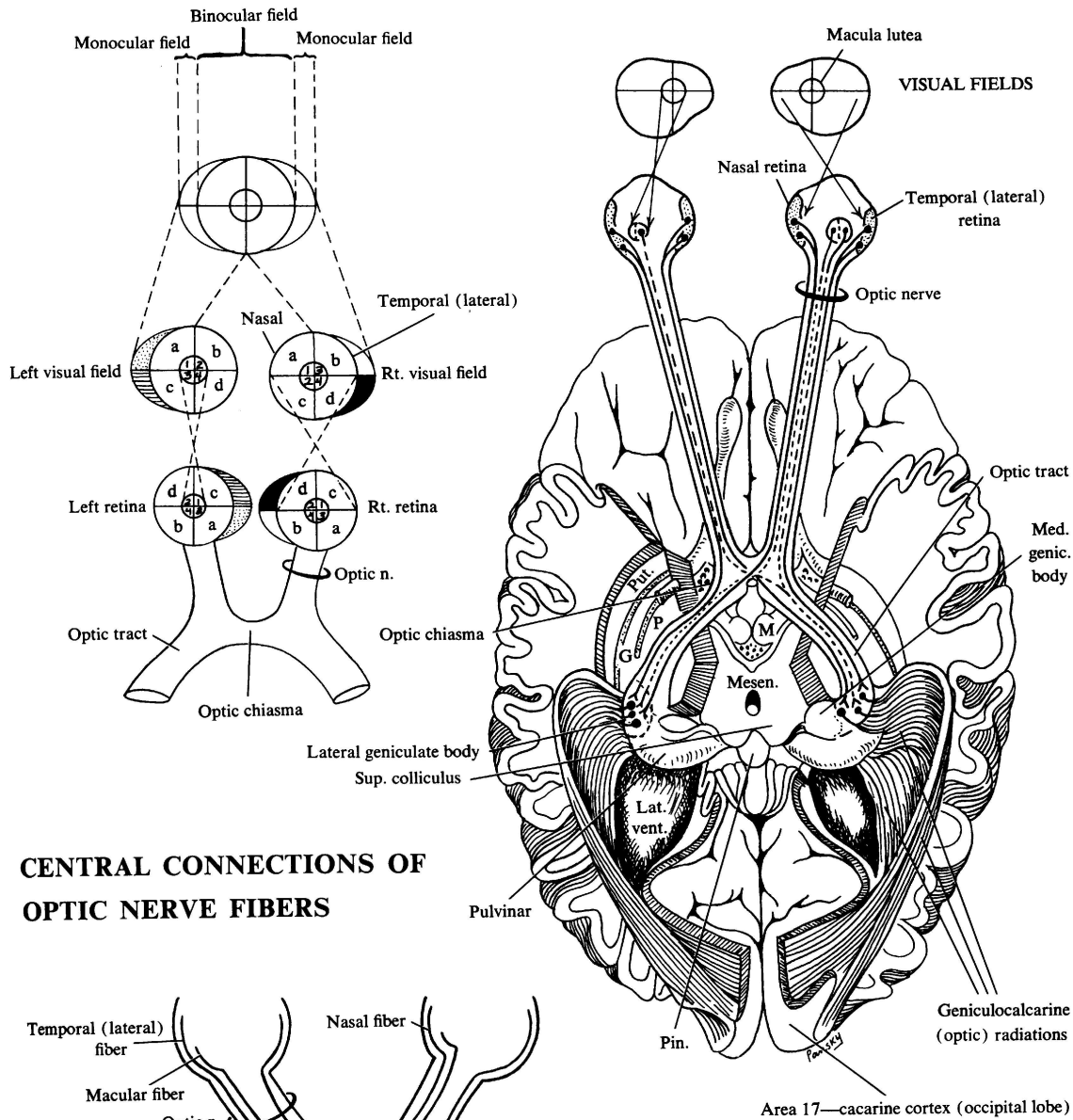


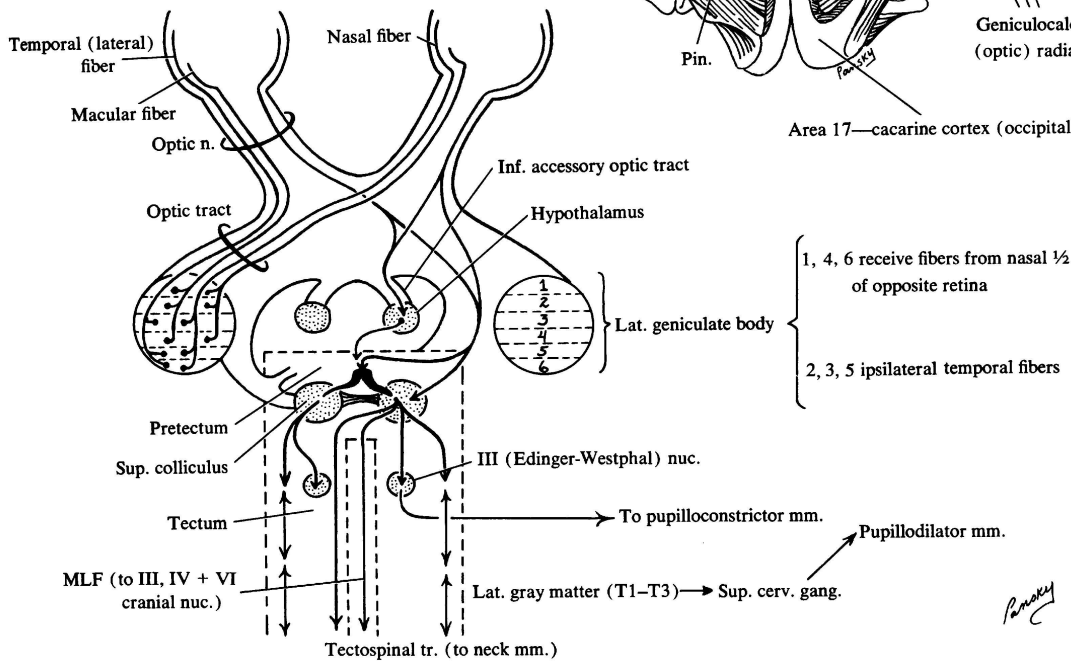
Aufgaben und Ziele der Bildverarbeitung

Photos	Qualitätsverbesserung, Herausheben von Details
OCR	Schrift-, Mustererkennung
Video	Bewegungsanalyse, Personenerkennung
Video	Erkennung von Gesten
Industrie	Automatisierte Fertigung
Industrie	Prozeßüberwachung, Robotik
Medizin	Erkennung pathologischer Prozesse
Medizin	Computer Aided Surgery
Forschung	Visualisierung nicht-sichtbarer Größen
Militär	Erkennung von Objekten, "Freund"- "Feind"
Militär	Steuerung von Flugzeugen und Waffen
Werbung, Kunst	Desktop Publishing

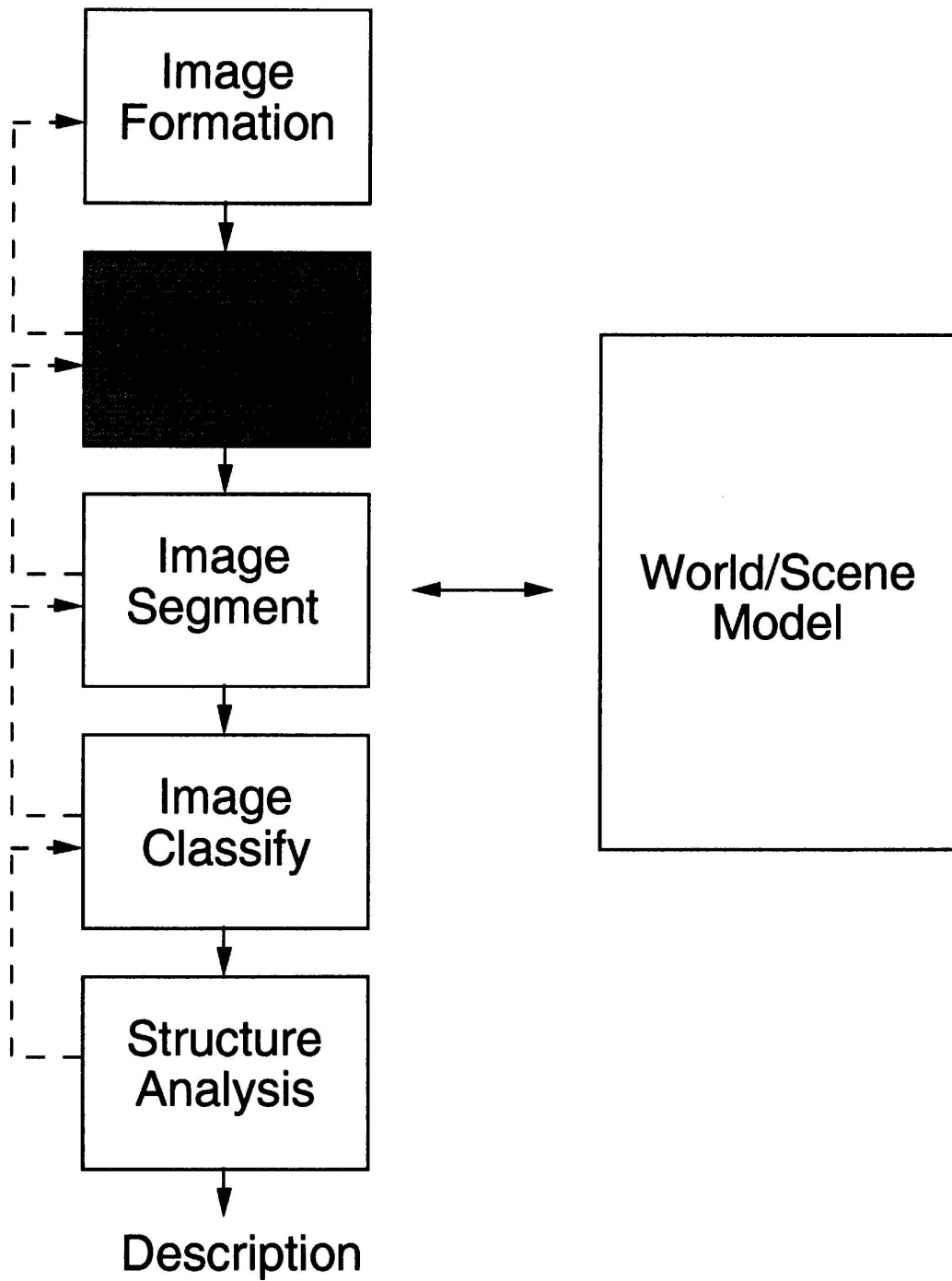
Modell: Das visuelle System



CENTRAL CONNECTIONS OF OPTIC NERVE FIBERS



Bildanalyse-System



Bildfunktionen

Ein Bild ist eine kontinuierliche Funktion $f(\cdot)$ von (i.a.) zwei bis vier Variablen, wobei zwei (drei) Variablen (x,y,z) den Ort, die dritte (vierte) die Zeit (t) bezeichnet. Beispiele:

$f(x, y)$	Photo
$f(x, y, t)$	Video
$f(x, y, z)$	Volumendatensatz aus der MR-Tomographie
$f(x, y, z, t)$	Zeitreihe eines Volumendatensatzes

Der Funktionswert $f(\cdot)$ entspricht dabei einer beliebigen physikalischen Größe (z.B. Intensität, Feldstärke, Niederschlag, Teilchendichte, ...).

Digitale Bilder

Digitale Bilder sind diskrete Repräsentationen einer Größe auf einem finiten räumlichen Gitter:

$$\begin{aligned}x_1 &\leq x \leq x_2 \\ y_1 &\leq y \leq y_2 \\ z_1 &\leq z \leq z_2 \\ t_1 &\leq t \leq t_2\end{aligned}$$

Wir werden im allgemeinen nur Bilder betrachten, bei denen die räumlichen Argumente auf ganzzahlige Werte beschränkt sind:

$$\begin{aligned}x &= n_1 \Delta x & n_1 &\in \mathbb{Z} \\ y &= n_2 \Delta y & n_2 &\in \mathbb{Z} \\ z &= n_3 \Delta z & n_3 &\in \mathbb{Z} \\ t &= n_4 \Delta t & n_4 &\in \mathbb{Z}\end{aligned}$$

Dabei werden $\Delta x, \Delta y, \Delta z$ als "Gitterweite" oder "Abtastintervall" bezeichnet. Δt wird "Zeitschritt" genannt. Die reziproken Größen $1/\Delta x, \dots$ werden auch "Auflösung" genannt.

Auch die Funktionswerte sind zumeist diskret (z.B. Grauwerte, Teilchenzahl).

Die Dirac-Verteilung

Ein idealer Impuls in der Bildebene wird durch die Dirac-Verteilung $\delta(x,y)$ definiert:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x,y) dx dy = 1,$$

wobei $\delta(x,y) = 0 \quad \forall \quad x,y \neq 0$.

Die Dirac-Verteilung kann im Ort verschoben werden:

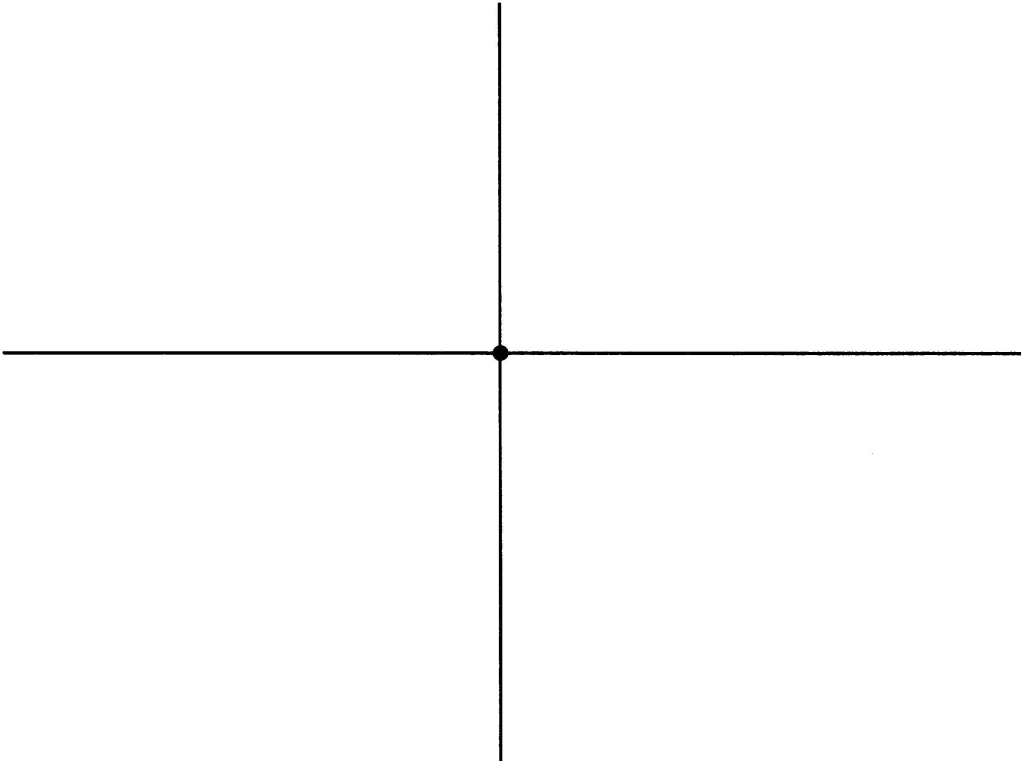
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \delta(x-\lambda, y-\mu) dx dy = f(\lambda, \mu),$$

womit der Funktionswert $f(x,y)$ am Ort λ, μ erhalten werden kann.

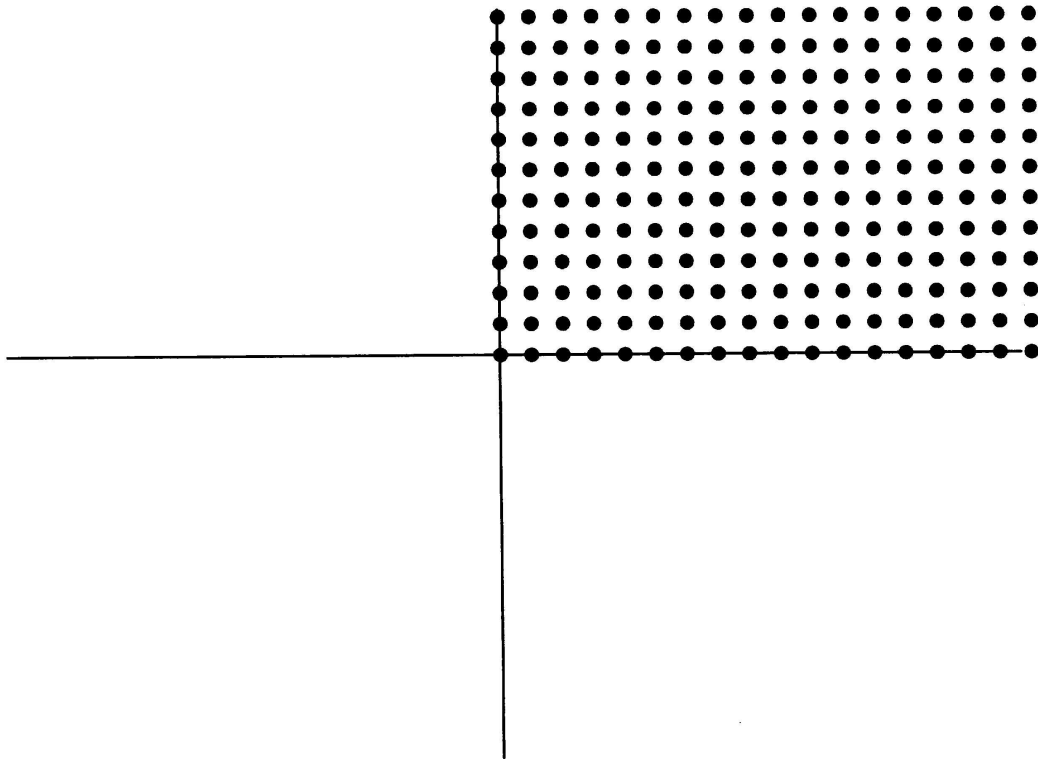
Ein Bild kann dann als Linearkombination von Dirac-Pulsen an allen Stellen a, b der Bildebene beschrieben werden, die durch die Bildfunktion $f(x,y)$ gewichtet sind:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a,b) \delta(a-x, b-y) da db = f(x,y).$$

Bild einer Dirac-Verteilung bei $\lambda = \mu = 0$



Linearkombination von Dirac-Pulsen



Konvolution

Die Konvolution (oder Faltung) ist eine wichtige Operation im linearen Ansatz der Bildverarbeitung. Die Konvolution $g(\cdot)$ der zweidimensionalen Funktionen $f(\cdot)$ und $h(\cdot)$ ist durch das Integral definiert:

$$\begin{aligned} g(x,y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a,b)h(x-a,y-b)dad b \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-a,y-b)h(a,b)dad b \\ &= (f \star h)(x,y) = h(\star f)(x,y), \end{aligned}$$

wobei das Zeichen $\star$ den Konvolutionsoperator bezeichnet. Die Konvolution ist translationsinvariant. Sie beschreibt einen linearen Filterprozeß mit dem Filter $h(\cdot)$.

