

Grundzüge der Kognition und Wahrnehmung

Zusammenfassung



Inhaltsverzeichnis

1 Wahrnehmung vs Kognition	1
1.1 Wahrnehmung	1
1.2 Percept	1
1.3 Kognition	1
1.4 Sensory System	1
1.5 Perception-Action Loop	1
1.6 Plenoptic Function	2
1.7 Wahrnehmungspsychologie	3
1.7.1 Physiological	3
1.7.2 Imaging	3
1.7.3 Psychophysical	3
1.7.3.1 Grenzmethode	3
1.7.3.2 Herstellungsmethode	3
1.7.3.3 Konstanzmethode	3
2 Visual Sensitivity	4
2.1 Licht	4
2.2 Licht Energie	4
2.3 Acuity	4
2.3.1 Stäbchen	5
2.3.2 Zapfen	5
2.3.3 Luminous efficiency curves	5
2.4 Receptive Fields	6
2.5 Contrast	6
2.6 Contrast Sensitivity	6
2.7 Temporal Sensitivity	7
2.8 Masking	7
2.9 Dynamic Range	8
3 2D Image Features	8
3.1 Kante	8
3.2 Gradient Computation	9
3.2.1 Sobel Operator	9
3.3 Gradient Distribution	9
3.4 Gaussian Scale Space	10
3.5 Mach Bands	11
3.6 Craik-O'Brian-Cornsweet Illusion	11
3.7 Junctions	12
3.7.1 T-Junctions	12
3.7.2 X-Junctions	13
3.8 Gabor Patch	13
3.9 Orientation Hypercolumn	14

3.10 Gestalt Laws	15
4 Farbe und Räumliche Tiefe	16
4.1 Farbe	16
4.1.1 Metamerie	16
4.2 Farbe: Transduktion	16
4.3 Farbe: Percept	17
4.3.1 Farbmessung	17
4.3.1.1 dominante Wellenlänge	18
4.3.1.2 Erregungsreinheit	18
4.3.2 Farbton (hue)	18
4.3.3 Sättigung (saturation)	18
4.3.4 Helligkeit (lightness)	18
4.4 Farbe: Farbmessung	18
4.4.1 Darstellung am Monitor	18
4.4.2 Farbe: Abhängigkeiten	19
4.4.3 Opponent Processing	20
4.5 Farbe: Messung	21
4.6 Farbe: Color Spaces	22
4.6.1 CIE XY \bar{Z} (1932)	22
4.6.1.1 Spektraler Lokus	23
4.6.1.2 Purpurlokus	23
4.6.1.3 Weißpunkt	23
4.6.1.4 Gammut	23
4.6.1.5 Dominante Wellenlänge	24
4.6.1.6 Erregungsreinheit	24
4.6.1.7 Gegenfarbe	24
4.6.1.8 Farbdifferenzen	24
4.6.2 Farbmischung	25
4.6.3 Munsell Farbsystem (1905)	25
4.6.4 HLS	26
4.6.5 CIE u'v'Y (1976)	26
4.6.6 CIE L*u*v* (1976)	27
4.6.7 CIE L*a*b* (1976)	27
4.6.8 LMS	27
4.7 Farbe: Fehlsichtigkeit	27
4.7.1 Arten	27
4.7.2 Verwirrungslinien	28
4.8 Farbe: Color Constancy	28
4.9 Räumliche Tiefe	29
5 2 Dimensionale Bewegungswahrnehmung u. Optischer Fluss	31
5.1 Static Bias: Motion Calculation	31
5.2 Change Detection/Motion Detection	32
5.3 Apparent Motion	34
5.4 Phi Motion	34
5.5 Wagon Wheel Illusion	34
5.6 Tunneling-Effect	35
5.7 First Order Motion	35

5.8 Second Order Motion	35
5.9 Motion Contrast	36
5.10 Hierachical and Biological Motion	36
5.11 Motion Aftereffect	36
5.12 Aperture Problem	37
5.13 Optic Flow	38
5.13.1 Anwendung	39
5.13.2 Einschränkungen	39
5.14 Reichardt-Detector	40
6 Textur Wahrnehmung	40
6.1 Textur Definition	40
6.2 Structural	41
6.2.1 Erkennung	42
6.3 Statistical	42
6.3.1 Neuronale Netze	42
6.3.2 Deep Learning	43
6.3.3 Convolution Neural Networks	44
6.3.3.1 Convolution	44
6.4 Oberflächen Orientierung	46
7 Beleuchtung und Schatten	47
7.1 Kumulierte Licht	48
7.2 Ambientes Licht	48
7.3 Zusammenhang Helligkeit Winkel	49
7.4 Lambertsches Gesetz für Reflektion	49
7.5 Diffuse Beleuchtung	50
7.6 Vigneting	50
7.7 Hemispherical Diffuse Illumination	51
7.8 Realität	51
7.9 Shape From Shading	52
7.9.1 Computer Vision	52
7.9.2 Human Vision	52
7.10 Estimating Lighting	52
7.11 Intrinsic Image Analysis	53
7.12 Schatten	54
8 Material Eigenschaften	55
8.1 Surface Appearance	55
8.2 Source Separation	56
8.3 Approaches	56
8.4 BRDF	57
8.5 Lightness	57
8.6 Brightness	57
8.7 Simulaneous Contrast	58
8.8 Gelb-Effekt	58
8.9 Specular Reflection	58
8.10 Transparency	59
8.11 Subsurface Scattering	60
9 Aufmerksamkeit	61
9.1 Experimental Design	61

9.2 Reizinduzierte vs. Zielgesteuerte Auswahl	62
9.3 menschliche Informationsverarbeitung	62
9.4 Psychologische Refraktärperiode nach Welford (1952)	63
9.5 Shadowing	65
9.6 Broadbents Filtertheorie (1958)	65
9.6.1 Kritik	65
9.7 Shadowing-Experiment von Gray & Wedderburn (1960)	66
9.8 Dämpfungstheorie nach Treisman (1964)	66
9.9 Late Selection Theorie nach Deutsch & Deutsch (1963)	67
9.10 Experiment von Treisman & Geffen (1967)	67
9.11 Aktuelle Erklärung	68
9.12 Visuelle Aufmerksamkeit	68
9.12.1 Visuelle Aufmerksamkeit im visuellen Feld	68
9.12.2 Spatial-Cueing-Experiment von Posner (1980)	69
9.12.2.1 Cues	69
9.12.3 Metapher des "Attentional Spotlight"	70
9.12.4 "Zooming In"	70
9.12.5 Visual Neglect	71
9.12.6 Raum- und Objektgebundenheit	71
9.12.7 Verarbeitung visueller Reize	71
9.13 Feature Integration Theory	72
9.13.1 Annahme	72
9.13.2 Prozesse	72
9.13.2.1 Merkmalsanalyse	72
9.13.2.2 Merkmalsintegration	73
9.13.2.3 Merkmalsvergleich	74
9.13.3 Experiment von Treisman & Gelade	74

1. Wahrnehmung vs Kognition

1.1. Wahrnehmung

Aufnahme von Daten → Extrahieren, Verarbeiten, Repräsentieren und Organisieren von Informationen aus der Sinneswahrnehmung als **Percept**, nicht unbedingt eine genaue Abbildung der Realität

1.2. Percept

Ergebnis des gesamten Wahrnehmungsprozess, der Vorgänge wie Zusammenfügen, Schätzen, Urteilen, Erinnern, Vergleichen und Assoziieren umfasst

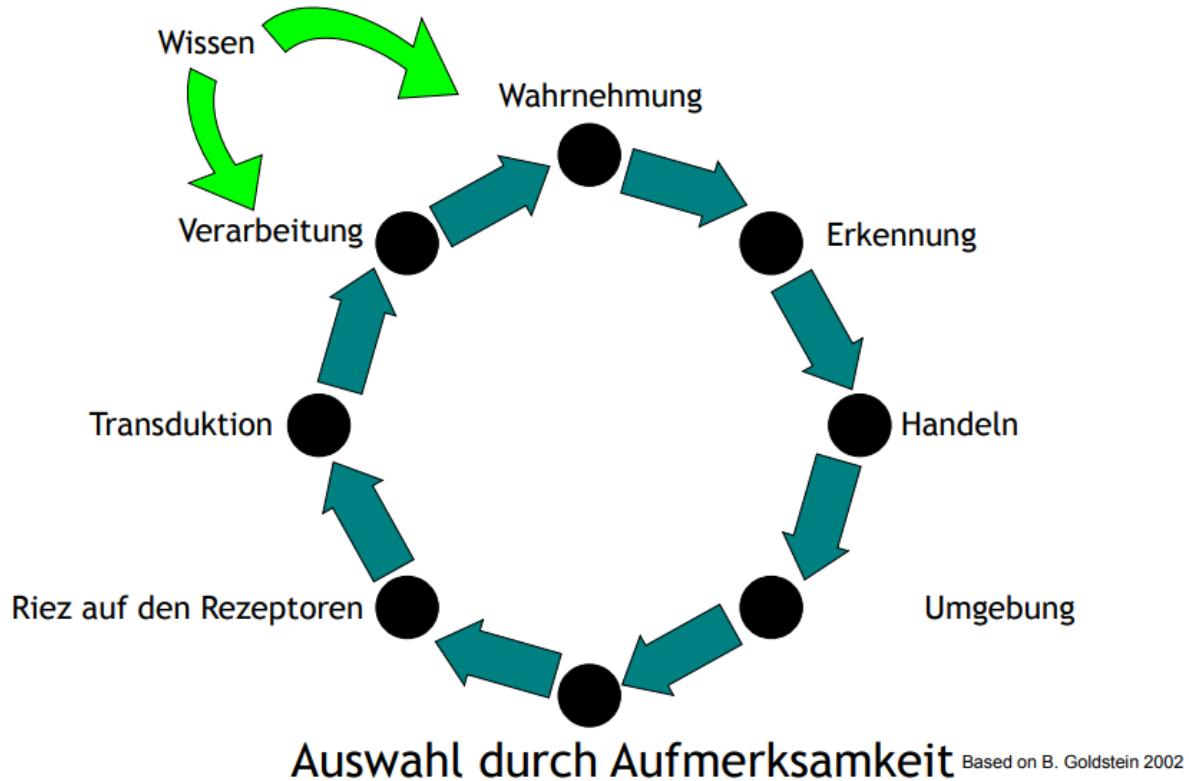
1.3. Kognition

Verarbeitung, Zusammenführung, Kategorisierung und Einordnung von **Percept** → Erkenntnis/Erkennung

1.4. Sensory System

Unsere Sinnesorgane fungieren als Wahrnehmungssystem → Wahrnehmung steuert das Verhalten → Verhalten bedingt die Eingabe für die Wahrnehmung ⇒ Perception-Action Loop

1.5. Perception-Action Loop



- Umgebung: alles was uns umgibt
- Auswahl durch Aufmerksamkeit: bewusste oder unbewusste, gelenkte Auswahl der wahrgenommenen Reize
- Reiz auf Rezeptoren: Sinnesorgane (Augen, Ohren, Haut, Nase, Zunge, Balance) nehmen Reize auf, auf denen die Wahrnehmung beruht
- Transduktion: physische Signal (z.B. Licht-> Richtung, Intensität, Wellenlänge) wird in neurochemische Informationen übersetzt/umwandeln
- Verarbeitung: 50-60% Auslastung des Gehirn für Wahrnehmung
- Erkennung: Kognition
- Handeln: Reaktion auf verarbeitetes Wissen, basierend auf früherer Erfahrungen

1.6. Plenoptic Function

beschreibt jeden Lichtstrahl in einer Szene → **Light Field**
 basiert auf Lochkamera → Loch bündelt alle Strahlen in einem Punkt

$$l(x, y, z, \theta, \omega, \lambda, t) \quad (1.1)$$

-
- $x, y, z \rightarrow$ Position des Lochs
 - $\theta, \omega \rightarrow$ Winkel des Strahls im Raum
 - $\lambda \rightarrow$ Wellenlänge des Lichts
 - $t \rightarrow$ Zeitpunkt

1.7. Wahrnehmungpsychologie

1.7.1. Physiological

Wirkung des Reiz/Stimulus im Kopf mittels Nadeln messen

1.7.2. Imaging

Bild/Stimulus zeigen und im MRT Gehirn-Aktivität messen

1.7.3. Psychophysical

Stimulus zeigen und Reaktion beobachten

Es gibt die 3 Methoden zum Messen:

1.7.3.1. Grenzmethode

per Fragen die Grenze abtasten

1.7.3.2. Herstellungsmethode

Grenze per Regler selber abtasten

1.7.3.3. Konstanzmethode

feste Bilder abfragen

2. Visual Sensitivity

2.1. Licht

Licht ist eine in Photonen gequantezte transversale elektromagnetische Welle

Für die Computergrafik sind drei Eigenschaften von Relevanz:

- Richtung
- Intensität
- Wellenlänge

Von Bedeutung sind dabei aber **nicht absolute** Werte, sondern **relative** Bezüge zwischen benachbarten Punkten bzw. deren Unterschiede

