

Bildanalyse und Bildverstehen

1. Einleitung

"Bildverstehen"

engl. "computer vision", "image understanding"

Bildverstehen ist ein *Prozess*:

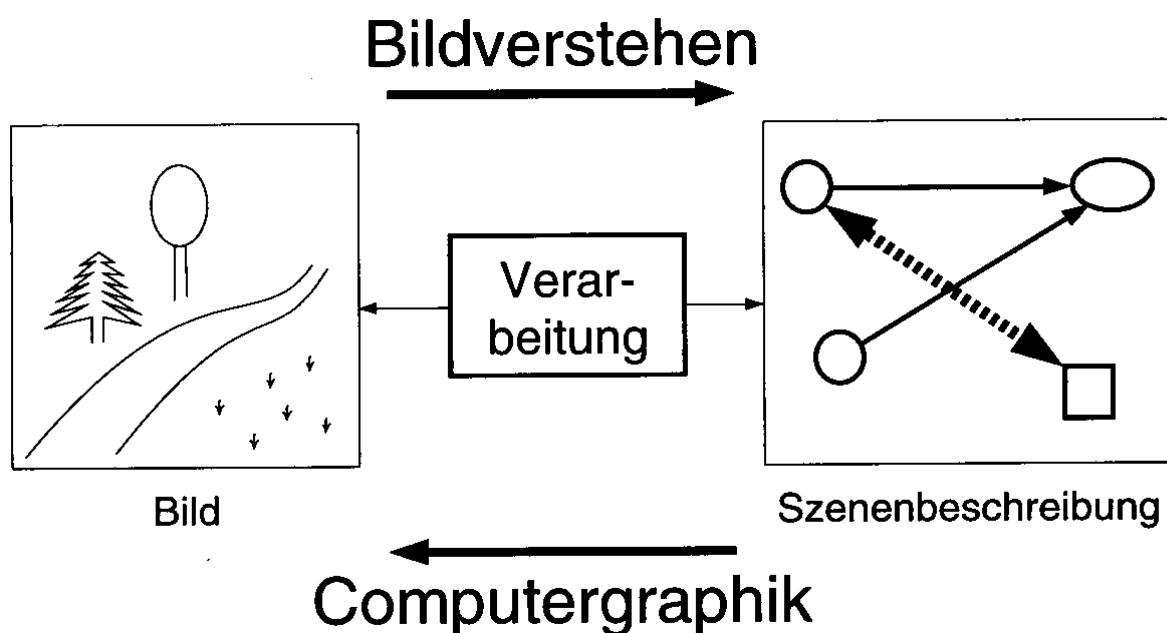
- ausgehend von 1 oder mehreren Bildern und einer Fragestellung
- Resultat ist eine Beschreibung

(genaugenommen: Transformation einer Form von Beschreibung in eine andere)

Bild ist oft 2D-Abbild einer realen 3D-Szene

⇒ für die korrekte Beschreibung ist es oft nötig, diese Szene zu rekonstruieren

Prozess des Bildverstehens komplementär zur synthetischen Computergrafik:



Definitionen des Prozesses "Bildverstehen" aus der Literatur:

"Das Gewinnen einer Repräsentation der Form" (Marr)

"Vom Bild zum Objektmodell" (Winston)

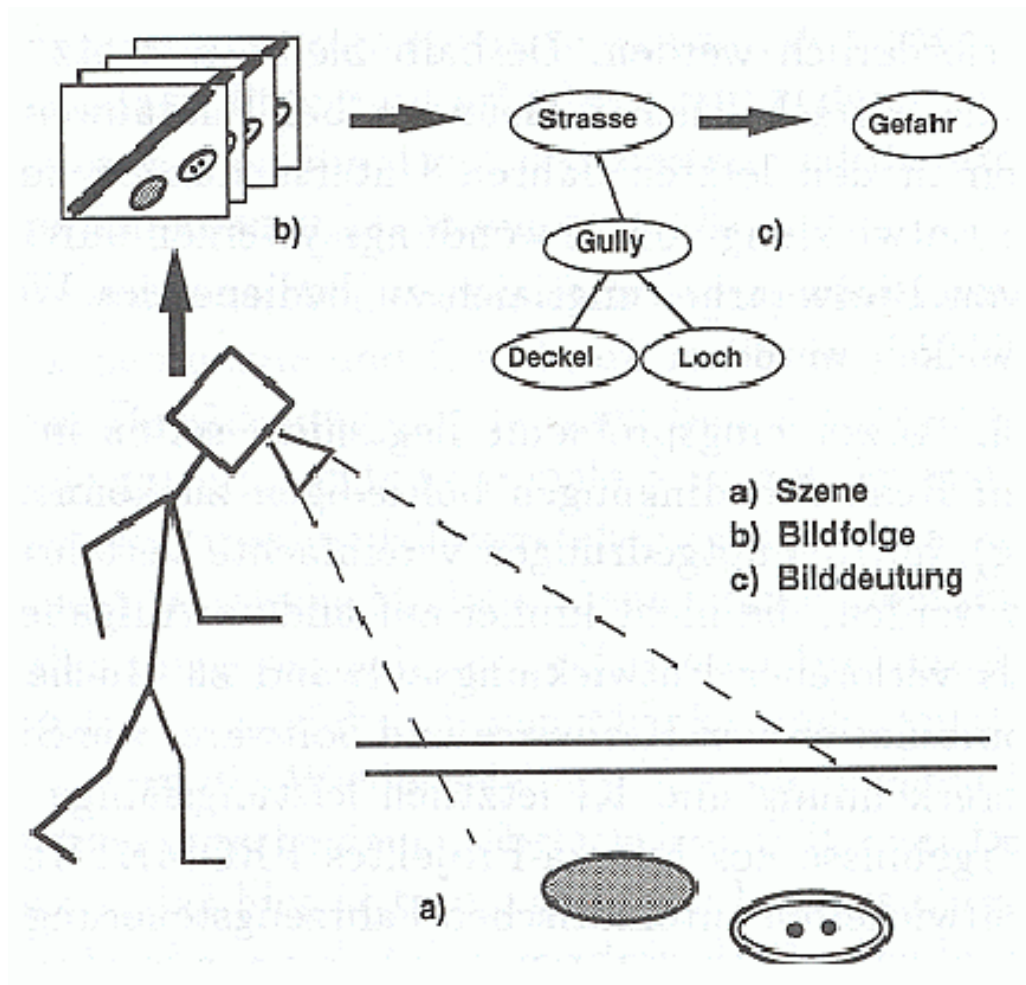
"Die Bedeutung von Bildern mit Hilfe von exakt definierten Prozessen ermitteln" (Niemann & Bunke)

"Bildverstehen ist die Rekonstruktion und Deutung einer Szene anhand von Bildern, so dass mindestens eine der folgenden operationalen Leistungen erbracht werden kann:

- Ausgabe einer sprachlichen Szenenbeschreibung
- Beantworten sprachlicher Anfragen bezügl. der Szene
- kollisionsfreies Navigieren eines Roboters
- planmäßiges Greifen und Manipulieren von Objekten in der Szene" (Neumann 1995)

Verschiedene Ebenen von Realität:

- physikalische Objekte einer Szene (Baum, Haus, Straße, Gully, Deckel, Loch, Regentropfen...)
- Bilder der Szene (als Eingabe): Einzelbild, Stereo-Bildpaar, Bildfolge... je nach bildgebendem Verfahren
- computerinterne Beschreibung der Szene (als Ausgabe): Listen, Merkmalsvektoren, komplexe Objekt- und/oder Szenenbeschreibungen ("is a"- oder "part of"-Hierarchien, Szenengraph, CSG-Modelle, Polygonnetze...)



Ziele des Bildverstehens:

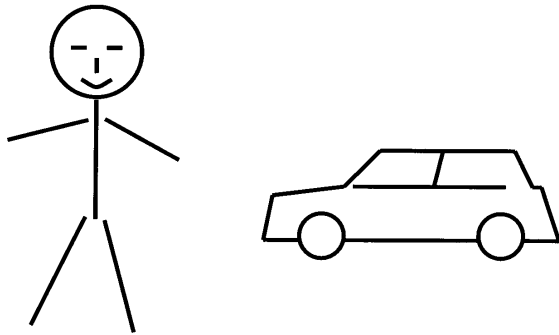
"Wo ist was"

Wo: Rekonstruktion der Szenengeometrie (räumlich-zeitliche Information)

Was: Deutung des Szeneninhalts (insbes.: Objekterkennung)

Die Art der gewünschten Beschreibung und Deutung hängt stark von der jeweiligen Fragestellung / vom Kontext ab.

Beispiel:



(aus Pinz 1994)

Mögliche Interpretationen dieses Bildes:

- ein Binärbild (ein Bild mit nur 2 Grauwerten)
- 3 Kreise, 1 Bogen und 20 Strecken
- "Mensch" und "Auto"
- ein lachender Mensch
- ein Fußgänger überquert die Straße

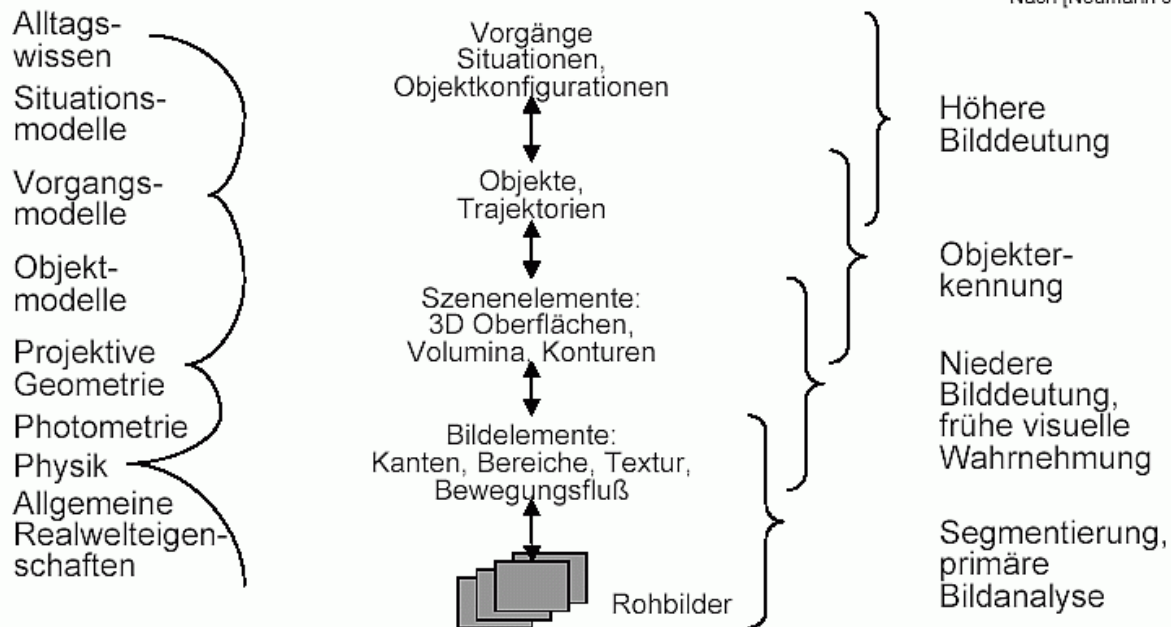
⇒

es gibt keine allgemeingültige "korrekte" Arbeitsweise eines bildverstehenden Systems

die korrekte Arbeitsweise hängt von der jeweiligen Fragestellung ab

Konzeptueller Rahmen eines Bildverstehenden Systems

Nach [Neumann 95]