Abtastung

Ein kontinuierliches (zweidimensionales) Bild kann auf ein diskretes Gitter von Abtastpunkten in der Ebene abgebildet werden. Das Bild wird im Abstand $x = n_1\Delta x, y = n_2\Delta y$ mit $n_1 = 1 \dots N_1, n_2 = 1 \dots N_2$ in den Intervallen $\Delta x, \Delta y$ abgetastet. Die Matrix der Werte $f(n_1\Delta x, n_2\Delta y)$ entsprechen dem diskreten Bild. Die ideale Abtastung $s(x, y)$ in einem regulären Gitter kann durch eine Summe von Dirac-Verteilungen repräsentiert werden:

$$s(x, y) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \delta(x - n_1\Delta x, y - n_2\Delta y).$$

Das abgetastete Bild $f_s(x, y)$ ist das Produkt des kontinuierlichen Bildes $f(x, y)$ und der Abtastfunktion $s(x, y)$:

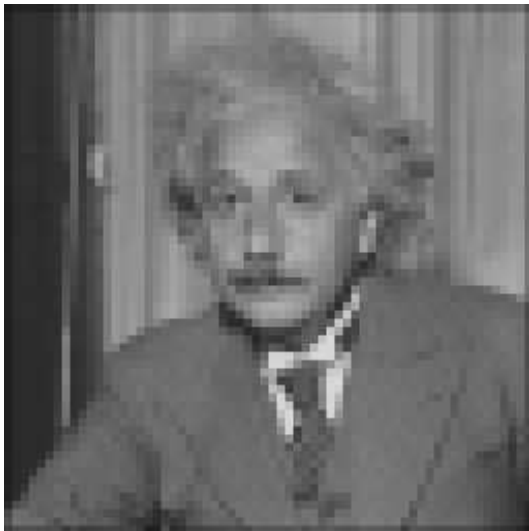
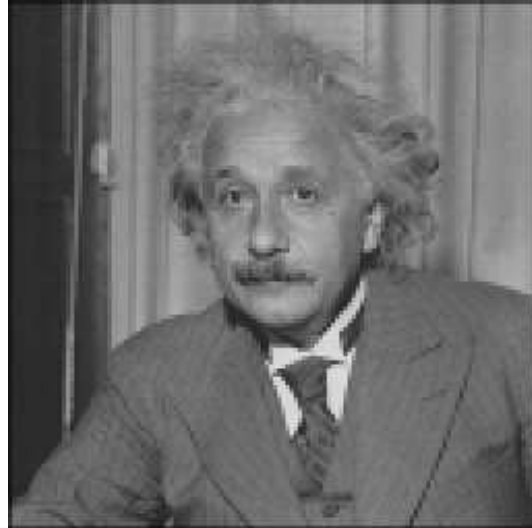
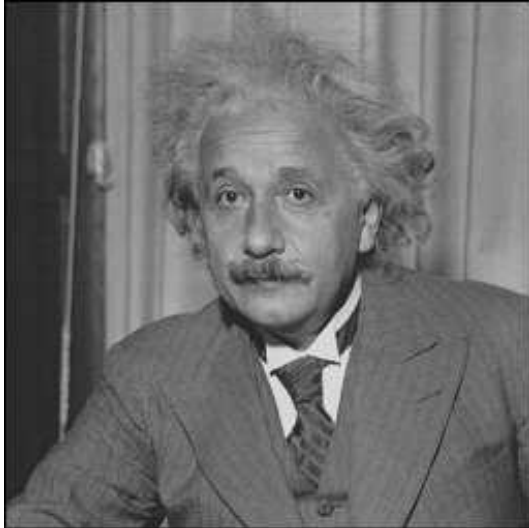
$$\begin{aligned} f_s(x, y) &= (fs)(x, y) \\ &= \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} f(x, y) \delta(x - n_1\Delta x, y - n_2\Delta y) \\ &= f(x, y) \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \delta(x - n_1\Delta x, y - n_2\Delta y). \end{aligned}$$

In der Praxis weist die Abtastfunktion andere spektrale Eigenschaften als die Dirac-Verteilung auf; insbesondere können die Verteilungen "überlappen". Hierdurch entsteht eine "Verschmierung" des Bildes, die "Aliasing" genannt wird.

Ein (unendlich kleiner) Abtastpunkt im Gitter entspricht einem Bildelement ("Pixel" oder "Voxel").

Abtastung - Beispiel

Das Bild links oben wurde mit $\Delta x = \Delta y = 2$ (rechts oben) bzw. 4 (links unten) und 8 (rechts unten) abgetastet.



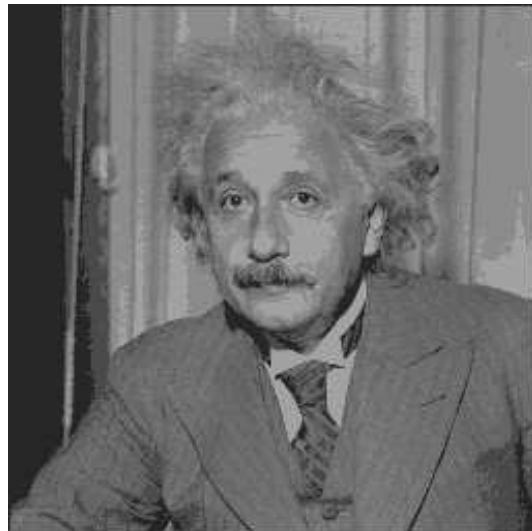
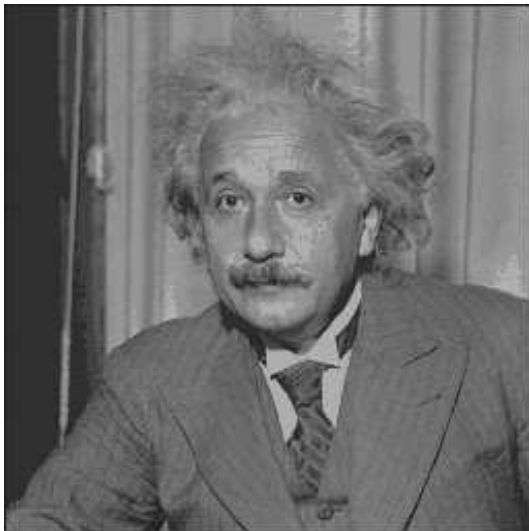
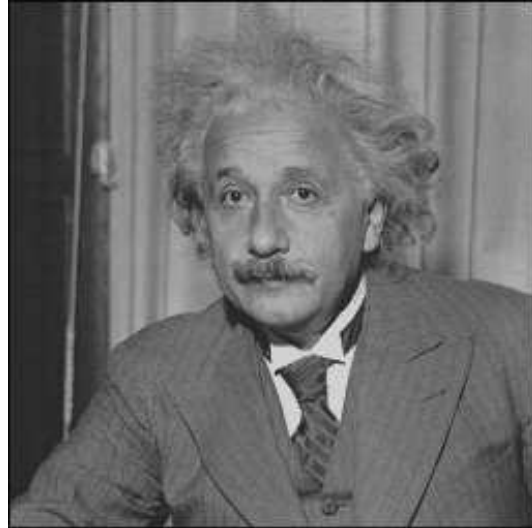
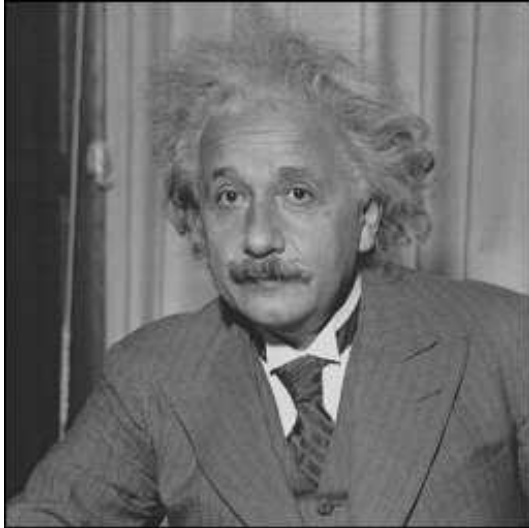
Quantisierung

Die Größe $f_s(x,y)$ wird als digitaler Wert in der Bildverarbeitung ausgedrückt. Der Übergang von den kontinuierlichen Werten der Bildfunktion zu seinem digitalen Äquivalent wird als "Quantisierung" bezeichnet.

Digitale Bilder werden meist mit k gleichen Intervallen quantisiert. Wenn b bits benutzt werden, um die Intensität auszudrücken, so ist die Anzahl der verfügbaren Intensitätsstufen $k = 2^b$. Der Mensch kann etwa 64 Intensitätsstufen unterscheiden, so daß zumeist 8 bit benutzt werden, um ein Grauwertbild darzustellen.

Quantisierung - Beispiel

Das Bild links oben wurde mit 64 (rechts oben) bzw. 16 (links unten) und 8 (rechts unten) Intensitätsstufen quantisiert.



Metrische Eigenschaften von digitalen Bildern

Manche einleuchtenden Eigenschaften kontinuierlicher Bilder haben kein Äquivalent in der Domäne der digitalen Bilder - die Distanz zwischen zwei Pixeln ist ein wichtiges Beispiel.

Der Abstand zwischen zwei Punkten (i, j) und (h, k) kann auf unterschiedliche Weise definiert werden:

Die Euklidische Distanz $D_E(\cdot)$ wird definiert als:

$$D_E((i, j), (h, k)) = \sqrt{(i - h)^2 + (j - k)^2}.$$

Die "City-Block"-Distanz $D_4(\cdot)$ wird definiert als:

$$D_4((i, j), (h, k)) = |i - h| + |j - k|.$$

Die "Schachbrett"-Distanz $D_8(\cdot)$ wird definiert als:

$$D_8((i, j), (h, k)) = \max\{|i - h|, |j - k|\}.$$

Zwei Pixel in der Ebene nennt man vierfach-verbunden, wenn sie die Distanz $D_4 = 1$ besitzen (also eine Kante gemeinsam haben).

Zwei Pixel in der Ebene nennt man achtfach-verbunden, wenn sie die Distanz $D_8 = 1$ besitzen (also eine Ecke oder Kante gemeinsam haben).

Pfade und Regionen

Ein **Pfad** von Pixel P zu Pixel Q ist eine Sequenz von Punkten $A_1, A_2, \dots, A_n$ mit $P = A_1$ und $Q = A_n$, wobei A_{i+1} ein Nachbar von A_i ist, für alle $i = 1, \dots, n-1$. Gibt es einen Pfad von P nach Q , werden sie als **zusammenhängend** bezeichnet.

Ein **einfacher Pfad** ist ein Pfad, bei dem kein Punkt wiederholt wird (außer ggf. den Endpunkten), bei dem kein Punkt mehr als zwei Nachbarn hat.

Ein **geschlossener Pfad** ist ein Pfad, bei dem der erste Punkt ein Nachbar des letzten ist.

Eine **Region** ist eine Menge von Punkten, für die es zwischen jedem Paar von Punkten einen Pfad gibt, dessen Punkte ebenso zur Menge gehören.

Oder:

Eine **Region** ist eine Menge von Punkten, in der jedes Pixel-paar zusammenhängend ist.

Eine Region in einem Bild wird oft auch als **Objekt** bezeichnet. Der Prozeß, der Bildregionen einem Objekt zuordnet, wird **Segmentation** genannt.

Der **Rand** einer Region ist die Menge der Pixel innerhalb der Region, die einen oder mehrere Nachbarn außerhalb der Region besitzen.

Histogramm

Das Intensitätshistogramm $h_f(z)$ eines Bildes gibt die Häufigkeit des Intensitätswertes z im Bild wieder - ein Histogramm eines Bildes mit Z Intensitätsstufen wird durch ein ein-dimensionales Feld mit Z Zellen repräsentiert.

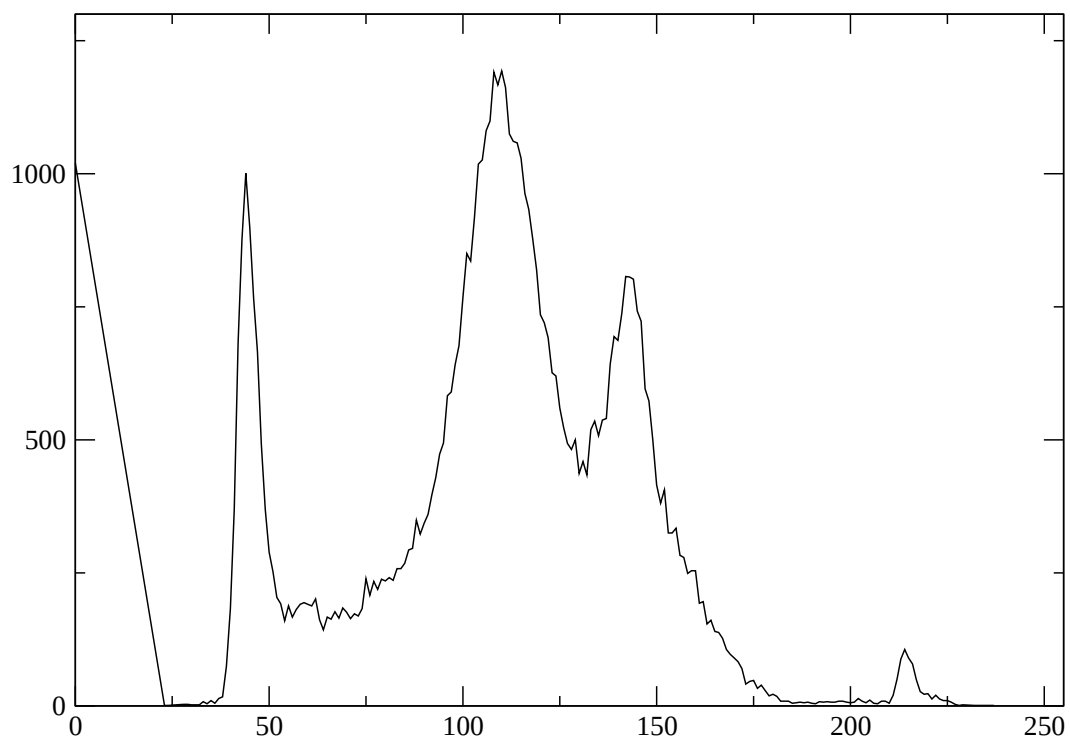
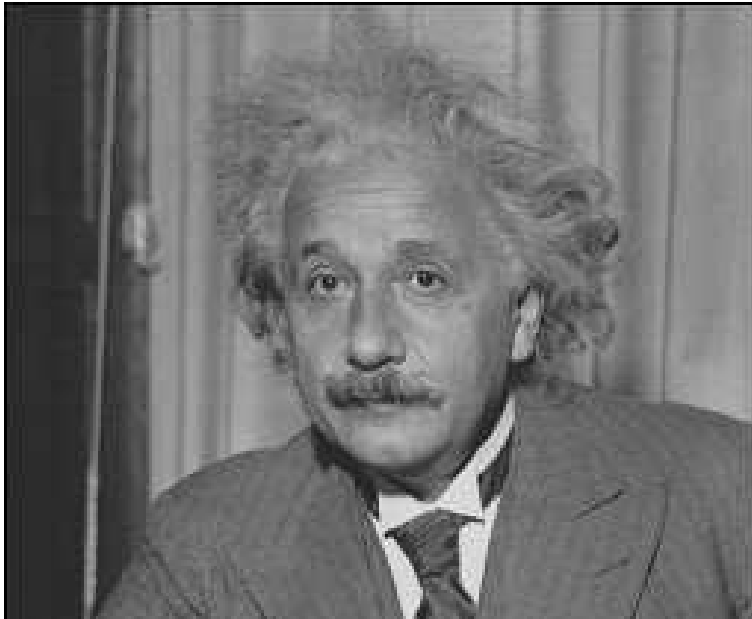
Algorithmus I: Berechnung eines Intensitätshistogramms

1. Setze alle Zellen des Feldes $h_f(z)$ auf null.
2. Für alle Pixel (x,y) des Bildes f , erhöhe $h_f(f(x,y))$ um eins.

Wird ein Bild als ein Ergebnis eines stochastischen Prozesses verstanden, so kann das Histogramm als Abschätzung der Dichtefunktion $p_1(z)$ benutzt werden. Hiermit kann die **Entropie** H eines Bildes berechnet werden:

$$H = - \sum_{z=1}^Z p_1(z) \log_2 (p_1(z)).$$

Histogramm - Beispiel



Wahrnehmung von Bildern

Unter **Kontrast** wird eine lokale Änderung der Intensität verstanden. Sie wird als das Verhältnis von mittlerer Intensität eines Objekts zur Intensität des Hintergrundes definiert.

Schärfe bezeichnet die Fähigkeit, Details im Bild zu erkennen. Bei einer Beleuchtung von 500 Lux (z.B. durch eine 60 Watt-Birne im Abstand von 40 cm) kann ein normalsichtiger Mensch zwei Objekte in einer Entfernung von 25 cm als getrennt erkennen, wenn sie mehr als 0.16 mm Abstand haben.

Rauschen

Reale Bilder sind oft durch zufällige Einflüsse gestört - dieser Einfluß wird als **Rauschen** bezeichnet. Rauschen kann durch die Bilderzeugung, die Übertragung oder durch die Bearbeitung entstehen. Es kann vom Bildinhalt abhängig oder unabhängig sein.

Rauschen wird üblicherweise durch seine Zufallsverteilung beschrieben. Das idealisierte Rauschen wird "weißes Rauschen" genannt; es besitzt eine konstante spektrale Verteilung. Es stellt den "worst case" einer Rauschquelle dar.

Eine häufig gewählte Rauschcharakteristik ist das Gauß'sche Rauschen. Seine Dichteverteilung entspricht einer Gaußfunktion:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

wobei μ der Mittelwert und σ die Standardabweichung der Rauschverteilung sind. Das Gauß'sche Rauschen stellt eine gute Annäherung an das natürliche Rauschen dar.

In der Nuklearmedizin und der MR-Tomographie ist das Rauschen anders zu beschreiben (durch eine Poisson- bzw. Rayleigh-Verteilung).

Ist das Signal $g(\cdot)$ vom Rauschen $v(\cdot)$ unabhängig, so wird es als **additives Rauschen** bezeichnet:

$$f(x,y) = g(x,y) + v(x,y).$$

Ist die Rauschamplitude dagegen von der Signalamplitude abhängig, so spricht man von **multiplikativem Rauschen**:

$$f = g + vg = g(1 + v) \approx gv.$$

Repräsentation von Bilddaten

Ikonische Bilder stammen aus Originaldaten (z.B. Matrizen mit Werten, die Intensitäten entsprechen) oder sind Ergebnisse von Bildverarbeitungsprozeduren (z.B. Filter, 3D-Rendering).

In **Segmentbildern** sind Teile eines Bildes in Regionen zusammengefaßt, die vermutlich zu demselben Objekt gehören. Es ist sinnvoll, bei der Segmentation Vorwissen über die Anwendungsdomäne zu verwenden. Dies erleichtert die Behandlung von Rauschen, Artefakten und Bildfehlern.

Geometrische Repräsentationen enthalten Wissen über 2D und 3D Formen. Die Quantifizierung von Formen ist schwierig, aber für viele Anwendungen wichtig.

Relationale Modelle ermöglichen eine effiziente Behandlung von Daten auf einer höheren Abstraktionsstufe. Zumeist wird Vorwissen benutzt, um Probleme auf dieser Ebene zu lösen.

Beispiele:

- Flugzeuge auf einem Flughafen sollen auf der Basis von Luftaufnahmen gezählt werden. Hierbei kann Vorwissen über die Position und Ausdehnung des Gebäudes im Verhältnis zu anderen Objekten (Straßen, Teiche, Dörfer) sowie die Struktur der Flugzeuge verwendet werden.
- Körperorgane sollen aus 3D-Tomographien segmentiert werden. Dabei kann ihre relative Lage verwendet werden, um ihren Erwartungsraum einzuschränken.

Datenstrukturen: Matrizen

Die häufigste Repräsentationsform von digitalen Bildern sind 2D (3D, 4D) Matrizen. Elemente dieser Matrizen sind z.B.:

- *binäre Werte*: S/W-Druckbilder, Zwischenprodukte von BV-Prozeduren
- *ganzzahlige Werte*: CCD-Bilder (Kamera, Scanner), Tomographen
- *Realzahlen*: gemessene physikalische Größen wie Druck, Temperatur
- *komplexe Zahlen*: Fourier-transformierte Bilder, k-Raum-Bilder aus der Tomographie
- *Vektoren, Tensoren*: Zwischenprodukte von BV-Prozeduren

```
template <class T> class iconic_image {
protected:
    T      *p;          // Zeiger auf Datenfeld
    int    nx, ny;      // Feldgrenzen

public:
    iconic_image(int _nx, int _ny) : nx(_nx), ny(_ny)
    { p = new T [nx*ny]; init(); };
    iconic_image(iconic_image const &a)
    { nx = a.nx; ny = a.ny; p = new T [nx*ny];
      memcpy(p, a.p, nx*ny*sizeof(T)); };
    ~iconic_image() { delete p; };
    T& operator() (int x, int y) const
    { return p[y*nx+x]; };
    iconic_image<T>& operator=(const iconic_image<T>& a)
    { delete p; nx = a.nx; ny = a.ny; p = new T [nx*ny];
      memcpy(p, a.p, nx*ny*sizeof(T)); return *this; };
    void init()
    { for (int i = 0; i < nx*ny; i++) p[i] = T(0.0); };
    ...
};
```

Datenstrukturen: Ketten

Ketten werden zur Repräsentation von Objektgrenzen benutzt. Symbole in einer Kette kodieren die Nachbarschaftsbeziehungen der Bildpunkte.

