

Kognitive Robotik

Prof. Dr. sc. Hans-Dieter Burkhard
Vorlesung Winter-Semester 2005/06

Verarbeitung visueller Daten

Mit Dank

■ an Teilnehmer Proseminar
WS02/03, speziell

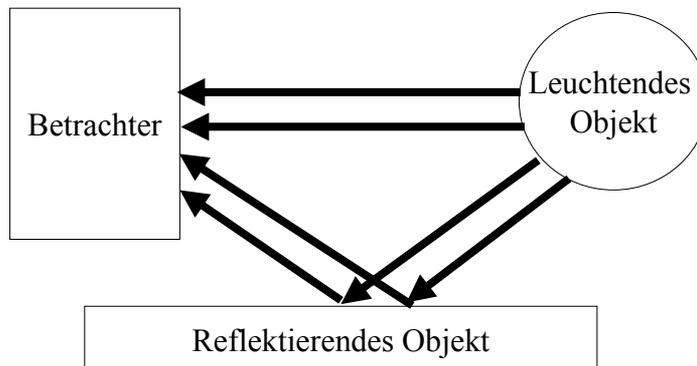
Michael Hempel
Henryk Plötz
Hagen Strauss
Alberto Bettin
Maximilian Buder
Volker Janetschek
Sebastian Lindstaedt
Tobias Meyer zu Hoberge
Franz Andert

an Mitarbeiter aus HU:

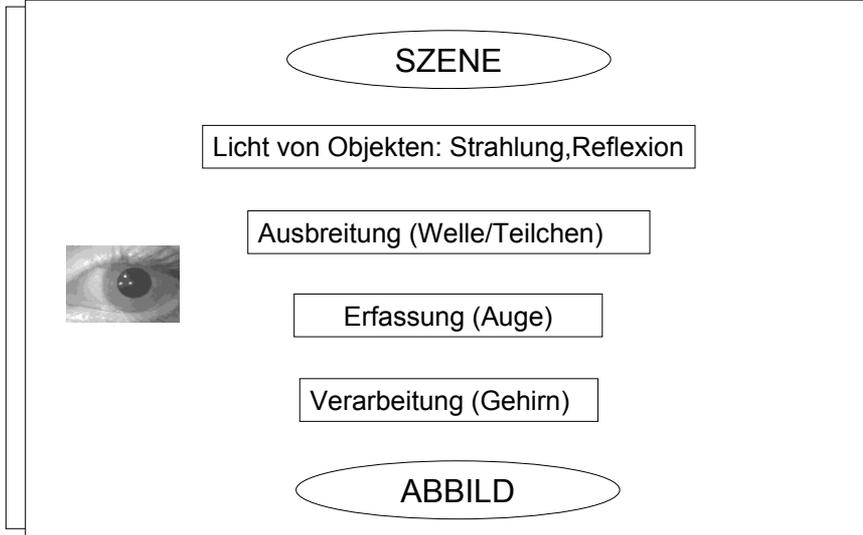
Jan Hoffmann
Joscha Bach
Matthias Jünger
und weitere Autoren

Licht und Farbe

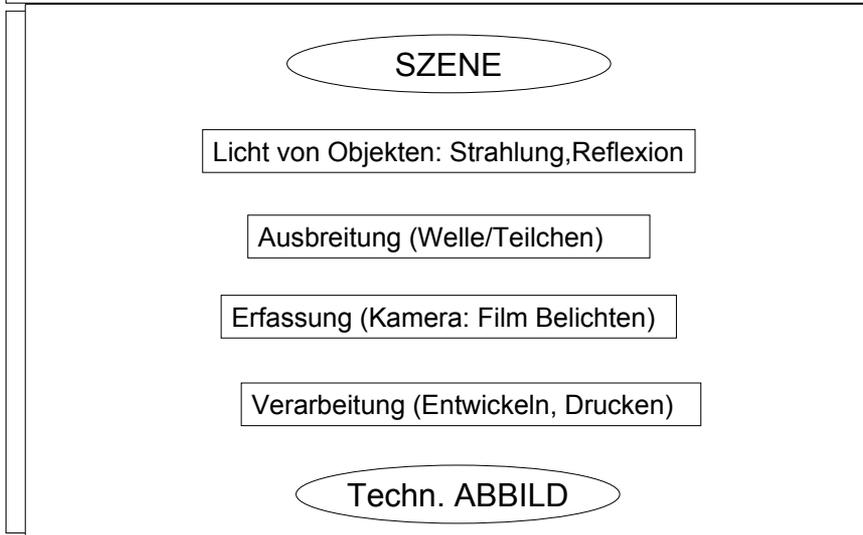
3 Beteiligte:



Betrachter: Mensch



Betrachter: Technisch/Chemie



Betrachter: Technisch/Elektronisch

SZENE

Licht von Objekten: Strahlung, Reflexion

Ausbreitung (Welle/Teilchen)

Erfassung (Kamera: Pixel)

Verarbeitung (Bildschirm, Drucken)

Techn. ABBILD

Licht, Abbildung

SZENE

Was passiert tatsächlich?

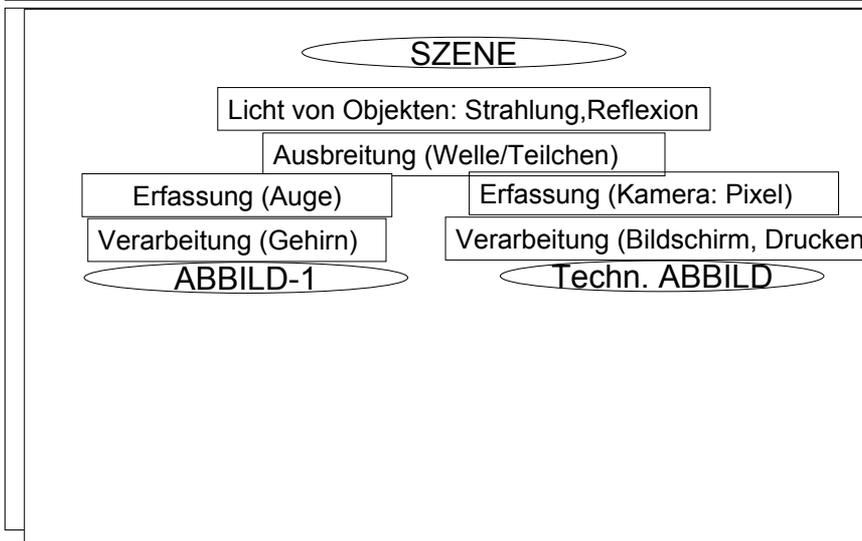
Stimmen SZENE und ABBILDER überein?

(Techn.) ABBILD

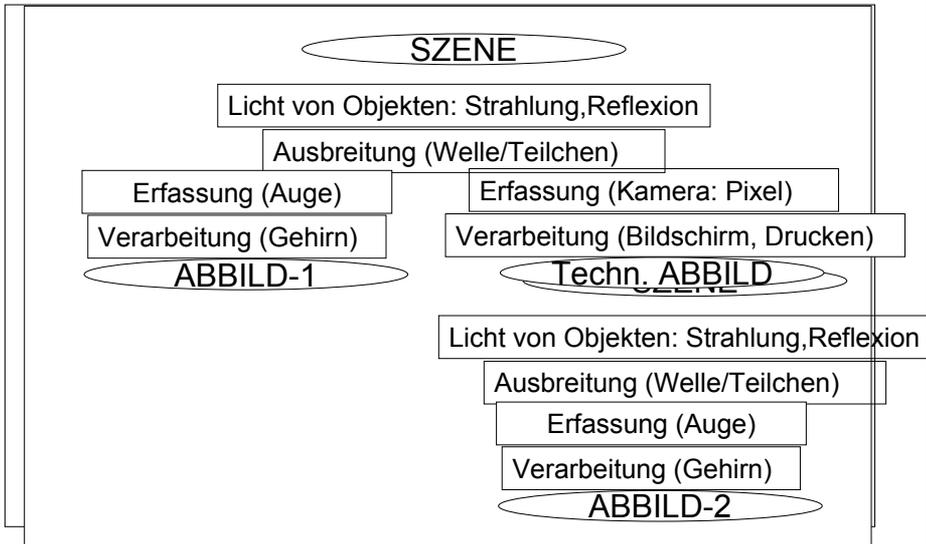
Licht, Abbildung



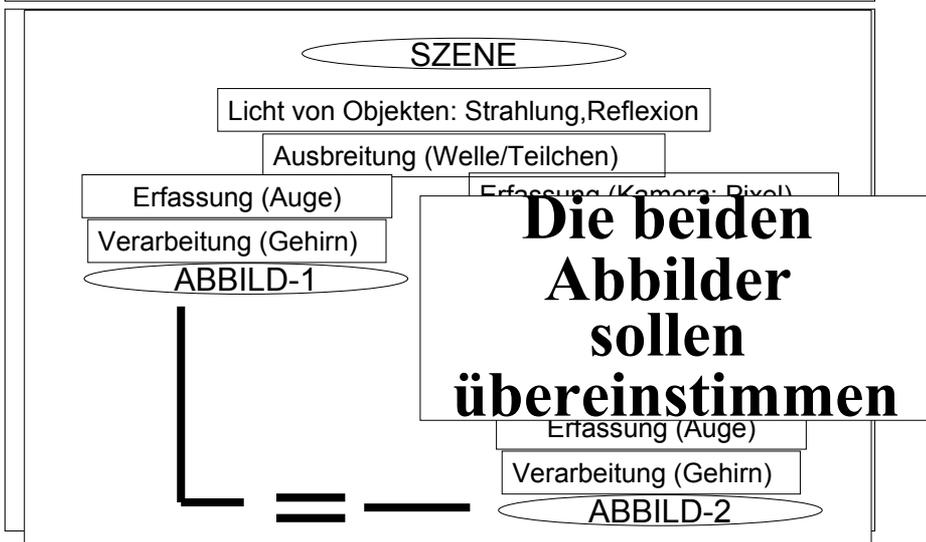
Licht, Abbildung



Licht, Abbildung



Licht, Abbildung: Anforderung



Licht, Abbildung: Anforderung

Ausgangspunkt für Übereinstimmung ist das menschliche Empfinden.

Technik soll Bild wiedergeben, das vom Menschen ähnlich wahrgenommen wird wie die Szene.

Darüber hinaus wird i.a. keine Übereinstimmung gefordert.

Ein Roboter könnte aber auch mit anderen Sehfähigkeiten ausgestattet werden.



Künstler

SZENE

Licht von Objekten: Strahlung, Reflexion

Ausbreitung (Welle/Teilchen)

Erfassung (Auge)

Verarbeitung (Gehirn)

Verarbeitung (Zeichnen, ...)

BILD

Künstler und Betrachter

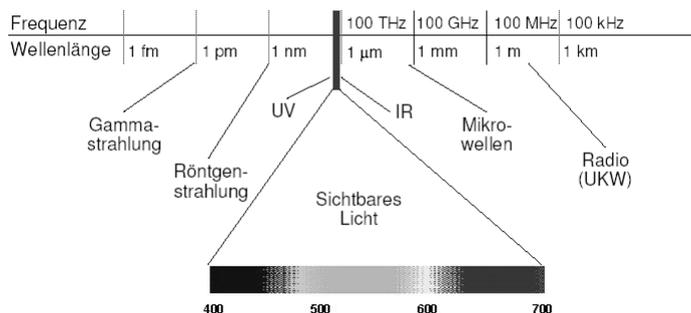


Licht

- Elektromagnetische Strahlung
- Für Menschen sichtbares Licht:

Wellenlängen zwischen 380 nm und 780 nm

Spektrum der elektromagnetischen Strahlung



Licht

Dualer Charakter (2 Modelle):

- Welle:
Brechung, Beugung, Dispersion, Streuung
- Teilchen (Photonen):
Absorption, Emission

Wellenlänge λ , Frequenz f , Periodendauer T

$$f = 1 / T \text{ [Hz]}$$

$$T = \lambda / c \text{ [s]}$$

$$f = c / \lambda \text{ [Hz]}$$

$$c = 2,998 \cdot 10^8 \text{ m/s (Lichtgeschwindigkeit)}$$

Licht

Geradlinige Ausbreitung

Absolute Reflexion: Einfallswinkel = Ausfallswinkel

Oberflächen reflektieren unterschiedlich:

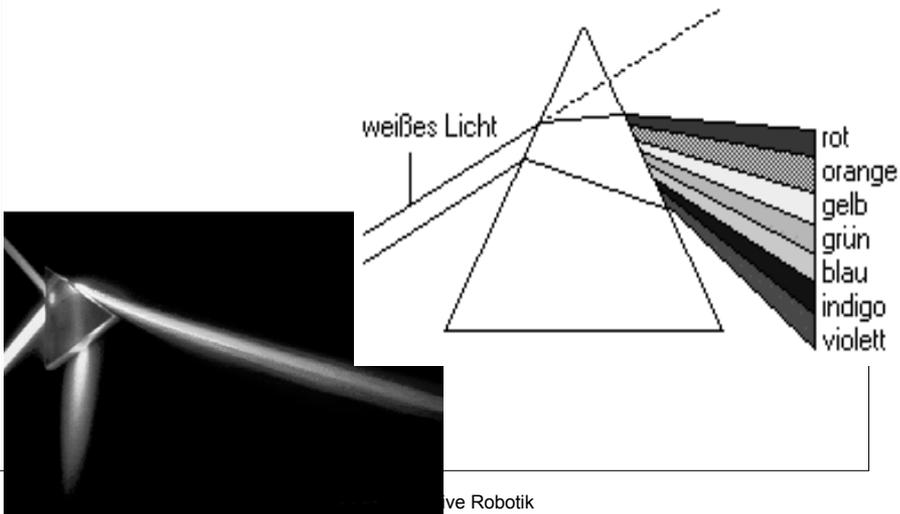
- Reflexionskoeffizient (abhängig von der Wellenlänge)
- Rauigkeit

Optische Dichte von Materie:

- Brechungsindex beschreibt niedrigere Ausbreitungsgeschwindigkeit gegenüber Lichtgeschwindigkeit
- Bei Eintritt in optisch dichtere Materie erfolgt Beugung des Lichtstrahls zur Senkrechten
- Dispersion: Abhängigkeit des Brechungsindex von der Wellenlänge (Regenbogeneffekt)

Licht

Zerlegung von weißem Licht in Spektralfarben



Winter-Semester 2005/06

Verarbeitung visueller Daten

17

Licht

- Reale Strahlungsquellen strahlen fast immer eine Mischung verschiedener Frequenzen aus.
- Das von Körpern reflektierte Licht hängt ab vom eingestrahlt Licht und von den Absorptionseigenschaften (nicht alle Frequenzen werden reflektiert).

Formal:

Intensität als Funktion der Wellenlänge: $I(\lambda)$

Unendlich-dimensionaler Raum \mathcal{L} von Möglichkeiten.

Basis: Licht einer Wellenlänge

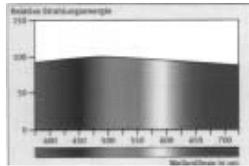
H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

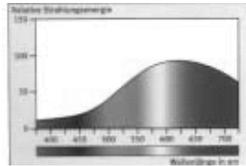
18

Licht

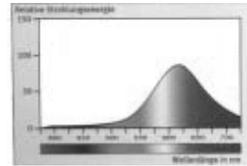
Intensität als Funktion der Wellenlänge: $I(\lambda)$



Tageslicht



Glühlampe



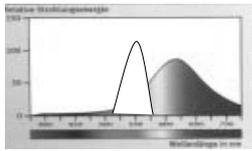
Rotes Licht

Sensor, Rezeptor

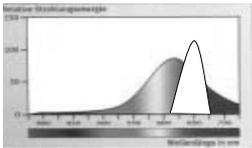
Unterschiedliche Farbempfindlichkeit

- Technisch:
CCD (Ladung proportional zur Lichtstärke) + Filter (nur durchlässig für eingeschränkten Frequenzbereich)
- Auge:
Zapfen, Stäbchen (durch Licht ausgelöste chemische Prozesse an Nerven übermitteln)
- Reduktion des unendlich dimensional Lichtraums auf eine Dimension (Erregungsskala)

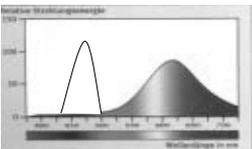
Sensor, Rezeptor



Grünempfindlicher Sensor



Rotempfindlicher Sensor



Blauempfindlicher Sensor

Sensor, Rezeptor

Messung:

Reaktion auf Licht in bestimmten Frequenz-Bereichen

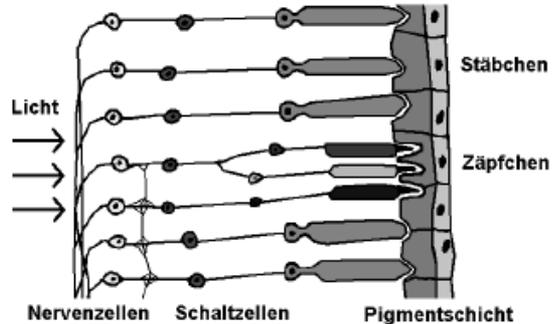
- Empfindlichkeit abhängig von Frequenz: $\phi(\lambda)$
- Erregung e insgesamt abhängig von eingestrahlem Licht $I(\lambda)$ und Empfindlichkeit $\phi(\lambda)$.

Bei Annahme linearer Beiträge und ohne Berücksichtigung von Schwelle/Sättigung realer Sensoren näherungsweise beschreibbar durch

$$e = \int \phi(\lambda) I(\lambda) d\lambda$$

Menschl. Auge: Aufbau der Netzhaut

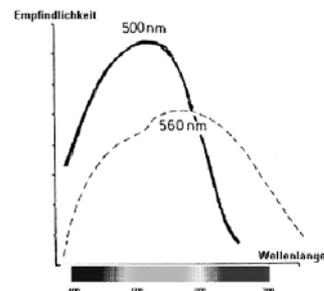
- Stäbchen (*Rods*)
Schwarzweiß-Sehen (Sehr empfindlich!)
- Zapfen (*Cones*)
Farben-Sehen
- Nerven
- Blutgefäße



Helligkeitsempfinden

Durchgezogene Linie: Stäbchen

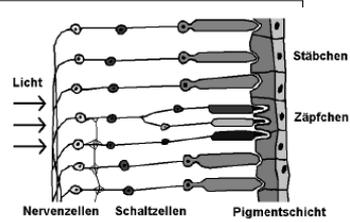
- Stäbchen sind "farbenblind",
aber lichtempfindlicher als Zapfen
- Dämmerungssehen mit Stäbchen:
Blau = hellgrau,
rot = schwarz
„Nachts sind alle Katzen grau.“



Gestrichelte Linie: Zapfen

- Zapfen für Farbsehen
- Tagsehen mit Zapfen:
Größte Empfindlichkeit für Gelb/orange-Töne
Scharfes Sehen mit Zapfen (Größte Häufigkeit in Fovea)

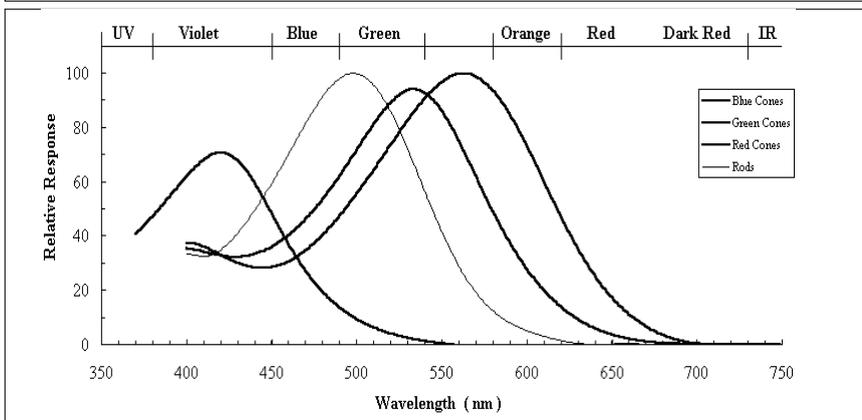
Farbsehen mit Zapfen



- Drei Sorten
 - Rot
 - Grün (höchste Empfindlichkeit)
 - Blau
- Drei-Dimensionaler Sensor-Raum für Erfassung des unendlich-dimensionalen Raums \mathcal{L} des Licht:
Unterschiedliches Licht (Spektralverteilungen) kann gleiche Reaktion der Zapfen hervorrufen.