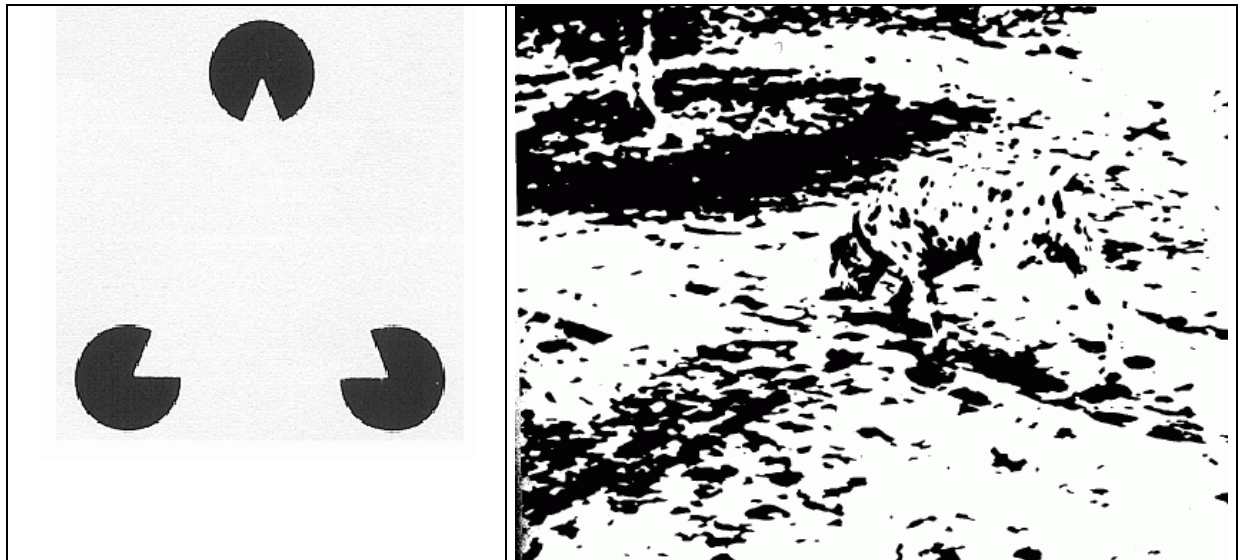
(aus Hermes 2001)

BV-System

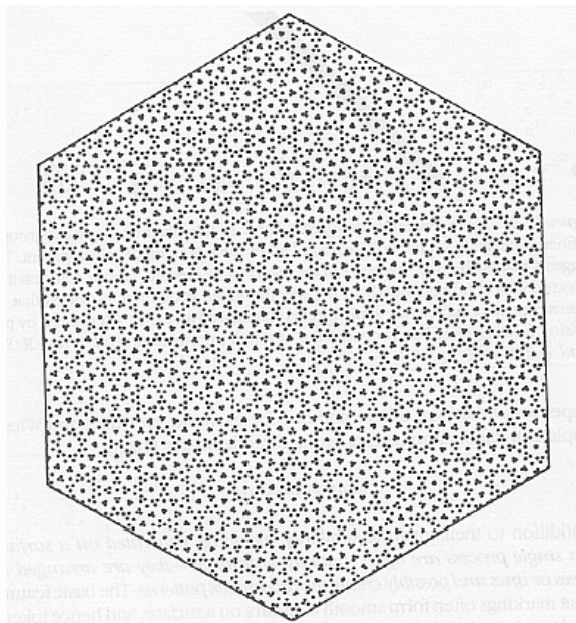
- **Bilderfassung**
 - Kamera, Satelliten, NMR/CT,...
- **Vorverarbeitung**
 - Rauschunterdrückung, Kontrastverstärkung,...
- **Segmentierung**
 - Konturen, Flächen,...
- **Merkmalsextraktion**
 - Kettencode, Umfang, Schwerpunkt, Farbe,...
- **Klassifikation**
 - Objekterkennung, Einteilung in Klassen, ...

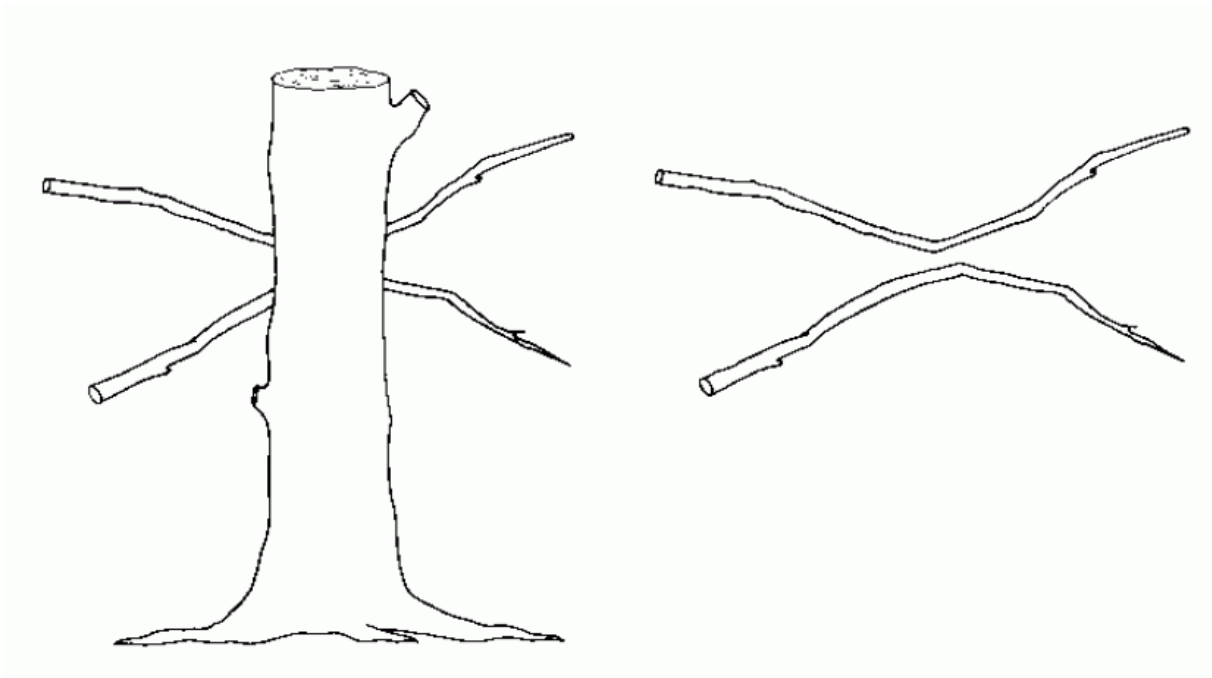
erforderliche Spezialleistungen, die das menschliche visuelle System "ohne Schwierigkeiten" erbringt:

Ergänzen "subjektiver Konturen"



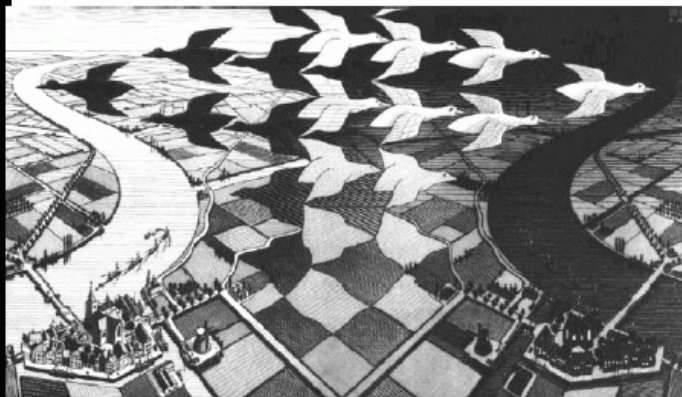
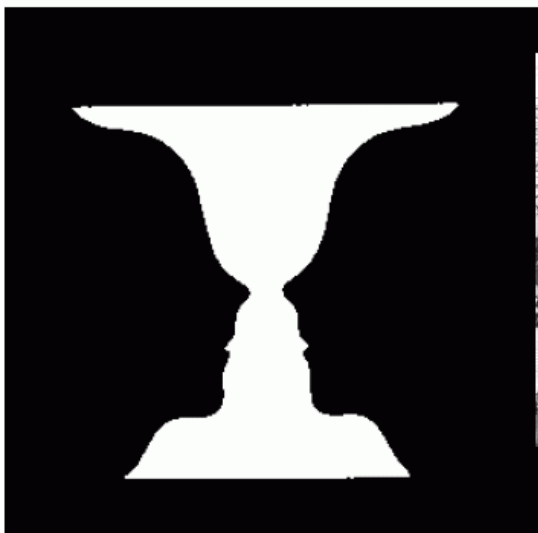
aktive Gruppierungsprozesse:





Separation in Figur und Grund

- Ermittlung von Grenzlinien
- Bestimmen zusammenhängender Bereiche
- Unterscheidung Hintergrund/Vordergrund



(Woessner, o.J.)