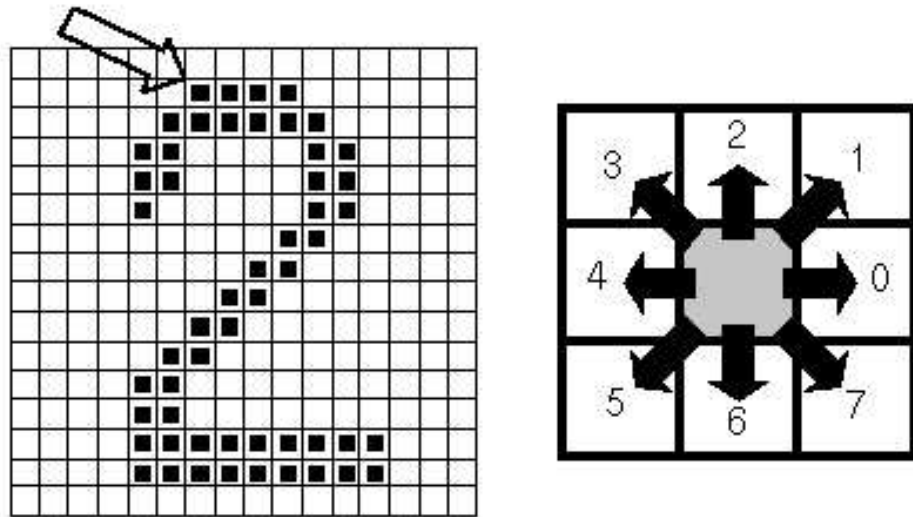


Figure 3.1 An example chain code; the reference pixel is marked by an arrow:
00007766555555660000000644444442221111112234445652211.

Ketten sind durch (dynamische) 1D-Felder repräsentiert. Lokale Informationen sind einfach durch Suche zu erhalten. Globale Informationen (z.B. Flächeninhalt) sind schwierig zu berechnen.

Datenstrukturen: Lauflängenkodierung

Die Lauflängenkodierung (run length encoding - RLE) wird oft zur Kodierung von Symbolen verwendet (z.B. in Faxmaschinen). In Binärbildern werden nur Objekte kodiert. Jede Bildzeile wird durch eine Liste beschrieben, deren erstes Element die Bildzeile ist, danach folgen Koordinatenpaare, die dem Beginn und dem Ende des Objektes entsprechen. RLE kann auch zur Kodierung von Grauwertbildern verwendet werden.

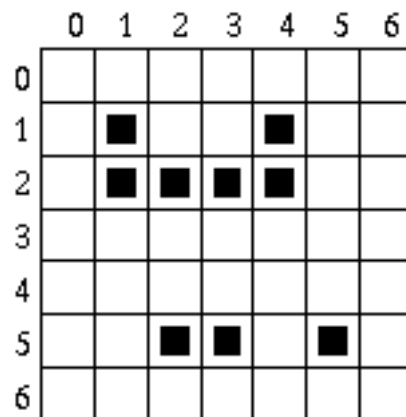


Figure 3.2 Run length coding; the code is $((11144)(214)(52355))$.

Datenstrukturen: Topologische Strukturen

Ein Bild kann auch als eine Menge von Elementen und deren Beziehungen beschrieben werden. Repräsentationsformen sind z.B. Graphen, gewichtete Graphen und Nachbarschaftsgraphen.

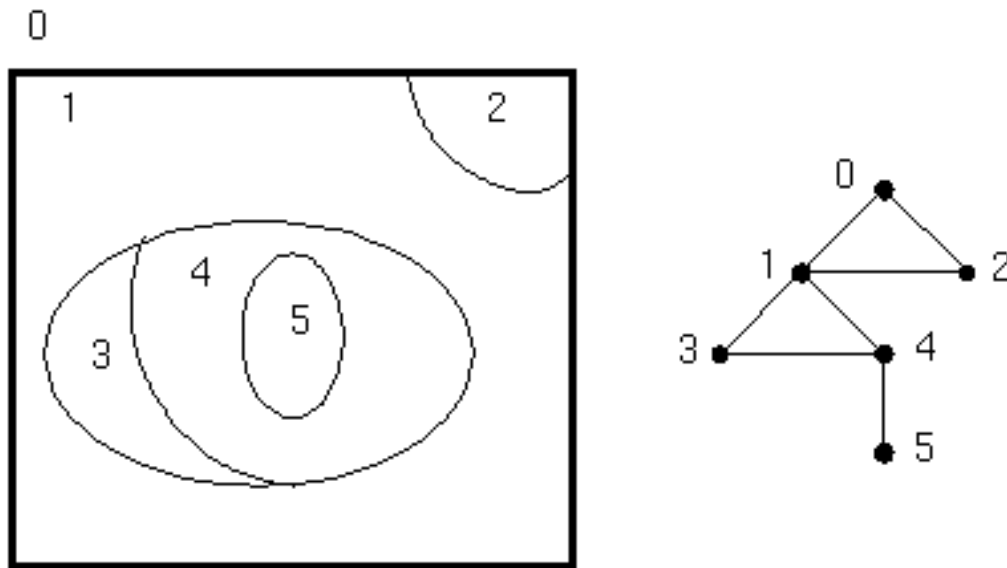


Figure 3.3 An example region adjacency graph.

Beispiel:

In einer Serie von mikroskopischen Schnittbildern wurden die Grenzen einer Zelle als Konturen markiert, die z.B. als Ketten repräsentiert sind. Ein Nachbarschaftsgraph stellt die räumlichen Zusammenhänge (also die topologischen Beziehungen) zwischen den Konturen der Schichten dar. Zwischen zwei Schichten können sich Konturen aufspalten und sich vereinigen. Ein solcher Graph wird auch **Reeb-Graph** genannt.

Datenstrukturen: Relationale Strukturen

Die Bildinformation kann auch als Relation zwischen symbolisch wichtigen Bildbereichen erfaßt werden. Solche Bereiche entsprechen Objekte als Ergebnis einer Segmentation. Relationale Strukturen sind für die symbolische Bildverarbeitung geeignet.

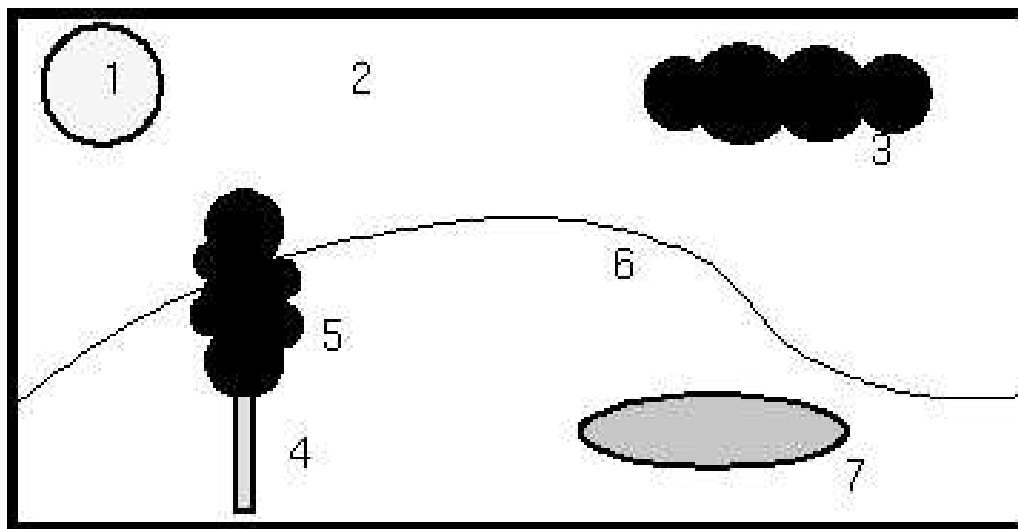


Figure 3.4 Description of objects using relational structure.

Nr.	Objektname	Farbe	min. Zeile	min. Spalte	Innerhalb
1	Sonne	weiß	5	40	2
2	Himmel	blau	0	0	-
3	Wolke	grau	20	180	2
4	Baumstamm	braun	95	75	6
5	Baumkrone	grün	53	63	-
6	Hügel	hellgrün	97	0	-
7	Teich	blau	100	160	6

Datenstrukturen: Hierarchische Strukturen

Die BV ist rechenintensiv, allein durch die Menge der zu verarbeitenden Daten. Eine (simple) Lösung ist die Nutzung von Parallelrechnern. Allerdings sind viele Algorithmen der BV schwierig zu partitionieren.

Hierarchische Strukturen ermöglichen die Nutzung von Bearbeitungsstrategien, die auf einer relativ kleinen Datenmenge basieren. Sie arbeiten nur dann auf der höchsten Bildauflösungsstufe, wenn es erforderlich ist.

Datenstrukturen: Auflösungspyramiden

Eine **Auflösungspyramide** (oder M-Pyramide) ist eine Sequenz $\{M_L, M_{L-1}, \dots, M_0\}$ von Bildern. M_L besitzt dieselbe Auflösung wie das Originalbild. M_{i-1} leitet sich von M_i durch Halbierung der Auflösung ab.

M-Pyramiden werden benutzt, wenn mit Bildern auf verschiedenen Auflösungsstufen gearbeitet werden soll. M_{i-1} enthält viermal weniger Daten als M_i , so daß Berechnungen etwa viermal schneller erfolgen.

Die Anzahl von Bildpunkten einer M-Pyramide ist

$$N^2(1 + \frac{1}{4} + \frac{1}{16} + \dots) \approx 1.33N^2.$$

Beispiel: siehe Vorlesung 01 - Abtastung - Beispiel

```
template <class T> class iconic_pyramid {
protected:
    iconic_image<T> *p;    // Feld mit Bildern
    int level;             // Anzahl der Stufen

public:
    ...
    void downscale(int i);
    void upscale(int i);
    ...
};
```

Eine Auflösungspyramide kann auch mehrere Bildbereiche auf einer Stufe getrennt enthalten. Diese Bildbereiche können geometrisch oder inhaltlich (z.B. als Segmente) definiert sein.

Datenstrukturen: Quadtrees

In einem **Quadtree** ist ein Bild in einer Baumstruktur repräsentiert. Jeder Knoten des Baumes (außer den Blättern) hat vier Nachfolger, die den vier Quadranten des Bildbereichs entsprechen (NW: Nordwest, NO: Nordost, SW: Südwest, SO: Südost). Wenn ein Vorfahr vier Nachfolger mit derselben Intensitätsstufe besitzt, braucht er nicht gespeichert werden.

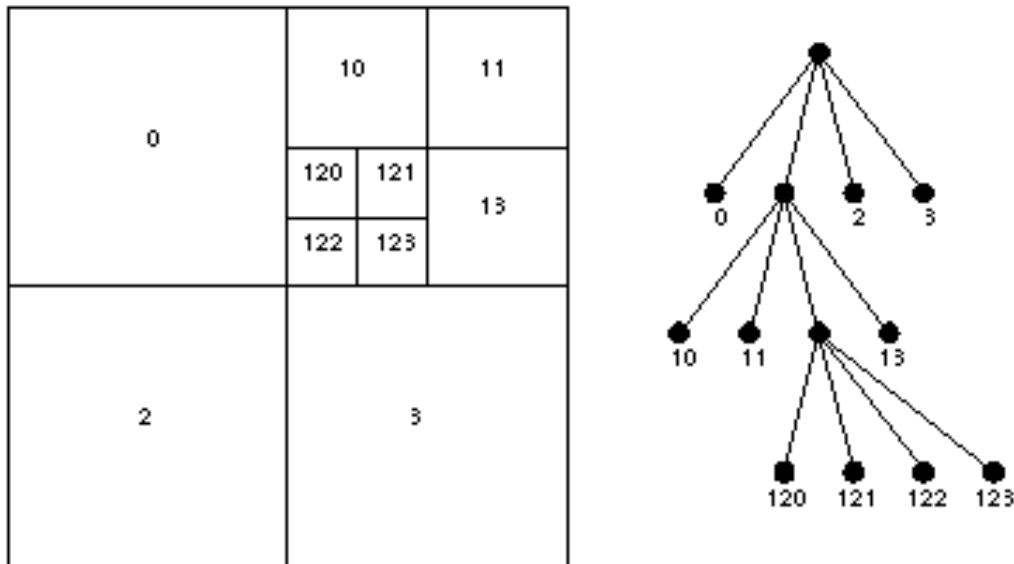


Figure 3.6 Quadtree.

Probleme:

Die Repräsentation eines Objektes hängt von der Position, Größe und Ausrichtung ab. Zwei ähnliche Bilder können sehr verschiedene Quadrees besitzen (z.B. bei einer kleinen Verschiebung eines Objektes).

Transformation von Bildern

Die Analyse von Bildern mit kontinuierlichen oder diskreten Bildtransformationen ist eine grundlegende Verarbeitungstechnik. Transformationen finden eine breite Verwendung in der Filterung, Bildkompression, Bildanalyse usw.

Ein Bild $f(\cdot)$ sei als eine Matrix von $M \times N$ Werten repräsentiert:

$$f = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}.$$

Die allgemeine Transformation $\mathbf{F} = \mathbf{P} \mathbf{f} \mathbf{Q}$ kann umgeschrieben werden als:

$$F(u,v) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} P(u,m) f(m,n) Q(n,v)$$
$$u = 0, 1, \dots, M-1; \quad v = 0, 1, \dots, N-1.$$

Sind $\mathbf{P}$ und $\mathbf{Q}$ nicht singulär, existieren die inversen Matrizen, und es gilt $\mathbf{f} = \mathbf{P}^{-1} \mathbf{F} \mathbf{Q}^{-1}$.

Wenn $\mathbf{P}$ und $\mathbf{Q}$ beide symmetrisch ($\mathbf{M} = \mathbf{M}^T$), real und orthogonal ($\mathbf{M} \mathbf{M}^T = \mathbf{I}$) sind, dann gilt:

$$\mathbf{F} = \mathbf{P} \mathbf{f} \mathbf{Q} \quad \text{und} \quad \mathbf{f} = \mathbf{P} \mathbf{F} \mathbf{Q},$$

und die Transformation ist *orthogonal*.

Die kontinuierliche Fourier-Transformation in 2D

Die Fourier-Transformation (FT) benutzt harmonische Funktionen zur Zerlegung eines Bildes. Im 2D-Fall ist sie durch das Doppelintegral definiert:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp(-2\pi i(xu + yv)) dx dy.$$

Mit der FT wird ein Bild vom Ortsraum in den Frequenzraum transformiert. Wie von (x, y) als Ortskoordinaten gesprochen wird, so nennt man (u, v) Ortsfrequenzen. Die Bildfunktion $f(\cdot)$ wird als Linearkombination von periodischen Funktionen ausgedrückt, die Koeffizienten $F(\cdot)$ sind komplexe Wichtungsfaktoren, die aus dem Betrag (Realteil, Amplitude der harmonischen Schwingung) und der Phase (Imaginärteil, Phase der harmonischen Schwingung) bestehen.

Ihre Inverse lautet:

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) \exp(2\pi i(xu + yv)) du dv.$$

Die Formeln für eine n-dimensionale Transformation lassen sich analog ableiten.

Die diskrete Fourier-Transformation

Analog zur 1D diskreten Fourier-Transformation (DFT) lautet die 2D-Form:

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \exp \left[-2\pi i \left(\frac{mu}{M} + \frac{nv}{N} \right) \right]$$
$$u = 0, 1, \dots, M-1; \quad v = 0, 1, \dots, N-1.$$

Sei ϕ_{JJ} eine Transformationsmatrix der Größe $J \times J$:

$$\phi_{JJ}(k, l) = \frac{1}{J} \exp\left(-\frac{2\pi i}{J} kl\right) \quad k, l = 0, 1, \dots, J-1.$$

Dann ist die diskrete Fourier-Transformation definiert als eine orthogonale Transformation der Form:

$$\mathbf{F} = \phi_{MM} \mathbf{f} \phi_{NN}.$$

Ihre Inverse lautet:

$$f(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(u, v) \exp \left[2\pi i \left(\frac{mu}{M} + \frac{nv}{N} \right) \right]$$
$$u = 0, 1, \dots, M-1; \quad v = 0, 1, \dots, N-1.$$

Die inverse Transformationsmatrix ist:

$$\phi_{JJ}(k, l)^{-1} = \exp\left(\frac{2\pi i}{J} kl\right) \quad k, l = 0, 1, \dots, J-1,$$

d.h., die inverse diskrete Fourier-Transformation lautet in Matrixschreibweise:

$$\mathbf{f} = \phi_{MM}^{-1} \mathbf{F} \phi_{NN}^{-1}.$$

Eigenschaften der diskreten Fourier-Transformation

Die Eigenschaften einer 2D DFT lassen sich aus der 1D Transformation ableiten.

Die Fourier-Transformation sei nun durch den Operator $\mathcal{F}$ ausgedrückt, d.h. $\mathcal{F}\{f(x,y)\} = F(u,v)$. Damit lassen sich die folgenden Eigenschaften der Fourier-Transformation darstellen:

- *Linearität:*
$$\mathcal{F}\{af_1(x,y) + bf_2(x,y)\} = aF_1(u,v) + bF_2(u,v)$$
- *Translation im Ortsbereich:*
$$\mathcal{F}\{f(x-a, y-b)\} = F(u,v) \exp(-2\pi i(au + bv))$$
- *Translation im Frequenzbereich:*
$$\mathcal{F}\{f(x,y) \exp(2\pi i(u_0x + v_0y))\} = F(u - u_0, v - v_0)$$
- *Symmetrie:*
Für eine reelle Bildfunktion $f(x,y)$ gilt:
$$F(-u, -v) = F^\#(u, v),$$

wobei # die komplex Konjugierte bezeichnet. Da die Bildfunktion reell ist, sie durch einen Quadranten der Fourier-Transformation vollständig repräsentiert.
- *Dualität der Faltung:*
Die Faltung und die Fourier-Transformation sind verbunden durch:
$$\mathcal{F}\{(f \star h)(x,y)\} = F(u,v)H(u,v)$$

$$\mathcal{F}\{f(x,y)h(x,y)\} = (F \star H)(u,v)$$

Die schnelle Fourier-Transformation I

Die DFT kann durch den "Fast Fourier Transformation"- (FFT) Algorithmus effizient berechnet werden.

Zur Implementierung kann die diskrete Transformation umgeschrieben werden:

$$F(u, v) = \frac{1}{M} \sum_{m=0}^{M-1} \left[\frac{1}{N} \sum_{n=0}^{N-1} \exp\left(\frac{-2\pi i v}{N}\right) f(m, n) \right] \exp\left(\frac{-2\pi i u}{M}\right)$$
$$u = 0, 1, \dots, M-1; \quad v = 0, 1, \dots, N-1.$$

Dabei entspricht der Term in den eckigen Klammern einer ein-dimensionalen DFT der m -ten Zeile, die mit einer Standard-FFT berechnet werden kann. Jede Bildzeile wird also durch ihre Fourier-transformierte ersetzt, worauf eine 1D-DFT in jeder Spalte berechnet wird.

Dieses Schema kann in höheren Dimensionen entsprechend fortgesetzt werden.

Die schnelle Fourier-Transformation II

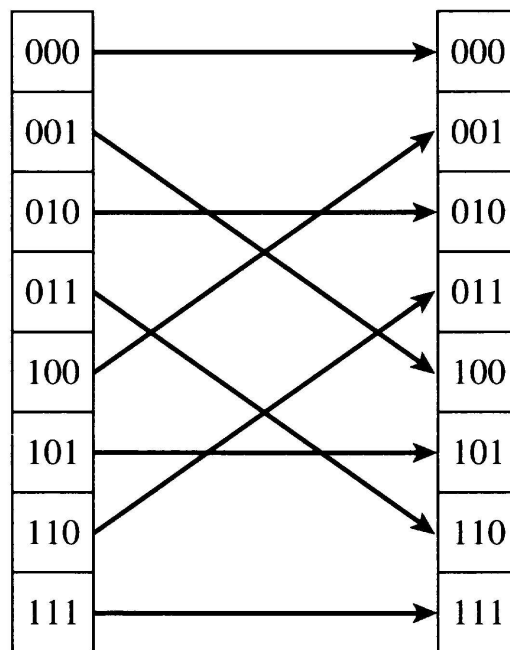
Die 1D diskrete Transformation kann in einen "geraden" und "ungeraden" Anteil getrennt werden:

$$\begin{aligned}
 F(v) &= \sum_{n=0}^{N-1} \exp\left(\frac{-2\pi i v}{N}\right) f(n) \\
 &= \sum_{n=0}^{N/2-1} \exp\left(\frac{-2\pi i 2v}{N}\right) f(2n) + \sum_{n=0}^{N/2-1} \exp\left(\frac{-2\pi i (2v+1)}{N}\right) f(2n+1) \\
 &= \sum_{n=0}^{N/2-1} \exp\left(\frac{-2\pi i v}{N/2}\right) f(2n) + w(v) \sum_{n=0}^{N/2-1} \exp\left(\frac{-2\pi i v}{N/2}\right) f(2n+1) \\
 &= F^g(v) + w(v)F^u(v).
 \end{aligned}$$

Wenn $N = 2^k$ gilt, dann kann die Zerlegung bis zur einer Transformation der Länge 1 durchgeführt werden. Eine solche Transformation entspricht genau einem Element der Ausgangsdaten:

$$F^{uggguggug...ugg}(v) = f(n) \quad \text{für ein } n.$$

Welcher Wert n ist nun einem Muster $uggguggug...ugg$ zugeordnet? Setzt man $g = 0$ und $u = 1$ und kehrt die Reihenfolge um, so erhält man den Platz von n als Binärwert.



