


---

# Visual Computing

Niclas Kusenbach

LaTeX version:  SCHÖNER VERLAG

---

## Table of Contents

---

<b>1 Einführung in Visual Computing</b>	<b>2</b>	<b>3.3 Naive Bayes Klassifikator</b>	<b>10</b>
1.1 Grundlagen Visual Computing . . . . .	2	3.3.1 Das Problem mit vielen Merkmalen	10
1.2 Vier exemplarische Themenbereiche . . . .	2	3.4 Fallstudie: Gesichtsdetektion . . . . .	11
1.3 Wichtige Konzepte und Begriffe . . . . .	2	3.4.1 Sliding Window Ansatz . . . . .	11
1.4 Empfohlene Literatur . . . . .	2	3.5 Erkennungsarten (Biometrie) . . . . .	11
1.5 Zusammenfassung . . . . .	3		
<b>2 Wahrnehmung (Perception)</b>	<b>4</b>		
2.1 Motivation und Kognition (Motivation and Cognition) . . . . .	4		
2.2 Das Visuelle System (The Visual System)	4		
2.2.1 Aufbau des Auges (Structure of the Eye) . . . . .	4		
2.2.2 Photorezeptoren (Photoreceptors)	5		
2.3 Vorverarbeitung & Helligkeit (Preprocessing & Brightness) . . . . .	6		
2.3.1 Signalverarbeitung in der Retina .	6		
2.3.2 Optische Täuschungen (Optical Illusions) . . . . .	6		
2.3.3 Helligkeitswahrnehmung (Brightness Perception) . . . . .	7		
2.3.4 Auflösung und Kontrast (Resolution and Contrast) . . . . .	7		
2.4 Informationsextraktion: Tiefenwahrnehmung (Depth Perception) . . . .	7		
2.5 Aufmerksamkeit und Gedächtnis (Attention and Memory) . . . . .	8		
<b>3 Erkennung</b>	<b>9</b>		
3.1 Das Problem der Objekterkennung . . . .	9		
3.1.1 Herausforderungen der Erkennung	9		
3.1.2 Intuition: Wie funktioniert Erkennung? . . . . .	9		
3.2 Bayes Decision Theory (BDT) . . . . .	9		
3.2.1 Bayes-Entscheidungsregel . . . . .	10		
3.2.2 Likelihood Ratio Test (für 2 Klassen)	10		
3.2.3 Bedeutung von Prior und Likelihood in der Praxis . . . . .	10		

---

# 1 Einführung in Visual Computing

---

## 1.1 Grundlagen Visual Computing

---

### Definition

Visual Computing ist die Kombination mehrerer Informatikbereiche, die im Wesentlichen mit Bildern und Modellen arbeiten.

- Umfasst **Computergrafik, Computer Vision, Mensch-Maschine-Interaktion, Mustererkennung und Maschinelles Lernen**.
- Enge Verbindung zwischen Computer Vision und Computer Graphics – keine getrennte Betrachtung.
- Zentrale Fragestellungen: Informationsgewinnung aus Daten, effiziente Extraktion relevanter Information.

## 1.2 Vier exemplarische Themenbereiche

---

1. **3D Internet:** Erweiterung des Dokumentbegriffs auf 3D-Modelle; Anwendungen in Bildung, Medizin, Kultur.
  - 3D-Web, AR/VR, digitale Kulturgüter, Metaverse.
  - Retro-Digitalisierung vs. Digital Creation (CultLab3D, Flickr-Photogrammetrie).
2. **Skalierbare Objektmodellierung und -erkennung:**
  - Ziel: Erkennung zehntausender Kategorien mittels semantischer Hierarchien.
  - Nutzung von Deep Learning (Convolutional NNs, Transferlernen, 3D-Objektdatenbanken).
3. **Big Data / Visual Analytics:**
  - Kombination von Datenanalyse, ML und Visualisierung zur explorativen Erkenntnisgewinnung.
  - Verarbeitung großer heterogener Datenmengen (z. B. 300 000 Flickr-Bilder von Rom).
4. **Scene Understanding (3D/4D):**
  - Modellierung und Analyse von Szenen in Raum und Zeit.
  - Kombination von Objekterkennung, Tracking, Semantik und Bewegungsanalyse.
  - Anwendungen: Autonomes Fahren, Robotik, Sicherheitsüberwachung, AAL.