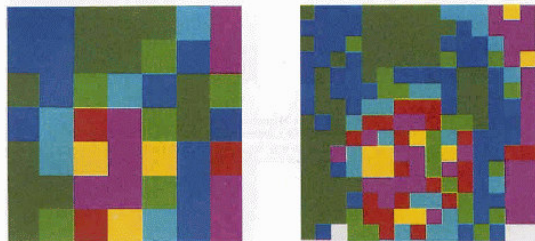
Erkennen "subjektiver Objekte" (insbes. Gesichter, menschliche Figuren):



<http://www.bnr-art.com/doolitt/foresteves.htm>

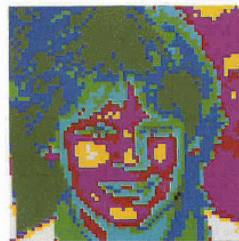
Wichtige Verarbeitungsschritte in bildverstehenden Systemen:

1. Quantisierung der Ortskoordinaten (**Rasterung**)



(A)

(B)



(C)



(D)

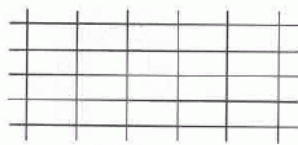
- Digitalisierung ist die Rasterung des Bildes auf 2D-Bildmatrix

$$g(x_1, x_2) \xrightarrow{D} g_{m,n}$$

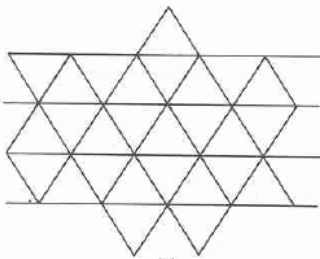
$$x_1, x_2 \in R \quad m, n \in N$$

- Größe des Rasters hat wesentlichen Einfluss auf Detailgenauigkeit
- N Spalten und M Zeilen und m Zeilenindex sowie n Spaltenindex

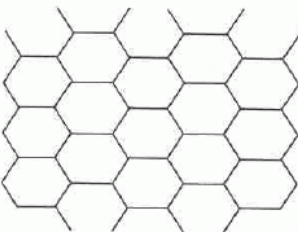
[Abmayr 1994], S. 289



(a)



(b)

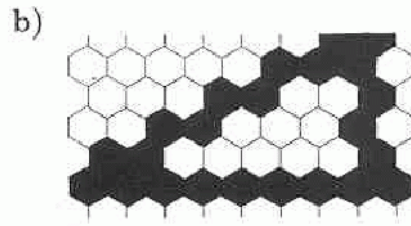
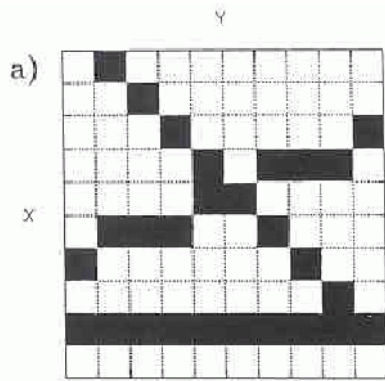


(c)

- Tessellation: Rasterung/Nachbarschaften

- rechteckig
- dreieckig
- hexagonal
- ...

[Ballard & Brown 82], S. 39



- Problem gerader Linien
- Drehungen nur um bestimmte Winkel möglich, z.B.:
 - rechteckig:
Vielfaches von 90°
 - hexagonal:
Vielfaches von 60°



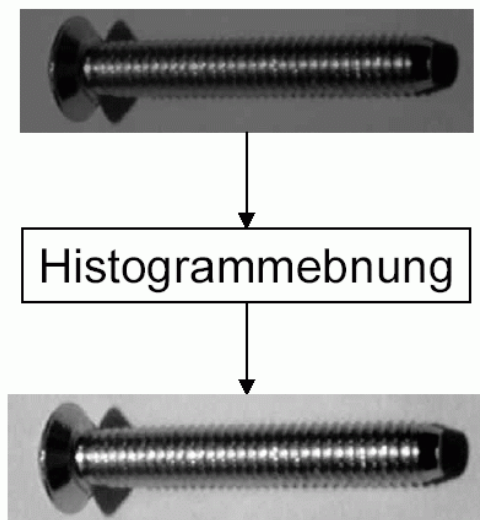
- Anzahl der Bits pro Pixel bestimmt Detailgenauigkeit
- „contouring“ bei geringer Anzahl

2. Bildvorverarbeitung:

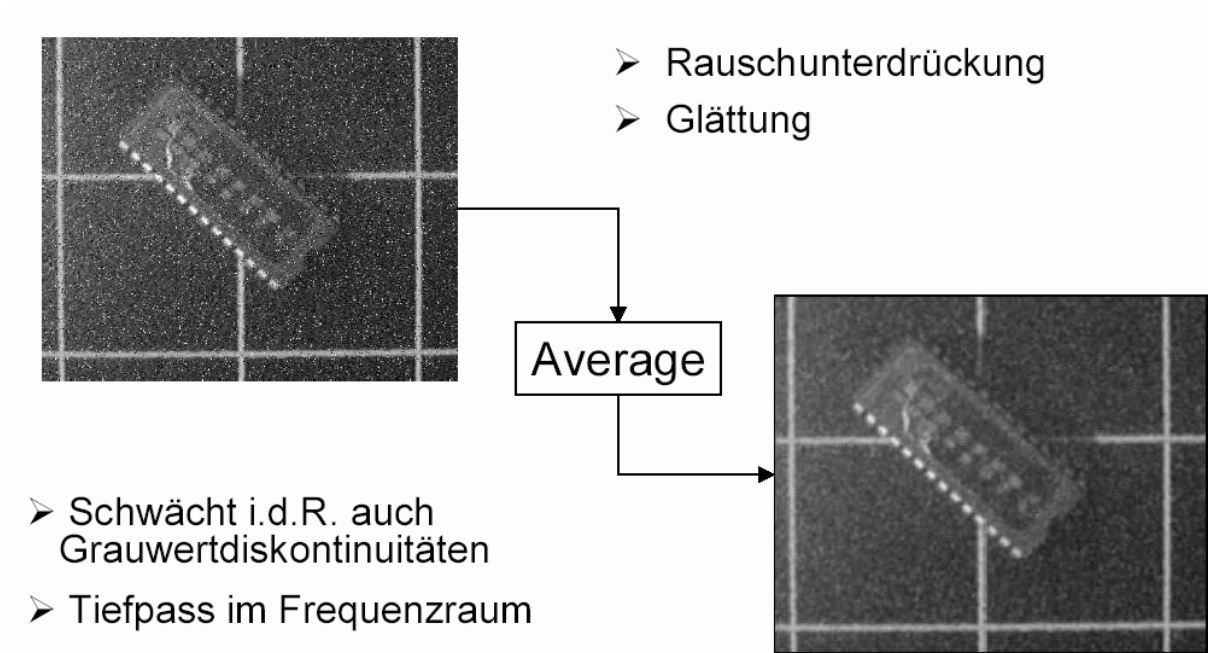
- Bildmodifikation, die den Informationsgehalt nicht wesentlich verändert
- Bildtransformation mit Eingabe- und Ausgabebild

Ziele:

- Beleuchtungskorrekturen bei inhomogener Beleuchtung
- Glättung
- Beseitigung von Digitalisierungsfehlern
- Kontrastverstärkung
- Normierung (Größe, Form, Farben)
- Korrektur von Grauwertverzerrungen (fotografischer Effekt)
- Ausgleich von kamerabedingten Verzerrungen
- selektive Filterung bestimmter Frequenzen



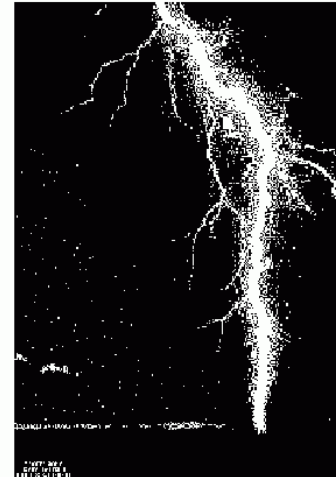
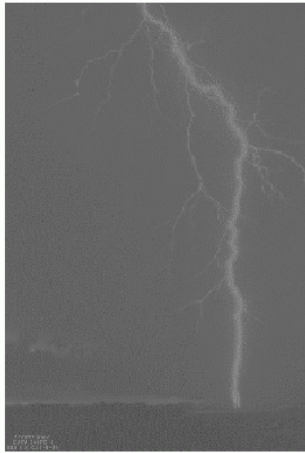
- Grauwertäqualisation
- Gleichverteilung der Grauwerte über der Skala
- I.d.R.: Verbesserung der visuellen Erkennbarkeit
- Kontrastverstärkung



3. **Segmentation** (auch **Segmentierung**):

- Trennung der zu untersuchenden Objekte von den übrigen Bildstrukturen
- Trennung von sich berührenden Objekten voneinander
- Zerlegung in Teilobjekte
- Klassifikation auf Pixelebene (Objektpixel / Nichtobjektpixel)
⇒ geeignete Merkmale auf Pixelebene?
- einfache und komplexe Verfahren

- Punktorientierte Verfahren
- Einfaches Schwellwertverfahren
 - Ideal: bimodales Histogramm
 - Schwellwert zwischen den Maxima



Zerlegung des Bildes in zu interessierende Strukturen (Objekte) und nicht zu interessierende Strukturen (Hintergrund) gemäß:

- Diskontinuitätskriterium (Kanten)
- Homogenitätskriterium (Intensitätswert, Farbe)
- hybride Verfahren (Kombination z.B. durch Region-Growing-Verfahren, vgl. Saatfüll-Algorithmus der 2D-Computergrafik)

4. Merkmalsextraktion

Flächen:

Intensitätswert / Farbe

Umfang, Größe

Schwerpunkt

Hauptachsen

Polygonzerlegung

...

Konturen:

Kettencode

Polygon

Länge

Orientierung

...