Die schnelle Fourier-Transformation III

Damit lautet der Algorithmus für eine 1D-FFT:

1. Sortiere die Eingangsdaten um, indem eine Bitumkehr der Platznummern durchgeführt wird.
2. Für $k = 2, 4, \dots, N$:
Fasse die entsprechenden Koeffizienten der "geraden" und "ungeraden" Anteile zusammen.

Geben ist ein komplexes Feld `data` der Länge N , worin der imaginäre Anteil 0 ist. In der Ausgabe steht in den Feldern $0 \dots N/2$ die Fourier-Transformierte.

```
int j = 0; int N2 = N/2;
for (int i = 0; i < N2; i++) {
    if (j < i) SWAP(data[j], data[i]);
    for (int m = N; m >= 2 && j > m; m >>= 1) j -= m;
    j += m;
}
for (int k = 1; k < N2; k *= 2) {
    complex wk(-2*sin(M_PI/k)*sin(M_PI/k), sin(2*M_PI/k));
    complex wi(1, 0);
    for (int m = 0; m < k; m++) {
        for (int i = m; i < N2; i += 2*k) {
            data[i+k] = data[i] - wi * data[i+k];
            data[i]    = data[i] + wi * data[i+k];
        }
        wi *= wk;
    }
}
```

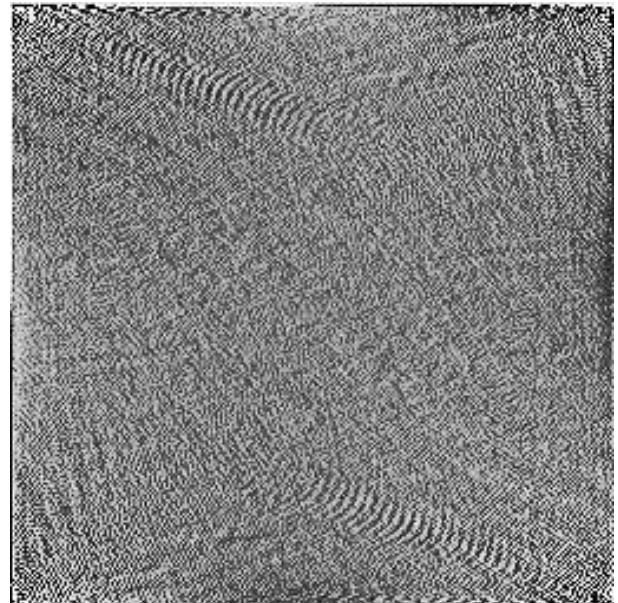
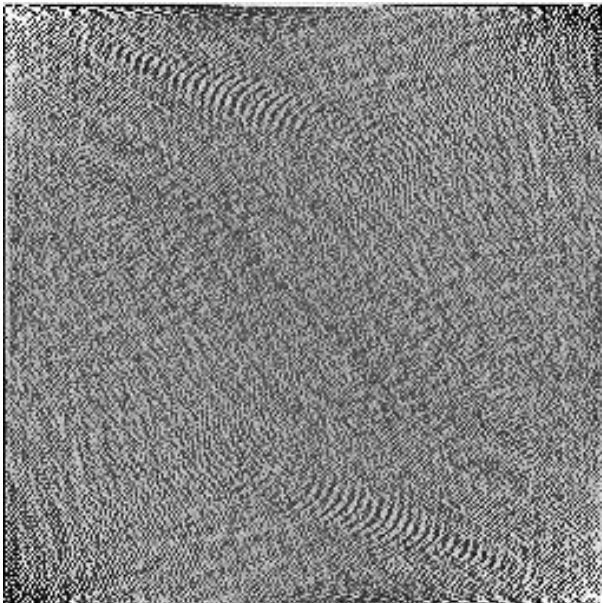
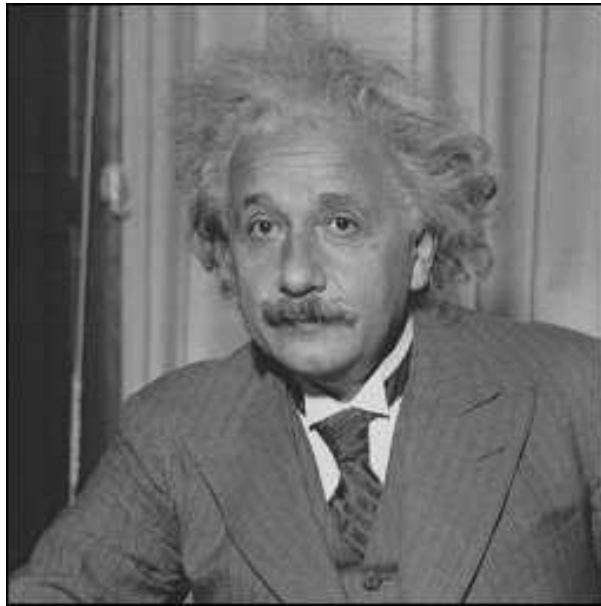
Eine effiziente Implementierung ist unter <http://www.fftw.org> zu finden.

Die diskrete Fourier-Transformation - Beispiel

Die Fourier-Transformation einer reellen Funktion ist eine komplexe Funktion:

$$F(u, v) = R(u, v) + iI(u, v).$$

Originalbild (oben mitte), Realteil (unten links) und Imaginärteil (unten rechts) des Fourier-transformierten Bildes.



Beachten Sie die Punktsymmetrie zum Bildmittelpunkt.

Spektrum und Phase eines Bildes

Der Betrag $|F(u, v)|$ wird *Frequenzspektrum* des Bildes $f(m, n)$ genannt:

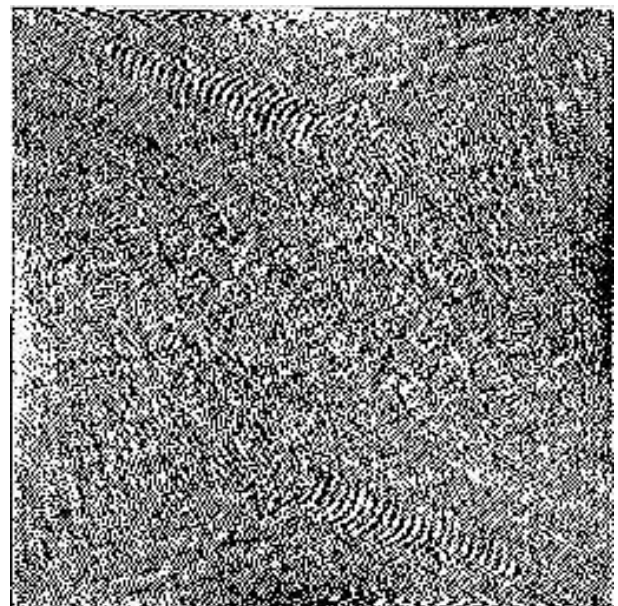
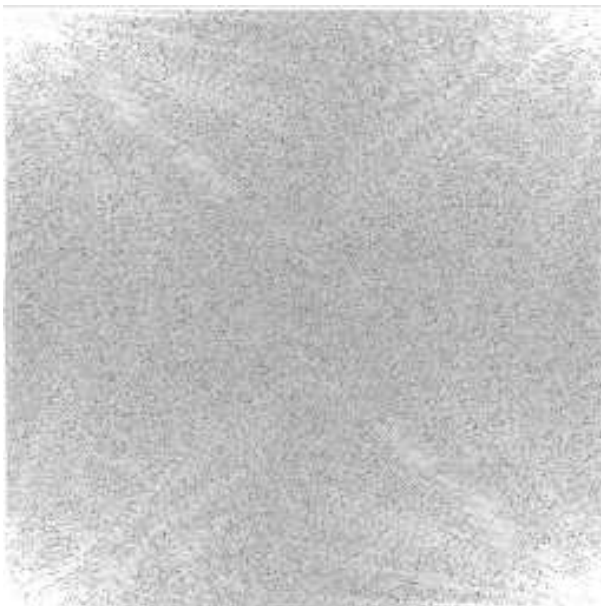
$$|F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)}.$$

Das *Phasenspektrum* $\phi(u, v)$ und *Energiespektrum* $P(u, v)$ sind definiert als:

$$\phi(u, v) = \tan^{-1} \left[\frac{I(u, v)}{R(u, v)} \right].$$

$$P(u, v) = |F(u, v)|^2 = R^2(u, v) + I^2(u, v).$$

Frequenzspektrum (links) und Phasenspektrum (rechts) des Fourier-transformierten Bildes.



Fourier-Transformation: Faltung

Die Fourier-Transformation läßt sich gut zur Faltung (Konvolution) eines Bildes benutzen:

$$g(a,b) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n)h(a-m,b-n).$$

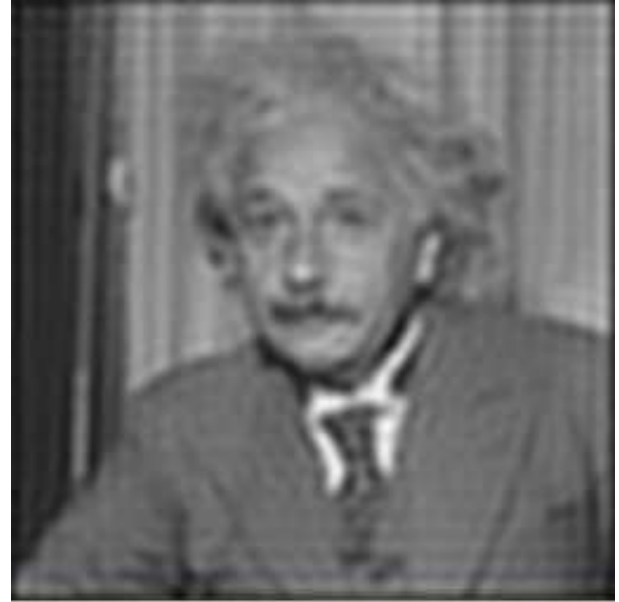
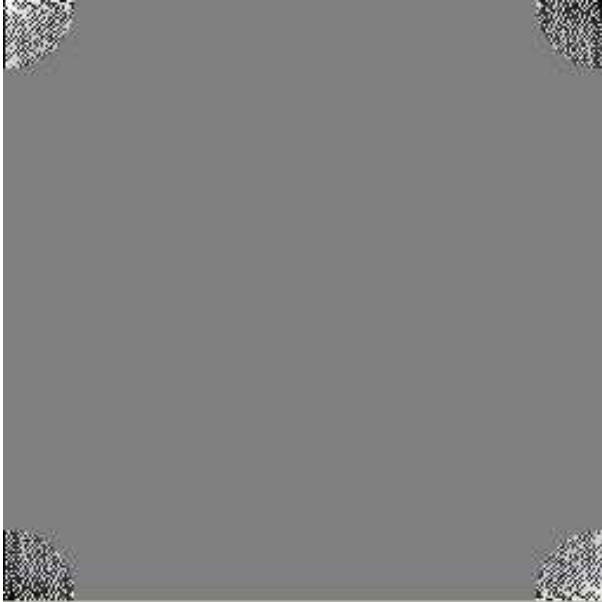
Mit dem Faltungstheorem $G(u,v) = F(u,v)H(u,v)$ folgt:

$$g(a,b) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v)H(u,v) \exp[2\pi i(\frac{au}{M} + \frac{bv}{N})].$$

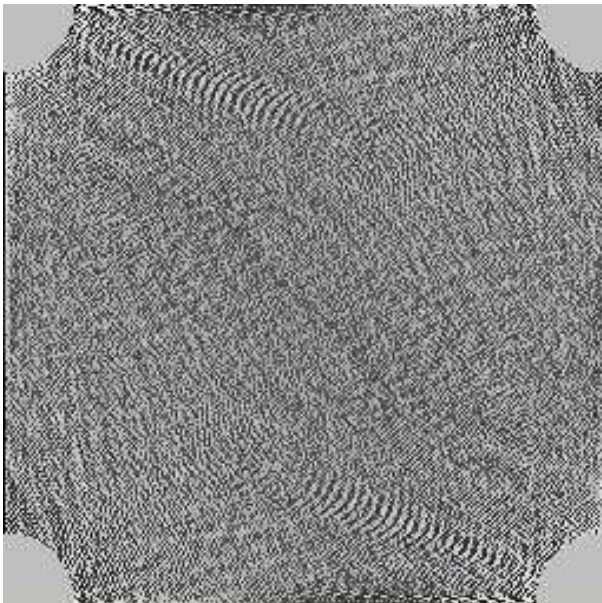
Die Faltung eines Bildes $f(\cdot)$ läßt sich also durch punktweise Multiplikation seiner Fourier-Transformierten $F(\cdot)$ mit der Fourier-Transformierten des Faltungskerns $H(\cdot)$ und anschließender Rücktransformation durchführen.

Faltung - Beispiel

Tiefpaßfilterung eines Bildes:



Hochpaßfilterung eines Bildes:



Die Hadamard-Transformation

Wenn als Satz orthogonaler Funktionen zur Repräsentation einer orthogonalen Transformation Rechteck-Funktionen (sog. Walsh-Funktionen) benutzt werden, erhält man die Hadamard-Transformation. Die Walsh-Funktionen sind reell und besitzen nur die Funktionswerte ± 1 .

Die Hadamard-Matrix der zweiten Ordnung ist:

$$H_{22} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}.$$

Eine Hadamard-Matrix der Ordnung 2^k ist:

$$H_{2J2J} = \begin{vmatrix} H_{JJ} & H_{JJ} \\ H_{JJ} & -H_{JJ} \end{vmatrix}.$$

Die inverse Matrix ist einfach:

$$H_{JJ}^{-1} = \frac{1}{J} H_{JJ}$$

Damit gilt für die Hadamard-Transformation:

$$F = H_{MM} f H_{NN}, \quad f = \frac{1}{MN} H_{MM} F H_{NN}$$

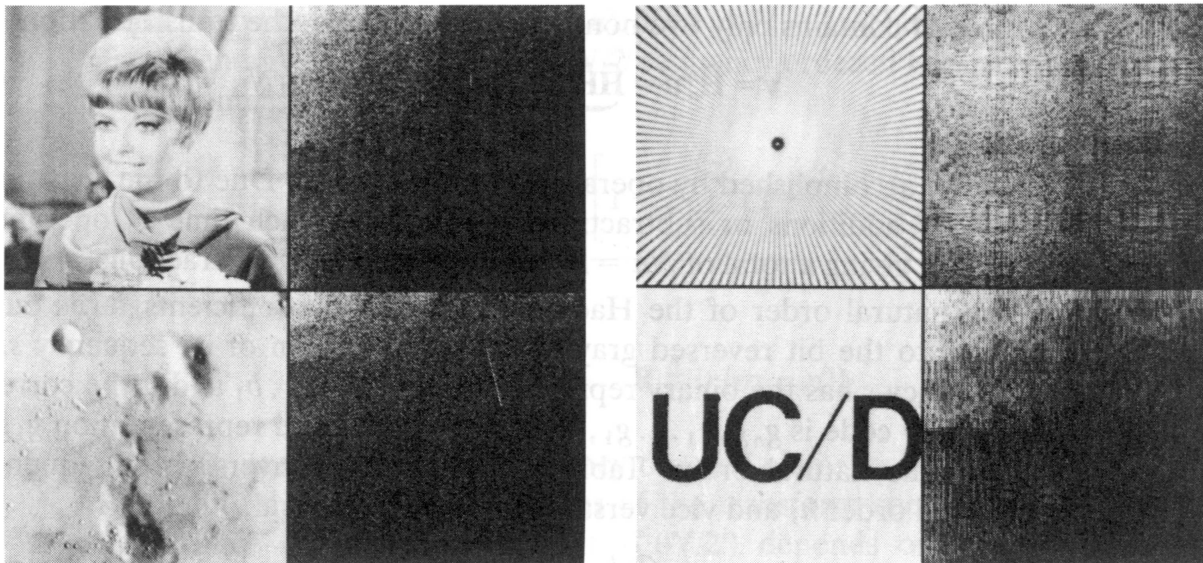
Die Hadamard-Transformation ist reell, symmetrisch und orthogonal:

$$H = H^* = H^T = H^{-1}$$

Die Hadamard-Transformation ist einfach zu berechnen - es sind nur Additionen und Subtraktionen notwendig.

Hadamard-Transformation - Beispiel

Die Hadamard-Transformation besitzt eine sehr gute Energiekompaktierung. 97.5% der Energie eines Bildes sind in der Hälfte der Koeffizienten enthalten. Daher besitzt diese Transformation in der Bildkodierung eine Bedeutung.



Die diskrete Cosinus-Transformation

Es gibt vier verschiedene diskrete Cosinus-Transformationen, von denen die zweite (DCT-II) Bedeutung in der Bildkodierung hat (JPEG, MPEG-Formate).

Für ein Bild der Größe $N \times N$ kann die DCT ausgedrückt werden als:

$$C_{NN}(k, l) = \begin{cases} \frac{1}{\sqrt{N}} & \text{für } l = 0, \\ \frac{2}{\sqrt{N}} \cos\left[\frac{(2k+1)l\pi}{2N}\right] & \text{für alle anderen } k, l. \end{cases}$$

Im 2D-Fall gilt für die Vorwärtstransformation der DCT-II:

$$F(u, v) = \frac{2c(u)c(v)}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m, n) \cos\left(\frac{2m+1}{2N}u\pi\right) \cos\left(\frac{2n+1}{2N}v\pi\right)$$
$$u = 0, 1, \dots, N-1; \quad v = 0, 1, \dots, N-1.$$

$$c(k) = \begin{cases} \frac{1}{\sqrt{2}} & \text{für } k = 0, \\ 1 & \text{für } k \neq 0. \end{cases}$$

Die inverse Transformation der DCT-II ist:

$$f(m, n) = \frac{2}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} c(u)c(v)F(u, v) \cos\left(\frac{2m+1}{2N}u\pi\right) \cos\left(\frac{2n+1}{2N}v\pi\right)$$
$$m = 0, 1, \dots, N-1; \quad n = 0, 1, \dots, N-1.$$

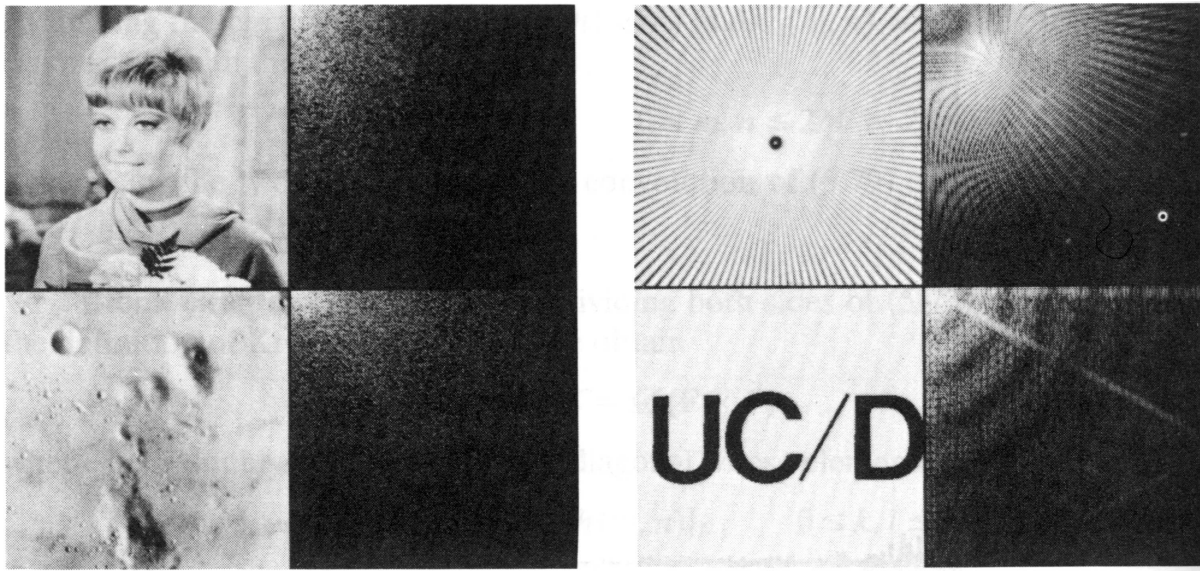
Die DCT kann auf der Basis der Fourier-Transformation berechnet werden. Alle N Koeffizienten können durch eine $2N$ -FFT bestimmt werden.