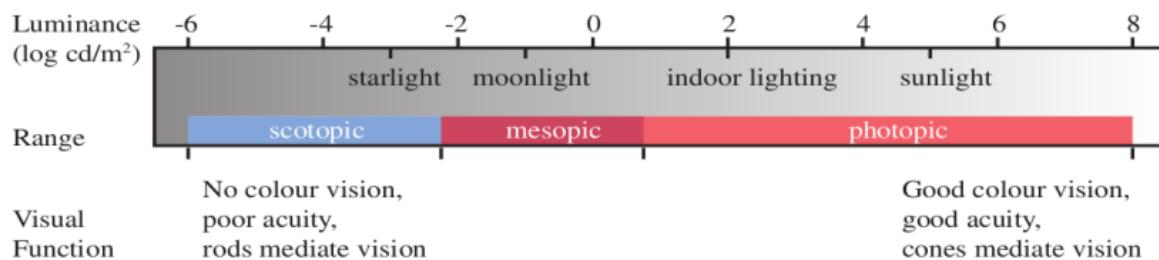
2.2. Licht Energie

- Radiometric → Physik von Strahlung
- Photometric → basiert auf menschlicher Wahrnehmung
- Photometric Brightness → Kraft der emittierenden Lichtquelle, gewichtet mit der visuellen, spektralen Wahrnehmungsempfindlichkeit, Einheit: Lumen
- Candela (Einheit) → Lichtkraft (Lumen) (durch spektrale Empfindlichkeit gewichtet) pro Richtung

2.3. Acuity

visuelle Sehkraft → wie viele Zyklen einer Sinuswelle pro Sehwinkelgrad im Auge aufgelöst werden können (am Besten bei 60 Cycles per Degree / 100 $\frac{\text{cd}}{\text{m}^2}$, Innenraum Beleuchtung)

- scotopic: Helligkeit wie bei Sternenlicht → keine Farbe, schlechte Auflösung, nur die Stäbchen arbeiten
- mesopic: Helligkeit wie bei Mondschein
- photopic: Helligkeit wie bei Innenraum/Sonnenlicht → Farben gut sehen, gute Auflösung, die Zapfen arbeiten primär



2.3.1. Stäbchen

- große Fläche abdecken -> sammeln über zeitlichen Bereich
- Hell/Dunkel -> kein Farbsehen
- schnelle Reaktion

2.3.2. Zapfen

- nehmen Farben wahr (3 Arten: Rot, Grün, Blau)
- genaue Lokalisation
- langsame Reaktion, nehmen wenig Licht auf (sammeln nicht)

2.3.3. Luminous efficiency curves

Scotopic Peak bei 507nm, Photopic Peak 555 nm → im Dunkeln besser Blau und Grün erkennen

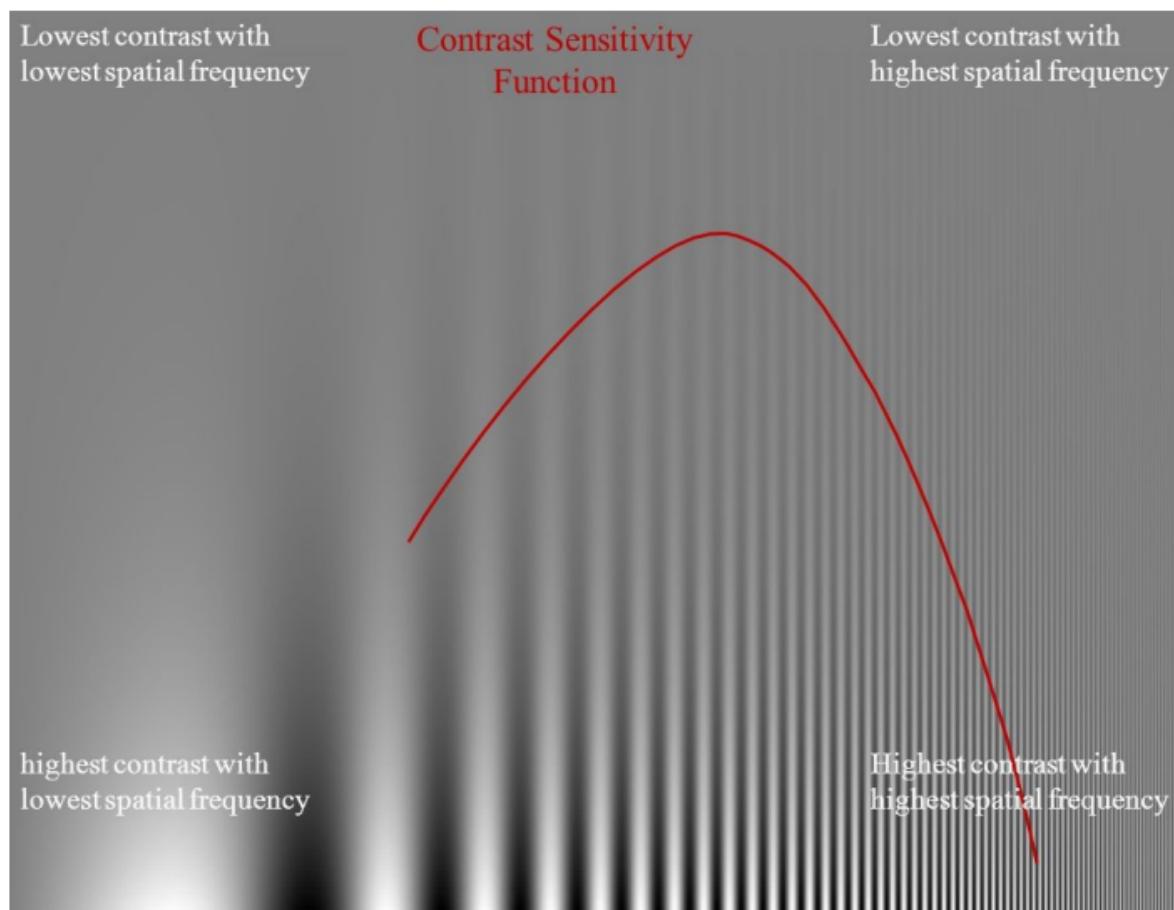
2.4. Receptive Fields

Center-Surround → schmaller Gauß-Filter

2.5. Contrast

Kontrast ist der relative Unterschied zwischen benachbarten Punkten
Michelson Contrast: zwischen maximalen und minimalen Werten (0 min., 1 max. Kontrast)

2.6. Contrast Sensitivity



- Kontrast Empfindlichkeit bei mittel hohen Frequenzen am besten

- Kontrast Empfindlichkeit steigt mit der Helligkeit

2.7. Temporal Sensitivity

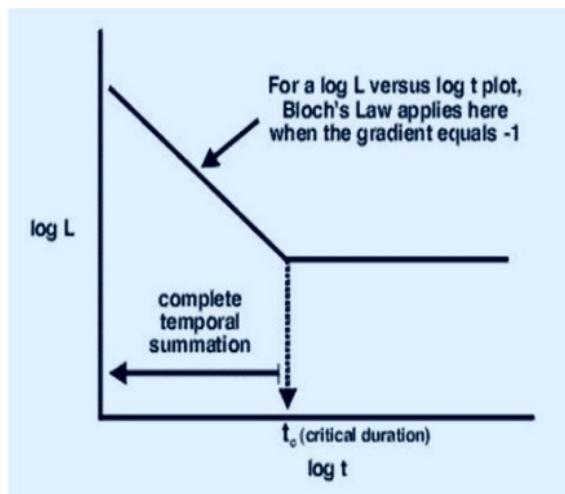
Verarbeitung des Signals im Gehirn bedarf Zeit → wenn die Veränderung des Lichts zu langsam ist, sieht man ein Flackern

Bei einem Licht im photopic light Level reagieren wir am empfindlichsten auf zu langsame Veränderungen

Blochs Gesetz:

Belichtungszeit (80ms) → Photonen sammeln

wenn nur für kurze Zeit sichtbar → weniger Photonen, wirkt dunkler
ab 80ms linear → gleichbleibend hell

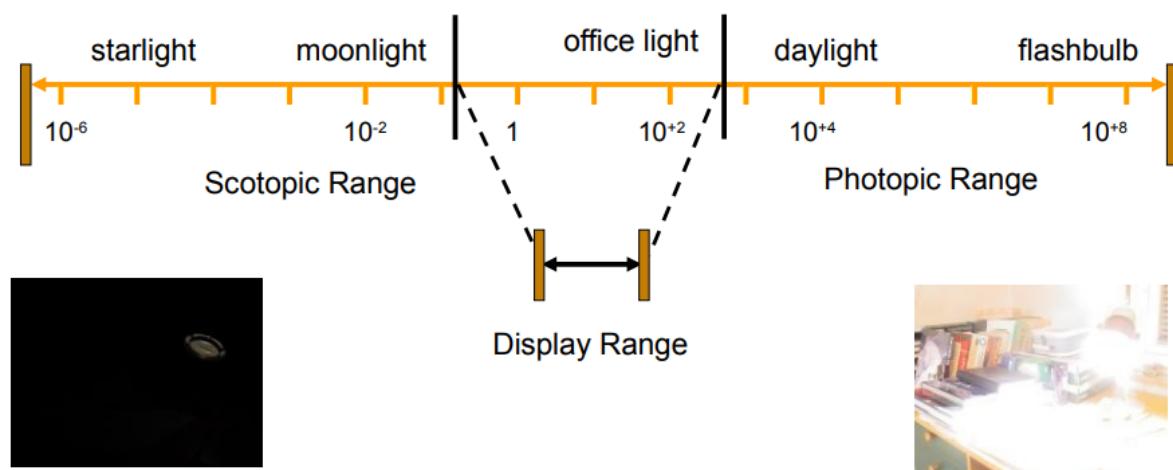


2.8. Masking

Die Kombination von zwei Mustern, die mit einander interagieren, beschreibt man mit einem Rauschen und einem Signal:

- Rauschen verschlechtert das Signal → maskiert
- Rauschen ähnlich zum Signal → konstruktive Interferenz → verbessert

2.9. Dynamic Range



Bildschirme/Kameras haben eine Display Range von ungefähr 10^{-1} bis 10^3 , alles unterhalb wird schwarz, alles oberhalb wird weiß



Mithilfe von **Tone Mapping Functions** versucht man, alle Helligkeitsbereiche einer Szene durch Bilder mit verschiedenen Belichtungszeiten in einem gesamten Bild zu integrieren

3. 2D Image Features

3.1. Kante

Oberflächen → große homogene Bereiche
Kante → Ort, wo sich zwei Oberflächen treffen,
→ schnelle, große Änderung der Helligkeit mit räumlicher Ausdehnung

3.2. Gradient Computation

1. Ableitung → Steigung

Kante → große Änderung → Maximum 1. Ableitung → Nulldurchgang 2. Ableitung

Bestimmung 2. Ableitung rechenintensiv → Gradienten Approximation
Gradient → speichert für einen Pixel in welche Richtung (Direction) und mit welcher Stärke (Magnitude) die Helligkeit steigt (→ Ableitung in einem Punkt).

3.2.1. Sobel Operator

Der Sobel Operator ist eine Approximation für einen Gradienten

Sobel Operator

$$I_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}$$
$$I_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

man schiebt den Operator über die Matrix und addiert die Werte der Matrix, multipliziert mit den Koeffizienten des Operator, auf. Am Rand muss man dann für leere Werte 0 nehmen.

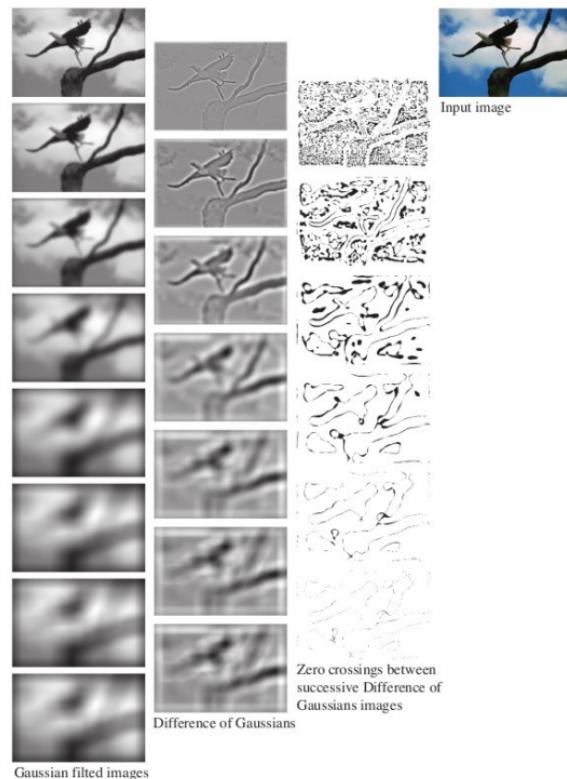
3.3. Gradient Distribution

Bei der Gradientenverteilung in Bildern von Natur sowie von menschlichen Bauten gibt es Gemeinsamkeiten:

- Scharfer Peak bei null: Dies bedeutet, dass die Welt größtenteils aus homogenen Flächen besteht, also aus Bereichen ohne starke Helligkeitsunterschiede.
- Schneller Abfall: Es gibt nur wenige Kanten, und diese sind meist von geringem Kontrast.
- Symmetrischer Abfall: Flächen mit ähnlichem Hintergrund erzeugen eine symmetrische Gradientenverteilung.

