


---

# Visual Computing

Niclas Kusenbach

LaTeX version:  SCHOUTER

---

## Table of Contents

---

### Contents

<b>1 Einführung in Visual Computing</b>	<b>2</b>	3.2.2 Likelihood Ratio Test (Spezialfall 2 Klassen) . . . . .	10
1.1 Grundlagen Visual Computing . . . . .	2	3.2.3 Prior und Likelihood in der Praxis . . . . .	10
1.2 Vier exemplarische Themenbereiche . . . . .	2	3.3 Naive Bayes Klassifikator . . . . .	10
1.3 Wichtige Konzepte und Begriffe . . . . .	2	3.3.1 Das Problem hoher Dimensionen . . . . .	10
1.4 Empfohlene Literatur . . . . .	2	3.3.2 Naive Bayes Entscheidungsregel . . . . .	11
1.5 Zusammenfassung . . . . .	3	3.4 Fallstudie: Gesichtsdetektion . . . . .	11
<b>2 Wahrnehmung (Perception)</b>	<b>4</b>	3.4.1 Sliding Window Ansatz . . . . .	11
2.1 Motivation und Kognition (Motivation and Cognition) . . . . .	4	3.5 Erkennungsarten (Biometrie) . . . . .	12
2.2 Das Visuelle System (The Visual System)	4	<b>4 Fouriertheorie</b>	<b>13</b>
2.2.1 Aufbau des Auges (Structure of the Eye) . . . . .	4	4.1 Motivation . . . . .	13
2.2.2 Photorezeptoren (Photoreceptors)	5	4.2 Mathematische Grundlagen . . . . .	13
2.3 Vorverarbeitung & Helligkeit (Preprocessing & Brightness) . . . . .	6	4.3 Die Fourier-Reihe (Fourier Series) . . . . .	13
2.3.1 Signalverarbeitung in der Retina . . . . .	6	4.3.1 Analogie zum Vektorraum . . . . .	13
2.3.2 Optische Täuschungen (Optical Illusions) . . . . .	6	4.3.2 Formeln der Fourier-Reihe . . . . .	14
2.3.3 Helligkeitswahrnehmung (Brightness Perception) . . . . .	7	4.4 Die Fourier-Transformation (FT) . . . . .	14
2.3.4 Auflösung und Kontrast (Resolution and Contrast) . . . . .	7	4.4.1 Wichtige Fourier-Transformationspaare . . . . .	15
2.4 Informationsextraktion: Tiefenwahrnehmung (Depth Perception) . . . . .	7	4.5 Faltung und Filterung . . . . .	15
2.5 Aufmerksamkeit und Gedächtnis (Attention and Memory) . . . . .	8	4.5.1 Anwendung: Filterung . . . . .	16
<b>3 Erkennung</b>	<b>9</b>	4.6 Abtastung von Signalen (Sampling) . . . . .	16
3.1 Das Problem der Objekterkennung . . . . .	9	4.6.1 Modell der Abtastung . . . . .	16
3.1.1 Herausforderungen der Erkennung . . . . .	9	4.6.2 Abtastung im Frequenzraum . . . . .	16
3.1.2 Intuition: Wie funktioniert Erkennung? . . . . .	9	4.6.3 Aliasing und das Abtasttheorem . . . . .	16
3.2 Bayes Decision Theory (BDT) . . . . .	9	4.7 Zusammenfassung der Konzepte . . . . .	17
3.2.1 Bayes-Entscheidungsregel (Fehlerminimierung) . . . . .	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)