## 1.3 Wichtige Konzepte und Begriffe

---

- **Deep Learning / Convolutional Networks** – Grundprinzipien, GPU-basierte Berechnungen.
- **Interpretierbarkeit (White-Box vs. Black-Box)** – KI-Erklärbarkeit, Bias und Ethik.
- **Visualisierung und Informationsvisualisierung** – Darstellung komplexer Daten.
- **3D Output und AR/VR Technologien** – z. B. Shapeways, Wikitude.
- **Digitale Kultur und Erhalt von Kulturgütern** – z. B. Project Mosul, 3D-Rekonstruktionen.

## 1.4 Empfohlene Literatur

---

Autor / Titel	Relevanz
Szeliski – <i>Computer Vision: Algorithms and Applications</i>	Fundamentale Algorithmen zur Bildverarbeitung und -analyse.

Blundell – <i>An Introduction to Computer Graphics</i>	Grundlagen der 3D-Darstellung.
Dix et al. – <i>Human Computer Interaction</i>	Basis für Mensch-Maschine-Schnittstellen.
Burger & Burke – <i>Digitale Bildverarbeitung – algorithmische Einführung</i>	Mathematische Grundlagen für Bildverarbeitung.

---

## 1.5 Zusammenfassung

---

- Visual Computing verbindet Grafik, Vision und Interaktion.
- Vier Kernbereiche (3D Internet, Objekterkennung, Visual Analytics, Scene Understanding) bilden die strukturierende Grundlage der Vorlesung und sind klausurrelevant.
- Verständnis der praktischen Beispiele und Übungen ist essentiell für die Klausur.

## 2 Wahrnehmung (Perception)

### 2.1 Motivation und Kognition (Motivation and Cognition)

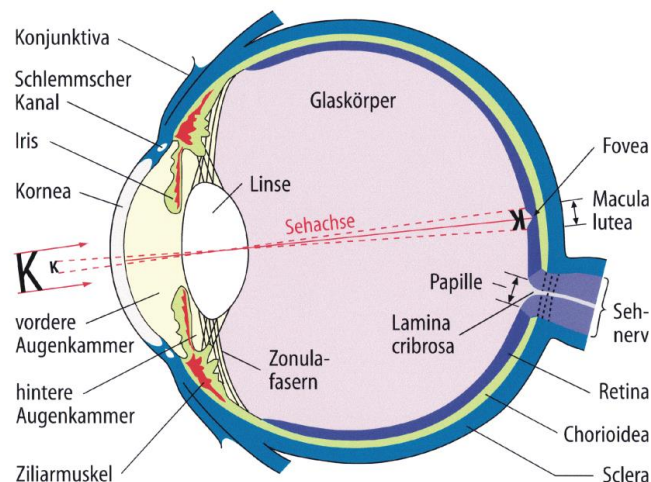
- **Warum VC? (Why VC?)** Die Leistungsfähigkeit von Rechnern wächst exponentiell (**Moore's Law**), aber die Kapazität von Menschen ist (fast) konstant (**Darwin's Law**). VC hilft, diese Lücke zu überbrücken.
- **Kognition (Cognition):** Sammelbegriff für alle Prozesse des Wahrnehmens und Erkennens (Denken, Erinnern, Lernen, etc.).
- **Modell der Informationsverarbeitung (Model of Information Processing):** Ein modulares 3-Stufenmodell:
  1. **Perception** (Wahrnehmung durch Sinne)
  2. **Decision** (Entscheidungsfindung im Gehirn)
  3. **Response** (Reaktion durch Körper)
- **Bearbeitungszeiten (Processing Times):** Jedes Untersystem benötigt Zeit.
  - Wahrnehmung (Perception):  $\approx 100$  ms
  - Entscheidung (Cognition):  $\approx 70$  ms
  - Reaktion (Motor):  $\approx 70$  ms
- **Wahrnehmung vs. Realität:** Was wir wahrnehmen, ist kein direktes Abbild der Realität, sondern eine partielle Hypothese, die auf unvollständiger Information basiert.