Texturen:

statistische Merkmale

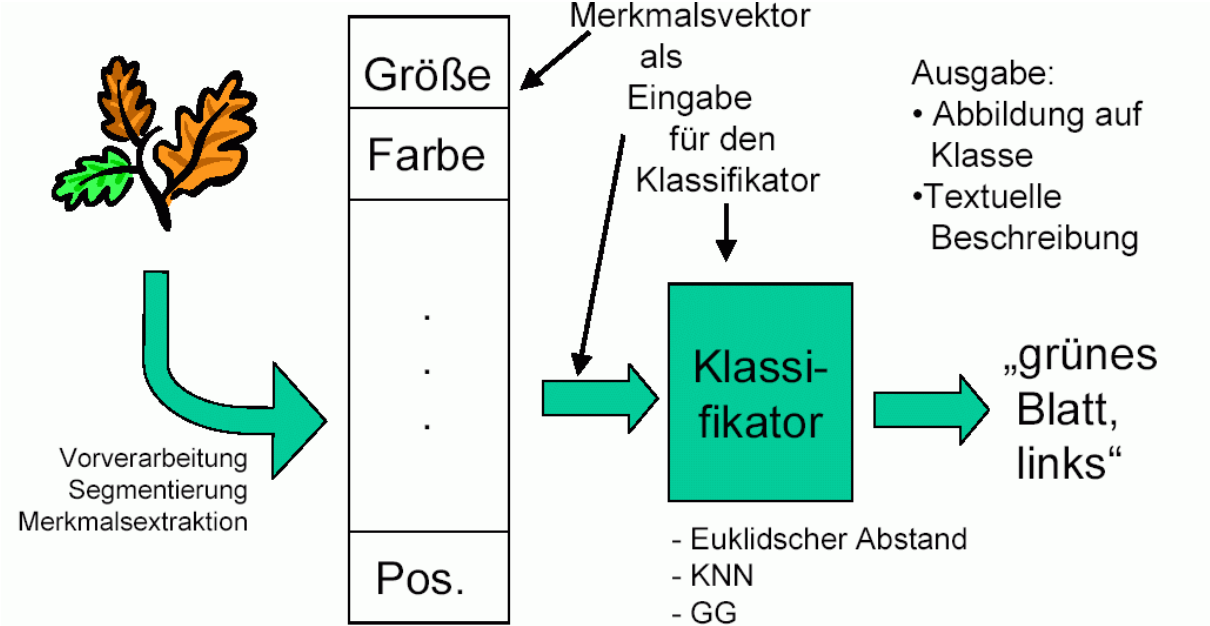
visuelle Eigenschaften

...

statistische Merkmale:

Informationsgehalt, Momente (Mittelwert, Varianz, Schiefe, Kurtosis... einer Verteilung), Anzahl der Pixel

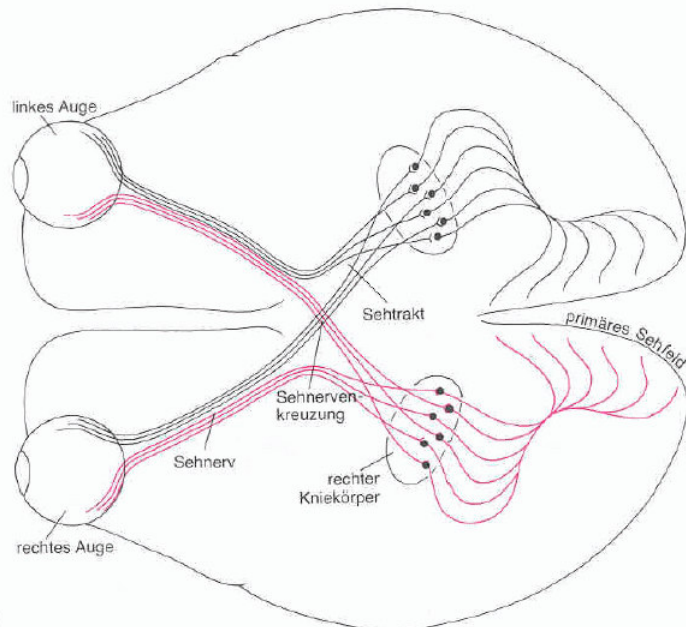
5. Klassifikation



Wie löst das menschliche visuelle System diese Aufgaben?

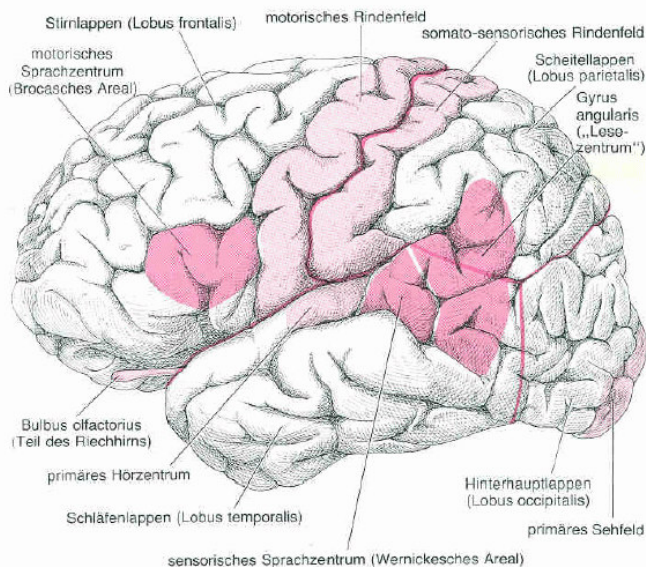
Biologische Grundlagen

Sehbahn



[Spektrum 86], S. 40

Die Großhirnrinde:



[Spektrum 86], S. 29

- 2mm dicke Schicht
- Ca. 100.000 Neurone pro mm^2
- entfaltet ca. 1350 cm^2
- besteht aus 6 Schichten
- Beobachtungen (Ende 19tes Jhd.): verschieden örtliche Schädigungen rufen unterschiedliche Störungen hervor

Teile des Körpers entsprechen bestimmten sensorischen und motorischen Feldern der Großhirnrinde: besonders empfindliche Regionen werden auf relativ große Bereiche abgebildet

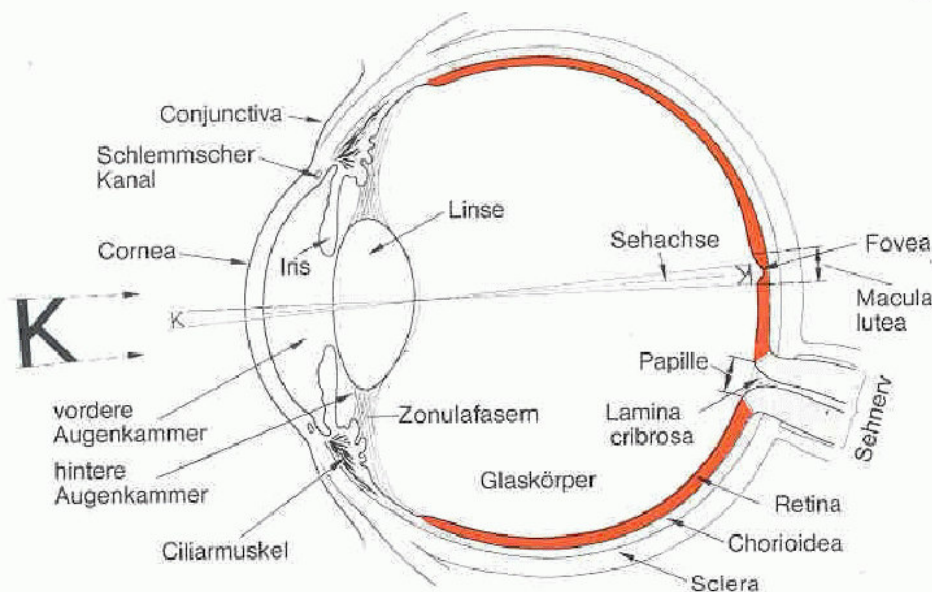
Die Sehbahn beginnt im Auge (vgl. Fachpraktikum Computergrafik, Kap. 2)

Aufgabe der Augen: auf beiden Netzhäuten ein klares, scharfes Bild festhalten

- 3 Muskelpaare; Nachführen der Augen in wenigen Bogenminuten
- Hornhaut (Cornea) und Linse dienen der Scharfstellung
- Durchmesser der Pupille bestimmt einfallende Lichtmenge (Öffnen und Schließen durch spezielle Muskelfasern)
- Selbstreinigungsfunktion

Auge: schematisch

[Schmidt 80], S. 156

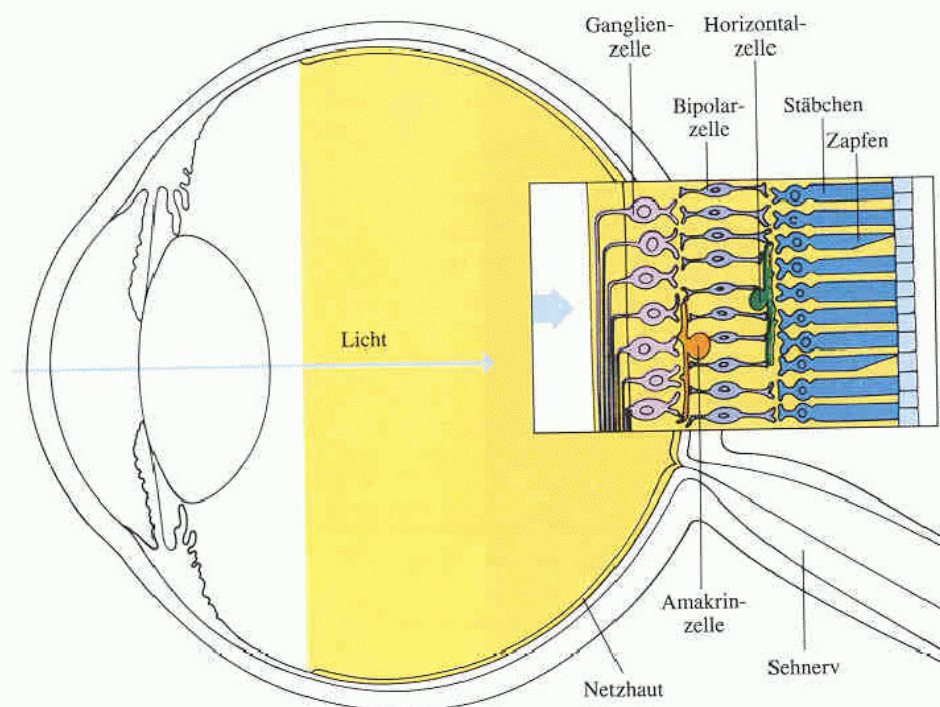


Die Netzhaut (Retina): wandelt Lichtsignale in Nervensignale (elektrische Signale) um, unterscheidet verschiedene Wellenlängen breiter Arbeitsbereich (Sonnenlicht / Sternenlicht) Präzision: aus ca. 1 m Entfernung kann noch ein Haar gesehen werden

Netzhaut ist Teil des Gehirns (verbunden über den Sehnerv)

- Dicke der Netzhaut: ca. $\frac{1}{4}$ mm
- 3 Schichten Nervenzellen + 2 Zwischenschichten
- Photorezeptoren: Zapfen und Stäbchen
- Stäbchen für Dämmerungssehen, Zapfen für Farbsehen
- in der Mitte der Retina (Fovea centralis) nur Zapfen
- Zellreihe hinter den Photorezeptoren enthält Farbstoff Melanin (schluckt einfallendes Licht)
- Schichten vor den Rezeptoren sind nahezu durchsichtig

Netzhaut
(Retina)



[Hubel 90], S. 46

Netzhaut (Retina)

➤ Mittlere Schicht besteht aus

➤ Bipolarzellen:

Input von den Rezeptoren, viele Bipolarzellen projizieren direkt auf die Ganglienzellen

➤ Horizontalzellen:

Verknüpfungen von Bipolar und Rezeptoren (rel. Lange Verbindungen, parallel zu den Retinaschichten)

➤ Amakrinenzellen:

Verknüpfungen von Bipolar. und Ganglien

➤ vorderste Schicht besteht aus Ganglienzellen

➤ Bündelung in der Papille

➤ Verlassen dort als Sehnerv das Auge

➤ Ca. 125 Mio. Stäbchen und Zapfen aber nur ca. 1 Mio. Ganglienzellen

➤ Frage: *Wie kann trotz dieser Datenreduktion detaillierte visuelle Information erhalten bleiben?*

➤ Antwort: *Untersuchung der Verbindungen zwischen den Netzhautzellen.*

➤ Direkter und indirekter Informationsfluss

➤ Direkt:

➤ Rezeptoren -> Bipolar -> Ganglien

➤ Sehr kompakt und hochspezialisiert

➤ Indirekt:

➤ Rezeptoren -> *Horizontal* -> Bipolar -> *Amakrinen* -> Ganglien

➤ Wg. Lateraler Verbindungen diffuser und ausgedehnter

Rezeptives Feld:

Gesamtfläche der Rezeptoren auf der Netzhaut, die auf eine Ganglienzelle abbilden

- das rezeptive Feld beeinflusst das Feuern dieser Ganglienzelle
- Spontanentladung (unregelmäßig) 1–20 mal pro Sek.
- Verschaltung variiert von der Fovea bis zum Rand der Netzhaut

Entdeckung von Kuffler (1950) bei systematischem Absuchen der Netzhaut mit kl. Lichtfleck:

On-Zentrum- und Off-Zentrum-Neuronen

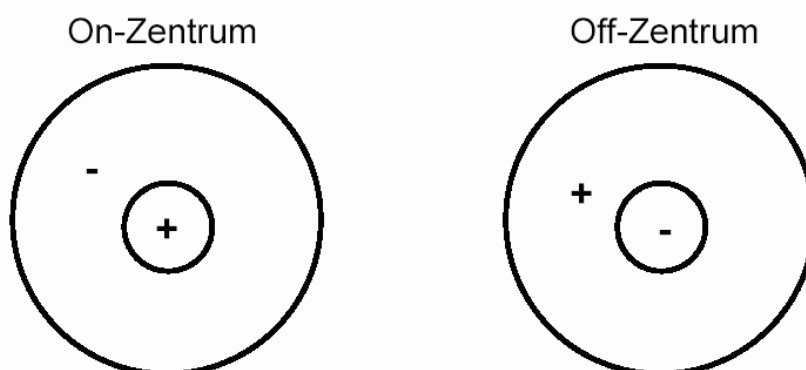
On-Zentrum: erhöhte Impulsrate, wenn das Zentrum des rezeptiven Feldes mit Lichtfleck gereizt wird ("On-Reaktion")

Off-Zentrum: Unterdrückung der Aktivität, starke Entladung nach Ausschalten des Lichtreizes ("Off-Reaktion")

je mehr von bestimmter On/Off-Region mit passendem Stimulus bedeckt, desto heftiger die Reaktion

Netzhaut (Retina)

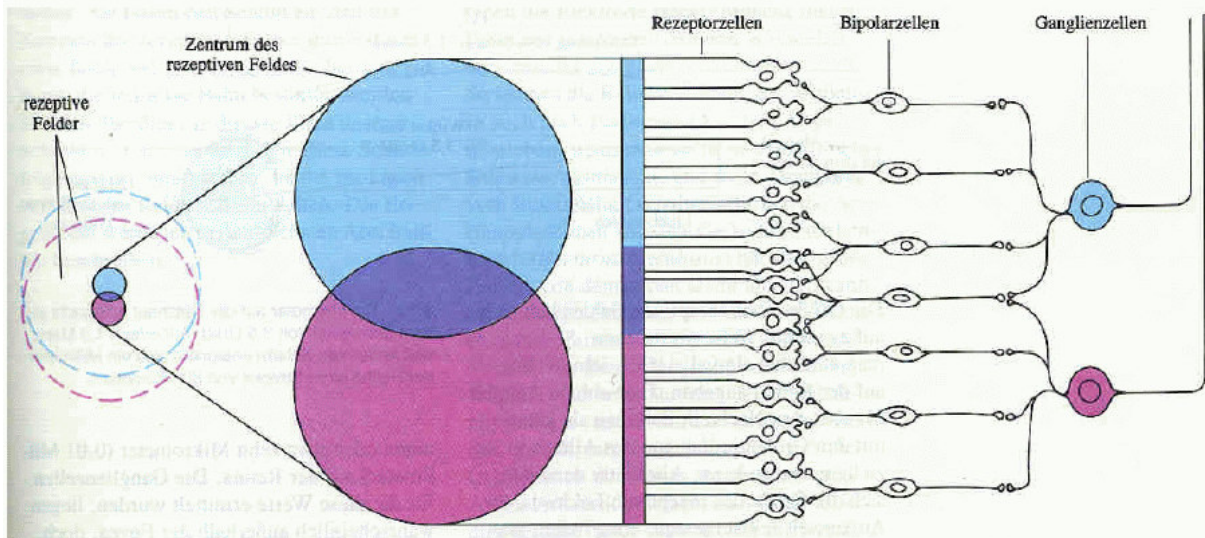
➤ Rezeptive Felder (RF)



je tiefer man ins Zentralnervensystem vordringt, desto komplexer die rezeptiven Felder der einzelnen Neuronen

Überlappung einzelner rezeptiver Felder von benachbarten Ganglienzellen

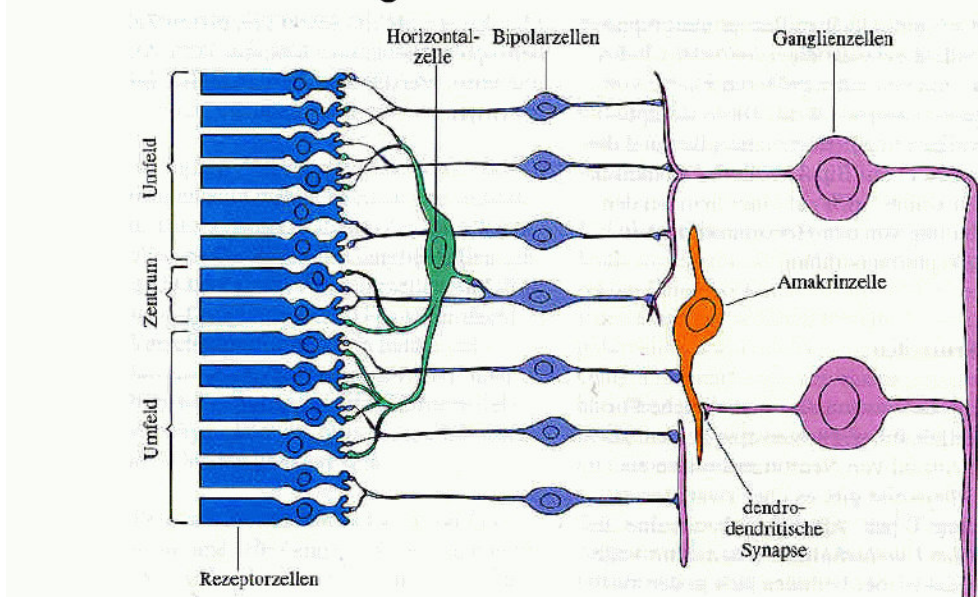
Überlappende RFs



[Hubel 90], S. 53

auch die Bipolarzellen besitzen rezeptive Felder

RF & Verschaltung von BPZ



Chiasma opticum: Sehnervkreuzung
beide Sehnerven teilen sich auf beide Gehirnhälften auf
"seitliche Kniehöcker": Zwischenverarbeitungsbereich
zwischen Chiasma und visuellem Cortex

Visueller Cortex (Sehrinde):

6 Schichten

- simple cells: sensitiv gegenüber Lichtschlitz, Balken, Gerade
- complex cells: sensitiv gegenüber kurzem Aufleuchten und evtl. Bewegung
- hyper complex cells: sensitiv gegenüber Bewegung und Orientierung

Zusammenfassung zur Sehbahn:

- Hierarchischer Aufbau
- massiv parallele Signalverarbeitung
- Datenreduktion bei Merkmalsextraktion
- viele Einzelheiten noch ungeklärt

2 Ansätze zur Untersuchung der Funktionsweise des visuellen Systems:

- bottom-up: von einzelnen Neuronen und ihrer Funktion ausgehend (siehe oben)
- top-down: von den Wahrnehmungsleistungen ausgehend (Versuchspersonen) – kognitive Psychologie, Gestaltwahrnehmung

idealerweise sollten sich beide Ansätze "in der Mitte treffen" – davon ist man noch weit entfernt

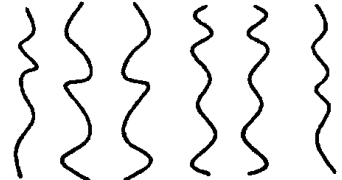
Gruppierungsprinzipien der Gestaltwahrnehmung:

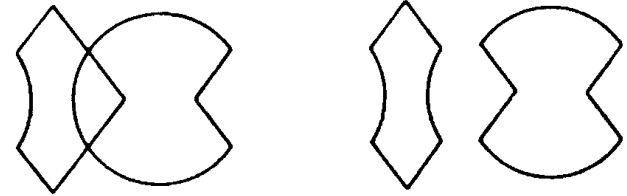
a) Proximity

b) . . ● ● . . ● ● . . ● ● . . Similarity

c) [] [] [] [Closure

d)  Continuation

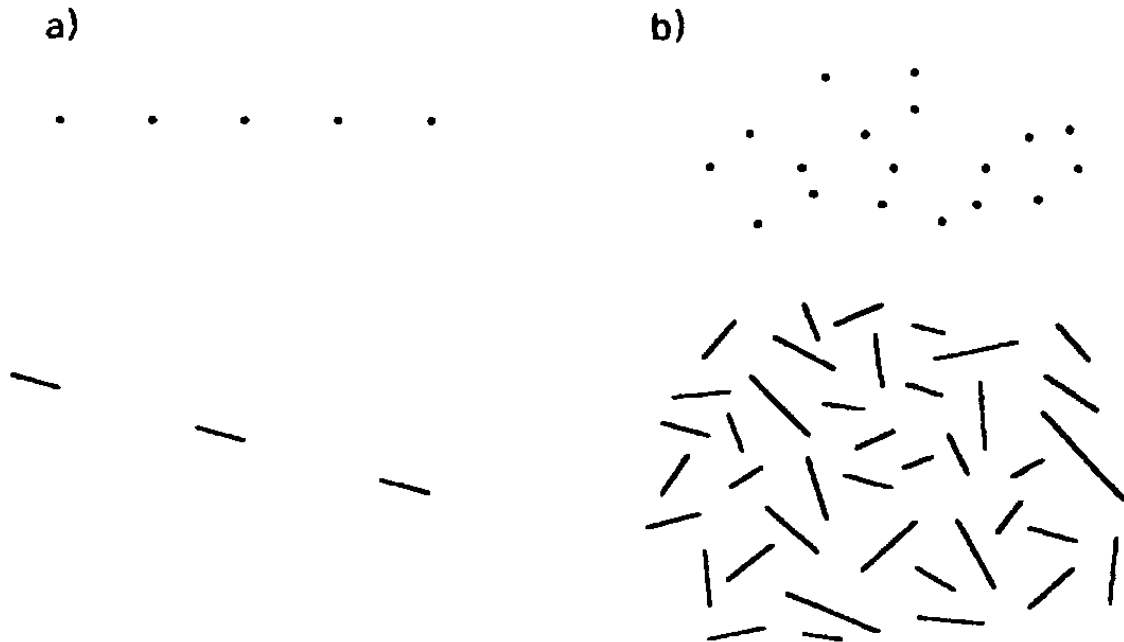
e)  Symmetry

f) 