Unterschiedliches Licht kann gleiche
Farbempfindung verursachen: metameres Licht

Farb-Empfindlichkeit



Retina enthält 3 Typen farbsensitiver Zapfen mit maximaler Sensitivität bei:

- 430 nm (ca. 4% der Zapfen, blausensitiv)
- 530 nm (ca. 32% der Zapfen, grün-sensitiv)
- 560 nm (ca. 64% der Zapfen, rot-sensitiv)

Farbwahrnehmung im Auge

Farbwahrnehmung wird bestimmt durch die Rezeptoren für rot (gelb), grün, blau.

Deren Reizung wird in Gehirn weiterverarbeitet und führt zu unterschiedlichen Wahrnehmungen.

Bei der Weiterverarbeitung werden die ursprünglichen Empfindungen miteinander verrechnet (bessere Ausnutzung der Informationskanäle).

Außerdem gibt es wieder „Ergänzungen“. Diese können zu Farbillusionen führen.

Die **Farbwahrnehmung** ist ein physiologisches Phänomen.
(Farben existieren nur im Gehirn, nicht in der Natur.)

Farbwahrnehmung im Auge



Bild: Joseph Albers

Die **Farbwahrnehmung** ist ein physiologisches Phänomen.
(Farben existieren nur im Gehirn, nicht in der Natur.)

Farbwahrnehmung im Auge

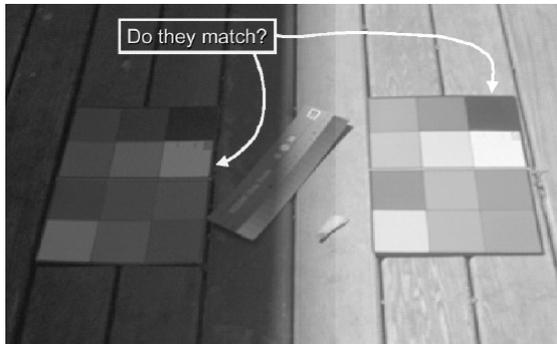
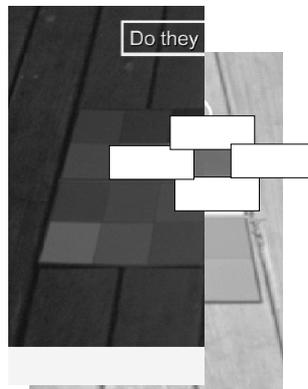


Bild: John M. Cann

Farbwahrnehmung im Auge



Cann

Menschliches Farbempfinden

Gegründet auf sensorischen Fähigkeiten des Auges:

- Kombination der Reizstärke der drei Zapfen-Sorten

Verarbeitung im Gehirn unter Einfluss von Erfahrung/Kultur
(unterschiedlicher Namensvorrat in verschiedenen Kulturen, Sprachen)

Verbindung von Farben mit

- anderen Empfindungen
- Stimmungen
- Situationen
- Politik ...

Wahrnehmbare Farben

Metamerie: Unterschiedliches Licht kann gleiche Wahrnehmung hervorrufen

- Hell-dunkel: eindimensionaler Empfindungsraum
- Farben: drei-Dimensionaler Empfindungsraum

Anzahl wahrnehmbarer Farben

Der menschliche Sehapparat kennt näherungsweise

- 128 verschiedene Farbtöne (hues)
- 130 verschiedene Farbsättigungen (Farbreinheit)
- 16 (im gelben Bereich) – 26 (im blauen Bereich) verschiedene Helligkeitswerte

Insgesamt ca. 380 000 verschiedene Farben

- Sichere Unterscheidung gleichzeitig dargestellter Farben in Experimenten nur bei ca. 15 Farben

Subjektive Farbmerkmale

- **Helligkeit** (Value, Brightness): physiologisch-psychologisches Maß für die Stärke der wahrgenommenen Intensität.
- **Farbton** (Hue): physiologisch-psychologischer Begriff zur Unterscheidung verschiedener charakteristischer Spektralmuster; unterscheidet zwischen reinen Farben (rot, gelb, grün, blau, usw.).
- **Sättigung** (Saturation): physiologisch-psychologisches Maß für den Grad, in dem der Farbton eines gegebenen Lichtes von dem Farbton eines weißen Lichtes gleicher Luminanz abweicht. Gesättigte Farbe enthält maximal 2 Grundfarben: Hinzufügen der dritten Farbe bewirkt (zusammen mit den anderen beiden Farben) einen Unbunt-Anteil: schwarz bzw. grau bzw. weiß.

Beispiel: Gesättigtes Rot in RGB: (255,0,0) -- Sättigung = 255

„Verdünnung in Richtung Weiß“ durch Zumischen von Blau und Grün ergibt Rosa: (255,127,127) --Sättigung = 128

Objektive Farbbeurteilung

Physiologische Ausgangssituation

- Drei Rezeptortypen
- Licht unterschiedlicher Eigenschaften erzeugt Farbwahrnehmung
- Die (chemischen) Prozesse der Reizaufnahme und die (kognitiven) Prozesse der Verarbeitung sind subjektiv

- Hilfsmittel: Farbtabelle
- Nachmischung von Farben zielt auf gleiche Farbempfindung

Objektive Farbbeurteilung

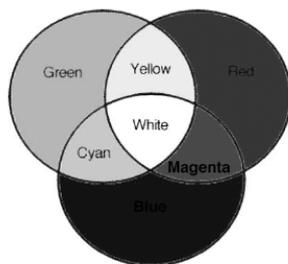
Physikalische Ausgangssituation

- Farbmischung:
 - Additive Mischung (Strahlungsquellen)
 - Subtraktive Mischung (Absorption, Filter)
- Farbmodelle auf der Basis von drei Rezeptoren
- Farbräume:
 - Technische Basis für Farbverarbeitung
 - Druck, Foto, TV, ...

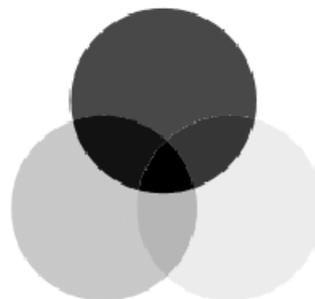
Festlegung von Standards:

- CIE (Commission Internationale de l'Éclairage)
 - Internationale Organisation mit Sitz in Wien
 - <http://www.cie.co.at>

Farbmischung



Additive Mischung
(Strahlungsquellen)



Subtraktive Mischung
(Absorption, Filter)

Farbmodelle

Eindeutige Beschreibung von Farben

- Allgemeine Farbmodelle: CIE-Farbraum, CIE-L*a*b
- Hardwarebezogene Farbmodelle: RGB, CMY, CMYK, YUV, YIQ
- Physiologisch orientierte Farbmodelle: HLS, HSV, HSI
- Ein Farbmodell muss nicht alle wahrnehmbaren Farben enthalten.
- Die Kombination von Farben soll dem Mischen von Farben entsprechen.

Modellierung

Wegen der Eigenschaften des Auges reichen 3 Parameter:

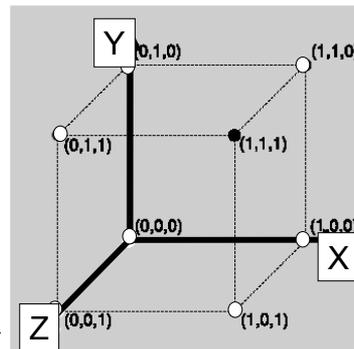
- Drei-dimensionaler Farbraum mit Koordinaten X,Y,Z.

Beim Mischen kommen nur Farben hinzu:

- Kombinationen mit positiven Koeffizienten.

Untergrenzen und Obergrenzen:

- Normierung
z.B. auf Intervall $[0,1]$



Modellierung aller Farben

Ausgangssituation entsprechend Farbwahrnehmung:

- Die Rezeptorempfindlichkeiten überschneiden sich: nicht der ganze Würfel ist besetzt.
- Alle Farben aus Spektralfarben kombinierbar mit positiven Komponenten: Farben liegen in der konvexen Hülle der Spektralfarben.

Beschreibung der Reizaufnahme (näherungsweise) für drei Sensoren X,Y,Z mit gegebenen Empfindlichkeiten:

$$e_x = \int \phi_x(\lambda) I(\lambda) d\lambda$$

$$e_y = \int \phi_y(\lambda) I(\lambda) d\lambda$$

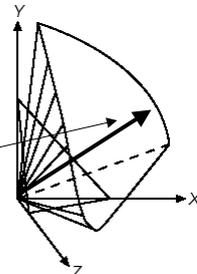
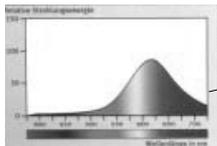
$$e_z = \int \phi_z(\lambda) I(\lambda) d\lambda$$

Modellierung aller Farben

$$e_x = \int \phi_x(\lambda) I(\lambda) d\lambda$$

$$e_y = \int \phi_y(\lambda) I(\lambda) d\lambda$$

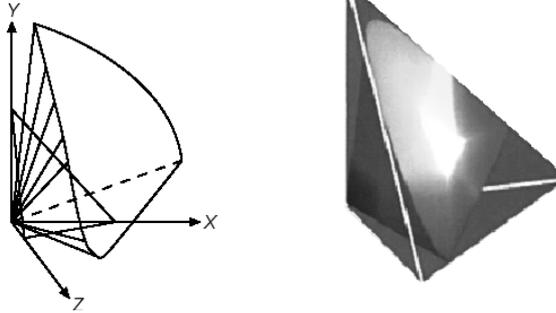
$$e_z = \int \phi_z(\lambda) I(\lambda) d\lambda$$



Bei gegebenen Grundfarben Φ_x, Φ_y, Φ_z kann für jede Intensitätsverteilung $I(\lambda)$ der darstellende Vektor (e_x, e_y, e_z) im Farbraum berechnet werden.

Modellierung aller Farben

Die Spektralfarben liegen am rechten und linken Rand.
Die anderen Farben innerhalb der konvexen Hülle.
Untere Begrenzung: „Purpurlinie“



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

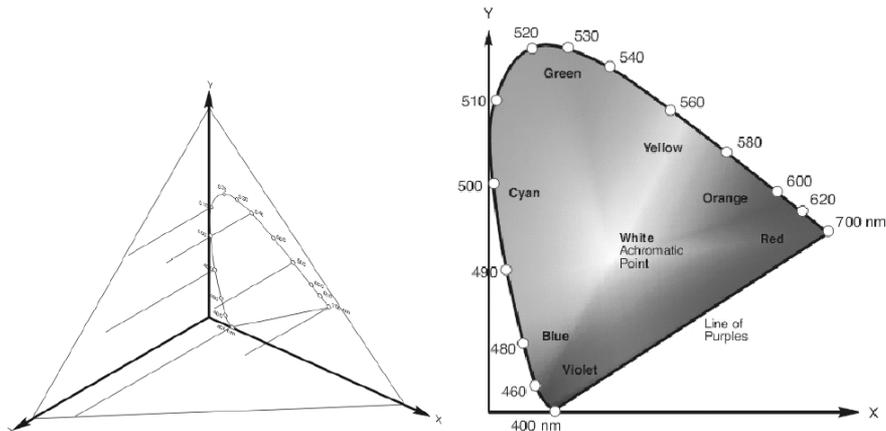
Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

41

Farbtafel (*chromaticity diagram*)

Verzicht auf Darstellung der Helligkeit,
d.h. Beschränkung auf die Ebene $X+Y+Z = 1$:

- $x := X/(X + Y + Z)$
- $y := Y/(X + Y + Z)$

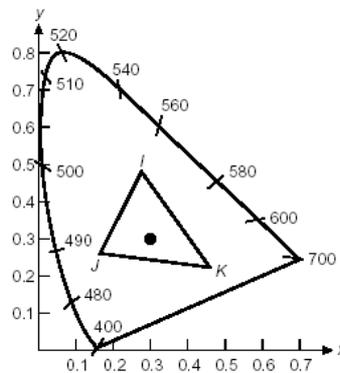


Farbtafel: Additive Farbmischung

Linearer Raum:

- Jede Farbe auf der Linie IJ kann aus I und J gemischt werden.
- Insbesondere können alle Farben aus zwei Spektralfarben gemischt werden.
- Alle Farben innerhalb des Dreiecks können aus I,J,K gemischt werden.
- Daraus folgt eine Einschränkung:

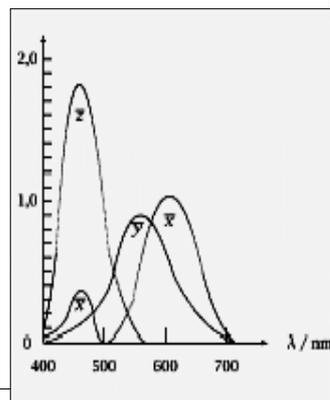
Bei additiver Farbmischung können nicht alle Farben aus 3 Farben erzeugt werden.



CIE-Standard (1931, 1976)

Festlegung von X,Y,Z („Standard-Beobachter“):

- Normspektralfunktionen x, y, z in Ahnlehnung an RGB-System.
- Koordinatentransformation zur besseren Normierung:
 - x und y orthogonal
 - y entspricht Helligkeitssehen (Zapfen)
 - in x -Richtung ist Helligkeit konstant
 - Empfindlichkeitskurven entsprechen nicht dem natürlichen Sehen (insbesondere Helligkeit nicht durch Summierung)
 - Weißpunkt bei $(1/3, 1/3)$



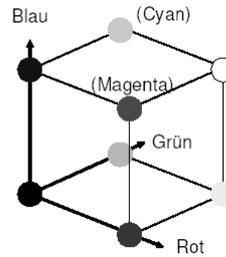
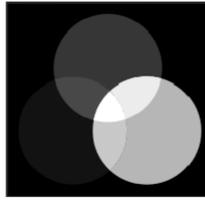
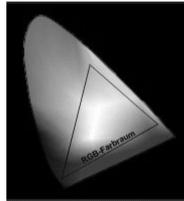
RGB-Modell

R=700 nm

G=546,1 nm

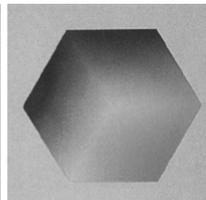
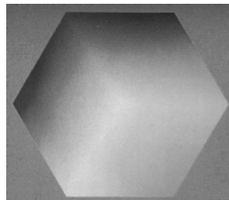
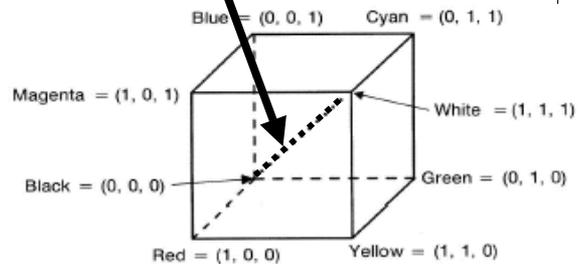
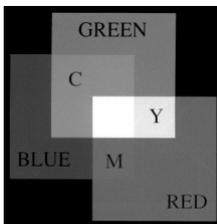
B=435,8 nm

- Additives Modell
- Meistverwendetes Modell für aktiv lichterzeugende Ausgabemedien (z.B. Displays)
- Spektrale Intensitäten der Komponenten werden addiert
- Bestimmte sichtbare Farben können nicht im RGB-Modell dargestellt werden.



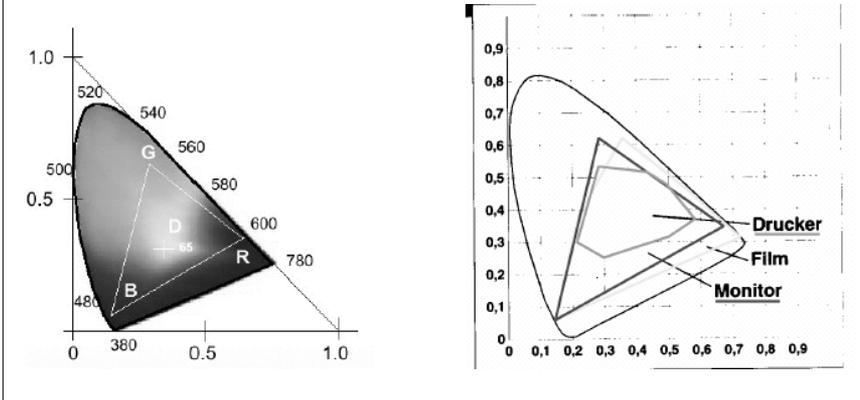
RGB-Modell

Unbunt-Linie: Schwarz-grau-weiß



RGB-Modell

- RGB-Darstellungen in der Praxis unterschiedlich
- ggf. Kalibrierung



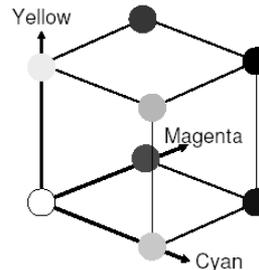
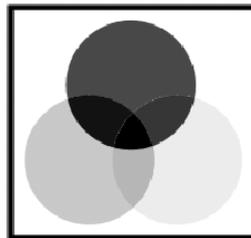
H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

47

CMY(K)

- Subtraktives Modell
- Meistverwendetes Modell zur Ausgabe auf reflektierenden Ausgabemedien (z.B. Farbdrucker)
- Anschaulich: Farbfilter subtrahieren Farbwerte
- Für Drucker oft vierte Komponente "schwarz" (black), deshalb CMYK (schwarz aus CMY nur unvollständig und mit viel Farbauftrag erzeugbar).



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

48

CMY(K)

Konvertierung RGB ↔ CMY

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

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

- gilt nur theoretisch
- kompliziertere Verfahren in der Druckindustrie

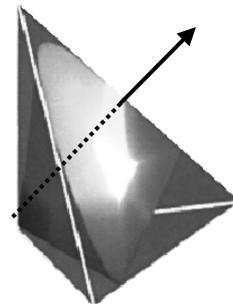
HSB-System

An subjektiver Wahrnehmung orientiertes Modell:

- H = Hue (Farbton)
- S = Saturation (Sättigung)
- B = Brightness (Helligkeit)

Im ursprünglichen
drei-dimensionalen XYZ-Modell:

- B = Länge des Farbvektors
- H = „Winkel“ ausgehend vom Weißpunkt
- S = Entfernung vom Weißpunkt



(Vom Prinzip ähnlich: HSV mit V=„Value“, HSI mit I = Intensität)

YUV

Abgeleitet aus RGB.

