

Chapter 5

Object Recognition



Distributed Algorithms
for Image and Video Processing

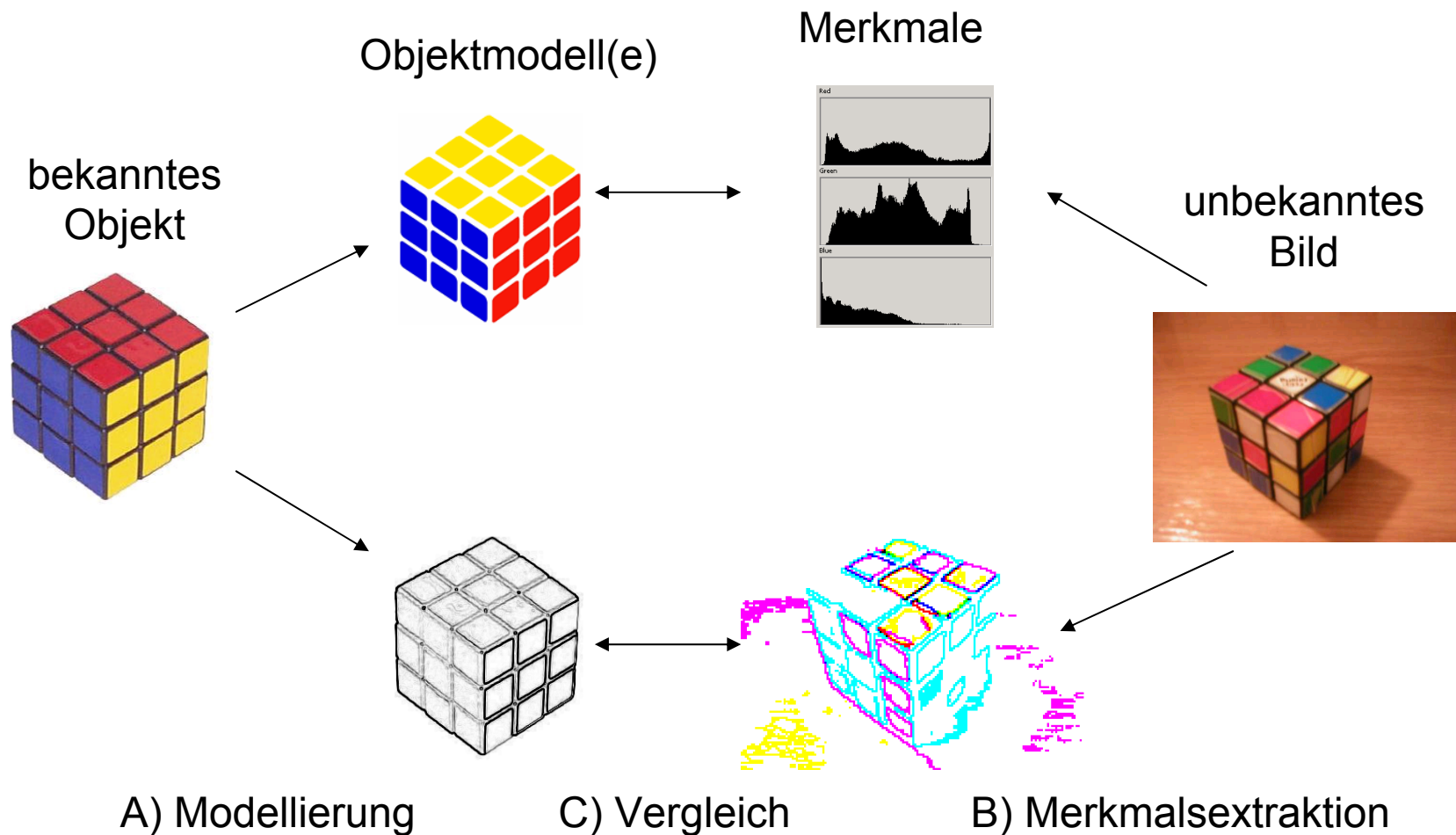
Übersicht

- Motivation
- Anforderungen
- Objekterkennung mittels
 - Farben
 - Konturen
 - Texturen
 - Bewegungen
- Zusammenfassung

Warum Objekterkennung ?

- **Schrifterkennung**
 - OCR-Software zur Texterkennung
 - Automatische Postleitzahlenerkennung
- **Video-Überwachung:**
 - Identifikation verdächtiger Personen (Diebstahl, Bedrohung)
 - Erkennung verdächtiger Gegenstände (unbeaufsichtigte Koffer auf Flughäfen oder Bahnhöfen)
- **Inhaltsbasierte Bildsuche**
 - Internetsuche (z.B. [Google Bilder](#))
 - Bild- und Videoarchive (z.B. [gettyimages](#))

Modell-basierte Objekterkennung



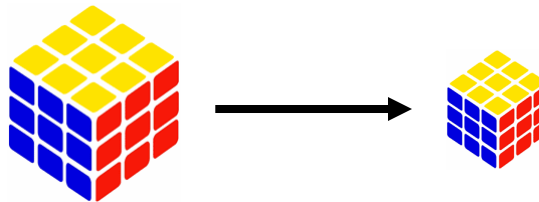
Was ist ein gutes Objektmodell ?

- *Sensitivity*: Das Modell muss alle relevanten Merkmale darstellen und zwischen diesen unterscheiden können.
- *Uniqueness*: Das Modell soll in der Lage sein, Objekte möglichst eindeutig zu beschreiben; gleichartige Objekte sollen die gleiche Beschreibung haben.
- *Stability*: Kleine Veränderungen am Objekt sollen kleine Auswirkungen auf das Modell haben.
- *Efficiency*: Es muss möglich sein, Merkmale effizient aus Daten zu berechnen und Merkmale effizient zu vergleichen.

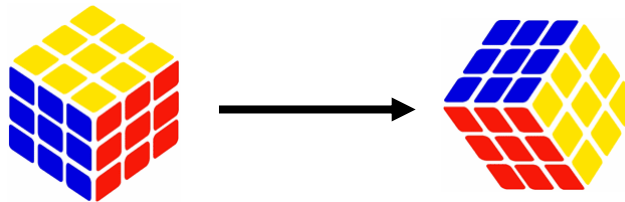
Anforderungen (I)

Das Modell sollte invariant sein gegenüber:

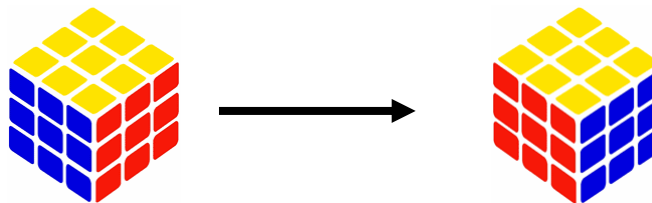
- Skalierungen



- Rotationen



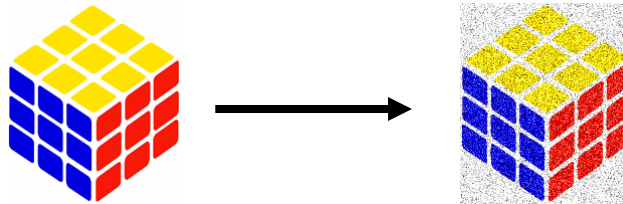
- Spiegelungen



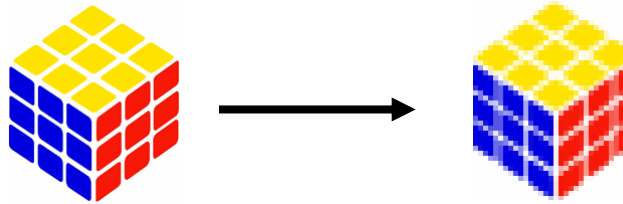
Anforderungen (II)

Das Modell sollte invariant sein gegenüber:

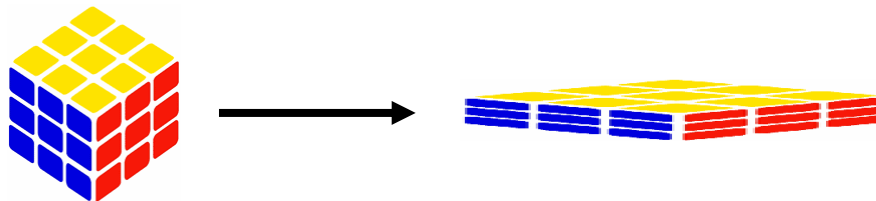
- Rauschen



- Unschärfe



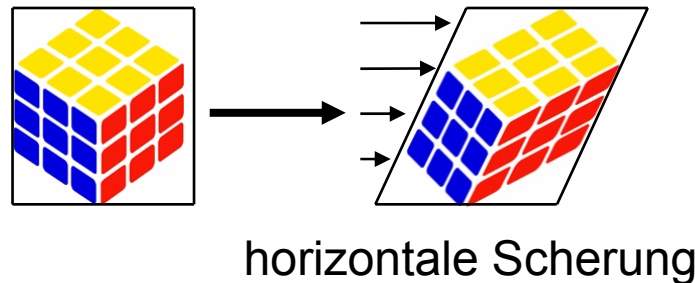
- Stauchungen



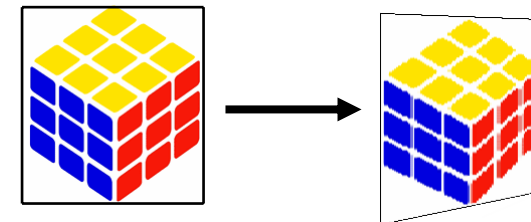
Anforderungen (III)

Das Modell sollte invariant sein gegenüber:

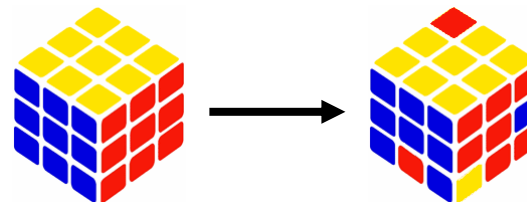
- Scherungen



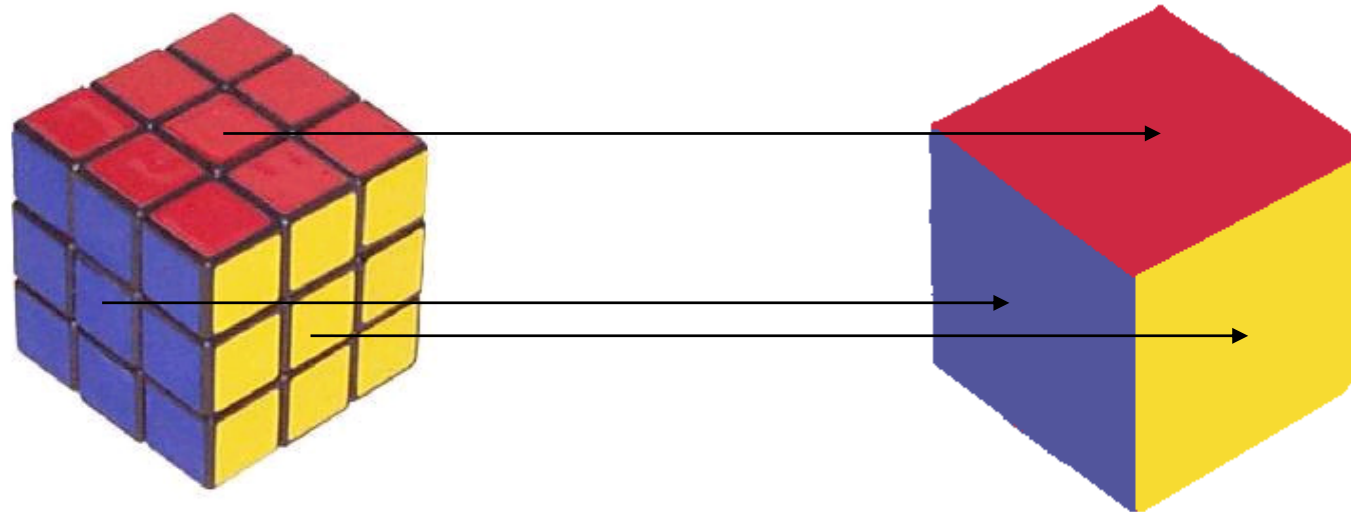
- Perspektivischen Verzerrungen



- Objektverformungen



Objekterkennung durch Vergleich von Farben



Farbräume (I)

Definitionen:

- **Farbe:** Sinnesempfindung (keine physikalische Eigenschaft), falls Licht einer bestimmten Wellenlänge auf die Netzhaut des Auges fällt. Die Sinneszellen im menschlichen Auge (Zapfen für Farben, Stäbchen für Helligkeitswerte) leiten Impulse an das Gehirn, das diese als Farben wahrnimmt. Mensch besitzt drei Arten von Zapfen (empfindlich für rot, grün und blau).
- **Sichtbares Licht:** zwischen 400 – 700 nm



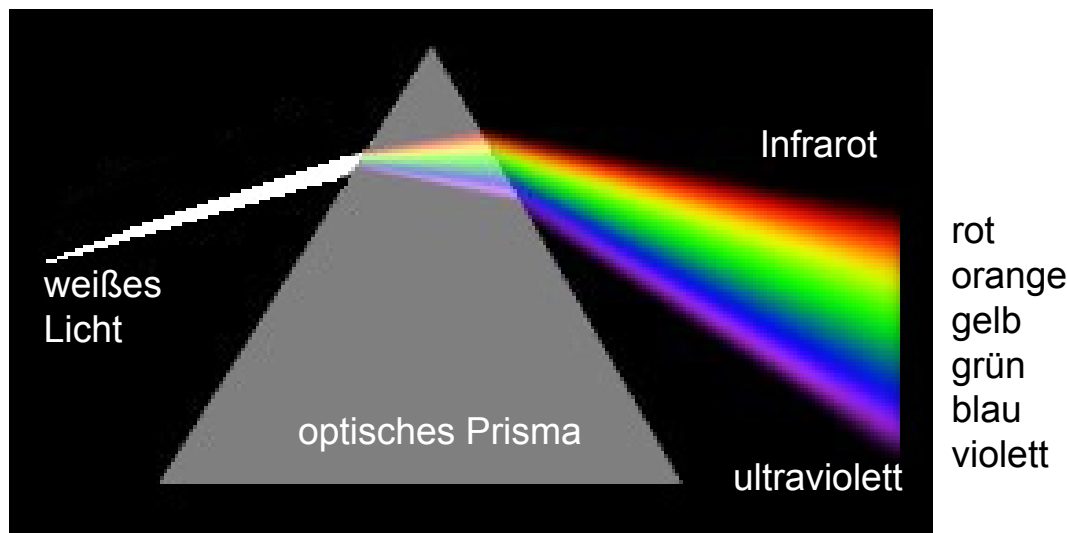
Quelle: Wikipedia.de

- **Farbraum:** Menge von Farben
- **Farbkörper:** Farbraum des menschlichen Auges
- **Farbmodell:** Beschreibt den Farbraum, den der Sehsinn oder ein Ein- oder Ausgabegerät (Display, Scanner, Drucker, Projektor, Foto, Kamera, Fernseher) darstellen kann.