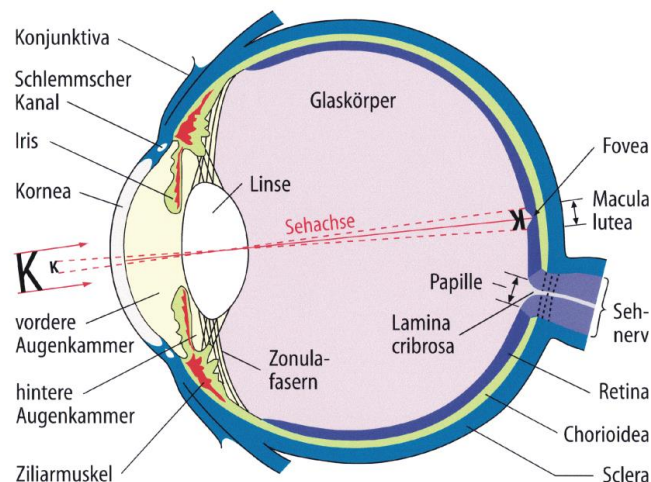
- **Warum VC? (Why VC?)** Die Leistungsfähigkeit von Rechnern wächst exponentiell (**Moore's Law**), aber die Kapazität von Menschen ist (fast) konstant (**Darwin's Law**). VC hilft, diese Lücke zu überbrücken.
- **Kognition (Cognition):** Sammelbegriff für alle Prozesse des Wahrnehmens und Erkennens (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.

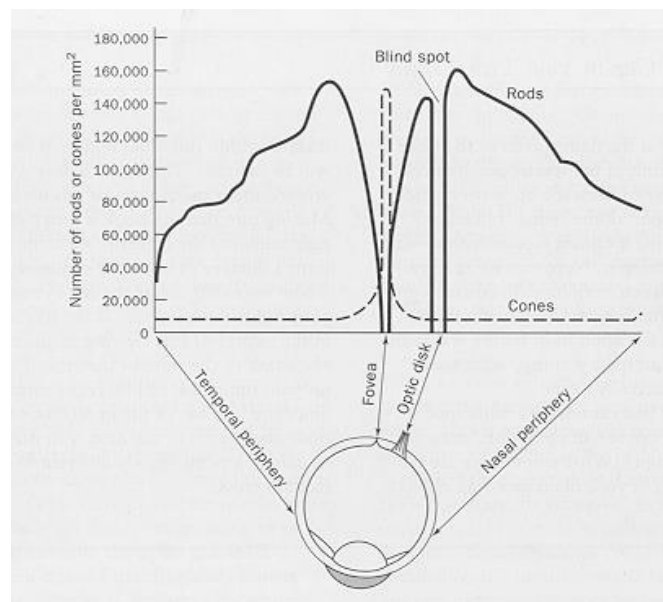
- **Linse (Lens):** Akkomodation (Scharfeinstellung).
- **Retina (Netzhaut):** Enthält die Photorezeptoren.
- **Fovea Centralis:** Bereich der höchsten Auflösung (im "gelben Fleck" / Macula lutea).
- **Blinder Fleck (Blind Spot):** Papilla nervi optici; Austrittspunkt des Sehnervs, keine Rezeptoren.

### 2.2.2 Photorezeptoren (Photoreceptors)

---

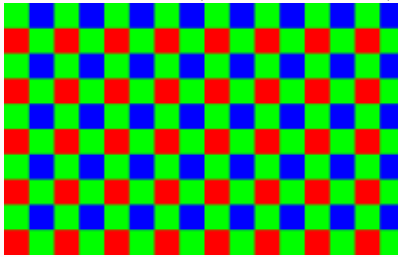
Es gibt zwei Haupttypen von Photorezeptoren auf der Retina:

- **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).