Bildinformation getrennt in:

- Helligkeitssignal

$$Y := 0,59 G + 0,30 R + 0,11 B$$

Wichtung entspricht Helligkeitsempfinden (grün am stärksten)

- Zwei Chrominanz-Signale:

$$U := 0,493 (B-Y) \quad (\text{Differenz Blau} - Y)$$

Farbanteil zwischen Gelb und Blau

$$V := 0,877 (R-Y) \quad (\text{Differenz Rot} - Y)$$

Farbanteil zwischen Cyan und Rot

Konvertierungen RGB \Leftrightarrow YUV

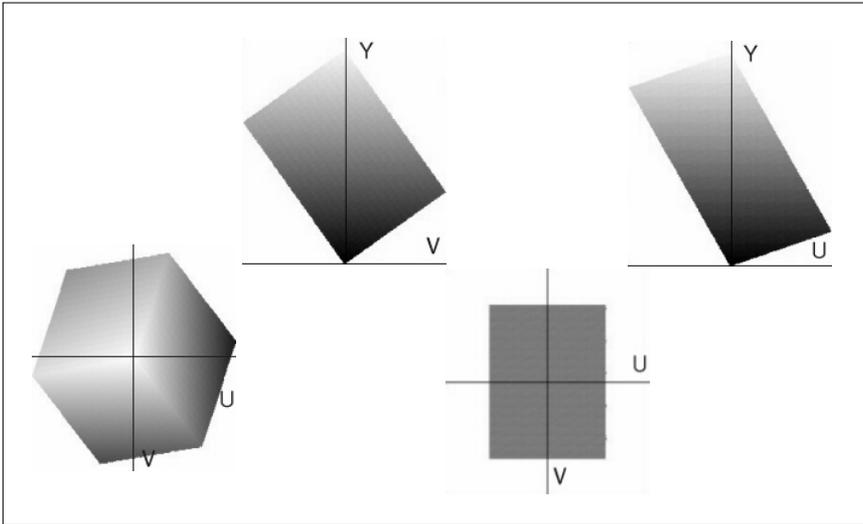
Transformations-Matrizen

(leicht unterschiedliche Werte auch in der Literatur,
dabei auch U und V manchmal vertauscht)

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.437 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.140 \\ 1 & -0.394 & -0.581 \\ 1 & 2.028 & 0 \end{bmatrix} \begin{bmatrix} Y \\ U \\ V \end{bmatrix}$$

YUV

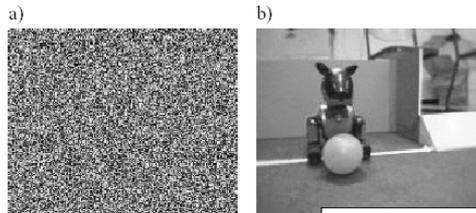


H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

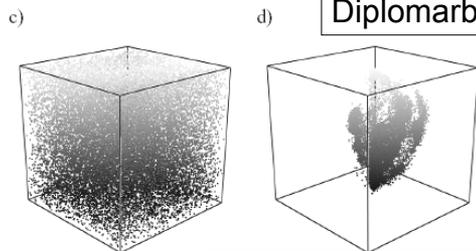
53

YUV



AIBO benutzt YUV

Diplomarbeit Jüngerl



Vertikal: Y-Achse (Helligkeit)

(c) YUV-Anteile des Zufallsbildes (a)

(d) YUV-Anteile des Bildes (b) - nur ein Teil des Raums wird benutzt

H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

54

YUV

Helligkeit $Y := 0,59 G + 0,30 R + 0,11 B$

Wertebereich 8 bit: [0...255]

Chrominanz-Signale:

$U := 0,493 (B - Y)$ (Differenz Blau – Y)

$V := 0,877 (R - Y)$ (Differenz Rot – Y)

Wertebereiche [-127...128], es reicht auch Kodierung mit 6 Bit.

Für RGB-Werte (x,x,x) ergibt sich YUV (x,0,0).

Alle Grauwerte von Schwarz bis Weiß liegen auf der Y-Achse.

Anwendung: TV- und Videotechnik

Kompatibilität zwischen Farb- und Schwarz-weiß-Geräten:

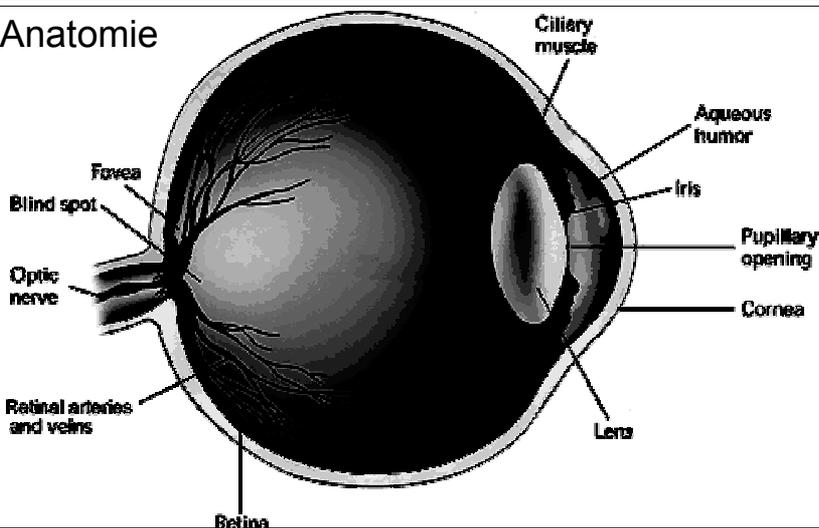
Luminanzsignal ermöglicht Abwärtskompatibilität zu SW.

YUV im PAL-Fernsehstandard

(analog: YIQ im amerikanischen NTSC-Fernsehstandard)

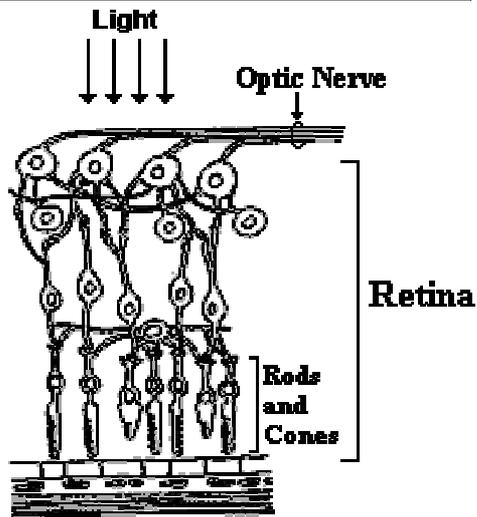
Menschliches Sehen: Das Auge

• Anatomie



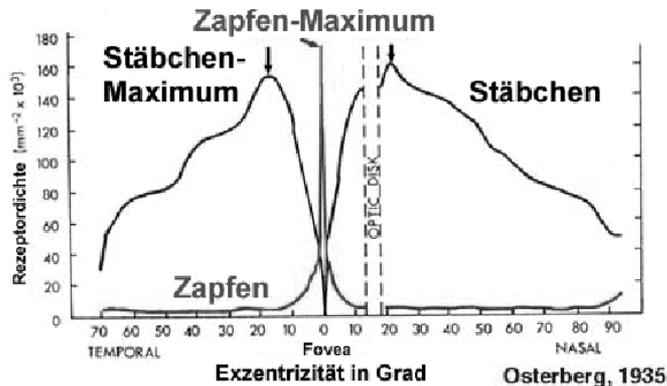
Aufbau der Netzhaut (Retina)

- Stäbchen (*Rods*)
Schwarzweiß-Sehen
Sehr empfindlich!
ca. 120 Millionen
- Zapfen (*Cones*)
Farben-Sehen
ca. 7 Millionen
- Nerven
- Blutgefäße



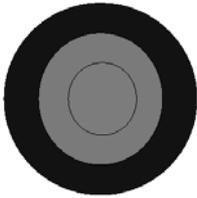
Verteilung Stäbchen/Zapfen

- Stäbchen (SW) in der Peripherie
- Zapfen (Farbe) im Zentrum (fovea centralis)
- Dämmerungssehen in der Peripherie besser als im Zentrum



Verteilung der Zapfen

- Blaurezeptoren:
Ca. 4%, nur peripher, nicht im Zentrum
- Grünrezeptoren:
Ca. 32 %, im zentralen Bereich konzentriert
- Rot-(bzw. Gelb-)rezeptoren:
Ca. 64%, mittlerer Abstand vom Zentrum

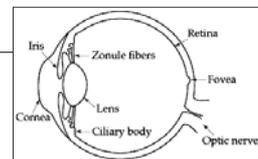


Detailreiche Information (Text) in reinem Blau ist schwer wahrzunehmen.
In der Peripherie sind rot und grün schwerer wahrzunehmen als blau.

Blinder Fleck: Austritt des Sehnervs

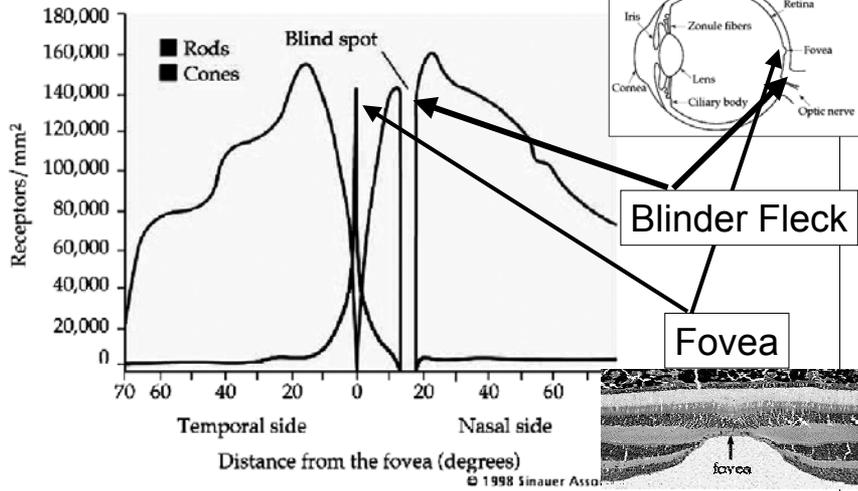


- Kreis mit dem rechten Auge fixieren
- Abstand zum Diagramm variieren
- Bei einem bestimmten Abstand verschwindet das Kreuz

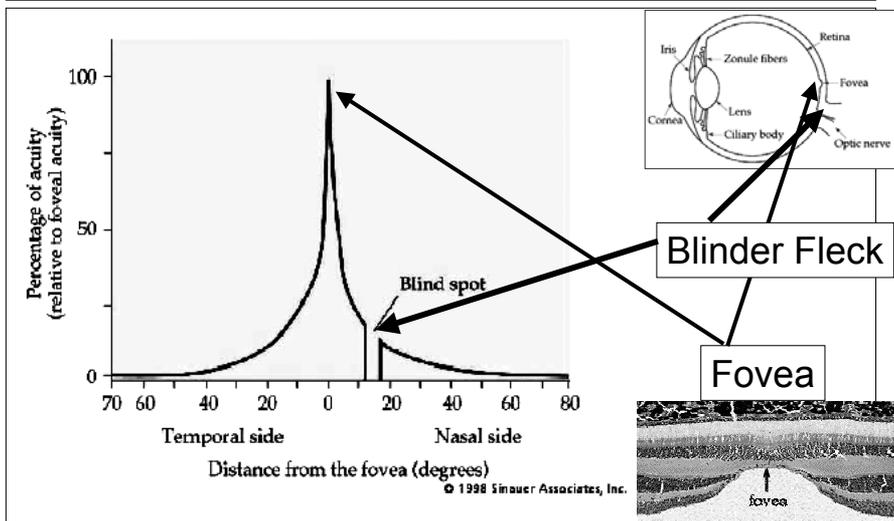


Fovea: Bereich höchster Auflösung

(a) Distributions of rods and cones across the retina



Fovea: Bereich höchster Empfindlichkeit



Anpassung

Anpassung der Augen an die Helligkeit der Umgebung
(mit Verzögerung)

- Helladaptation:
 - Pupillenverengung
 - Übergang vom Stäbchen- zum Zapfensehen
- Dunkeladaptation:
 - Übergang vom Zapfen- zum Stäbchensehen
 - Stäbchenadaptation
 - Stäbchen außerhalb Zentrum: Bei Dunkelheit im Zentrum schwächere Unterscheidung als daneben

Brennweite der Augenlinse

abhängig von der Wellenlänge

- groß im roten Bereich
- klein im blauen Bereich

Bild mit roten und blauen Bereichen: Auge ermüdet.

Rot vor blauem Hintergrund und umgekehrt vermeiden!

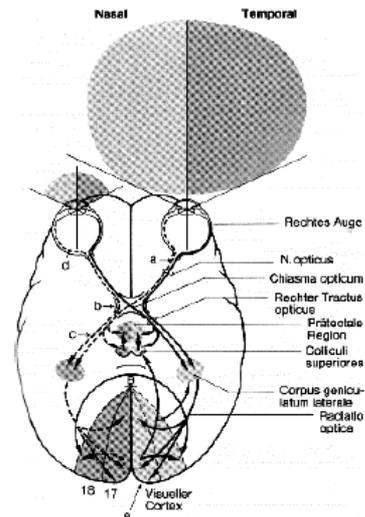
Räumliche Wahrnehmung: Rot vorn, blau hinten



Menschliches Sehen: Das Gehirn

Verarbeitung in mehreren Stufen

Steuerung der Bewegungen



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

65

Verarbeitung im menschlichen Gehirn

Im Zwischenhirn werden die separaten Signale für Rot (R), Grün (G), Blau (B) in drei neue Signale umgewandelt:

- Luminanzsignal: $Y = R + G$ Helligkeit (Gelb)
- Differenzsignal: $R - G$ Rot/Grün-Unterscheidung
- Differenzsignal: $Y - B$ Blau/Gelb-Unterscheidung

Gelb-Anteil wesentlich für Helligkeitswahrnehmung

Blau-Anteil unwesentlich für Helligkeitswahrnehmung

Farbkontraste Rot/Grün und Blau/Gelb besonders klar

"Komplementärfarben"



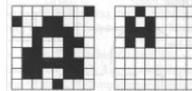
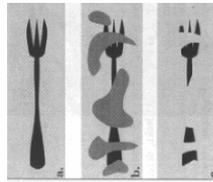
H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

66

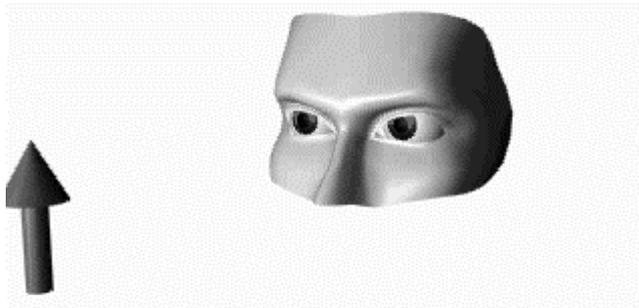
Verarbeitung im menschlichen Gehirn

- Ergänzung
- Identifikation
- Interpretation
- ...



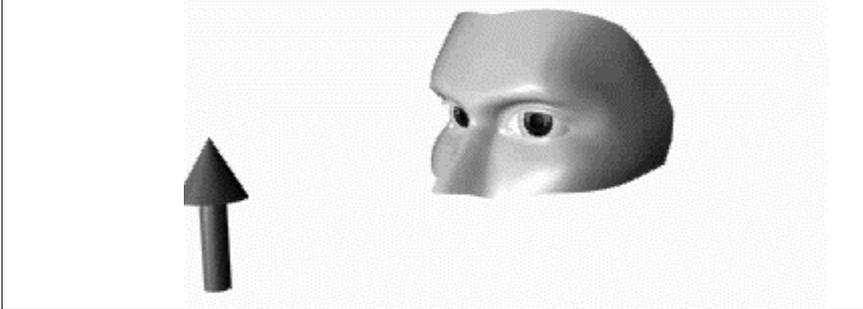
Sakkaden

- Rasches Anvisieren von interessierenden Merkmalen:
(Korrektur der Sichtlinie)
- Während der Augenbewegung keine Wahrnehmung
- Unwillkürlich (schnell) und willkürlich möglich



Vestibulo-Okulärer Reflex

- Kompensation von Eigenbewegungen
- Augen werden entgegen der Kopfbewegung bewegt
- Blickrichtung bleibt annähernd konstant
- Rückkopplung mit Gleichgewichtssinn (schneller)



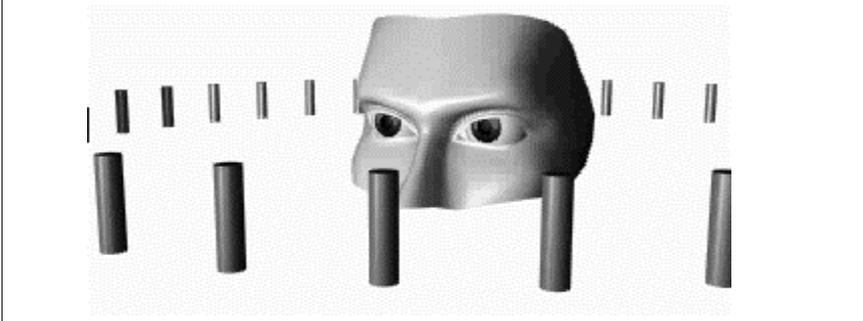
H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

69

Optokinetischer Reflex

- Ausgleich langanhaltender Bewegungen (wenn Gleichgewichtsänderung abklingt)
- Rückkopplung über visuelles System



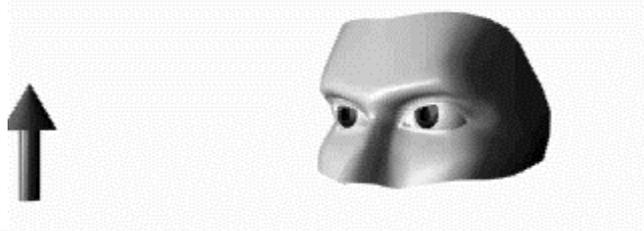
H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

70

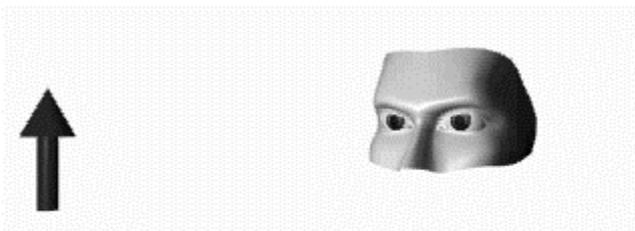
Augenfolgebewegung (*Smooth Pursuit*)

- gleitende Nachführung des Auges
- dadurch wird das Bild von bewegten Objekten dauerhaft auf einer Stelle der Netzhaut gehalten
- Rückkopplung mit unterschiedlichen Bereichen



Stereo-Sehen, Vergenz

- Koordination zwischen den Bewegungen beider Augen
- binokuläres Erfassung von Objekten
(Korrespondenz-Problem, entfernungsabhängig)
- räumliches Sehen



CCD-Kamera

CCD = Charge Coupled Device:

Lichtempfindlicher Sensor:

- Ladung proportional zur Lichtstärke
- Nach Aufnahme (Verschluss schließen) Übertragung des Signals (Ladungs-Kopplung) in einen Speicher zur weiteren Verarbeitung.

