine MRA definiert wird. Ein einfaches Beispiel für eine Skalierungsfunktion, die (1.36) erfüllt, ist die unstetige Funktion

$$\phi(x) = H(x)H(1 - x), \quad (1.37)$$

wobei  $H$  die Heaviside-Funktion ist. In diesem Fall gilt

$$\phi(x) = \phi(2x) + \phi(2x - 1).$$

Skalierungsfunktionen höher Glattheit (sogenannte Daubechies<sup>22</sup>-Wavelets.) erhält man, wenn man eine Lösung von (1.36) für eine höhere (aber endliche) Anzahl nichtverschwindender Koeffizienten  $\gamma_k$  sucht (cf. [5]).

Nun hat man durch die Folge der Räume  $V_j$  immer noch zuviel Redundanz, die von den Wavelets eliminiert wird. Wavelets entsprechen in der MRA den Details, die beim Übergang

---

<sup>22</sup>Ingrid Daubechies (\*1954), belgische Physikerin und Mathematikerin. Arbeitet in der Signal- und Bildverarbeitung. Professorin in Princeton, Mitglied der Academy of Sciences, cf. [12]

von einer Skala zur nächsten verloren gehen. Wegen  $\mathcal{V}_j \subset \mathcal{V}_{j+1}$  existiert ein Unterraum  $\mathcal{W}_j \subset \mathcal{V}_{j+1}$  mit der Zerlegung

$$\mathcal{V}_{j+1} = \mathcal{V}_j \oplus \mathcal{W}_j. \quad (1.38)$$

Wegen der Vollständigkeit gilt dann auch

$$L^2(\mathbb{R}) = \mathcal{V}_J \oplus \sum_{j \geq J} \mathcal{W}_j = \sum_{j=-\infty}^{\infty} \mathcal{W}_j. \quad (1.39)$$

In diesem Fall erhalten wir eine direkte Summe der Unterräume, d.h. die Redundanz in der Darstellung ist eliminiert. Die dyadische Ähnlichkeit bleibt vollständig erhalten, es gilt  $v(x) \in \mathcal{V}_j \Leftrightarrow v(2x) \in \mathcal{V}_{j+1}$ .

Da die Unterräume  $\mathcal{W}_0$  und  $\mathcal{V}_0$  aufeinander orthogonal stehen, folgt

$$\int_{\Omega} v(x)\phi(x - \ell) dx = 0, \quad \forall v \in \mathcal{W}_0, \forall \ell \in \mathbb{Z}.$$

Für  $v \in \mathcal{V}_0 \subset \mathcal{W}_0$  gibt es eine Darstellung der Form

$$v(x) = \sum_k \beta_k \sqrt{2} \phi(2x - k)$$

und  $\phi(x - \ell)$  kann wegen (1.36) als

$$\phi(x - \ell) = \sum_m \gamma_m \sqrt{2} \phi(2x - 2\ell - m)$$

geschrieben werden. Die obige Orthogonalitätsrelation wird dann zu

$$2 \sum_{k,m} \beta_k \gamma_m \int_{\Omega} \phi(2x - k) \phi(2x - 2\ell - m) dx = \sum_{k,m} \beta_k \gamma_m \langle \phi_{1,k}, \phi_{1,2\ell+m} \rangle = 0.$$

Da die  $\{\phi_{1,k}\}_{k \in \mathbb{Z}}$  eine ONB sind, ist  $\langle \phi_{1,k}, \phi_{1,2\ell+m} \rangle = 0$  für  $k \neq 2\ell + m$ . Also folgt

$$\sum_m \beta_{2\ell+m} \gamma_m = 0. \quad (1.40)$$

Bestimmt man den Nullraum dieser Gleichung, so lässt man sich zeigen, dass jede Funktion in  $\mathcal{W}_0$  in der Form

$$v(x) = \sum_k c_k \psi(x - k), \quad c_k = \int_{\Omega} v(x) \psi(x - k) dx \quad (1.41)$$

schreiben lässt, mit einer gegebenen Funktion  $\psi$  der Form

$$\psi(x) = \sum_k d_k \psi(2x - k). \quad (1.42)$$

Die Funktion  $\psi$  nennt man dann das eigentliche Wavelet, die Menge  $\{\psi(\cdot - k)\}_{k \in \mathbb{Z}}$  ist eine ONB von  $\mathcal{W}_0$ . Wegen der dyadischen Ähnlichkeit lässt sich wiederum zeigen, dass  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  eine ONB von  $\mathcal{W}_j$  ist, wobei  $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$ . Damit erhalten wir bezüglich dem Wavelet  $\psi$  eine Multiskalendarstellung ohne jegliche Redundanz, die Menge  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  ist

eine ONB von  $L^2(\mathbb{R})$ . Die Koeffizienten der Darstellung einer Funktion  $u$  bezüglich dieser ONB

$$C_{jk} = \langle u, \psi_{jk} \rangle = 2^{j/2} \int_{\Omega} u(x) \psi(2^j x - k) dx$$

nennt man auch die Wavelet-Koeffizienten.

