

# Kognitive Systeme

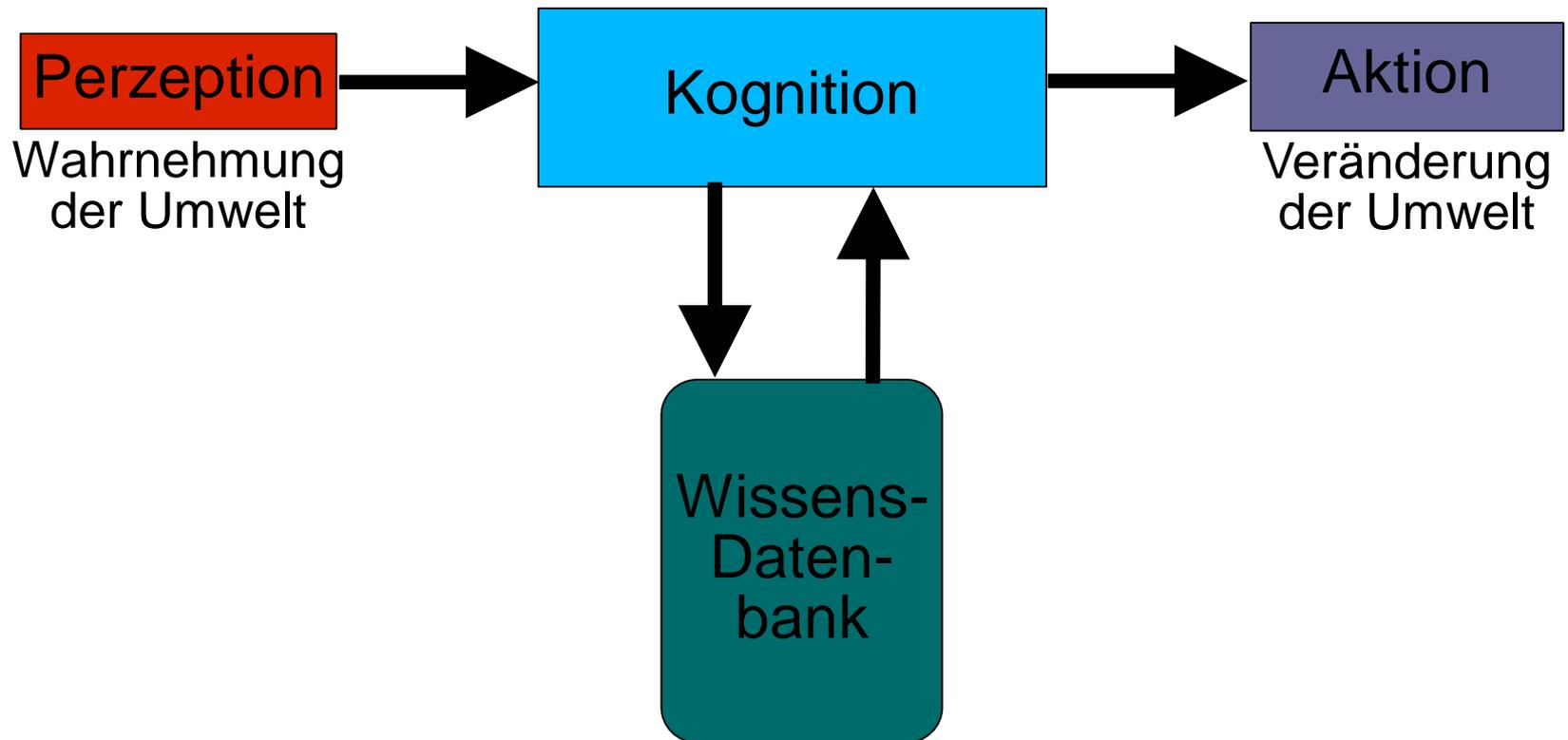
## 2D-Bildverarbeitung

Montag, 26. Juni 2017

Zunächst: kurze Wiederholung

## Elemente eines Kognitiven Systems

Ingenieurmäßiger Ansatz:



- Modelliert den visuellen Sinn des Menschen
- Basiert auf Halbleiter-Kameras und Laser-Scanner
- Erlaubt reichhaltige sensorische Information
  - Farbe
  - Merkmale
  - Strukturen
  - Relationen
  - Textur
  - Bewegungen, Handlungen
- Für den Menschen einfach nachvollziehbar

# Sensorische Erfassung: Szene

## Aufgaben der sensorischen Umwelterfassung:

- Wiedererkennung bekannter Sachverhalte
  - Objekte und Relationen
  - Personen
  - Orte von Objekten und Szenen
- Erlernen neuer Sachverhalte
- Erkennung der eigenen Bewegung

## Verfahren zur Lösung dieser Aufgaben:

- Sensorische Primitive, Segmentierung
- Annahmen, Einschränkungen
- Lernverfahren

# Sensorische Erfassung: Situation

- Erkennung von Handlungen und Wechselwirkungen
- Beobachtung von dynamischen Szenarien
- Erkennung von Bewegungsabläufen
- Interaktionen zwischen Mensch und Maschine
- Zeithorizont für dynamische Abäufe
- Einbindung von Kontext und Hintergrundwissen
- Lernverfahren

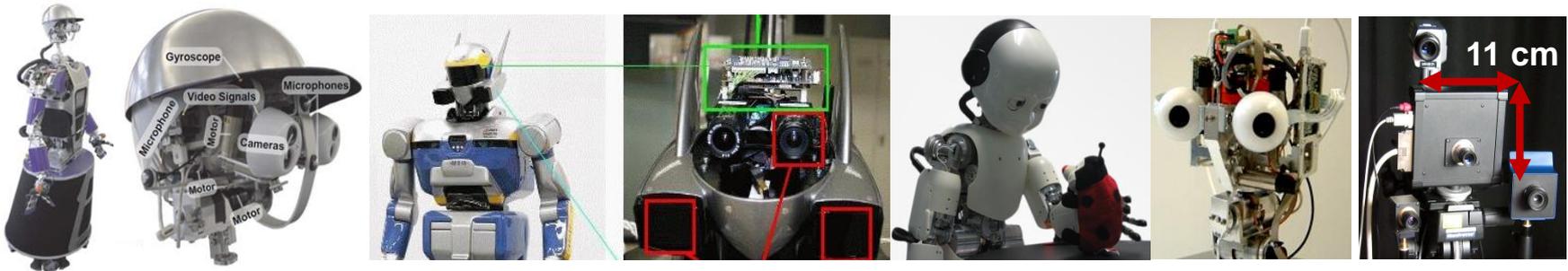
# Zustandsprädiktion auf Basis von Beobachtungen

- Kalman-Filter
- Bayes'sche Schätzer
- POMDPs
- Lernverfahren

# Visual Sensing Problems

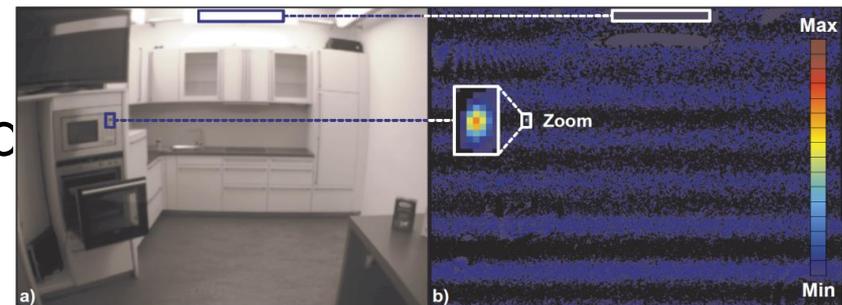
## ■ Humanoid robot embodiment restrictions

- Limited camera size
- Complex perturbations



## ■ Environmental adverse conditions

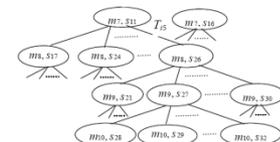
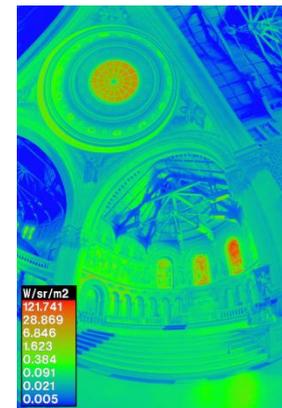
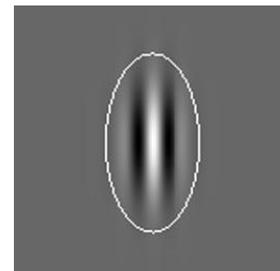
- Lighting flickering
- Dynamic day lighting



## ■ Complex materials/structures

# Critical Path for Cognitive Vision

- Humanoid robots
  - Self-localization
    - Thompson et al. 2005
    - Stasse et al. 2006
  - Environmental object recognition
    - Okada et al. 2008
    - Yokoi et al. 2011
- Computer vision
  - Radiometric calibration
    - Debevec et al. 2001
    - Krawczyk et al. 2007
  - Edge extraction
    - Grigorescu et al. 2003



# HDR: Image Capturing

- Avoid over- and under-exposure artifacts
- Overcome limitations of analog-to-digital converter



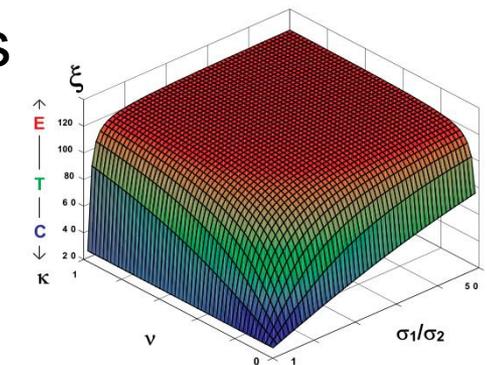
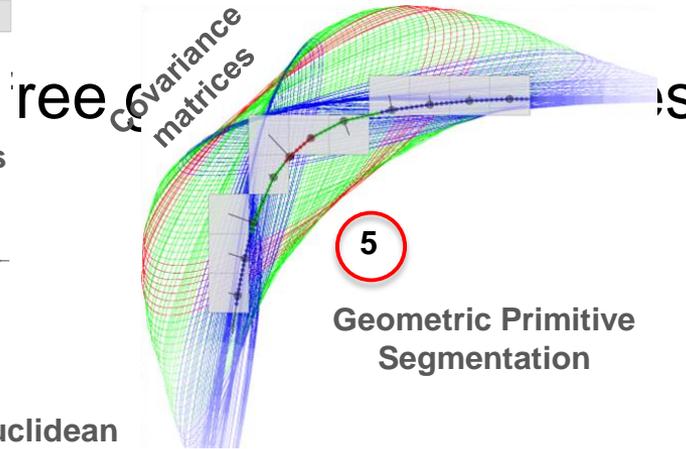
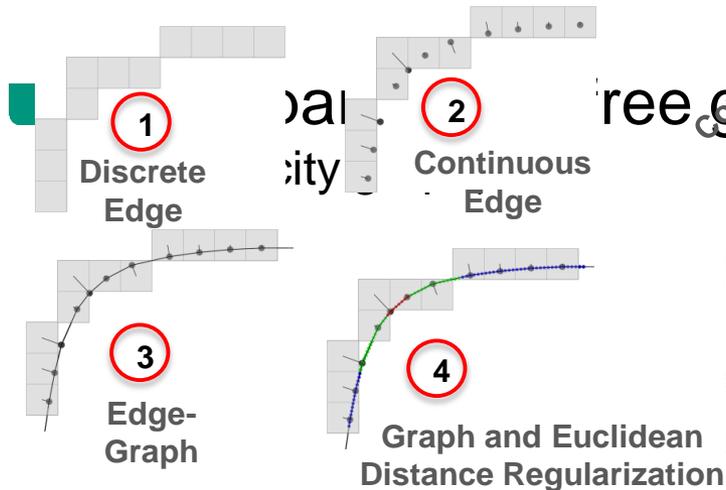
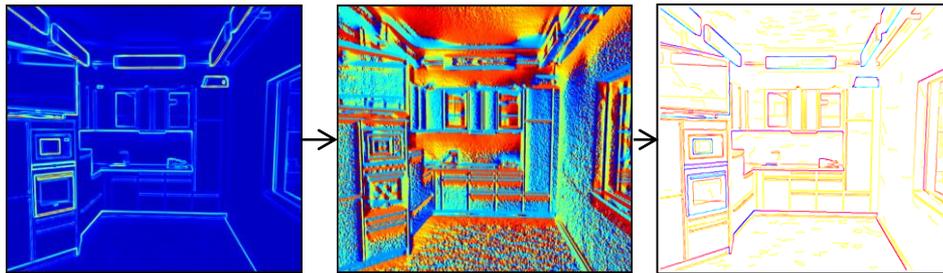
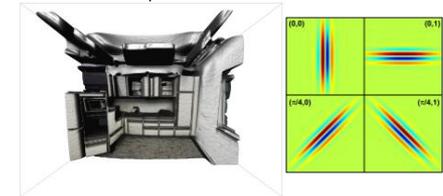
# Visual Feature Extraction

## Edge extraction [Grigorescu et al. 2006]

$$\Gamma(\mathbf{x}) = \sum_{p=0}^n \sum_{q=0}^1 U_{\langle p\pi/2n \rangle} \left\{ \Phi(\mathbf{x}) \otimes \Psi(p\pi/2n, q) \right\}$$