CCD-Kamera: Rasterförmige (Zeile, Feld) Anordnung von CCD-Zellen zur Aufnahme von Bildern.

- Das gespeicherte Bild (frame) ist die Matrix der von jeder Zelle empfangenen Werte.
- Die Bildpunkte heißen Pixel (=picture element).

CCD-Kamera

Farbaufnahmen:

- CCD-Zellen mit Farbfiltern
Anordnung z.B. 4 Zellen quadratisch mit Empfindlichkeit für rot(1)-grün(2)-blau(1) entsprechend RGB-Farbraum mit höherer Empfindlichkeit für grün (Einbuße an Schärfe gegenüber Schwarz-Weiß)
- Bildpunkte (Pixel) kodieren die Farb-Werte („Kanäle“).
- Aufnahmen auch jenseits des sichtbaren Lichts möglich.

Softwaretechnische Aufbereitung:

- Verbesserung des Bildes
(Interpolation, Rauschunterdrückung usw.)
- Kompressionsalgorithmen

Abbildungsvorgang

1. Lichtstrahlung aus verschiedenen Quellen im 3D-Raum:
2. Lichtstrahlen treffen auf Abbildungsfläche
 - Im Auge: Netzhaut
 - In der Kamera: Film, CCD-Chip
 - *Modellierung als Projektion 3D nach 2D (abhängig von Kamera-Konstruktion)*
3. Lichtstrahlen werden in Signale umgewandelt
 - CCD-Kamera:
 - *Quantisierung gemäß*
 - *Rasterung (Anordnung der Zellen)*
 - *Diskretisierung der Signalwerte*

Quantisierung

2 Vorgänge bei Erzeugung:

- Hardwaremäßig vorgegebene Rasterung (rechteckig, quadratisch, sechseckig)
Auslesen z.B. bei CCD 30mal pro Sekunde 640 x 480 Intensitätswerte
- Softwaremäßige Quantisierung (Digitalisierung):
Analog-Digital-Wandler („Frame Grapper“)
transformiert Intensitätswerte in 8-bit oder 16-bit-Werte
 - Grauwerte bei „Schwarz-Weiß“-Kanal
 - Farbwerte in Farbkanälen

Digitale Bilder (1-Kanal-Bilder)

- Felder von „Pixeln“ („picture elements“), jeweils x-y-Position, „Grauwert“
- Darstellung als Matrix $[g(x,y)]_{x,y}$ von Grauwerten mit
 - Anordnung: (1,1) ist oben links
 - x nach unten, y nach rechts
 - Wertebereich z.B. 8 bit, 16 bit

1,1	1,2	1,3	1,4	1,5
2,1	2,2	2,3	2,4	2,5
3,1	3,2	3,3	3,4	3,5
4,1	4,2	4,3	4,4	4,5

8-bit-Darstellung:

- 0 für Schwarz
- 255 für Weiß
- Binärbild: 1 bit (0=Schwarz)

Digitale Bilder (1-Kanal-Bilder)

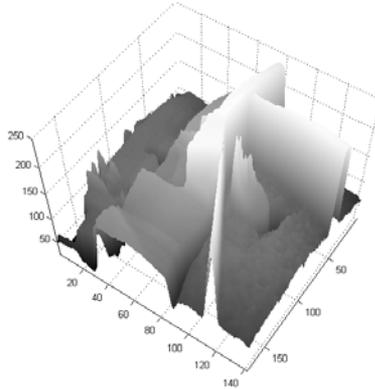
- Felder von „Pixeln“ („picture elements“), jeweils x-y-Position, „Grauwert“

```

*** **
118 120 122 122 120 128 122 121 120 120 121 121 122 121 129 121 123 123 123 123 121 117 116 112 108 1
120 121 119 121 119 122 120 120 119 121 121 121 121 121 123 123 123 123 125 121 115 114 115 114 1
120 119 121 122 120 119 118 121 121 120 121 120 122 120 120 120 120 120 120 120 120 120 117 112 110 118 119 1
117 118 122 121 117 118 119 118 120 120 122 119 120 122 120 118 121 121 123 121 122 119 112 104 116 123 1
117 121 119 120 120 120 120 123 122 122 120 123 122 121 122 121 126 121 121 121 112 104 111 120 1
118 118 122 120 122 128 124 123 124 125 122 122 121 121 122 122 126 123 123 121 112 104 110 122 1
120 121 124 124 126 130 131 128 127 129 126 128 124 121 120 123 122 122 121 118 111 108 111 103 1
122 128 129 129 127 128 131 131 129 135 135 128 128 128 126 123 122 122 119 115 107 106 115 119 1
125 127 130 130 128 129 129 129 131 136 138 134 132 133 131 126 121 122 117 111 103 100 108 109 1
129 130 130 132 128 128 129 128 129 131 130 133 134 131 128 124 116 104 108 102 89 88 90 90
124 130 130 130 130 128 128 127 126 122 121 119 119 118 116 111 109 102 98 94 96 92 89 87 84 80 71 62 42 46
120 127 128 125 124 121 119 115 116 111 109 102 98 94 96 92 89 87 84 80 71 62 42 46
117 116 123 117 107 103 102 97 97 93 91 84 80 76 76 76 71 73 72 73 66 53 51 55
94 96 100 100 91 85 83 79 76 71 70 71 66 63 62 62 41 60 41 60 43 56 51 50
74 76 79 78 75 71 68 63 69 62 57 62 57 55 53 55 56 54 54 57 59 57 53 53
66 64 64 65 67 64 57 53 52 52 56 54 53 50 52 49 53 52 52 54 56 51 53 54
65 59 55 55 60 58 51 48 47 51 53 54 48 48 50 52 54 54 53 53 55 50 51 56
64 58 53 55 56 54 48 48 48 52 50 49 46 48 50 52 55 54 51 54 54 51 53 54
63 61 56 56 53 50 48 46 46 48 47 48 48 51 49 51 53 54 57 55 51 54 50
63 61 59 58 60 58 54 47 47 43 44 44 46 48 50 51 51 48 49 52 52 53 53 58 58
63 61 59 56 58 60 52 44 47 43 43 48 48 53 53 50 52 51 49 49 52 51 52 53 54
62 61 62 60 57 56 53 44 47 44 47 48 50 52 54 53 51 51 51 52 53 54 52 55
61 59 65 62 57 53 53 48 50 49 45 48 47 47 48 52 52 51 51 52 54 53 55 56
57 59 62 60 58 54 52 50 52 54 44 45 45 50 51 51 51 52 52 56 54
57 58 62 60 57 56 56 52 49 52 48 47 47 45 49 49 49 51 50 52 51 55 53 51
60 59 62 58 56 57 54 51 51 52 51 48 49 48 50 48 51 49 51 54 53 52 50
59 60 61 60 61 60 60 54 53 55 51 49 49 50 50 49 49 53 52 49 50 52 51
56 50 57 61 61 60 58 55 53 53 54 52 48 51 50 48 51 54 51 49 49 51 53 52
60 59 59 56 56 57 56 56 55 56 58 48 47 50 52 49 51 50 48 48 52 50 51 50
63 60 62 60 58 54 57 58 56 56 58 51 50 48 52 48 50 48 49 48 50 51 51 51
64 59 61 63 60 55 56 57 59 55 59 52 53 49 49 48 48 48 50 51 52 51 52 55
61 62 64 60 59 59 57 57 60 54 54 49 54 51 51 49 49 51 53 52 53 52 53
58 61 61 58 56 60 59 57 57 55 51 50 53 53 52 53 52 48 51 51 52 54 51
59 59 59 61 59 58 53 54 55 56 57 54 53 55 52 54 53 48 53 53 53 52 53
59 60 60 60 58 60 60 54 57 53 53 53 53 51 48 52 51 53 52 55 52 50 49
59 61 61 61 59 57 56 54 55 55 55 52 53 56 52 53 53 53 53 54 48 50 49
59 59 59 61 59 58 53 54 55 56 57 54 53 55 52 54 53 48 53 53 53 52 52
63 59 60 57 57 58 53 54 58 58 58 57 53 57 55 52 50 54 59 56 52
62 60 56 58 58 58 59 56 57 54 53 53 55 55 57 55 53 53 54 55 58 54
59 59 57 58 62 58 58 56 57 56 55 56 54 55 52 55 57 54 55 57 55 54 53
62 58 57 59 61 59 55 54 54 56 60 59 55 51 52 54 52 56 54 55 55 55 53 51
    
```

Digitale Bilder (1-Kanal-Bilder)

- Felder von „Pixeln“ („picture elements“), jeweils x-y-Position, „Grauwert“



Digitale Bilder (1-Kanal-Bilder)

Andere Darstellungen

- **Daten-Komprimierung: Originalbild rekonstruierbar**
 - RLE (run length encoding, Lauflängenkodierung)
für Binär-Bilder z.B. nur Anfangspunkt und Länge von horizontalen schwarzen Linien angeben.
 - Weitere Möglichkeiten: Baumstrukturen, Richtungsketten.
Oft angepasst an spezielle Formen der Weiterverarbeitung.
- **Daten-Reduktion: Originalbild nicht vollständig rekonstruierbar**
 - Reduktion der Grauwertskala
 - Informationstheoretische Verfahren
 - Physiologische Unterscheidungsfähigkeit

Quantisierung mit

- 128, 64,
 - 32, 16,
 - 8, 4
- Grauwerten



Histogramme

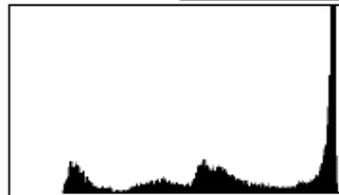
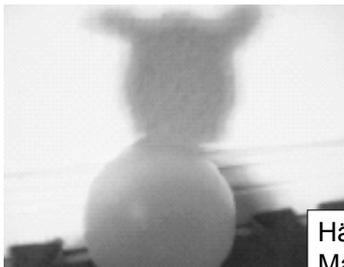
Relative Häufigkeit von Grauwerten $a \in [0 \dots 255]$

$\text{Hist}(a) := A/M$

A := Anzahl Pixel (x,y) mit $g(x,y)=a$

M := Anzahl aller Pixel

Diplomarbeit Jüngel



Häufigkeiten für Grauwerte als Abszissen.
Maximum bei hohem Weißanteil.

Histogramme für Zeilen/Spalten

Relative Häufigkeit von Grauwerten $a \in [0 \dots 255]$ für Zeilen bzw. Spalten

Spalten:

$$\text{hist}_{\text{Spalte}_i}(a) := A_{\text{Spalte}_i} / M_{\text{Spalte}_i}$$

$A_{\text{Spalte}_i} :=$ Anzahl Pixel (x,y) mit $g(x,y)=a$ in Spalte i

$M_{\text{Spalte}_i} :=$ Anzahl aller Pixel in Spalte i

Zeilen:

$$\text{hist}_{\text{Zeile}_j}(a) := A_{\text{Zeile}_j} / M_{\text{Zeile}_j}$$

$A_{\text{Zeile}_j} :=$ Anzahl Pixel (x,y) mit $g(x,y)=a$ in Zeile j

$M_{\text{Zeile}_j} :=$ Anzahl aller Pixel in Zeile j

Frage: gilt $\text{Hist}(a) = \sum_i \text{hist}_{\text{Spalte}_i}(a) = \sum_j \text{hist}_{\text{Zeile}_j}(a)$?

Histogramme: Anwendungen

Relative Häufigkeit im Bild z.B. anwendbar für

- Quantisierung
- Schwellwertverfahren
(Schwellwerte für Klassifizierung bestimmen)

Relative Häufigkeit für Zeilen bzw. Spalten

- Position von Bildinhalten
(in Schnittpunkten bzgl. Histogramm)
- Ähnlichkeit von Bildern

Analoge Verfahren mit Histogrammen für Farbwerte

Statistische Merkmale von Bildern

- Mittlerer Grauwert (Mittelwert):

$$m = (1/ LR) \cdot \sum_{i=1...L} \sum_{j=1...R} g(i,j)$$

Maß für Helligkeit eines Bildes

- Mittlere quadratische Ableitung (Varianz):

$$\sigma^2 = (1/ LR) \cdot \sum_{i=1...L} \sum_{j=1...R} (g(i,j) - m)^2$$

Maß für Kontrast eines Bildes

- Entropie:

$$H = \sum_{a=0...255} \text{Hist}(a) \log_2 \text{Hist}(a)$$

Maß für mittlere Unsicherheit des Grauwerts eines Bildpunktes

Maß für mittlere Anzahl von Bits pro Bildpunkt

Mehrkanal-Bilder

- Kanäle mit unterschiedlichen Informationen

- Farbbilder (3 Kanäle, entsprechend Farbmodell)
- Multispektralbilder (weitere Kanäle, z.B. für erweiterte Farbempfindlichkeit oder andere Informationen)

- Bildelement (Pixel):

- x-y- Position, Intensitätswerte je Kanal

- Darstellung als Matrix

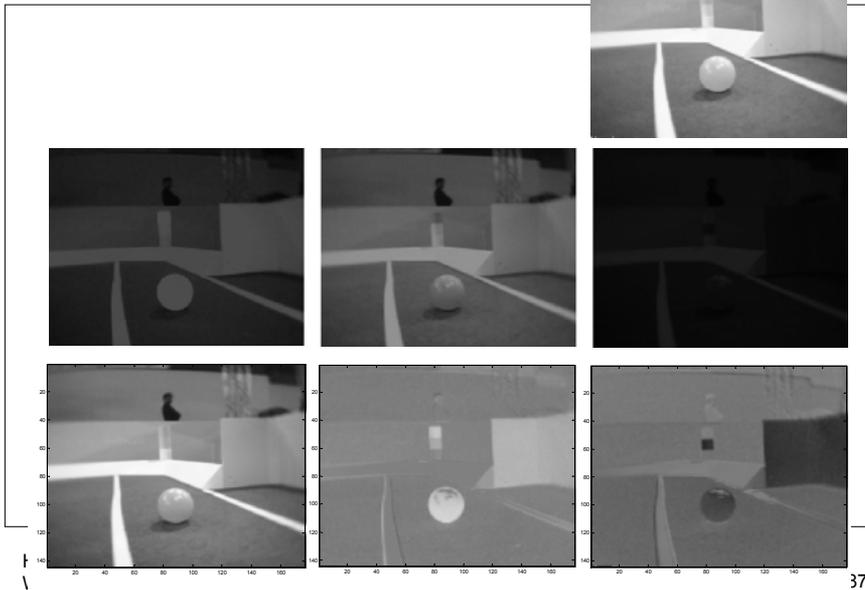
$$[f_1(x,y), f_2(x,y), f_3(x,y)]_{x,y}$$

bzw. je Kanal eine Matrix

$$[f_i(x,y)]_{x,y}$$

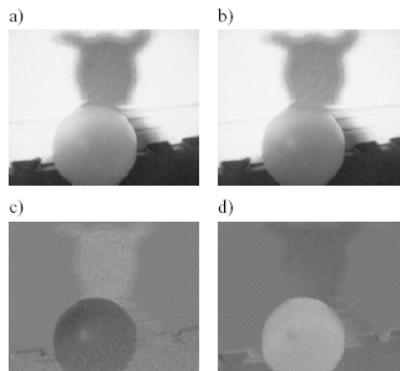
(interpretierbar als je ein „Graubild“ pro Kanal)

Farbkanäle in RGB



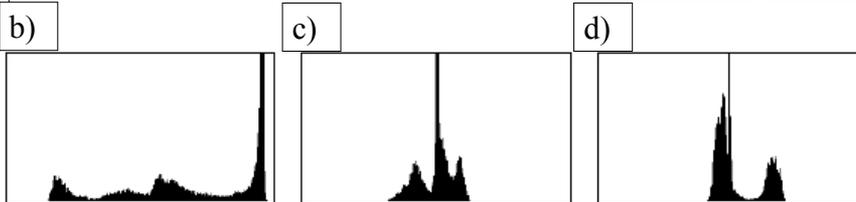
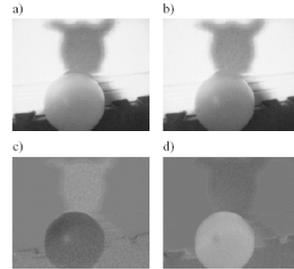
Farbkanäle in YUV

- a) Bild-Wiedergabe insgesamt
- b) Y-Kanal (Helligkeit) als Graubild
- c) U-Kanal (Blau) als Graubild
(blauer Schatten: hell,
orange Ball: dunkel)
- d) V-Kanal (Rot) als Graubild
(orange Ball
hat höchste Intensität)



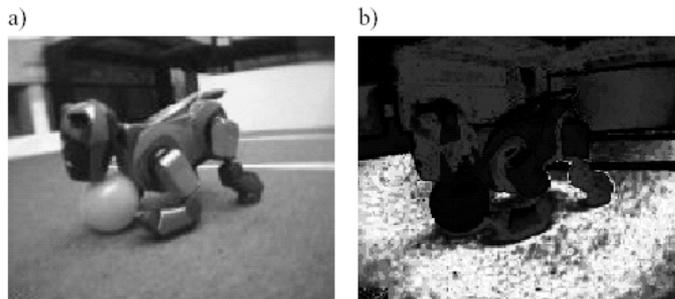
Histogramme der YUV-Kanäle

Maxima entsprechend den hohen Weißanteilen



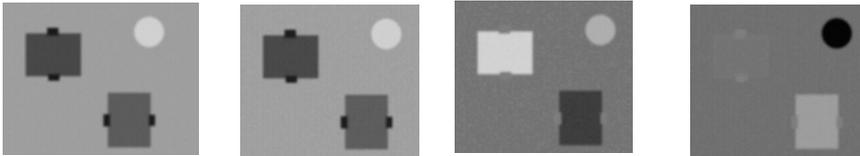
Häufigkeits-Bild

- Original-Pixel ersetzt durch Grauwert-Pixel entsprechend der Häufigkeit seiner Farbe:
- Pixel mit selten benutzter Farbe sind dunkel.
- Pixel mit oft benutzter Farbe sind hell.



Farbkanäle in YUV

- Starke Helligkeit (Y) bzw. starkes Rot bzw. starkes Blau erscheinen hell in der SW-Wiedergabe der Kanäle.
- Die einzelnen Kanäle können wie Graubilder behandelt werden (Kompression/Reduktion, Verarbeitung)
- Angaben über Nutzen der Kanäle aus Histogramm, Mittelwert, Varianz, Entropie ableitbar



Y (hell)

V (rot)

U (blau)

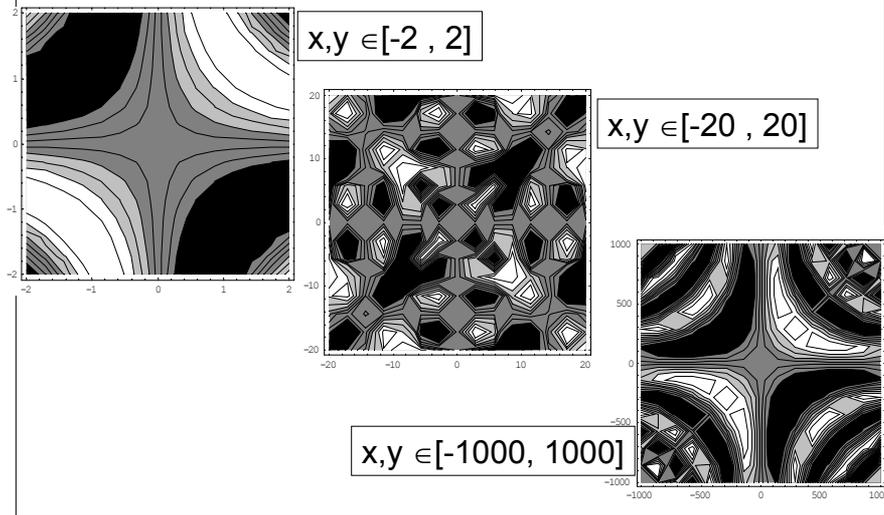
Bildqualität

Qualität bestimmt durch

- Wertebereiche der Intensitätswerte
(z.B. Anzahl der Grauwert-Töne: 8 bit, 16 bit)
- Rastergröße:
 - Die Rasterpunkte wirken als diskrete Abtastpunkte:
Nur an bestimmten Stellen wird gemessen.
 - Die Rastergröße muss ausreichen, um insbesondere kleine Strukturen korrekt wiederzugeben.