3.4. Gaussian Scale Space

Um Kanten zu bestimmen, muss man auch benachbarte Punkte berücksichtigen. Dies kann man auch umsetzen, in dem man über das Bild mit einem Gauß-Filter schmiert:

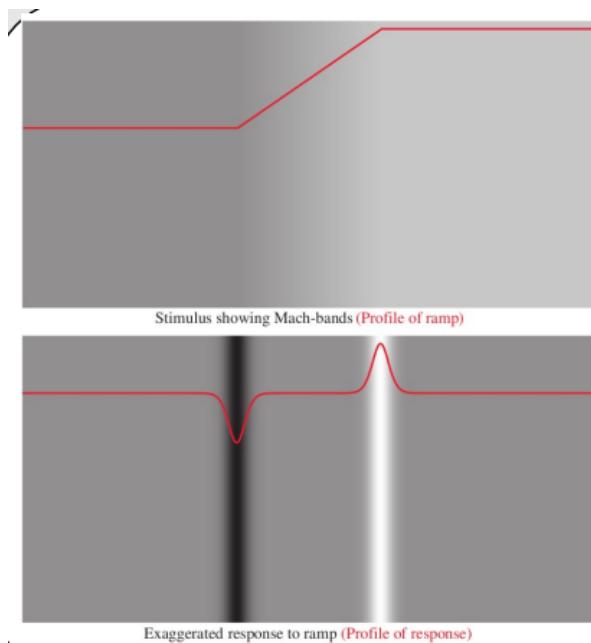


1. Über das Bild mit einem Gauß-Filter schmieren (\rightarrow Rausch verschwindet)
2. Schmiere über das bereits geschmierte Bild erneut, und wiederhole bis der Scale Space aufgebaut ist (\rightarrow verschieden starke Schmier-Stufen)

3. benachbarte Stufen von einander subtrahieren
4. Kanten detektieren (\rightarrow Zero-Crossing \rightarrow Gauß-Schmieren entspricht
2. Ableitung, Nulldurchgang finden)

3.5. Mach Bands

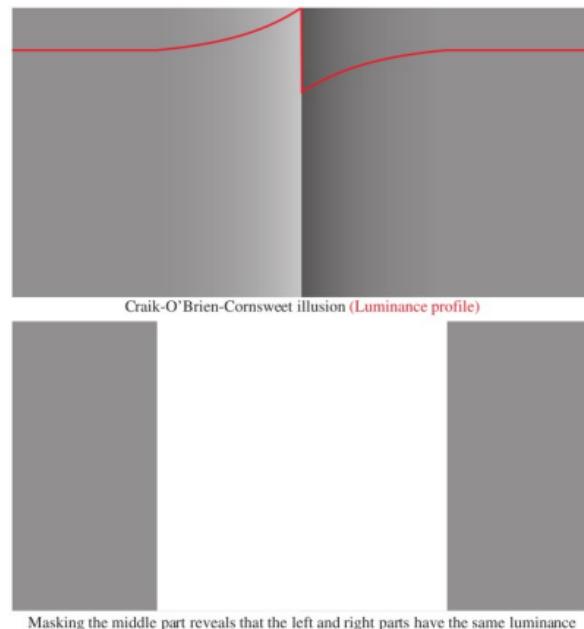
- Beschreibung: Wenn ein gleichmäßiger Helligkeitsgradient plötzlich ändert, kommt es zu einer wahrgenommenen Über- oder Unterschreitung der Helligkeit, den sogenannten Mach-Bändern.
- Problem: Die rezeptiven Felder des visuellen Systems betonen Kanten, was zu einer übertriebenen Wahrnehmung von Kanten führt. Dadurch erscheinen Kanten heller oder dunkler, als sie tatsächlich sind, was eine visuelle Illusion erzeugt.



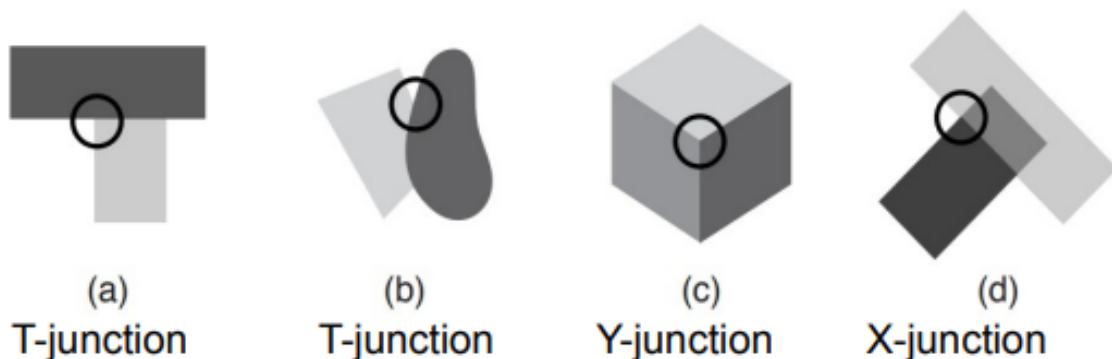
3.6. Craik-O'Brian-Cornsweet Illusion

- Beschreibung: Diese Illusion zeigt einen Stufenübergang, der von sanft variierenden Rampen flankiert wird. Visuell entsteht der Eindruck, dass die linke Seite des Bildes heller ist als die rechte.

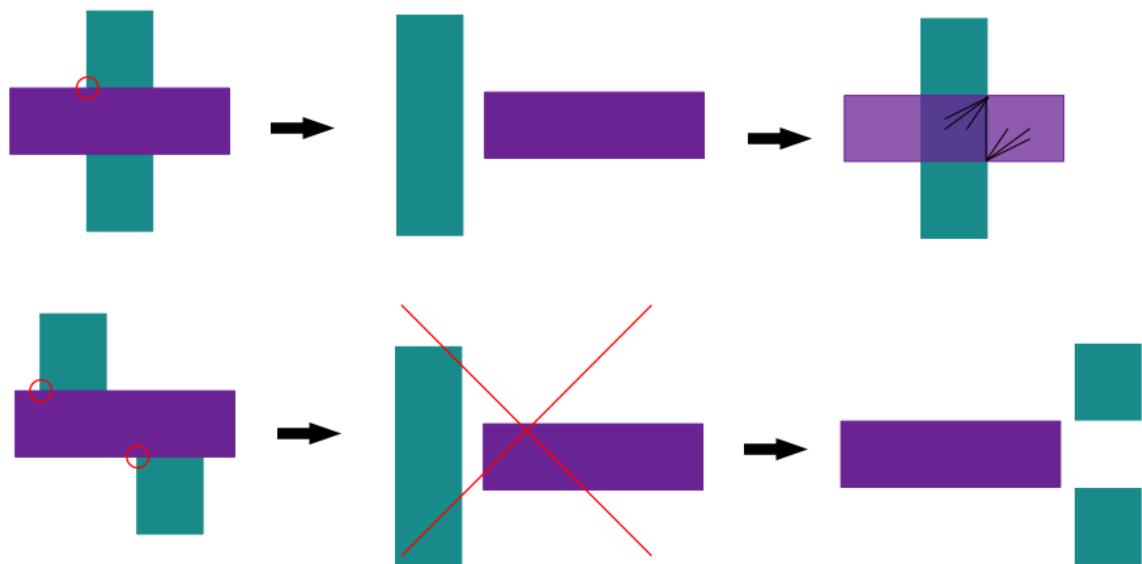
- Problem: Auch hier betonen die rezeptiven Felder Kanten, wodurch ein falscher Eindruck von Helligkeit entsteht. Wenn man den mittleren Bereich der Rampen maskiert, stellt sich heraus, dass die beiden Seiten tatsächlich dieselbe Helligkeit haben.



3.7. Junctions

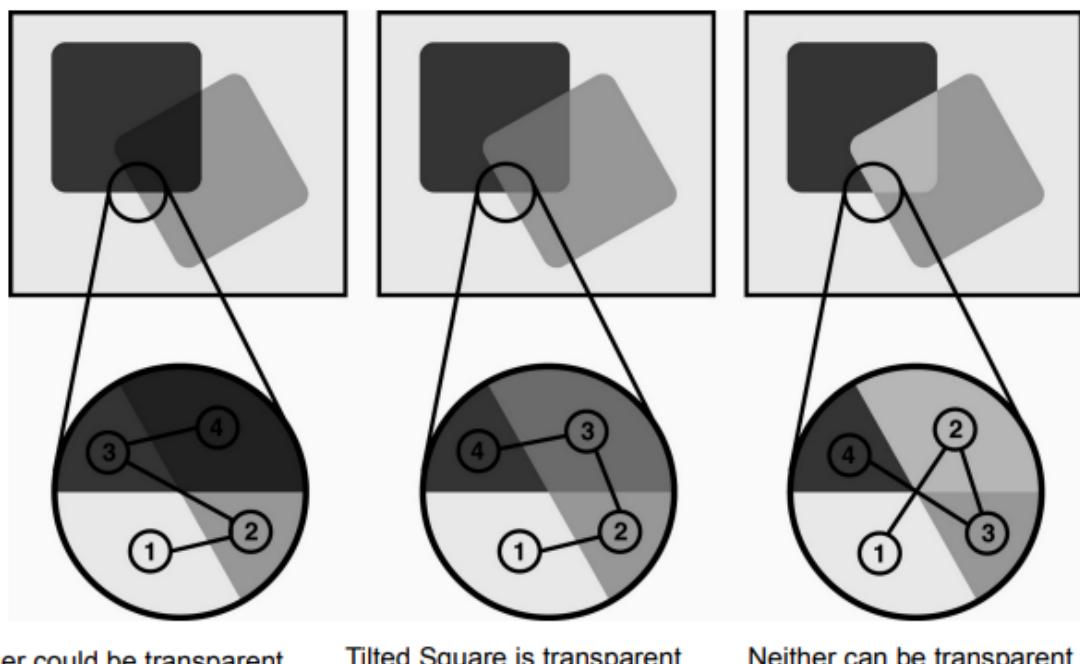


3.7.1. T-Junctions



T-Junction → Objekt geht hinten weiter

3.7.2. X-Junctions



X-Junction → entscheidet über Transparenz

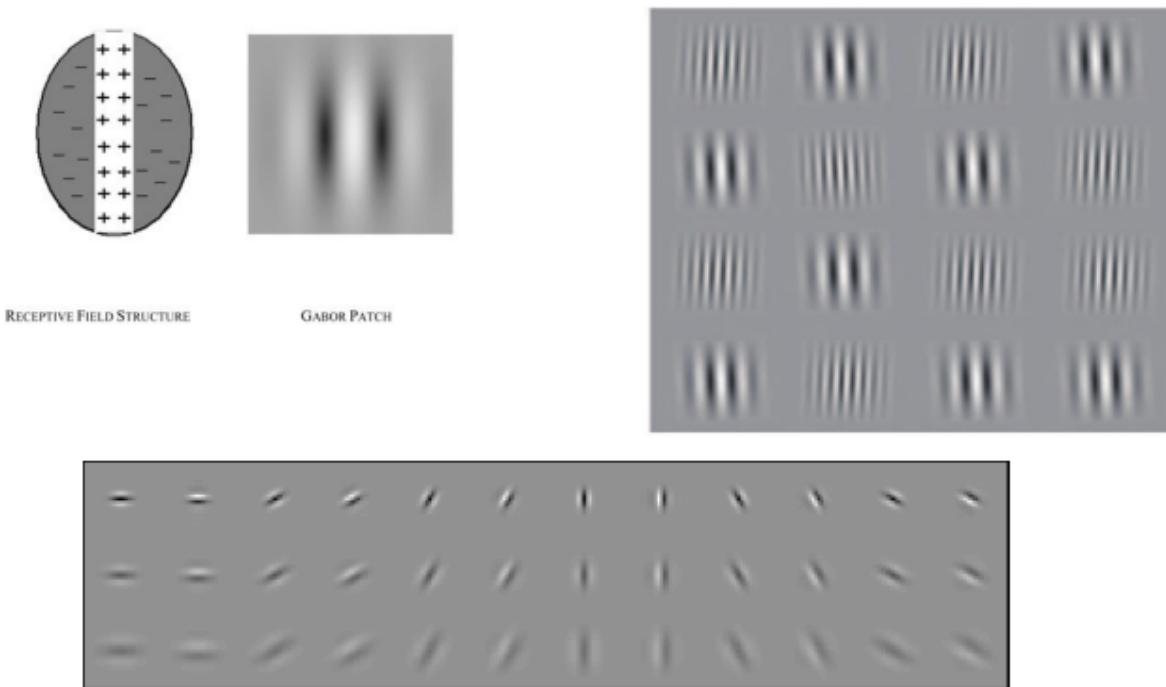
3.8. Gabor Patch

Ein Gabor Patch ist ein Gradienten Fleck und speichert 3 Eigenschaften:

- Orientierung (Ausrichtung)
- Kontrast (Unterschied Bänder und Hintergrund)
- Frequenz (Anzahl der sichtbaren Bänder)

Er findet Verwendung bei:

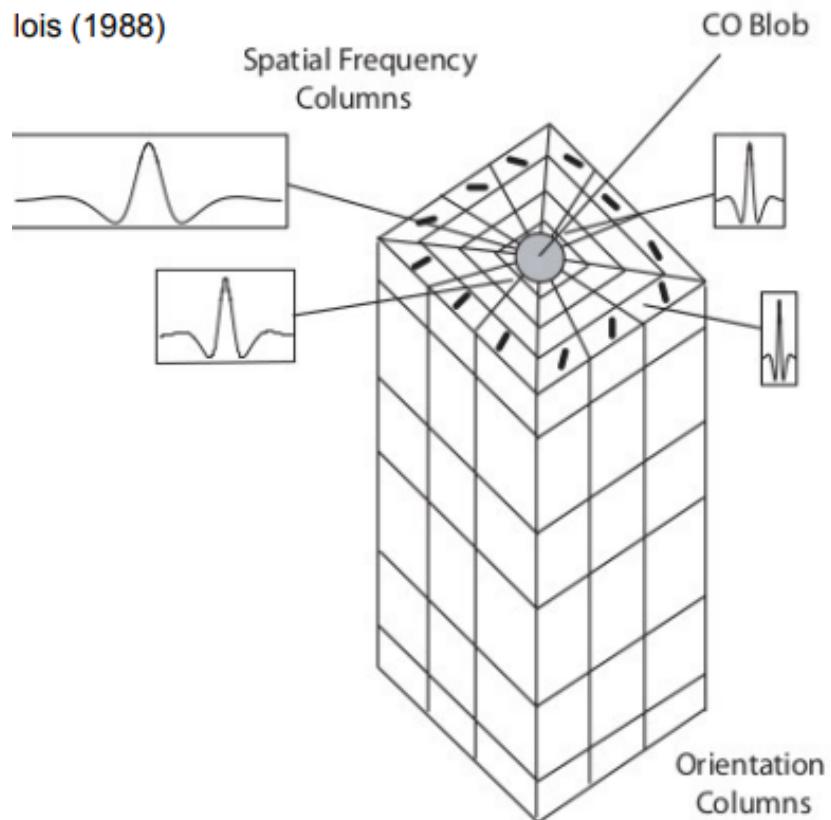
- Kantendetektion
- Textur Erkennung → Features der Textur extrahieren
- Wahrnehmungsexperimente



3.9. Orientation Hypercolumn

- Hyperkolumne: Jede Hyperkolumne verarbeitet Informationen von einem einzelnen Punkt des visuellen Feldes.
- Orientierungsspezifität: Zellen innerhalb einer Hyperkolumne, die auf eine bestimmte Orientierung empfindlich sind, haben fördernde Verbindungen zu Zellen in benachbarten Hyperkolumnen mit derselben Orientierungsspezifität.