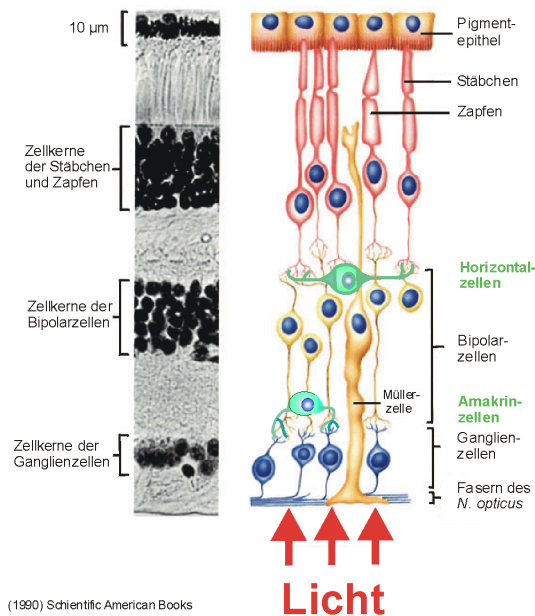


## 2.3 Vorverarbeitung & Helligkeit (Preprocessing & Brightness)

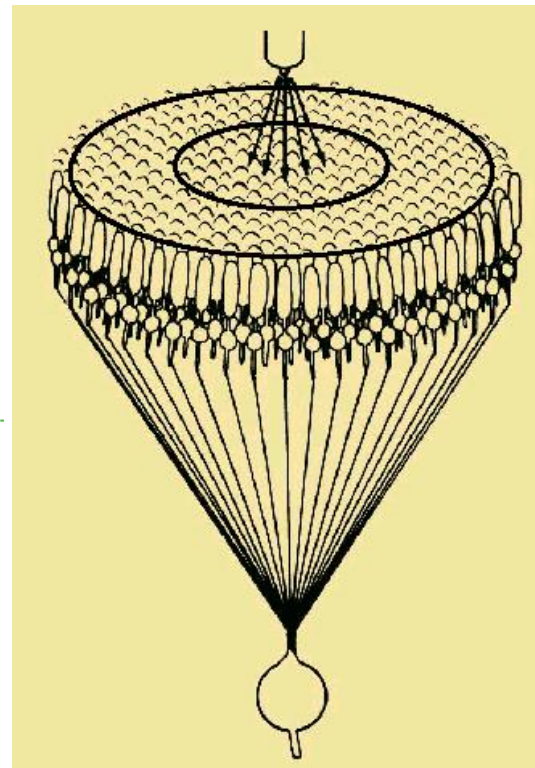
### 2.3.1 Signalverarbeitung in der Retina

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. Punktschärfe 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 ist ein mathematisches Framework zur optimalen Entscheidungsfindung unter Unsicherheit. Ihr Ziel ist es, Klassifikationsentscheidungen so zu treffen, dass die **erwartete Fehlklassifikationsrate minimiert** wird. Sie nutzt dazu Wahrscheinlichkeiten als formale Beschreibung von Unsicherheit.

Die drei Kernkonzepte der BDT

1. **A Priori (Prior)  $P(C_k)$ :** Die Wahrscheinlichkeit einer Klasse  $C_k$ , *bevor* wir Daten  $x$  beobachtet haben. Der Prior kodiert unser Vorwissen oder unsere Erwartungen über die Häufigkeit von Klassen. Wichtig: Priors wirken wie ein "Bias" zugunsten bestimmter Klassen.
2. **Klassenspezifische Dichte (Likelihood)  $p(x|C_k)$ :** Die Wahrscheinlichkeit (genauer: Wahrscheinlichkeitsdichte), die Merkmalsausprägung  $x$  zu beobachten, *unter der Annahme*, dass die wahre Klasse  $C_k$  ist. Sie beschreibt also, wie typisch ein bestimmtes Datenmuster  $x$  für eine Klasse ist. Hinweis: Likelihood ist eine Funktion in  $x$ , nicht in der Klasse.
3. **A Posteriori (Posterior)  $P(C_k|x)$ :** Nachbeobachtete Wahrscheinlichkeit einer Klasse, *nachdem* wir  $x$  kennen. Dies ist genau die Größe, die wir für die eigentliche Entscheidung benötigen.

Bayes' Theorem (Verbindung der Konzepte)

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

Die Evidenz

$$p(x) = \sum_j p(x|C_j) P(C_j)$$

stellt sicher, dass die Posterior-Wahrscheinlichkeiten normiert sind. Wichtig: Für die Entscheidung ist  $p(x)$  oft irrelevant, da es für alle Klassen gleich ist.

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidenz}}$$

### 3.2.1 Bayes-Entscheidungsregel (Fehlerminimierung)

Wir wählen die Klasse mit der größten posterioren Wahrscheinlichkeit:

$$C^*(x) = \arg \max_k P(C_k|x).$$

Da der Nenner  $p(x)$  konstant ist, genügt:

$$C^*(x) = \arg \max_k p(x|C_k) P(C_k).$$

**Intuition:** Die Likelihood misst, wie gut die Daten zu einer Klasse passen; der Prior berücksichtigt, wie plausibel diese Klasse im Voraus ist. Der Schnittpunkt der Produkte  $p(x|C_k)P(C_k)$  bildet die **Entscheidungsgrenze**.