- Gabor kernels

- Structural selection instead of thresholds



Eccentricity Segmentation Model

# Principle: Programming by Demonstration

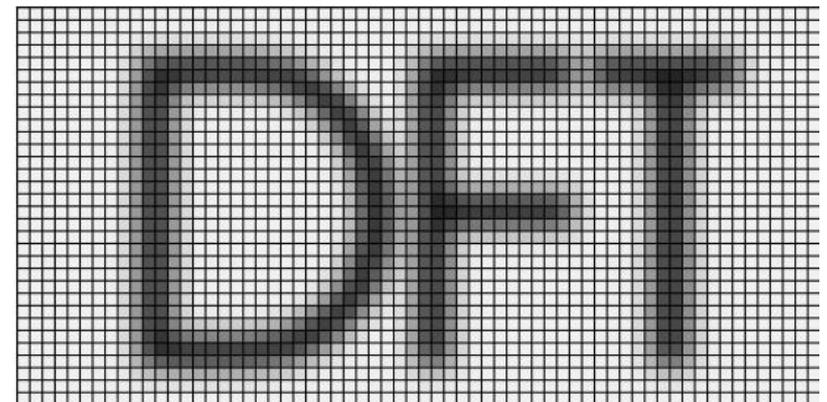
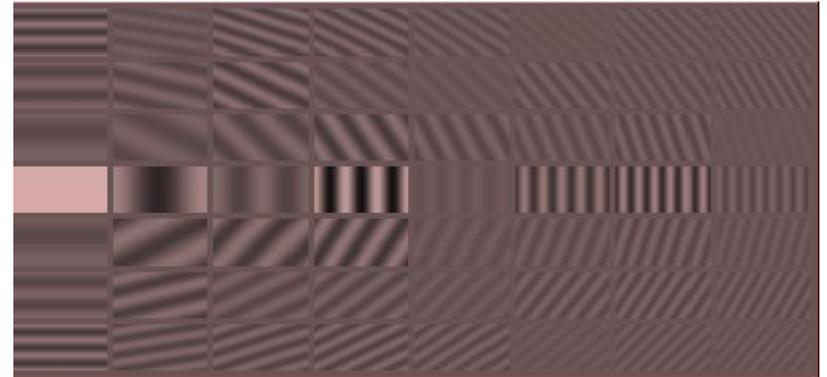
- A human demonstrates a task naturally, incorporating all implicit and explicit domain knowledge
- The demonstration is observed by sensors and robot task knowledge is generated from that data



Specific PbD  
methodology



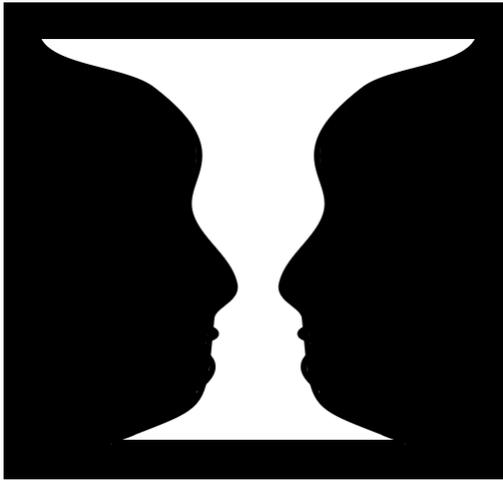
- (Grauwert-)Bilder lassen sich signaltheoretisch als Summe verschiedenfrequenter Signale betrachten
- Niedrige Frequenzen:  
Schwache  
Grauwertübergänge
- Hohe Frequenzen:  
Scharfe Grauwertübergänge



# Inhalt der heutigen Vorlesung

- Segmentierung von Bilddaten
  - Schwellwertsegmentierung
  - Segmentierung von Farbflächen
  - Segmentierung von Kanten
  - Segmentierung auf der Basis von Punktmerkmalen
- Geometrische 2D-Transformationen

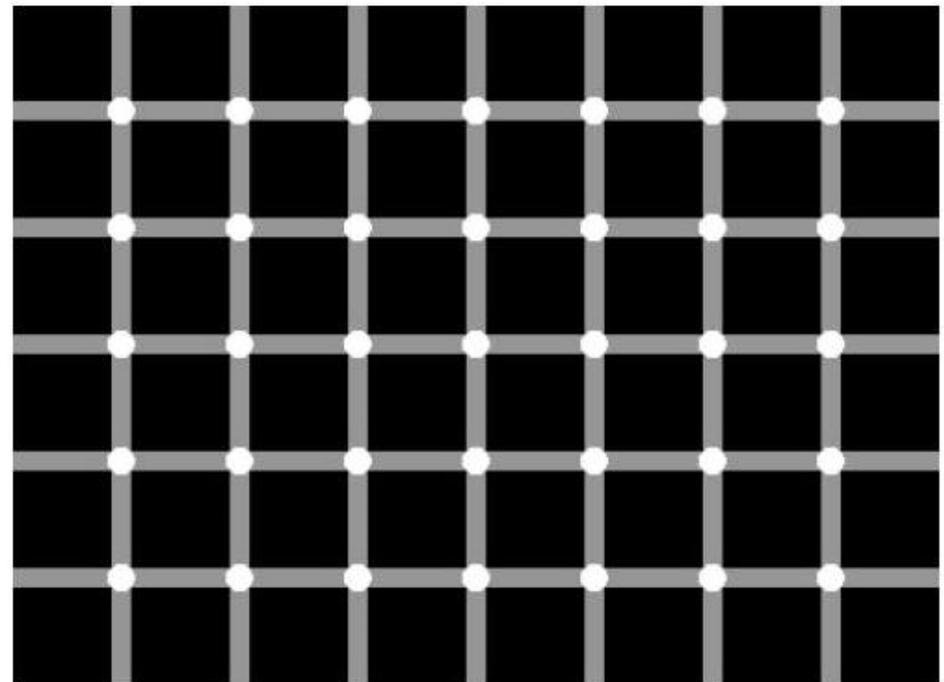
# Segmentierung



Menschliche Wahrnehmung



Erkennung durch den Computer

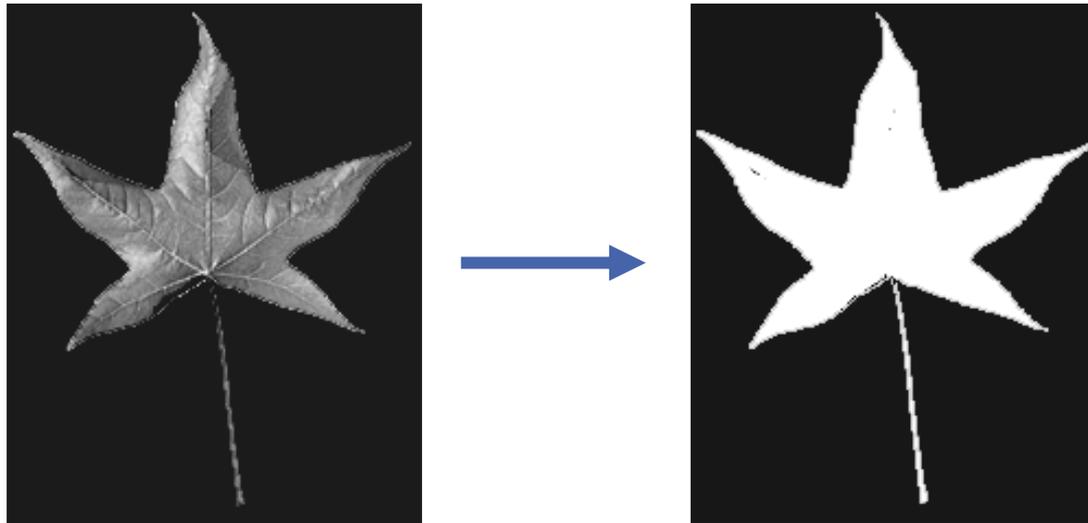


- Segmentierung ist die Aufteilung eines digitalen Bildes in aussagekräftige Bestandteile
- Gesucht:
  - Aussagen über das Bild
  - Reduktion der Datenmenge
  - Verfolgung von Merkmalen über
    - die Zeit
    - bei mehrere Sensoren
- Üblich sind unter Anderem:
  - Kanten, Ecken, Löcher, Formelemente
  - Textur, Farbe, Beschriftung

# Segmentierung: Schwellwertfilterung

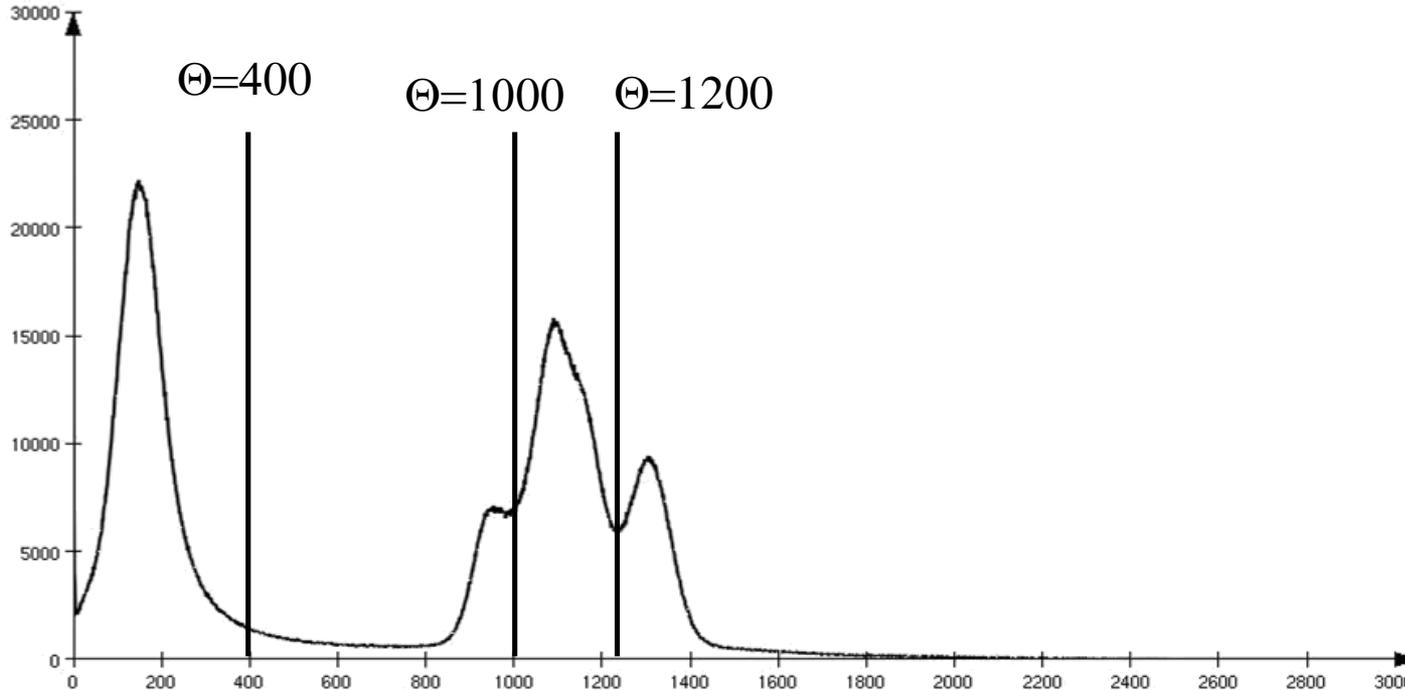
- Schwellwertfilterung zur Konvertierung eines Grauwertbildes in ein binäres Bild
- Ziel:  
Trennung interessanter Objekte vom Hintergrund

$$\text{Img}'(u,v) = \begin{cases} q & \text{falls } \text{Img}(u,v) \geq T \\ 0 & \text{sonst} \end{cases}$$



# Schwellwertoperationen

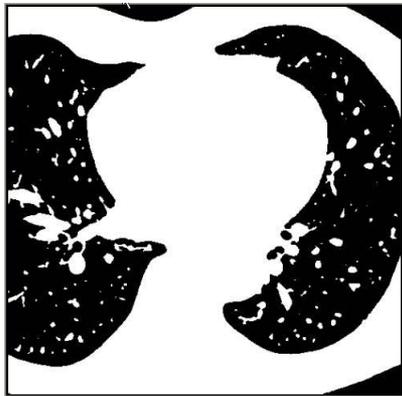
- Bestimmung des Schwellwertes
  - Von Hand nach Erfahrung
  - Betrachtung des Histogramms