### 2.2 Das Visuelle System (The Visual System)

#### Visueller Reiz (Visual Stimulus)

Ein äußerer visueller Reiz ist **elektromagnetische Strahlung**. Sichtbares Licht liegt im Wellenlängenbereich von ca. 400nm (violett) bis 700nm (rot). Die Frequenz  $\nu$  und Wellenlänge  $\lambda$  hängen über  $\nu \cdot \lambda = c$  zusammen.

#### 2.2.1 Aufbau des Auges (Structure of the Eye)



- **Optische Elemente:** Hornhaut (Kornea), Linse, Iris (Blende, 2-8mm), Glaskörper.

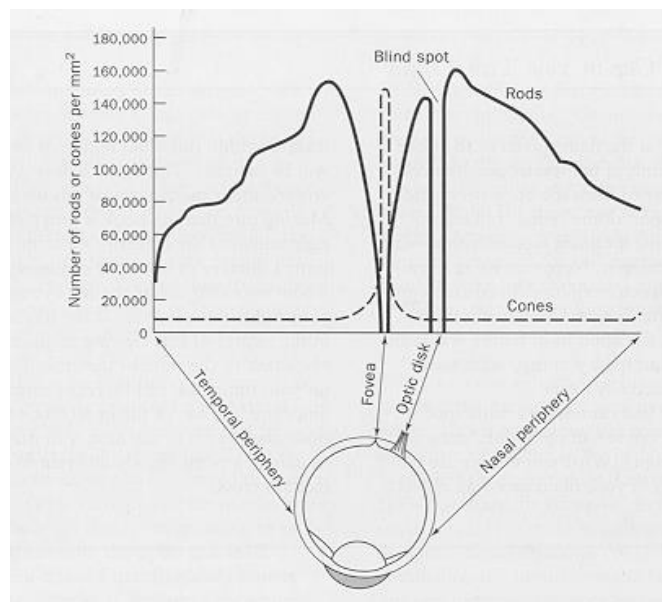
- **Linse (Lens):** Akkomodation (Scharfeinstellung).
- **Retina (Netzhaut):** Enthält die Photorezeptoren.
- **Fovea Centralis:** Bereich der höchsten Auflösung (im "gelben Fleck" / Macula lutea).
- **Blinder Fleck (Blind Spot):** Papilla nervi optici; Austrittspunkt des Sehnervs, keine Rezeptoren.

### 2.2.2 Photorezeptoren (Photoreceptors)

---

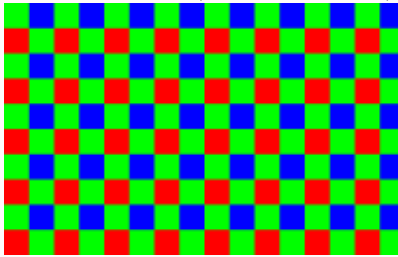
Es gibt zwei Haupttypen von Photorezeptoren auf der Retina:

- **Stäbchen (Rods):**
  - ca. 100-120 Mio.
  - Hauptsächlich außerhalb der Fovea.
  - Für **Nachtsehen (skotopisches Sehen)**.
  - Sehr lichtempfindlich, kein Farbsehen.
  - Empfindlichkeitsmaximum bei 498 nm (grün).
- **Zapfen (Cones):**
  - ca. 7-8 Mio.
  - Hauptsächlich **in der Fovea** (Bereich des schärfsten Sehens).
  - Für **Tagsehen (photopisches Sehen)**.
  - 3 Typen für Farbsehen:
    - \* **S-Zapfen** (Short): Max. bei 420 nm (Blau).
    - \* **M-Zapfen** (Medium): Max. bei 534 nm (Grün).
    - \* **L-Zapfen** (Long): Max. bei 564 nm (Rot).