Farbräume (II)

Grundlagen von Farben:

- Entdeckung 1666 von Newton: Sonnenlicht, das durch ein Prisma fällt, wird in ein kontinuierliches Farbspektrum aufgespalten.



Farbräume (III)

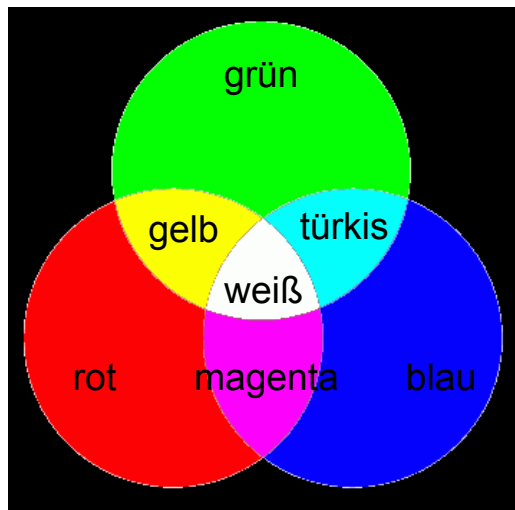
Aufbau eines Farbraumes

- Koordinatensystem, bei dem einzelne Farben die Achsen definieren (durch Aufbau des menschlichen Auges meistens 3 Dimensionen).
- Physikalische Farbmodelle (Mischung von Farben): RGB, CMYK
 - Anordnung als Würfel
 - Änderung einer Farbe → gleichzeitige Änderung der Helligkeit, Sättigung und des Farbtons
 - Im Jahr 1931 Definition der Primärfarben durch die CIE (International Commission on Illumination): rot (=435,8 nm), grün (=546,1 nm) und blau (=700 nm)
- Wahrnehmungsorientierte Farbmodelle (Beschreibung durch Helligkeit, Sättigung, Farbton): HSV, HSI
 - Beschreibung durch Zylinderkoordinaten (Winkel definiert die Farbe)

Farbräume (IV)

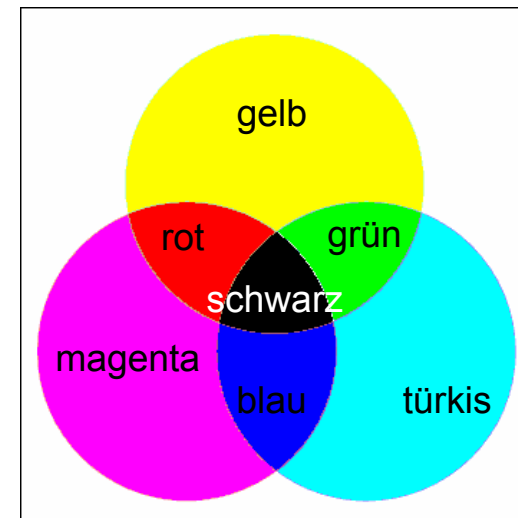
Additiver Farbraum

- Grundfarben addieren sich zu Weiß
- Bsp: Displays oder Beamer nutzen die Grundfarben RGB



Subtraktiver Farbraum

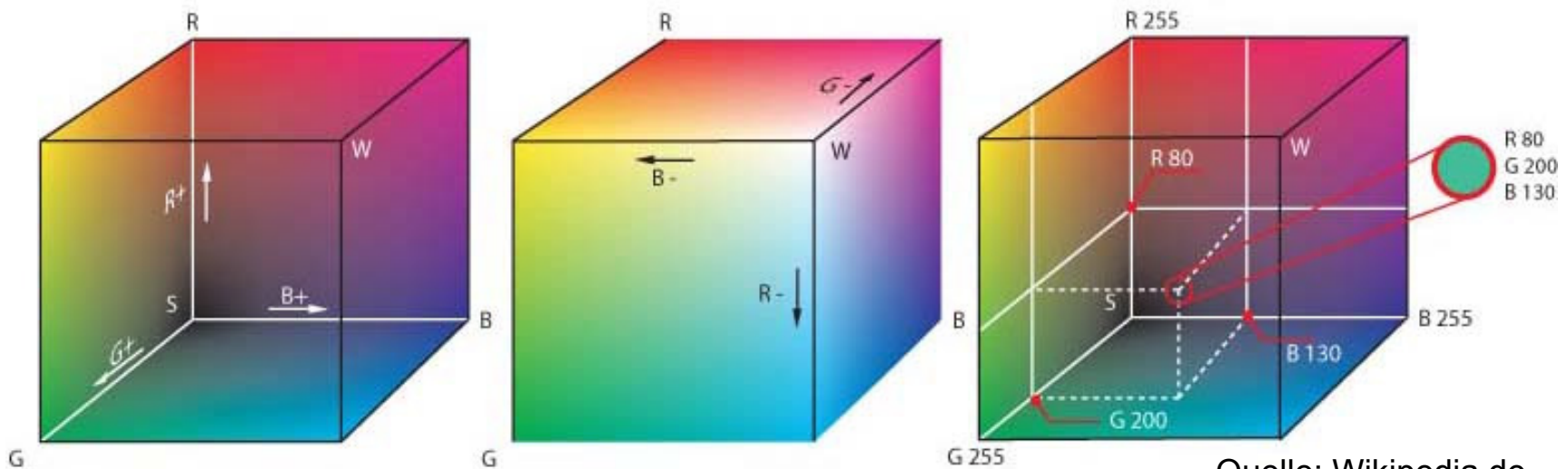
- Von Weiß werden Grundfarben subtrahiert
- Bsp: Tintenstrahldrucker (Pixel absorbieren weißes Licht), Folien vor weißer Lampe filtern einzelne Farbkomponenten



Farbräume (V)

RGB-Farbraum

- Idee: aus farbigem Licht dreier Primärfarben kann man beliebige Farben mischen
- Additiver Farbraum (Farben addieren sich zu Weiß)
- 8 Bit / 16 Bit pro Farbkanal

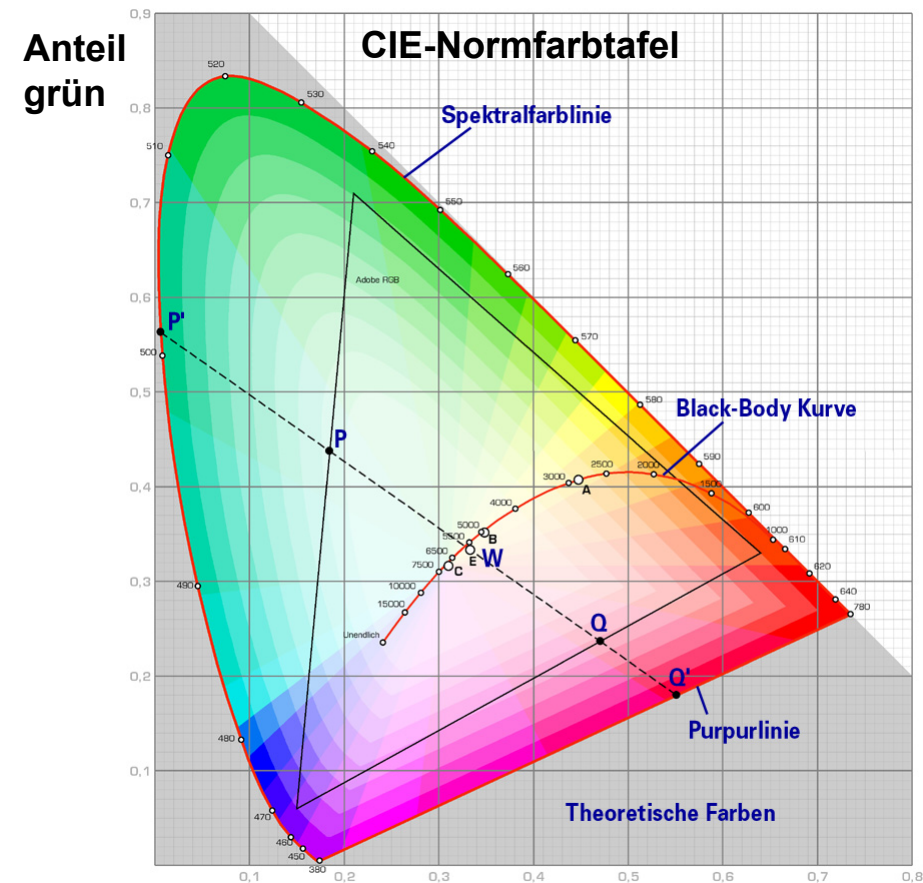


Quelle: Wikipedia.de

Farbräume (VI)

RGB-Farbraum

- Parabelförmig begrenzt (theoretische darstellbare Farben)
- Reine (vollständig gesättigte) Farben liegen auf der Spektralfarblinie
- Mischfarben: im Inneren des Diagrams
- Mittlerer Weißpunkt: W
- Mischfarben (Kombination zweier Farben) liegen auf einer Geraden definiert durch die beiden Farben
- Drei Farben spannen den RGB-Farbraum des Computers auf (schwarzes Dreieck)
- Ecken werden durch Darstellungsmöglichkeiten von Monitoren definiert



Anteil rot + Anteil grün + Anteil blau = 1 Anteil rot
→ Anteil blau = 1 - Anteil rot - Anteil grün

Quelle: Wikipedia.de

UNIVERSITY OF
MANNHEIM

Farbräume (VII)

Farbräume bei Displays

- Die beiden Displays decken einen sehr ähnlichen Farbraum ab; sie stellen aber unterschiedliche maximale Helligkeiten dar.
- Das Gerät mit der geringeren Helligkeit scheint einen geringeren Farbbereich abzudecken.



Quelle: Gabriel Marcu, Apple.

Farbräume (VIII)

Vergleich von CRT- und TFT-Displays

identische Farben



Farbräume (IX)

Vergleich von CRT- und TFT-Displays (gleiche Farbbereiche)



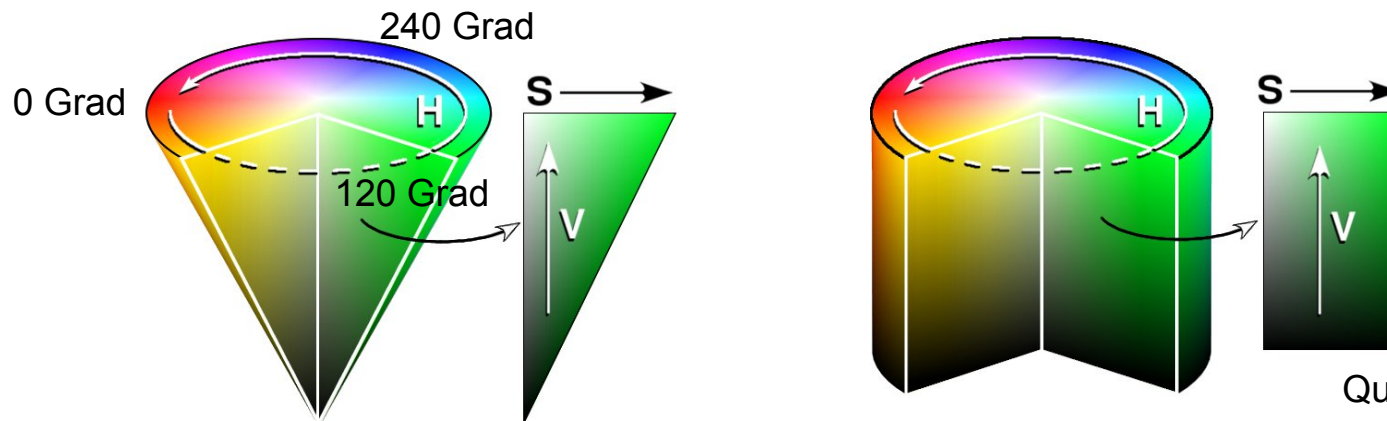
Quelle: Gabriel Marcu, Apple.

UNIVERSITY OF
MANNHEIM

Farbräume (X)

HSV-Farbraum:

- Hue (Farbe): dominante Wellenlänge (dominante Farbe, mit der ein Mensch ein Objekt beschreibt)
- Saturation (Sättigung): Beschreibt wie kräftig eine Farbe ist, d.h. das Mischungsverhältnis einer Farbe mit einer anderen Farbe (welche Wellenlängen sind neben der dominanten Wellenlänge vorhanden?).
- Value (Intensität/Farbwert): Helligkeit einer Farbe



Quelle: Wikipedia.de

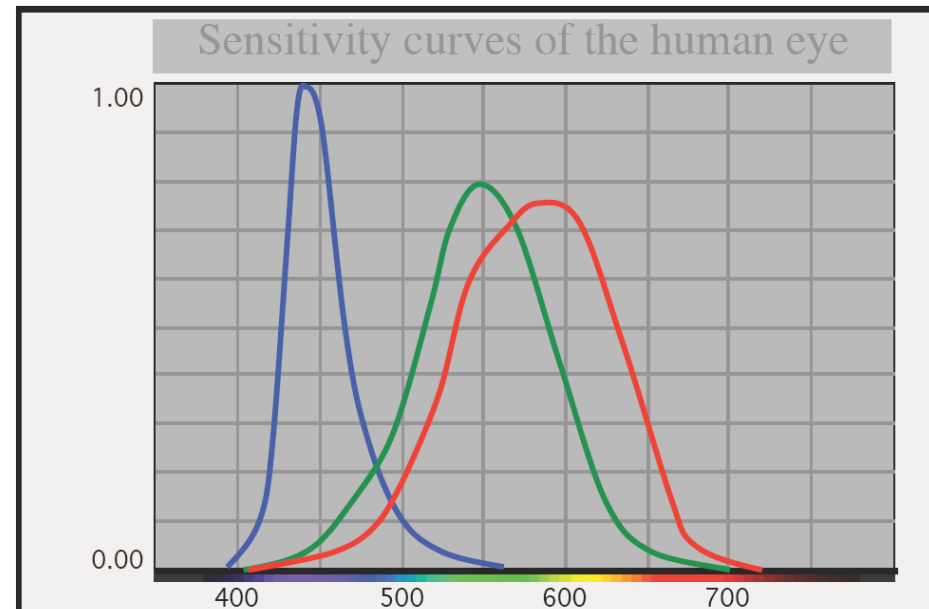
Menschliche Wahrnehmung (I)

Die Netzhaut des menschlichen Auges besteht aus:

- Zapfen (cone cells): Erkennung von Farben
- Stäbchen (rod cells): Erkennung von Helligkeitswerten

Zapfen enthalten drei Arten von Farbpigmenten, die sensitiv gegenüber bestimmten Farben sind:

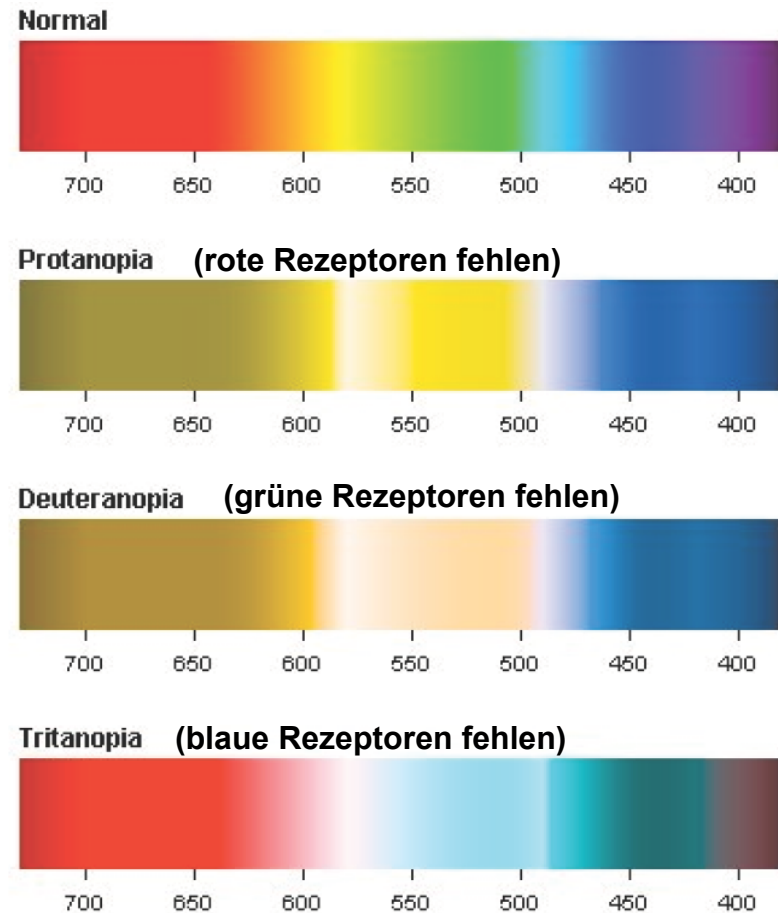
- niedrige Frequenzen: blaue Farbtöne
- mittlere Frequenzen: grüne/gelbe Farbtöne
- hohe Frequenzen: rote Farbtöne



Quelle: Gabriel Marcu, Apple

Menschliche Wahrnehmung (II)

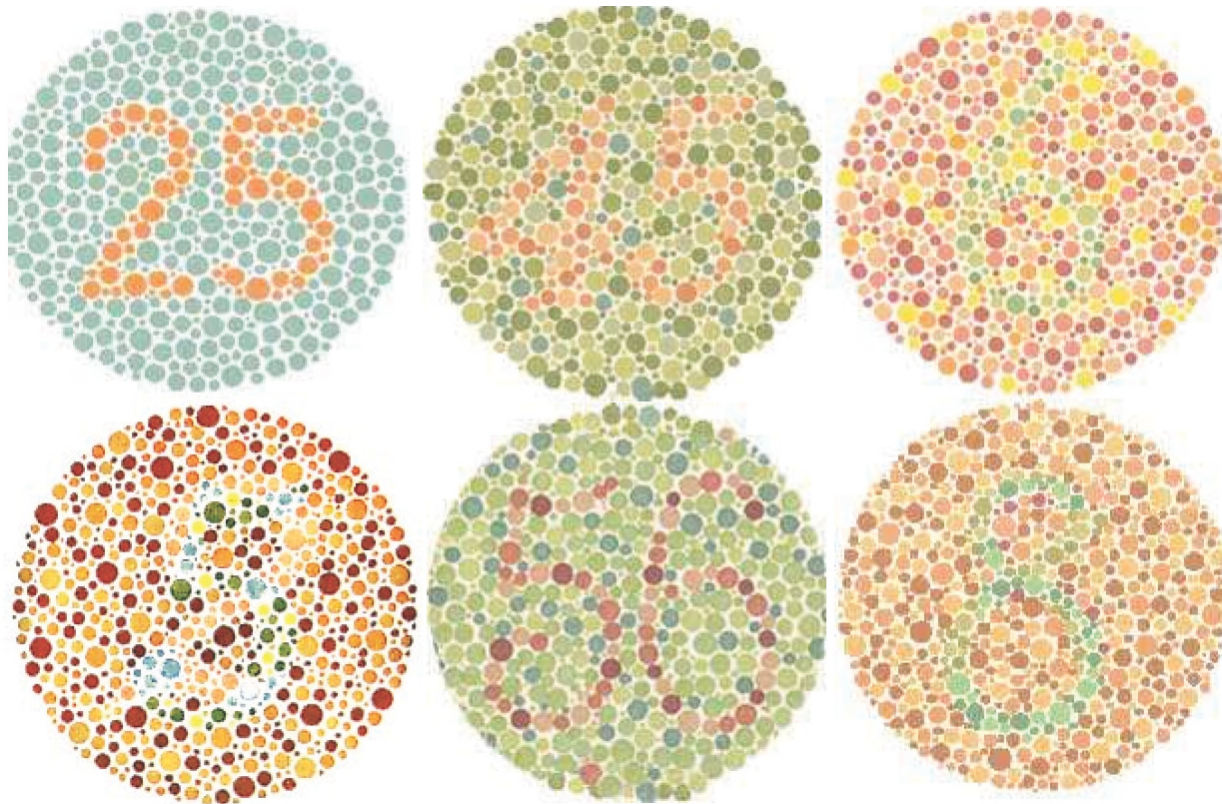
- Falls eins der Farbpigmente keine vollständigen Signale übermittelt, spricht man von Farbenblindheit
- 10 % der Bevölkerung können Farben nur eingeschränkt wahrnehmen



Quelle: Gabriel Marcu, Apple

Menschliche Wahrnehmung (III)

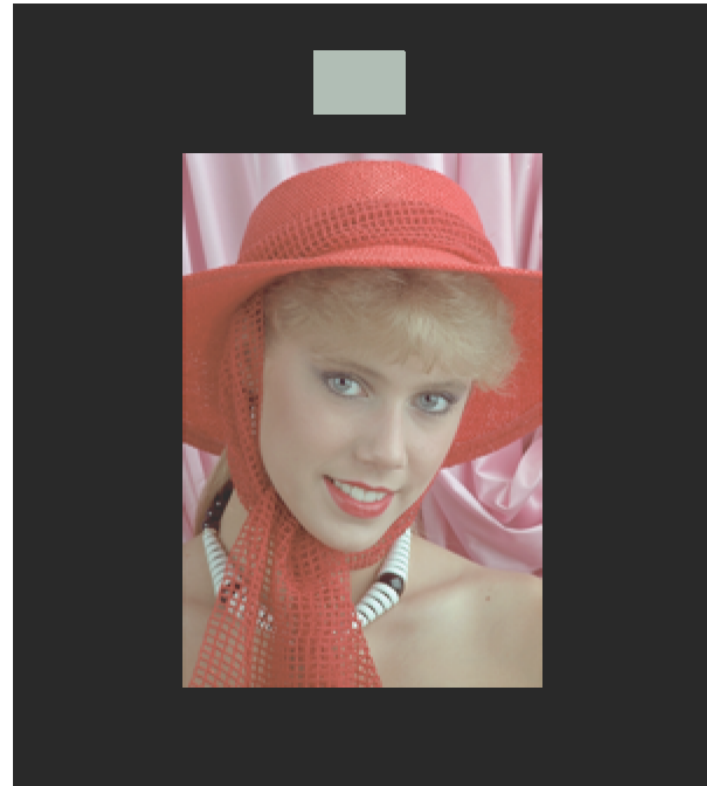
Ishihara-Test zur Erkennung von Farbenblindheit



Quelle: Gabriel Marcu, Apple

Menschliche Wahrnehmung (IV)

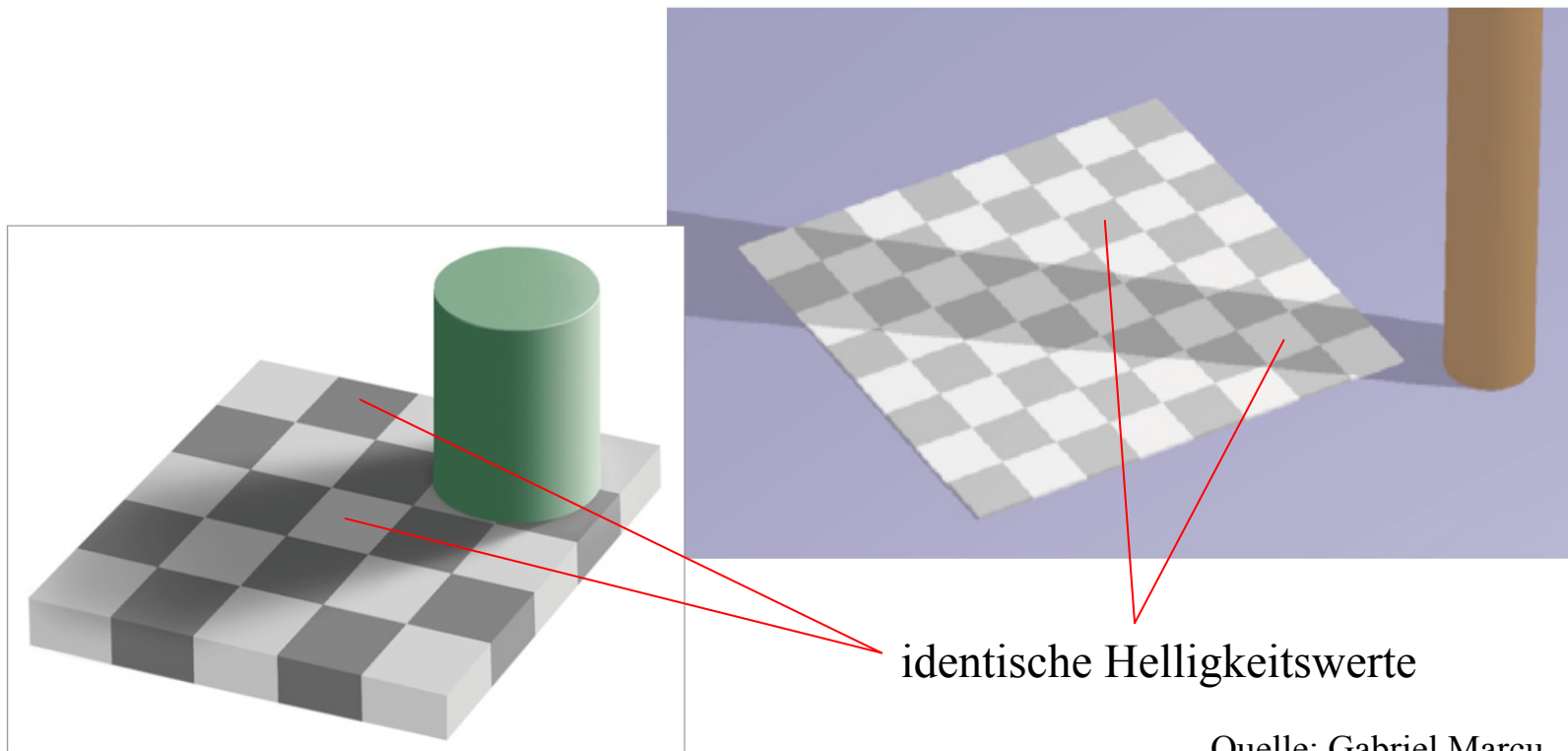
Empfundene Farbe wird durch ihre Umgebung beeinflusst



Quelle: Gabriel Marcu, Apple

Menschliche Wahrnehmung (V)

Helligkeiten werden durch Formen, Objekte und deren Schatten beeinflusst

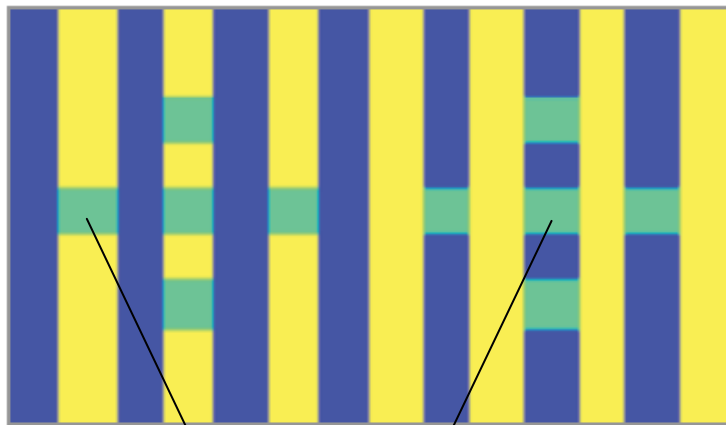


Quelle: Gabriel Marcu, Apple

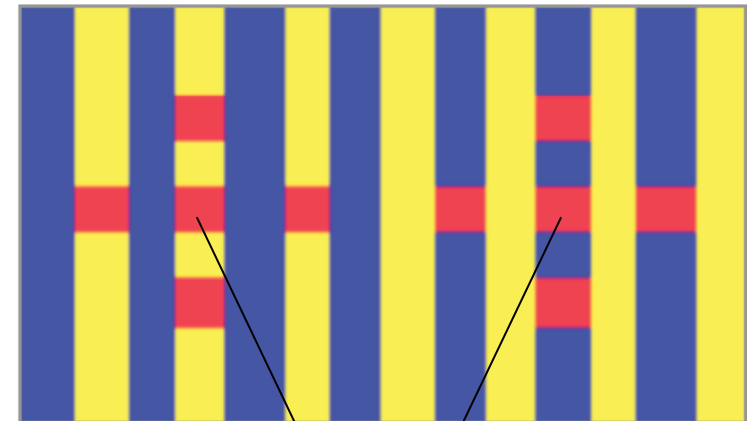
UNIVERSITY OF
MANNHEIM

Menschliche Wahrnehmung (VI)

Farben werden durch benachbarte Farbregionen beeinflusst



identische Farbe



identische Farbe

Quelle: Gabriel Marcu, Apple

Objekterkennung mittels Farben (I)

Erkennung von Farben beim Menschen

- Signale der drei Farbpigmente der Netzhaut
- Psycho-physikalische Einflussfaktoren (Umgebung, absolute Helligkeit, Textur, Form, Rauschen, Größe)
- Psychologische Einflussfaktoren (Erinnerung an Farben, Farbpräferenzen)

Erkennung von Farben durch den Rechner

- Sensoren (Scanner, Kamera, Messgeräte)
- Farbkorrektur, Anpassung der Helligkeit
- Analysealgorithmen

Objekterkennung mittels Farben (II)