Bei ungenügender Rastergröße
entstehen neue Muster: „Alias“

Darstellung von $\sin(xy)$ in ungenügender Rasterung



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

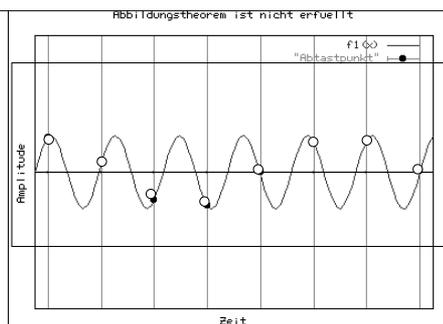
Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

93

Abtasttheorem

Problem:

Die rote Kurve wird nur an wenigen Punkten gemessen:
Es werden nur die schwarzen Punkte ermittelt.



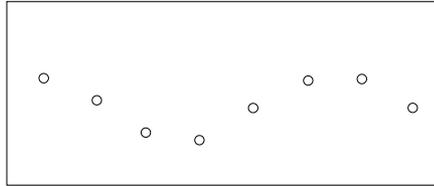
H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

94

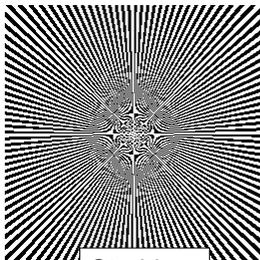
Abtasttheorem

Der Verlauf der schwarzen Punkte wird als eine Kurve niedrigerer Frequenz interpretiert: Alias

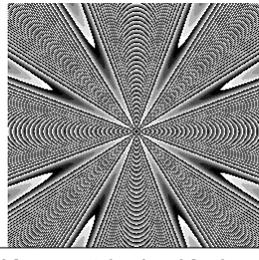


Moire

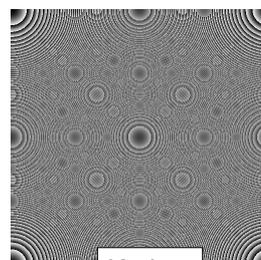
- Moire-Muster entstehen z.B. durch Filterwirkung (=Abtastpunkte) feiner Stoffe (Gardinen).



Strahlen



Konzentrische Kreise



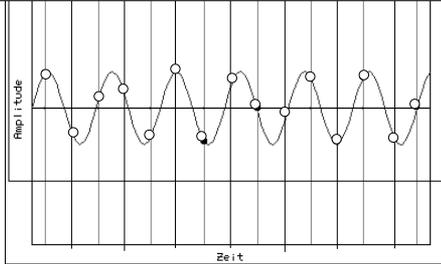
Kreise

Pixelgröße jeweils nicht ausreichend für exakte Wiedergabe.
Es entstehen tief-frequente Muster.

Abtasttheorem

Wie oft müsste man messen?

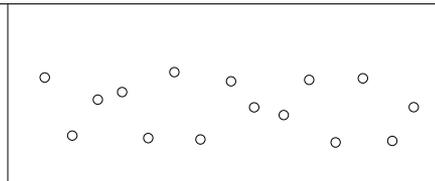
Beispiel: Verkleinerung der Messintervalle
-- Mehr der Messpunkte



Abtasttheorem

Wie oft müsste man messen?

Beispiel: Verkleinerung der Messintervalle
-- Mehr der Messpunkte



Abtasttheorem

Abtasttheorem

Für korrekte Wiedergabe muss gelten:

Pro Wellenlänge T mindestens 2 Abtastpunkte,
d.h. Abtastrate $\Delta x < T/2$. (Nyquist-Bedingung)

oder:

Abtast-Frequenz mindestens doppelt so groß
wie höchste vorkommende Frequenz.

Beispiel:

Sinusförmige Intensitätsfunktion $\sin(\omega x)$ mit Frequenz $\omega = 2\pi / T$
und Wellenlänge T .

Bei Abtastrate Δx werden Frequenzen $\omega > \pi / \Delta x$ nicht mehr korrekt
wiedergegeben.

Abtasttheorem

Die Bedingung muss in jeder Richtung des Bildes gelten.

Wenn die Bedingung nicht erfüllt ist, passiert zweierlei:

1. Korrekte Bildinhalte werden nicht übermittelt.
2. Es entstehen zusätzliche störende Muster.

Die störenden Muster können durch vorherige
Tiefpassfilter verhindert werden:

Nicht korrekt abtastbare hochfrequente Bildinhalte
werden eliminiert.

Kamera, Abbildung, Projektion

Voraussetzung für scharfe Darstellung:

- Jeder Bildpunkt wird nur von einem Licht-Strahl aus der Umgebung getroffen (näherungsweise: Licht aus einem engen Raumwinkel)

Das wird erreicht durch

- Lochkamera
- Kollimator (Bündel von Röhren)
- Linsensysteme (Bündelung von Lichtstrahlen: Aus dem gleichen Ursprung stammende Strahlen werden auf den gleichen Bildpunkt gelenkt.)

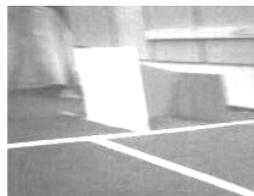
Bei großer Linse (großer Öffnung der Blende):
höhere Lichtausbeute, aber geringere Tiefenschärfe.

Kalibrierung + Korrekturen

Die Bilder (bzw. allgemein: Sensordaten) sind nicht immer so, wie man glauben möchte ...



Fig. 2. Barrel distortion of the field with a 4.2 mm lens



Kameramodellierung

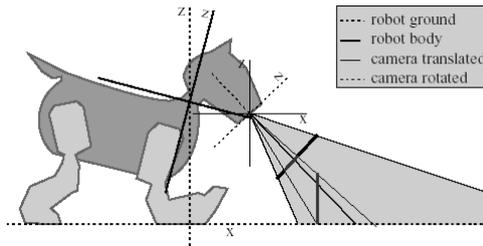
Unterschiedliche Koordinatensysteme.

Grauer Kegel:

Blickrichtung, optische Achse, virtuelle Bildebene

Rote Linien:

Ein Objekt und seine Projektion auf die Bildebene

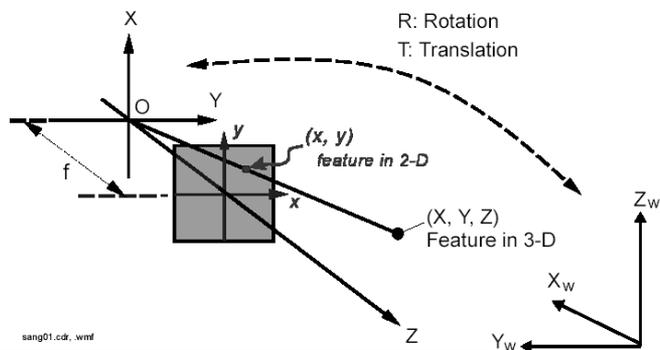


H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

103

Kameramodell („Vorwärtsmodell“)



X, Y, Z : Kamera-Koordinatensystem

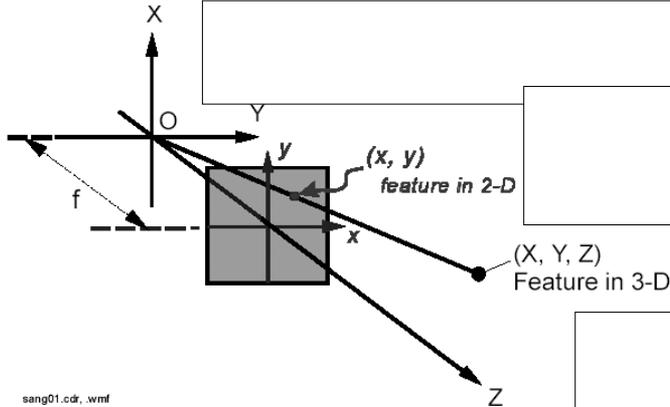
X_w, Y_w, Z_w : Welt-Koordinatensystem

f : Brennweite

Die Bildkoordinaten (x, y) sind durch die
Objektkoordinaten (X, Y, Z) eindeutig bestimmt.

104

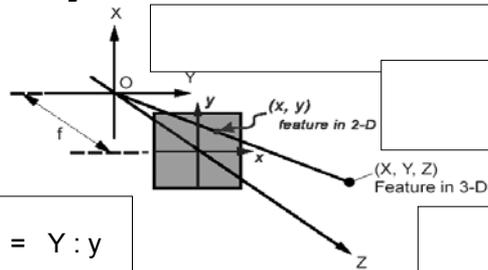
Perspektivische Projektion



sang01.cdr, wmf

X, Y, Z : Kamera-Koordinatensystem.
 Z -Achse = optische Achse, Bildebene senkrecht dazu.
 Bildkoordinaten (x, y) durch Objektkoordinaten (X, Y, Z) eindeutig bestimmt.

Perspektivische Projektion



Strahlensatz: $Z : f = X : x = Y : y$

Umrechnung $x = f \cdot Z/X$ $y = f \cdot Z/Y$

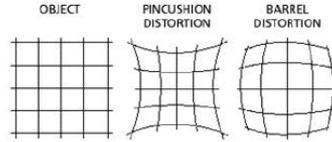
Abbildung:
$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \Rightarrow \frac{f}{Z} \begin{pmatrix} x \\ y \end{pmatrix}$$

Kamera, Abbildung, Projektion

Tatsächlich ist das Kameramodell komplexer.

Intrinsische Parameter:

- Brechung des Lichts bei Eintritt und Austritt
- Unterschiedliche Gegenstandsentfernung
- Linsenmodelle (Farbe, Verzerrung)
- Sensor-Eigenschaften (Norm-Abweichungen einzelner Sensoren)



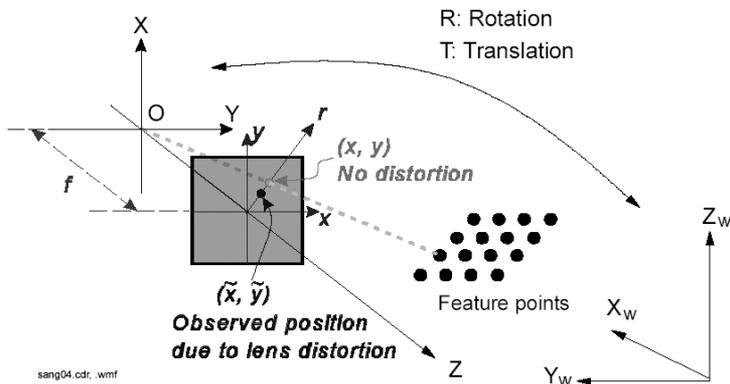
Kalibrierung einer konkreten Kamera:

Genauere Bestimmung der tatsächlichen Abbildungsverhältnisse
bzgl. geometrischer Abbildung und Farbwiedergabe

Kamera-Kalibrierung bzgl. Projektion

Ziel: Verzerrungen ausgleichen

Methode: Abbildungsparameter bestimmen mittels
zugeordneter vermessener Punkte im Raum und im Bild.



Kamera-Kalibrierung bzgl. Projektion

Abbildung hängt ab von

- Intrinsischen Parameter (Kamera intern)
- Extrinsische Parameter (Position Kamera im Raum)

Abbildungsmodell:

Zusammenhang zwischen

X, Y, Z : Kamera-Koordinatensystem und

X_w, Y_w, Z_w : Welt-Koordinatensystem

als Projektionsmatrix \mathbf{P} mit $[X, Y, Z] = \mathbf{P} \cdot [X_w, Y_w, Z_w]$

\mathbf{P} bestimmen aus einander zugeordneten vermessenen Punkten im Raum ($[X_w, Y_w, Z_w]$) und im Bild ($[X, Y, Z]$).

Kamera-Kalibrierung

Beispiel Deckenkamera Small-Size-League
(FU-Fighters Berlin)

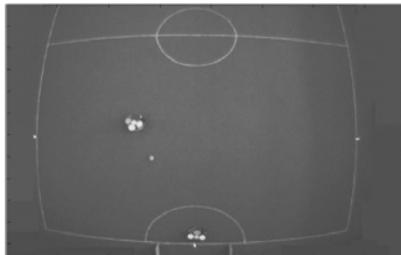


Fig. 2. Barrel distortion of the field with a 4.2 mm lens

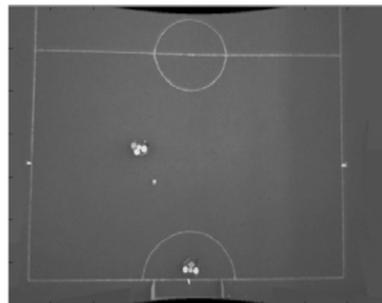


Fig. 6. Corrected field image

Kamera-Kalibrierung bzgl. Farbe

Farb-Korrekturen der intrinsischen Parameter

- Optik-bedingt insbesondere am Bildrand
(AIBO: Am Rand blau verstärkt)
- Sensorbedingt: Unterschiedliche Empfindlichkeit
- Für Kalibrierung: Farbwerte Soll/Ist vermessen

Varianten für Auswertung (Projektion und Farbe):

- Neues Bild berechnen
- Bei Bild-Verarbeitung Korrekturen berücksichtigen
(z.B Scan-Linie entsprechend modifizieren)

Kalibrierung bzgl. Farbe der Umgebung

Bzgl. Farbe/Farbtabelle (analog: Grauwerte)

abhängig von Beleuchtung

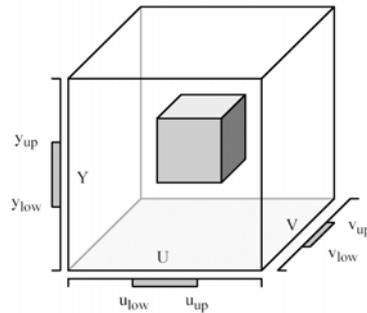
- „wie sieht orange heute aus“
- „wie sieht das Objekt heute aus“
- Beispielbasiert:
 - Referenzfarben im Raum (bzw. am Roboter) mit Kamera erfassen.
 - Im Bild auswählen für Eintrag in Farbtabelle:
 - Manuell („anklicken“)
 - Automatisiert: Lernverfahren

Normierung anhand von Farbtafeln.

Problem bei Manueller Auswahl: Metamere Farben
(Kamera kann anders reagieren als Auge).

Farbtabelle

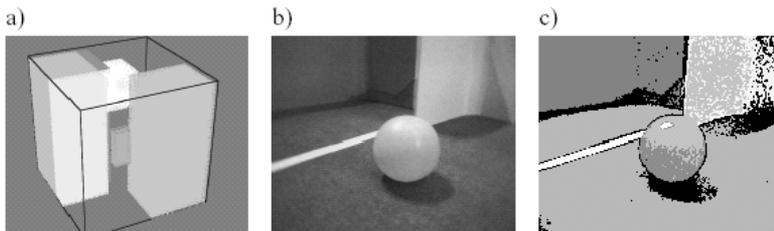
- Color-look-up-table:
Pixelwerte für „grün“ im Farbraum festlegen
- Ziel: als „grün“ klassifizierte Objekte erkennen



Diplomarbeit Jüngel

Farbtabelle

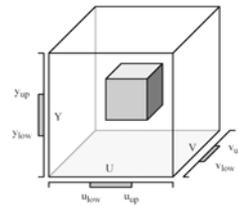
- a) Farben im Farbraum
- b) Bild
- c) Klassifiziertes Bild



Farbtabelle

Beschreibung der Farben:

- oft durch Intervalle
- bzw. Mengen von Intervallen

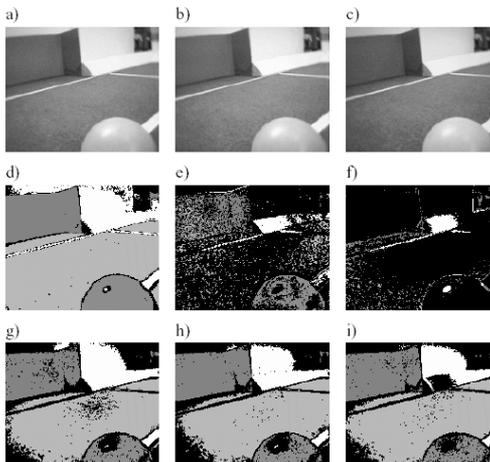


Ziel: Gute Klassifizierung,
d.h. möglichst großer Abstand zwischen Bereichen
unterschiedlicher Farben (insbesondere keine
Überschneidungen).

Unterschiedliche Farbräume mit unterschiedlicher Eignung.

Anpassung an Lichtverhältnisse

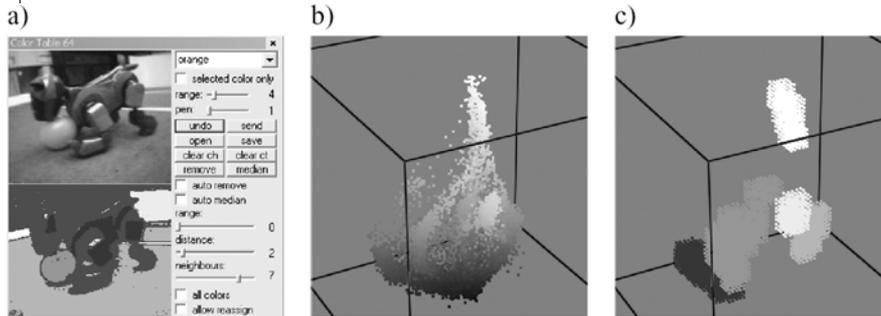
(a,b,c) Bilder, unter verschiedenen Bedingungen aufgenommen
(d,e,f) Farbklassifizierungen anhand einer für (a) erzeugten Farbtabelle,
(g,h,i) Farbklassifizierungen mit spezifischen Farbtabelle



Farbkalibrierung

Manuelle Farbkalibrierung (Aibo Team Humboldt)

- a) Interface: Bild und Klassifikationsresultat
- b) Bildraumbelegung des Bildes
- c) Bildraumbelegung der Farbtabelle



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

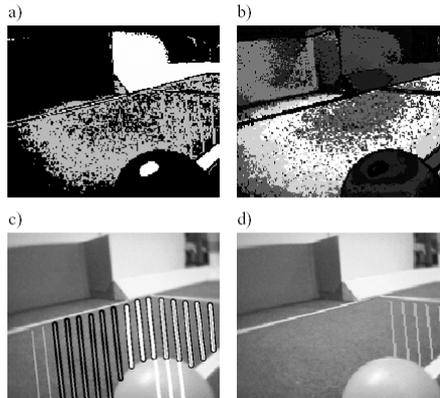
Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

117

Farbkalibrierung

Automatische Farbkalibrierung (Aibo Team Humboldt)

- (a) Identifizierte Referenzwerte für „grün“ (vgl. Bild (d))
- (b) Häufigkeitsverteilung der Farben.
- (c) Scanlinien mit unterschiedlicher Bewertung bzgl. „grün“
- (d) Für Kalibrierung benutzte Scanlinien



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung k
Verarbeitun

Farbkalibrierung

Ein allgemeiner modellbasierter Ansatz.