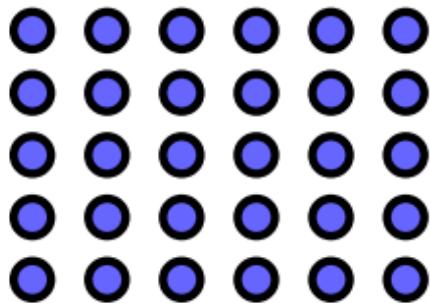
- Bedeutung: Diese Struktur ist wichtig für die Wahrnehmung von Linien, Konturen und Kurven sowie für die Gruppierung von Wahrnehmungen. Ein Teil der Hyperkolumnen könnte auch Farbverarbeitung übernehmen.



3.10. Gestalt Laws

Das Ganze ist verschieden von der Summe seiner Teil

Prinzip der Prägnanz: Mensch wählt bevorzugt als mentales Bild das einfachste mit den Informationen übereinstimmende aus



Man sieht ein Viereck mit Kanten, obwohl das Viereck aus Kreisen geformt ist, die keine Kanten besitzen

4. Farbe und Räumliche Tiefe

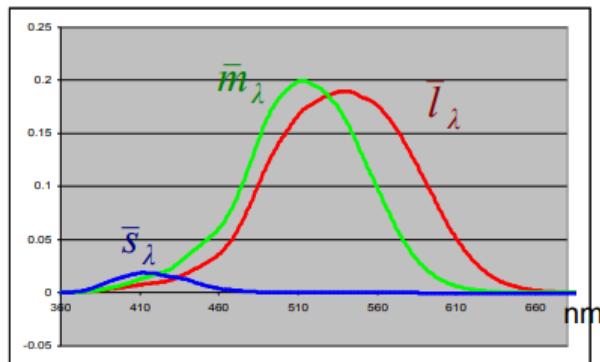
4.1. Farbe

4.1.1. Metamerie

Lichtreize mit unterschiedlicher spektralen Leistungsverteilung können nicht unterschieden werden, wenn gleiche Reize in Zapfen erzeugt werden

$$\rightarrow 540 \text{ nm} \stackrel{\wedge}{=} 7R + 33G + 39B$$

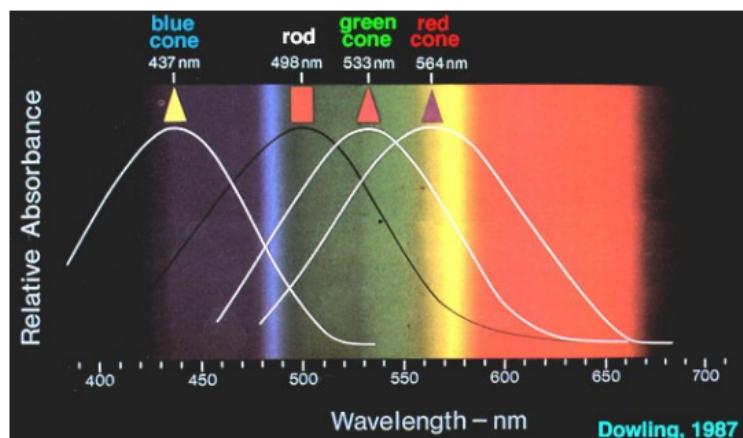
4.2. Farbe: Transduktion



Die Kurven beschreiben, wie stark die Zapfarten bei einer gewissen Wellenlänge feuern. Dadurch, dass die Kurven unterschiedlich verlaufen, kann man aus den relativen Bezügen zwischen den Intensitäten dann auf die Wellenlänge geschlossen werden.

Wenn man nur eine Zapfenart hat, kann keine Farbe sehen, weil keine relativen Bezüge möglich sind, und eine Intensität alleine wird als Hell/Dunkel interpretiert.

Die meisten Zapfen liegen in der Fovea, welche den Bereich des schärfsten Sehens repräsentiert.



relative spektrale Empfindlichkeit der Stäbchen und der drei verschiedenen Zäpfchenarten

4.3. Farbe: Percept

4.3.1. Farbmessung

4.3.1.1. dominante Wellenlänge

Wellenlänge λ eines Metamers, das aus weiß und einer Spektrallinie zusammengesetzt ist

4.3.1.2. Erregungsreinheit

$$E = \frac{e_2}{e_1}$$

4.3.2. Farbton (hue)

Perzept von der dominante Wellenlänge
Gerade noch erkennbare Differenz zweier Wellenlängen ca. $\Delta\lambda$ 4nm insgesamt ca. 128 verschiedene Farbtöne

4.3.3. Sättigung (saturation)

Intensität eines Farbton.
bei 575nm ca. 16 Schritte unterscheidbar, an den Ränder des Spektrums ca. 23 Schritte

4.3.4. Helligkeit (lightness)

bei grünem Licht ist die relative Lichtausbeute am größten

4.4. Farbe: Farbmessung

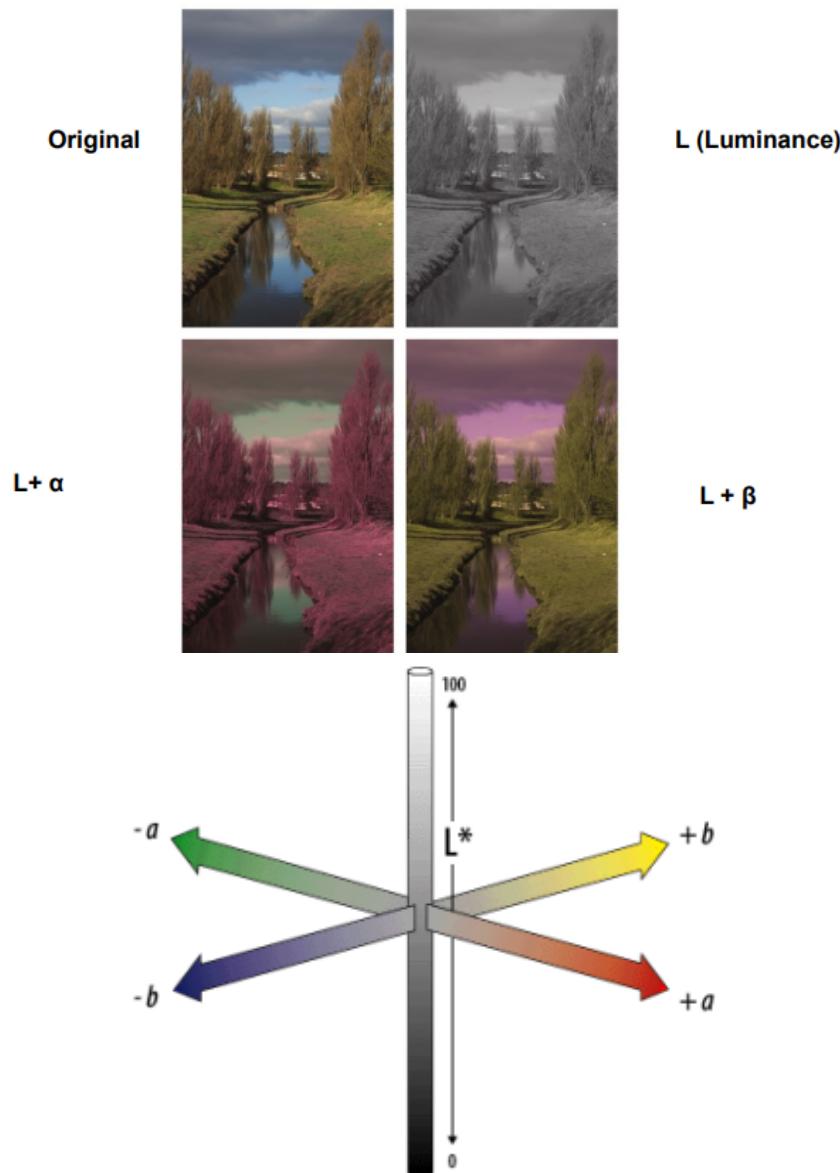
4.4.1. Darstellung am Monitor

- **Phosphor:** Zur Darstellung von Farben sind drei unterschiedliche Phosphore nötig, die jeweils unabhängige Leistungsspektren besitzen (zum Beispiel für Rot, Grün und Blau: $P_R(\lambda)$, $P_G(\lambda)$, $P_B(\lambda)$).

-
- **Überlagerung von Spektren:** Jede auf dem Monitor darzustellende Lichtleistungsverteilung $P(\lambda)$ wird als Überlagerung (Superposition) von spektralen Lichtquellen interpretiert, die durch die Leistungsspektren der Phosphore repräsentiert werden.
 - **Berechnung der RGB-Werte:** Um die Farben auf dem Monitor darzustellen, müssen die Intensitäten der spektralen Lichtquellen für die drei Phosphore bestimmt werden. Daraus lassen sich die RGB-Werte (Rot, Grün, Blau) berechnen.

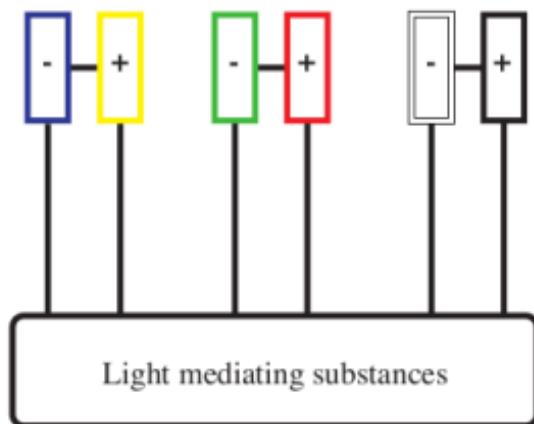
4.4.2. Farbe: Abhängigkeiten

- **Überlappung der Kanäle:** Die drei Kanäle von LMS (Long, Medium, Short)/RGB überlappen sich, was dazu führt, dass ihre Werte korrelieren. Änderungen in einem Kanal beeinflussen somit auch die anderen Kanäle.
- **Ziel:** Um die Kanäle unabhängig voneinander zu machen, sollte der Farbraum so transformiert werden, dass die drei Dimensionen voneinander unabhängig sind.
- **LAB:** Bei LAB (Luminance, Alpha (Rot Grün), Beta (Blau Gelb)) hat mein orthogonales System, sodass man in eine Richtung gehen kann, ohne gleichzeitig in eine andere mitzugehen



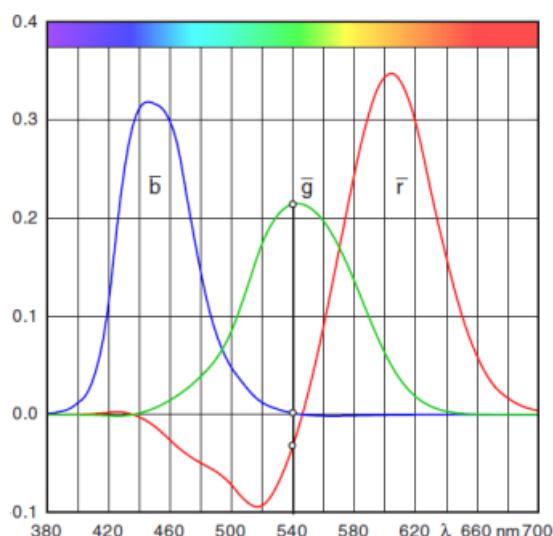
4.4.3. Opponent Processing

Die Color Opponent Repräsentation vermindert Korrelationen zwischen Werten, wodurch man eine effizientere Darstellung mit weniger Redundanzen bekommt → LAB



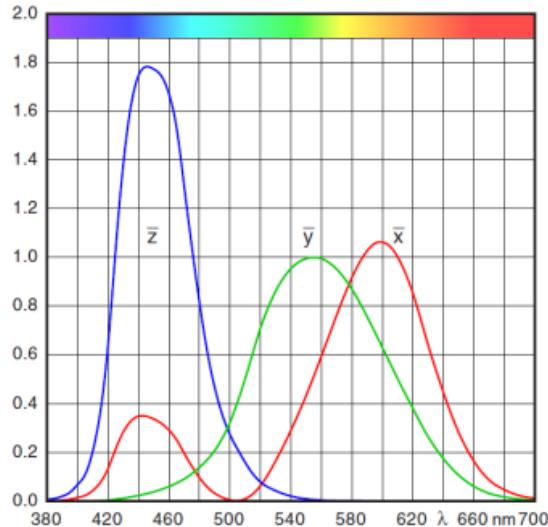
4.5. Farbe: Messung

Wissenschaftlich bestimmt, wie man alle Wellenlängen durch Roten, Grünen und Blauen Phosphor darstellen kann:



Problem: negative Farbvalenzen

- 1931 entwirft Commission Internationale d'Éclairage (**CIE**) ein chromatisches Farbsystem mit den Primärfarbvalenzen XYZ
- Alle drei Farbabgleichsfunktionen sind positiv
- Y entspricht photopischer spektraler Empfindlichkeit (Luminanz)
- X und Z sind ungefähr rot/grün und blau/gelb Unterschied; auf Basis von Versuchen standardisiert.