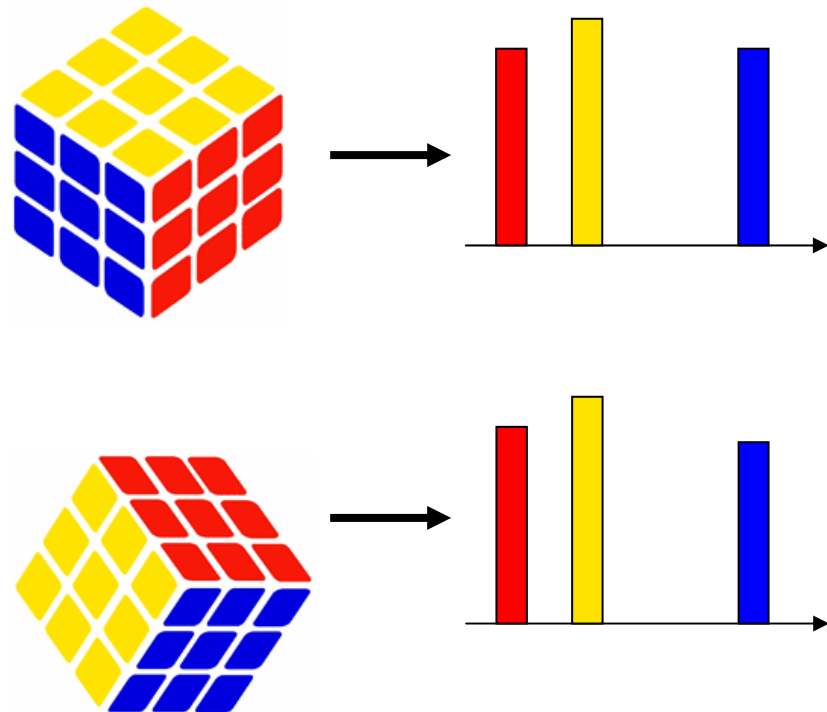
Ablauf

- Auswahl eines Farbraumes
- Abbildung der Pixel eines Objektes auf Merkmalswerte
 - Histogramm
 - Dominante Farbe
- Vergleich mit Merkmalswerten bekannter Objekte durch ein Entfernungsmaß

Objekterkennung mittels Farben (III)

Histogramme

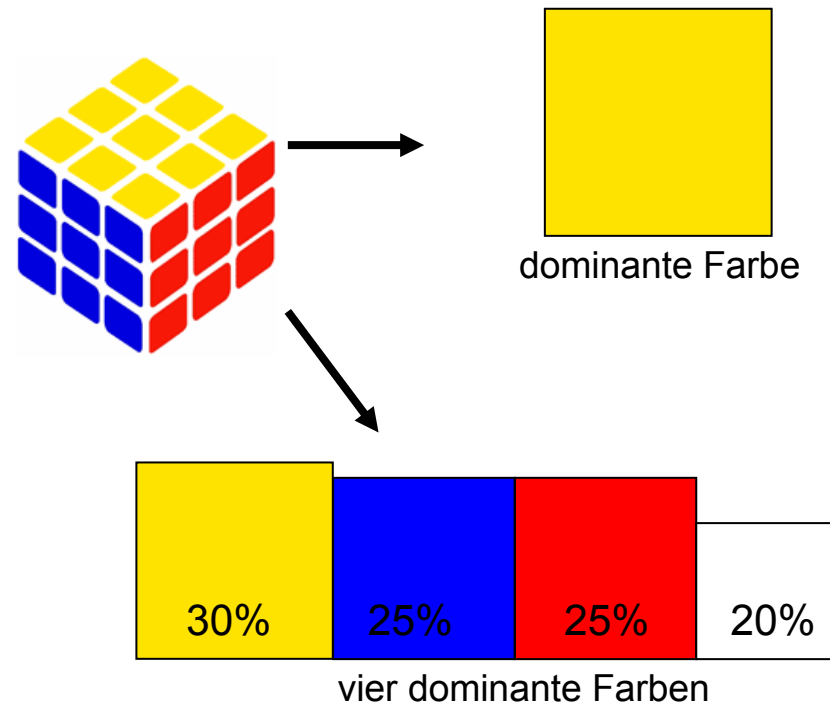
- Beschreibung eines Objektes durch Anzahl der Pixel der einzelnen Farben
- Vergleich zweier Histogramme mittels L1- oder L2-Norm
- Zuverlässige Ergebnisse bei affinen Transformationen des Objektes



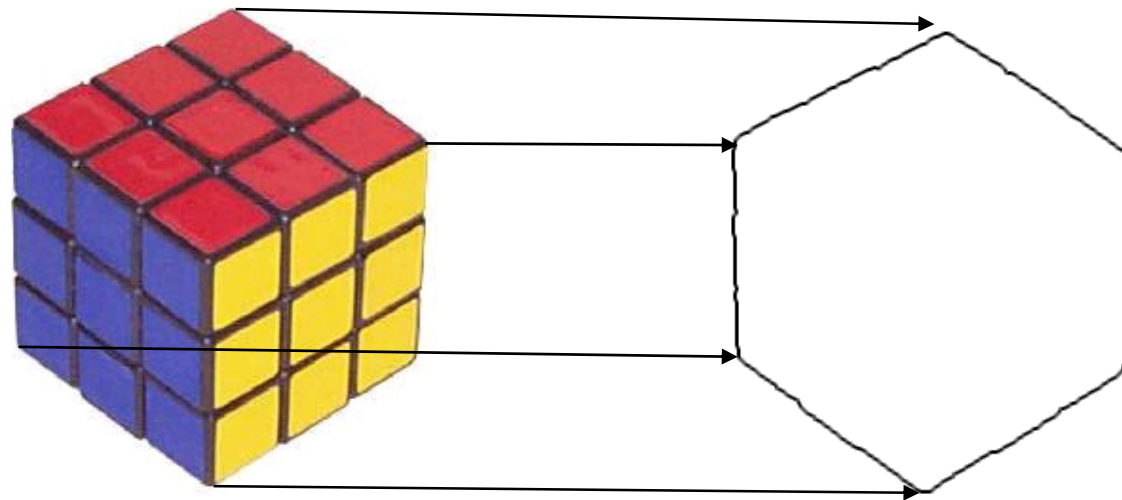
Objekterkennung mittels Farben (IV)

Dominante Farbe

- **Annahme:** Die am häufigsten in einem Bild enthaltene Farbe liefert charakteristische Informationen.
- Problem, falls mehrere Farben ähnlich häufig auftreten
→ Beschreibung eines Objektes durch wenige dominante Farben und den Anteil jeder Farbe



Objekterkennung durch Vergleich von Konturen (I)

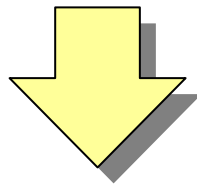


Ablauf der Objekterkennung

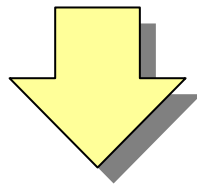
1. Parametrisierung (Abtastung) der Kontur mit einer vordefinierten Anzahl an Konturpunkten
2. Identifiziere Merkmalspunkte zur Beschreibung der Kontur
 - Kompaktheit
 - Exzentrizität
 - Curvature-Scale-Space-Merkmale
3. Vergleiche die Merkmale mit Merkmalen von bekannten Objekten (Shape Matching)

Parametrisierung einer Kontur

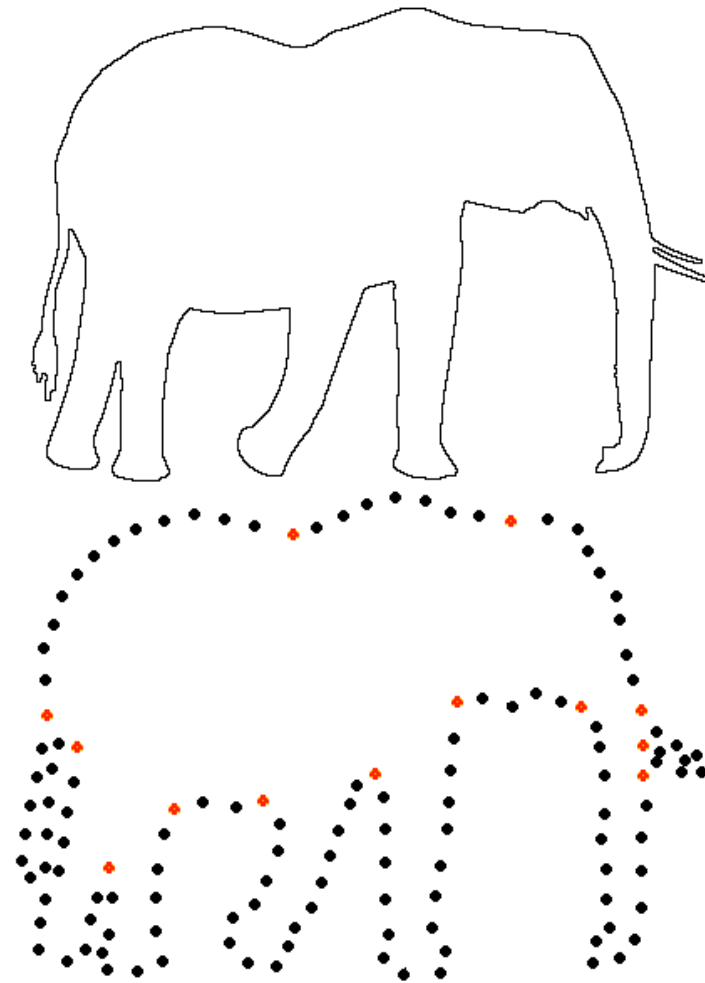
Segmentierung



Parametrisierung
der Kontur



Berechnung der
Konturmerkmale



Vergleich von Konturen (I)

Kompaktheit (*compactness*)

- Globaler Konturdeskriptor, der einen aggregierten Wert für die komplette Kontur liefert (beschreibt die Ähnlichkeit einer Kontur mit einem Kreis)
- Einfach zu berechnen
- Eignet sich nur für eine grobe Abschätzung der Ähnlichkeit zweier Konturen
- Invariant gegenüber geometrischen Transformationen (Rotation, Skalierung)
- Der Wert für die Kompaktheit ist bei einem Kreis minimal

Vergleich von Konturen (II)

Kompaktheit (*compactness*)

- c_i : Kompaktheit
- i : segmentiertes Objekt
- U : Länge der Kontur
- F : Fläche des Objektes

$$c_i = \frac{U^2}{4 \cdot \pi \cdot F}$$

- Unterschiede zwischen zwei Konturen i und j

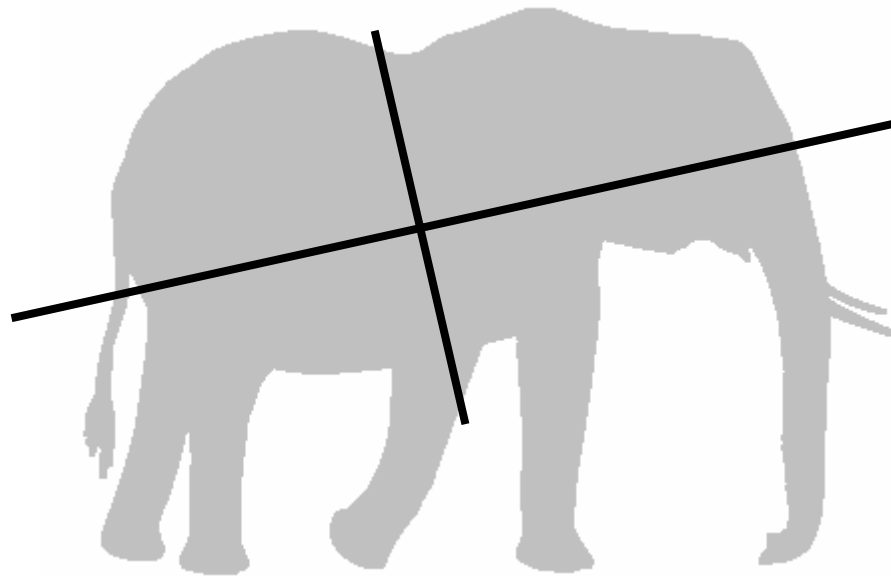
$$\alpha_c(i, j) = \frac{|c_i - c_j|}{\max(c_i, c_j)}$$

Vergleich von Konturen (III)

Exzentrizität (*eccentricity*)

Verhältnis der Längen der Hauptachsen bezogen auf die zentralen Momente der Konturpixel

→ invariant gegenüber geometrischen Transformationen.



Vergleich von Konturen (IV)

Exzentrizität (*eccentricity*)

zentralen Momente der Konturpixel

$$M_{n,m} = \sum_{x,y} (\bar{x} - x(u))^n (\bar{y} - y(u))^m$$

Schwerpunkt der Konturpixel

$$\bar{x} = \frac{1}{N} \sum_{u=0}^{N-1} x(u) \quad \text{und} \quad \bar{y} = \frac{1}{N} \sum_{u=0}^{N-1} y(u)$$

Vergleich von Konturen (V)

Exzentrizität (*eccentricity*)

e_i : Exzentrizität

$$e_i = \frac{(M_{2,0} - M_{0,2})^2 + 4 \cdot M_{1,1}}{F}$$

Vergleich zweier Objekte i und j :

$$\alpha_e(i, j) = \frac{|e_i - e_j|}{\max(e_i, e_j)}$$

Vergleich von Konturen (VI)

Skalenraumabbildung (*curvature scale space*)

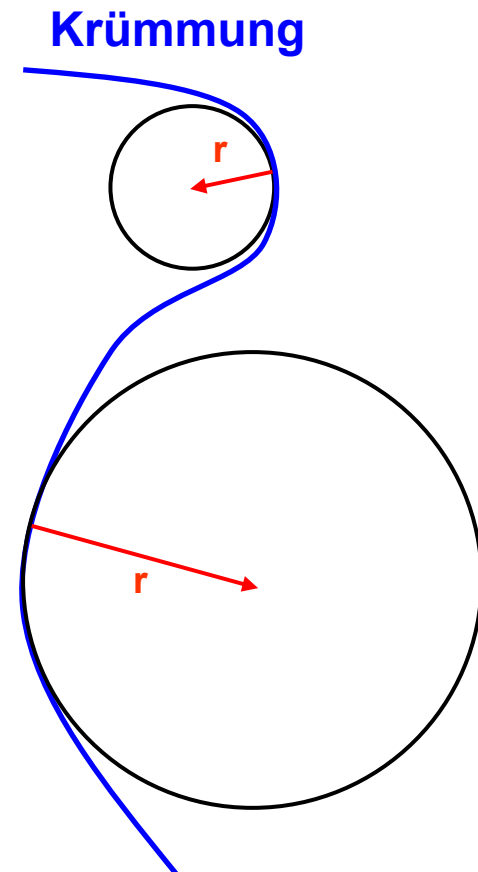
- Beschreibung der Krümmung der Kontur eines Objektes
 - Stark und lang gekrümmte Bereiche sollen durch hohe Merkmalswerte beschrieben werden
 - Die Größe eines Objektes soll keinen Einfluss auf die Merkmalswerte haben
- Vergleich von Konturen mittels Skalenraumabbildungen

Definition der Krümmung (I)

- Die **Stärke der Krümmung** in einem Punkt entspricht dem Kehrwert des Radius eines sich anschmiegenden Kreises (der Kreis berührt die Kurve):

$$K = \frac{1}{r}$$

- Die Krümmung ist ein Vektor, der in die Richtung der Kreismittelpunktes zeigt
- Ein kleiner Kreis repräsentiert eine hohe Krümmung; eine Gerade hat eine Krümmung von Null



Definition der Krümmung (II)