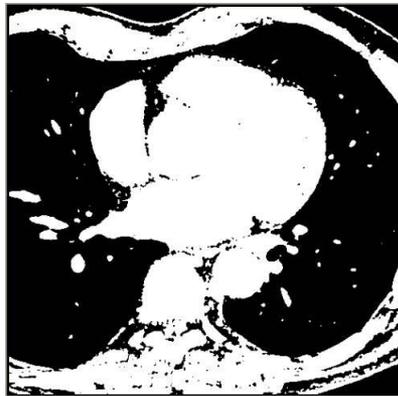
# Schwellwertoperationen



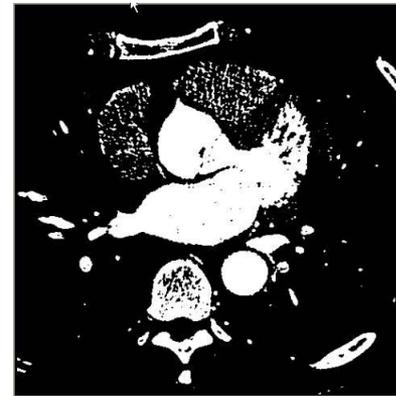
$\Theta=400$



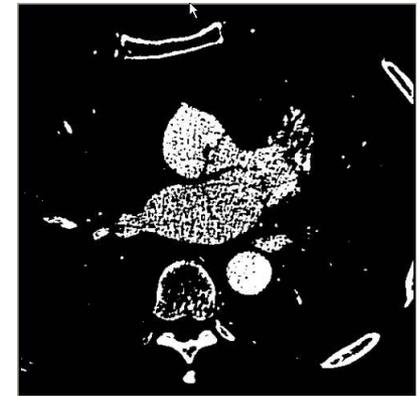
$\Theta=1000$



$\Theta=1200$



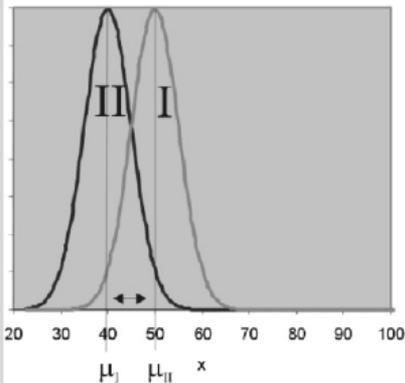
$\Theta=1300$



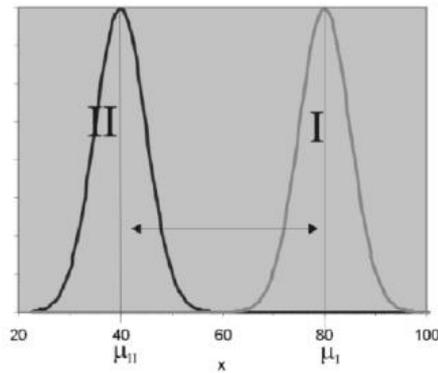
- Automatisches multimodales Schwellwertverfahren
- Aufgrund der Grauwertverteilung findet das Otsu Verfahren die optimalen Schwellwerte zur Klassifizierung aller Bildpunkte
- Bei einer Vorgabe von  $k$  Schwellwerten wird das Bild in  $k+1$  Klassen unterteilt
- Unterteilung erfolgt mit Hilfe des Diskriminanzkriteriums

# Multilevel Otsu Verfahren

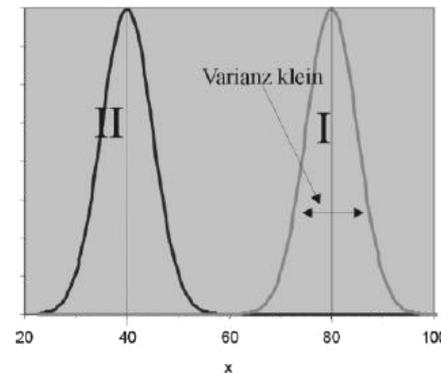
- Interklassenvarianz  $\sigma_b^2$  : Maß für Streuung zwischen den Klassen
- Intra-Klassenvarianz  $\sigma_w^2$  : Maß für Streuung in den Klassen
- Für eine möglichst gute Trennung sollte die Interklassenvarianz möglichst groß und die Intra-Klassenvarianz möglichst klein sein.



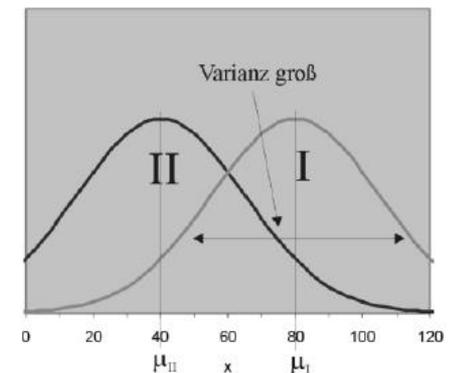
Kleine  
Interklassenvarianz



Große  
Interklassenvarianz



Kleine  
Intra-Klassenvarianz



Große  
Intra-Klassenvarianz

Quelle: Wikipedia

- Diskriminanzkriterium:

$$\lambda(t_1, \dots, t_k) = \frac{\sigma_b^2(t_1, \dots, t_k)}{\sigma_w^2(t_1, \dots, t_k)}$$

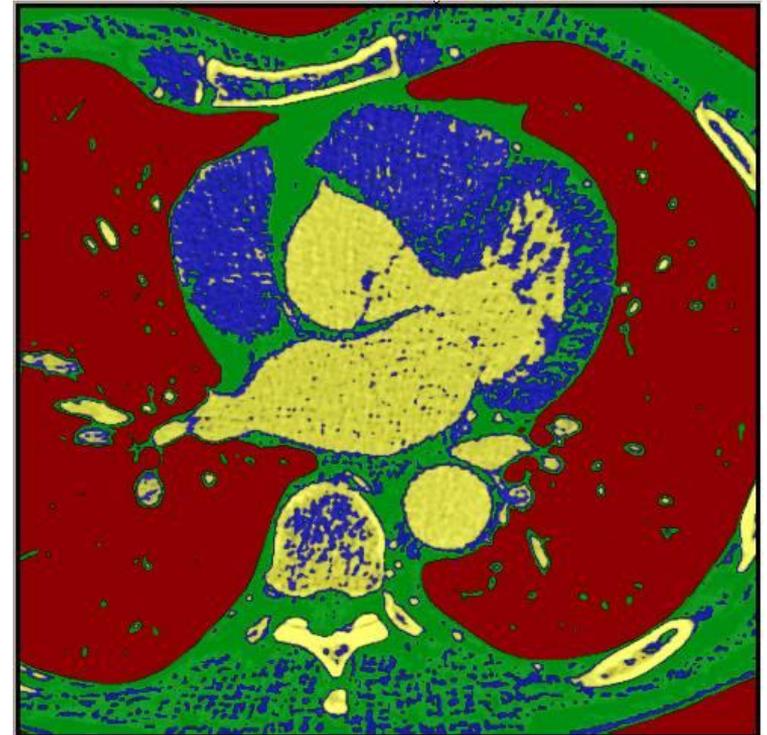
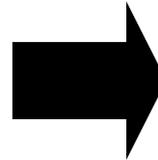
- Normiert:

$$\eta(t_1, \dots, t_k) = \frac{\sigma_b^2(t_1, \dots, t_k)}{\sigma^2}$$

- Da  $\sigma^2$  unabhängig von den Schwellwerten ist, reicht es  $\sigma_b^2$  zu maximieren
- Das Multilevel-Otsu Verfahren berechnet die k optimalen Schwellwerte, die die Interklassenvarianz maximieren

# Multilevel Otsu Verfahren

- Ergebnis bei Einteilung in vier Klassen



# Segmentierung: Farbe

- Oft können Objekte über ihre Farbe segmentiert werden:
  - Menschliche Haut
  - Einheitlich gefärbte Objekte
- Problem:
  - Wechselnde Lichtbedingungen
  - Reflexionen, Schattenwürfe
- Verfahren:
  - Histogrammbasiert (z.B. in RGB, HSV bzw. RG, HS)
  - Mit Hilfe der Mahalanobis-Distanz (z.B. in RGB)
  - Klassifikation durch Verwendung von Neuronalen Netzen
  - Durch Intervallschranken im HSV-Farbraum

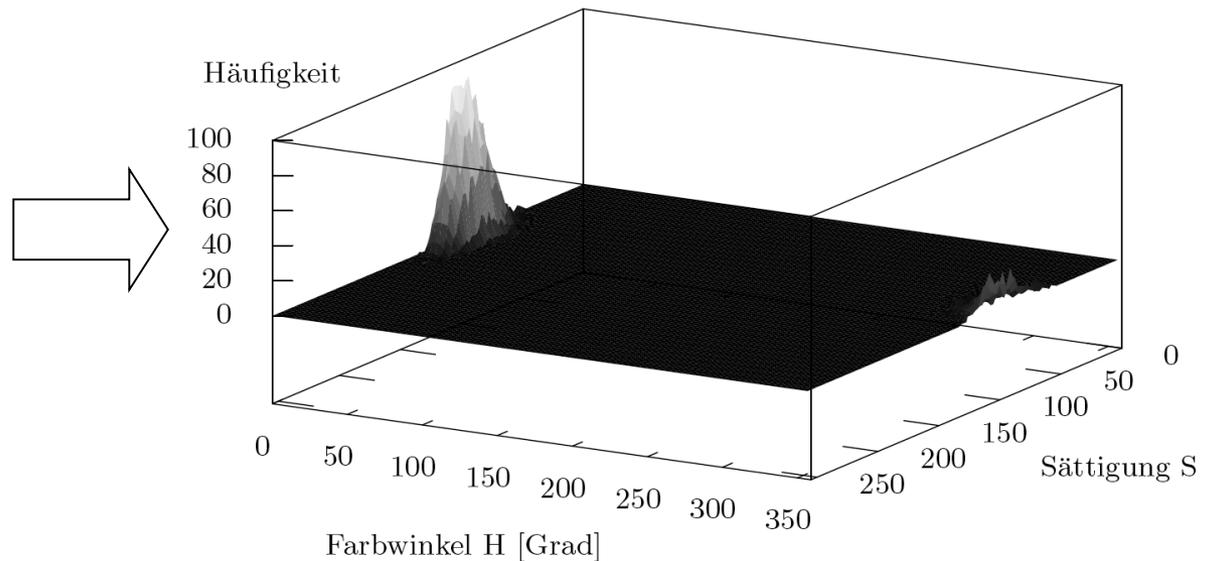
# Segmentierung: Farbe II

## ■ HS-Farbhistogramm

- Weglassen des I-Kanals ergibt 2D-Histogramm
- Training eines Klassifikators auf dem Histogramm
- Beispiel Hautfarbe:



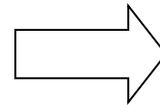
Manuell klassifiziertes  
Trainingsbild



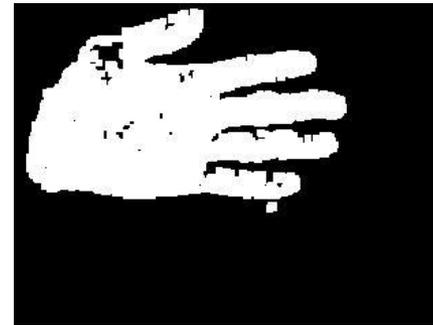
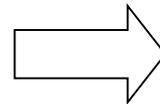
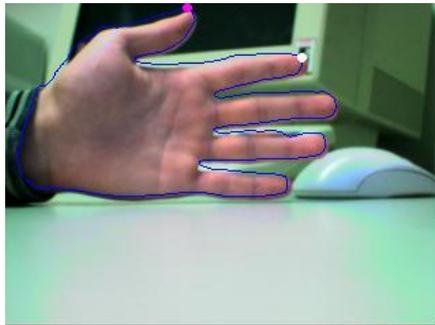
# Segmentierung: Farbe III

Histogrammbasierte Klassifikation:

- z.B. durch Bayes-Schätzer  $\Rightarrow$  Wahrscheinlichkeitsbild:



- Und anschließender Binarisierung:



- Intervallschranken im HSV-Farbraum:

$$f(H, S, V) = H \geq H_{\min} \text{ AND } H \leq H_{\max} \text{ AND } S \geq S_{\min} \\ \text{ AND } S \leq S_{\max} \text{ AND } V \geq V_{\min} \text{ AND } V \leq V_{\max}$$

- Auf der Basis der Mahalanobis-Distanz:

- Gegeben:  $x_i = (R_i \ G_i \ B_i)^T$  sind manuell positiv klassifizierte Pixel

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad \text{Kovarianzmatrix}$$

$$p(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2} \mathbf{x}^T C^{-1} \mathbf{x}\right) \quad \text{Berechnung Farbwahrscheinlichkeit}$$

# Morphologische Operatoren

- Morphologie = **Lehre von der Form**
- Morphologische Operatoren = Binäre Nachbarschaftsoperatoren zur **Veränderung von Flächen** mit Strukturelementen
- **Anwendung** auf Binärbildern:
  - Entfernung einzelstehender Pixel
  - Entfernung einzelner, dünner Linien
  - In einem weißen Objekt sind einzelne Pixel schwarz geblieben

# Morphologische Operatoren

- Prinzip: Über das Binärbild wird ein Strukturelement/Maske geschoben



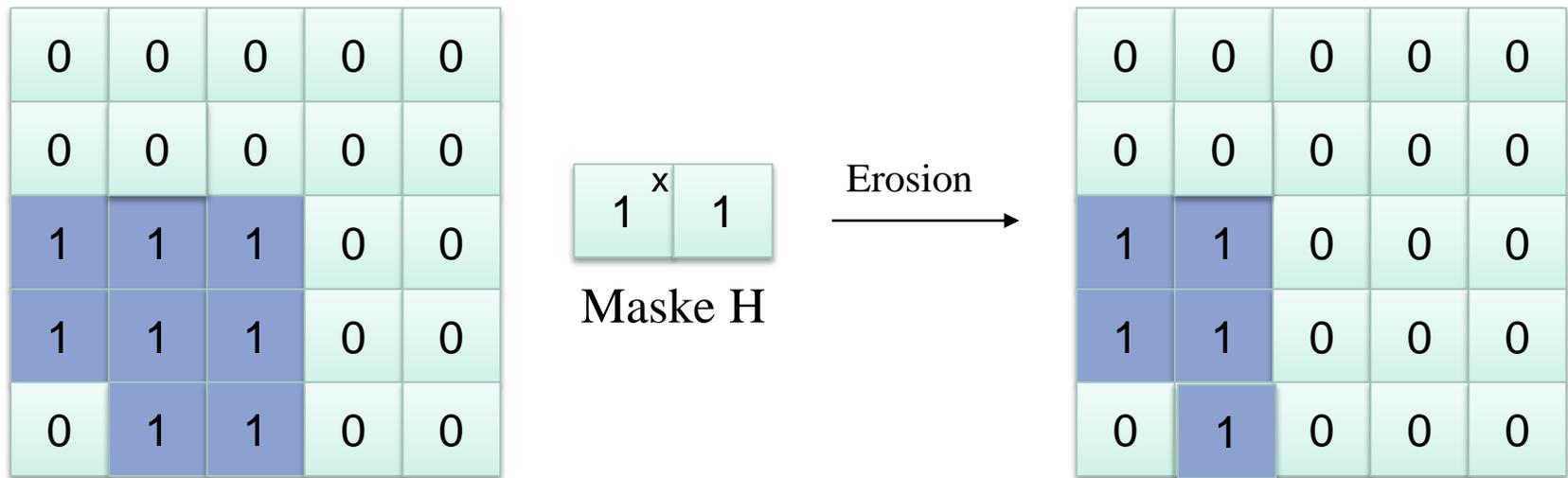
- Grundlegende Operationen

- Dilatation: Aufblasen des Objektes, vergrößert Pixel zu größeren Bereichen
- Erosion: Schrumpfen des Objektes, entfernt vereinzelte Pixel und schwach zusammenhängende Pixelgruppen