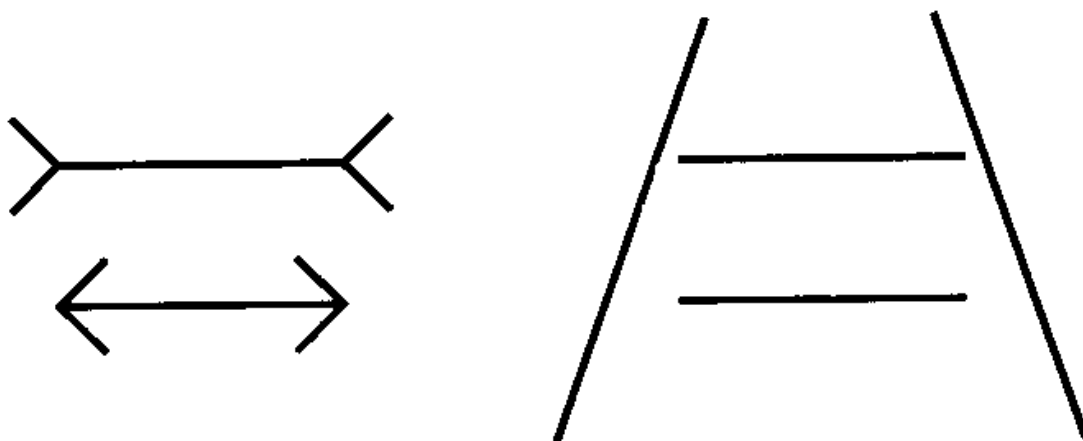
(aus Pinz 1994)

f): kontinuierl. Fortsetzung gruppiert stärker als Symmetrie

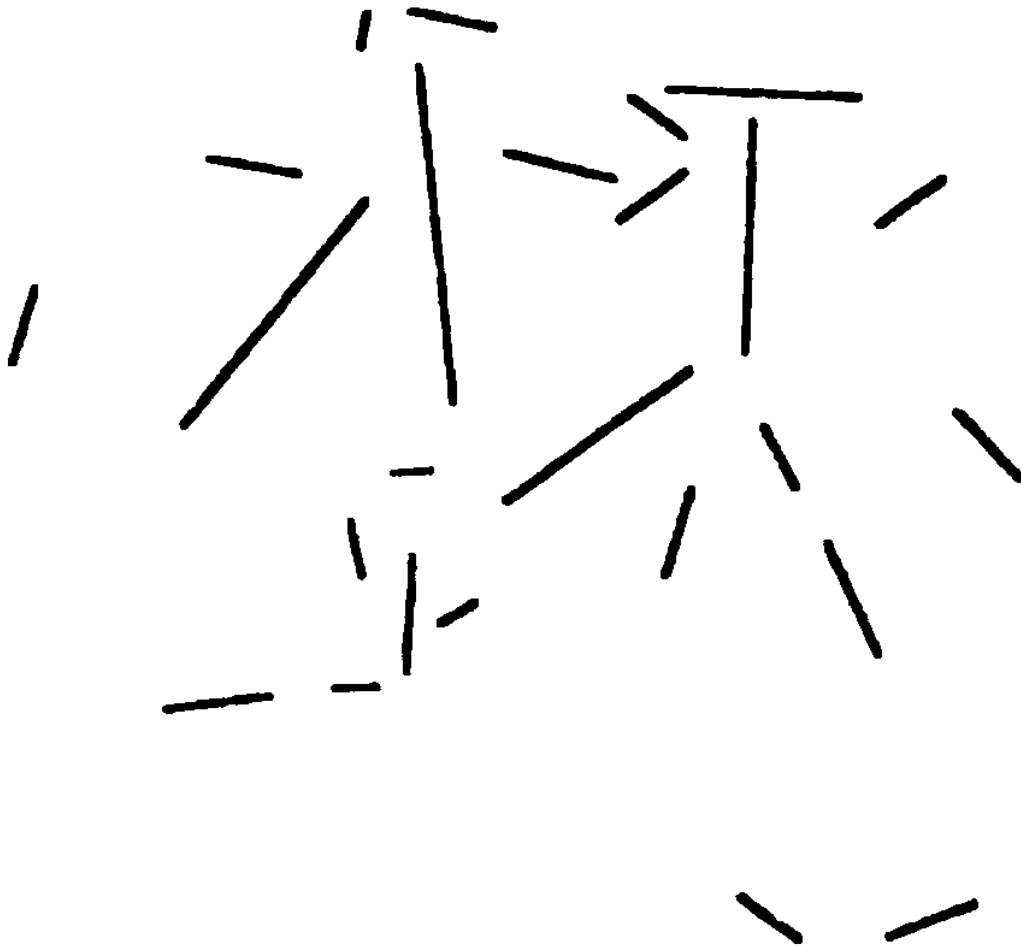
Es gibt auch Merkmale, die vom menschlichen visuellen System *schlecht* extrahiert werden können, z.B. Kollinearität:



Gestaltwahrnehmung berücksichtigt automatisch Tiefeneffekte, was zu "optischen Täuschungen" führen kann (eigentl. Täuschungen des visuellen Systems, entstehen bei der neuronalen Verarbeitung):



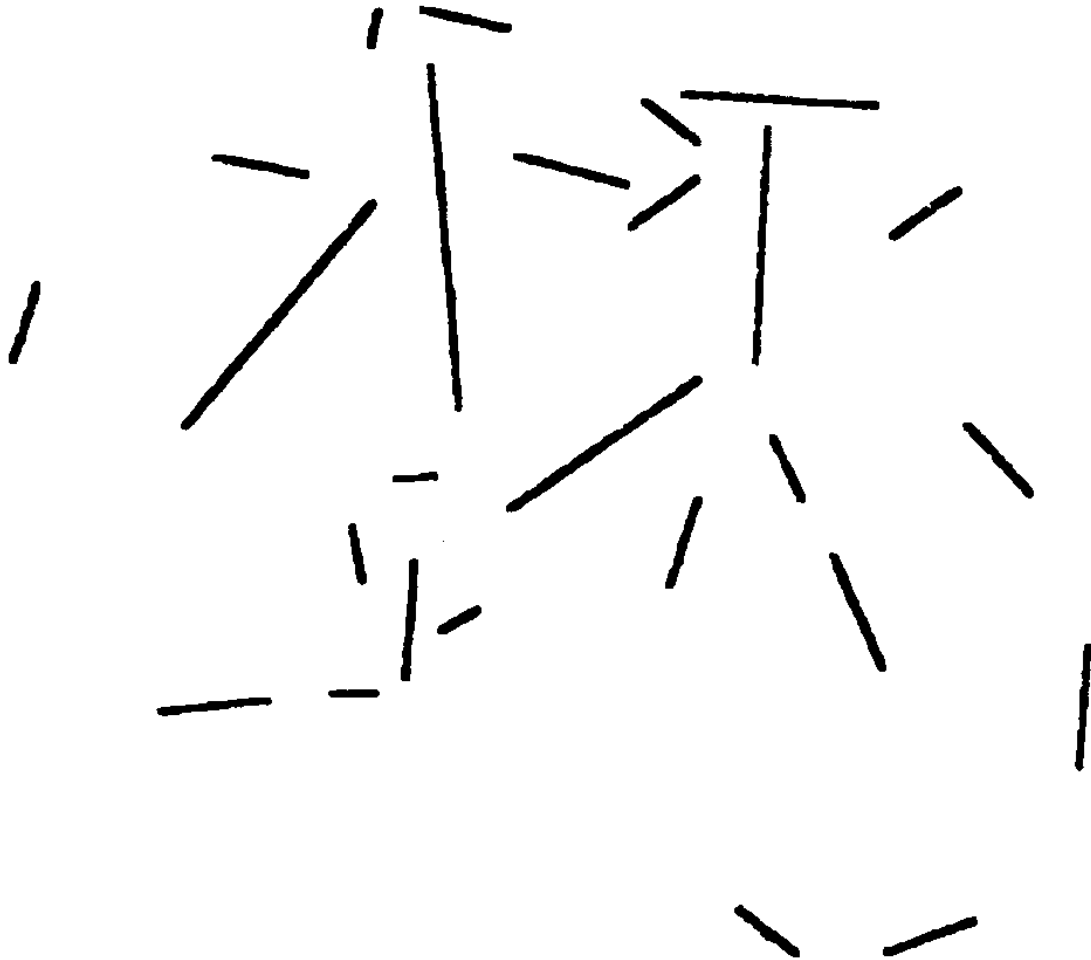
wie rekonstruiert das menschl. visuelle System
Bildinformationen? Hinweise aus Versuchsreihen:



was ist das?

Die meisten Testpersonen benötigen zum Erkennen
eines Objekts mehr als 1 Minute.

Hier sind die Erkennungszeiten deutlich niedriger,
obgleich nur 1 Strich (rechts unten) hinzugefügt wurde:



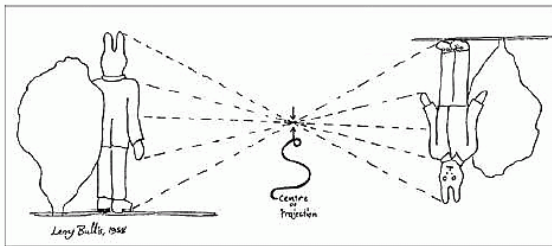
Vermutung:

Das zusätzliche Liniensegment ermöglicht eine höhere
Wahrscheinlichkeit für eine kreisförmige Gruppierung;
von da aus wird schrittweise weiter rekonstruiert
(Kreis → Rad → Fahrrad).