## Bayer-Sensor

Digitale Kamerasensoren nutzen oft ein **Bayer-Muster**. Dies ist ein Farbfilter-Array, meist mit **50% Grün**, **25% Rot** und **25% Blau**. Grün ist privilegiert, da das menschliche Auge für Grün den größten Beitrag zur **Helligkeits- und Kontrastwahrnehmung** leistet (72% Grünanteil).

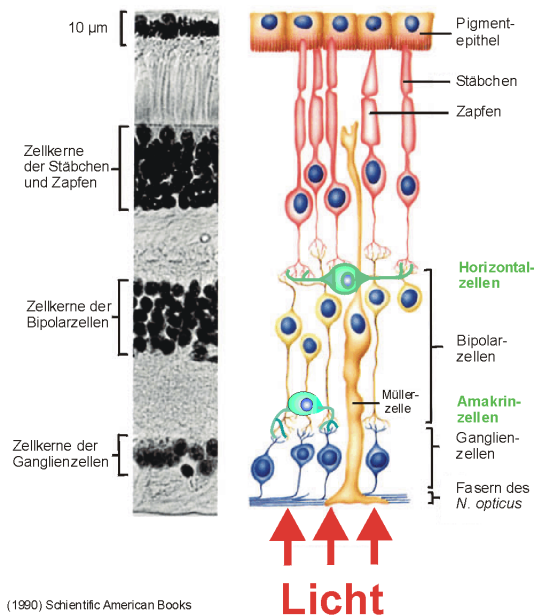


## 2.3 Vorverarbeitung & Helligkeit (Preprocessing & Brightness)

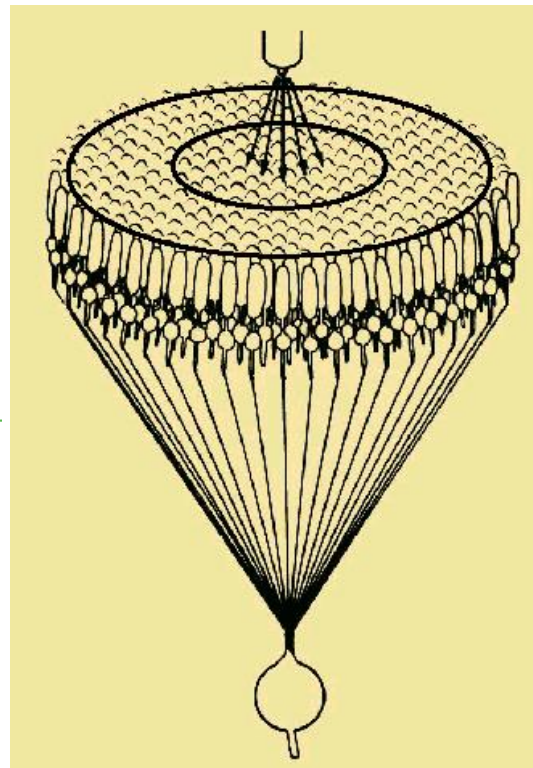
### 2.3.1 Signalverarbeitung in der Retina

Das Licht trifft (paradoxe Weise) erst auf die Ganglien- und Bipolarzellen, bevor es die Stäbchen und Zapfen erreicht.

- **Bipolar-Zellen:** Sammeln, gewichten und leiten Informationen weiter.
- **Horizontal- & Amakrin-Zellen:** Kombinieren Signale mehrerer Rezeptoren (räumlich) bzw. verarbeiten zeitliche Änderungen.
- **Ganglien-Zellen:** Integrieren Informationen, z.B. für **Kontrastwahrnehmung** durch Unterschied zwischen Zentrum und Peripherie (Center-Surround-Antagonismus).



(1990) Scientific American Books



### 2.3.2 Optische Täuschungen (Optical Illusions)

Diese frühe Signalverarbeitung führt zu Täuschungen, die zeigen, dass Wahrnehmung nicht objektiv ist:

- **Hermann-Gitter:** Graue Flecke erscheinen in den Kreuzungen eines weißen Gitters auf schwarzem Grund.