-> „Computer Vision – Das Praxisbuch“

# Erosion

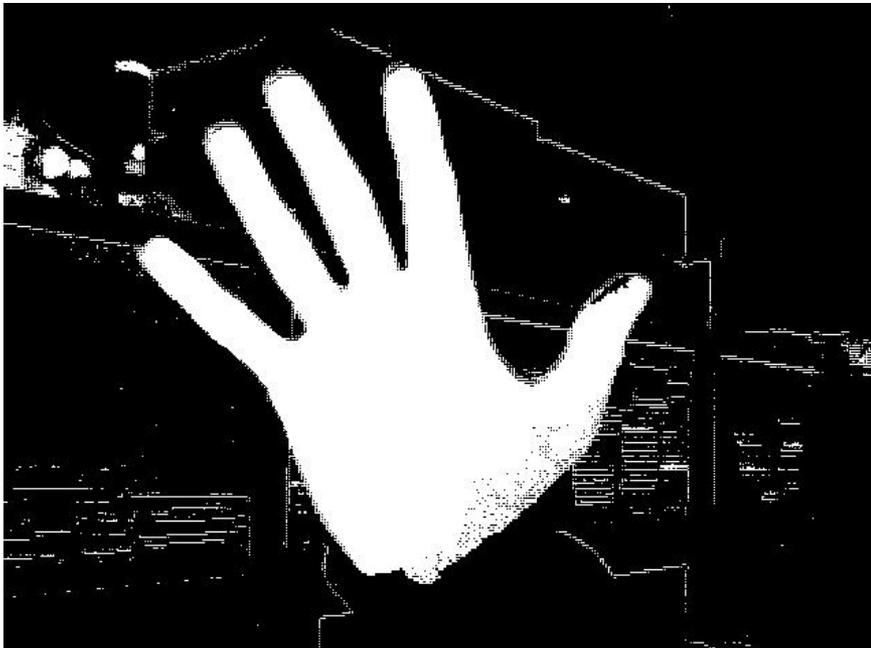
- Schiebe Maske H über das Bild
- Prüfe an jeder Position ob H Teilmenge von B ist



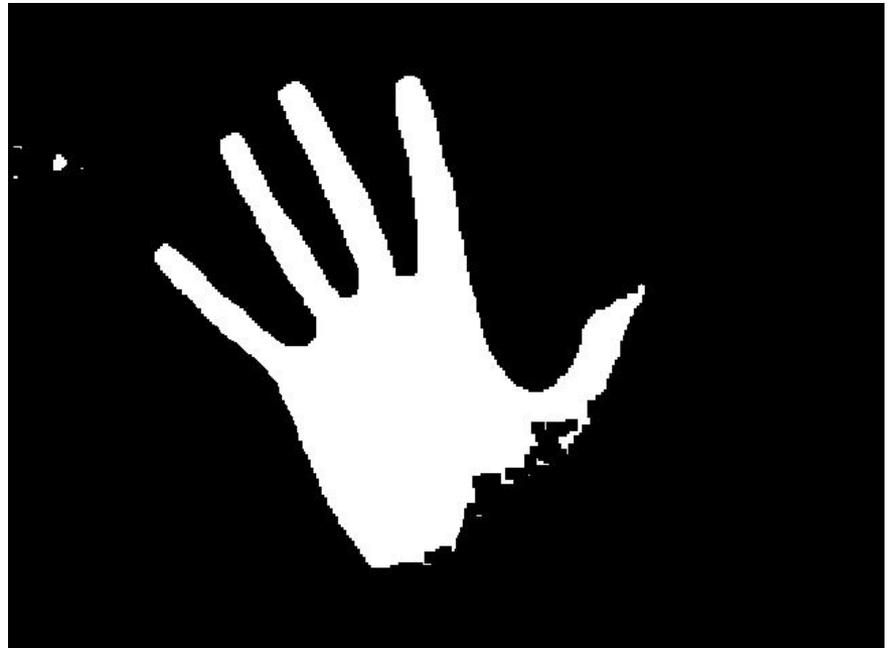
- Ergebnis abhängig von der Größe des Strukturelements
- Objekte werden verkleinert
- Dünne Linien verschwinden

# Morphologische Operatoren

- Beispielanwendung einer Erosion



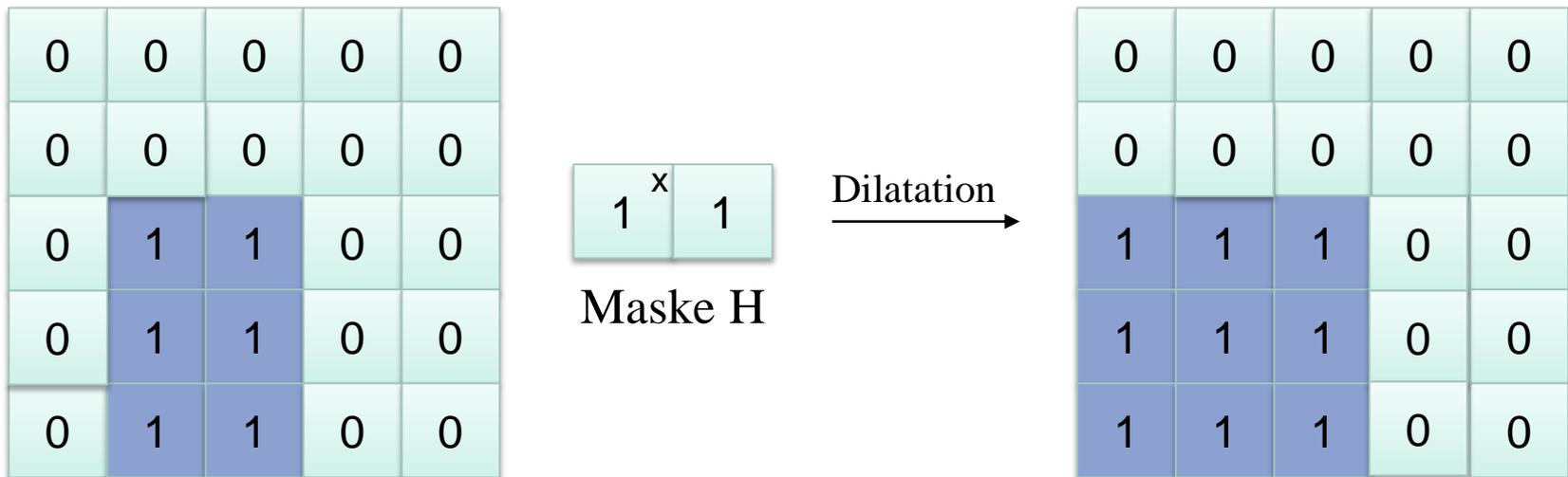
Eingabebild



Ergebnisbild

# Dilatation

- Schiebe Maske H über das Bild
- Prüfe an jeder Position, ob die Schnittmenge von B und H nicht leer ist



Ergebnis abhängig von der Größe des Strukturelements

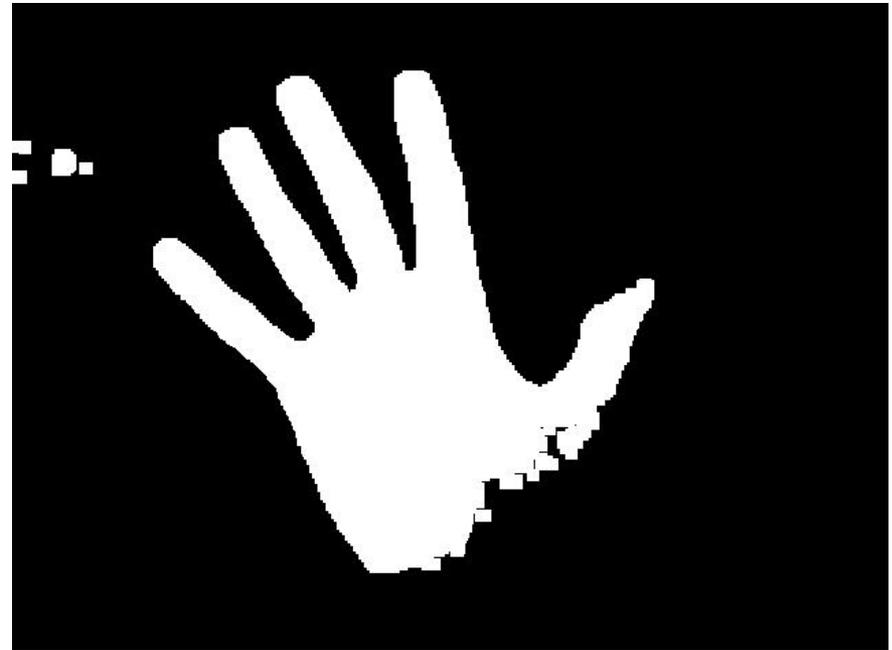
- Objekte werden vergrößert bzw. verbunden
- Hohlräume werden geschlossen

# Morphologische Operatoren

- Beispielanwendung einer Dilatation



Eingabebild



Ergebnisbild

# Opening und Closing

- Erosion verringert Ausdehnung des Objekts, Dilatation vergrößert Ausdehnung
  - Vermessung wird verfälscht
- Vermeidung: n-fache Kombination von Erosion und Dilatation
- Opening: Erosion gefolgt von Dilatation
- Closing: Dilatation gefolgt von Erosion

# Opening

- „Extremitäten“ des Objekts werden eliminiert
- Dünne Verbindungen werden aufgetrennt
- Kleine Strukturen werden beseitigt

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	1	1	1	0
0	1	1	1	1	1	1	1	1	0
0	1	1	1	0	0	1	1	1	0

1
1

Maske H

Opening →

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	1	1	1	0
0	1	1	1	0	0	1	1	1	0
0	1	1	1	0	0	1	1	1	0

# Closing

- Lücken in der Umrandung werden aufgefüllt
- Dicht beieinander liegende Objekte werden verbunden
- Anzahl der Elemente wird nur wenig verändert

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	1	0
0	1	1	1	0	0	1	1	1	0
0	1	1	1	1	1	1	1	1	0

1
1

Maske H

Closing →

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	1	0

- Einfacher Ansatz: Differenzbilder
  - Subtraktion aufeinander folgender Bilder einer Video-Sequenz:
$$\text{Img}_t(u, v)' = |\text{Img}_t(u, v) - \text{Img}_{t-1}(u, v)|$$
  - Anschließend kann auf  $\text{Img}_t'$  Schwellwertfilterung durchgeführt werden
  - Regionen, in denen sich etwas bewegt, erscheinen weiß; ruhige Regionen erscheinen schwarz
  - Bewegung in homogenen Regionen wird nicht erkannt (Kanten, Textur sind notwendig)
  - Richtung der Bewegung wird nicht erkannt
- Differenzbilder werden auch für Hintergrundsubtraktion verwendet, dann wird  $\text{Img}_{t-1}$  durch ein festes  $\text{Img}_0$  ersetzt
- Weitere Ansätze: *Optical Flow* und Erweiterungen