4.6. Farbe: Color Spaces

Farbraum: System das Farben mathematisch repräsentiert, um damit rechnen zu können. Benötigt wird es zur Darstellung (Wiedergabe) und Bearbeitung (Veränderung (künstlich erschaffen, verändern)).

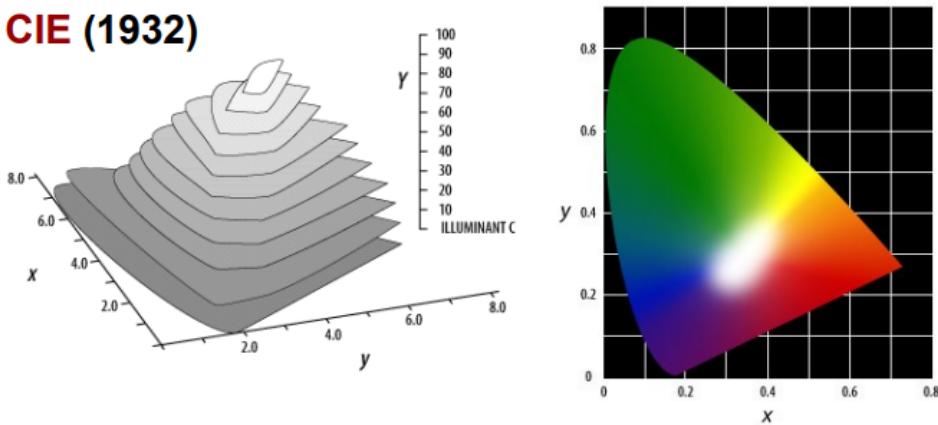
Medium: Bildschirm (additiv) → Pixel leuchten, addieren sich (Metameie), Drucker, Papier, Beamer (subtaktiv) → Blatt reflektiert
Farbräume können in einander umgerechnet werden!

4.6.1. CIE XYZ (1932)

Farben in die Ebene projizieren:

$$x = \frac{X}{X + Y + Z} \quad y = \frac{Y}{X + Y + Z}$$

→ Farbvariable: (x,y,Y)



Links: dreidimensionale Darstellung des xyY-Systems
Rechts: xy-Ebene maximaler Sättigung

4.6.1.1. Spektraler Lokus

U-förmiger Rand des CIE-Diagrams auf dem die Spektrallinien liegen

CIE (1932)

4.6.1.2. Purpurlokus

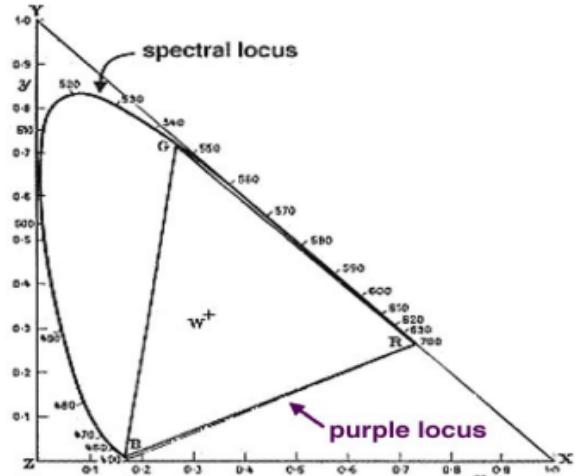
untere Grenze des Diagrams, das aus Mischung von blau und rot hervorgeht

4.6.1.3. Weißpunkt

Ort, wo die Farbe Weiß liegt

4.6.1.4. Gammut

Bereich der z.B. von einem Monitor dargestellt werden kann



4.6.1.5. Dominante Wellenlänge

Schnitt der Linie vom Weißpunkt durch eine Farbe mit dem spektralen Lokus

4.6.1.6. Erregungsreinheit

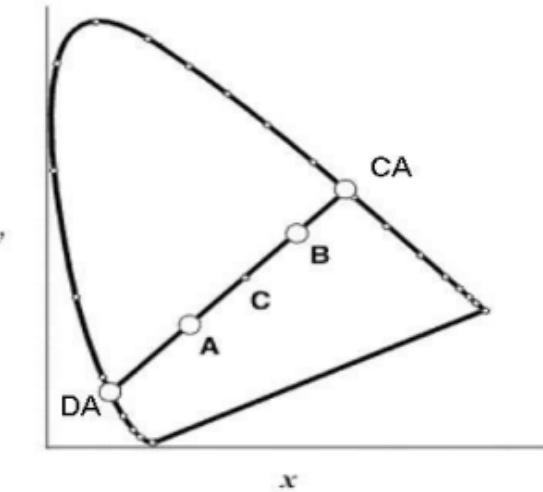
Länge Weißpunkt \leftrightarrow Farbe / Länge Weißpunkt \leftrightarrow dominanter Wellenlänge

4.6.1.7. Gegenfarbe

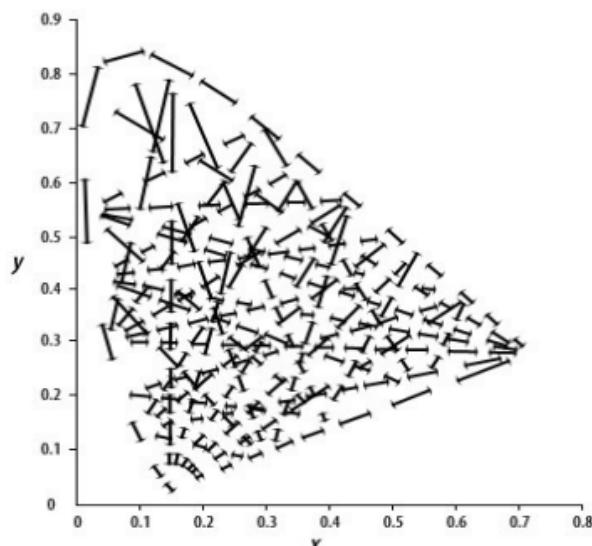
Farbe auf anderer Seite des Weißpunkts mit gleichem Abstand zum Weißpunkt

4.6.1.8. Farbdifferenzen

Wie gut sind Farbdifferenzen wahrnehmbar:



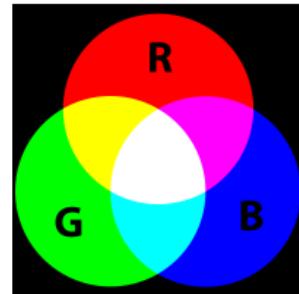
CIE (1964/76)



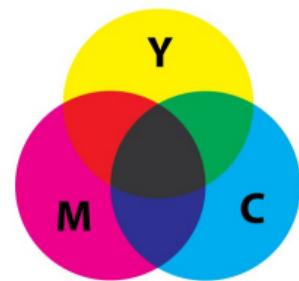
4.6.2. Farbmischung

Farbmischung

additiv: zwei Farbreize überlagern sich durch Addition der spektralen Leistungsspektren (z.B. Bildschirm)



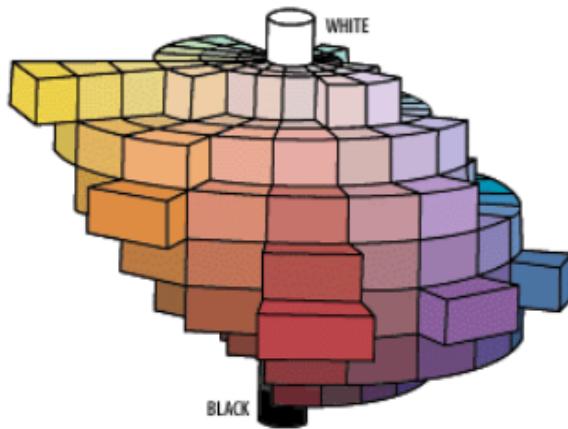
subtraktiv: aus dem weißen Spektrum werden mehrere Farben absorbiert (z.B. Drucker)



4.6.3. Munsell Farbsystem (1905)

Farbsystem für die Kunst, Farbglobus nach menschlicher Unterscheidungsmöglichkeit abgestuft:

- Farbton (hue) um Äquator: 100 Schritte, 10 Buchstabenkategorien zu je 10 Unterteilungen
- Helligkeit (value) von unten nach oben: 10 Unterteilungen: 1 = schwarz, 10 = weiß
- Sättigung (chroma) von Achse horizontal nach außen: von 0 bis zum Maximum, das von Helligkeit abhängt

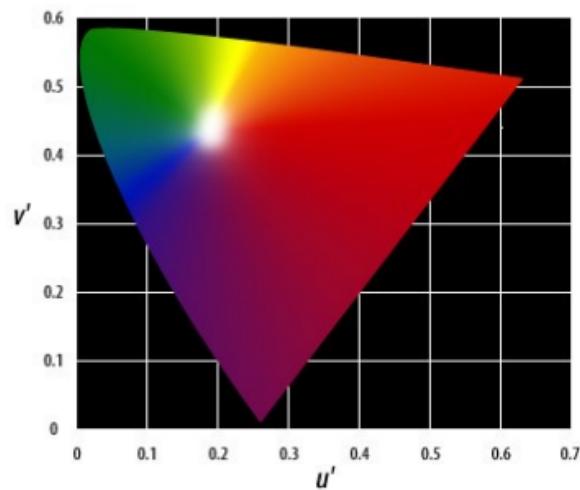


4.6.4. HLS

- Hue, 0..360 [rot..rot]
- Luminance, L : 0..100
- Saturation, S : 0..100

4.6.5. CIE u'v'Y (1976)

Bei Grün erkennt man wenige Farbdifferenzen, daher mehr nach oben verschoben (mehr Rot, X-Wert verändert)



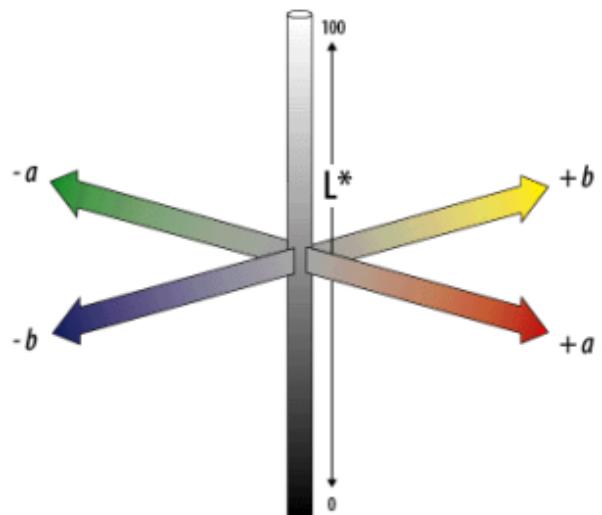
4.6.6. CIE L*u*v* (1976)

wie CIE u'v'Y (1976), nur Weiß bei (0,0)

4.6.7. CIE L*a*b* (1976)

Wie L*u*v*, nur dass a*b* den Farbton und Sättigung empfindungsbezogen auf **subtraktive** Mischung darstellen

Wenn der Farbabstand 1 beträgt, dann kann man den Farbunterschied noch gerade wahrnehmen



4.6.8. LMS

die Primärvalenzen werden entsprechend der spektralen Empfindlichkeiten der drei Zapfenarten gewählt (S/M/L ... short / medium / long wavelength)

4.7. Farbe: Fehlsichtigkeit

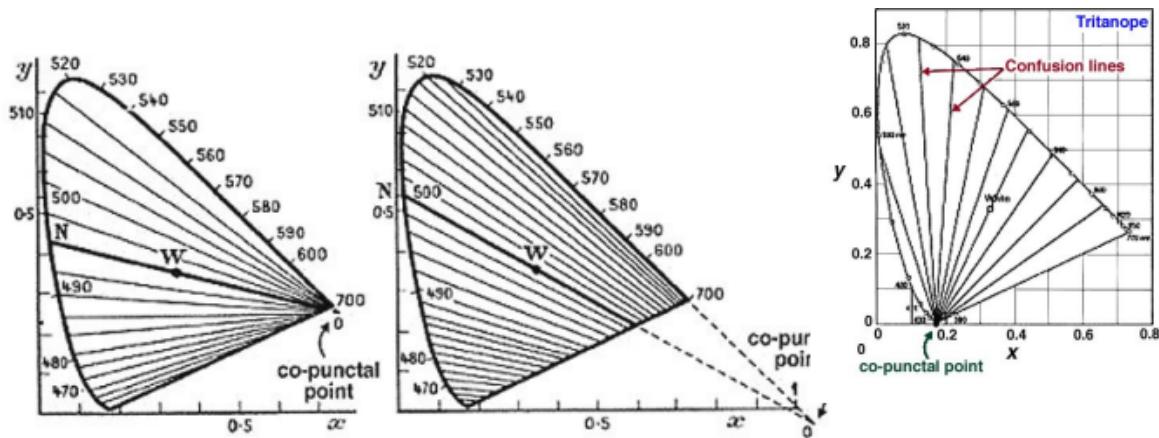
4.7.1. Arten

- anomale Trichromasie (Unterempfindlichkeit)
 - Protanomalie: Rot-Schwäche (1%)
 - Deuteranomalie: Grün-Schwäche (5-6%)
 - Tritanomalie: Blau(-Gelb)-Schwäche
- Dichromasien (partielle Farbblindheit)

- Protanopie: Rot-Blindheit (1%)
- Deutanopie: Grün-Blindheit (1%)
- Tritanopie: Blau-Gelb-Blindheit (0.004%)
- Monochromasie: keine Farben (0.003%)

4.7.2. Verwirrungslinien

bei Dichromasien können im CIE-Diagramm Farben entlang von Verwirrungslinien nicht unterschieden werden:



Rot-Blindheit | Grün-Blindheit | Blau-Gelb-Blindheit

4.8. Farbe: Color Constancy

- Farbkonstanz beschreibt das Phänomen, dass die wahrgenommene Farbe eines Objekts trotz Veränderungen in der Beleuchtung konstant bleibt.
- Bei einem Experiment, wo zwei Flächen mit demselben Licht auf der Netzhaut wahrgenommen werden, werden aufgrund des Kontexts diese als unterschiedliche Farben erscheinen. Dies illustriert, wie das visuelle System Beleuchtungseffekte herausfiltert.
- Der Weißabgleich in einer Szene ist komplex. Obwohl das Durchschnittsbild einer großen Anzahl von Bildern eine neutrale Farbe (Grau) ergeben würde, können einzelne Bilder durch dominante Farben verzerrt werden.