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

Für zwei Klassen  $C_1$  und  $C_2$ :

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

Links: datengetriebene Evidenz, Rechts: Schwellwert aus Vorwissen. So sieht man explizit, wie ein starker Prior eine Entscheidung "umkippen" kann.

### 3.2.3 Prior und Likelihood in der Praxis

Die Trennung zwischen Datenmodell (Likelihood) und Vorwissen (Prior) ermöglicht robuste Modelle:

- **Spracherkennung:** Likelihood = akustisches Modell; Prior = Sprachmodell. Das Sprachmodell zieht uns aus akustischen Mehrdeutigkeiten heraus.
- **Bildverarbeitung:** Likelihood = Appearance Model; Prior = Kontextwissen über sinnvolle Objekte am Ort.

## 3.3 Naive Bayes Klassifikator

### 3.3.1 Das Problem hoher Dimensionen

Für  $d$  Merkmale:

$$x = (x_1, \dots, x_d)$$

müssten wir die gemeinsame Dichte  $p(x|C_k)$  im  $d$ -dimensionalen Raum schätzen. Dies benötigt exponentiell viele Daten: **Fluch der Dimensionalität**. Schon wenige Dutzend Dimensionen machen klassische Dichteschätzung praktisch unmöglich.

Die Naive-Unabhängigkeitsannahme Um das Problem zu umgehen, wird angenommen:

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

Das heißt: Die Merkmale sind *bedingt unabhängig* gegeben die Klasse. Diese Annahme ist oft falsch, aber in vielen Anwendungen erstaunlich effektiv, weil:

- Abhängigkeiten zwischen Merkmalen sich teilweise gegenseitig ausgleichen,
- Die Entscheidung meist nur eine Rangordnung braucht (MAP), nicht exakte Wahrscheinlichkeiten.

### 3.3.2 Naive Bayes Entscheidungsregel

Durch Einsetzen der Zerlegung:

$$C^*(x) = \arg \max_k P(C_k) \prod_{i=1}^d p(x_i|C_k).$$

Lernen bedeutet:

1. Schätzen der Priors  $P(C_k)$  (Häufigkeiten).
2. Schätzen der eindimensionalen Dichten  $p(x_i|C_k)$  für alle Merkmale und Klassen.

**Log-Space (Numerische Stabilität)** Da Produkte kleiner Wahrscheinlichkeiten numerisch instabil sind:

$$C^*(x) = \arg \max_k \left[ \log P(C_k) + \sum_{i=1}^d \log p(x_i|C_k) \right].$$

Logarithmen ersetzen Multiplikationen durch Additionen, ohne die Entscheidung zu verändern.

## 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,  $m \ll n$ ).

## 4 Fouriertheorie

### 4.1 Motivation

Die Fouriertheorie bietet eine neue Perspektive, um Signale und Funktionen zu analysieren, indem sie vom **Ortsraum** (z.B. Position, Zeit) in den **Frequenzraum** (z.B. Frequenz, Wellenlänge) wechselt.

- **Beispiele:**

- **Optik/Physik:** Lichtbeugung an einem Spalt. Ein Spalt kann mathematisch als **Rechteckfunktion** ( $rect(x)$ ) beschrieben werden. Das resultierende Beugungsmuster (die Amplitudenverteilung)  $B(\varphi)$  hat die Form einer **sinc-Funktion** ( $\frac{\sin x}{x}$ ). Die gemessene Intensität  $I(x)$  ist proportional zum Quadrat der Amplitude ( $I \propto B^2$ ), also eine  $sinc^2$ -Funktion.
- **Medizintechnik (MRT):** Ein MR-Scanner misst direkt Frequenzmuster im Fourierraum.
- **Menschliche Wahrnehmung:** Die Kontrastempfindlichkeit des Auges wird im Frequenzraum gemessen, oft mittels sinusförmiger Muster (sinusoidal gratings).

- **Kernidee:** Es besteht ein direkter Zusammenhang zwischen der Gestalt eines Objekts (Ortsraum) und seiner Amplitudenfunktion (Frequenzraum). Dieser Zusammenhang ist die **Fourier-Transformation**.

- **Erstes Fourier-Paar:**  $rect(x) \xrightarrow{FT} sinc(u)$ .