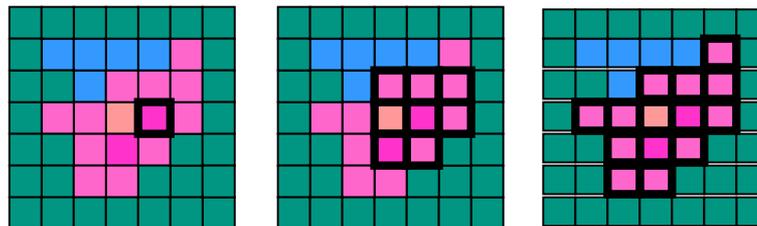
# Segmentierung: Region Growing

Gegeben: Graustufen-Bild

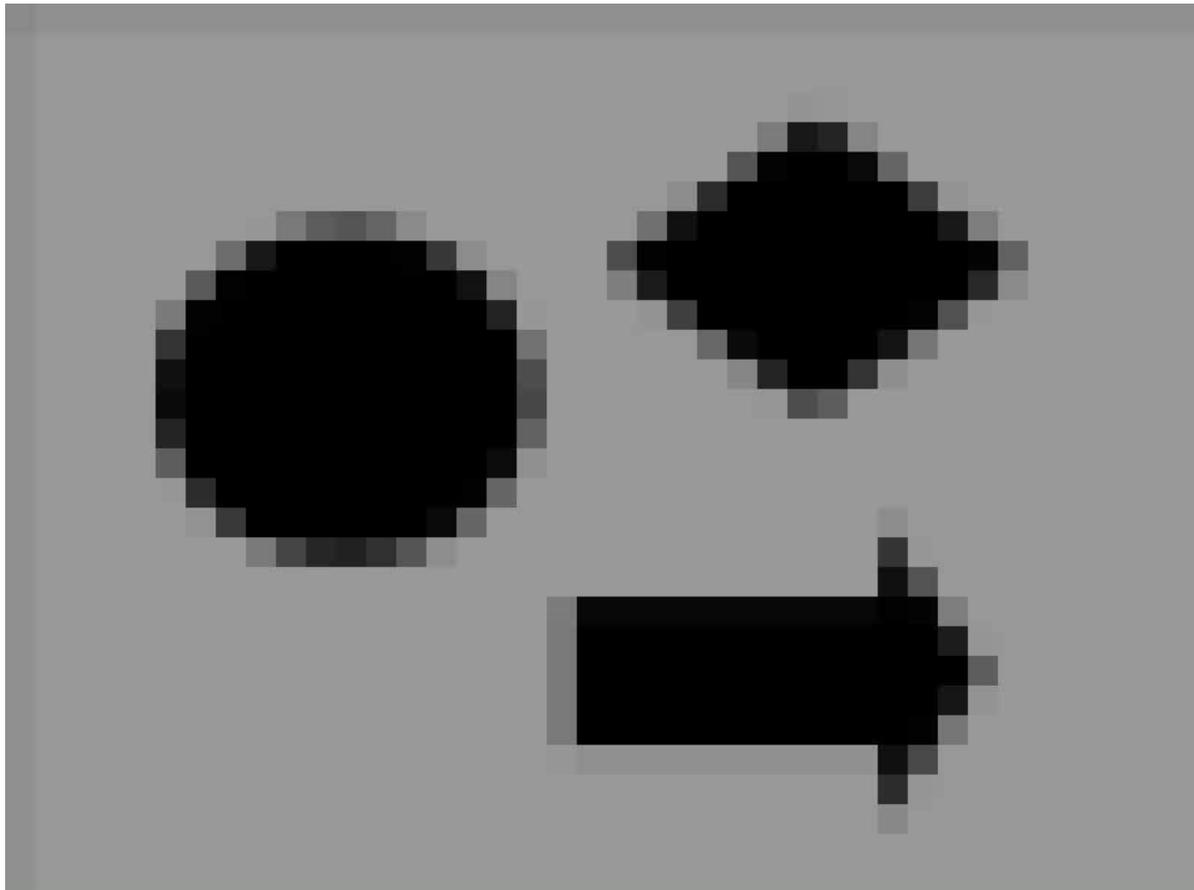
Gesucht: Zusammenhängende Regionen

Algorithmus in Pseudocode:

1. Wähle Saatpunkt  $p_0 = (u_0, v_0)$
2. Initialisiere Region  $R = \{p_0\}$ , wähle Schwelle  $\varepsilon$
3. while  $\exists p \in R, q \notin R$  mit  $\|p - q\| \leq 1$  und  $|\text{Img}(p) - \text{Img}(q)| \leq \varepsilon$ :  
 $R = R \cup \{q\}$

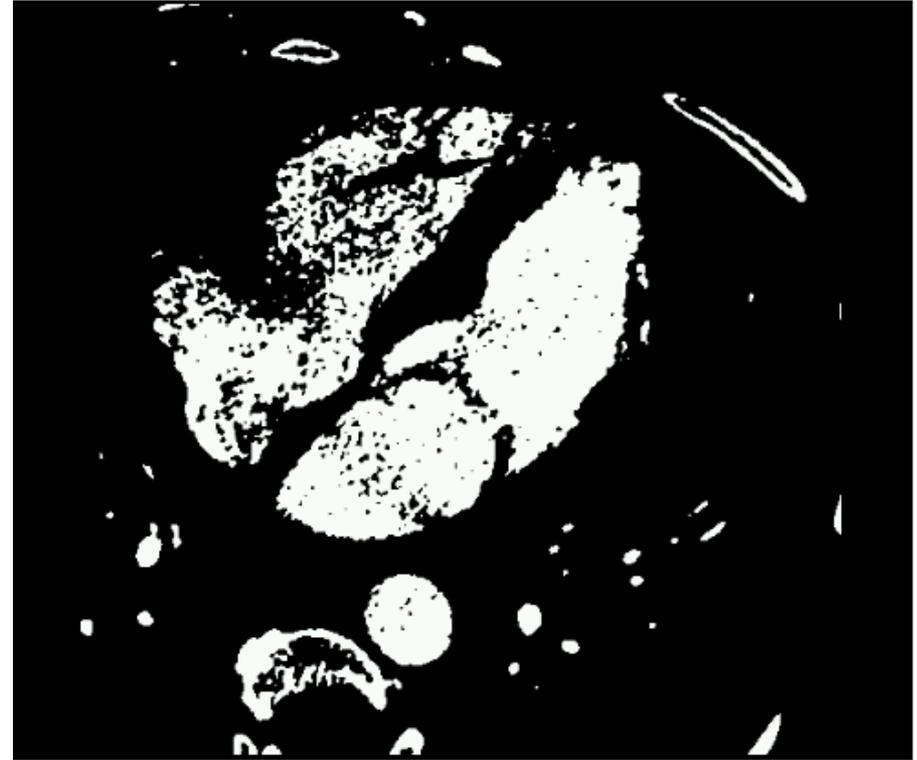
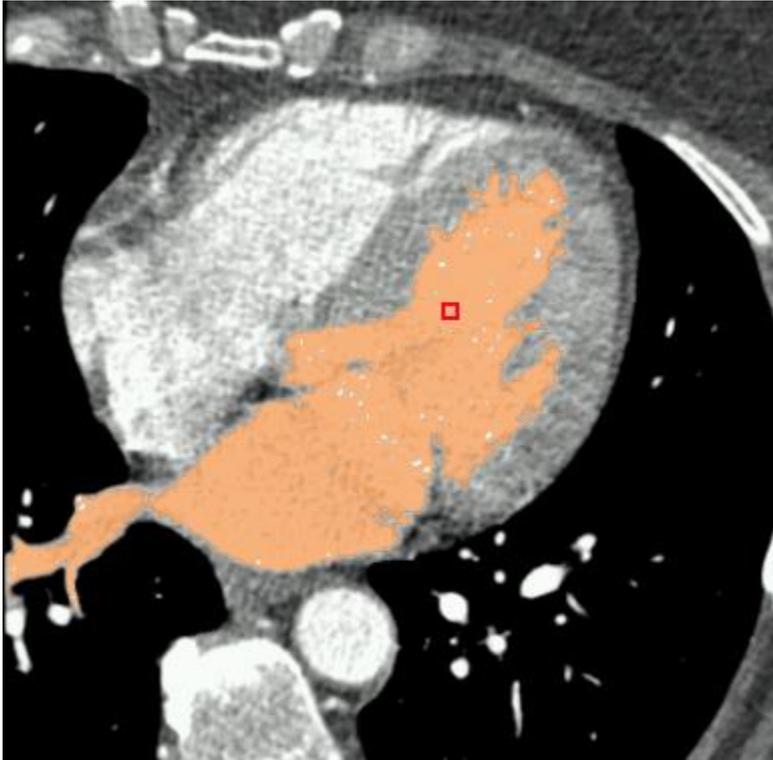


# Segmentierung: Region Growing II



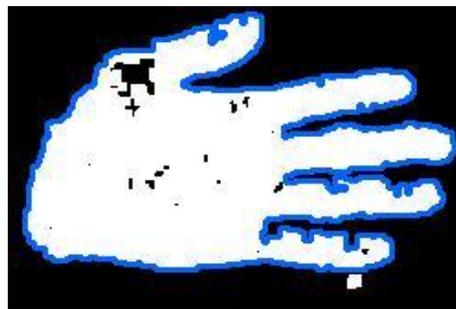
Video: Region Growing mit automatischem Setzen der Saatpunkte

# Regiongrowing vs. Threshold

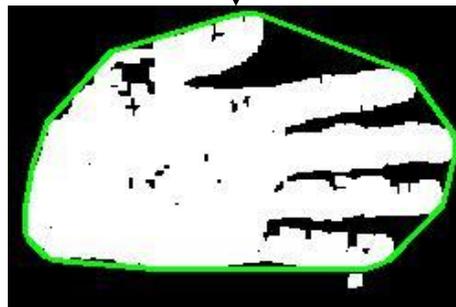


# Segmentierung: Beispielanwendungen

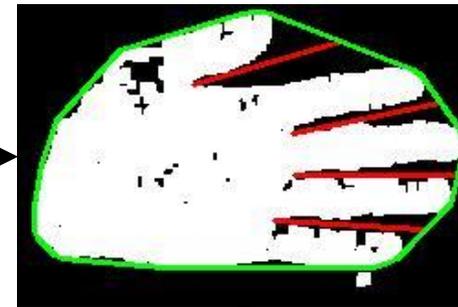
- Beispielanwendung: Finden der Fingerspitzen



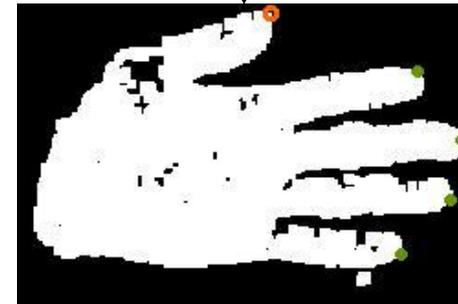
Kontur



Konvexe Hülle



Konvexitätsdefekte



Fingerspitzen

# Segmentierung: Beispielanwendungen II

- Tracking von Kopf und Händen auf der Basis von Hautfarbe und Region-Growing



Eingabe-Video

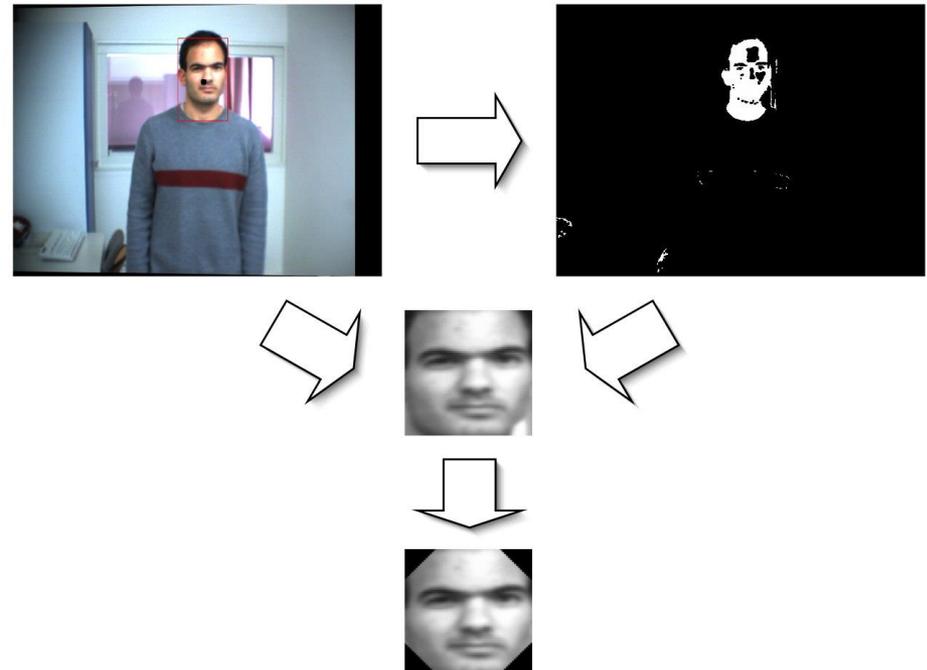


Ergebnis Segmentierung und Tracking

# Segmentierung: Beispielanwendungen III

## ■ Ausblick: Identifikation von Personen

- Auf der Basis segmentierter Hautfarbbereiche können mit Hilfe von Algorithmen zur Mustererkennung Personen über globale Ansätze identifiziert werden
- Reine Gesichtsdetektion ist auch auf Graustufenbildern und ohne Segmentierung möglich (z.B. Viola-Jones-Algorithmus)