Die DCT ist reell und orthogonal:

$$C = C^* \leftarrow C^T = C^{-1}$$

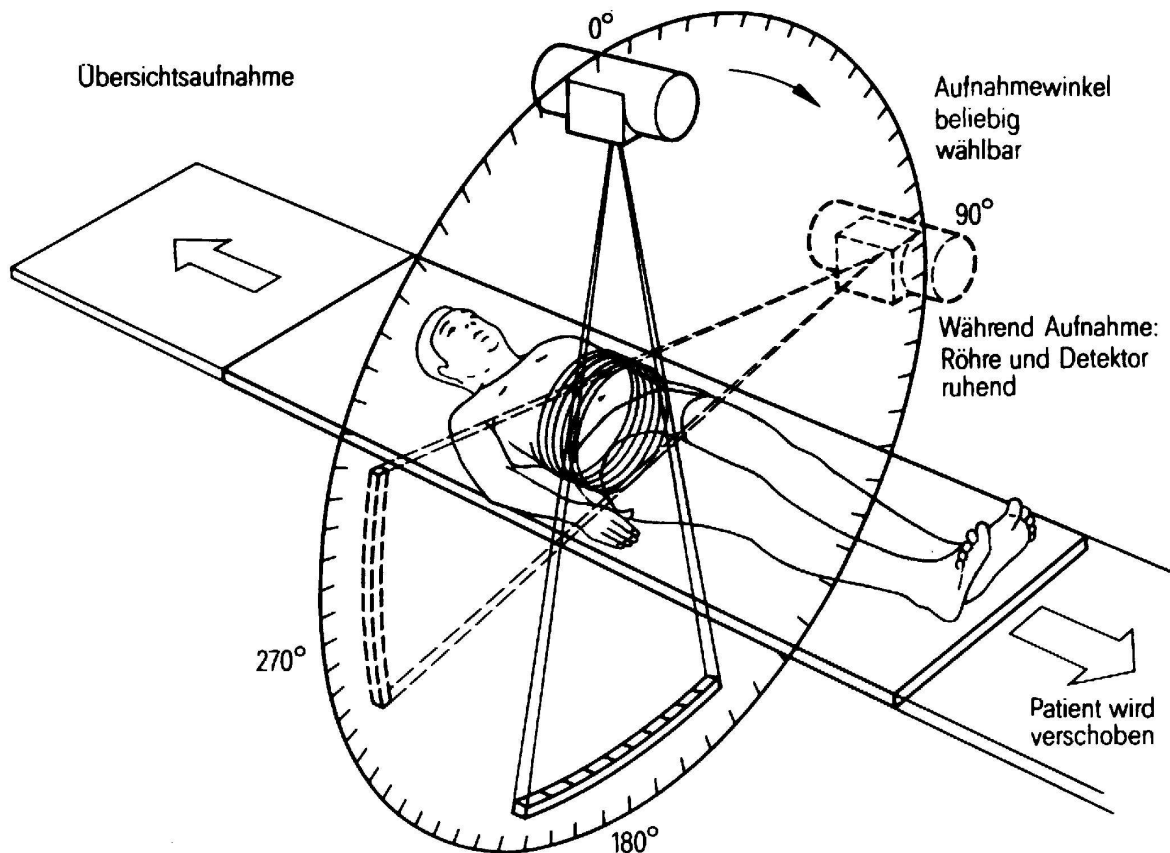
DCT - Beispiel

Auch die DCT besitzt eine sehr gute Energiekompaktierung. Auch diese Transformation in der Bildkodierung eine Bedeutung, z.B. im JPEG- und MPEG-Format.



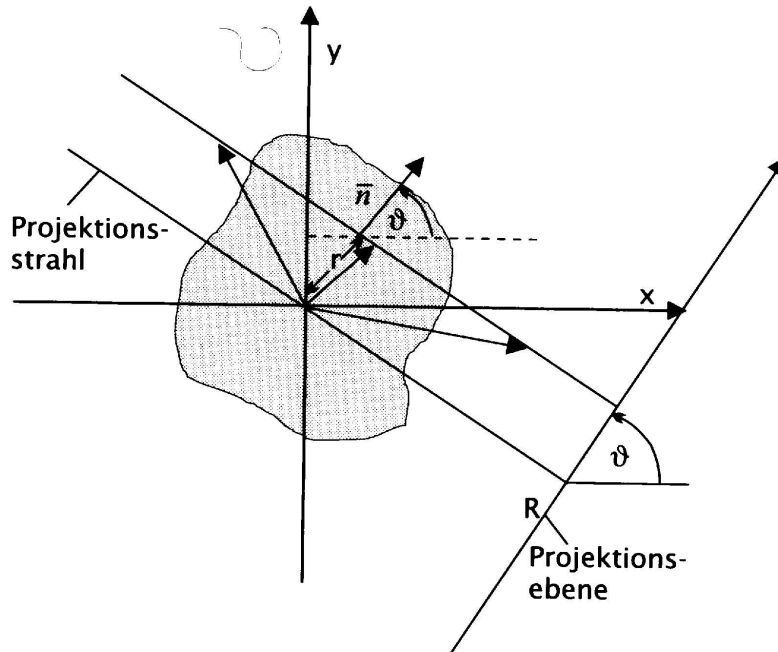
Bilderzeugung im Tomographen

Der Patient liegt im Detektorring eines Tomographen. Ein fächerförmiger Röntgenstrahl durchdringt ihn, der auf der gegenüberliegenden Seite von Detektoren gemessen wird. Der Detektorring dreht sich um 180 Grad um den Patienten, um so Projektionen aus allen Richtungen zu erhalten.



Die Radon-Transformation

Ein Tomograph erzeugt ein Bild durch eine Serie von Projektionen in unterschiedlichen Richtungen. Diese Projektionen können als eine weitere Art eines 2D-Bildes verstanden werden: eine Koordinate ist die Position im Projektionsprofil r , die andere der Winkel ϑ . Somit entspricht die Rekonstruktion eines tomographischen Bildes der Anwendung einer inversen Transformation.



Die Geradengleichung für den Projektionsstrahl lautet:

$$\vec{x}\vec{n} - r = x\cos(\vartheta) + y\sin(\vartheta) - r = 0.$$

Die projizierte Intensität $g(r, \vartheta)$ wird durch die Integration entlang des Projektionsstrahls gegeben:

$$g(r, \vartheta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x\cos(\vartheta) + y\sin(\vartheta) - r) \, dx dy.$$

Die δ -Funktion reduziert dabei das Doppelintegral auf ein Segment des Projektionsstrahls in Richtung ϑ im Abstand r vom Koordinatenursprung. Die Funktion $g(\cdot)$ beschreibt die materialabhängige Absorption entlang des Strahls. Diese projektive Transformation der Funktion $f(x, y)$ (d.h., des Bildes) auf $g(r, \vartheta)$ (d.h., auf die Detektoren des Rings) wird *Radon-Transformation* genannt.

Die inverse Radontransformation I

Wenn Projektionen aus allen Richtungen vorliegen, kann das tomographische Schnittbild kann aus den Projektionen durch eine inverse Radontransformation rekonstruiert werden. Zum Verständnis dieser inversen Transformation wird zunächst das Fourier-Scheibentheorem eingeführt.

Einschub: Das Fourier-Schnitt-Theorem

Die 1D-Fourier-Transformierte der Projektion $g(r, \vartheta)$ entlang r im Winkel ϑ ist gleich einem Schnitt in diesem Winkel durch eine 2D-Fourier-Transformierte $F(u, v)$ des Bildes $f(x, y)$:

$$\mathcal{F}(g(r, \vartheta)) = F(\eta \cos(\vartheta), \eta \sin(\vartheta)).$$

Beweis:

$$\begin{aligned} G(\eta, \vartheta) &= \mathcal{F}(g(r, \vartheta)) \\ &= \int_{-\infty}^{\infty} g(r, \vartheta) \exp(-2\pi i \eta r) dr \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r \cos(\vartheta) - u \sin(\vartheta), r \sin(\vartheta) + u \cos(\vartheta)) \\ &\quad \exp(-2\pi i \eta r) dr du. \end{aligned}$$

Nun wird das Koordinatensystem (r, u) nach (x, y) transformiert:

$$\begin{aligned} x &= r \cos(\vartheta) - u \sin(\vartheta) \\ y &= r \sin(\vartheta) + u \cos(\vartheta) \\ r &= x \cos(\vartheta) + y \sin(\vartheta), \end{aligned}$$

womit folgt:

$$\begin{aligned} G(\eta, \vartheta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp[-2\pi i (x \eta \cos(\vartheta) + y \eta \sin(\vartheta))] dx dy \\ &= F(\eta \cos(\vartheta), \eta \sin(\vartheta)). \end{aligned}$$

Die inverse Radontransformation II

Wenn Projektionen aus allen Richtungen vorliegen, decken die resultierenden Schnitte des Spektrums das komplette Spektrum des Objektes ab. Die inverse FT liefert dann genau das Originalbild.

Da die Strahlen bei kleinen Radien dichter als bei großen Radien liegen, ist allerdings die spektrale Dichte bei kleinen Wellenzahlen zu hoch. Bevor das Spektrum der Projektionsstrahlen zum Gesamtspektrum aufaddiert wird, muß daher ein geeigneter Wichtungsfaktor berücksichtigt werden.

Dieser Wichtungsfaktor ist im kontinuierlichen Fall $|\eta|$, da die Dichte der Projektionsstrahlen proportional zu $|\eta|^{-1}$ ist.

Die inverse FT:

$$f(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u,v) \exp(2\pi i(xu + yv)) du dv$$

lautet in Polarkoordinaten:

$$f(x,y) = \frac{1}{2\pi} \int_0^{\pi} \int_{-\infty}^{\infty} F(\eta \cos(\theta), \eta \sin(\theta)) \exp(-2\pi i \eta (x \cos(\theta) + y \sin(\theta))) |\eta| d\eta d\theta,$$

wobei $du dv = |\eta| d\eta d\theta$.

Die inverse Radontransformation III

Mit dem Fourier-Scheibentheorem $G(\eta, \vartheta) = F(\eta \cos(\vartheta), \eta \sin(\vartheta))$ folgt:

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi} \int_0^\pi \left[\int_{-\infty}^\infty G(\eta, \vartheta) \exp(-2\pi i \eta) |\eta| d\eta \right] d\vartheta \\ &= \frac{1}{2\pi} \int_0^\pi \hat{g}(r, \vartheta) d\vartheta, \end{aligned}$$

worin

$$\hat{g}(r, \vartheta) = \int_{-\infty}^\infty G(\eta, \vartheta) \exp(-2\pi i \eta) |\eta| d\eta.$$

Man setzt $G(\eta, \vartheta) |\eta| = \eta G \operatorname{sign}(\eta)$ und wendet das Faltungstheorem an:

$$\begin{aligned} \hat{g}(r, \vartheta) &= \mathcal{F}^{-1}[\eta G] \star \mathcal{F}^{-1}[\operatorname{sign}(\eta)] \\ &= \left[\frac{1}{2\pi i} \frac{\delta g}{\delta r}(r, \vartheta) \right] \star \left[-\frac{1}{2\pi i r} \right] \\ &= \frac{1}{2\pi^2} \int_{-\infty}^\infty \left[\frac{\delta g(t, \vartheta)}{\delta t} \right] \frac{1}{r-t} dt. \end{aligned}$$

Einsetzen von $g(r, \vartheta)$ ergibt dann die Gleichung für die *inverse Radon-Transformation*:

$$f(x, y) = \frac{1}{2\pi^2} \int_0^\pi \int_{-\infty}^\infty \frac{(\delta g / \delta r)(r, \vartheta)}{x \cos(\vartheta) + y \sin(\vartheta) - r} dr d\vartheta.$$

Die inverse Radontransformation IV

Die Rekonstruktion eines tomographischen Bildes besteht also aus drei Schritten:

- FT der Projektionen $g(r, \vartheta) \rightarrow G((\eta, \vartheta))$,
- Inverse FT der gefilterten $G(\cdot)$: $|\eta|G((\eta, \vartheta)) \rightarrow \hat{g}(r, \vartheta)$,
- Rückprojektion mit $r = x \cos(\vartheta) + y \sin(\vartheta)$:
$$f(x, y) = \frac{1}{2\pi} \int_0^\pi \hat{g}(r, \vartheta) d\vartheta.$$

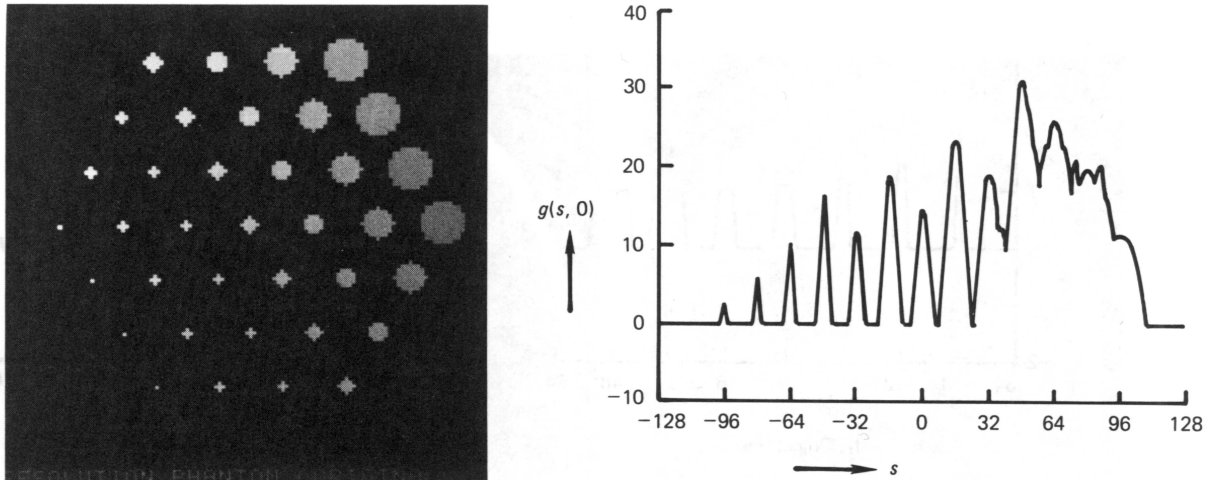
Also:

- Gemessen wurde eine Sequenz von Projektionen:
 $g_n(m), -M/2 \leq m \leq (M/2) - 1$.
Erweitere die Sequenz, so daß:
 $g_n(m)_c, 0 \leq m \leq K - 1$, mit $K = 2M$.
Führe eine 1D-FFT durch, um $\hat{G}_n(k), 0 \leq m \leq K - 1$ zu erhalten.
- Taste die Filterfunktion $H(\eta)$ ab:
 $\hat{H}(k) = H(k\Delta\eta), \hat{H}(K - k) = \hat{H}^*(k), 0 \leq k < K/2$.
- Berechne die Faltung: $\hat{G}_n(k)\hat{H}(k), 0 \leq m \leq K - 1$.
Berechne die inverse FT, um $\hat{g}_n(m), -K/2 \leq m \leq (K/2) - 1$ zu erhalten.
- Berechne eine Approximation des Rückprojektionsintegrals:
$$f(x, y) \approx \sum_{n=0}^{N-1} \hat{g}(x \cos(n\Delta) + y \sin(n\Delta), n\Delta).$$

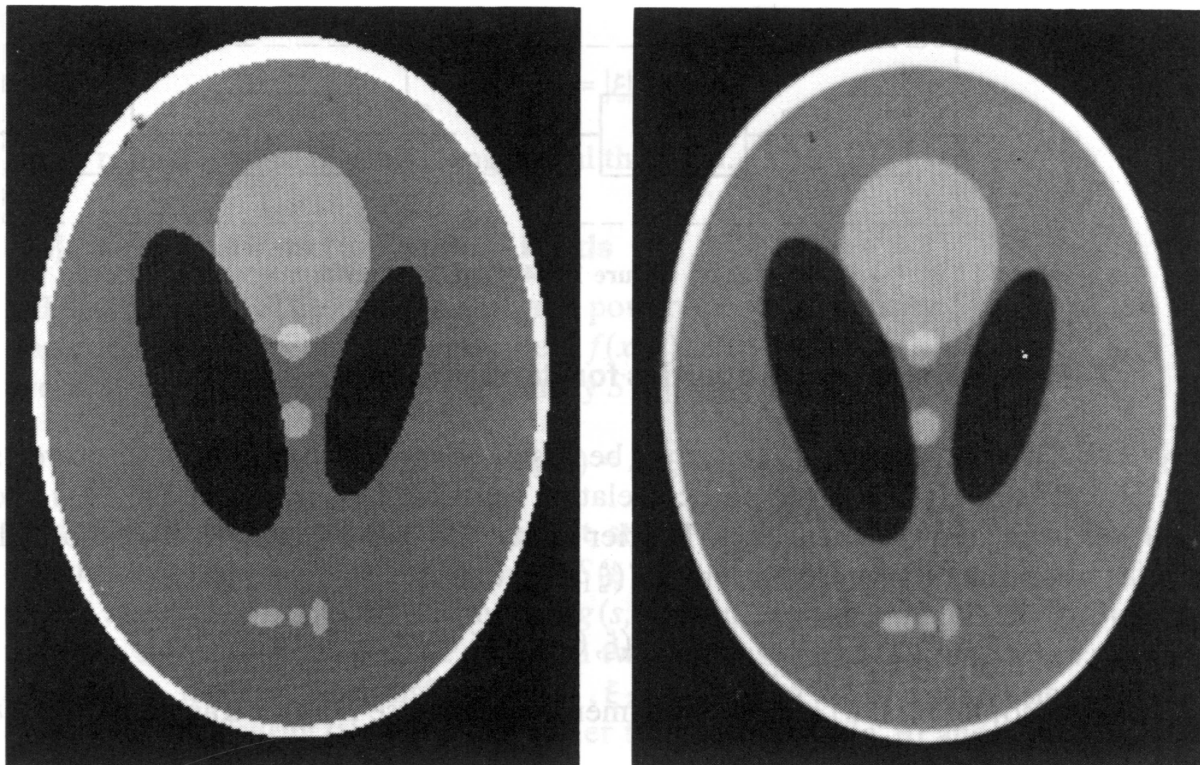
Um $f(\cdot)$ auf jedem Gitterpunkt berechnen zu können, muß $\hat{g}(\cdot)$ in der Regel interpoliert werden.

Die Rückprojektion - Beispiele

Testbild (links) und Projektion $g(r, 0)$:

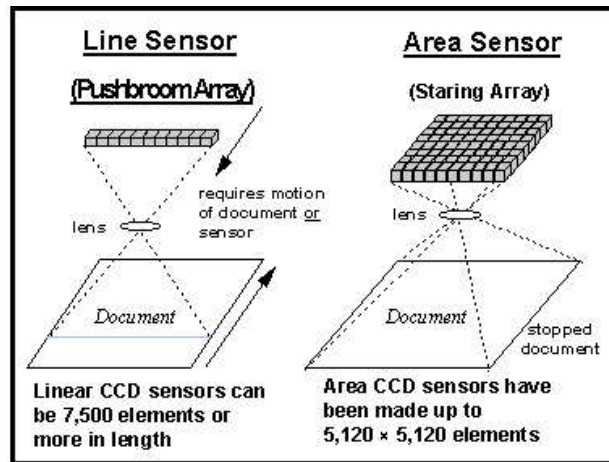


Phantombild (links) und Rekonstruktion (rechts):

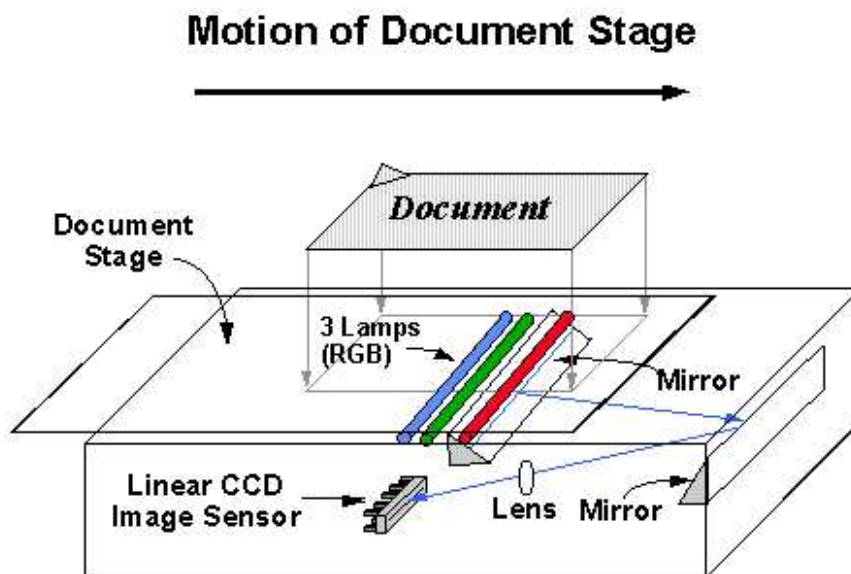


CCD Sensoren

Die meisten Scanner und Digitalkameras verwenden heute **charge coupled devices** (CCD) als Sensoren, die in linearer oder Array-Konfiguration aufgebaut sind. Lineare CCDs haben 7500 Pixel oder mehr, Array-Konfigurationen erreichen heute Auflösungen von 4kx4k Pixeln. Die Pixelgröße liegt bei $9 \times 9 \mu\text{m}$.