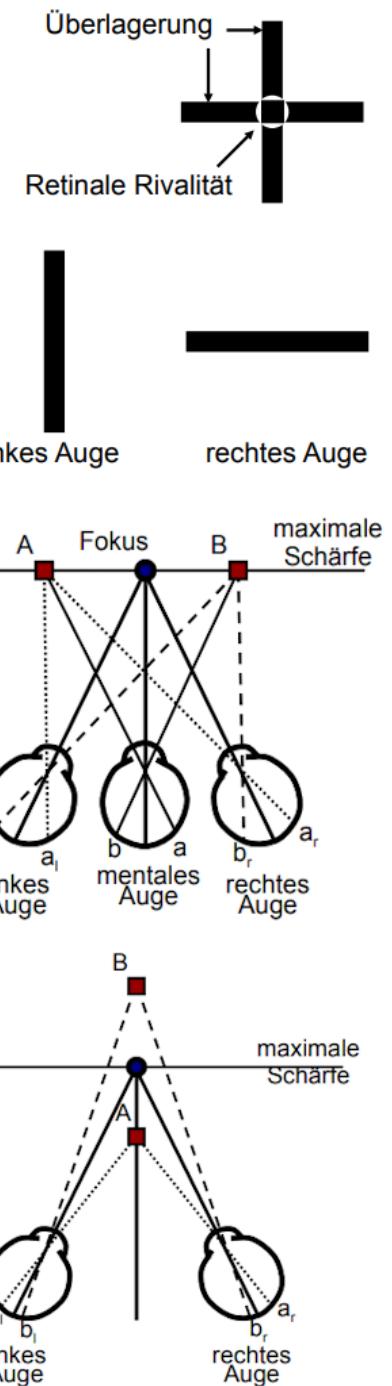
-
- Kameras versuchen, diesen Effekt durch „**White-Balancing**“ nachzuahmen, aber dies ist oft weniger genau als die Verarbeitung durch das menschliche Gehirn.
 - Die dargestellten Beispiele unterstreichen, wie das Gehirn Kontextinformationen nutzt, um die wahrgenommene Farbe stabil zu halten, obwohl sich die spektralen Eigenschaften des einfallenden Lichts ändern.

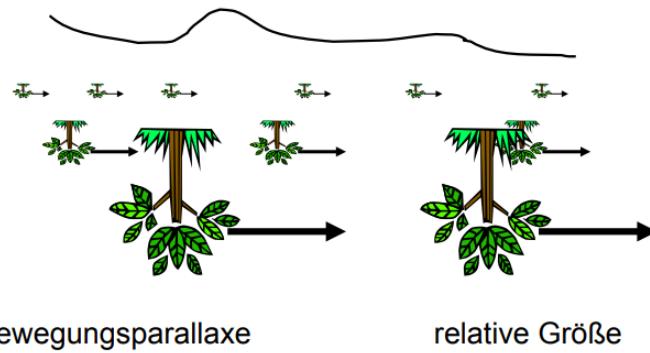
4.9. Räumliche Tiefe

Monokulare Faktoren (1 Auge) Binokulare Faktoren (2 Augen)

- Bewegungsfaktoren
 - Bewegungsparallaxe
- Bildliche Faktoren
 - relative Größe
 - Perspektive
 - Beleuchtung
 - Schatten
 - Verdeckung
- Fusion
- Disparität

- **Fusion:** Neuronaler Prozess zur Zusammenführung des linken und rechten Bildes innerhalb des Panumschen Fusionsbereichs.
- **Unterdrückung:** blendet unvereinbare Regionen aus.
- **Überlagerung:** vereint gleiche Regionen.
- Bei **Retinaler Rivalität** wechselt unvereinbares linkes und rechtes Bild.
- **Fokus:** Raumpunkt, auf den Augen gerichtet und akkomodiert sind.
- **Retinale Position:** Für jedes Auge die relative Position zur opt. Achse
- **Übereinstimmung:** Bildpunkte auf gleicher retinaler Position liegen auf Bildebene und sind Kandidaten für Fusion
- **Mentales Auge:** Durch Fusion kombinierte Infos
- **Überkreuzung:** Retinale Position eines Raumpunkts auf unterschiedlicher Seite der opt. Achse
- **Disparität:** Unterschied der retinalen Position von fusionierten Punkten im mentalen Auge. Aus Disparität kann relative Entfernung vom Fokus berechnet werden.



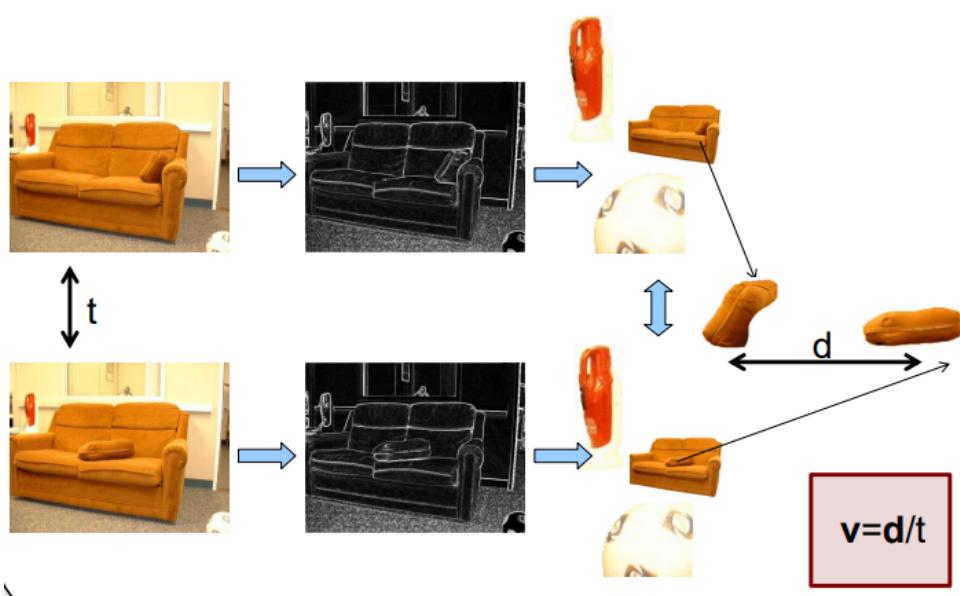


Bewegungsparallaxe: Objekte im Hintergrund bewegen sich verhältnismäßig langsamer

relative Größe: Objekte im Hintergrund sind kleiner bei eigentlich gleicher Größe

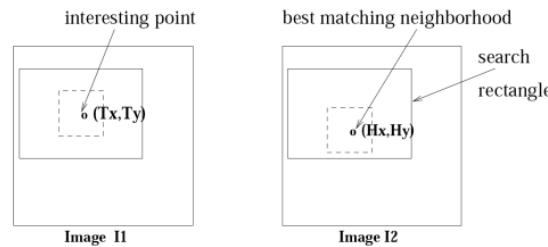
5. 2 Dimensionale Bewegungswahrnehmung u. Optischer Fluss

5.1. Static Bias: Motion Calculation

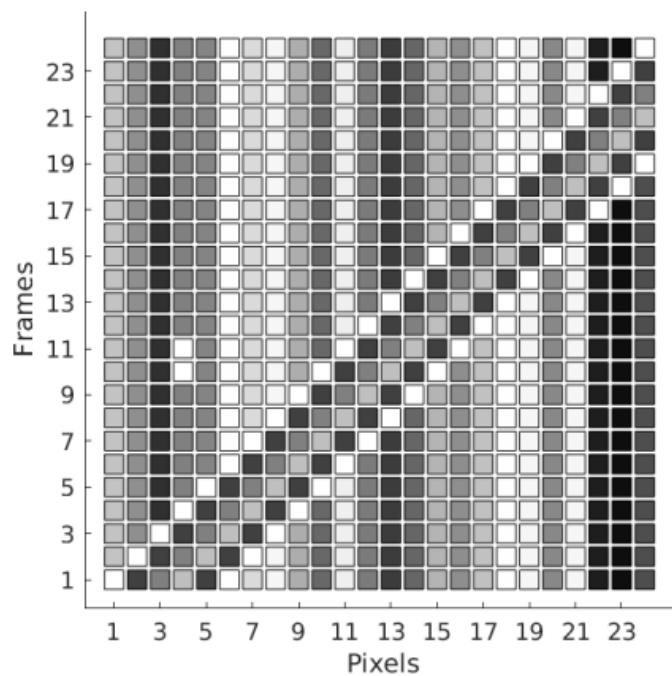


man sieht zwei statische Bilder, wo sich aber was ändert, und man hat die Voreingenommenheit, dass es sich bewegt haben muss. Die Geschwindigkeit lässt sich dabei aus der zurückgelegten Strecke durch die benötigte Zeit bestimmen.

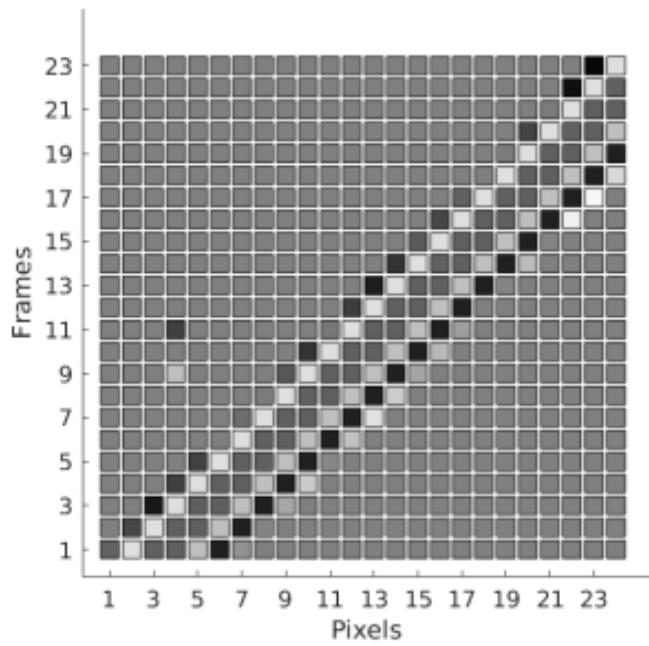
- 1) Detect points of interest in image 1 (e.g., corner points)
- 2) Search in image 2 for this point (e.g., using cross correlation).
- 3) Distance and time give motion vector



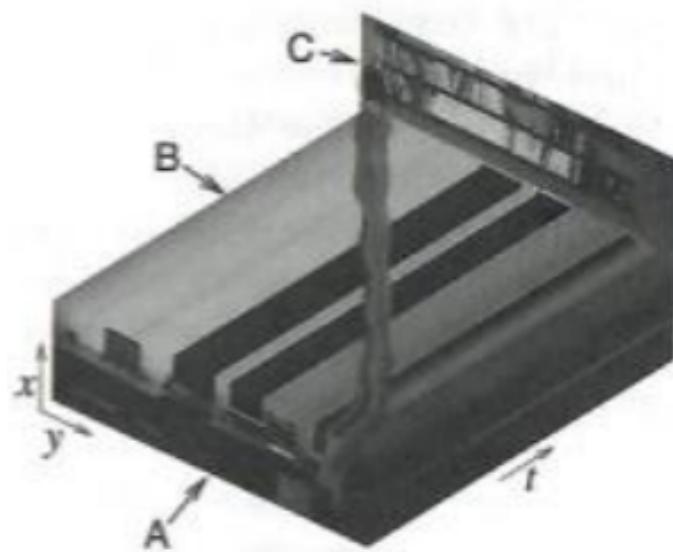
5.2. Change Detection/Motion Detection



→ man sieht von einer Pixel-Reihe, und über die Zeit hinweg (Frames) kann eine Veränderung erkennen.



wenn man von die 1. Ableitung bezüglich der Zeit betrachtet, kann die Veränderung besser erkennen, wobei der Rausch weiterhin lokal bleibt. Diese Methodik lässt sich auch auf das komplette Bild anwenden, wobei die Bewegung eine schlagartige Änderung ist, die nicht lokal in der Zeit und Raum ist. Dabei entspricht die Steigung in der zeitlichen Ebene der Geschwindigkeit, je schräger desto schneller.