- **Mach-Bänder (Mach Bands):** An Kanten zwischen unterschiedlich hellen, aber homogenen Flächen werden helle/dunkle Bänder wahrgenommen, wo keine sind (eine Art "Überschwingen" der Wahrnehmung).
- **Simultankontrast (Simultaneous Contrast):** Die wahrgenommene Helligkeit einer Fläche hängt von der Helligkeit ihrer Umgebung ab. Ein identisches Grau erscheint auf schwarzem Grund heller als auf weißem Grund.

### 2.3.3 Helligkeitswahrnehmung (Brightness Perception)

---

- Helligkeit (Brightness) ist **keine absolute Größe**, sondern subjektiv.
- Sie ist u.a. abhängig von der Reizstärke (Leuchtdichte), der Adaption an vorherige Leuchtdichten und der Umgebungsleuchtdichte.
- **Weber-Fechnersches Gesetz:** Beschreibt den Zusammenhang zwischen Reizintensität ( $R$ ) und Hellempfindung ( $L$ ).
  - **Webersches Gesetz (Schwelle):**  $\Delta L = \frac{\Delta R}{R} = \text{const.}$  (minimaler Kontrast für Wahrnehmung ca. 0.8%)
  - **Fechnersches Gesetz:**  $L = c_1 \times \log R$
  - **Stevensches Gesetz (State-of-the-Art):**  $E = c_2 \times R^k$  (für Licht  $k = 0.3$ )

### 2.3.4 Auflösung und Kontrast (Resolution and Contrast)

---

- **Sehschärfe (Visual Acuity):** Die Fähigkeit, kleine Details zu erkennen, ist begrenzt. Z.B. Punktsehschärfe ca. 1 Bogenminute ( $1'$ ).
- **Kontrastempfindlichkeit (Contrast Sensitivity):** Gemessen mit Sinus-Mustern (sinusoidal gratings).
- **Contrast Sensitivity Function (CSF):** Beschreibt die Auflösung im Frequenzraum. Zeigt, dass das Auge für mittlere Ortsfrequenzen (ca. 2-5 Zyklen/Grad) am empfindlichsten ist und die Empfindlichkeit zu sehr hohen (Details) und sehr niedrigen Frequenzen (langsame Übergänge) abfällt.

## 2.4 Informationsextraktion: Tiefenwahrnehmung (Depth Perception)

---

Das visuelle System nutzt verschiedene Hinweisreize (**Depth Cues**), um Raumwahrnehmung zu erzeugen.

- **1. Binokulare Cues (Zwei Augen):**
  - **Disparität / Parallaxe:** Der Haupt-Cue. Da die Augen getrennt sind, sehen sie leicht unterschiedliche Bilder. Das Gehirn fusioniert diese.
  - **Positive Parallaxe:** Objekte erscheinen *hinter* der Bildebene.
  - **Negative Parallaxe:** Objekte erscheinen *vor* der Bildebene.
  - **Akkommodation** (Linsenanpassung) und **Konvergenz** (Augenstellung).
- **2. Pictorial Depth Cues (Monokular / Bildlich):**
  - **Linearperspektive:** Parallele Linien konvergieren in der Ferne.
  - **Verdeckung (Occlusion):** Ein Objekt, das ein anderes verdeckt, wird als näher wahrgenommen.
  - **Texturgradient:** Texturen werden mit der Entfernung dichter und feiner.
  - **Atmosphärische Tiefe:** Entfernte Objekte erscheinen blasser und bläulicher.
  - **Schattenwurf (Shadows):** Wichtig für Position und Form; Annahme: Licht kommt von oben.
  - Weitere: Fokus/Blur, Vertraute Größe, Höhe im Gesichtsfeld, etc..
- **3. Dynamische Depth Cues (Bewegung):**
  - **Bewegungsparallaxe (Motion Parallax):** Objekte, die näher sind, bewegen sich bei einer Kopfbewegung scheinbar schneller als entfernte Objekte.

- **Kinetischer Tiefeneffekt (Kinetic Depth Effect):** 3D-Struktur wird aus der 2D-Projektion einer Bewegung extrahiert (z.B. "Structure from Motion").

Die Depth Cues sind nicht redundant, sondern **additiv** und werden je nach Aufgabe (Task) **flexibel gewichtet**.