### 4.2 Mathematische Grundlagen

- **Vektorraum** ( $R^n$ ): Ein Raum, in dem Vektoren addiert und skalar multipliziert werden können.
- **Skalarprodukt:** Definiert Längen und Winkel.  $\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^n v_i w_i$ .
- **Basis:** Ein Satz linear unabhängiger Vektoren (z.B.  $\vec{e}_1, \vec{e}_2$ ), die den Raum aufspannen. Jeder Vektor  $\vec{v}$  ist eine Linearkombination der Basis:  $\vec{v} = a_1 \vec{e}_1 + a_2 \vec{e}_2$ .
- **Funktionenräume:**
  - Die Elemente sind nun **Funktionen** statt Vektoren.
  - Diese Räume sind i.d.R. **unendlich-dimensional**.
  - **Ziel:** Finde **Basisfunktionen**, um (beliebige) Funktionen als Linearkombination dieser Basen darzustellen. Dies leistet die **Fouriertheorie**.

### 4.3 Die Fourier-Reihe (Fourier Series)

#### Grundidee der Fourier-Reihe

Jede  **$2\pi$ -periodische Funktion**  $f(x)$ , die die **Dirichlet-Bedingungen** erfüllt, kann als unendliche Summe (Überlagerung) von Sinus- und Kosinusfunktionen dargestellt werden.

#### Dirichlet-Bedingungen :

1. Endliche Anzahl von Unstetigkeiten pro Periode.
2. Endliche Anzahl von Maxima und Minima pro Periode.
3. In jeder Periode integrierbar (d.h.  $\int |f(x)| dx < \infty$  pro Periode).

#### 4.3.1 Analogie zum Vektorraum

- **Skalarprodukt für Funktionen:**  $\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)g(t)dt$ .

- **Orthogonale Basis:** Die Funktionen  $u_n(t) = \cos(nt)$  und  $v_n(t) = \sin(nt)$  bilden eine **orthogonale Basis** im Funktionenraum  $H$ . Das bedeutet, ihr Skalarprodukt ist 0, wenn sie nicht identisch sind (z.B.  $\int_{-\pi}^{\pi} \cos(nt) \sin(mt) dt = 0 \forall n, m$ ).

#### 4.3.2 Formeln der Fourier-Reihe

##### Reelle Fourier-Reihe

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

Die **Fourier-Koeffizienten**  $a_n, b_n$  werden durch Projektion von  $f(x)$  auf die Basisfunktionen berechnet:

- $a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx$  (Der "Gleichanteil" oder Mittelwert)
- $a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(mx) dx$  (für  $m > 0$ )
- $b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(mx) dx$

- **Symmetrie-Eigenschaften:**

- **Gerade Funktion** ( $f(-t) = f(t)$ , z.B.  $\cos(x)$ ): Alle  **$b_n = 0$** .
- **Ungerade Funktion** ( $f(-t) = -f(t)$ , z.B.  $\sin(x)$ ): Alle  **$a_n = 0$**  (inkl.  $a_0$ ).

- **Beispiel (Rechteck-Schwingung):** Dies ist eine ungerade Funktion. Daher sind alle  $a_n = 0$ . Die  $b_n$  Koeffizienten sind  $\frac{4k}{n\pi}$  für ungerade  $n$  und 0 für gerade  $n$ .

$$f(x) = \frac{4k}{\pi} \left( \sin(x) + \frac{1}{3} \sin(3x) + \frac{1}{5} \sin(5x) + \dots \right)$$

(Die Folien 39-43 visualisieren, wie diese Summe die Rechteckfunktion approximiert).

##### Komplexe Fourier-Reihe

Mit der Euler-Identität ( $e^{i\phi} = \cos \phi + i \sin \phi$  und  $e^{inx} = \cos(nx) + i \sin(nx)$ ) lässt sich die Reihe kompakter schreiben :

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

#### 4.4 Die Fourier-Transformation (FT)

##### Motivation der Fourier-Transformation

Erweiterung der Fourier-Reihe auf **nicht-periodische Funktionen**.

Dies geschieht durch einen Grenzübergang: Die Periode  $L$  der Funktion wird unendlich groß ( $L \rightarrow \infty$ ).

Dabei wird die **diskrete Summe** der Fourier-Reihe (über  $n$ ) zu einem **kontinuierlichen Integral** (über  $u$ ).