Modell für Reflexion:

Für jede Wellenlänge λ ist die Reflexion $I(p, \lambda)$ eines Oberflächenpunktes p proportional zu

- Eingestrahlem Licht auf Oberfläche $E(p, \lambda)$
(Lichtstärke, Orientierung zum Licht)
- Reflektivität der Oberfläche $S(p, \lambda)$
(Anteil, Farbe, Orientierung zum Beobachter)

$$I(p, \lambda) = E(p, \lambda) \cdot S(p, \lambda)$$

Farbkalibrierung

Einfacher linearer Zusammenhang für einzelne Wellenlänge

$$I(p, \lambda) = E(p, \lambda) \cdot S(p, \lambda)$$

Aber: Komplexer Zusammenhang für die Sensoren wegen Integration über alle Wellenlängen

$$e_x(p) = \int \phi_x(\lambda) I(p, \lambda) d\lambda = \int \phi_x(\lambda) E(p, \lambda) S(p, \lambda) d\lambda$$

$$e_y(p) = \int \phi_y(\lambda) I(p, \lambda) d\lambda = \int \phi_y(\lambda) E(p, \lambda) S(p, \lambda) d\lambda$$

$$e_z(p) = \int \phi_z(\lambda) I(p, \lambda) d\lambda = \int \phi_z(\lambda) E(p, \lambda) S(p, \lambda) d\lambda$$

Farbkalibrierung

$$e_x(p) = \int \phi_x(\lambda) I(p, \lambda) d\lambda = \int \phi_x(\lambda) E(p, \lambda) S(p, \lambda) d\lambda$$

$$e_y(p) = \int \phi_y(\lambda) I(p, \lambda) d\lambda = \int \phi_y(\lambda) E(p, \lambda) S(p, \lambda) d\lambda$$

$$e_z(p) = \int \phi_z(\lambda) I(p, \lambda) d\lambda = \int \phi_z(\lambda) E(p, \lambda) S(p, \lambda) d\lambda$$

Für Farbkalibrierung am Farbpunkt p wird benötigt:

Abweichung von $[e_x(p), e_y(p), e_z(p)]$

in Abhängigkeit von (unbekannten) Einstrahlungswerten $E(p, \lambda)$

bei fixierter (aber unbekannter) Reflektivität $S(p, \lambda)$.

Annahmen zur Vereinfachung z.B.:

$S(p, \lambda)$ als Linearkombination von 3 Basisvektoren (Farbintervall)

Weiteres: Siehe Literatur

Vorwärtsmodell der Projektion

Projektion Raum \rightarrow Bildebene

Abbildungsmodell (Projektion)

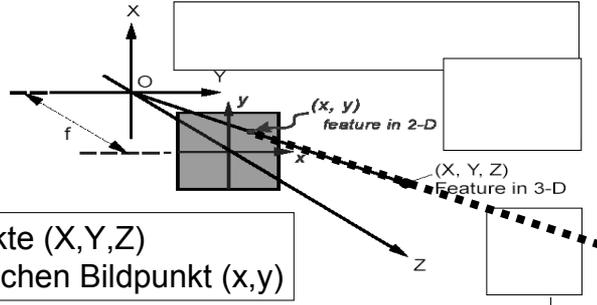
Reduktion der 3D-Welt auf 2D-Abbildung

mit räumlicher Information bzgl.

- Orientierung der Objekte
(Lage der Bildebene relativ zu Objekten)
- Lichtverhältnisse (Beleuchtung, Schatten)
- Relative Größe der Objekte
- Beziehungen zwischen Objekten

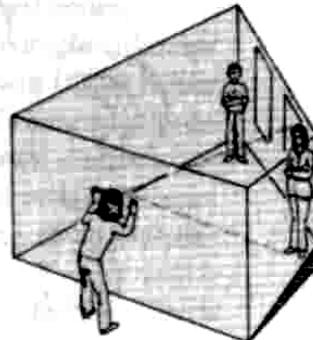
Rückwärtsmodell der Projektion

Rekonstruktion des 3D-Raumes aus 2D-Bild.
Mehrdeutigkeiten („schlecht gestelltes Problem“):
Urbilder eines Bildpunktes auf einer Geraden im Raum
Geometrie allein nicht ausreichend für Rekonstruktion



Rückwärtsmodell der Projektion

Größen-Illusion: Ames-Zimer



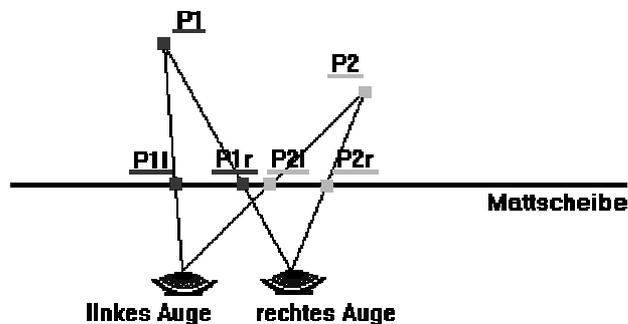
Rückwärtsmodell der Projektion

Stereo-Sehen:

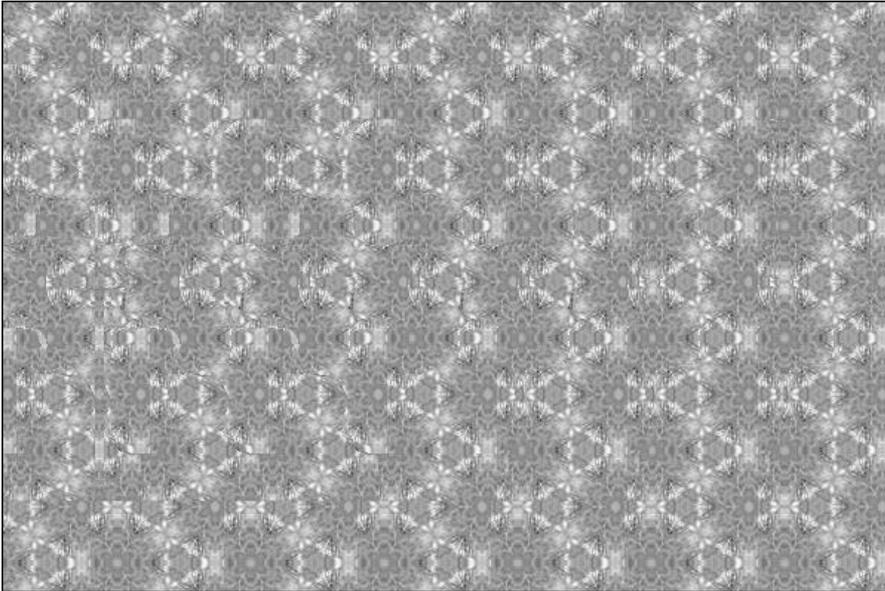
- 2 Kamerabilder aus unterschiedlichen Positionen (2 Kameras oder Kamera bewegen)
- Urbild von korrespondierenden Bildpunkten im Schnittpunkt der jeweiligen Geraden
 - ermöglicht Rückrechnung auf Tiefe
- Korrespondenz-Problem:
Welche Bildpunkte gehören zusammen?
 - Vergleich von Bildmerkmalen
 - Korrelationsverfahren

Autostereogramm

Korrespondenz durch korrespondierende Muster



Autostereogramm

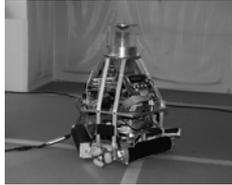


Rückwärtsmodell der Projektion

- Weitere technische Möglichkeiten:
 - Kombination mit Entfernungsmessung
(z.B. Laser: misst exakte Entfernung,
liefert aber keine Helligkeits- oder Farbinformation)
 - Bildfolgen, Bewegung
- Weitere Möglichkeiten bei der Interpretation
Zusatzwissen ausnutzen
(Größe, Umfeld, räumliche Beziehungen, ...)

Spezielle Bilderzeugung

Omnidirektionale Kamera
Spiegel



Unterschiedliche Oberflächenkrümmung erlaubt z.B.
zusätzlich vergrößerte Auflösung im Nahbereich

Anpassung der Algorithmen zur Auswertung

Bildverarbeitung

Ziele:

- effiziente Speicherung bzw. Verarbeitung
- verbesserte Darstellung
(Kontrastveränderung, Helligkeitsveränderung,
Elimination von Störungen, Ausgleich ...)
- Interpretation
Klassifikation von Bildelementen usw.

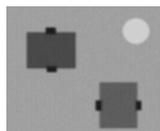
Bildverarbeitung

In der Regel mehrere Verarbeitungsschritte, z.B.

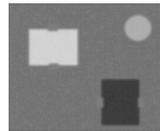
- Vorverarbeitung
 - z.B. Kalibrierung, Kontrastverstärkung, ...
- Komponenten bestimmen
 - z.B. Segmentierung: Farbbereiche, Umrisse, ...
- Objekte klassifizieren
 - z.B. Ball, Spieler
- Szenenbeschreibung
 - z.B. Dribbeln

Beispiel: Objekterkennung

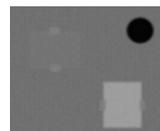
- welcher Farbkanal enthält Informationen
(z.B. Auswertung der Histogramme der Kanäle)



Y (hell)



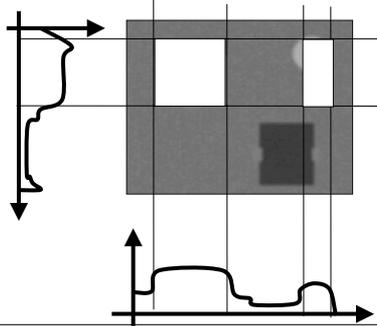
V (rot)



U (blau)

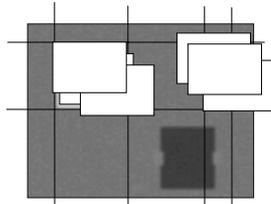
Beispiel: Objekterkennung

- Hypothesen für mögliche Orte des Objekts
- Hypothesen für mögliche Größe des Objekts
- Hypothesen für Farberscheinung des Objekts
(Intensitäts-Mittelwerte pro Zeilen/Spalten auswerten)



Beispiel: Objekterkennung

Vergleich des gesuchten Objekts
mit hypothetischen Positionen im Bild



Maß für Ähnlichkeit: Korrelation

Korrelation

Die Korrelation $\text{corr}(p)$ zwischen einem Objekt h und einem entsprechenden Ausschnitt des Bildes g wird an verschiedenen Stellen $p=(u,v)$ berechnet.

Sie kann als Wahrscheinlichkeit dafür verwendet werden, dass sich das Objekt am Platz p befindet.

Vergleich von zwei Grauwert-Bildern $[g(x,y)]_{x,y}$, $[h(x,y)]_{x,y}$.

Der Abstand

$$\begin{aligned}d^2 &:= \sum_{x=-k..+k} \sum_{y=-l..+l} (g(x,y) - h(x,y))^2 \\ &= \sum_{x=-k..+k} \sum_{y=-l..+l} (g(x,y)^2 + h(x,y)^2 - 2g(x,y) \cdot h(x,y))\end{aligned}$$

wird minimiert indem die Korrelation

$$\text{corr} = \sum_{x=-k..+k} \sum_{y=-l..+l} g(x,y) \cdot h(x,y)$$

maximiert wird.

Bildverarbeitung: Partielle Verfahren

Zeit-effiziente Verfahren:

Verzicht auf Verarbeitung des gesamten Bildes,
Partielle Verarbeitung

- Verarbeitung von reduzierten Bilder
- Untersuchungen längs ausgewählter Scanlinien:
 - Histogramme, Farbklassifikation, ...
 - Histogramme nur für ausgewählte Farbwerte
- Bei Bedarf:
Detailliertere Untersuchung in begrenzten Bereichen

Bildverarbeitung: Transformationen

Ein Bild $g = [g(x,y)]_{x,y}$

– Matrixdarstellung:

$g(x,y)$ (Grauwert oder) Farbinformation

wird transformiert in ein Bild $g' = [g'(r,s)]_{r,s}$

– sowohl Angaben zu Pixeln als auch Bildgröße können verändert werden (meist aber Bildgröße gleich)

Methoden

$$[g'(r,s)]_{r,s} = \text{Trafo}([g(x,y)]_{x,y})$$

- Punktweise Verfahren
- Verfahren im Ortsraum
- Verfahren im Ortsfrequenzraum

Arten von Transformationen

Geometrische Ortsabhängigkeit

- Punktbezogen
- Lokale Nachbarschaft U
- Global (z.B. Fouriertrafo)

Inhaltliche Ortsabhängigkeit

- Homogen: unabhängig von Bildkoordinate
- Inhomogen: abhängig von Bildkoordinate

Linearität: linear/nicht-linear

Beispiele:

- punktbezogen, homogen: Negativbild $g'(x,y) = 1 - g(x,y)$
- lokal, inhomogen: Kompensation von Kamera-Bewegung

Bildverarbeitung: Punktweise Verfahren

Transformationen der Farbinformation erfolgt isoliert für jedes einzelne Pixel.

$$[g'(x,y)]_{x,y} = \text{Trafo}([g(x,y)]_{x,y})$$

realisiert durch punktweise Umformung

$$g'(x,y) = \text{Trafo}(g(x,y)) \text{ für alle Punkte } (x,y)$$

Bildverarbeitung: Punktweise Verfahren

Lineare Skalierungsverfahren (Trafo linear)

- Helligkeit verändern: Additiv $g'(x,y) := g(x,y) + c$
(ggf. Abschneiden an Grenzen)
- Kontrast verändern: Multiplikativ $g'(x,y) := c \cdot g(x,y)$
(ggf. Abschneiden an Grenzen)
- Vorgegebene Mittelwerte/Varianzen erzielen
(z.B. eine mittlere Helligkeit)

Beispiel:

Maximaler Kontrast unter Ausnutzung des verfügbaren Bereiches
(Hilfsmittel: Histogramme)

Bildverarbeitung: Punktweise Verfahren

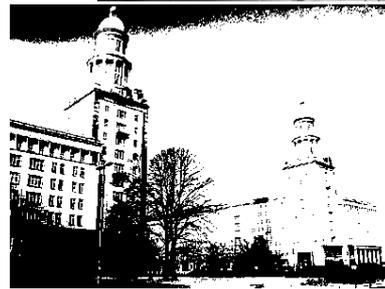
Stückweise konstante Trafo

- Äquidensiten (vergrößerte Quantisierung), Binärbilder
- Klassifikation/Segmentierung für gegebene Zerlegung (Klasseneinteilung) K_1, \dots, K_n des Farb-/Grauwertbereichs:
$$g'(x,y) := K_i \text{ für } g(x,y) \in K_i$$

bzw.
$$g'(x,y) := w_i \text{ für } g(x,y) \in K_i \text{ und „typisches“ } w_i \in K_i$$
- Transformation von Farbräumen

Bildverarbeitung: Punktweise Verfahren

Quantisierung der Intensität (Äquidensiten)

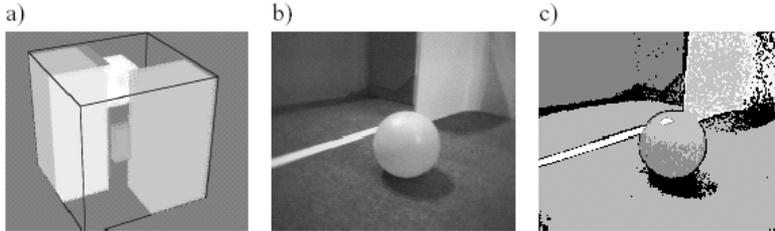


Bildverarbeitung: Punktweise Verfahren

Beispiel:

Segmentierung (c) eines Bildes (b) für gegebene Zerlegung (a) des Farbraums

$$g'(x,y) := w_i \text{ für } g(x,y) \in K_i \text{ und „typisches“ } w_i \in K_i$$



Zerlegung z.B. gemäß Farbtabelle, Klassen als Intervalle oder Mengen.

Bildverarbeitung: Punktweise Verf

Äquidensiten:

- stückweise konstante Skalierung von Grauwertbildern analog zu Farbsegmentierung:

$$g'(x,y) := w_i \text{ für } g(x,y) \in K_i \text{ und „typisches“ } w_i \in K_i$$

Bestimmung

geeigneter Intervalle K_i und $w_i \in K_i$

ausgehend z.B.

von Mittelwert, Varianz, Histogramm, ...

- Äquidensiten (insbesondere Binärbilder) als Vorstufe für weitere Verarbeitung, z.B. Kantenbildung:
- Äquidensiten 2. Ordnung (Kantenbild der Äquidensiten)



Bildverarbeitung: Punktweise Verfahren

Verknüpfung von Bildern

Addition

- z.B. Überlagerung von Luftaufnahmen mit topographischer Information, politischen Grenzen usw.

Subtraktion

- Zeitliche Differenzbilder: Erkennung von Bewegung
- Differenz verschobener Bilder: Erkennung gerichteter Kanten (parallel zur Verschiebung)

Differenz verschobener Bilder:



Bildverarbeitung: Ortsbereich

Verfahren im Ortsbereich:

Transformationen der Farbinformation erfolgt für jedes einzelne Pixel in Abhängigkeit von seiner Umgebung.
(Beispiel: Dilatation/Erosion)

$$[g'(x,y)]_{x,y} = \text{Trafo}([g(x,y)]_{x,y})$$

realisiert durch „**Diskrete Filteroperation**“

$$g'(x,y) = \text{Trafo}(\{g(u,v) / (u,v) \in U(x,y)\}) \text{ für alle Punkte } (x,y)$$

Mit $U(x,y)$ =Umgebung des Punktes (x,y) , z.B. alle Nachbarn

Beispiele

Rangordnungsoperatoren

- Median: $g'(x,y) := \text{Median}_{(x,y) \in U} g(x,y)$

Anwendung: Elimination isolierter Störungen

- Morphologische Operatoren (Speziell für Binärbilder)

- Dilatation: $g'(x,y) := \text{Max}_{(x,y) \in U} g(x,y)$

- Erosion: $g'(x,y) := \text{Min}_{(x,y) \in U} g(x,y)$

Anwendungen:

- Kanten (Differenzbildung)

- Elimination isolierter Punkte durch Dilatationen und danach Erosionen („Closing“) bzw. Erosionen und Dilatationen („Opening“)

Dilatation/Erosion: Kantenextraktion

Für Binärbilder (in gewissem Maße auch für andere Bilder)

Dilatation:

Bild G' entsteht aus G durch Ausdehnung weißer Flächen in jede Richtung um ein Pixel.

Erosion

Bild G'' entsteht aus G durch Ausdehnung schwarzer Flächen in jede Richtung um ein Pixel.