## 2.5 Aufmerksamkeit und Gedächtnis (Attention and Memory)

---

- **Frühe Wahrnehmung (Preattentive Processing):** Bestimmte Merkmale (Farbe, Größe, Richtung, Schattierung) werden sehr schnell (ca.  $\approx 10$  ms) und parallel verarbeitet, bevor die bewusste Aufmerksamkeit greift. **Verbindungen** von Merkmalen (z.B. "roter Kreis") erfordern Aufmerksamkeit.
- **Aufmerksamkeit (Attention):** Dient als **Filter** oder "Gateway to Memory".
- **Veränderungsblindheit (Change Blindness):** Unfähigkeit, große Änderungen in einer Szene zu bemerken, wenn die Aufmerksamkeit abgelenkt ist (z.B. durch Flimmern). Dies zeigt, dass wir kein vollständiges Bild der Welt im Kopf haben.
- **Arbeitsgedächtnis (Working Memory):**
  - Schneller Zugriff (ca. 70 ms), schneller Verfall (ca. 200 ms).
  - Sehr begrenzte Kapazität:  $7 \pm 2$  "**Chunks**" (Miller, 1956).
  - "Chunks" sind sinnvolle Einheiten (z.B. DA, TU, VC, VL statt DATUVCVL).
- **Langzeitgedächtnis (Long-term Memory):**
  - Nahezu unbegrenzte Kapazität.
  - Langsamerer Zugriff (ca. 100 ms).



## 3 Erkennung

### 3.1 Das Problem der Objekterkennung

#### 3.1.1 Herausforderungen der Erkennung

Menschliche Wahrnehmung ist Computern weit überlegen. Die Erkennung ist schwierig aufgrund von:

- **Mehrdeutigkeit & Illusionen:** (z.B. Don Quixote-Gesicht, Mona Lisa-Mosaik, ambivalente Elefantenbeine).
- **Tarnung & Verdeckung:** (z.B. Bev Doolittles Pferde im Schnee).
- **Kontext:** Die Interpretation von Teilen hängt stark vom Gesamtkontext ab (z.B. verschwommene Straßenszene, Punktwolken-Wörter).
- **Fehler im Kontext:** M.C. Escher-Bilder sind *lokal* konsistent, aber *global* unmöglich.

#### 3.1.2 Intuition: Wie funktioniert Erkennung?

Objekterkennung basiert auf zwei Hauptkomponenten:

1. **Lokale Beschreibung / Merkmale:** (z.B. Augen, Nase, Mund bei einem Gesicht).
2. **Globale Anordnung der Merkmale:** (z.B. relative Positionen und Größen der Merkmale zueinander).

Weitere wichtige Aspekte sind Segmentierung und Szenenkontext.

Pictorial Structure (Fischler & Elschlager, 1973) Ein klassisches Modell zur Objekterkennung, das diese Intuition formalisiert.

- **Teile (Parts):** 2D-Bildfragmente (lokale Merkmale).
- **Aufbau (Structure):** Die Anordnung der Teile, oft modelliert durch "Federn", die die relative Position und Deformation beschränken.

Herausforderungen für dieses Modell sind **Deformationen** (Teile ändern ihre Position) und **Durcheinander** (Clutter, irrelevante Merkmale im Hintergrund).

### 3.2 Bayes Decision Theory (BDT)

Die BDT bietet ein mathematisches Framework, um Klassifikationsentscheidungen zu treffen, die die **Wahrscheinlichkeit einer Fehlklassifikation minimieren**.

Die drei Kernkonzepte der BDT

1. **A Priori (Prior)  $P(C_k)$ :** Die Wahrscheinlichkeit einer Klasse  $C_k$  *bevor* wir irgendwelche Daten (Merkmale) gesehen haben. Dies ist unser "Vorwissen". (Bsp:  $P(a) = 0.75$ ,  $P(b) = 0.25$ ).
2. **Bedingte W'keit (Likelihood)  $p(x|C_k)$ :** Die Wahrscheinlichkeitsdichte, das Merkmal  $x$  zu beobachten, *gegeben* dass es zur Klasse  $C_k$  gehört. (Bsp: Wie ist die Verteilung der "Anzahl schwarzer Pixel" für den Buchstaben 'a'?).
3. **A Posteriori (Posterior)  $P(C_k|x)$ :** Die (aktualisierte) Wahrscheinlichkeit, dass es sich um Klasse  $C_k$  handelt, *nachdem* wir das Merkmal  $x$  beobachtet haben. Dies ist, was wir für die Entscheidung wissen wollen.