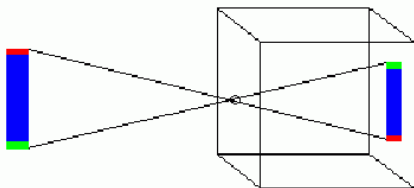
Grundlage der maschinellen Bildauswertung: *Bildgebende Verfahren*

schon hier Unterschied zum menschlichen Sehen –
Kamera sieht anders als das Auge

Wichtige bildgebende Verfahren (nur grober Überblick):



<http://www.bingo-ev.de/~db106/faq.htm#BM1>



<http://users.hol.gr/~dsathen/d/projekte/optik/fotoapp/lochk.htm>

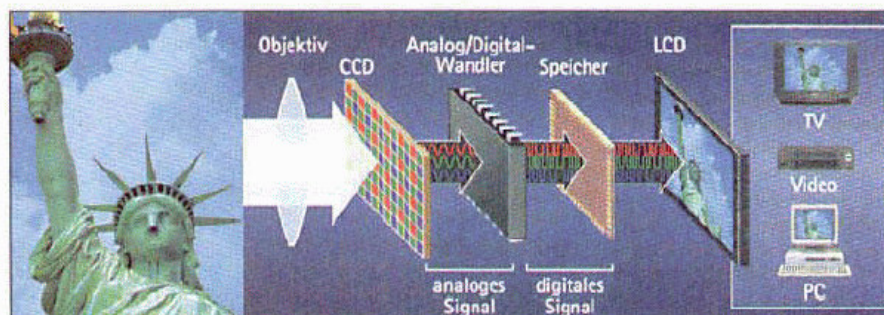
Fotografie

- Hohe Auflösung
- (beinahe) beliebige Brennweiten
- kostengünstig
- Aufwendige Entwicklung
- => Digitalkameras

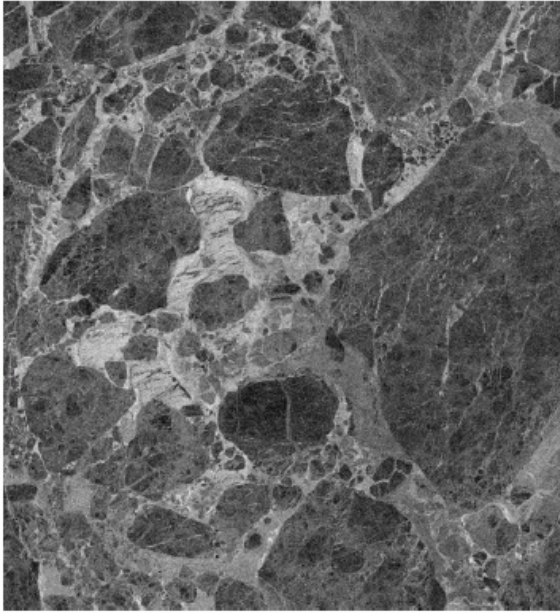
Digitalkamera

<http://www.burg-halle.de/~albrecht/burg/Vorlesung/DigitaleBilder/sld019.htm>

Funktionsweise von Digitalkameras



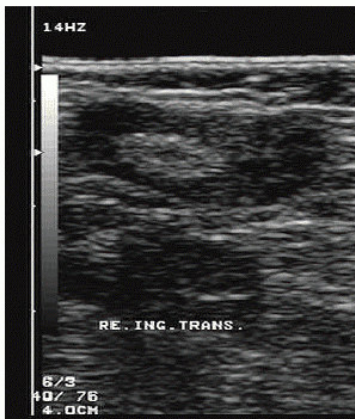
Das Bildmotiv wird vom Objektiv auf den CCD-Chip geleitet. Das in den Fotozellen des Chips festgehaltene analoge Bild wird im A/D-Wandler in ein digitales Signal umgewandelt. Das digitalisierte Bild wird auf speziellen Speicherkarten festgehalten.
(Aus: »D-Cam-Fibel«, SANYO FISHER)



Satellitenbilder

- Landschaftsfernerkundung
- Luftarchäologie
- Landmarkenerkennung zur Navigation
- Eisdriftdetektion
- Wellenlängen
 - UV: 0,1...0,4 mm
 - VIS (visible): 0,4...0,7 mm
 - IR: 1...30 mm
 - MW: größer 2 mm

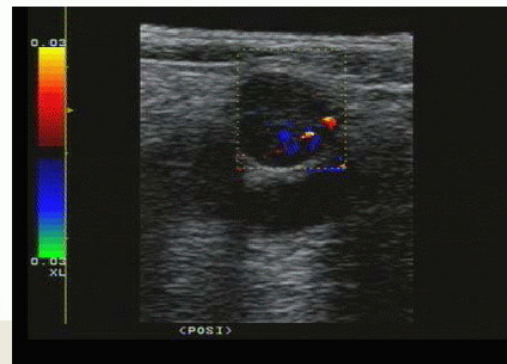
Ultraschall



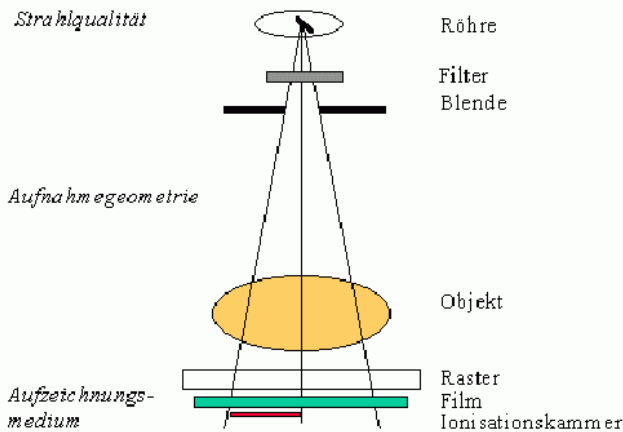
- Keine Strahlenbelastung
- Schmerzlos / nicht invasiv
- hohe Ortsauflösung (mod. Geräte)

<http://isgnw.cs.uni-magdeburg.de/~regina/Ultraschall.html>

- Klinische Diagnostik
- Aussagen über
 - Gewebestrukturen und -veränderungen (Vorhandensein von Flüssigkeiten)
 - Organabmessungen (Lage und Veränderungen)
 - Bewegungs- und Funktionsabläufe und Blutströmungen



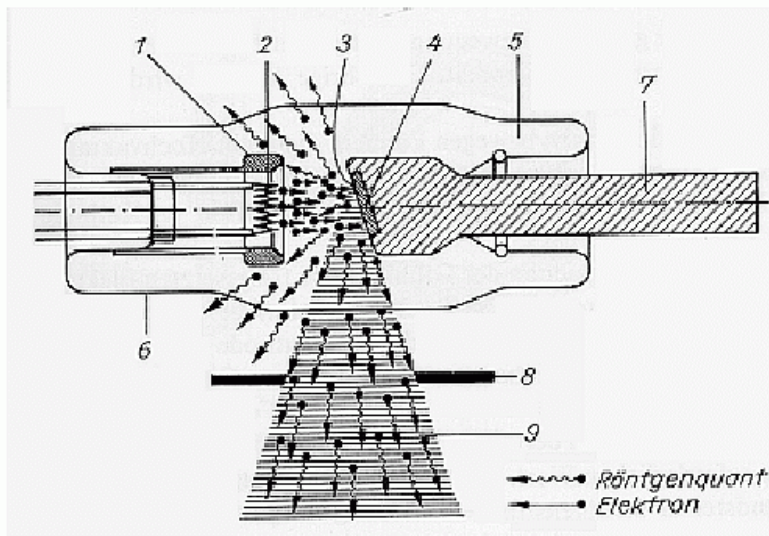
Röntgen



http://isgnw.cs.uni-magdeburg.de/~regina/cv1_vl2.htm

- medizinische Diagnostik
- gute Visualisierung von „harten“ Strukturen
- Aufprall energie-reicher Elektronen auf ein Material
- Wechselwirkung der Atome: Umwandlung der kinetischen Energie in Röntgenstrahlung

Röntgen

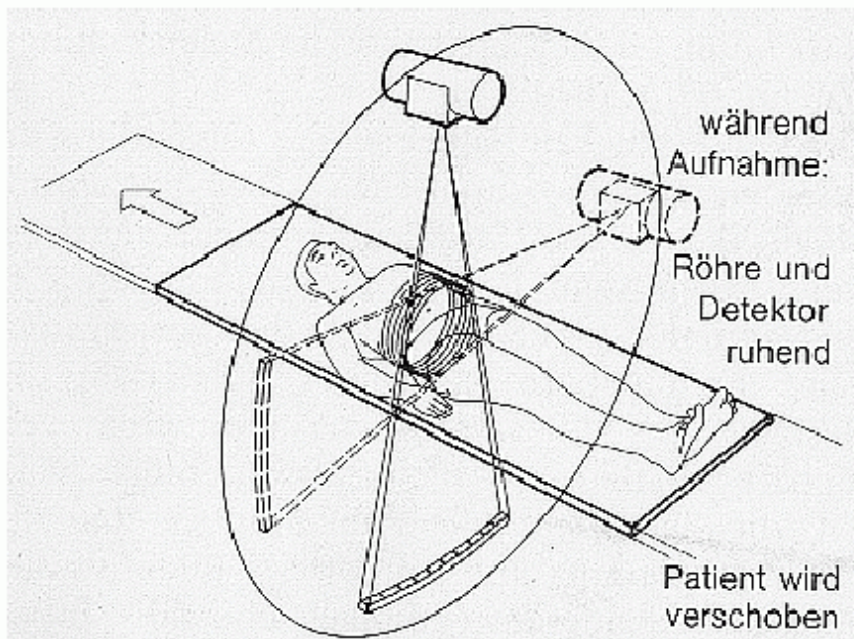


http://isgnw.cs.uni-magdeburg.de/~regina/cv1_vl2.htm

1. Kathode
2. Elektronenquelle (Glühfaden)
3. thermischer Brennfleck
4. Wolframscheibe
5. Hochvakuum
6. Glaskolben
7. Anode
8. Primärstrahlblende
9. Nutzstrahlenkegel