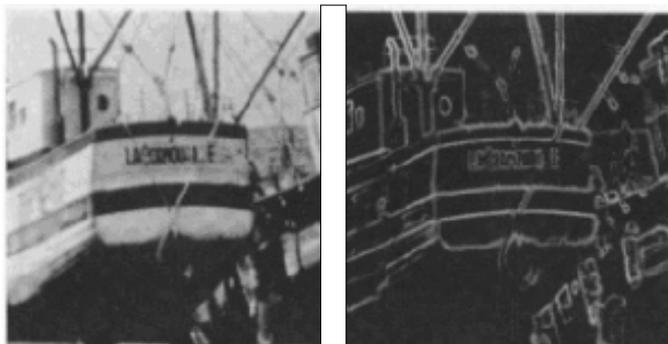
Bildung der symmetrischen Differenz $G' \text{ XOR } G''$

Ergebnis:

Bild der Kanten an den Übergängen Schwarz/Weiß.

Dilatation/Erosion: Kantenextraktion

Erosion/Dilatation und symmetrische Differenz



Bildverarbeitung: Ortsbereich

Quantisierung der Ortsinformation bzw. der Auflösung

(abhängig von Umgebung, dagegen Äquidensiten nur lokal)



H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

151

Bildverarbeitung: Ortsbereich

Diskrete **ortsinvariante** Filteroperationen

- Umgebungen $U(x,y)$ („Maske“, „Fenster“) immer in **gleicher Form** z.B. quadratische Umgebung 3×3

$(x-1,y-1)$	$(x,y-1)$	$(x+1,y-1)$
$(x-1,y)$	(x,y)	$(x+1,y)$
$(x-1,y+1)$	$(x,y+1)$	$(x+1,y+1)$

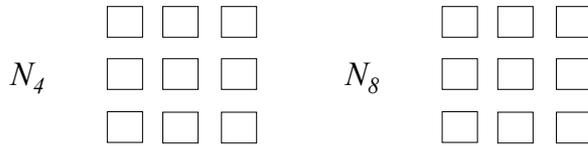
- Trafo **unabhängig vom Ort** (x,y)
- Am Bildrand spezielle Konventionen (z.B. Spiegelung, konstanter Wert, periodische Fortsetzung ...).

H.D.Burkhard, HU Berlin
Winter-Semester 2005/06

Vorlesung Kognitive Robotik
Verarbeitung visueller Daten

152

Umgebungen (Nachbarschaften, Masken)



- 4-Nachbarschaft eines Punktes $p=(r,c)$

$$N_4 = \{(r, c-1), (r-1, c), (r, c+1), (r+1, c)\}$$

$$= \{(u, v) \mid |u-r| + |v-c| = 1\}$$

- 8-Nachbarschaft eines Punktes

$$N_8 = \{(u, v) \mid \max[|u-r|, |v-c|] = 1\}$$

Faltungsoperationen

Beschreibung der Maske als Matrix $[h(u,v)]$ mit

- $h(u,v)$ als Wichtung („Kern“, „Maske“)
- Bezeichnung (u,v) gemäß Abständen zu (x,y) in U

$h(-1,-1)$	$h(0,-1)$	$h(+1,-1)$
$h(-1,0)$	$h(0,0)$	$h(+1,0)$
$h(-1,+1)$	$h(0,+1)$	$h(+1,+1)$

- Transformation als „Diskrete Faltung“ (mit Normierung):

$$g'(x,y) = g(x,y) * h(u,v)$$

$$= (1/|U|) \cdot \sum_{u=-k\dots+k} \sum_{v=-l\dots+l} g(x+u,y+v) \cdot h(u,v)$$

Faltungsoperationen: Anwendungen

- Glättung mit 5x5 - Maske:

$$g'(x,y) = (1/25) \cdot \sum_{u=-k\dots+k} \sum_{v=-l\dots+l} g(x-u,y-v)$$

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

- Glättungen mit anderen Masken z.B.

0	1	0
1	1	1
0	1	0

0	1	0
1	4	1
0	1	0

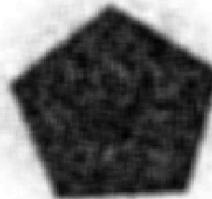
- Kontrastverschärfung z.B.

-1	-1	-1
-1	1	-1
-1	-1	-1

0	-1	0
-1	4	-1
0	-1	0

Gewichteter Mittelwertfilter, Gauß-Filter

- Mittelwertfilter, z.B. $\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$



Tiefpassfilter: Die Bilder werden weicher, „Glättung“

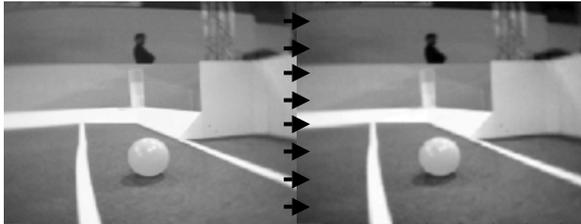
Gewichteter Mittelwertfilter, Gauß-Filter

Gauß-Filter

= Gewichteter Mittelwertfilter zur Approximation von

$$h(x,y) = 1/2\pi \exp(- (x^2+y^2)/2)$$

0.01	0.08	0.01
0.08	0.61	0.08
0.01	0.08	0.01



Gaußfilter mit
Maskengröße 3x3)

Tiefpassfilter: Die Bilder werden weicher, „Glättung“

Differenzenoperatoren

Richtungsabhängige Kantenselektierung

Richtungsabhängige Differenzenbildung: Unterschiede erkennen in
Analogie zu Differentialquotient $(g(x+1)-g(x)) / (x+1-x)$

In x-Richtung:

0	-1	1
0	-1	1
0	-1	1

bzw.:

-1	0	1
-1	0	1
-1	0	1

Sobel-Operator in x- bzw. y-Richtung

1	2	1
0	0	0
-1	-2	-1

1	0	-1
2	0	-2
1	0	-1

Weitere Operatoren z.B. für „Kompass-Richtungen“ siehe Literatur

Sobel-Operator

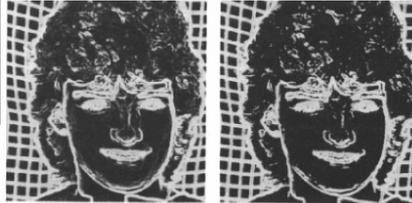
x-Richtung



y-Richtung



Gradienten-Bild
durch Kombination
der beiden Sobel-
Operatoren



Kanten:

Mit Schwellwert
aus Gradientenbild
erzeugtes Binärbild

Diskreter Laplace-Operator

Richtungsunabhängige Kantenselektierung

Richtungsunabhängige Differenzenbildung:

- Stärke von Änderungen
in Analogie zum kontinuierlichen Laplace-Operator
$$d^2g(x,y)/dx^2 + d^2g(x,y)/dy^2$$
- auch: Wendepunkte des Originals als Nullstellen,
Dabei 3. Ableitung nicht Null,
- auf Ebenen verschwindet Laplace-Operator.

Diskreter Laplace-Operator

Approximation durch diskrete 3x3-Maske z.B.

0	1	0
1	-4	1
0	1	0

Weitere Varianten:

-1	-1	-1
-1	8	-1
-1	-1	-1

1	-2	1
-2	4	-2
1	-2	1



Kombination Glättung/Kantenextraktion

Problem Kanten und Störungen sind beide hochfrequent

- Differenzbildung reagiert auf Störungen
- Rauschen variiert sogar stärker als Kantenübergang

Lösung:

- Erst Rauschunterdrückung (Tiefpass-Filter)
- dann Kantenextraktion (Hochpass-Filter)

Kombination in einem Operator:

GoL = Gauss-of-Laplacian

$$(g(x,y) * h_{\text{Gauss}}(u,v)) * h_{\text{Laplace}}(s,t)$$

$$= g(x,y) * (h_{\text{Gauss}}(u,v) * h_{\text{Laplace}}(s,t))$$

„Mexican hat“

Auch als Differenz von Gauss-Operatoren approximierbar:

DoG = Diffence of Gaussian

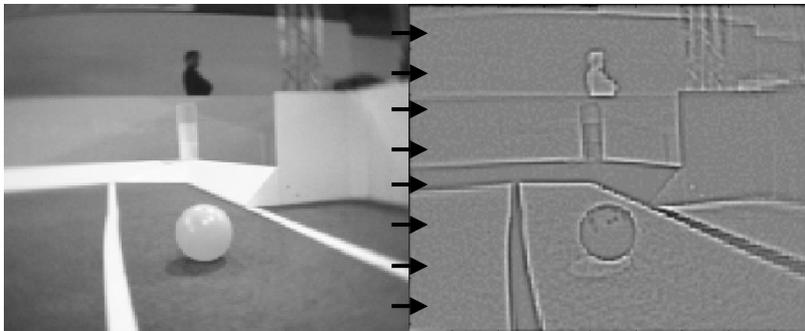
GoL = Gauss-of-Laplacian

$$\text{Gau\ss-Kernel: } K_G = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad \text{Laplace-Kernel: } K_L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Kombinierter Filterkern K :

$$K = K_G * K_L = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 \\ 1 & 0 & -2 & 0 & 1 \\ 2 & -2 & -8 & -2 & 2 \\ 1 & 0 & -2 & 0 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{bmatrix}$$

Laplacian-of-Gaussian (LoG)



(Maskengröße 5x5)

Bildverarbeitung: Ortsbereich (Forts.)

- Modifikation der Ortskoordinaten

$$x' = f_1(x, y)$$

$$y' = f_2(x, y)$$

ggf. in Verbindung mit umgebungsbezogenen Intensitätsberechnungen (Mittelung, Ausgleich)

- räumliche Skalierung (Pyramiden)
- Verschiebung, Drehung
- Entzerrung

Passpunktmethode zur Bestimmung von f_1 , f_2

Ortsfrequenzbereich

- Übergang zu Fouriertransformierten
- Bilder als überlagerte Schwingungen

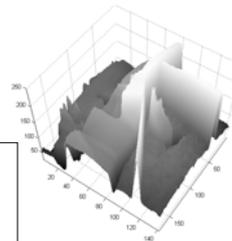
Hochfrequente Schwingungen:

- Störungen (global)
- Kanten (lokal)

Muster durch Spektren beschreiben

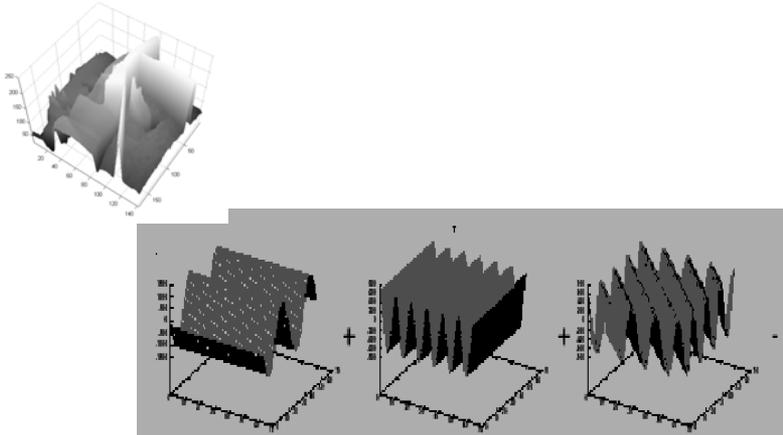
Grauwertgebirge:

Darstellung von $g(x, y)$ als Funktion



Ortsfrequenzbereich

- Bilder als überlagerte Schwingungen



Diskrete Fouriertransformation

Fouriertransformation:

Darstellung von periodischen Funktionen als unendliche trigonometrische Reihen:

$$f(x) = \sum c_k \cdot \exp(i 2\pi kx) = \sum (a_k \cdot i \sin(2\pi kx) + b_k \cdot \cos(2\pi kx))$$

Diskrete Fouriertransformation:

Approximation einer Funktion $f(x)$ im Intervall $[0, X]$ durch Abtastung an diskreten Punkten

$$x_0 = 0, x_1 = 1 \cdot X/N, x_2 = 2 \cdot X/N, \dots, x_{N-1} = (N-1) \cdot X/N$$

als endliche Summe

$$\sum_{k=0, \dots, N-1} c_k \cdot \exp(i 2\pi kx/X)$$

mit den Fourier-Koeffizienten

$$c_k = 1/N \cdot \sum_{k=0, \dots, N-1} f(x_k) \cdot \exp(-i 2\pi kx/X)$$

DiskreteFouriertransformation

Diskrete Fouriertransformation eines Bildes

$$g(x,y) , \quad x=0,\dots,L-1 \quad y=0,\dots,M-1$$

analog durch Abtastung an den diskreten Bild-Punkten

$$x_u = u/L , y_v = v/M \quad \text{für } u=0,\dots,L-1 \quad v=0,\dots,M-1$$

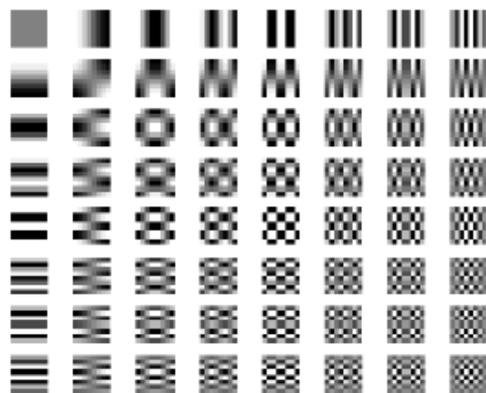
als endliche Summe

$$g(x,y) = \sum_{u=0,\dots,L-1} \sum_{v=0,\dots,M-1} G(u,v) \cdot \exp(i 2\pi (xu/L + yv/M))$$

mit den Fourier-Koeffizienten

$$G(u,v) = 1/LM \cdot \sum_{x=0,\dots,L-1} \sum_{y=0,\dots,M-1} g(x,y) \cdot \exp(-i 2\pi(xu/L + yv/M))$$

DiskreteFouriertransformation



Beispiel für Basisfunktionen
(Cosinus, Zellengröße N=8)

Fouriertransformiertes Bild

$G(u,v)$ für $u=0,\dots,L-1$ $v=0,\dots,M-1$

hat wie das ursprüngliche Bild $g(x,y)$ die Form einer Bildmatrix der Dimension $L \times M$.

Die $G(u,v)$ heißen **Ortsfrequenzen**.

$G(u,v) = 1/LM \cdot \sum_{x=0,\dots,L-1} \sum_{y=0,\dots,M-1} g(x,y) \cdot \exp(-i 2\pi(xu/L + yv/M))$
heißt **Fouriertransformierte des Bildes** $g(x,y)$,

Umgekehrt ist

$g(x,y) = \sum_{u=0,\dots,L-1} \sum_{v=0,\dots,M-1} G(u,v) \cdot \exp(i 2\pi(xu/L + yv/M))$
die inverse **Fouriertransformierte von** $G(u,v)$.

Spektrum der Fouriertransformierten

Die Fouriertransformierte ist komplex:

$$G(u,v) = G_1(u,v) + i G_2(u,v)$$

$G(u,v)$

$$= 1/LM \cdot \sum_{x=0,\dots,L-1} \sum_{y=0,\dots,M-1} g(x,y) \cdot \exp(-i 2\pi(xu/L + yv/M))$$

$$= 1/LM \cdot \sum_{x=0,\dots,L-1} \sum_{y=0,\dots,M-1} g(x,y) (-i \sin(2\pi(xu/L + yv/M)) + \cos(2\pi(xu/L + yv/M)))$$

$$= -i \cdot 1/LM \cdot \sum_{x=0,\dots,L-1} \sum_{y=0,\dots,M-1} g(x,y) \sin(2\pi(xu/L + yv/M)) + 1/LM \cdot \sum_{x=0,\dots,L-1} \sum_{y=0,\dots,M-1} g(x,y) \cos(2\pi(xu/L + yv/M))$$

Der Betrag $|G(u,v)| = \sqrt{G_1^2(u,v) + G_2^2(u,v)}$

heißt **Fourier-Spektrum**

Spektrum der Fouriertransformierten

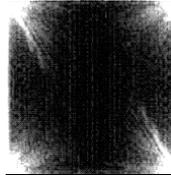
Das Fourier-Spektrum $|G(u,v)| = \sqrt{G_1^2(u,v) + G_2^2(u,v)}$ beschreibt den Anteil der Frequenzen im Bild.

Das Fourier-Spektrum ist als Grauwert-Bild darstellbar, dazu erfolgt meistens Normierung:

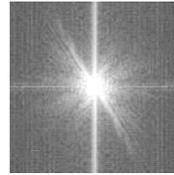
- Zentralsymmetrische Darstellung, geringe Frequenzen im Zentrum
- Logarithmische Skalierung ($\log(1+|F(u,v)|)$)



Original



$|G(u,v)|$ ohne Normierung



mit Normierung

Operationen im Ortsfrequenzraum

Direkte Manipulation von Frequenzen.

Beispiele:

- Tiefpassfilter (für Rauschunterdrückung)
 - Elimination der hohen Frequenzen
- Hochpassfilter (für Kantenextraktion)
 - Elimination der tiefen Frequenzen
- Bandpassfilter
 - speziellen Frequenzbereich erhalten

Operationen im Ortsfrequenzraum

Zusammenhang zum Ortsbereich

Faltung im Ortsbereich

= Multiplikation im Frequenzbereich

Bild: $g(x,y) \Rightarrow G(u,v)$

Maske: $h(x,y) \Rightarrow H(u,v)$

Faltung: $g(x,y) * h(x,y) \Rightarrow G(u,v) \cdot H(u,v)$

Operationen im Ortsfrequenzraum

- Anwendung von Filterverfahren im Frequenzraum
 - Berechnung der Fouriertransformierten $G(u,v)$
 - lokale Multiplikation von $G(u,v)$ mit einem Filter $H(u,v)$
$$G'(u,v) := G(u,v) \cdot H(u,v)$$
 - Berechnung der Inversionen Fouriertransformierten von $G'(u,v)$
- Analyse von Frequenzen
- Bildvergleich durch Vergleich des Spektrums
- Analyse/Vergleich von Texturen.