Bayes' Theorem (Verbindung der Konzepte) Das Theorem verknüpft Prior, Likelihood und Posterior:

$$P(C_k|x) = \frac{p(x|C_k)P(C_k)}{p(x)}$$

Wobei  $p(x) = \sum_j p(x|C_j)P(C_j)$  ein Normalisierungsfaktor ist (Summe über alle Klassen).

Merke: Posterior =  $\frac{\text{Likelihood} \times \text{Prior}}{\text{Normalization (Evidence)}}$

### 3.2.1 Bayes-Entscheidungsregel

Um den Fehler zu minimieren, wählen wir die Klasse  $C_k$ , die die **höchste A-Posteriori-Wahrscheinlichkeit** hat:

- Entscheide  $C_k$  wenn:  $P(C_k|x) > P(C_j|x)$  für alle  $j \neq k$ .
- Da der Nenner  $p(x)$  für alle Klassen gleich ist, ist dies äquivalent zu:
- Entscheide  $C_k$  wenn:  $p(x|C_k)P(C_k) > p(x|C_j)P(C_j)$  für alle  $j \neq k$ .

Wir vergleichen also nicht die reinen Likelihoods, sondern die mit dem Prior **skalierten Likelihoods**. Der Punkt, an dem diese skalierten Kurven sich schneiden, ist die **Entscheidungsgrenze**.

### 3.2.2 Likelihood Ratio Test (für 2 Klassen)

Eine äquivalente Formulierung ist der Likelihood-Ratio-Test:

$$\text{Entscheide } C_1 \text{ wenn: } \frac{p(x|C_1)}{p(x|C_2)} > \frac{P(C_2)}{P(C_1)} = \lambda$$

### 3.2.3 Bedeutung von Prior und Likelihood in der Praxis

Die Trennung von  $p(x|C_k)$  und  $P(C_k)$  ist extrem nützlich:

- **Beispiel Spracherkennung:**

- $x$  = Audiosignal
- $C_k$  = Satz (z.B. "This machine can recognize speech")
- $p(x|C_k)$  = **Akustisches Modell** (Wie wahrscheinlich klingt das Audio  $x$ , wenn Satz  $C_k$  gesagt wurde?)
- $P(C_k)$  = **Sprachmodell (Language Model)** (Wie wahrscheinlich ist der Satz  $C_k$  in der Sprache?)

Zwei Sätze ("recognize speech" vs. "wreck a nice beach") klingen evtl. ähnlich (ähnliches Likelihood), aber das Sprachmodell (Prior) wird dem unwahrscheinlichen Satz eine viel geringere Wahrscheinlichkeit zuweisen.

- **Beispiel Bildverarbeitung:**

- $p(x|C_k)$  = **Low-Level Image Measurements** (Likelihood)
- $P(C_k)$  = **High-Level Model Knowledge** (Prior)

## 3.3 Naive Bayes Klassifikator

### 3.3.1 Das Problem mit vielen Merkmalen

Bei  $d$  Merkmalen  $(x_1, \dots, x_d)$  müssten wir die  $d$ -dimensionale Verteilung  $p(x_1, \dots, x_d|C_k)$  schätzen. Dies ist extrem datenaufwändig und oft unmöglich (**Fluch der Dimensionalität**).

Die "Naive" Annahme Der Naive Bayes Klassifikator trifft eine vereinfachende (naive) Annahme:

- Alle Merkmale  $x_i$  sind **statistisch unabhängig** voneinander, *gegeben die Klasse  $C_k$* .
- Dies erlaubt es, die gemeinsame Wahrscheinlichkeit zu faktorisieren:

$$p(x_1, \dots, x_d | C_k) = \prod_{i=1}^d p(x_i | C_k)$$

Diese Annahme ist oft *falsch*, aber der resultierende Klassifikator funktioniert in der Praxis überraschend gut.