- Gegeben ist eine planare Kurve $\mathbf{u}(t)$ innerhalb eines 2D Raumes. $\mathbf{u}(t)$ wird durch die Bogenlänge t parametrisiert.
- Die Kurve $\mathbf{u}(t)$ wird durch die beiden Funktionen $\mathbf{x}(t)$ und $\mathbf{y}(t)$ definiert:

$$\mathbf{u}(t) = (\mathbf{x}(t), \mathbf{y}(t)).$$

- Die Krümmung K für die Kurve $\mathbf{u}(t)$ ist definiert als:

$$K = \frac{\dot{x} \cdot \ddot{y} - \dot{y} \cdot \ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}$$

\dot{x} und \dot{y} definieren die erste Ableitung (Gradient),

\ddot{x} und \ddot{y} definieren die zweite Ableitung (Änderung des Gradienten)

Definition der Krümmung (III)

Definition der Krümmung für Funktionen:

- verwende explizite planare Kurven: $y = f(x)$.
- Krümmung im Punkt $(x, f(x))$:

$$K = \frac{f''(x)}{\left(1 + (f'(x))^2\right)^{3/2}}$$

Definition der Krümmung (IV)

Beispiel

- Beispielfunktion $\mathbf{u}(t) = (\mathbf{x}(t), \mathbf{y}(t)) = (t, t^2)$
Explizite Definition der Funktion: $\mathbf{y} = \mathbf{f}(\mathbf{x}) = \mathbf{x}^2$
- Krümmung basierend auf der **parametrisierten Kurve**:
Erste und zweite Ableitung: $\dot{\mathbf{x}} = 1, \ddot{\mathbf{x}} = 0, \dot{\mathbf{y}} = 2t, \ddot{\mathbf{y}} = 2$

$$K(t) = \frac{\dot{x} \cdot \ddot{y} - \dot{y} \cdot \ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}} = \frac{1 \cdot 2 - 2t \cdot 0}{(1^2 + (2t)^2)^{3/2}} = \frac{2}{(1 + 4t^2)^{3/2}}$$

- Krümmung basierend auf der **expliziten Definition**:

$$f'(x) = 2x, \quad f''(x) = 2 \quad K(x) = \frac{f''(x)}{(1 + (f'(x))^2)^{3/2}} = \frac{2}{(1 + 4x^2)^{3/2}}$$

Definition der Krümmung (V)

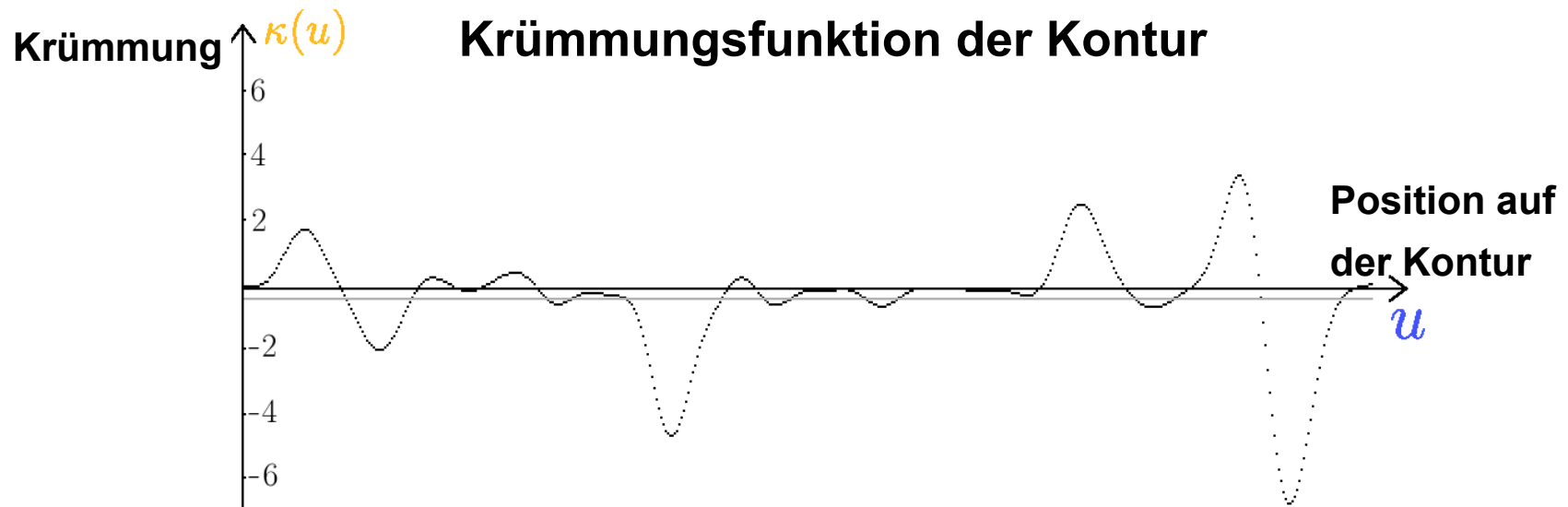
- Approximation der Ableitung für diskrete Werte (parametrisierte Kurve):

$$\dot{x}(t) = \frac{x(t+1) - x(t-1)}{2 \cdot hx}$$

$$\dot{y}(t) = \frac{y(t+1) - y(t-1)}{2 \cdot hy}$$

- Der Parameter t ist für ganze Zahlen definiert ($t \in \mathbb{N}$).
- hx und hy normalisieren die Ableitungen abhängig von der Entfernung der Abtastpunkte

Konturvergleich mit Hilfe von Krümmungen

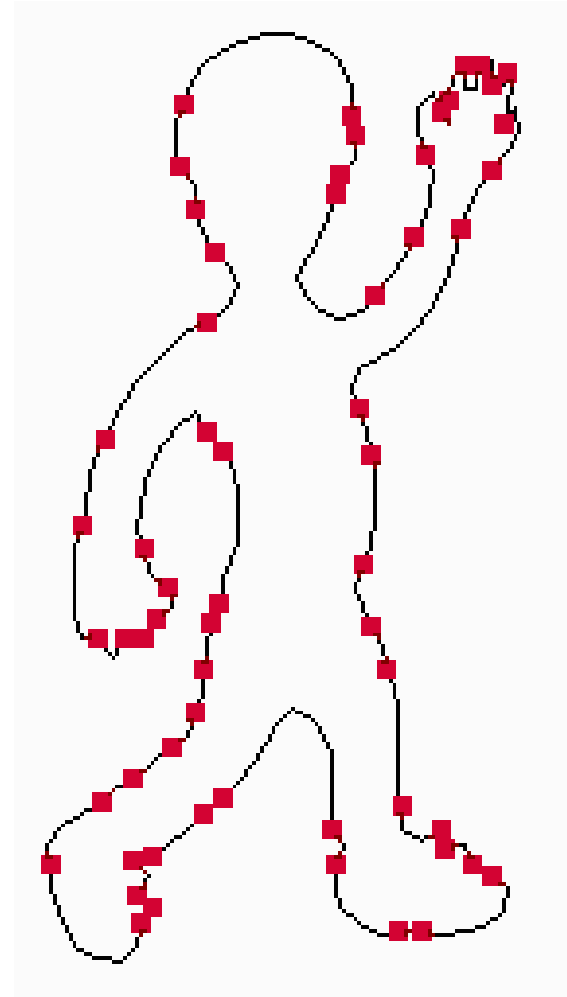


Problem: Zwei Krümmungsfunktionen können nur sehr schwer miteinander verglichen werden

- Identifiziere signifikante Punkte der Krümmungsfunktion
- das Curvature-Scale-Space-Verfahren liefert gute Merkmalspunkte

Curvature-Scale-Space-Verfahren (I)

- Analysiere äußere Kontur eines Objektes
- Glätte Kontur (Gaußfilter)
- Merkmalspunkte: Wendepunkte der Kontur



Curvature-Scale-Space-Verfahren (II)

Ablauf

1. Iterative Glättung der Kontur
2. Berechnung der Krümmung
3. Betrachte Nullstellen der Krümmungsfunktion
(entspricht den Wendepunkten der Kontur) :

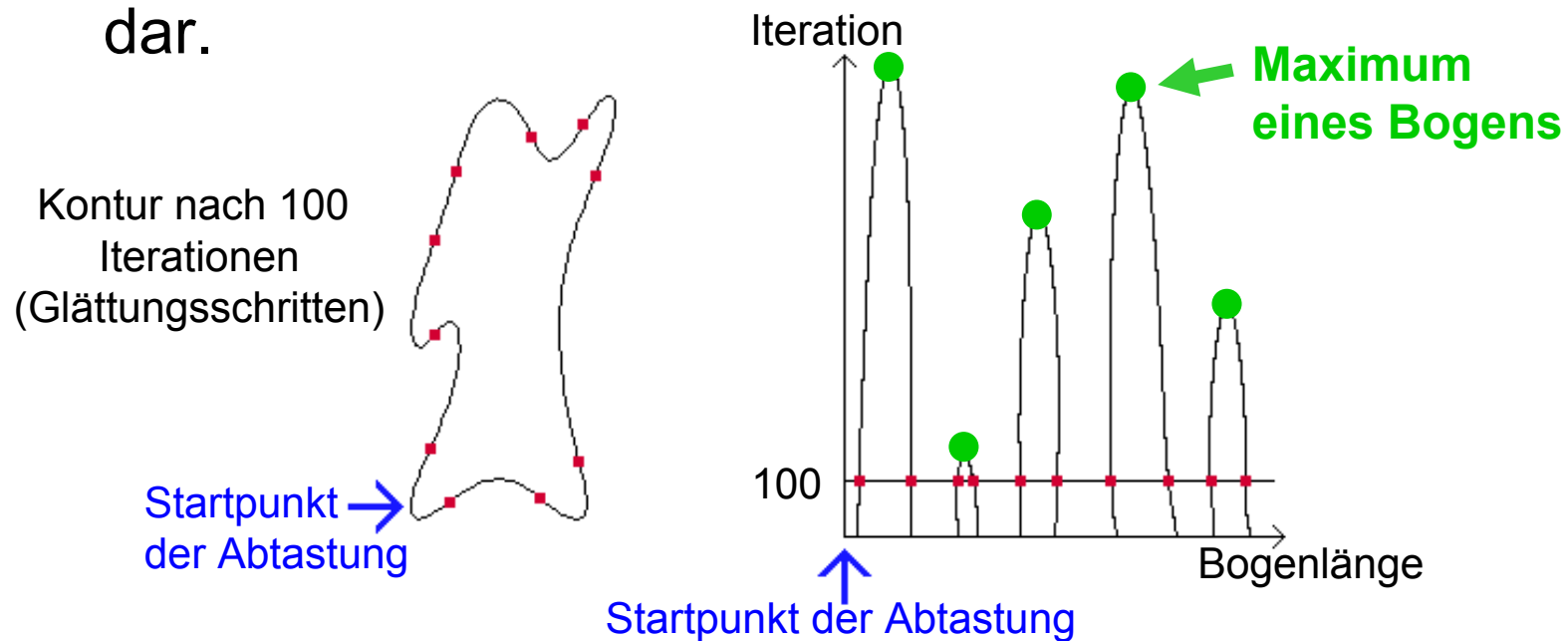
$$\kappa(u, n) = 0$$

4. Erzeuge Skalenraumabbildung (Curvature-Scale-Space-Bild):

$$I(u, n) = \{(u, n) \mid \kappa(u, n) = 0\}$$

Curvature-Scale-Space-Verfahren (III)

- Ein Curvature-Scale-Space-Bild (CSS-Bild) stellt die Wendepunkte, die bei der Glättung entstehen, visuell dar.



Die lokalen Maxima der **Bögen (peaks)** werden als Merkmale zur Beschreibung des Objektes verwendet

Curvature-Scale-Space-Verfahren (IV)

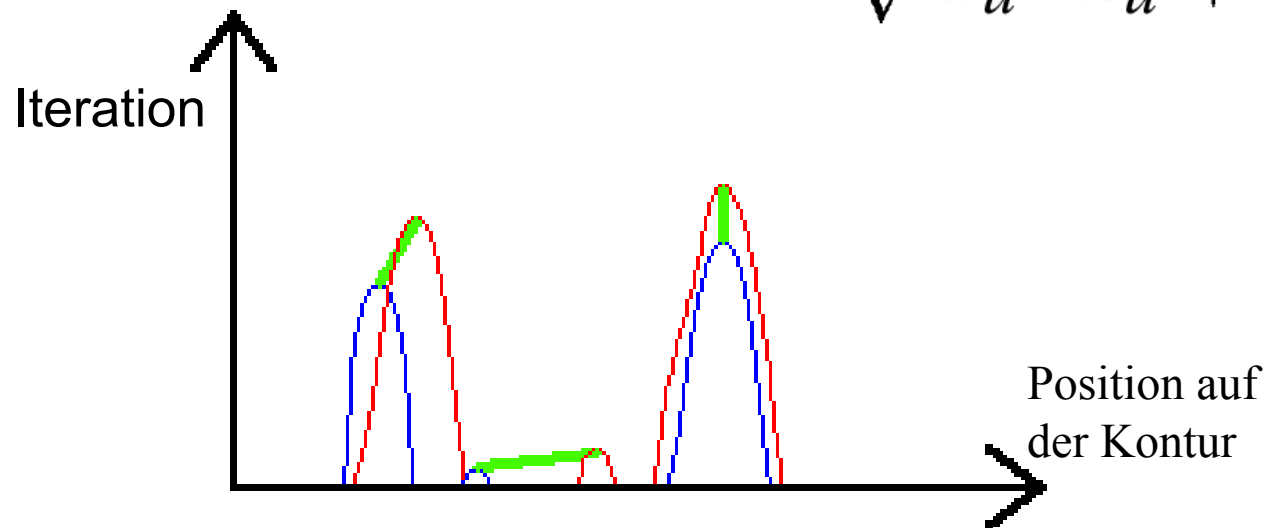
Merkmale von CSS-Bildern

- Ein Bogen im CSS-Bild beschreibt einen konkaven Bereich einer Kontur
- Die Bögen liefern die Merkmale zur Beschreibung einer Kontur.
- Jeder Bogen wird beschrieben durch
 - eine Position:
relative Position im Vergleich zu den anderen Bögen
 - eine Höhe (Anzahl der erforderlichen Glättungen):
Stärke eines konkaven Segmentes