Für Farbaufnahmen werden entweder Filter oder drei verschiedene Lichtquellen verwendet.



Der Abbildungsprozeß einer Kamera I

Wir betrachten eine Kamera mit einer dünnen Linse. Diese Kamera macht eine Perspektivprojektion. Die vertikale Linie wird **optische Achse** genannt, die Ebene π die Bildebene. Die Linse befindet sich senkrecht zur optischen Achse am Punkt C. Die Fokusslänge f ist eine Linseneigenschaft. Ein Strahl der Szene, der am Punkt X reflektiert wird, wird durch Punkt C auf U auf der Bildebene abgebildet.

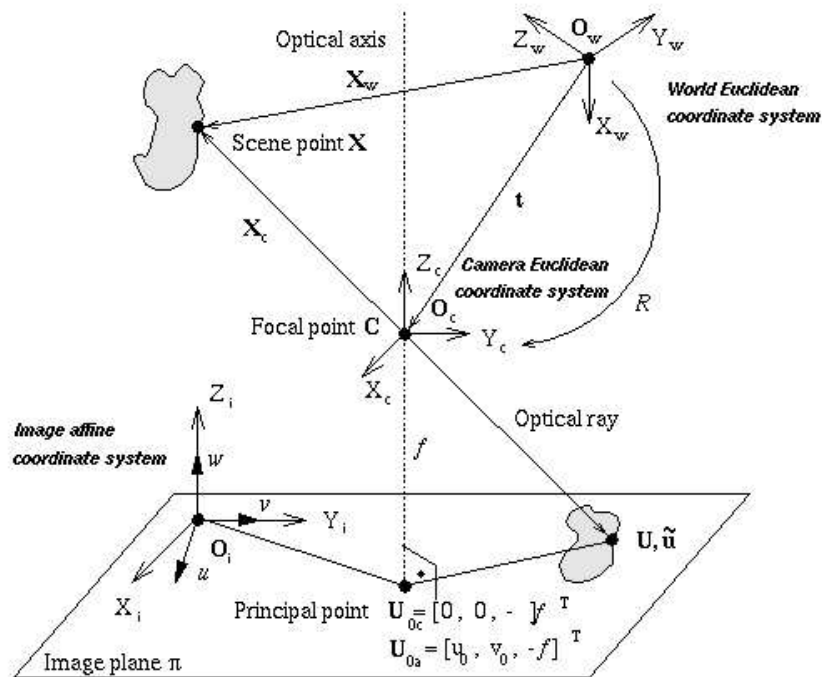


Figure 9.3 The geometry of a linear perspective camera.

Es werden vier Koordinatensysteme vereinbart:

- das euklidische Weltkoordinatensystem (z.B. X_w, U_w)
- das euklidische Kamerakordinatensystem (Ursprung C_c , die optische Achse Z_c)

Beide Systeme können durch eine euklidische Transformation aus Translation und Rotation abgebildet werden.

Der Abbildungsprozeß einer Kamera II

- das euklidische Bildkoordinatensystem besitzt Achsen, die in der Bildebene liegen (also X_i, Y_i).
- das affine Bildkoordinatensystem besitzt die Koordinaten u, v, w und den Ursprung O_i .

Ein Punkt X_w der Welt wird in der Kamera auf X_c abgebildet:

$$X_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R(X_w - t).$$

Der Punkt X_c wird in der Bildebene π als Punkt U_c abgebildet:

$$U_c = \left[-\frac{fx_c}{z_c}, -\frac{fy_c}{z_c}, -f \right]^T.$$

Es bleibt festzustellen, wo der Punkt U_c im affinen Bildkoordinatensystem liegt. Das Bildkoordinatensystem, mit dem Ursprung an der linken oberen Bildecke, wird durch eine Scherung und Skalierung dargestellt. Der Punkt U_0 stellt den Schnittpunkt der optischen Achse mit der Bildebene dar: $U_{0a} = [u_0, v_0, 0]^T$.

$$\tilde{u} = \begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} a & b & -u_0 \\ 0 & c & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{fx_c}{z_c} \\ -\frac{fy_c}{z_c} \\ 1 \end{bmatrix}.$$

Der Abbildungsprozeß einer Kamera III

Einsetzen aller Konstanten führt zur **Kamerakalibrationsmatrix** K :

$$\begin{aligned}
 z_c \tilde{u} &= z_c \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{x_c}{z_c} \\ \frac{y_c}{z_c} \\ 1 \end{bmatrix} \\
 &= \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -x_c \\ y_c \\ z_c \end{bmatrix} \\
 &= \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} R(X_w - t) = KR(X_w - t).
 \end{aligned}$$

Die so genannten sechs **extrinsischen** Parameter R und t verbinden das euklidische Kamerasystem mit dem euklidischen Weltsystem, es handelt sich um drei Rotationen und drei Translationen.

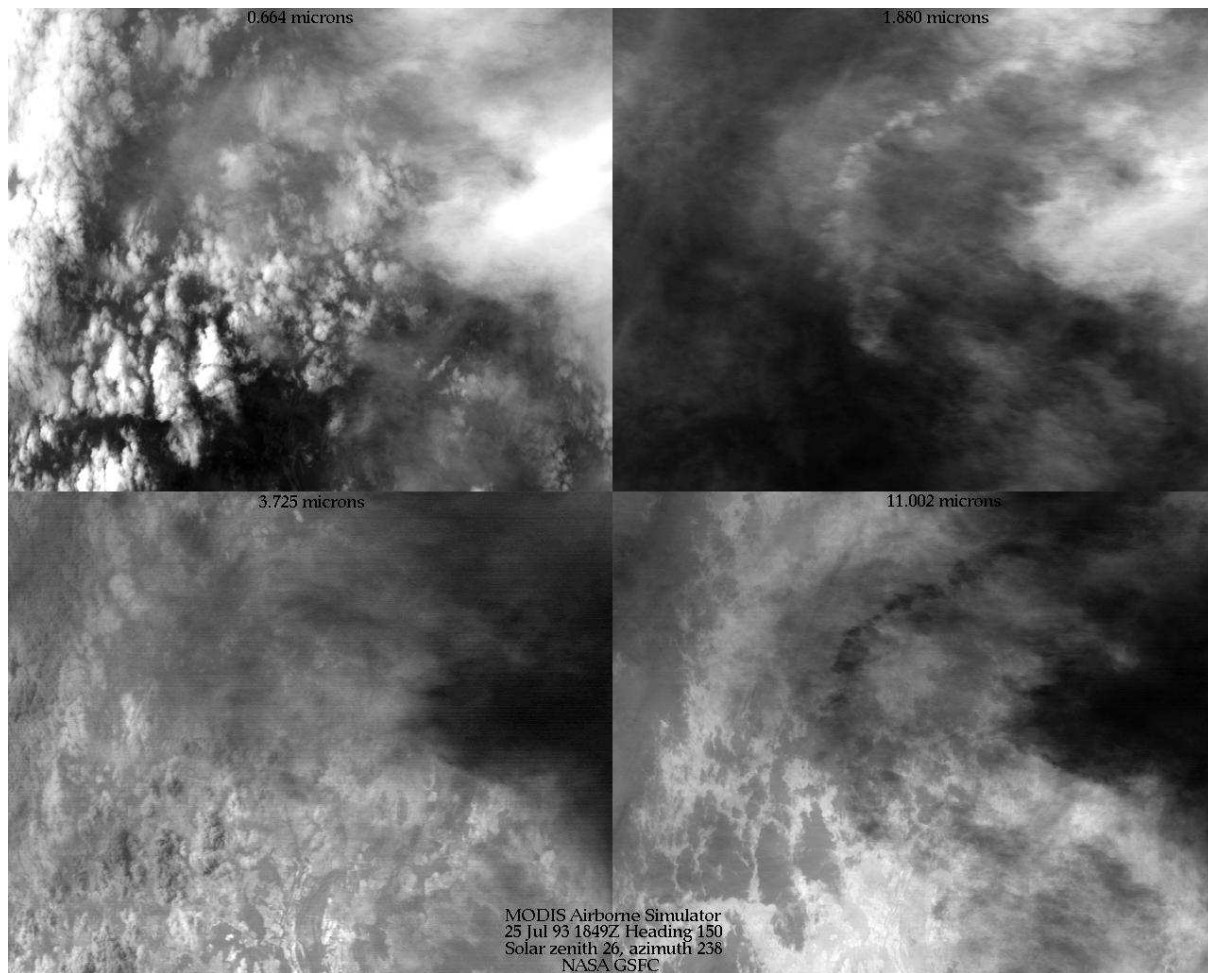
Die Kamerakalibrationsmatrix K umfaßt die **intrinsischen** Parameter der Kamera, die von ihrer Position unabhängig sind. Wenn die intrinsischen Parameter einer Kamera bekannt sind, können metrische Messungen auf Bildern gemacht werden.

Die Kameramatrix M folgt daraus als 4x4 Matrix:

$$\tilde{u} = \begin{bmatrix} U \\ V \\ W \end{bmatrix} = [KR | -KRt] \begin{bmatrix} X_w \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ 1 \end{bmatrix}.$$

Abbildung - Beispiel

Das Rückrechnen von Bildkoordinaten auf Weltkoordinaten hat z.B. bei Luftaufnahmen große Bedeutung:



Dabei wird der so genannte "ground truth" (Kalibrierungsmarken am Boden) zur Kalibrierung benutzt.

Fingerabdrücke

Eine frühe Anwendung der Bildverarbeitung ist der Vergleich von neu gefundenen mit bereits gespeicherten Fingerabdrücken, z.B. zur Identifizierung von Tätern. Weitere Beispiele sind zu finden unter:

<ftp://ftp.cs.columbia.edu/jpeg/other/uuencoded/>



Stereokamerabilder

Nach Kalibrierung einer Kamera kann aus einem Bildpunkt der Abbildungsstrahl berechnet werden. Wenn zwei Kameras denselben Punkt X betrachten, können seine Koordinaten in 3D als Schnittpunkt beider Strahlen berechnet werden.

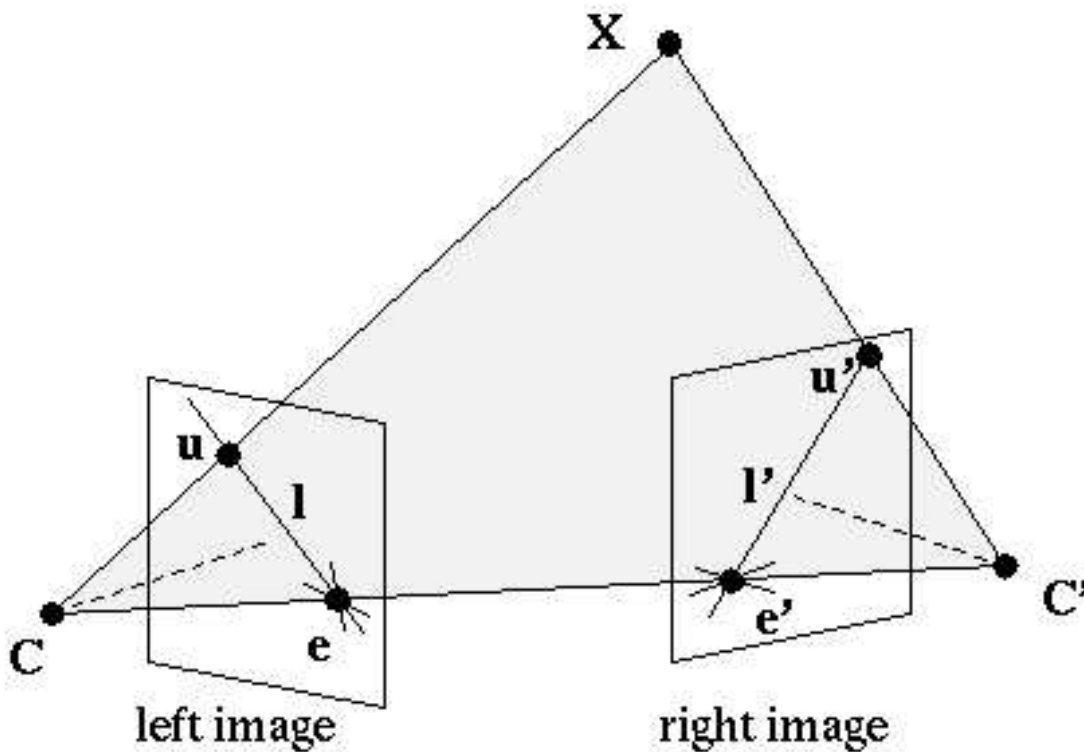


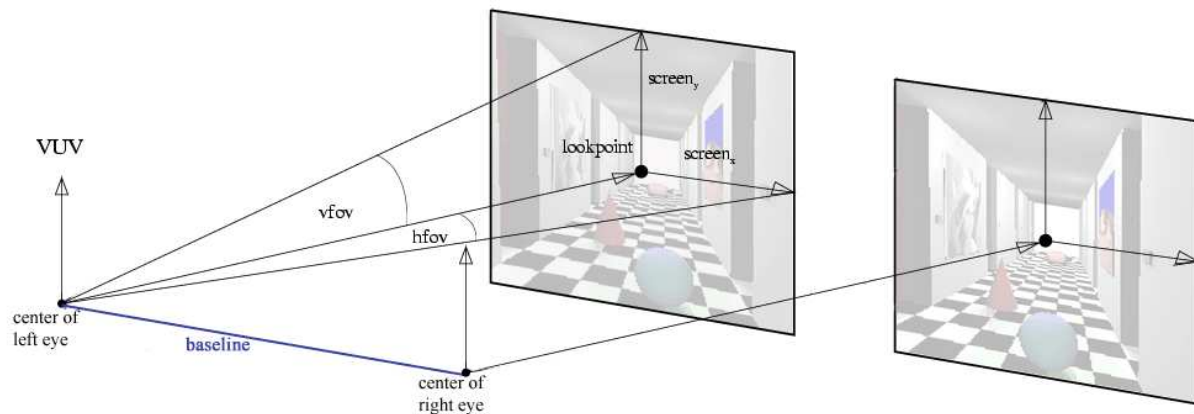
Figure 9.8 *Epipolar geometry in stereopsis.*

Eine Stereokamera nimmt zwei Bilder in einem bestimmten Abstand unter einem bestimmten Winkel auf. Die Aufnahmetechnik ist dem menschlichen Sehen nachgebildet. Das Prinzip des Stereosehens besteht aus drei Komponenten:

- die Kamerakalibrierung
- der Feststellung von Punktkorrespondenzen
- der Rekonstruktion von 3D Koordinaten in der Szene.

Stereokamerabilder II

Häufig wird eine **kanonische** Konfiguration gewählt, bei der die Bildebenen parallel sind:



Position linkes Auge: (250, -150, 170) blickt auf (100, 1600, 100)

Position rechtes Auge: (259.963,-149.146,170) blickt auf (109.963,1600.85,100)

Gesichtsfeld ($H_{FOV}=V_{FOV}=30$ deg) mit Auflösung (X=Y=2560)



Stereokamerabilder III

Betrachten wir eine einfache kanonische Konfiguration und er-
rechnen die Tiefe:

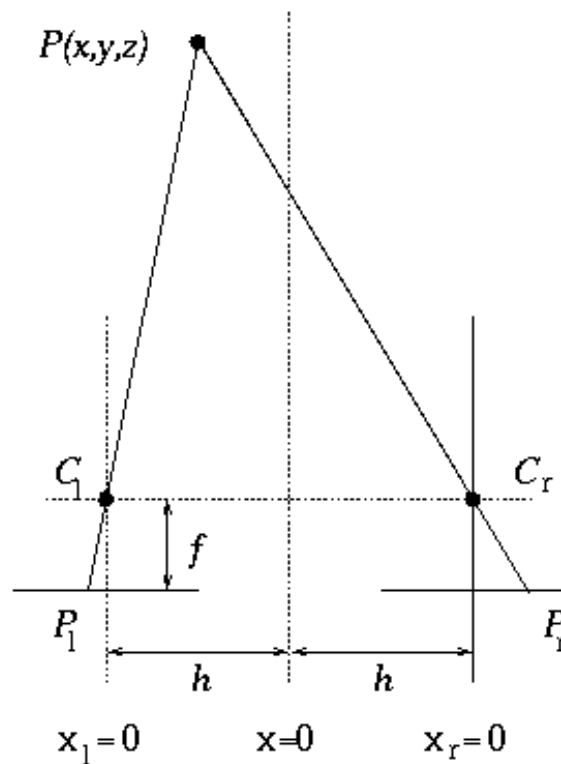


Figure 9.10 *Elementary stereo geometry in canonical configuration.*

$\vec{P_l C_l}$ und $\vec{C_l P}$ sind die Hypotenusen von ähnlichen rechtwinkli-
gen Dreiecken, und es gilt:

$$\frac{P_l}{f} = -\frac{h+x}{z} \quad \text{bzw.} \quad \frac{P_r}{f} = \frac{h-x}{z}.$$

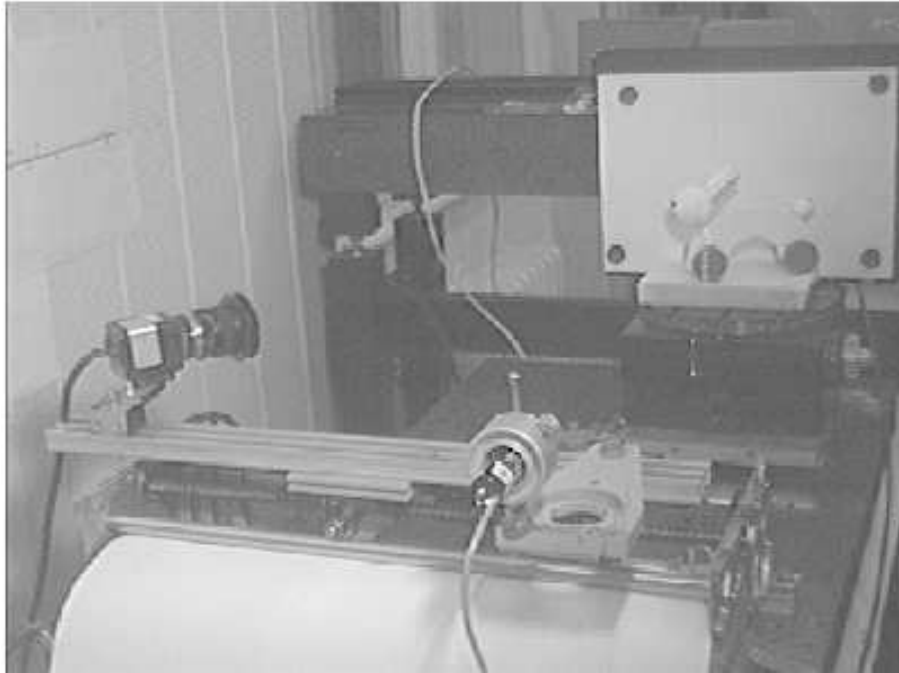
Eliminieren von x ergibt:

$$z(P_r - P_l) = 2hf \quad \text{bzw.} \quad z = \frac{2hf}{P_r - P_l}.$$

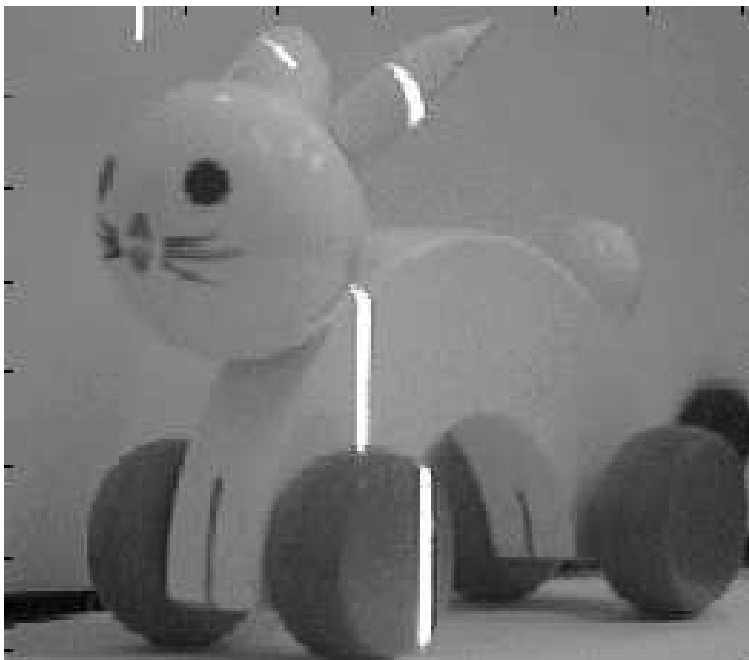
$P_r - P_l$ wird die Disparität zwischen den Stereobildern genannt.
Wenn $P_r - P_l = 0$, liegt der Punkt im Unendlichen.