5.3. Apparent Motion

Apparent Motion bezeichnet das Phänomen, bei dem zwei oder mehr stationäre Lichtquellen oder Objekte, die nacheinander an verschiedenen Positionen erscheinen, als kontinuierliche Bewegung wahrgenommen werden.

Beispiel: Apparent Motion

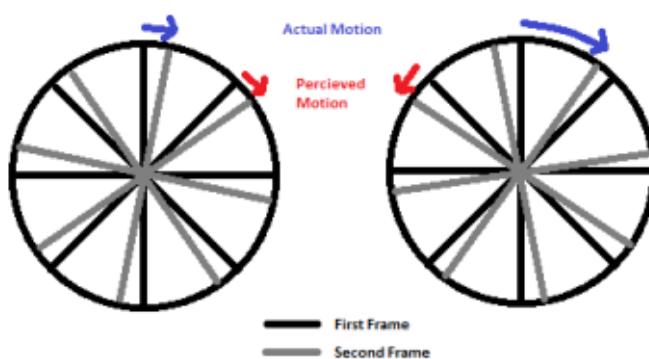
5.4. Phi Motion

Das Phi-Phänomen ist eine spezielle Form der scheinbaren Bewegung, bei der zwei Lichtquellen oder Objekte in schneller Folge aufleuchten, was eine Illusion von Bewegung erzeugt, ohne dass ein tatsächliches Objekt als bewegt wahrgenommen wird. Stattdessen wird die Bewegung als abstrakte Bewegung (ohne Objekt) zwischen den beiden Positionen gesehen.

Beispiel: Phi Motion

Phi-Motion hängt von der Geschwindigkeit ab: wenn es langsam abgespielt wird, scheint der blaue Kreis zu wandern (Apparent Motion), sofern es aber schnell abgespielt wird, sieht man einen scheinbaren, weißen Kreis, der die blauen Kreise verdeckt

5.5. Wagon Wheel Illusion



Beispiel: Wagon Wheel Illusion

Abtastrate und Bildfrequenz: Die Illusion entsteht, weil die Bildfrequenz (Frames per Second, FPS) einer Kamera oder die Abtastrate (80ms) des

menschlichen Auges dafür sorgt, dass die Bewegung des Rades in diskreten Schritten erfasst. Wenn das Rad sich zwischen zwei Frames um einen Winkel dreht, der nahe einem Vielfachen des Abstands zwischen den Speichen liegt, erscheint das Rad als stünde es still oder bewege sich langsamer. Rückwärtsbewegung: Wenn der Drehwinkel zwischen zwei Frames so ist, dass eine Speiche in einer Position erfasst wird, die minimal hinter der Position im vorherigen Frame liegt, erscheint das Rad als würde es rückwärts drehen.

5.6. Tunneling-Effect

Ein Objekt bewegt sich und wird durch einen Balken verdeckt. Das Objekt tritt auf der anderen Seite des Balken raus. Man hat die Vorstellung, dass es sich um das selbe Objekt handelt, weil jeweils T-Junctions auftraten und so das Objekt hinten durch geht.

5.7. First Order Motion

der Kontrast wandert nach rechts:

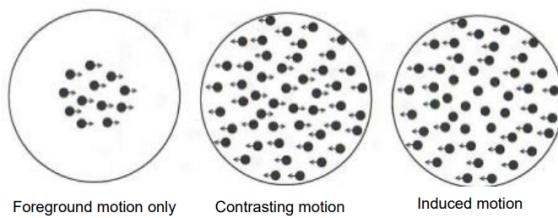


5.8. Second Order Motion

das Kontrast Muster ist statisch, aber Bänder von höheren und niedrigeren Kontrasten bewegt sich:

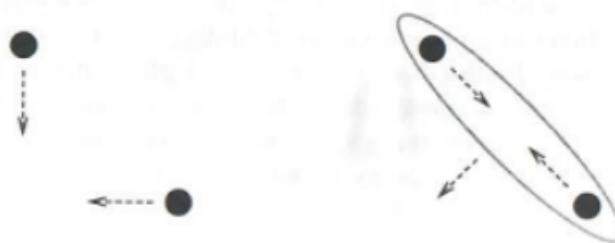


5.9. Motion Contrast



5.10. Hierarchical and Biological Motion

man denkt dass die Punkte in einer Ebene liegen, sich näher kommen, und die Ebene sich verschiebt:



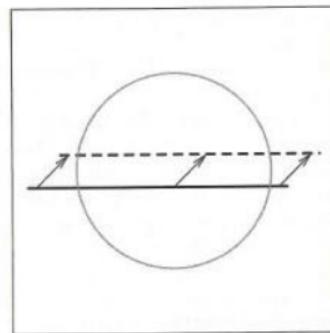
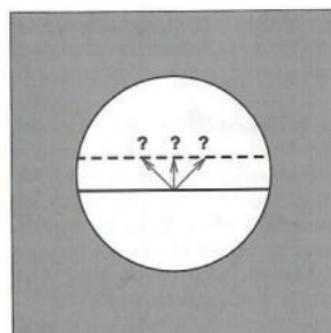
5.11. Motion Aftereffect

In einem Bild sucht man einen Nullpunkt, auf den geeicht wird. Alle anderen Geschwindigkeit/Bewegungen sind relativ zu diesem Nullpunkt.

Im Beispiel nimmt man als Nullpunkt den Wasserfall, wodurch der Felsen sich nach oben bewegt:

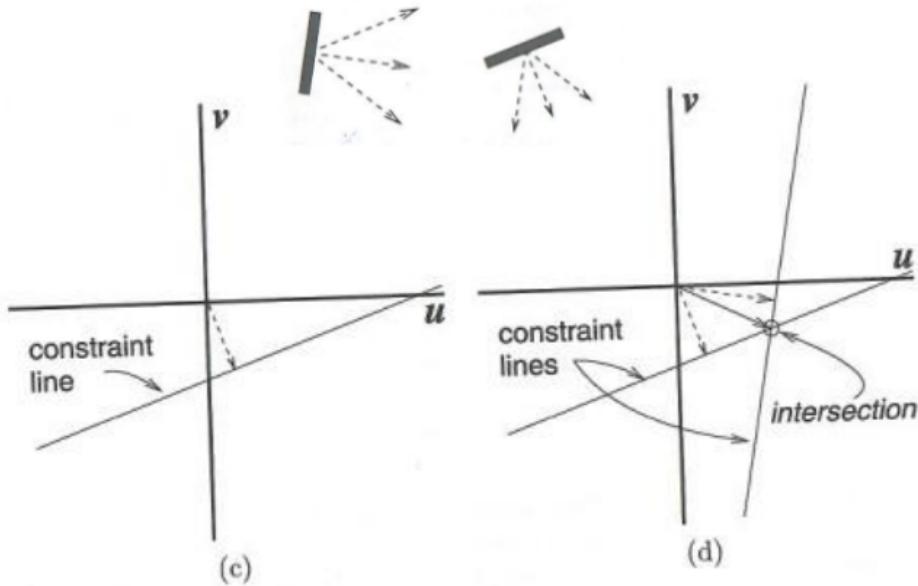


5.12. Aperture Problem



→ eine einzelne Zelle kann nicht genügend sehen, um die tatsächliche Richtung der orthogonale Bewegung zu erkennen.

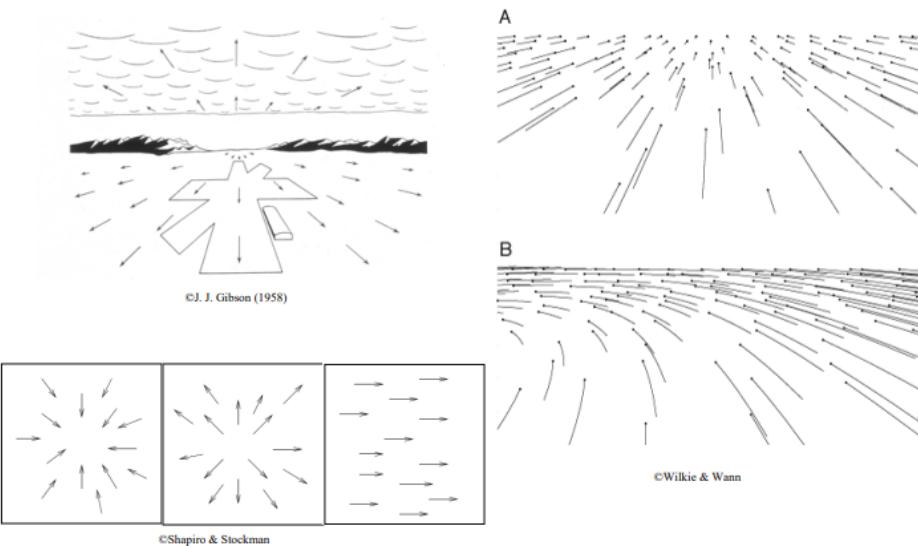
Da eine Zelle allein zu wenig Informationen hast, müssen mehrere Zellen zusammen arbeiten:



5.13. Optic Flow

Der Optische Fluss ist ein Vektorfeld, was Bewegungen von Bildpunkten beschreibt. Dabei ist ein Vektor ein Kante der Bewegung, welcher noch die Richtung der Bewegung speichert. Es konvergieren alle Vektoren zum Nullpunkt, was das Ziel der Bewegung darstellt. Bei einer Drehung gibt es aber eine Divergenz (Weg vom Punkt).

James J. Gibson (1958)



5.13.1. Anwendung

- Navigation
 - Richtung und Entfernung
 - Zeit bis zum Kontakt (für Landung, Spurwechsel, Kollisionsdetektion)
 - Trajektorienschätzung
 - Flugstabilisierung
- Visualisierung
- Bewegungsdetektoren
- Bildkomprimierung/-codierung
- Bildsegmentierung
- Bild-/Objektausrichtung
- Tiefenschätzung

5.13.2. Einschränkungen

- Die sich bewegenden Flächen sind flach
- Die Beleuchtung ist gleichmäßig über die Flächen verteilt (keine Schatten)
- Keine Verdeckung (keine Diskontinuitäten)
- Das Geschwindigkeitsfeld variiert überall gleichmäßig

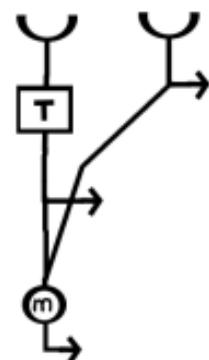
Dies bedeutet, dass die Helligkeit im Bild ausschließlich eine Funktion der Reflexion der Oberfläche ist und daher Helligkeitsänderungen eine Funktion der Bewegung sind. In Wirklichkeit wird die Helligkeit in einem Bild jedoch beeinflusst durch:

- Unterschiedliche Beleuchtungen
- Geometrie der Oberfläche
- Ausrichtung zur Kamera
- Materialeigenschaften (Textur)

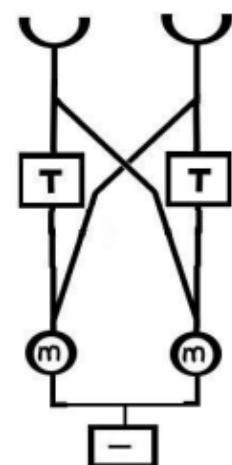
- Verdeckung
- Diskontinuitäten im Geschwindigkeitsfeld

5.14. Reichardt-Detector

Der Reichardt-Detektor kann eine Pixelreihe anschauen. Jeder Rezeptor kann nur Ein/Aus, deshalb reicht einer nicht aus, weil die Änderung könnte Rausch sein. Durch den Verzögerung (Delay/t) links hat die Bewegung Zeit, noch den rechten Rezeptor zu erreichen, bevor die Änderung links vergessen wird. Am Ende werden beide Werte multipliziert.



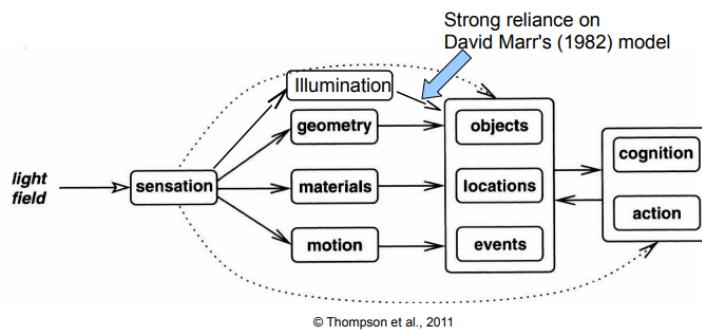
Dieser Detektor kann aber nur Bewegung von links nach rechts wahrnehmen, deshalb schaltet man zwei zusammen und subtrahiert am Ende, um die Richtung zu bekommen.



6. Textur Wahrnehmung

6.1. Textur Definition

- **Wahrnehmungspyschologie:** was vom Mensch visuell spontan als strukturiertes aber einheitliches Gebiet wahrgenommen wird.
- **Computer Graphik:** 2- oder 3-dimensionale Struktur eines Erscheinungsattributs zur Codierung feinen Details.
- **Visuelle Textur (Englert 93):** Gesichtsempfindung (Sinneseindruck des Auges), durch die sich angrenzende, möglicherweise strukturierte Gebiete des Gesichtsfeldes bei unbewegtem Auge unterscheiden
- **Perceptual Psychology:** Im Wesentlichen konzentriert sich die „Textur“ auf visuelle Attribute, die mit folgenden Eigenschaften zusammenhängen Oberflächeneigenschaften wie Material, Reflexion, Geometrie und Ausrichtung



So....die Muster des strukturierten Lichts, die sich auf Oberflächen Eigenschaften...