Vergleich zweier Konturen

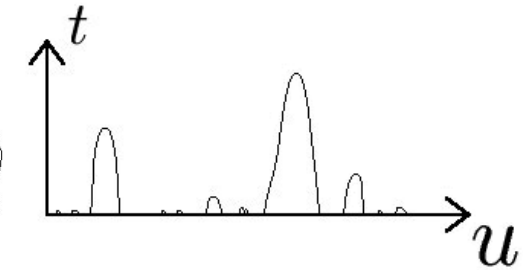
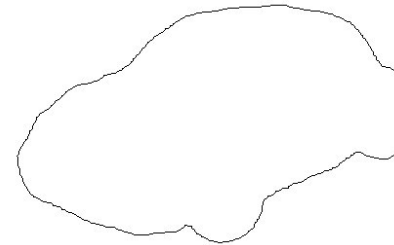
- Verschiebe ein CSS-Bild horizontal bis die höchsten Bögen übereinstimmen (Verfahren wird invariant gegenüber Rotationen)
- Berechne die Euklidische Distanz zwischen zwei lokalen Maxima im Skalenraumbild
- Summiere die Distanzen

$$D = \sqrt{d_u \cdot d_u + d_t \cdot d_t}$$

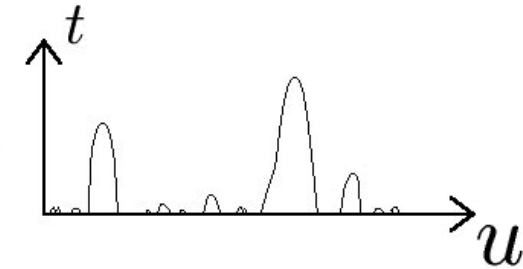
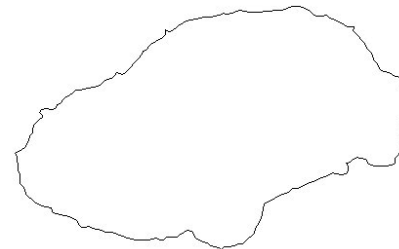


Merkmale von CSS-Bildern (I)

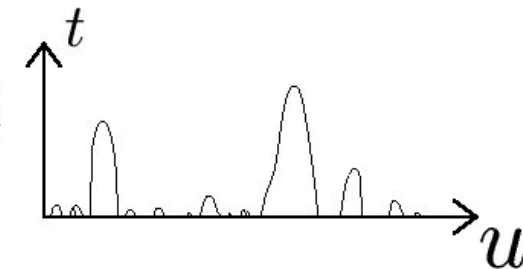
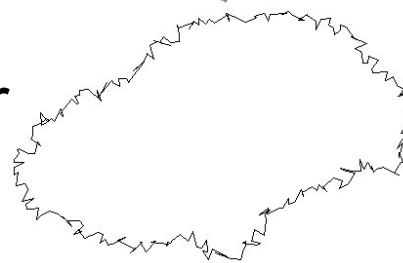
Originalkontur



Verrauschte Kontur



Stark verrauschte Kontur



→ Trotz Rauschen sind die CSS-Bilder sehr ähnlich

Merkmale von CSS-Bildern (II)

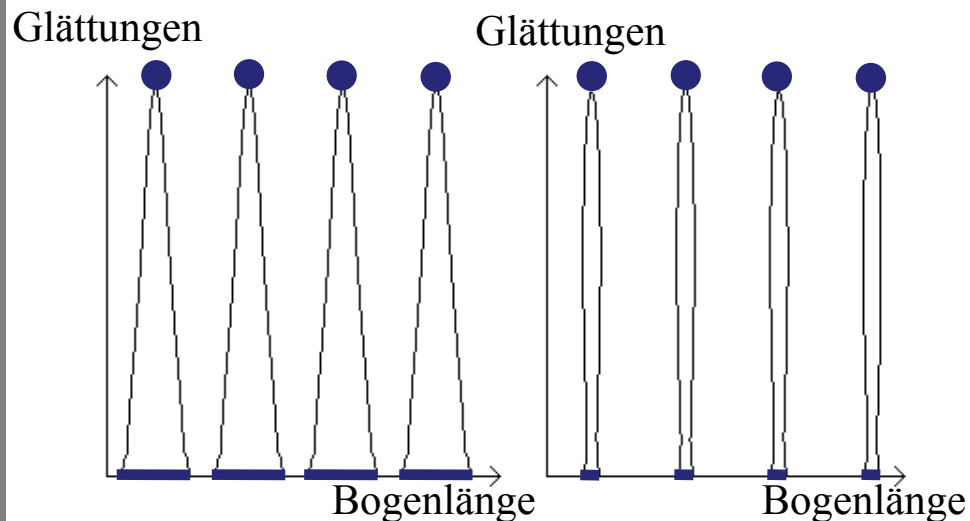
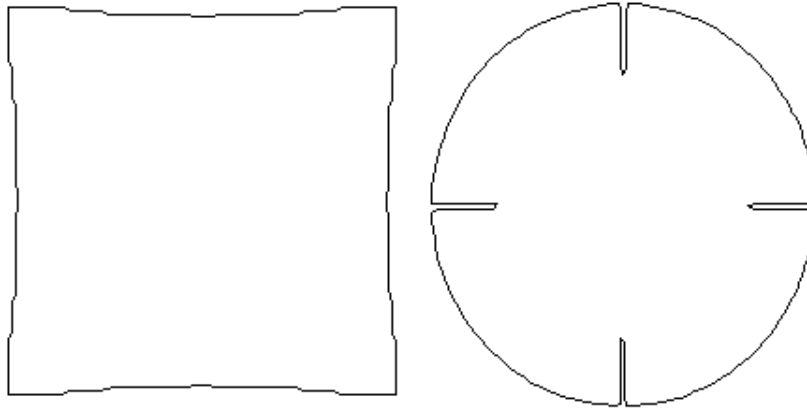
Vorteile

- Gute Klassifikationsergebnisse
- Nur wenige Werte werden zur Beschreibung komplexer Objekte benötigt
- Invariant gegenüber Rotationen / Skalierungen
- Robust bei Rauschen und perspektivischen Verzerrungen
- Schnelle Berechnung der Merkmale
- Schneller Vergleich zweier Konturen

Nachteile

- Schlechte Klassifikationsergebnisse bei einzelnen Konturen (Mehrdeutigkeiten, konvexe Objektregionen)

Mehrdeutigkeiten von CSS-Bildern (I)



- Die Höhe eines Bogens beschreibt die Länge der gekrümmten Region und die Stärke der Krümmung
- Für jeden Bogen wird zusätzlich dessen **Breite** gespeichert

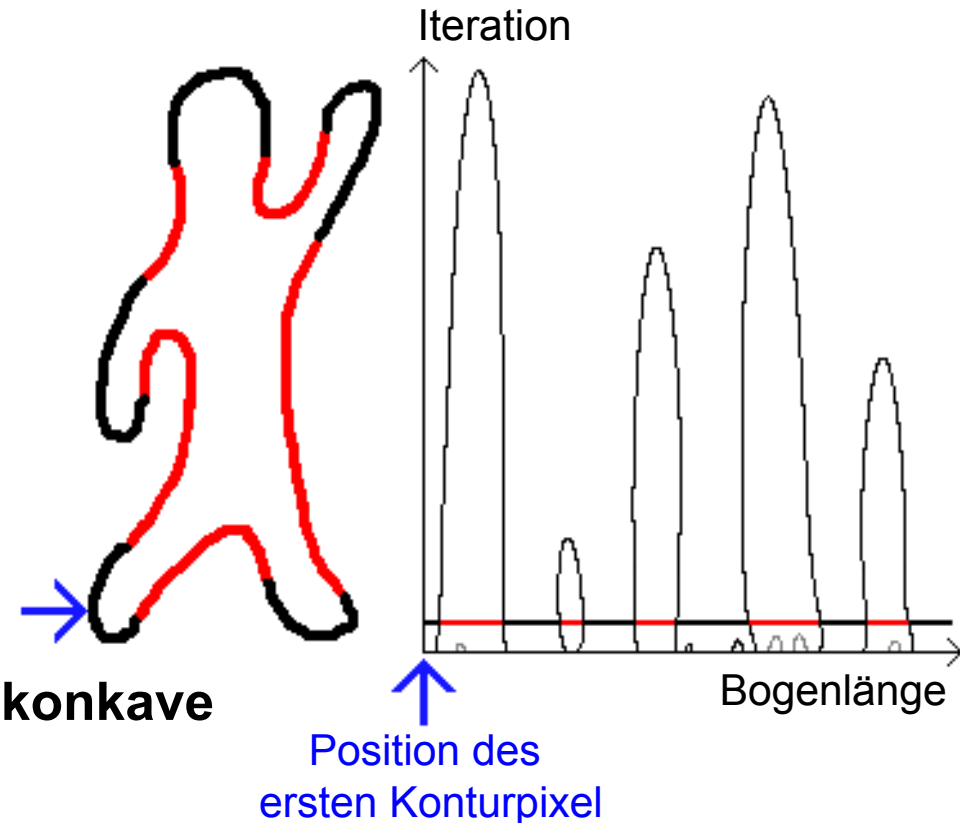
Mehrdeutigkeiten von CSS-Bildern (II)

Konvexe Regionen

- Schlechte Repräsentation konvexer Bereiche einer Kontur
- Konvexe Objekte können gar nicht unterschieden werden

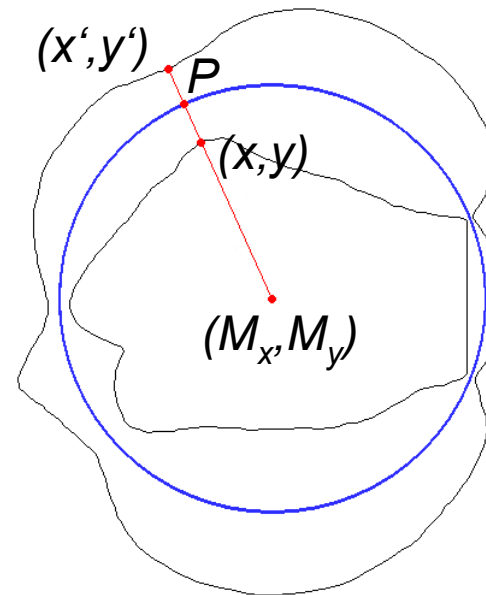
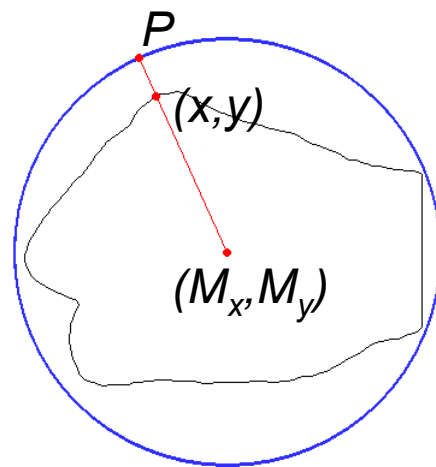
→ **Lösung:**
Umwandlung konvexer in konkave Bereiche und umgekehrt

→ **Verfahren:**
Spiegelung der Kontur

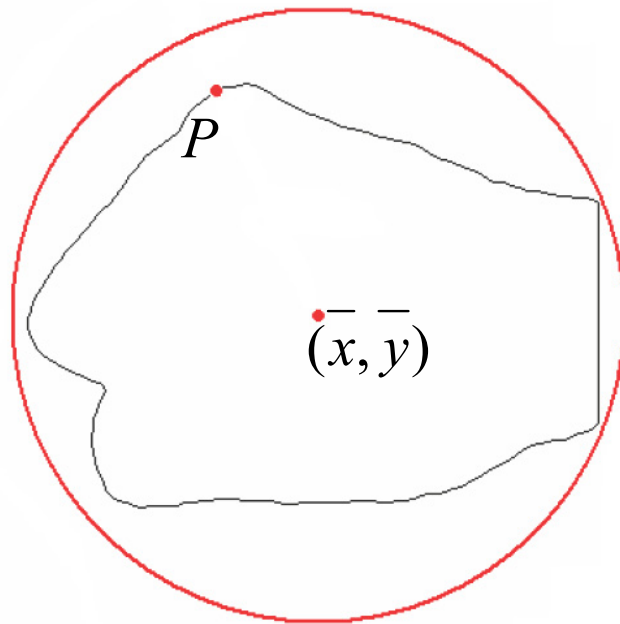


Gespiegelte Konturen (I)

- Idee: Erzeuge eine neue Kontur durch Spiegelung der Konturpixel an einem umgebenen Kreis
- Starke konvexe Regionen der ursprünglichen Kontur werden konkave Regionen in der gespiegelten Kontur



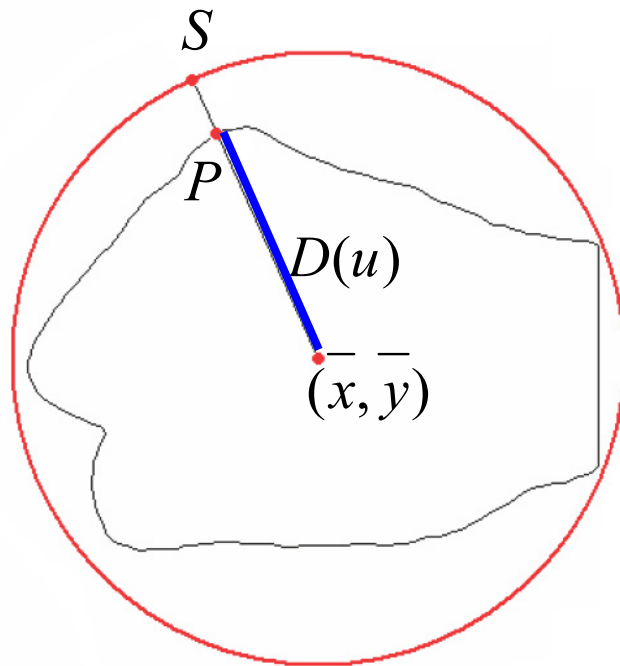
Gespiegelte Konturen (II)



(\bar{x}, \bar{y}) Schwerpunkt der Konturpixel
 P Konturpixel $(x(u), y(u))$

$$R = \max_u \{ \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2} \}$$

Gespiegelte Konturen (III)



(\bar{x}, \bar{y}) Schwerpunkt der Konturpixel

P Konturpixel $(x(u), y(u))$

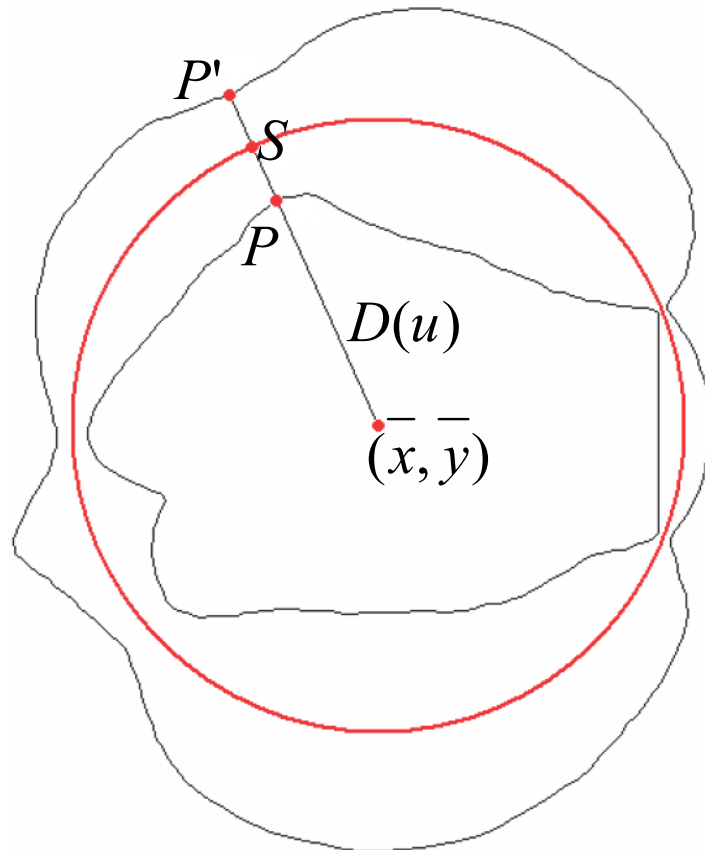
$D(u)$ Entfernung zum Mittelpunkt

S Punkt auf Kreislinie

$$R = \max_u \{ \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2} \}$$

$$D(u) = \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2}$$

Gespiegelte Konturen (IV)