Range-Kamera

Eine Rangekamera dient zur Tiefenabtastung eines Bildes. Ein Objekt wird mit einem linienförmigen Laserstrahl beleuchtet, das Bild wird von einer CCD-Kamera aufgefangen.

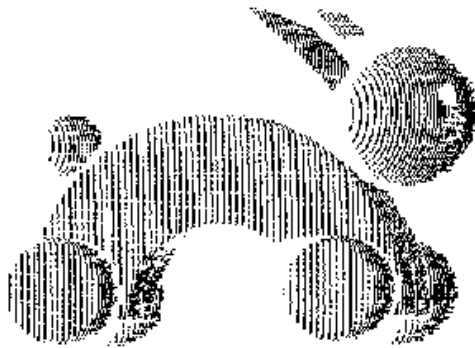


Das Licht trifft in unterschiedlichen Tiefen auf das Objekt, was zu einem Versatz in der seitlich sitzenden Kamera führt.



Range-Kamera II

Aus diesen Profilen kann ein Tiefenbild rekonstruiert werden. Üblicherweise wird dieser Abtastungsprozeß unter zahlreichen Winkeln wiederholt, und im Rechner ein 3D-Modell rekonstruiert.

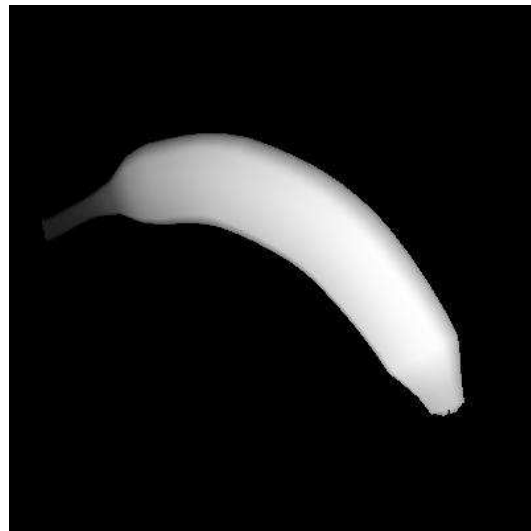
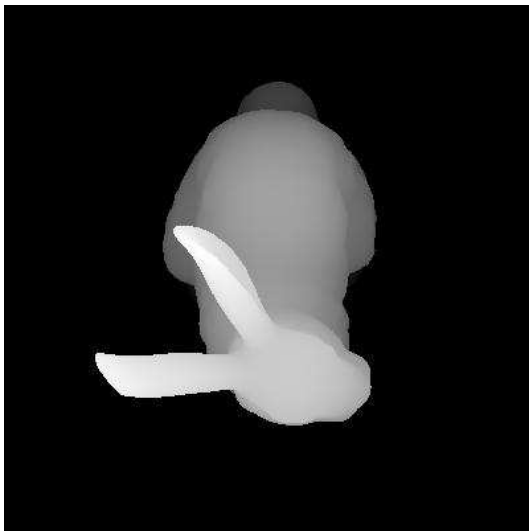
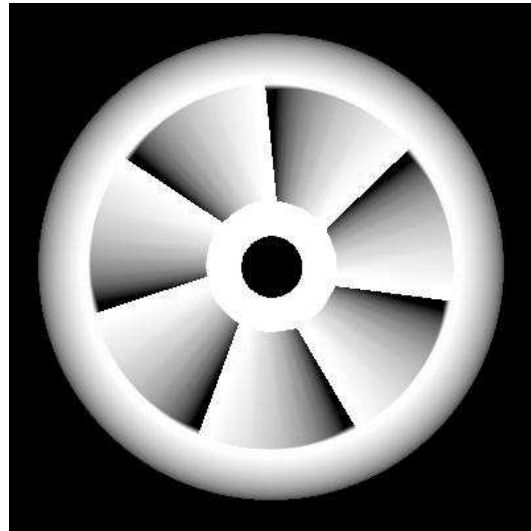
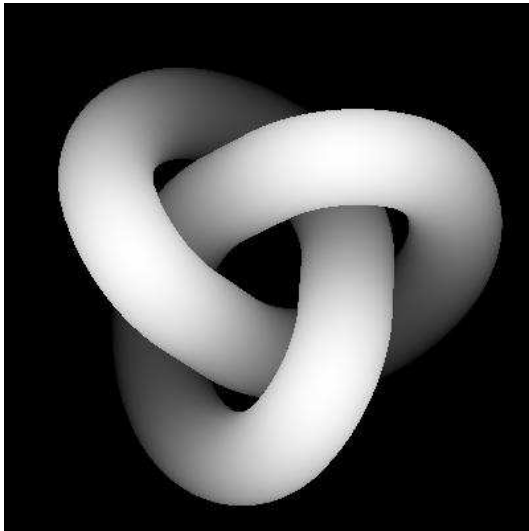


Range-Kameras haben heute große Bedeutung in der Film- und Spieleindustrie, wenn reale Personen oder Tiere als Modelle in den Rechner übertragen werden sollen.

Range-Bilder - Beispiele

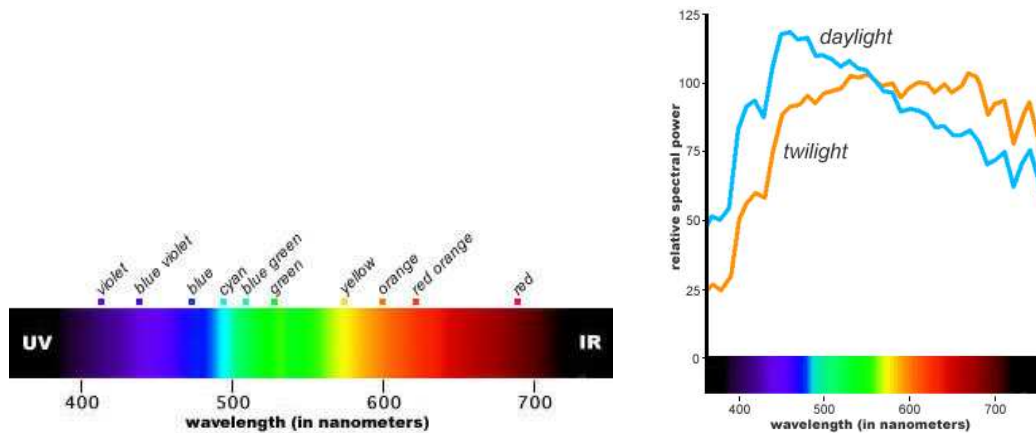
Reproduktionen von Range-Bildern sind unten dargestellt. Die Originale sind zu finden unter:

<http://range.informatik.uni-stuttgart.de>

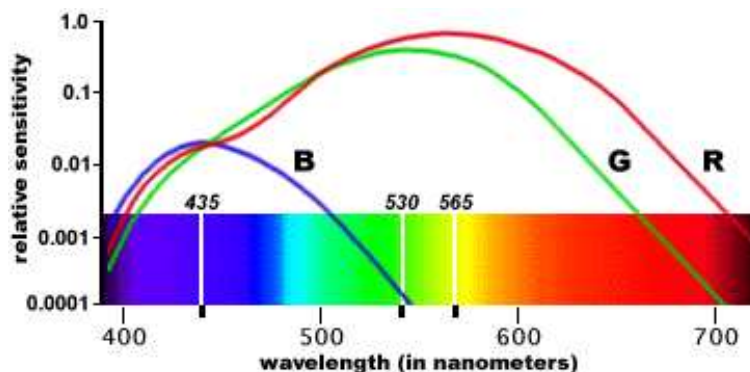


Spektren und Farbsehen

Das sichtbare Spektrum elektromagnetischer Wellen und die spektrale Verteilung des Sonnenlichtes bei unbedecktem Himmel und in der Dämmerung.



Die Fotorezeptoren des Auges setzen sich aus zwei Typen zusammen. Die Zäpfchen sind farbempfindlich; sie haben folgende Sensitivität:

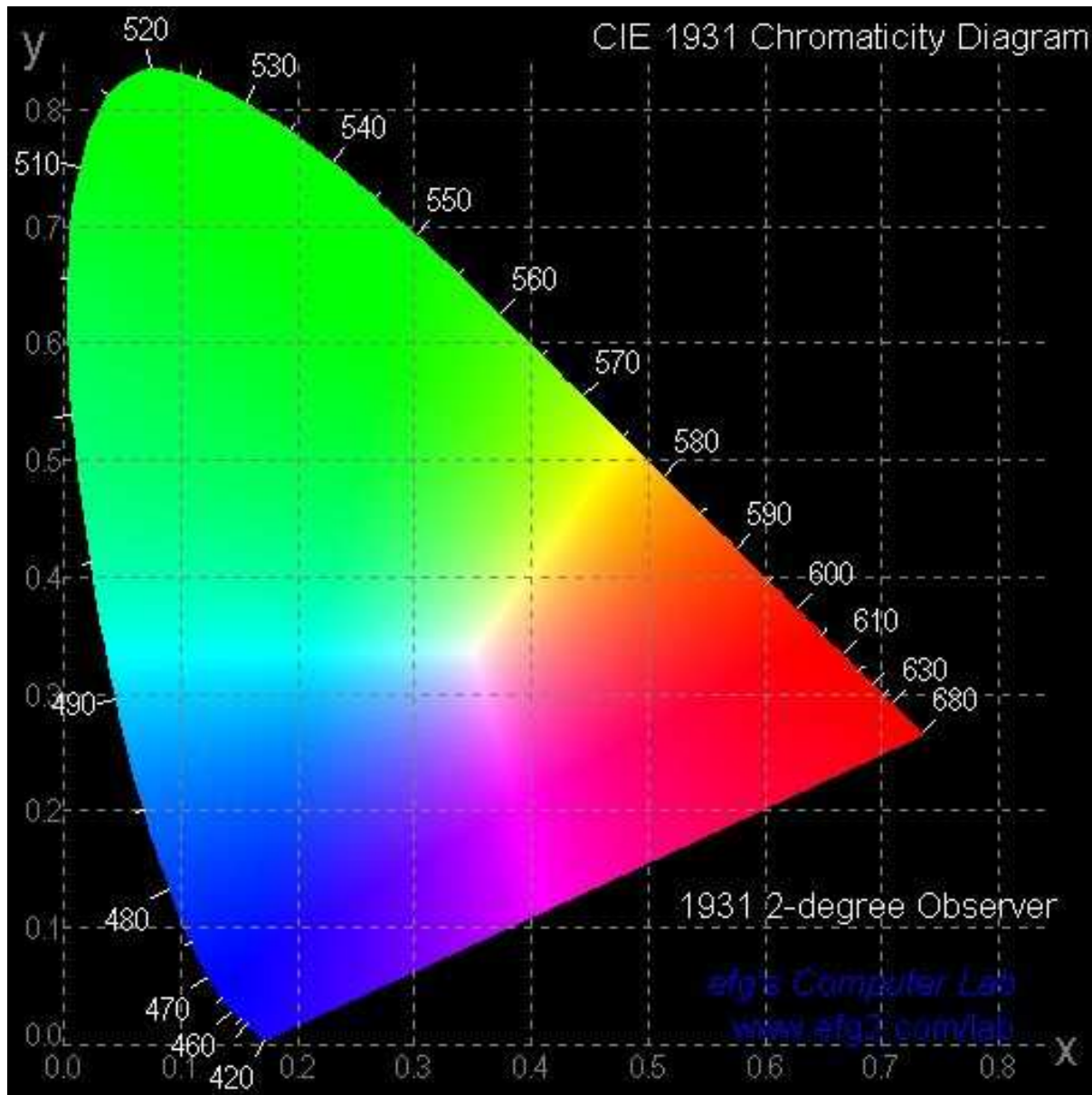


Daher kann Farbsehen als Mixtur von drei "Primärfarben" (RGB) angesehen werden. Die kombinierte Stimulation von R- und G-Zäpfchen erzeugt die Luminanz einer Farbe als dominantem visuellem Eindruck.

Wenn überwiegend R- und G-Zäpfchen stimuliert werden, entsteht der Eindruck von gelb. Wenn überwiegend B-Zäpfchen stimuliert werden, entsteht der Eindruck von blau. Werden alle Zäpfchen ähnlich stimuliert, entsteht ein grauer oder weißer Eindruck. Die Stäbchen haben eine hohe Lichtempfindlichkeit, aber keine Farbspezifität (Dunkelsehen).

Das CIE-Diagramm

Die **Commission Internationale L'Eclairage** hat 1931 drei Primärfarben (X , Y und Z) definiert, aus denen sich alle sichtbaren Farben zusammensetzen lassen. Das Farbdigramm für $X + Y + Z = 1$ (also konstante Luminanz) sieht so aus:



"Reine" Farben liegen auf der Peripherie des Diagramms. Komplementärfarben ergeben in der Mischung mit der Ausgangsfarbe weißes Licht.

Das RGB-Farbmodell

Rot, Grün und Blau können in einem kartesischen Koordinatensystem als einfachstes Farbmodell eingesetzt werden. Es wird **additiv** genannt, da Weiß durch Addition aller drei Kanäle entsteht. Das RGB-Modell wird vor allem in Röhrenmonitoren eingesetzt. RGB- und XYZ-Modell stehen in folgender Beziehung:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.18 & 0.81 & 0.01 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

Das CMY-Farbmodell

Das Cyan-Magenta-Yellow (CMY) Modell wird **subtraktiv** genannt, da eine Farbkomponente von weißem Licht abgezogen wird (z.B. als Druckfarbe). RGB- und CMY-System stehen in einem einfachen Zusammenhang:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

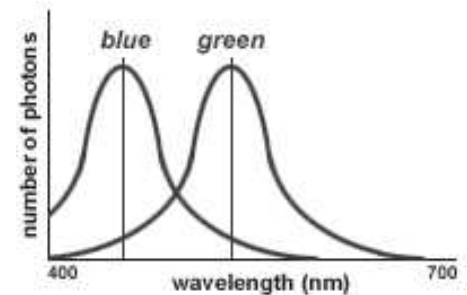
Das YIQ-Farbmodell

Das YIQ-Modell wird zur Übertragung von Fernsehbildern genutzt, um eine Kompatibilität mit Schwarz-Weiß-Geräten zu erhalten. Der Y-Kanal entspricht der Helligkeit. Die Umwandlung von RGB nach YIQ erfolgt durch:

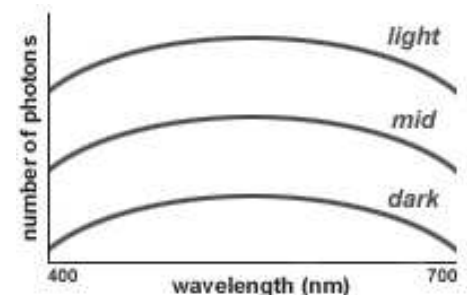
$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.28 & -0.32 \\ 0.21 & -0.52 & 0.31 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

Das Hue-Luminanz-Sättigung (HLS-) Modell I

Hue: die Qualität einer Farbe, die mit einem Namen (rot, grün, gelb, blau) identifiziert wird.

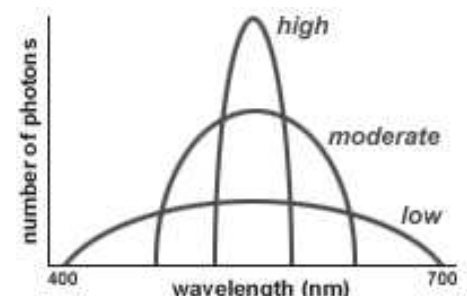


Sättigung und Luminanz sind konstant. Die Hue wird durch die mittlere Wellenlänge bestimmt.



Hue und Sättigung sind konstant. Die Luminanz beschreibt die Intensität einer Farbe.

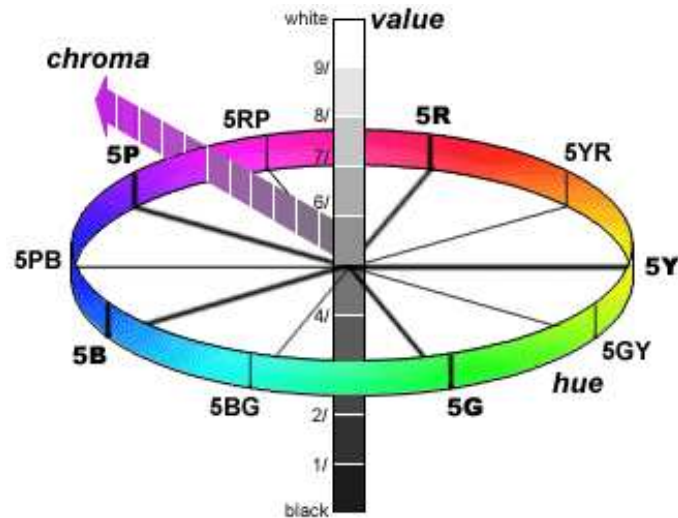
Sättigung: die Reinheit der Hue, unabhängig von der Intensität.



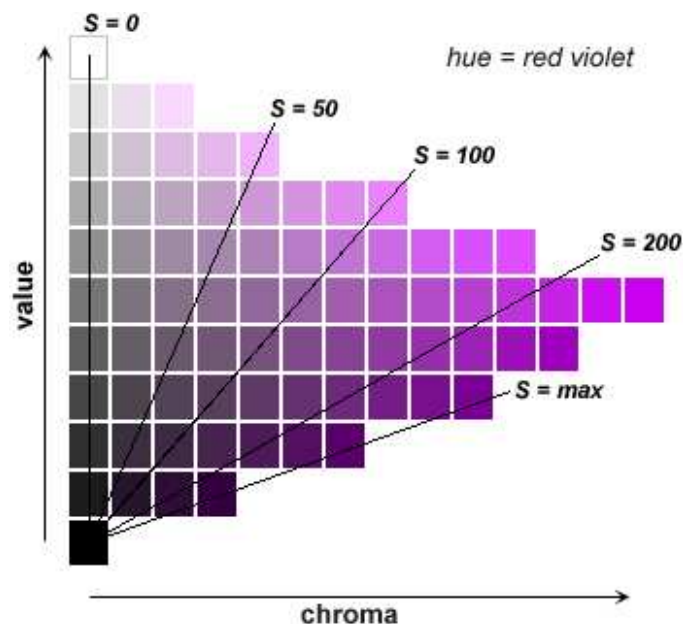
Hue und Luminanz sind konstant. Die Sättigung beschreibt die Farbintensität im Verhältnis zu ihrer Helligkeit.

Das Hue-Lightness-Sättigung (HLS-) Modell II

Das Munsell-Farbschema wird durch 5 Farbdimensionen in gleichen Abständen auf einem hue-Kreis bestimmt: rot, gelb, grün, blau und violett. Der Kreis ist in 100 Abschnitte unterteilt. Radial dazu wird die Sättigung einer Farbe in 10 Stufen definiert, und senkrecht dazu die Luminanz (ebenfalls in 10 Schritten).



Das untenstehende Diagramm erläutert den Unterschied zwischen Luminanz und Sättigung für eine bestimmte Hue (rotviolett).



Die Umwandlung von RGB nach HLS ist komplizierter (s. z.B. Foley, van Dam).

Kamerabilder - Beispiele

Digitale Farbbilder werden heute zumeist direkt durch Aufnahme mit einer CCD-Kamera erzeugt. Einfache Kameras erzeugen Bilder mit recht niedriger Auflösung und Farbtiefe (obere Reihe), während Kameras für professionelle Anwendungen heute Filmqualität erreichen oder übertreffen (untere Reihe).