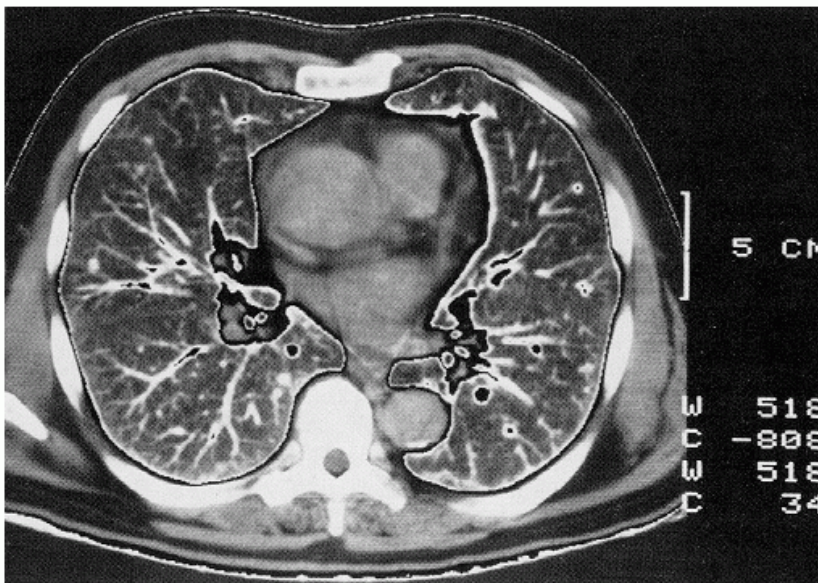
Computertomografie

CT



- Zunächst Übersichtsaufnahme dann Meßwertaufnahme

http://isqgw.cs.uni-magdeburg.de/~regina/skript_ct.html



- Strahlenbelastung nimmt mit größer werdender Schichtdicke ab

http://isqgw.cs.uni-magdeburg.de/~regina/skript_ct.html

ähnliches Aufnahmeverfahren, aber andere Strahlung: Magnetresonanztomografie (MRT)

3D-Scanner:

Laser-Kamera



EBK der ASTRIMUM GmbH

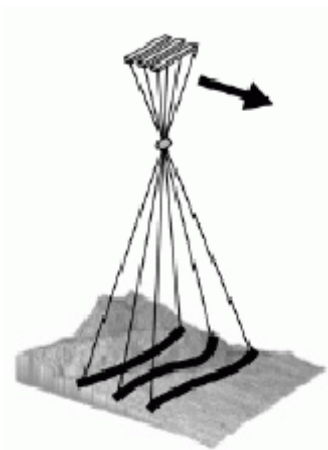
- Distanzbereich: 1 - 30 m
- Meßbereich: 1,5 - 15 m
- Bildwinkel: hor. $\pm 15^\circ$
ver. $\pm 22^\circ$
- Bildpunkte 640 * 480
- Entfernungsauflösung:
2% des Meßbereiches

- verschiedene Techniken der Tiefengewinnung (siehe später)

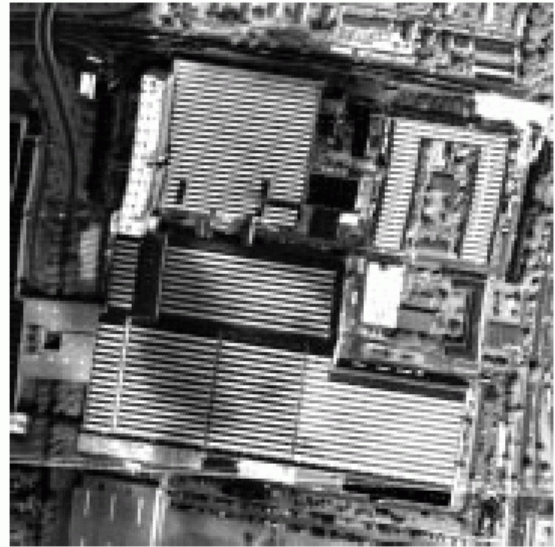
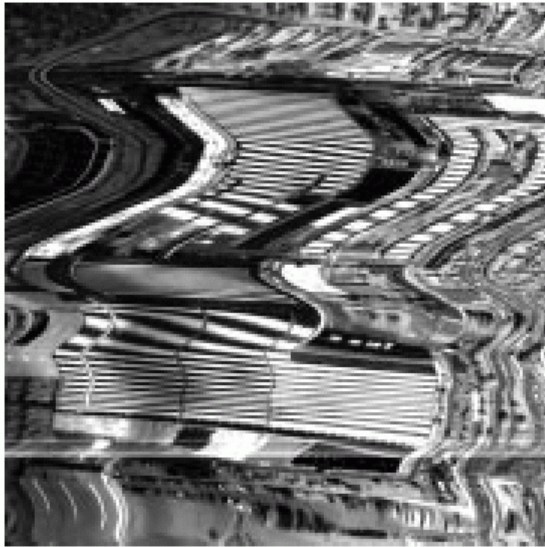
Zeilenkamera

(Zeile mit CCD-Elementen – lichtempfindliche Sensoren)

für Fernerkundung oft 3 parallele Zeilen:



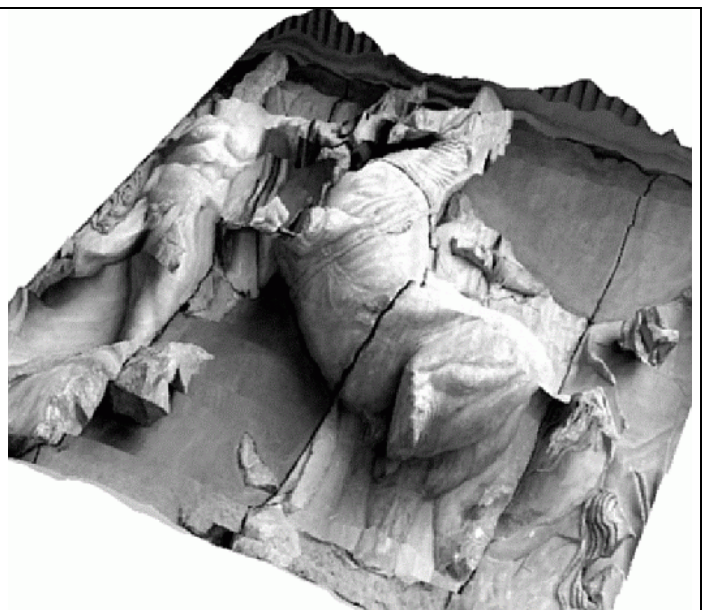
Luftbild unkorrigiert / korrigiert:



auch für terrestrische Aufnahmen / Gebäude- und Innenraum-Rekonstruktion, Archäologie:

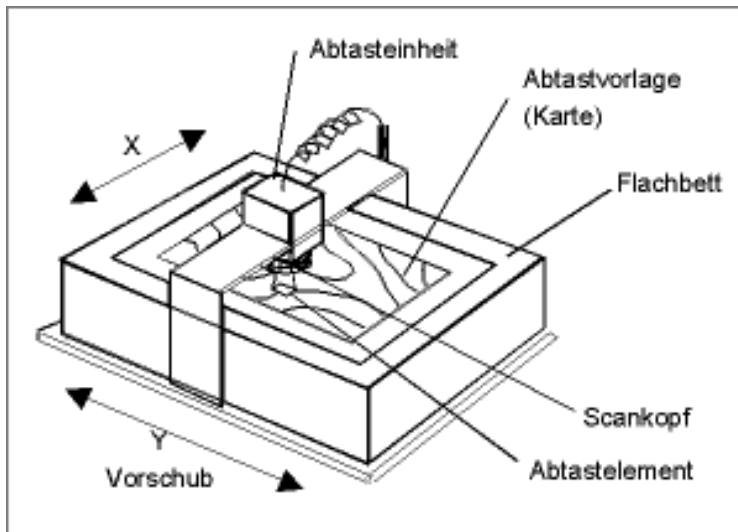


Panoramakamera des
DLR

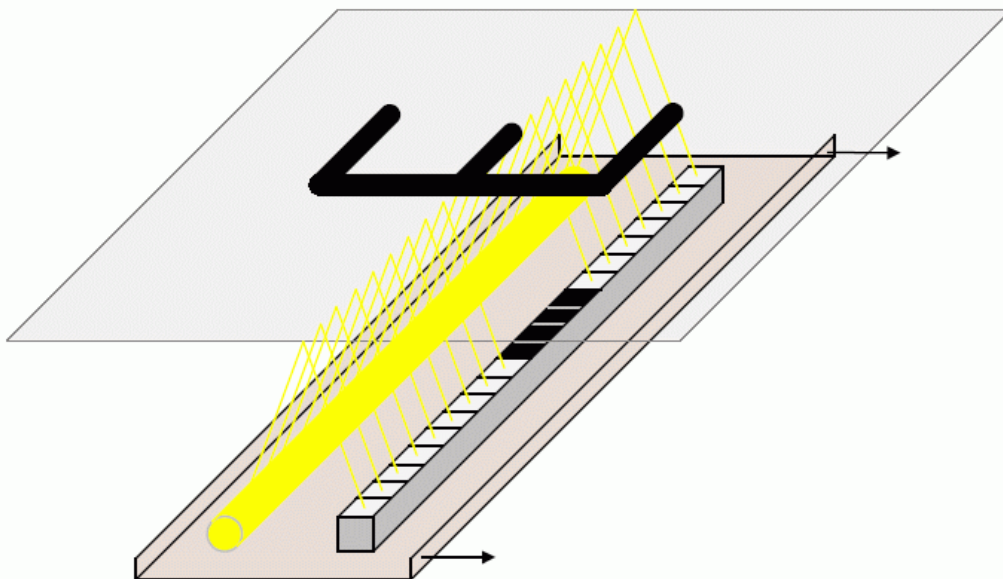


3D-Rekonstruktion von Teilen
des Pergamon-Altars

Flachbettscanner



Aufbau eines Scanners



In einem 600dpi-Farbscanner sitzt auf einem beweglichen Schlitten eine kleine Leuchtstofflampe und in einer Reihe typischerweise 3×4800 lichtempfindliche Zellen („3×“ wegen der drei Grundfarben Rot, Grün, Blau). Ein Schrittmotor schiebt nun diesen Schlitten in etwa 7000 Minischritten (DIN A4-Blatt) unter der Vorlage hinweg, dabei wird bei jedem (der mit dem Auge nicht wahrnehmbaren Stopps) ein „Foto“ gemacht: Die weißen Flächen des Papiers reflektieren das von der Leuchtstofflampe kommende Licht auf die lichtempfindlichen Zellen, die schwarzen Buchstaben auf dem Papier absorbieren das Licht – die darunter liegenden Zellen bekommen demzufolge kein Licht. Entsprechend verhält es sich bei grauen bzw. den farbigen Flächen des Papiers.

Bei einem DIN A4-Blatt fallen dabei rund $3 \times 4800 \times 7000 \approx 100$ Millionen Helligkeitsinformationen (entsprechend ca. 100 MByte!) an. Jede Helligkeitsinformation besteht aus der eingefallenen Lichtmenge in 256 Abstufungen. Aus diesem Raster setzt sich der Scan zusammen und erscheint aus genügend großem Abstand als nicht sichtbar.

(aus http://www.mathematikphysik.de/computer/compdoc1/docs/lac08_scannen.pdf)