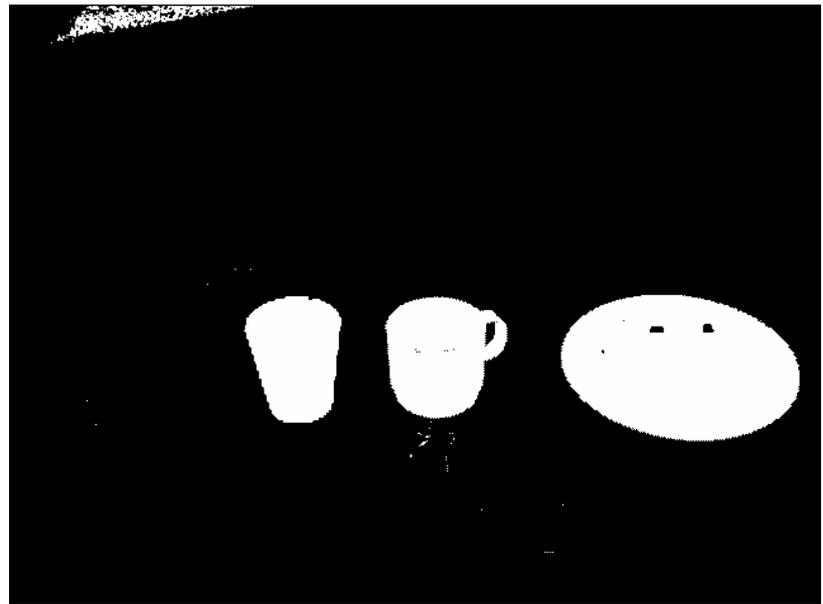


# Segmentierung: Beispielanwendungen IV

- Beispielanwendung: Objekterkennung und -lokalisierung



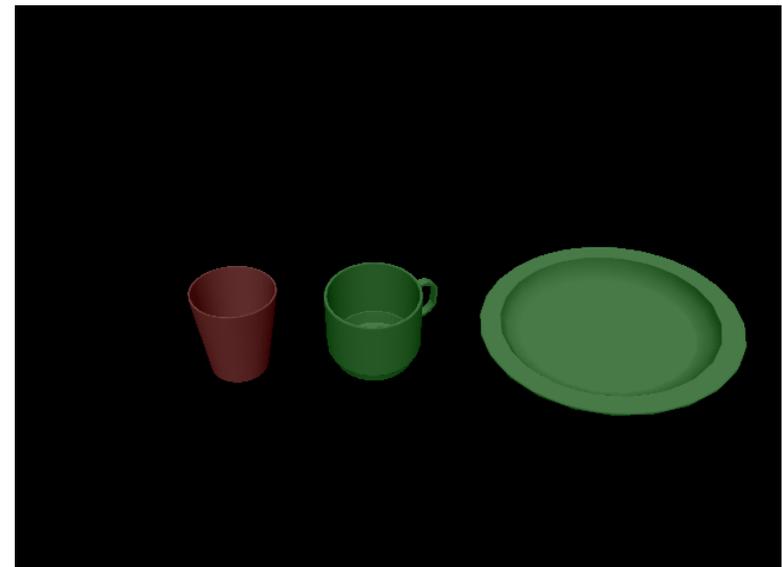
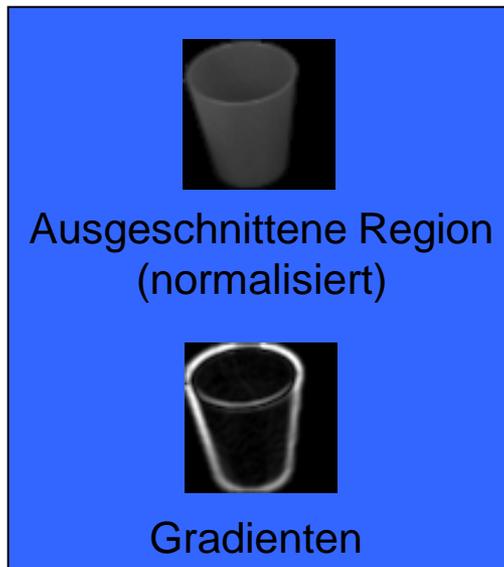
Eingabebild



Ergebnis der Segmentierung

# Segmentierung: Beispielanwendungen V

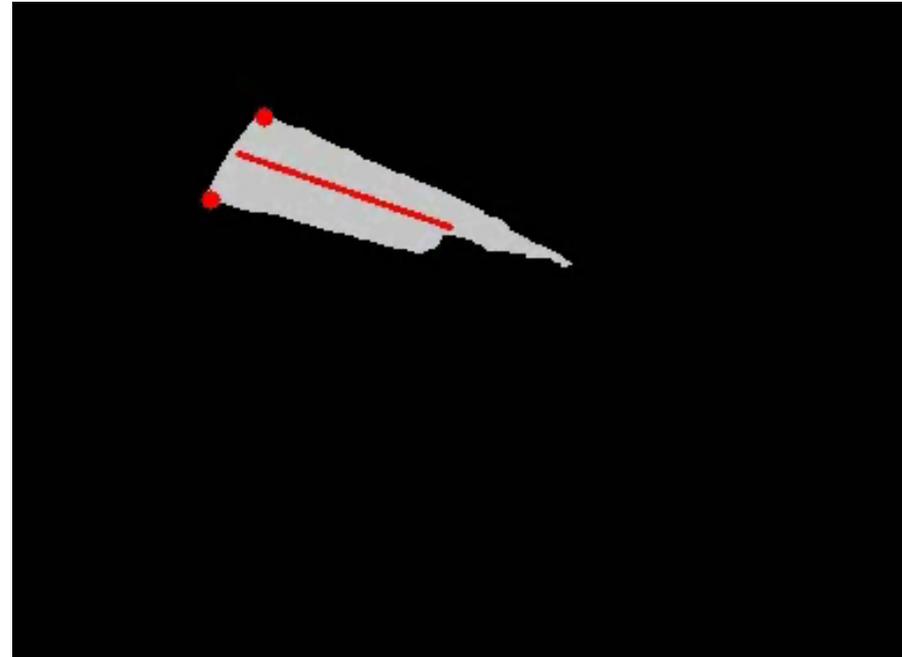
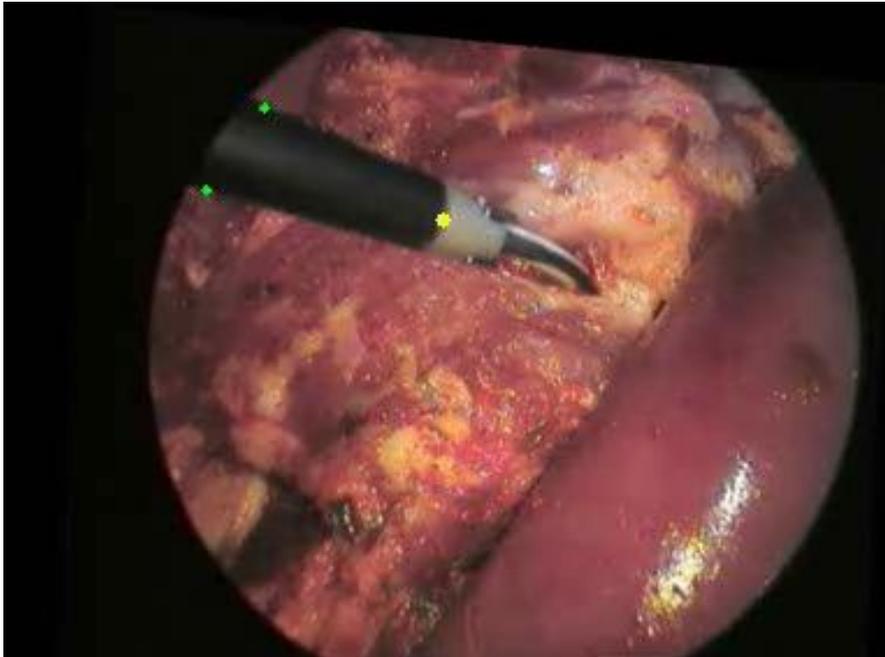
- Ausblick: Objekterkennung und -lokalisierung
  - Nach der Segmentierung können potenzielle Objektregionen ausgeschnitten werden
  - Erkennung und Lokalisierung erfolgt über Kombination aus Mustererkennungsverfahren auf Basis von Gradienten und Stereo-Sehen



Ergebnis der Szenenanalyse

Original

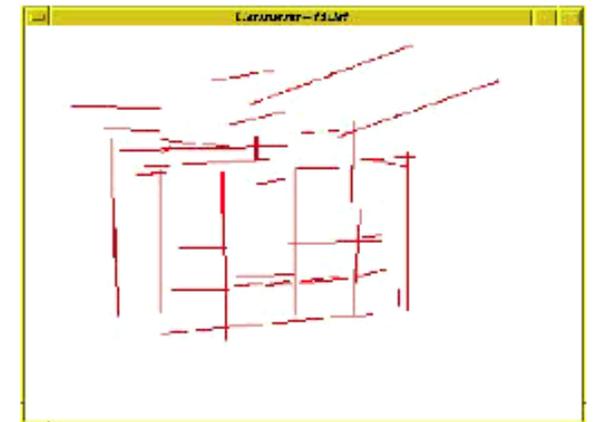
Bildbasiertes Tracking



# Segmentierung: Kanten

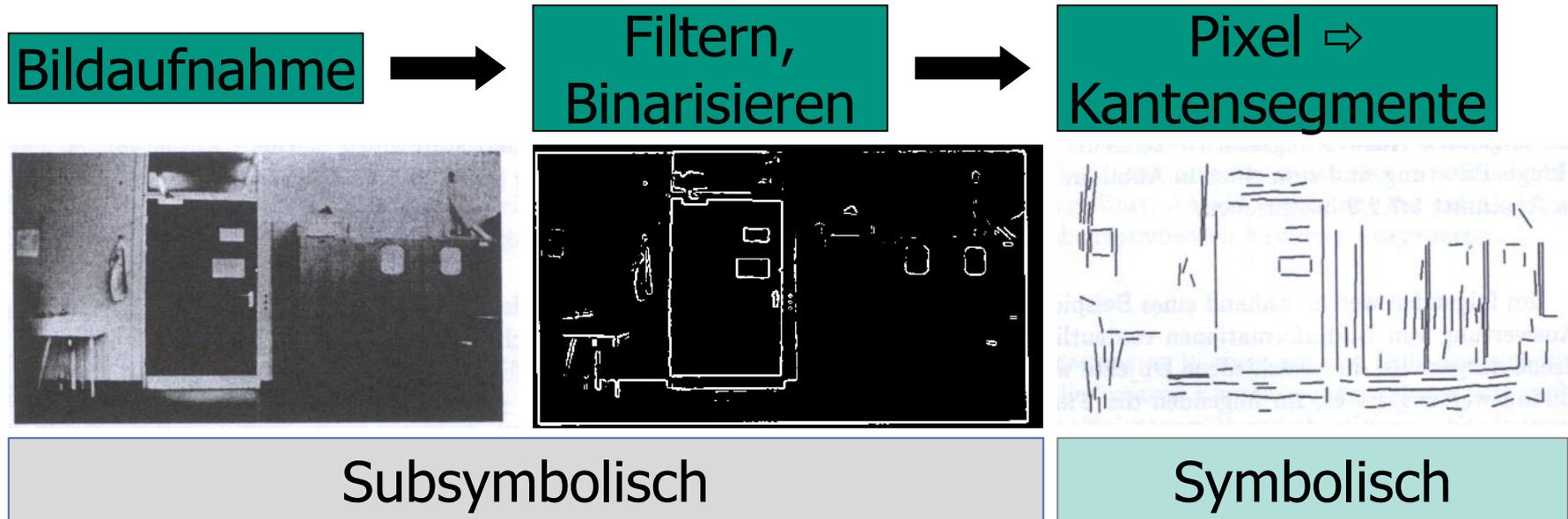
Vom Menschen konstruierte Umgebungen:

- Gut strukturiert
- Viele gerade Linien (Wände, Türen, Schränke)
- Einfache
  - Segmentierung
  - 3D-Rekonstruktion
- Viel Information in einem einzigen Merkmal



# Segmentierung: Kanten II

Vorgehensweise:



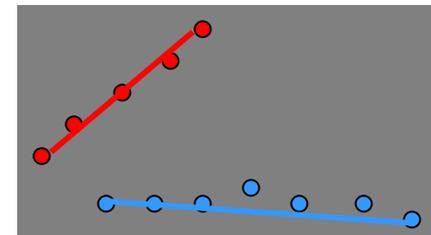
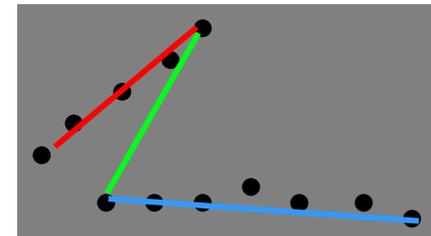
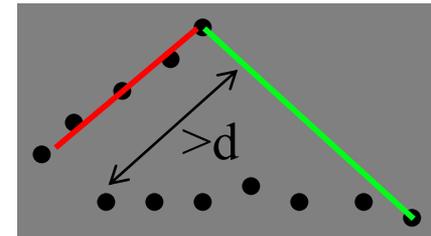
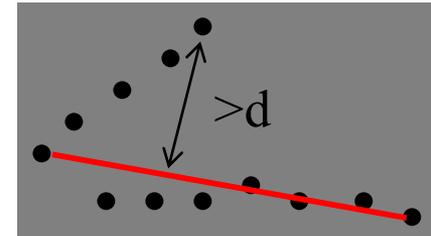
Pixel  $\Rightarrow$  Kantensegmente: Verschiedene Verfahren

# Segmentierung: Pixel $\Leftrightarrow$ Kanten

## Iterative Endpoint Fit:

Gegeben: Punkte  $P$ , Linien  $L = \{\}$ , Distanzschwelle  $d$

- Finde  $x_1, x_2 \in P$  mit  $\|x_1 - x_2\| = \max$ ;  
verbinde sie durch Linie  $l_0 = \{x_1, x_2\}$ ;  $L = L \cup \{l_0\}$
- Entferne  $x_1, x_2$  aus  $P$
- Für alle  $l \in L$ :
  - Finde  $x \in P$  mit  $\|l - x\| = \max$
  - Wenn  $\|l - x\| < d$ :
    - Ordne  $x$  als Mitgliedspunkt  $l$  zu
    - Entferne  $x$  aus  $P$
  - Sonst:
    - Brich  $l$  in  $l_1 = \{x_1, x\}$  und  $l_2 = \{x, x_2\}$  auf
    - Andere Mitgliedspunkte von  $l$  wieder in  $P$
- Lösche Linien mit weniger als  $n$  Punkten