Das Spektrum ist nicht mehr diskret (Vielfache einer Grundfrequenz), sondern **kontinuierlich**.

## Das Fourier-Transformationspaar

- **Fourier-Transformation (FT):** (Analyse: Ortsraum  $\rightarrow$  Frequenzraum)

$$F(u) = \int_{-\infty}^{\infty} f(t)e^{-2\pi iut} dt$$

- **Inverse Fourier-Transformation (iFT):** (Synthese: Frequenzraum  $\rightarrow$  Ortsraum)

$$f(x) = \int_{-\infty}^{\infty} F(u)e^{+2\pi iux} du$$

$f(x)$  ist oft reell, aber  $F(u)$  ist i.d.R. komplex:  $F(u) = \operatorname{Re}(F(u)) + i \operatorname{Im}(F(u))$ .

### 4.4.1 Wichtige Fourier-Transformationspaare

- **Rechteck-Funktion  $\leftrightarrow$  sinc-Funktion:**

–  $f(x) = \operatorname{rect}(x)$

–  $F(u) \propto \operatorname{sinc}(u)$

- **Dirac-Delta-Distribution  $\leftrightarrow$  Konstante:**

– Die **Dirac-Delta-Distribution**  $\delta(t)$  ist keine Funktion, sondern eine Distribution, definiert durch ihre **”Sampling-Eigenschaft”**:  $\int f(t)\delta(t-t')dt = f(t')$ .

–  $f(x) = \delta(x) \xrightarrow{FT} F(u) = 1$  (Ein Impuls im Ort enthält alle Frequenzen)

–  $f(x) = K \xrightarrow{FT} F(u) = K \cdot \delta(u)$  (Eine Konstante im Ort ist nur die Frequenz 0)

- **Kosinus  $\leftrightarrow$  Zwei Delta-Peaks:**

–  $f(x) = \cos(kx)$

–  $F(u) \propto (\delta(u-k) + \delta(u+k))$

– Eine reine Schwingung im Ortsraum entspricht zwei scharfen Peaks (diskreten Frequenzen) im Frequenzraum.

## 4.5 Faltung und Filterung

### Faltung (Convolution)

Das Faltungsintegral  $h(t)$  zweier Funktionen  $f$  und  $g$  ist definiert als:

$$h(t) = (f \circ g)(t) = \int_{-\infty}^{\infty} f(x)g(t-x)dx$$

Graphisch entspricht dies einer ”Spiegelung, Verschiebung, Multiplikation und Integration”.

- Bsp:  $\operatorname{rect}(x) \circ \operatorname{rect}(x) = \operatorname{triangle}(x)$ .

### Der Faltungssatz (Convolution Theorem)

Dies ist einer der wichtigsten Sätze der Fouriertheorie:

- Eine **Faltung im Ortsraum** entspricht einer **Multiplikation im Frequenzraum**.
- Eine **Multiplikation im Ortsraum** entspricht einer **Faltung im Frequenzraum**.

$$h(t) = f(t) \circ g(t) \xrightarrow{FT} H(u) = F(u) \cdot G(u)$$

### 4.5.1 Anwendung: Filterung

---

Der Faltungssatz ist die Grundlage für effiziente Filterung :

1. Ein Signal  $f(x)$  soll gefiltert werden.
2. Anstatt eine aufwändige Faltung im Ortsraum ( $f \circ g$ ) durchzuführen...
3. ...transformiert man Signal ( $f \rightarrow F$ ) und Filterfunktion ( $g \rightarrow G$ ) in den Frequenzraum.
4. Dort wird eine **einfache Multiplikation** durchgeführt:  $H(u) = F(u) \cdot G(u)$ .
5. Das Ergebnis  $H(u)$  wird zurück in den Ortsraum transformiert ( $H \rightarrow h$ ), um das gefilterte Signal  $h(x)$  zu erhalten.

Dies ist oft (via *Fast Fourier Transform*, FFT) viel schneller als die direkte Faltung.