Als Beispiel betrachten wir das *Haar*<sup>23</sup>-Wavelet, das wir aus der Skalierungsfunktion  $\phi(x) = H(x)H(1-x)$  erhalten. In diesem Fall haben wir  $\gamma_0 = \gamma_1 = 1$  und damit wird (1.40) zu  $\beta_{2\ell} + \beta_{2\ell+1} = 0$ . Also kann jede Funktion in  $\mathcal{W}_0$  in der Form ( $c_\ell = \frac{\beta_{2\ell}}{\sqrt{2}}$ )

$$v(x) = \sum_{\ell} c_{\ell} \sqrt{2} (\phi(2x - 2\ell) - \phi(2x - 2\ell - 1)).$$

Nach Normalisierung ( $\|\psi\| = 1$ ) erhält man das Haar-Wavelet

$$\psi(x) = \sqrt{2}(\phi(2x) - \phi(2x - 1)) = \sqrt{2}(H(2x)H(1-2x) - H(2x-1)H(2-2x)). \quad (1.43)$$

Besonders interessant ist die Wavelet-Darstellung von Funktionen mit beschränktem Träger (natürlich nur für Wavelets mit beschränktem Träger). In diesem Fall ist nämlich nur ein kleiner Teil der Wavelet-Koeffizienten wichtig, da für fast alle Werte von  $k$  die Träger von  $u$  und  $\psi_{j,k}$  disjunkt sind. Wir illustrieren dies wieder für das Haar-Wavelet und nehmen der Einfachheit halber an der Träger der Funktion  $u$  sei eine Teilmenge des Intervalls  $[0, 1]$  (andernfalls können wir das Gebiet oder die Wavelet-Basis passend reskalieren). Es ist dann völlig natürlich mit dem Raum  $\mathcal{V}_0$  zu beginnen, da die Skalierungsfunktion auf dem Intervall  $[0, 1]$  konstant ist und keine größeren Skalen auftreten. Es gilt dann

$$\langle u, \phi \rangle = \int_0^1 u(x) dx, \quad \langle u, \phi(\cdot - k) \rangle = 0, k \neq 0.$$

Auf der größten Skala speichern wir also nur den Mittelwert von  $u$ , alle anderen Koeffizienten müssen nicht gespeichert werden. Im Raum  $\mathcal{W}_0$  haben wir dann ebenfalls nur eine Funktion, deren Träger in  $[0, 1]$  liegt, es gilt  $C_{00} = 1$  und  $C_{0k} = 0$  für alle  $k \neq 0$ . Im Raum  $\mathcal{W}_{\infty}$  hat die Funktion  $\psi_{10}$  den Träger  $[0, \frac{1}{2}]$  und die Funktion  $\psi_{11}$  den Träger  $[\frac{1}{2}, 1]$ . Der Träger aller anderen Basisfunktion  $\psi_{1k}$  liegt ausserhalb des Intervalls  $[0, 1]$ , diese führen also zu verschwindenden Wavelet-Koeffizienten. Also müssen auf der ersten Skala nur die Koeffizienten  $C_{10}$  und  $C_{11}$  gespeichert werden. Analog folgern wir, dass auf der  $j$ -ten Skala nur die Koeffizienten  $C_{j0}, C_{j1}, \dots, C_{j,2^j-1}$  von Null verschieden sein können. Damit müssen wir zur Darstellung auf der  $j$ -ten Skala  $2^j$  Wavelet-Koeffizienten speichern, insgesamt zur Darstellung der größten  $j$  Skalen genügen  $1 + 1 + 2 + 4 + \dots + 2^j = 2^{j+1}$  Koeffizienten.

Im Allgemeinen werden von den Wavelet-Koeffizienten noch viele sehr klein sein, da ja nicht auf allen Skalen signifikante Objekte im Bild auftreten. Durch Weglassen der Wavelet-Koeffizienten mit zu kleinem Absolutbetrag erreicht man oft grosse Kompressionsraten bei geringem Qualitätsverlust. Dies ist die Basis des JPEG Formats für Bilder (wird über diskrete Cosinus-Transformationen realisiert, die sich aber als Haar-Wavelet interpretieren lassen). Eine explizite Wavelet-Kompression erfolgt im moderneren, aber noch wenig verbreiteten JPEG-2000 Format (Dateien .jp2, .jk2, .jpf).

<sup>23</sup> Alfred Haar (1885-1933) ungarischer Mathematiker. Professor in Szeged. Arbeitete über partielle Differentialgleichungen, Approximationstheorie und topologische Gruppen. Nach ihm ist auch das Haar-Maß auf lokalkompakten Gruppen, cf. [12]