

Visual Computing

Niclas Kusenbach
LaTeX version:  SCHOUTER

Table of Contents

1 Einführung in Visual Computing	2
1.1 Grundlagen Visual Computing	2
1.2 Vier exemplarische Themenbereiche	2
1.3 Wichtige Konzepte und Begriffe	2
1.4 Empfohlene Literatur	2
1.5 Zusammenfassung	3
2 Wahrnehmung (Perception)	4
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

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 strukturierte 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)

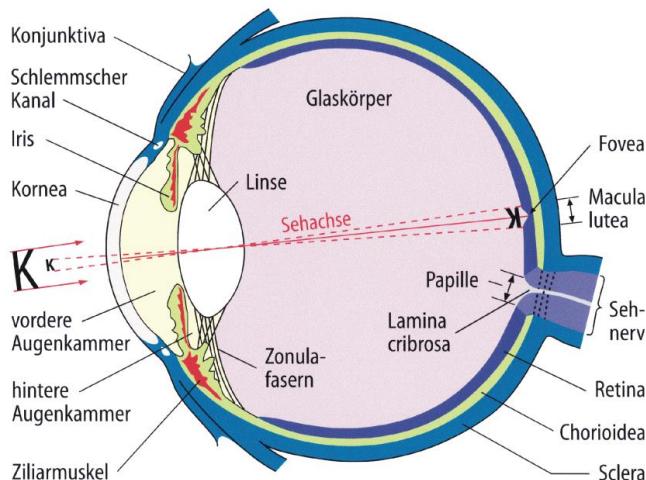
- **Warum VC? (Why VC?)** Die Leistungsfähigkeit von Rechnern wächst exponentiell (**Moore's Law**), aber die Kapazität von Menschen ist (fast) konstant (**Darwin's Law**). VC hilft, diese Lücke zu überbrücken.
- **Kognition (Cognition):** Sammelbegriff für alle Prozesse des Wahrnehmens und Erkennens (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): ≈ 100 ms
 - Entscheidung (Cognition): ≈ 70 ms
 - Reaktion (Motor): ≈ 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 v und Wellenlänge λ hängen über $v \cdot \lambda = c$ zusammen.

2.2.1 Aufbau des Auges (Structure of the Eye)



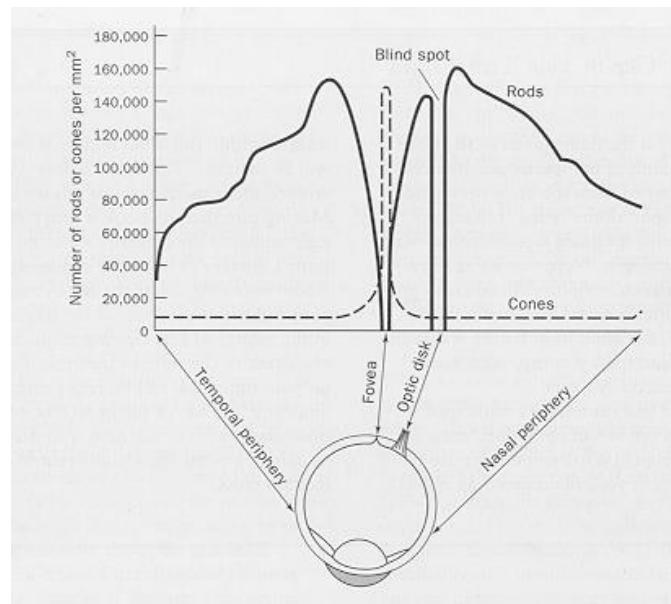
- **Optische Elemente:** Hornhaut (Kornea), Linse, Iris (Blende, 2-8mm), Glaskörper.

- **Linse (Lens):** Akkommodation (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)

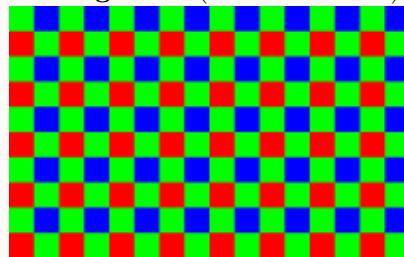
Es gibt zwei Haupttypen von Photorezeptoren auf der Retina:

- **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).

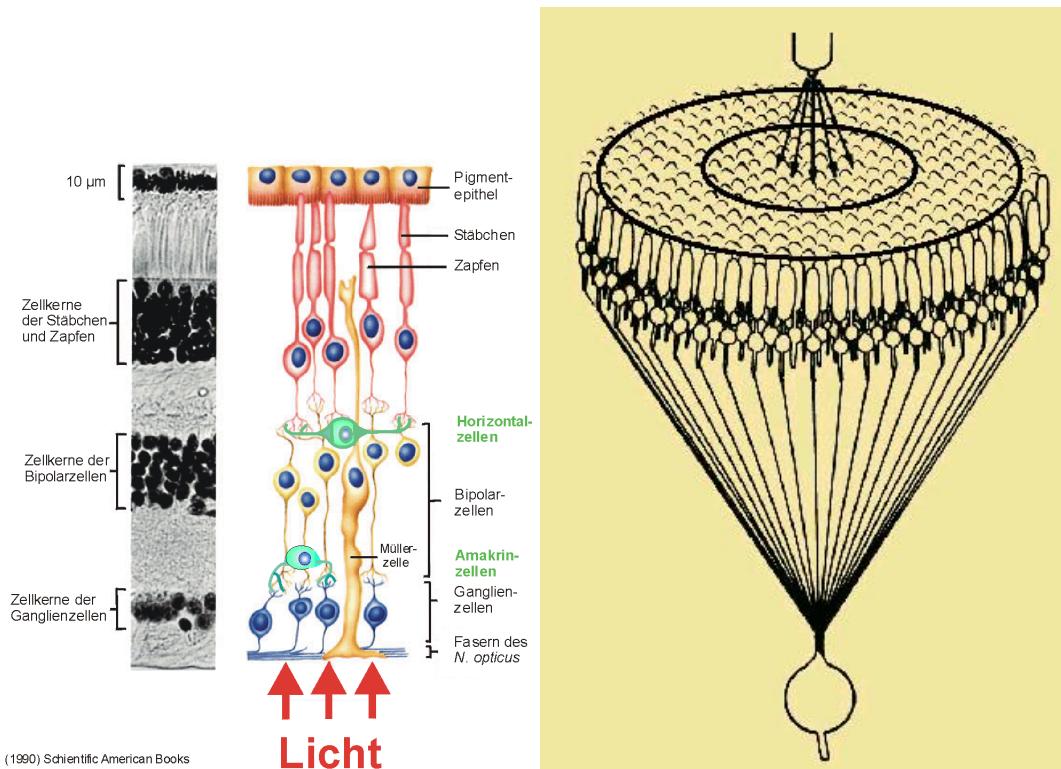


2.3 Vorverarbeitung & Helligkeit (Preprocessing & Brightness)

2.3.1 Signalverarbeitung in der Retina

Das Licht trifft (paradoxerweise) 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).



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 Hellemmpfindung (L).
 - **Webersches Gesetz (Schwelle):** $\Delta L = \frac{\Delta R}{R} = const.$ (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.
 - **Akkomodation** (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. 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 ± 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).