## 4.6 Abtastung von Signalen (Sampling)

---

### 4.6.1 Modell der Abtastung

---

- **Problem:** Ein kontinuierliches Signal  $f(x)$  muss in diskrete Werte  $f(n\Delta x)$  für die digitale Verarbeitung umgewandelt werden.
- **Mathematisches Modell:** Abtastung ist eine **Multiplikation** des Signals  $f(x)$  mit einer **Kamm-Funktion** (einer Kette von  $\delta$ -Impulsen im Abstand  $\Delta x$ ).

$$\hat{f}(x) = f(x) \cdot \sum_{n=-\infty}^{\infty} \delta(x - n \cdot \Delta x)$$

### 4.6.2 Abtastung im Frequenzraum

---

- Laut Faltungssatz (Multiplikation im Ortsraum  $\rightarrow$  Faltung im Frequenzraum) wird das Spektrum  $F(u)$  des Signals mit der FT der Kamm-Funktion gefaltet.
- Die FT einer Kamm-Funktion (Abstand  $\Delta x$ ) ist wieder eine Kamm-Funktion (Abstand  $1/\Delta x$ ).
- **Folge:** Das Spektrum  $\hat{F}(u)$  des abgetasteten Signals ist eine **periodische Wiederholung** des Originalspektrums  $F(u)$ , wobei die Kopien den Abstand  $1/\Delta x$  haben.

### 4.6.3 Aliasing und das Abtasttheorem

---

Angenommen, das Signal ist **bandbegrenzt**, d.h. seine höchste Frequenz ist  $u_G$  ( $F(u) = 0$  für  $|u| > u_G$ ).

- **Fall 1: Korrekte Abtastung** ( $1/\Delta x > 2u_G$ )
  - Die Abtastfrequenz ( $1/\Delta x$ ) ist mehr als doppelt so hoch wie die maximale Signalfrequenz ( $u_G$ ).
  - Die periodischen Kopien von  $F(u)$  im Frequenzraum **überlappen nicht**.
  - Das Originalsignal kann **fehlerfrei rekonstruiert** werden (z.B. durch einen Tiefpassfilter, der nur die zentrale Kopie isoliert).
- **Fall 2: Unterabtastung** ( $1/\Delta x < 2u_G$ )
  - Die Abtastfrequenz ist zu niedrig.
  - Die Kopien von  $F(u)$  **überlappen sich**.
  - In den Überlappungsbereichen addieren sich Frequenzen. Hohe Frequenzen "erscheinen" fälschlicherweise als niedrige Frequenzen.
  - Dieser irreversible Fehler wird **Aliasing** genannt. (z.B. Wagenrad-Effekt bei Filmen , Moiré-Muster ).



### Abtasttheorem von Whittaker-Shannon

Eine bandbegrenzte Funktion (höchste Frequenz  $u_G$ ) kann aus ihren Abtastwerten  $f(n\Delta x)$  fehlerfrei rekonstruiert werden, wenn die Abtastfrequenz  $f_s = 1/\Delta x$  **mindestens doppelt so hoch** wie die höchste Signalfrequenz  $u_G$  ist.

$$f_s > 2u_G \quad \text{oder} \quad \frac{1}{\Delta x} > 2u_G$$

Die Frequenz  $2u_G$  wird als **Nyquist-Frequenz** bezeichnet.

## 4.7 Zusammenfassung der Konzepte

- **Fourier-Reihe:**

- **Für:**  $2\pi$ -periodische Funktionen.
- **Spektrum:** **Diskret** (Vielfache einer Grundfrequenz).
- **Formel:**  $f(x) = \sum c_n e^{inx}$ .

- **Fourier-Transformation:**

- **Für:** Nicht-periodische Funktionen.
- **Spektrum:** **Kontinuierlich**.
- **Formel:**  $F(u) = \int f(t) e^{-2\pi i u t} dt$ .

- **Faltungssatz:**

- $f(t) \circ g(t) \longleftrightarrow F(u) \cdot G(u)$
- Ermöglicht effiziente **Filterung** im Frequenzraum.

- **Abtasttheorem:**

- Abtastung (Multiplikation mit  $\delta$ -Kamm) im Ortsraum  $\rightarrow$  Periodisierung (Faltung mit  $\delta$ -Kamm) im Frequenzraum.
- Zur Rekonstruktion muss die Abtastfrequenz  $f_s$  größer als die doppelte maximale Signalfrequenz  $u_G$  sein ( $f_s > 2u_G$ ), um **Aliasing** zu vermeiden.