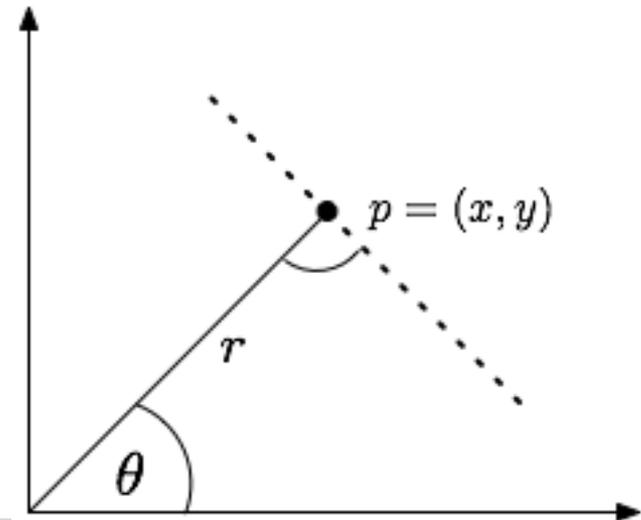
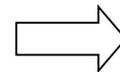


# Segmentierung: Pixel $\Rightarrow$ Kanten II

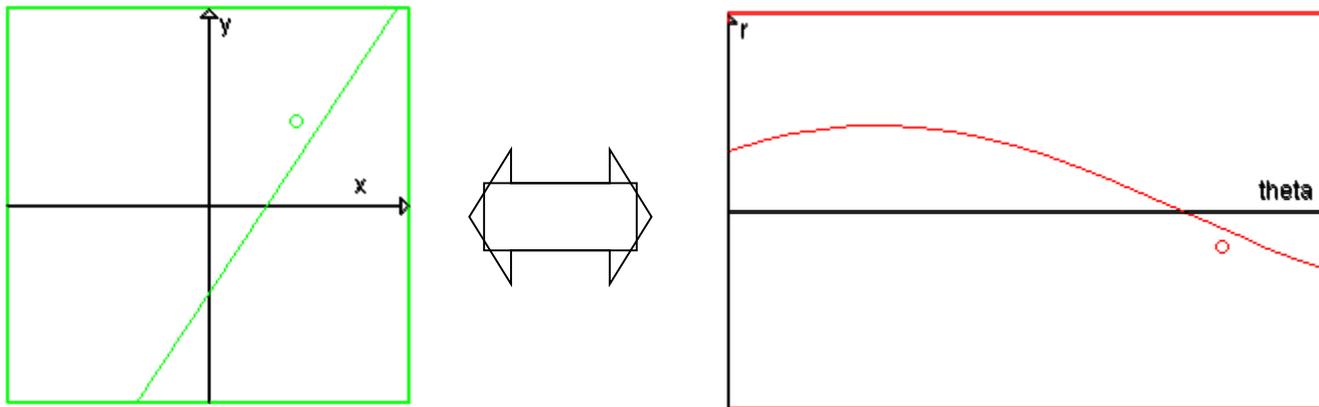
## Hough-Transformation

- Ziel: Erkennung gerader Linien im Bild
- Definition eines zum Bild dualen Parameterraums:
  - Dimensionalität des Parameterraums=Anzahl Parameter zur Beschreibung des Objekts
- Ansatz: Stelle Linie durch Normalenvektor (Länge, Winkel) in Polarkoordinaten dar (Sinus-Kosinus-Kurve)

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

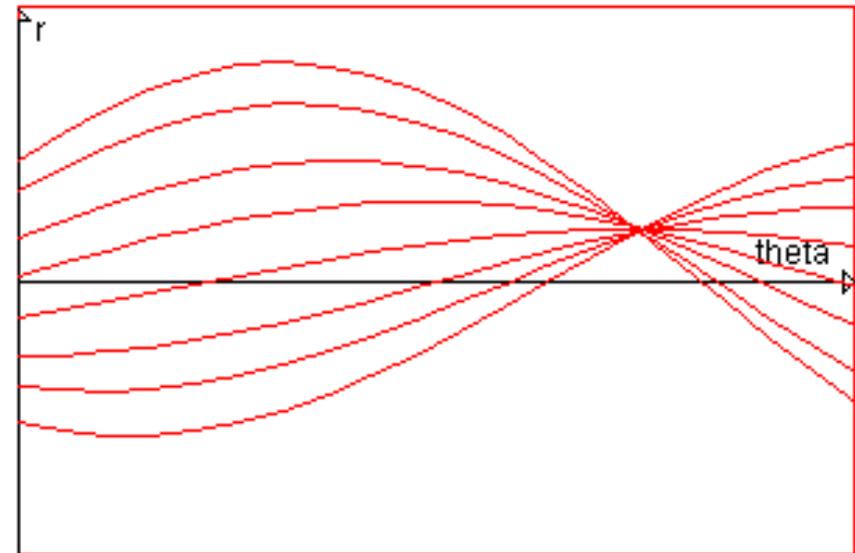
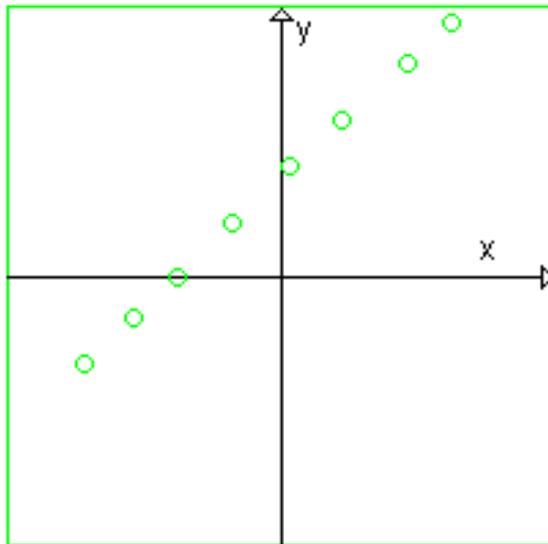


- Kantenpixel im Bildraum entspricht Sinusoid im Parameterraum.
- Punkt im Parameterraum entspricht Gerade im Bildraum.



# Hough-Transformation II

- Realisierung des Parameterraums als Akkumulator Array
- Für jedes Kantenpixel:
  - Erhöhen der Akkumulatorzellen auf dem zugehörigen Sinusoid
- Gesuchte Gerade entspricht Maximum im Parameterraum



- Erweiterung der Hough-Transformation für Geraden auf andere parametrische Formen möglich

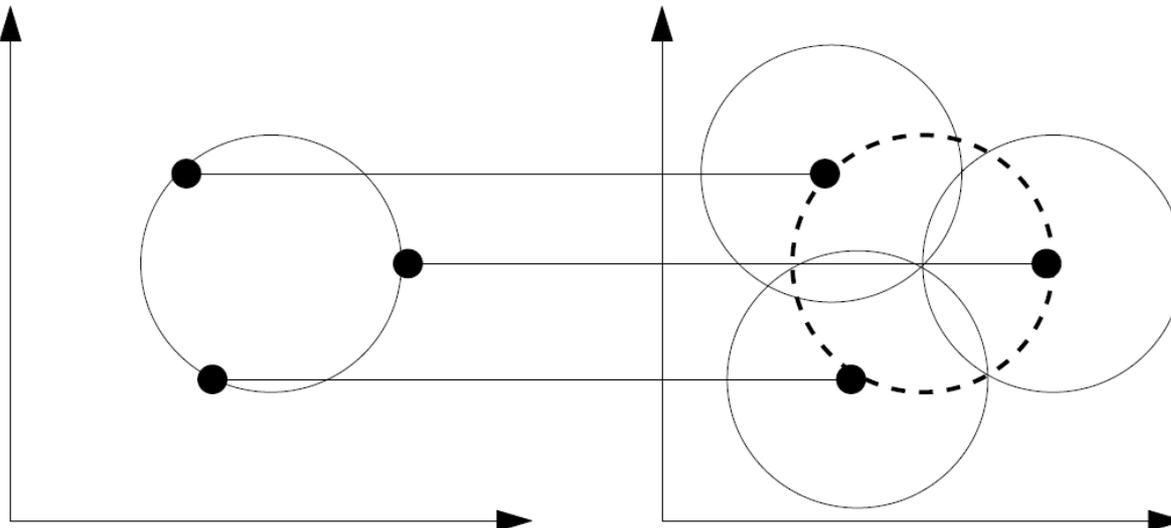
Anwendung zur Detektion von Kreisen

- Kreisgleichung:

$$r^2 = x^2 + y^2$$

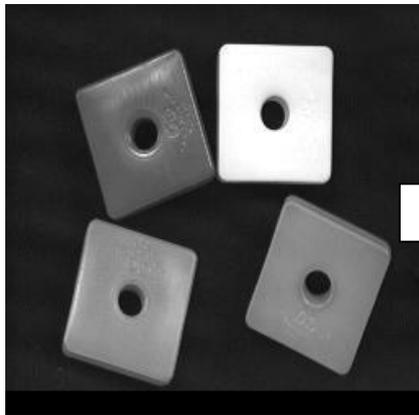
- Bei unbekanntem  $r \Rightarrow$  3D Parameterraum  $(x,y,r)$
- Für Kreise mit festem Radius  $r$  – analog zu Geradentransformation

- Für jedes Kantenpixel:
  - Erhöhen der Akkumulatorzellen auf dem zugehörigen Kreis
- Gesuchter Kreis entspricht Maximum im Parameterraum

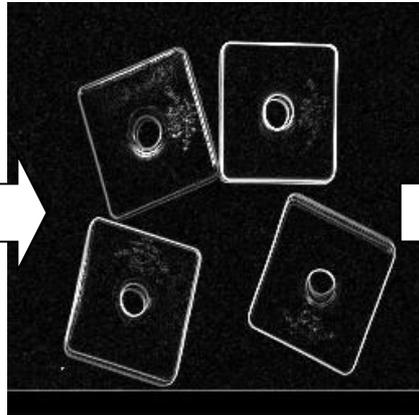


# Segmentierung: Pixel $\Rightarrow$ Kanten III

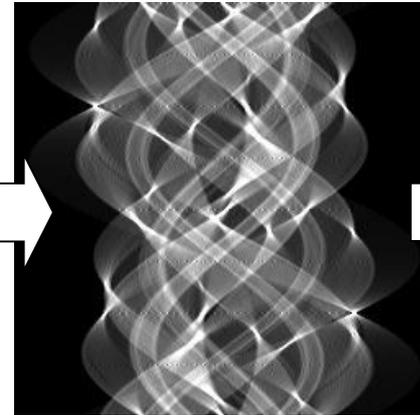
- Transformation in den Hough-Raum: Additives Eintragen aller Sinus-Kosinus-Kurven für alle Pixel in ein Histogrammbild
- Finden der Maxima bzw. Cluster von „Treffern“ im Hough-Raum



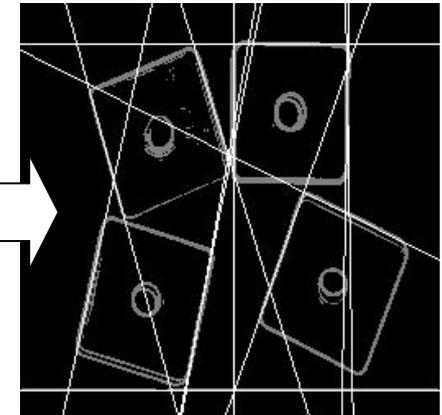
Eingabebild



Nach Anwendung  
Kantenfilter



Histogrammbild  
(=Hough-Raum)



Rücktransformation  
der berechneten  
Maxima bzw. Cluster

# Segmentierung: Pixel $\Leftrightarrow$ Kanten IV

- Gegeben:  $n > 2$  Pixel mit Koordinaten  $(u_i, v_i)$
- Gesucht: Regressionsgerade (bester „Fit“)
- Lösung: Methode der kleinsten Quadrate (*Least Squares*)
- Gerade kann modelliert werden als (einfachste Variante):  
 $u \rightarrow v: f(u) = mu + c$
- Formulierung als überbestimmtes Lineares Gleichungssystem

$Ax = b$ :

$$\begin{pmatrix} u_1 & 1 \\ \vdots & \vdots \\ u_n & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

- Optimale Lösung kann beispielsweise durch Lösung der Normalengleichung berechnet werden:

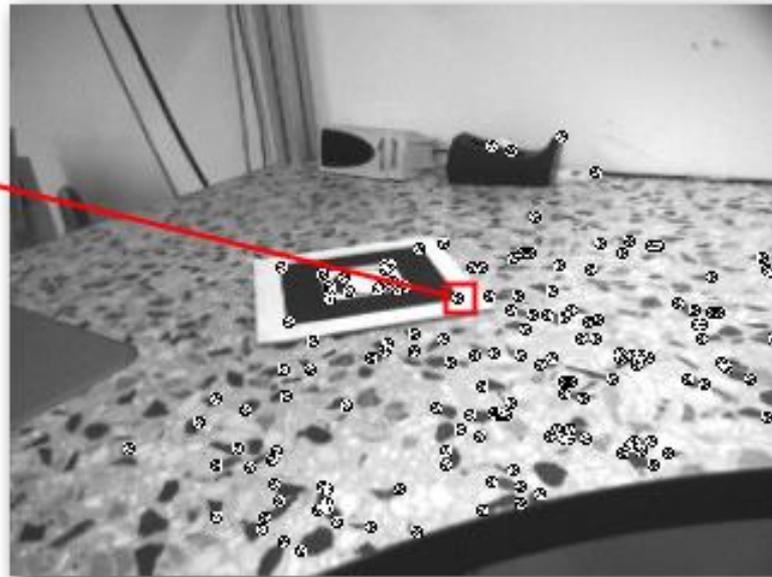
$$A^T A x = A^T b$$

# Segmentierung: Punktmerkmale

- Kanten/Konturen/Farbe können nicht immer für die Segmentierung herangezogen werden
- Texturierte Objekte lassen sich in der Regel nicht durch die bislang vorgestellten Verfahren segmentieren
- Lösung: Verwendung von **lokalen** Punktmerkmalen (auch genannt Texturmerkmale):
  - Harris Corner Detector
  - Shi-Tomasi Features
  - SIFT-Features
  - Maximally Stable Extremal Regions
  - ...