**Naive Bayes Entscheidungsregel** Die Entscheidungsregel (maximiere Posterior) vereinfacht sich zu:

$$\text{Entscheide } C_k \text{ wenn: } P(C_k) \prod_{i=1}^d p(x_i | C_k) \text{ maximal ist.}$$

Für den 2-Klassen-Fall wird der Likelihood Ratio Test zu:

$$\prod_{i=1}^d \frac{p(x_i | C_1)}{p(x_i | C_2)} > \frac{P(C_2)}{P(C_1)}$$

"Lernen" bedeutet hier **Density Estimation**: Man muss die 1D-Wahrscheinlichkeitsdichten  $p(x_i | C_k)$  für jedes Merkmal  $i$  und jede Klasse  $k$  aus den Trainingsdaten schätzen (z.B. durch Histogramme).

### 3.4 Fallstudie: Gesichtsdetektion

Ein Beispiel für **Appearance-Based Methods**: Modelle werden direkt aus (großen) Bilddatensammlungen gelernt.

#### 3.4.1 Sliding Window Ansatz

- Ein Fenster (z.B.  $19 \times 19$  Pixel) wird über das gesamte Bild geschoben.
- Bei jeder Position wird entschieden: "Gesicht" ( $C_1$ ) oder "Kein Gesicht" ( $C_2$ ).
- Das Bild wird skaliert (z.B. Faktor 1.2 verkleinert) und der Vorgang wiederholt sich, um Gesichter aller Größen zu finden.

Fallstudie: Schneiderman & Kanade (1998) Ein sehr erfolgreicher Gesichtsdetektor, der auf Naive Bayes basiert.

- **1. Repräsentation (Merkmale  $x_i$ ): Wavelet-Koeffizienten** an bestimmten Frequenzen, Orientierungen und **Positionen** ( $f_i, u_i, v_i$ ). Dies kodiert sowohl lokale Merkmale (Kanten) als auch deren globale Anordnung (Position).
- **2. Trainingsdaten:**
  - *Positive Beispiele* ( $C_1$ ): Tausende von Bildern, die Gesichter enthalten (normalisiert).
  - *Negative Beispiele* ( $C_2$ ): Tausende von Bildern, die *keine* Gesichter enthalten.
- **3. Klassifikator & Lernen:** Naive Bayes. Das "Lernen" besteht darin, die Wahrscheinlichkeiten  $P(x_i | C_{\text{face}})$  und  $P(x_i | C_{\text{non-face}})$  für jedes Merkmal  $x_i$  zu schätzen. Dies geschieht durch **Zählen (Erstellen von Histogrammen)** in den positiven und negativen Trainingsdatensätzen.
- **Multi-View:** Um Gesichter aus verschiedenen Winkeln zu erkennen, werden separate Detektoren trainiert (z.B. Frontal, Linksprofil, Rechtsprofil) und die Ergebnisse kombiniert.

### 3.5 Erkennungsarten (Biometrie)

Gesichtserkennung ist ein biometrisches Verfahren. Man muss zwischen verschiedenen Aufgaben unterscheiden:

- **Detektion:** (Face vs. Non-Face) Ist überhaupt ein Objekt (Gesicht) vorhanden?

- **Verifikation (1:1):** "Bin ich das?" (z.B. Smartphone entsperren).
  - Eine Person gibt ihre Identität an (z.B. Nutzer-ID).
  - Das System vergleicht die aktuelle Probe **nur mit dem einen** gespeicherten Template dieser ID.
  - Ausgabe: Ja / Nein.
- **Identifikation (1:n):** "Wer ist das?" (z.B. Überwachung).
  - Eine Person zeigt nur ihr Merkmal (Gesicht).
  - Das System vergleicht die Probe mit **allen  $n$**  Templates in der Datenbank.
  - Ausgabe: Eine Kandidatenliste (die  $m$  besten Treffer,  $m \ll n$ ).