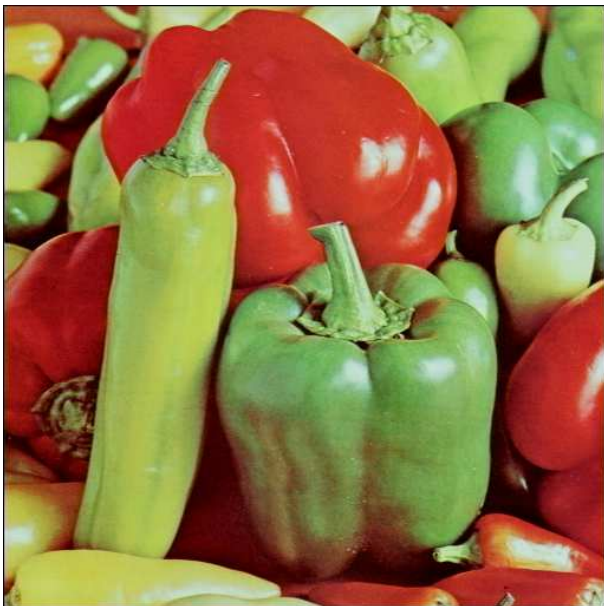
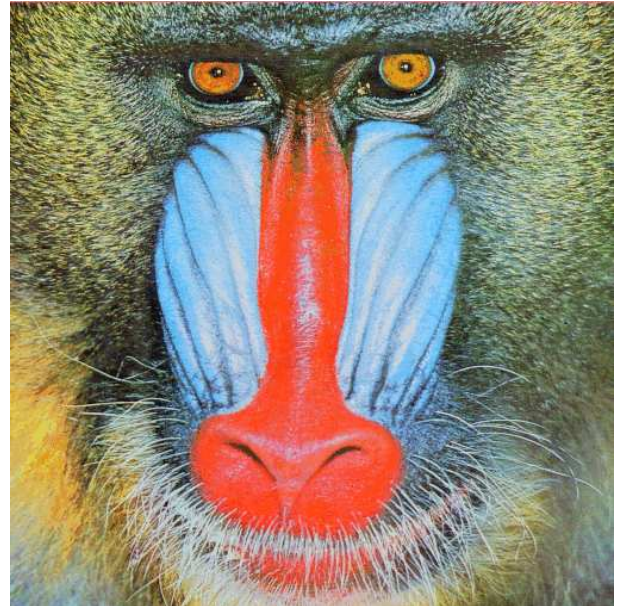


"Berühmte" Testbilder

Oft ist es sinnvoll, die Ergebnisse von Bildverarbeitungsprozeduren an allgemein benutzten Testbildern zu demonstrieren. Viel verwendete Bilder sind Lena (links oben), Mandrill (recht oben), Peppers (links unten) und Monarch (recht unten). Die Originale sind zu finden unter:

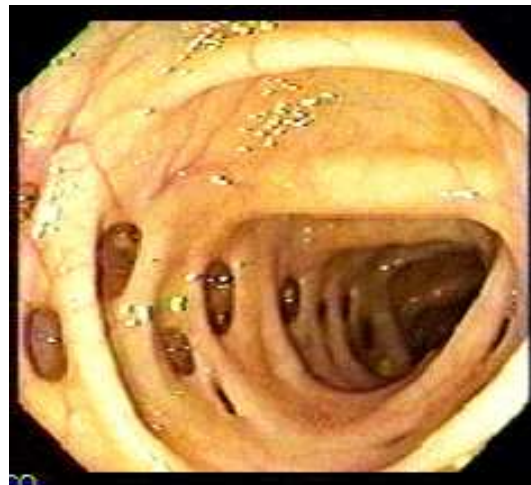
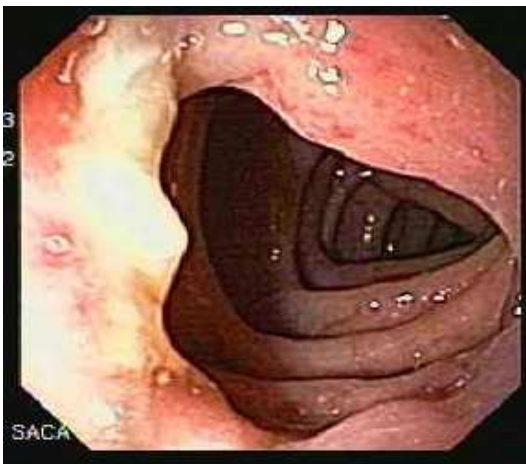
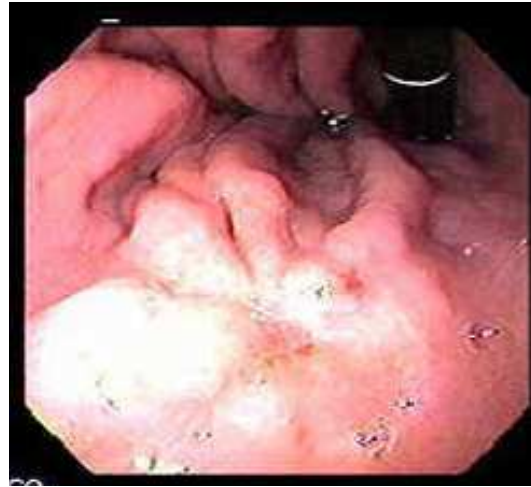
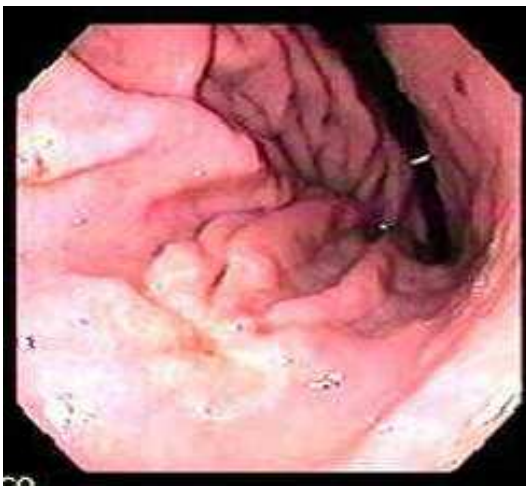
<http://sampl.eng.ohio-state.edu/>

[~sampl/data/stills/waterloo-colorset/index.htm](http://sampl/data/stills/waterloo-colorset/index.htm)



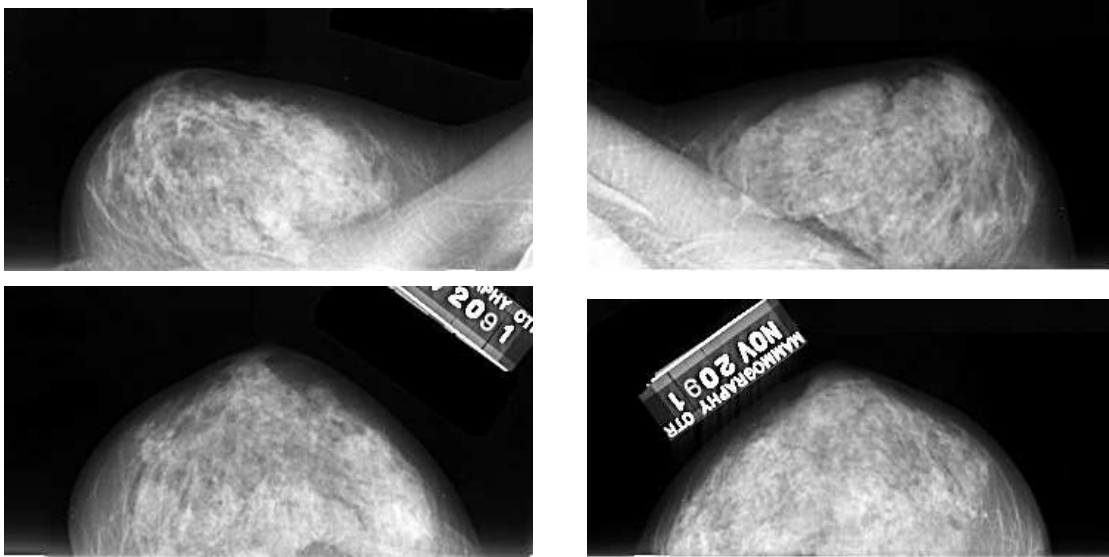
Endoskopie - Beispiele

Endoskopien sind Videoaufnahmen aus dem Inneren des Körpers. Sie dienen dem Chirurgen, um pathologische Befunde an schwer direkt zugänglichen Orten (z.B. im Magen-Darm-Trakt) zu untersuchen und ggf. auch Proben davon zu entnehmen. In der oberen Bildreihe ist ein beginnender Magenkrebs zu erkennen (der weiße Fleck), in der unteren Reihe die Folgen einer entzündlichen Darmerkrankung (die Ausstülpungen des Darms).



Mammographie - Beispiel

Bei der Mammographie wird ein Röntgenbild der weiblichen Brust angefertigt, ggf. unter Gabe von Kontrastmitteln.

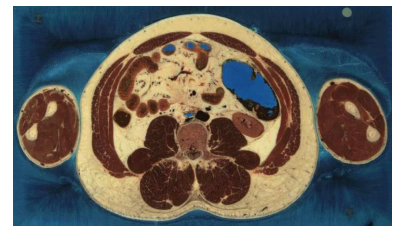
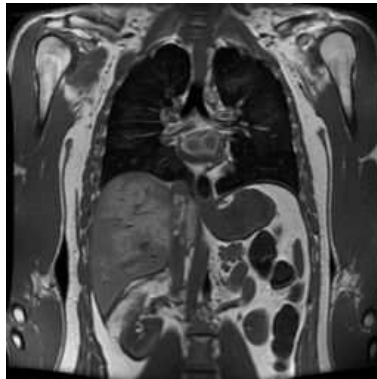
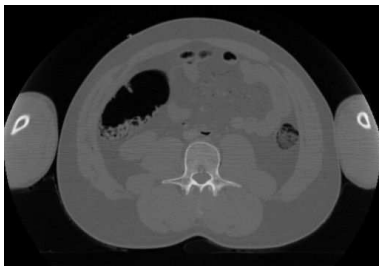
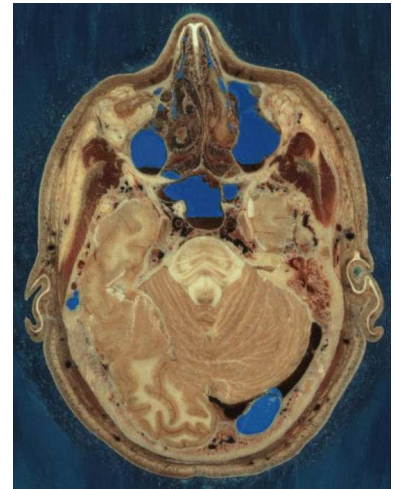
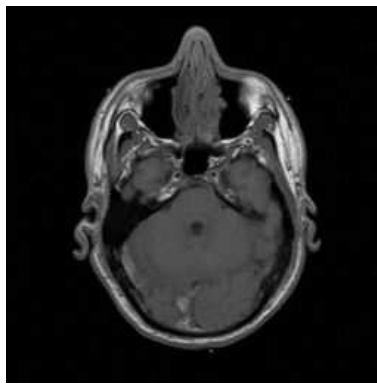
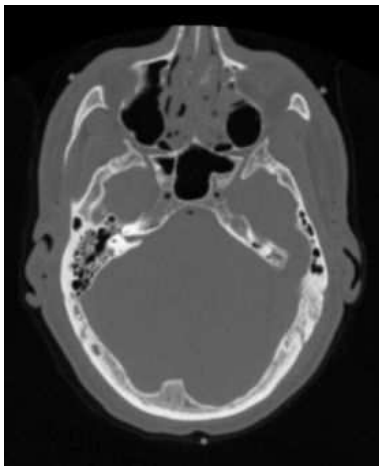


Eine Aufgabe ist z.B., die oberen Aufnahmen mit der Nachfolgeuntersuchung unten zu vergleichen. Hierbei ist die unterschiedliche Positionierung des Weichteilgewebes von echten Veränderungen (z.B. Tumorstadium) zu differenzieren.

Das "Visible Human"-Projekt

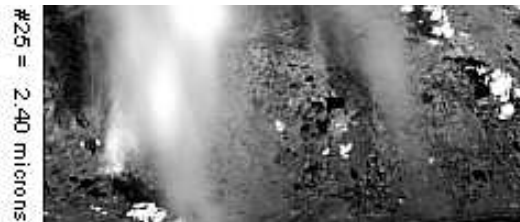
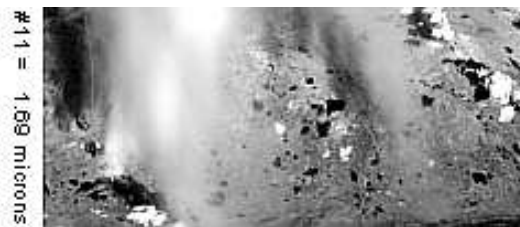
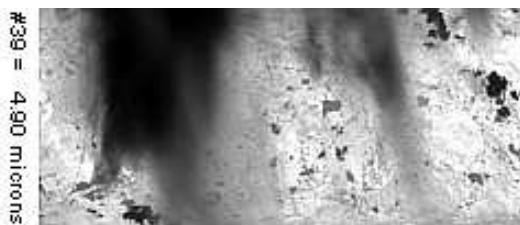
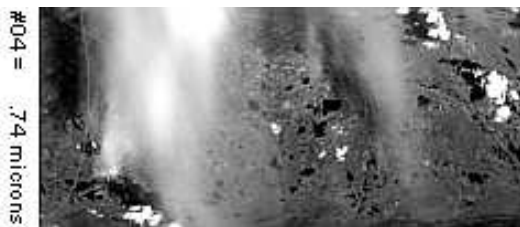
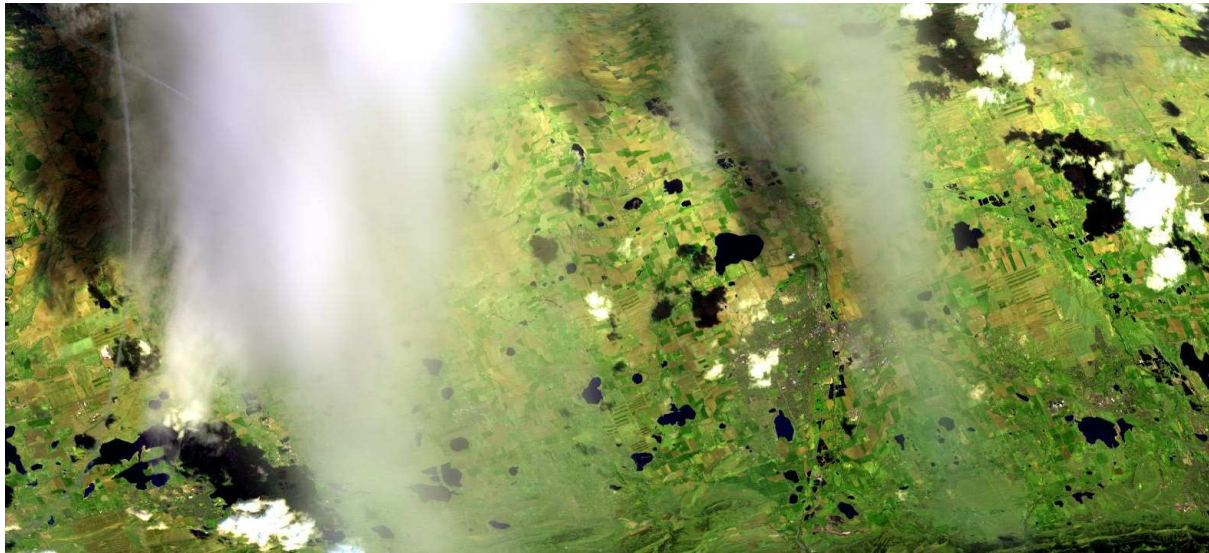
Das Visible Human Projekt der National Library of Medicine zielt darauf ab, eine anatomisch detaillierte, dreidimensionale Repräsentation eines normalen männlichen und weiblichen Körpers zu erhalten. Es wurden axiale Schnitte im CT, im MRT und Gefrierschnitte erzeugt. Die Originalbilder sind zu finden unter:

http://www.nlm.nih.gov/research/visible/visible_human.html



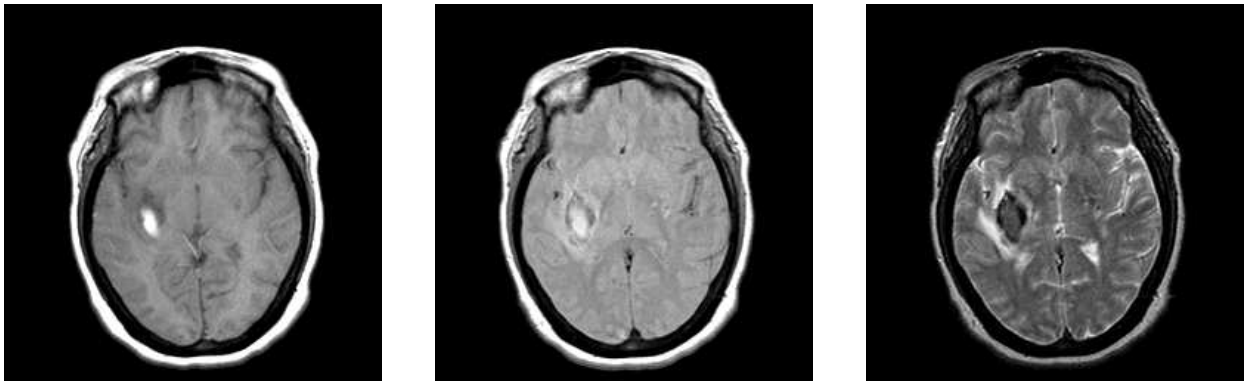
Multispektral-Kameras

Multispektralkameras nehmen mit speziellen Sensoren einzelne Bänder der elektromagnetischen Strahlung auf, auch außerhalb des sichtbaren Lichtes. Sie werden zur Luftüberwachung aus unterschiedlichen Gründen eingesetzt.



MR-Tomographie - Beispiel

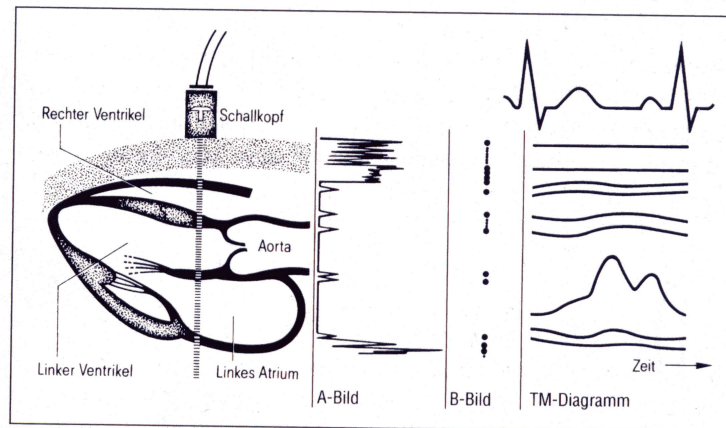
Im Gegensatz zur Röntgen-Computertomographie bietet die MR-Tomographie weiterreichende Möglichkeiten, durch Parametereinstellungen bei der Bildaufnahme den Gewebskontrast zu beeinflussen. So entstehen z.B. T_1 -gewichtete (links), PD -gewichtete (Mitte) oder T_2 -gewichtete Aufnahmen.



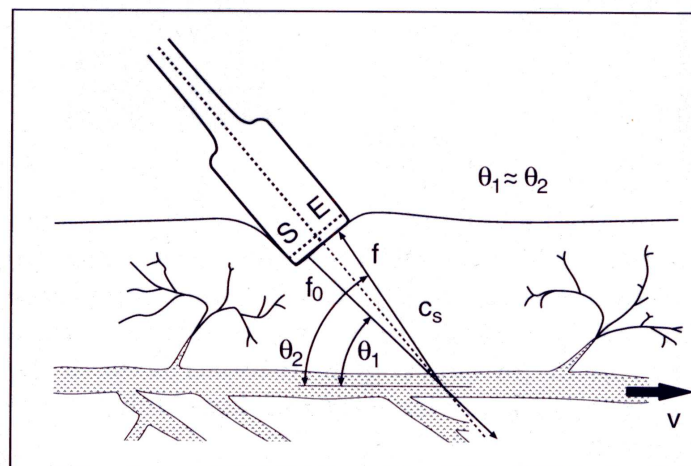
Auf diesem Schnitt durch den Kopf ist frisches Blut (der weiße Fleck im linken Bild) im Hirn als Folge einer Hirnblutung erkennbar. Diese Blutung erzeugt ein Ödem (Schwellung) des umliegenden Hirngewebes (weiße Zone um die Blutung im rechten Bild).

Ultraschall-Bildgebung

Die Ultraschall-Bildgebung ist eine relativ billige, schnelle und strahlungsfreie Bildgebungstechnik. Ein kleiner Transducer wird auf die Haut gesetzt, der im Bereich von 2.2-10 MHz Ultraschallwellen aussendet. Diese Wellen werden an Grenzflächen im Körper reflektiert. Aus der Laufzeit kann auf die Tiefe geschlossen werden.



Beim **Doppler-Verfahren** kann auch die Bewegungsgeschwindigkeit von Objekten gemessen werden, da die reflektierten Wellen eine Frequenzverschiebung aufweisen, die zur Geschwindigkeit proportional ist. Das Vorzeichen der Verschiebung ist durch die Richtung zum Beobachter gegeben.



Ultraschall-Bildgebung - Beispiele