Effiziente Verfahren:

Schnelle Fouriertransformation (Symmetrien nutzen)

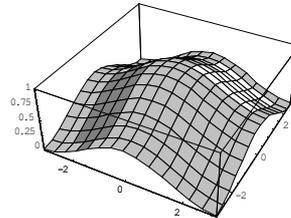
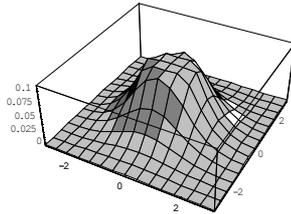
Mittelwert vs. Gauss

Mittelwert-Filter (Resultat):

$$g'(x,y) = 1/8 (4g(x,y) + g(x-1,y) + g(x,y-1) + g(x,y+1) + g(x+1,y))$$

Spektrum:

$$H(u,v) = 1/8 (4 + 2 \cos(u) + 2 \cos(v))$$



Gauss-Filter (kontinuierlicher Kern):

$$h(x,y) = 1/2\pi \exp(- (x^2+y^2)/2)$$

(Spektrum ebenfalls eine Gauss-Funktion)

Bildanalyse

- Identifizierung von Komponenten (Kanten, Flächen)
- Identifizierung von Objekten
- Beziehungen zwischen Objekten
- Interpretation der Szene
- ggf. Interpretation von Bildfolgen (Abläufe)

Schichtenmodelle

- streng sequentiell

- oder Hypothesenbasiert

(mit Backtracking, z.B. Blackboard-Architektur)

Schichtenmodell

- Vorverarbeitung (Aufbereitung, Anpassung)
 - Digitalisieren
 - Entzerren (Sensormodell, Kalibrieren)
 - Verbessern (ohne spezielles Störmodell):
Filteroperationen zur Glättung bzw. Verschärfung,
evtl. Ausführung im Ortsfrequenzraum (Fouriertransf.)
 - Restaurieren (mit bekanntem Störmodell)
z.B. Bewegungsunschärfe, atmosph. Störungen
 - Rekonstruktion
 - Normierung (Größe, Lage, Grauwerte, Farben)

Schichtenmodell

- **Segmentierung:**
Klassifizierung von Pixeln
Identifizierung von Bildprimitiven:
 - Kanten aus Kantenelementen
(„edge“ = edge element, Pixel auf einer Kante)
 - Bereiche (abgegrenzte Pixelmengen)
bestimmt durch Grenzlinie und/oder gemeinsame Textur
(„texel“ = texture element, Pixelmenge mit spezieller Textur,
d.h. spezieller Grauwertstruktur)Zerlegung des Bildes in Bereiche
- **Komponentenerkennung**
Kombination von Bildprimitiven zu Komponenten

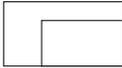
Kanten, Linien



- Grundlegend für menschliche Orientierung
- Begrenzung von Körpern bzw. Flächen
- Sprung der Intensität (Begrenzung)
- Maximum bzw. Minimum der Intensität (Linie)



Kanten, Linien

- Kante: Übergang zwischen unterschiedlichen Bereichen
 - Grauwertkante
großer Gradient in $g(x,y)$ 
 - Texturkante
durch Vorverarbeitung in Grauwertkante überführen 
 - Farbkante
Grauwertkanten in Farbkanälen 
- Linie = 2 naheliegende, parallele Kanten 

Differenzoperatoren
erzeugen Linien aus
Kanten



Kantenerkennung

Ausgangssituation: Bild $g = [g(x,y)]_{x,y}$

Ergebnis: Kanten

Repräsentationsformen für Kanten:

- Quantitativ, z.B.
 - Mengen von Punkten:
Koordinaten $p=(x,y)$ bzw. Koordinaten mit Richtung $p=(x,y,r)$
 - Kurvengleichung (Gerade, Kreis, ...)
 - Gerade mit Anfangspunkt, Richtung, Länge
- Qualitativ, z.B.
 - „großer Kreis am linken Bildrand“

Kantenerkennung

Mehrere Schritte, insbesondere

- 1) Identifizierung relevanter Bild-Punkte:
„Edgels“ = edge element: Punkt (+Richtung)
 - Merkmal z.B. Grauwert-Änderung im Originalbild
 - abgeleitete Merkmale nach Filterung
- 2) Zusammenfassung von Edgels zu Kanten
 - z.B. Hough-Transformation, Regression, Suchverfahren,
Korrelation mit Kanten-Maske

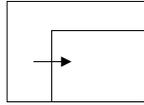
Zur Vorbereitung /als Zwischenschritte ggf. weitere Filteroperationen

- z.B. Rauschunterdrückung, Binärbild (mittels Schwellwert)
Kantenverstärkung, Vervollständigung

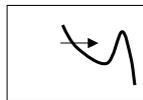
Kantenextraktion

„Kurvendiskussion“

Kante = Sprungstelle der Grauwertverteilung



Linie = Maximum (Minimum) in der Grauwertverteilung
bzw. zwei Sprungstellen der Grauwertverteilung



Kantenextraktion: Edgel

– Gradient: Betrag, Richtung

- Lokale Differenz-Operatoren
- Sobel-Operator



Sobel-Operator in x- bzw. y-Richtung

1	2	1
0	0	0
-1	-2	-1

1	0	-1
2	0	-2
1	0	-1

Kantenextraktion: Edgel

– Wendepunkt

Nullstelle der 2. Ableitung

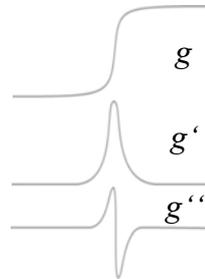
$$\nabla^2 g(x, y) = \frac{\partial^2 g(x, y)}{\partial x^2} + \frac{\partial^2 g(x, y)}{\partial y^2} = 0$$

- Diskreter Laplace-Operatoren z.B.

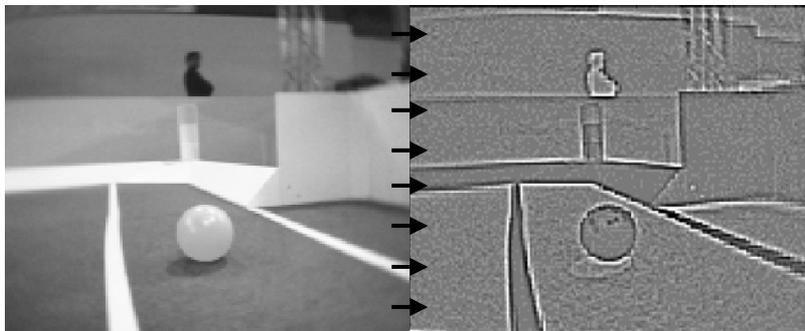
0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

1	-2	1
-2	4	-2
1	-2	1



Kantendetektion mit Laplace Operator



Kantenextraktion: Edgel

– Hochpass-Filter im Frequenzbereich

- Berechnung der Fouriertransformierten $G(u,v)$
- lokale Multiplikation von $G(u,v)$ mit einem Hochpass-Filter $H(u,v)$ zur Elimination der tiefen Frequenzen:

$$G'(u,v) := G(u,v) \cdot H(u,v)$$

- Berechnung der Inversionen Fouriertransformierten von $G'(u,v)$

Kantenextraktion: Edgel

- Morphologische Kantenoperatoren

Mit vorher geglättetem $g(x,y)$ und $U = 3 \times 3$ -Umgebung durch

• „Erosion“ $g_1(x,p) = \text{Min}_{(i,j) \in U(x,y)} \{g(i,j)\}$ (verkleinert)

oder

• „Dilatation“ $g_2(x,p) = \text{Max}_{(i,j) \in U(x,y)} \{g(i,j)\}$ (erweitert)

anschließende Differenzenbildung

$$g(x,p) - \text{Min}_{(i,j) \in U(x,y)} \{g(i,j)\}$$

oder

$$\text{Max}_{(i,j) \in U(x,y)} \{g(i,j)\} - g(x,p)$$

ergibt jeweils ein (evtl. versetztes) Kantenbild

(speziell geeignet für Binärbilder, Probleme evtl. bei Linien)

Kantenextraktion: Edgel

- Morphologische Kantenoperatoren

Kombination der beiden Verfahren (auch für Linien)

$$g^*(x,y) := \text{Min} \{ (g(x,p) - \text{Min}_{(i,j) \in U(x,y)} \{g(i,j)\}) , (\text{Max}_{(i,j) \in U(x,y)} \{g(i,j)\} - g(x,p)) \}$$

mit vorher geglättetem $g(x,y)$ und $U = 3 \times 3$ -Umgebung

als Kombination von

- „Erosion“ $g_1(x,p) = \text{Min}_{(i,j) \in U(x,y)} \{g(i,j)\}$ (verkleinert)
- „Dilatation“ $g_2(x,p) = \text{Max}_{(i,j) \in U(x,y)} \{g(i,j)\}$ (erweitert)
- Differenzenbildungen

$$g'(x,y) := g(x,p) - \text{Min}_{(i,j) \in U(x,y)} \{g(i,j)\} \quad g''(x,y) := \text{Max}_{(i,j) \in U(x,y)} \{g(i,j)\} - g(x,p)$$

- Minimum (scharfe Kanten) $g^*(x,y) := \text{Min} \{g'(x,y), g''(x,y)\}$

Realisierung durch zwei 3×3 -Masken (Rangordnungsoperatoren)

oder kombinierte 5×5 -Maske

Kantenfindung aus Edgels

- Schwellwertbildung: Binärbild der Edgels
 - Störungen beheben
 - „Verdünnen“
- Kombinationsverfahren:
 - Hough-Transformation
 - Regression
 - Suchverfahren (z.B. Bergsteigen, A^*)
 - Template-Vergleich (Korrelation)

Hough-Transformation

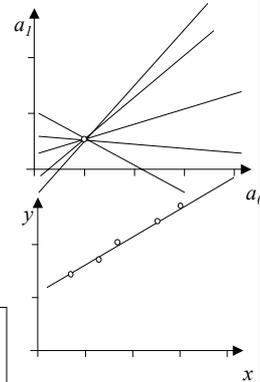
- Ausgangspunkt: Menge von Edgels
- Resultat: Menge von Geraden

Idee:

- Jedes Edgel kann auf unterschiedlichen Geraden liegen
- Voting-Verfahren: Edgel „stimmen ab“ für „ihre Geraden“
- Die Geraden mit den meisten Stimmen (Edgels) gewinnen

Vorteile:

- Gleichzeitig mehrere Kanten erkennbar
- Robust bzgl. fehlender/zusätzlicher Edgels

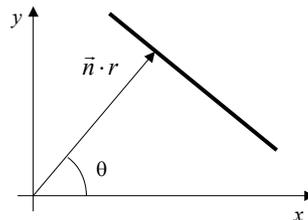


Hough Transformation

Darstellung der Geraden als Hesse'sche Normalform

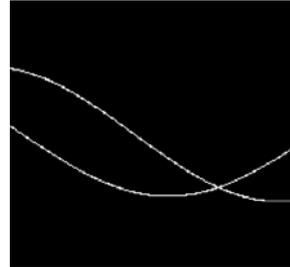
$$\vec{n} \cdot \vec{x} = r$$

$$x \cos \theta + y \sin \theta = r$$



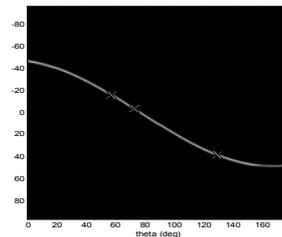
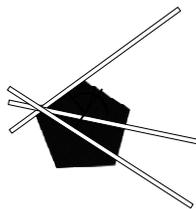
Hough Transformation

Kurven $\kappa_{(x,y)}$ von Punkten (x,y) auf $r = x \cos \phi + y \sin \phi$ im x - y -Raum schneiden sich im Punkt (r, ϕ) im r - ϕ -Raum



Hough Transformation

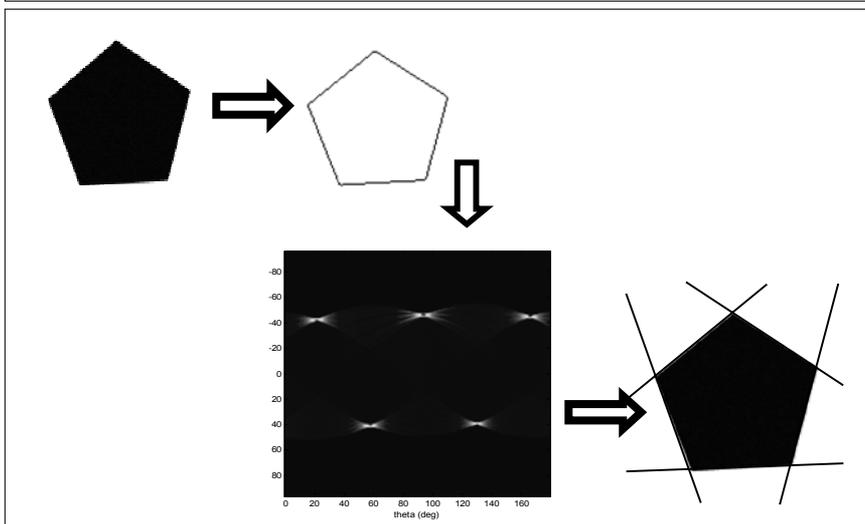
Geraden $r = x \cos \phi + y \sin \phi$ im x - y -Raum
werden dargestellt als Punkte (r, ϕ) im r - ϕ -Raum.
Geradenbündel durch (x,y) im x - y -Raum
werden dargestellt als Kurven $\kappa_{(x,y)}$ im r - ϕ -Raum



Hough-Transformation

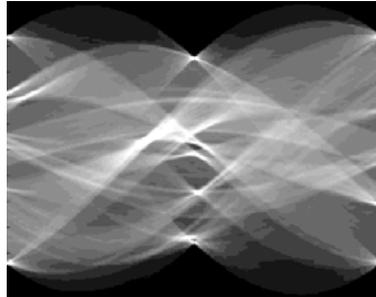
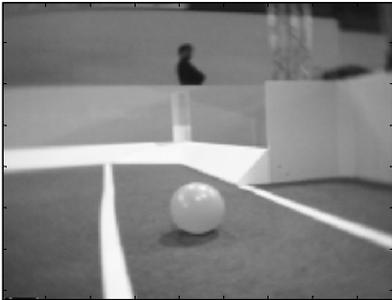
- Praktische Umsetzung:
 - Akkumulator-Grid im r - ϕ -Raum füllen („Voting“).
 - Für jedes Edgel entsprechende Einträge für Kandidaten (= Gerade (r, ϕ)) füllen.
 - Ggf. nur entsprechend Richtung der Edgels
 - Geraden mit hinreichend vielen Stimmen gelten als erkannt.

Hough Transformation



Hough Transformation

Beispiel:



Hough-Transformation

- Erweiterbar für andere Konturen (Kreise, ...)

- Kreise:

- 3-dimensionales Grid für Kreisparameter a , b , r

$$(x-a)^2 + (y-b)^2 = r^2$$

- Rechnungen entsprechend aufwändiger

Flächen

- Merkmale:
 - Textur, Farbe, Begrenzung
- Merkmale bestimmen:
 - Farbesegmentierung
 - Texturanalyse (Frequenzvergleich)
 - Kombination von Kanälen (Summe, Differenz)
- Erkennung
 - Segmentierung anhand von Merkmalen
 - Suchverfahren: Gebiete erweitern bzw. einschränken

Objektklassifizierung

- Klassifizierung anhand charakteristischer Merkmale
 - Kantenformen
 - Verzweigungen, Ecken
 - Flächenmerkmale (Farben, Textur), . . .und ihrer Beziehungen (z.B. als Graph)
- Constraintverfahren nach Waltz
- Klassifizierung in komprimierten Darstellungen
 - RLE, Nachbarschaftsgraphen
 - Quad-Trees
 - Richtungsketten, . . .

Objektklassifikation

- Vergleichsoperationen, Ähnlichkeit (Bildobjekt, Referenzobjekt)
 - lokaler Vergleich
 - Differenzbild (einfaches "Fehlerbild")
 - Korrelation (Summe quadratischer Fehler minimieren)
 - Filterung mit „Suchmaske“ (= Referenzobjekt)
 - Vergleich globaler Merkmale
 - Histogramme
 - Mittelwert, Varianz, Kovarianz, ...
 - Entropie

Ähnlichkeit

Vergleich auf unterschiedlichen
Verarbeitungsstufen möglich, z.B. auf

- Pixelebene
- struktureller Ebene (Beziehungen von Linien)

A

Dieser Text
enthält ein "A".

Dieser Text
enthält ein "A".

Ähnlichkeit

Vergleich auf Pixelebene (auch: Differenzbild) erfordert

- Angleichung von Position und Größe
 - Translation
 - Rotation
 - Skalierung
- Angleichung der Darstellung
 - Grauwerte/Farben
 - Ergänzung (verdeckter/unsichtbarer Teile)
 - Abstraktion (Verzicht auf individuelle Eigenschaften)
 - Rauschunterdrückung

Ähnlichkeit

Vergleich von allgemeinen Merkmalen

- Statistische Merkmale (Bilder als Zufallsvariable)
 - Mittelwert (durchschnittlicher Grauwert)
 - Varianz (Kontrast)
 - Kovarianz bei mehrkanaligen Bildern
- Histogramme
 - abgeleitete Werte
- Entropie

Ähnlichkeit

Vergleich von strukturellen Merkmalen

- Bildprimitive, Komponenten
Kanten, Ecken, Oberflächen
- Struktur:
Beziehungen von Bildprimitiven/Komponenten
z.B. Constraint-Verfahren von Waltz
- Beziehungen zu anderen Bildobjekten

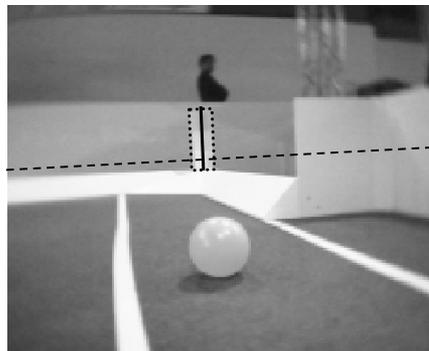
Ähnlichkeit: strukturelle Merkmale

Beispiel:

Erkennung der Flaggen

Bekannte Merkmale:

- Kanten: weiss-gelb und gelb-pink
- Lage im Bild relativ zum Horizont



Außenkanten als Maß für die Größe einer Landmarke

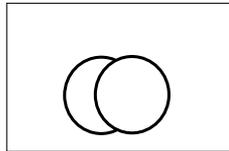
Bildfolgen verstehen

- **Zusätzliche Ziele:**
 - Abläufe interpretieren
 - Eigene Bewegung kontrollieren
- **Zusätzliche Möglichkeiten:**
 - Bezugnahme auf weitere Bilder
 - Bestätigung von Annahmen
 - Erkennen von Veränderungen

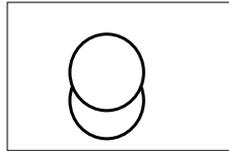
Bewegungen

- **Zeit als weitere Dimension**
 - Aperturproblem:
Bewegung nur senkrecht zu Kanten erkennbar
 - Methoden:
 - Differenzbild
 - Berechnung im Raum-Modell
 - Eigenbewegungen berücksichtigen
- **Aktives Sehen**
 - Hinsehen (mechanisch gesteuert)
 - Auswählen im Bild bzw. in Bilderfolgen
- **Stereosehen**

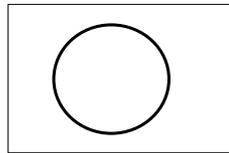
Bewegungen



nach rechts



nach
oben



nach
vorn



Bewegungen

Allgemein: Über Differenzbilder auswerten.

Anforderungen:

- Eigene Bewegung bestimmen
- Bewegung von Objekten bestimmen

Zusammenfassung

- Visuelle Information ist hochkomplex
- Rekonstruktion räumlicher Verhältnisse ist ein „schlecht gestelltes Problem“
- Objekte sind durch vielfältige Merkmale charakterisierbar
- Objekterkennung durch Vergleich von Merkmalen
- Merkmalsextraktion durch Bildverarbeitungsverfahren
- Filter/Masken im Ortsraum bzw. Frequenzraum

Entwurfsentscheidungen bzgl. Interpretation

- Ausgangsdaten
 - Format der Bilddaten
 - Kontext der Bildentstehung (Beleuchtung, Abbildungsmodell, Sensoreigenschaften, ...)
- Repräsentation von (Zwischen-)Modellen
 - Prototypen von Objekten und Eigenschaften
- Repräsentation der Zielbeschreibung
 - Formalismen der Wissensrepräsentation für Objekte mit Eigenschaften und Beziehungen