(\bar{x}, \bar{y}) Schwerpunkt der Konturpixel

P Konturpixel $(x(u), y(u))$

$D(u)$ Entfernung zum Mittelpunkt

S Punkt auf Kreislinie

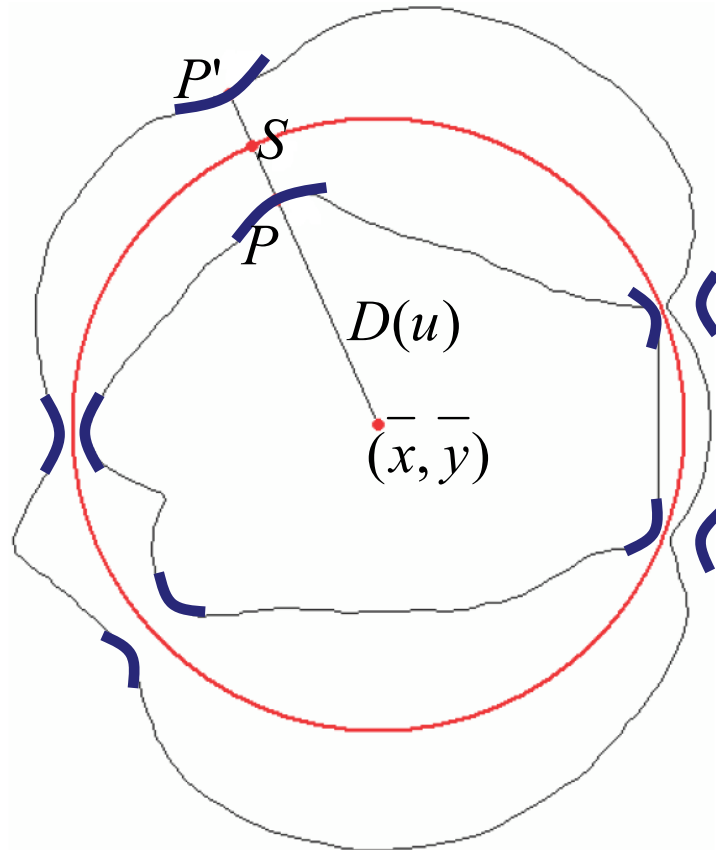
P' Gespiegeltes Pixel $(x'(u), y'(u))$

$$R = \max_u \left\{ \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2} \right\}$$

$$D(u) = \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2}$$

$$x'(u) = \frac{2R - D(u)}{D(u)} \cdot (x(u) - \bar{x}) + \bar{x}$$

Gespiegelte Konturen (V)



(\bar{x}, \bar{y}) Schwerpunkt der Konturpixel

P Konturpixel $(x(u), y(u))$

$D(u)$ Entfernung zum Mittelpunkt

S Punkt auf Kreislinie

P' Gespiegeltes Pixel $(x'(u), y'(u))$

$$R = \max_u \{ \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2} \}$$

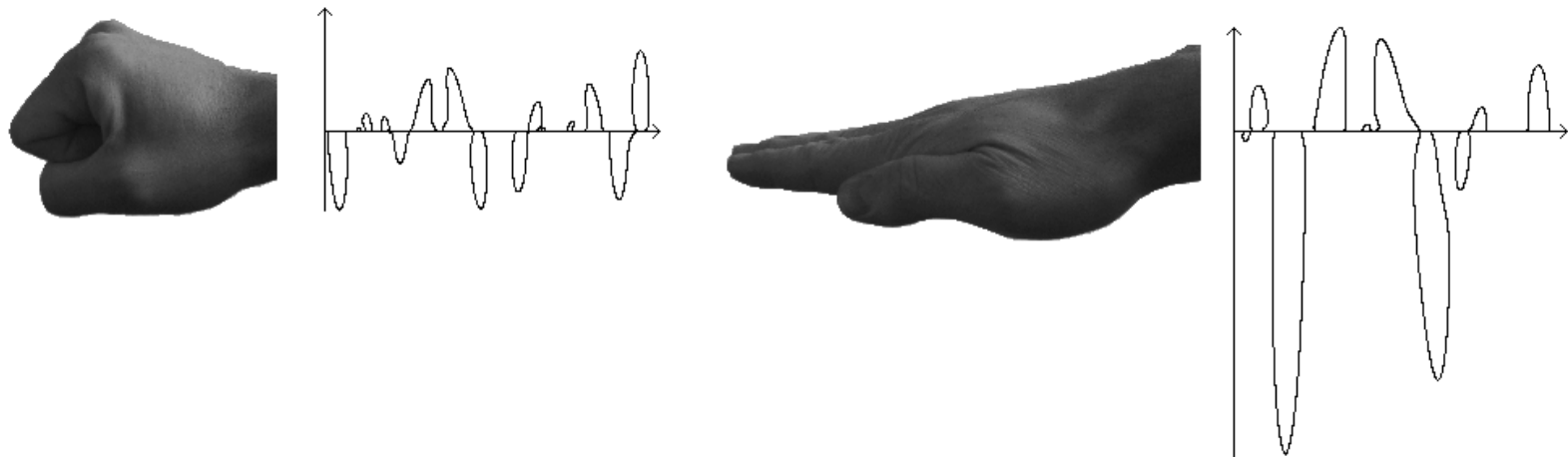
$$D(u) = \sqrt{(\bar{x} - x(u))^2 + (\bar{y} - y(u))^2}$$

$$x'(u) = \frac{2R - D(u)}{D(u)} \cdot (x(u) - \bar{x}) + \bar{x}$$

Gespiegelte Konturen (VI)

Erweiterter Konturvergleich

- Berechne ursprüngliche Merkmale der CSS-Bilder
- Berechne Merkmale für die gespiegelte Kontur.



Objekterkennung in Bildern (I)

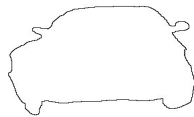
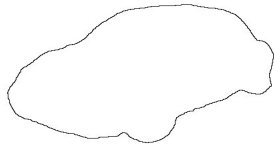
Ansatz

- Vergleiche unbekanntes Objekt mit bekannten Objekten einer Datenbank
- Berechne die durchschnittliche Differenz zwischen unbekanntem Objekt und allen Objekten einer Objektklasse (Zusammenfassung ähnlicher Objekte)
- Rückgabe des ähnlichsten Objektes

Objekterkennung in Bildern (II)

Objekte der Test-Datenbank

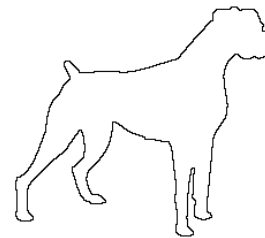
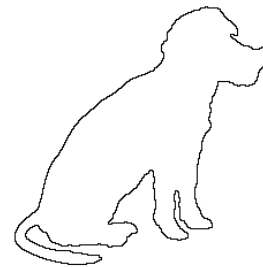
- Datenbank enthält mehr als 300 Objekte
- 13 Objektklassen gruppieren ähnliche Objekte



PKW



Person



Hund



Fisch

Objekterkennung in Bildern (III)

Kriterien zur Auswahl der Bilder der Datenbank: Kanonische Sichten

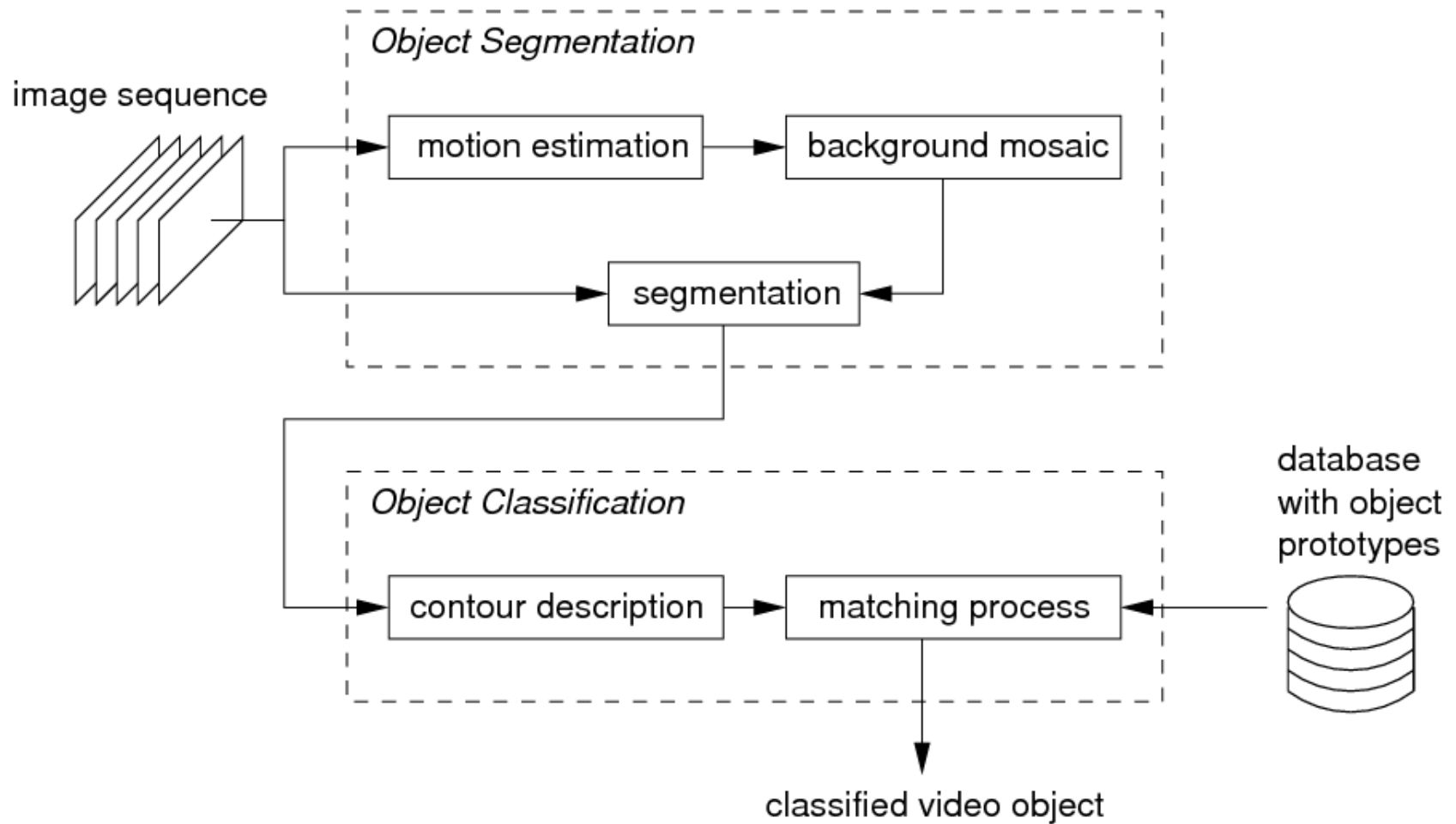
- Wie werden Objekte im menschlichen Gehirn repräsentiert?
→ unbekannt
- **aber:** dreidimensionale Objekte scheinen als zweidimensionale Ansichten abgebildet zu werden
- Die Drehung eines Objektes zur Kamera hat starken Einfluss, ob und wie schnell ein Mensch ein Objekt erkennt
- Einfach zu erkennende zweidimensionale Projektionen eines dreidimensionalen Objektes werden als kanonische Sichten (canonical view) bezeichnet

Objekterkennung in Bildern (IV)

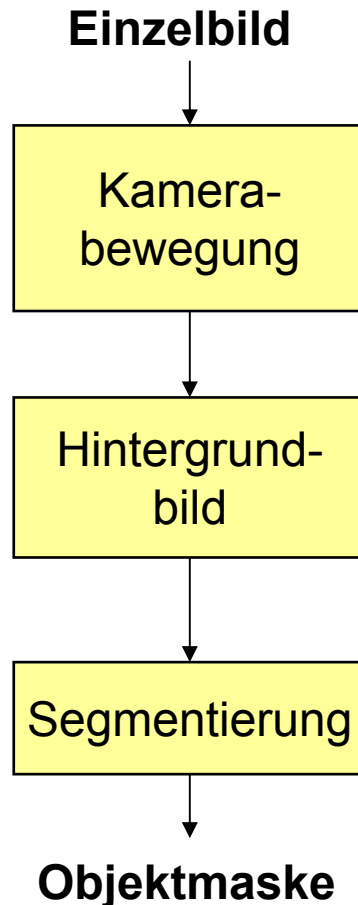
Beispiele für kanonische Sichten

- Ansichten im Profil oder leicht erhöhte Ansichten von schräg vorne
- Vertraute Perspektiven: Perspektiven, aus denen ein Objekt üblicherweise betrachtet oder im Fall von Gebrauchsgegenständen verwendet wird
- Als Objekte der Test-Datenbank wurden Objekte in kanonischen Sichten ausgewählt

Objekterkennung in Videos (I)



Objekterkennung in Videos (II)