# Segmentierung: Punktmerkmale II

- Punktmerkmal:  $(2n+1) \times (2n+1)$ -Pixel-Block um Pixel  $p$
- Fast immer basierend auf Grauwertbildern
- Gewünschte Eigenschaft: Wiedererkennbarkeit  
⇒ hoher Gradient in mehrere Richtungen



Beispiel: **Harris-Corner-Detector** (nach [Harris 88]):  
Sind die Eigenwerte der Matrix

$$A = \begin{pmatrix} \left( \frac{\partial \text{Img}(x,y)}{\partial x} \right)^2 & \frac{\partial \text{Img}(x,y)}{\partial x} \frac{\partial \text{Img}(x,y)}{\partial y} \\ \frac{\partial \text{Img}(x,y)}{\partial x} \frac{\partial \text{Img}(x,y)}{\partial y} & \left( \frac{\partial \text{Img}(x,y)}{\partial y} \right)^2 \end{pmatrix}$$

groß, dann verursacht eine kleine Bewegung in beliebiger Richtung eine große Grauwertänderung.

Finden von Ecken durch Suche nach lokalen Maxima in:

$$R = \det A - k \cdot \text{trace}(A)^2, k \approx 0.04$$

Häufiges Problem:

- Wiederfinden bzw. Zuordnung von Punktmerkmalen (*Korrespondenzproblem*) für:
  - Objekterkennung auf der Basis von Punktmerkmalen
  - Stereo-Sehen bzw. „Structure from Motion“
- Lösung des Korrespondenzproblems erfolgt für Punktmerkmale durch Korrelationsverfahren:
  - SSD (*Sum of Squared Differences*)
  - SAD (*Sum of Absolute Differences*)
  - Kreuzkorrelation
  - Erweiterungen: *Zero Mean Normalized Cross-Correlation*, ...

- Zwei Beispiele für Korrelationsverfahren zwischen zwei Bildausschnitten der Größe  $(2n+1) \times (2n+1)$  um den Mittelpunkt  $(u_1, v_1)$  im Bild  $\text{Img}_1$  und um den Mittelpunkt  $(u_2, v_2)$  im Bild  $\text{Img}_2$ :

- **Sum of Squared Differences (SSD)** wird *minimal* bei guter Übereinstimmung:

$$\sum_{i=-n}^n \sum_{j=-n}^n (\text{Img}_1(u_1 + i, v_1 + j) - \text{Img}_2(u_2 + i, v_2 + j))^2$$

- **Zero Mean Normalized Cross-Correlation (ZNCC)** wird *maximal* bei guter Übereinstimmung:

$$\sum_{i=-n}^n \sum_{j=-n}^n (\text{Img}_1(u_1 + i, v_1 + j) - \overline{\text{Img}_1}(u_1, v_1, n)) \cdot (\text{Img}_2(u_2 + i, v_2 + j) - \overline{\text{Img}_2}(u_2, v_2, n))$$

$$\sqrt{\sum_{i=-n}^n \sum_{j=-n}^n (\text{Img}_1(u_1 + i, v_1 + j) - \overline{\text{Img}_1}(u_1, v_1, n))^2 \sum_{i=-n}^n \sum_{j=-n}^n (\text{Img}_2(u_2 + i, v_2 + j) - \overline{\text{Img}_2}(u_2, v_2, n))^2}$$

$$\overline{\text{Img}}(u, v, n) = \frac{1}{(2n+1)^2} \sum_{i=-n}^n \sum_{j=-n}^n \text{Img}(u + i, v + j)$$

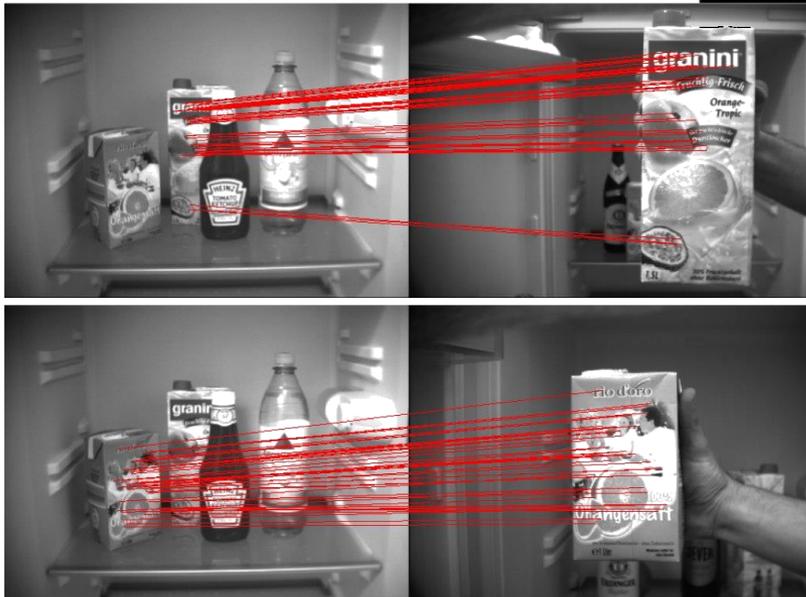
- Für die Objekterkennung existieren eine Vielzahl von Repräsentationen von lokalen Merkmalen:
  - SIFT (Scale Invariant Feature Transform)
  - SURF (Speeded Up Robust Features)
  - MSER (Maximally Stable Extremal Regions)
  - Repräsentation eines Bildausschnittes über eine Vielzahl (100-200) synthetisch generierter Ansichten, Matching über PCA (Principal Component Analysis)
  - ...

# Segmentierung: Punktmerkmale VII

## ■ Ausblick: Anwendung Objekterkennung

Korrespondenzen zwischen  
eingelernter und aktueller Sicht

Juice: Granini  
Juice: Rio  
Ketchup: Heinz  
Water: Apollinaris



Ergebnis einer Szenenanalyse  
im Kühlschrank

- Ansichten können synthetisch generiert werden durch
  - Verzerrungsabbildung (Affine Transformation)
  - Texturiertes 3D-Modell und Rendering



100 synthetisch generierte Ansichten eines  $32 \times 32$  Pixel großen Merkmals

## ■ Nachteil:

- Hoher Speicheraufwand:  $100 \cdot 32 \cdot 32 = 102400$  Bytes / Merkmal
- Hoher Rechenaufwand: 102400 Vergleiche pro Merkmal

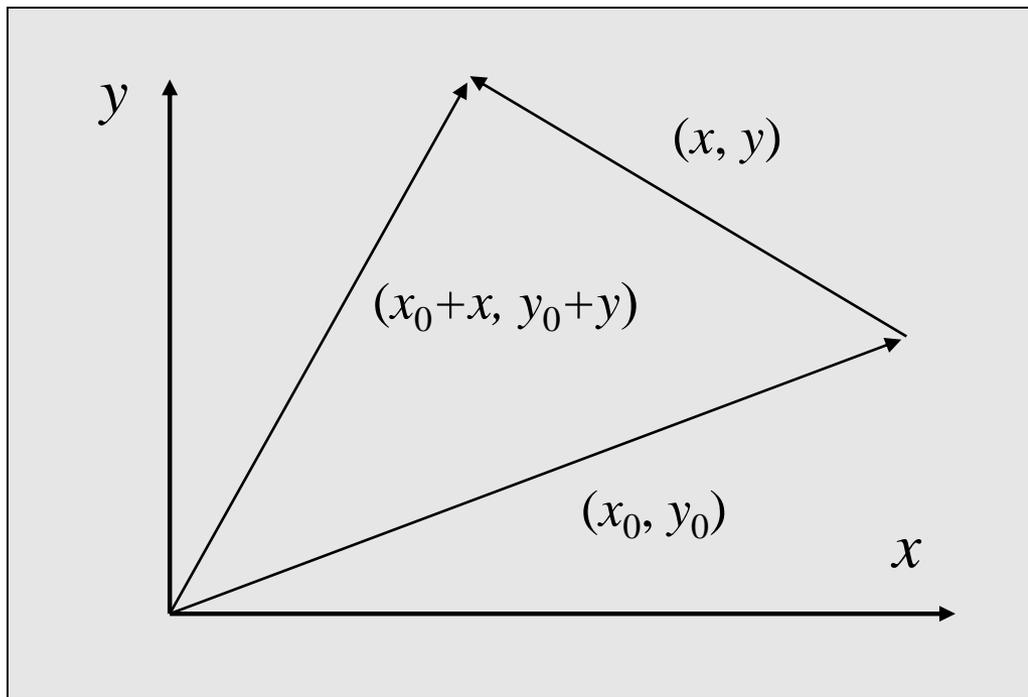
## ■ Lösungsansatz

- Komprimierung der Ansichten mithilfe der Hauptkomponentenanalyse (PCA)
- Reduktion von  $32 \cdot 32 = 1024$  auf 20 Dimensionen
- Dadurch Beschleunigung und Reduktion des Speicheraufwands um in etwa Faktor 50

- Zu unterscheiden sind
  - Merkmalsdetektoren
  - Merkmalsdeskriptoren
- Verfahren beschrieben, welche sowohl den Detektor als auch den Deskriptor spezifizieren:
  - Scale Invariant Feature Transform (SIFT)
  - Speeded Up Robust Features (SURF)
  - Maximally Stable Extremal Regions (MSER)

# Geometrische 2D-Transformationen

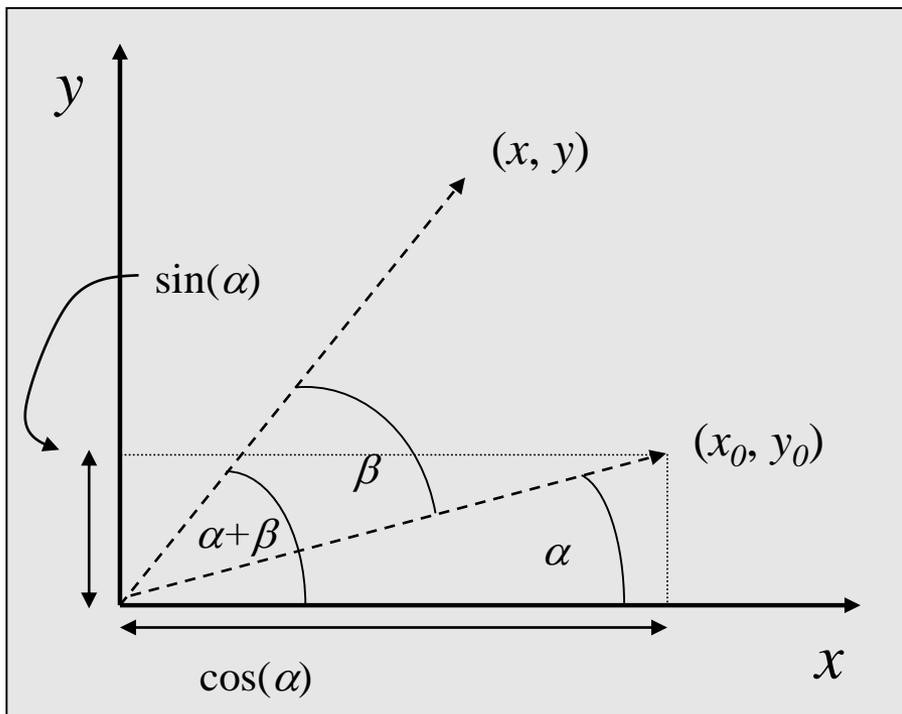
## ■ Translation eines 2D-Vektors:



$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 + x \\ y_0 + y \end{pmatrix}$$

# 2D-Transformationen: Rotation

- o.B.d.A. auf Einheitsvektor zurückführbar (Basistransformation)
- Konvention: *Rechtskoordinatensystem*
- Rotation von  $(x_0, y_0)$  um Winkel  $\beta$  mit Ergebnis  $(x, y)$ :



Aus Additionstheorem:

$$\begin{aligned}x &= \cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta \\y &= \sin(\alpha + \beta) = \sin \beta \cos \alpha + \cos \beta \sin \alpha\end{aligned}$$

und mit  $(x_0, y_0) = (\cos \alpha, \sin \alpha)$ :

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

# Homogene Koordinaten

Definition Homogene Koordinaten:

Homogene Koordinaten

$$\mathbf{h} = (h_1, \dots, h_i, h_{i+1})$$

eines Punktes  $p$  im  $R^i$  mit

$$\mathbf{p} = (p_1, \dots, p_i)$$

sind Zahlen, für die gilt:

$$p_k = \frac{h_k}{h_{i+1}} \quad \forall 1 \leq k \leq i.$$

$h_{i+1}$  ist ein Skalierungsfaktor; üblicherweise 1.

# Homogene 2D-Transformationen

- Transformation definiert durch Rotation  $R$  und Translation  $t$ :

$$\begin{pmatrix} x \\ y \end{pmatrix} = R \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + t = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

- Darstellung mit Hilfe homogener Koordinaten und einer geschlossenen Transformationsmatrix:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \left( \begin{array}{cc|c} & R & t \\ \hline 0 & 0 & 1 \end{array} \right) \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} = A \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix}$$

**Nächste Veranstaltung:**

**Mittwoch, 28. Juni 2017**

**Übung 4**