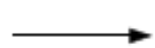
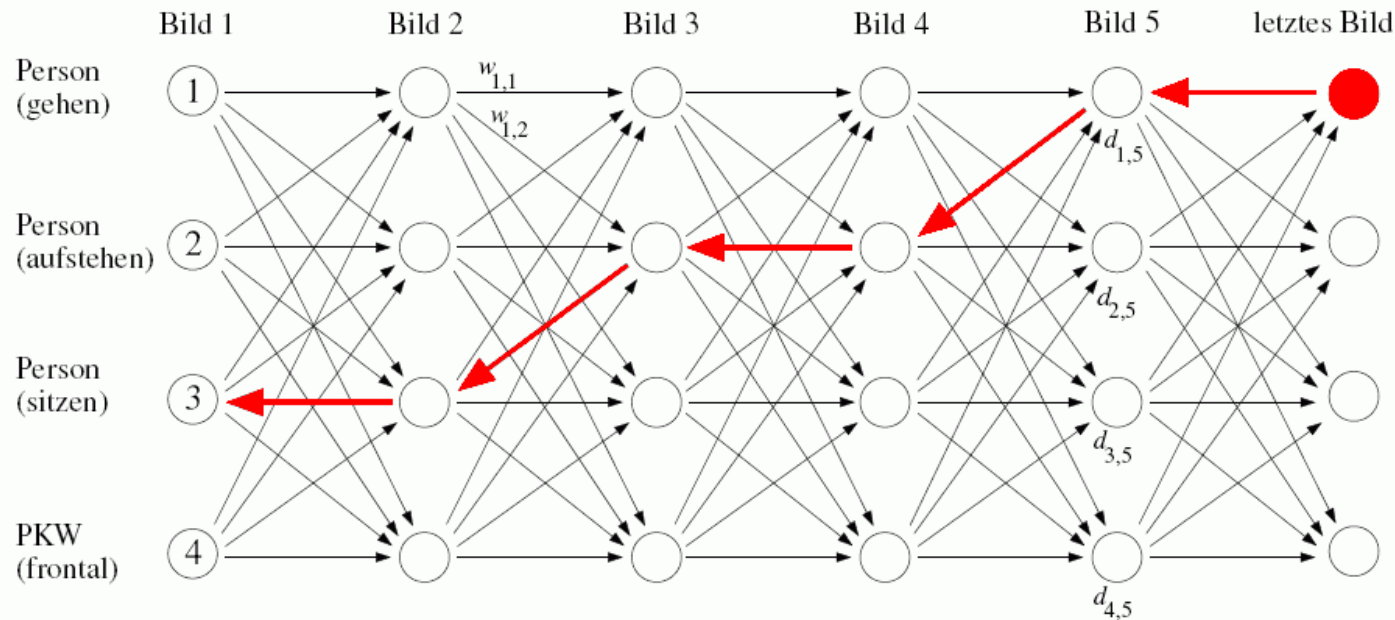


- **Annahme:** Die Hälfte aller Pixel eines Bildes sind Hintergrundpixel
- Berechne Kamerabewegung zwischen zwei benachbarten Bildern des Videos
- Richte alle Bilder einer Kameraeinstellung entsprechend des Kameramodells aus, so dass der Bildhintergrund deckungsgleich ist
- Verwende Medianfilter, um aus den transformierten Bildern ein Hintergrundbild zu erzeugen
- Vergleiche das Hintergrundbild mit den transformierten Einzelbildern

Objekterkennung in Videos (III)

Objekt-
klasse k

Eingabebild i der Kameraeinstellung



Übergänge zwischen Objektklassen: $w_{c_{i-1},i}$



Knoten d_{c_i} speichert Differenz des Objektes i zur Objektklasse c_i



Pfad mit minimalen Gesamtkosten



Letzter Knoten mit minimalen Kosten

$$K = \min_c \sum_{i=1}^N d_{c_i,i} + w_{c_{i-1},c_i}$$

Objekterkennung in Videos (IV)

Erkennungsraten

- Die Erkennungsraten liegen zwischen 25-95 % (abhängig von der Komplexität des Objektes und der Anzahl an verfügbaren Objekten in der Datenbank)
- Starre Objekte (z.B. Autos) werden deutlich zuverlässiger erkannt
- Das CSS-Verfahren ist invariant gegenüber Skalierungen und Rotationen; es ist sehr robust bei Rauschen
- Ein Vergleich von Konturen ist sehr effizient möglich (einmaliges Glätten, häufiger Vergleich mittels Euklidischer Distanz)

Objekterkennung in Videos (V)



standing

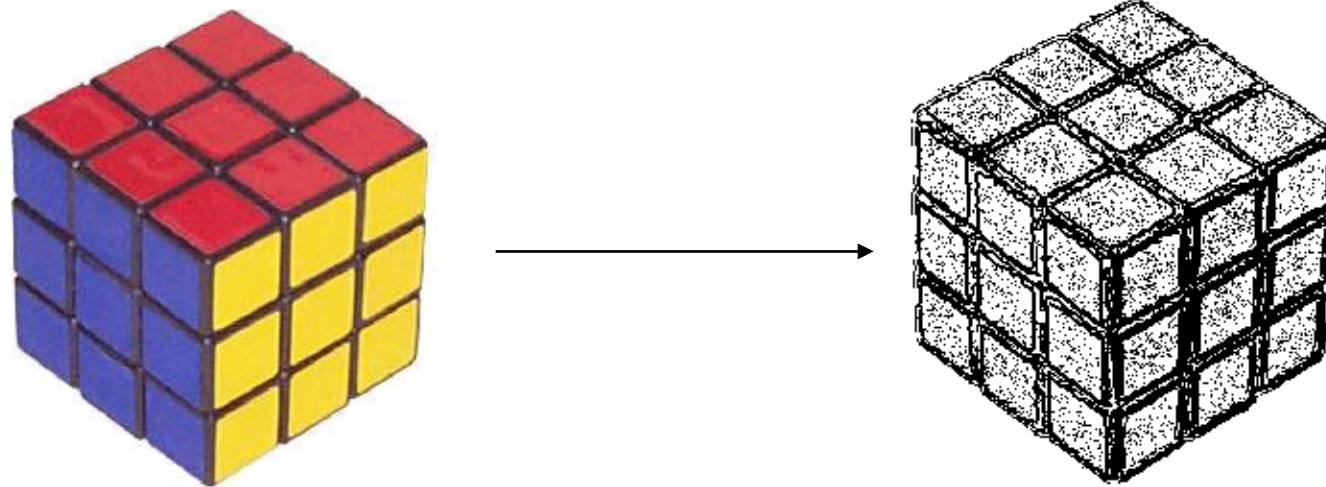
walking

turn around

sit down

sitting

Objekterkennung durch Vergleich von Texturen



Objekterkennung mittels Texturen (I)

Textur

- Visuelle Muster, die durch Farben bzw. Helligkeiten in einem Bild entstehen
- Entstehen durch Reflektion des Lichtes an einer Oberfläche (Gras, Holz, Metall, Stoff, aber auch Wolken)
- Enthalten Informationen über die Struktur der Oberfläche



Objekterkennung mittels Texturen (II)

Homogener Texturdeskriptor

- Beschreibt Richtung, Unebenheit und Regelmäßigkeit einer Textur
- Gut geeignet zur Beschreibung homogener Flächen
- Die im Bild enthaltenen Frequenzen werden ermittelt und deren Durchschnitt und Standardabweichung wird berechnet
- Invariant gegenüber Skalierung und Rotation

Objekterkennung mittels Texturen (III)

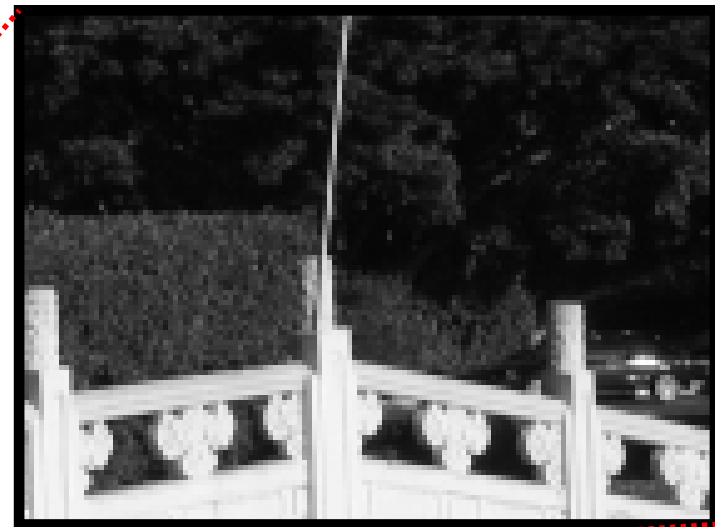
Kantenhistogramm (edge histogram)

- Beschreibt lokale (in einem Bildbereich) Verteilung von Kanten in Bildregionen (Texturen müssen nicht homogen sein)
- Einteilung des Bildes in 16 gleichgroße Blöcke (Skalierungsinvariant), Berechnung von Kanten für jeden Block
- 5 Typen von Kanten: vertikal, horizontal, 45 Grad, 135 Grad, ungerichtet
- Speicherung der Werte in einem Histogramm für jeden Typ und jeden Block ($5 \times 16 = 80$ Elemente)
- Bei einem Vergleich kann (muss aber nicht) die Rotation berücksichtigt werden
- MPEG-7: Es werden nur 3 Bits zur Beschreibung eines Histogrammelementes verwendet (insgesamt 240 Bits zur Beschreibung der Textur)

Objekterkennung mittels Texturen (IV)

Berechnung des Kantenhistogramms

1. Einteilung des Bildes in 16 Regionen

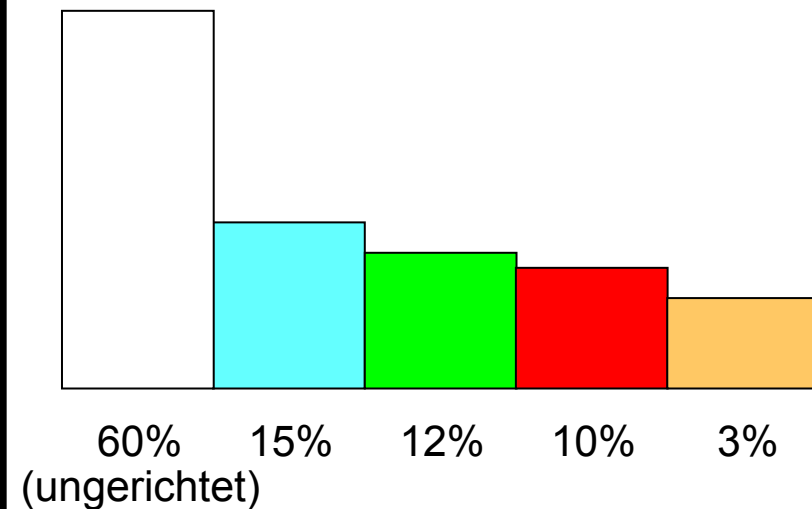
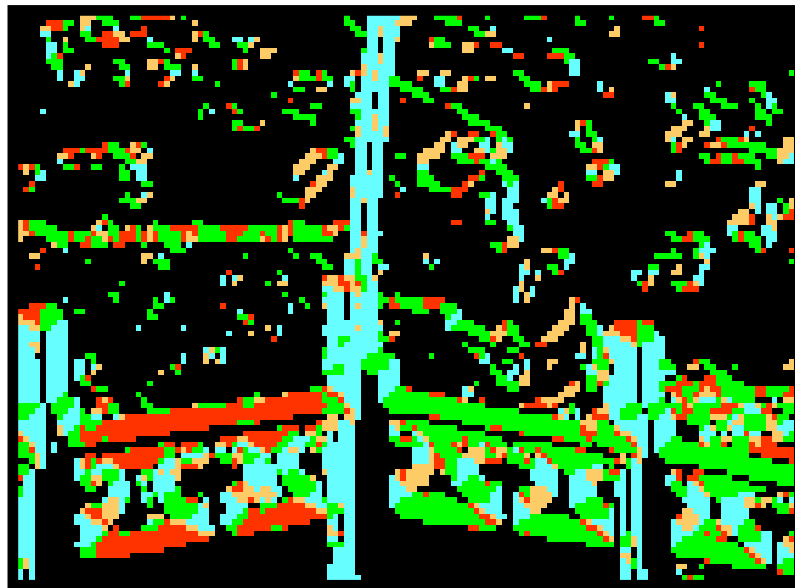


Berechnung der Kanten
für jede Region

Objekterkennung mittels Texturen (V)

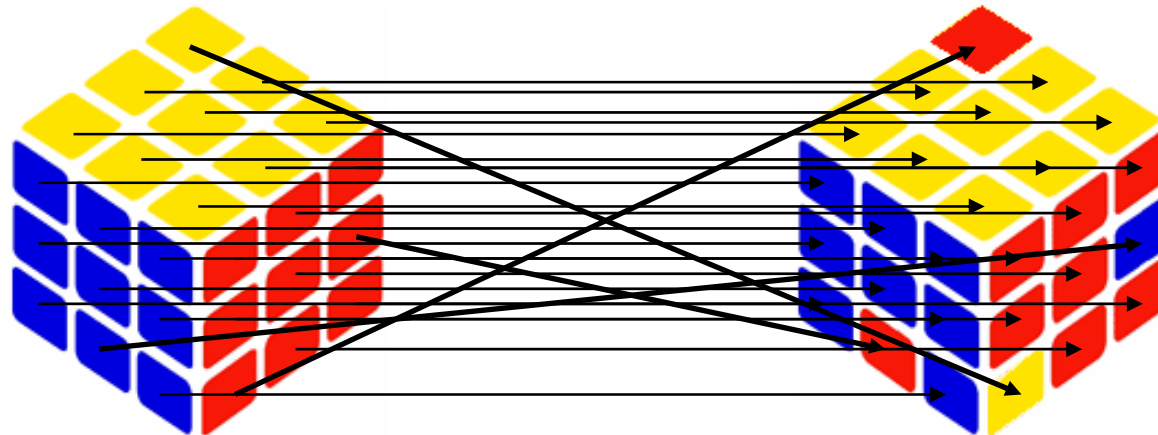
Berechnung des Kantenhistogramms

2. Berechnung des Kantenanstiegs des Blocks



0 Grad, 45 Grad, 90 Grad, 135 Grad

Objekterkennung durch Vergleich von Bewegungen



Objekterkennung durch Bewegungsanalyse (I)

Bewegungsaktivität (motion activity)

- Beschreibung der Bewegung in einem **Videosegment**
- Grobeinteilung in Kategorien:
 - langsam: Nachrichtensprecher
 - schnell: Strassenszene
 - rasend: Fußball, Basketball
- Maß: Standardabweichung der Längen der Bewegungsvektoren (eine von 5 Kategorien)
- Optionale Parameter:
 - Richtung der Bewegung
 - Bewegungsaktivität einer Bildregion
 - Zeitliche Dauer einer Bewegung

Objekterkennung durch Bewegungsanalyse (II)

Bewegungsbahn (motion trajectory)

- Beschreibung der Bewegung eines Bildbereiches in einem Videosegment
- Ähnlich wie Bewegungsaktivität, nur können mehrere Bewegungen beschrieben werden
- Beispiele:
 - Verkehrsüberwachung: Bei einer Verkehrsüberwachung wird für jedes einzelne Fahrzeug und jede Person im Bild eine Beschreibung der Bewegung gespeichert. Suchanfrage liefert Objekte in der Nähe.
 - Überwachung eines öffentlichen Platzes: Suchanfrage liefert Personen, die sich nach einem bestimmten Muster bewegen (z.B. sich langsam bewegen und plötzlich losrennen)

Zusammenfassung

- Modellbasierte Objekterkennung: Vergleich aus dem Bild extrahierter Merkmale mit einem Modell des gesuchten Objektes
- Die Wahl des Objektmodells orientiert sich an dem zu erkennenden Objekt sowie den zur Verfügung stehenden Daten
- Erkennung von Objekten durch Vergleich von:
 - **Farben** (Farbraum, menschliche Wahrnehmung, Histogramm, dominante Farbe)
 - **Konturen** (Kompaktheit, Exzentrizität, Krümmungen)
 - **Texturen** (homogener Texturdeskriptor, Kantenhistogramm)
 - **Bewegungen** (Bewegungsaktivität, Bewegungsbahn)

Zusammenfassung

Was wären gute Modelle für folgende Objekte?



Questions ?