Welche Muster?

6.2. Structural

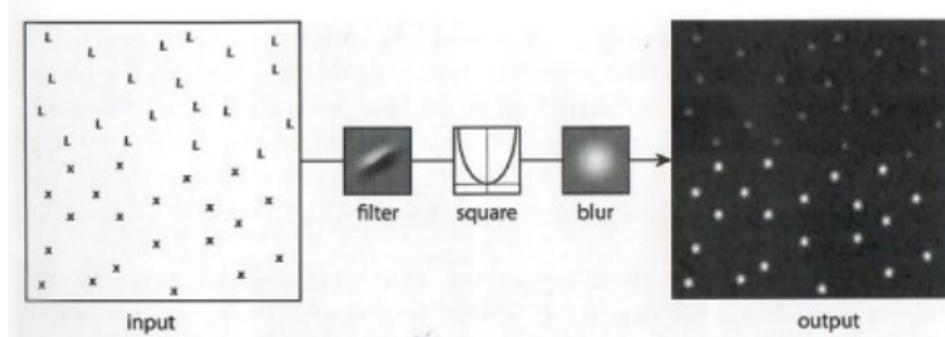
Kriterien zur Unterscheidung (Tamura 78):

- Grobheit (Coarseness)
 - Rauigkeit (Roughness)
 - Kontrast (Contrast)
 - Gerichtetheit (Directionality)
 - Linienartigkeit (Line-Likeness)
 - Regelmäßigkeit (Regularity)
- } (Rao 93)
} Komplexität
} Orientierung
} Wiederholungsgrad

Textur entspricht einer Sammlung von Gabor Patches (Orientierung → Gerichtetheit, Kontrast, Frequenz → Regelmäßigkeit), wobei jeder Pixel einen Gabor Patch hat.

6.2.1. Erkennung

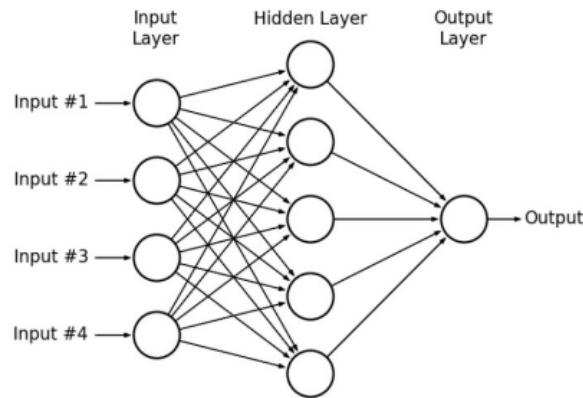
1. Bild falten mit Gabor Patch
2. Quadrieren → Betonung (klein wird kleiner, groß wird größer)
3. Schmieren mit Gauß-Filter



6.3. Statistical

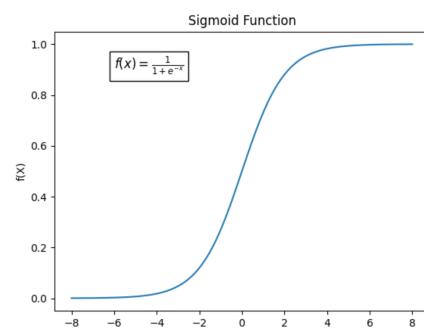
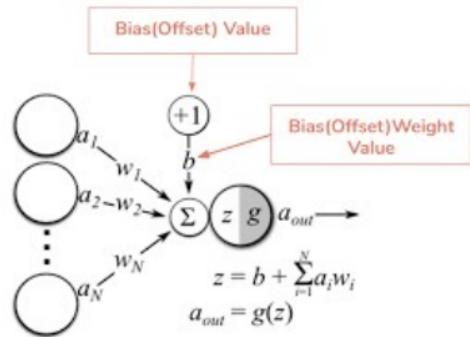
- Bei groben Texturen sind benachbarte Pixel ähnlich.
 - In feinen Texturen sind sie es nicht.
 - Wir betrachten die Unterschiede zu den benachbarten Pixeln, also die Beträge der Gradienten. Bei anliegenden Pixeln sind benachbarte Pixel identisch, der Gradient wäre 0.
 - Anschließend beleuchten wir die statistische Verteilung der Gradienten in einem Histogramm
- ⇒ Eine Verteilung mit kleinen Gradienten entspricht einer Textur

6.3.1. Neuronale Netze



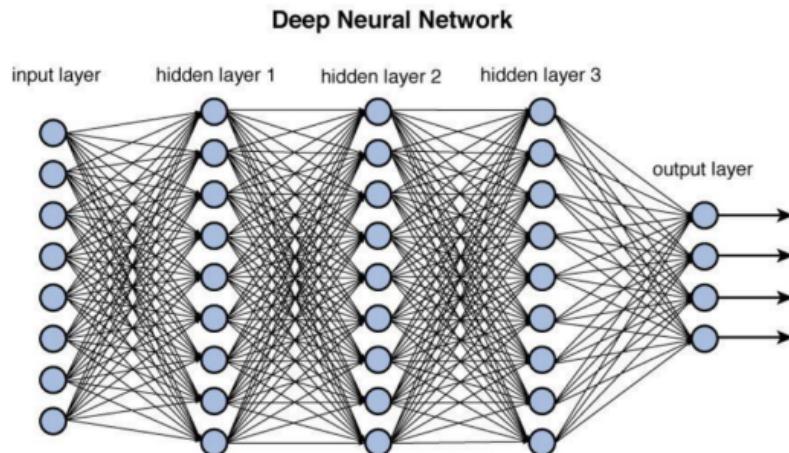
Das Bild stellt eine einfaches Multi-layer Perceptron (MLP), und dieses ist ein universeller Funktions-Approximator, dessen Genauigkeit exponentiell mit der Anzahl der Knotenpunkte steigt.

Ein Neuron ist mit allen Neuronen der hinteren Schicht verbunden. Deren Output (a) ist gewichtet (w) (Multiplikation $a \cdot w$), und dann werden alle Outputs aufsummiert. Der Bias Value sorgt dafür, dass immer eine gewisse Schwelle erreicht wird. Der Bias-Value ist ein zusätzlicher Parameter, der auf die gewichtete Summe addiert wird, bevor diese der Aktivierungsfunktion übergeben wird und sorgt für Flexibilität bei der Entscheidungsgrenze. Die Aktivierungsfunktion muss eine lokal begrenzte, stückweise kontinuierliche Funktion und NICHT ein Polynom sein, um Universalität zu gewährleisten. Die Aktivierungsfunktion gibt an, wie sicher sich das Neuron ist. Die natürlichste (menschenähnlich) Funktion ist die Sigmoid Funktion $g(z) = \frac{1}{1+e^{-z}}$

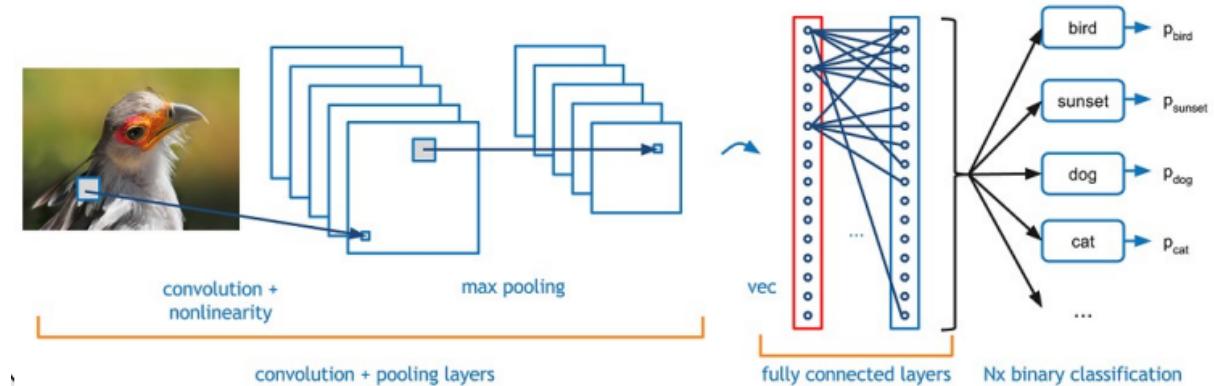


6.3.2. Deep Learning

Deep Learning hat 2 oder mehr verdeckte Schichten (und die Genauigkeit wächst linear).

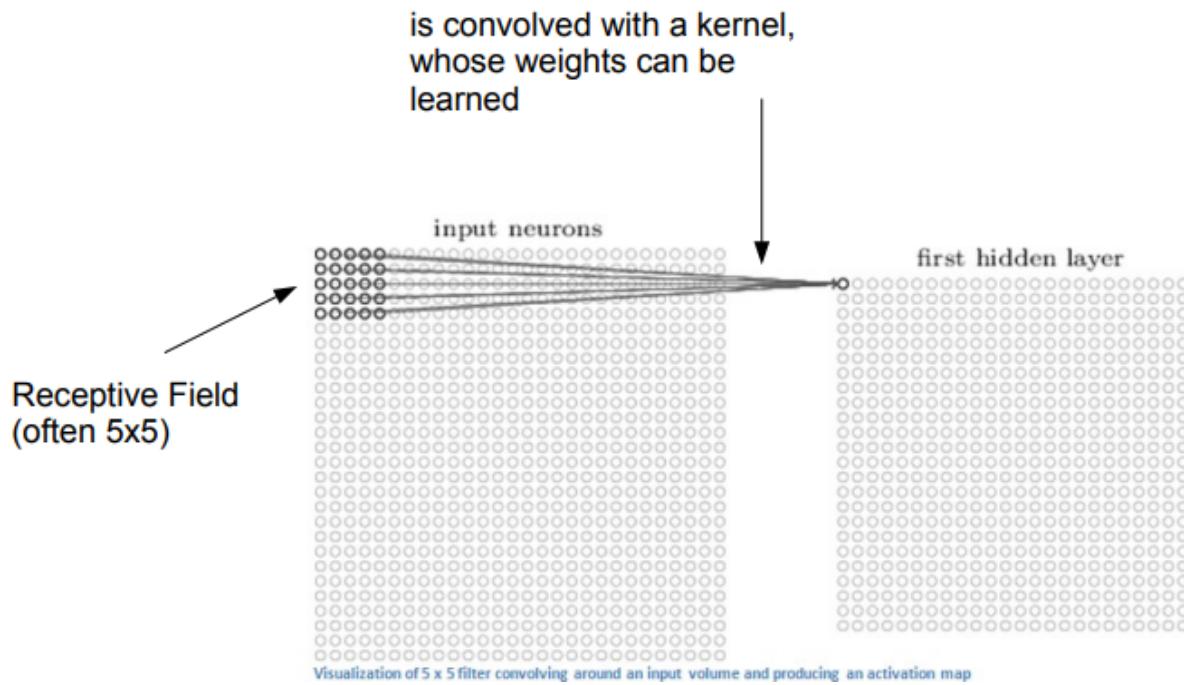


6.3.3. Convolution Neural Networks



Ein Convolution Neural Networks ist eine spezielle Form des Deep Neural Networks, der 2 neue Arten von Schichten innerhalb eines „Vorfilters“, gefolgt von einem „KlassifizierungsNN (in der Regel ein 3-schichtiger MLP)“ einfügt. Der Vorfilter verwendet abwechselnd Faltung und Pooling-Schichten. Ein CNN wendet primär auf Bilder (2D Strukturen) an. Dabei wählt man Pixelbereiche im Bild und filtert (convolution) mit z.B. Gabor Patches und nimmt die Max Werte (Max pooling). Am Ende wurde das Bild in einen großen Vektor dargestellt, den man als Input für das MLP verwenden kann.

6.3.3.1. Convolution

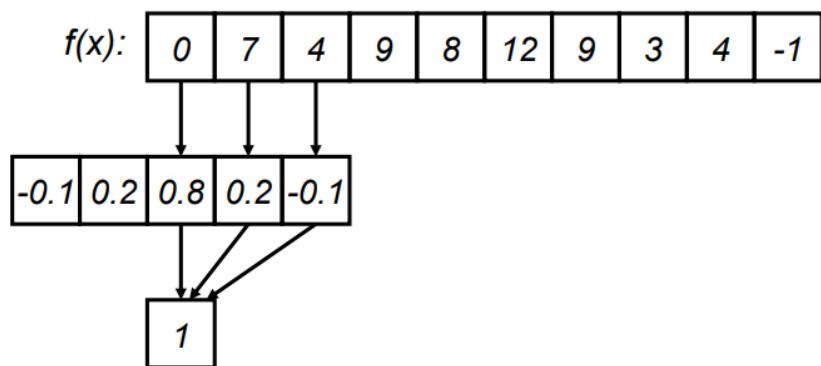


Man faltet mehrere Zellen mit einer Filterkern (Gewichte trainierbar) zu einer Zelle, wodurch Zellen am Rand wegfallen (Randbedingungsproblem), außer man setzt einen zusätzlichen Rand mit 0.

- Filterkern: Filterfunktion, die mit den Daten verknüpft wird.
- Man spricht von einer Faltung (Convolution) der Datenfunktion (Signal) mit der Filterfunktion
- Filterkerne haben im Allgemeinen einen eingeschränkten Wirkungsbereich (→ Randbedingungsproblem)

Filterung:

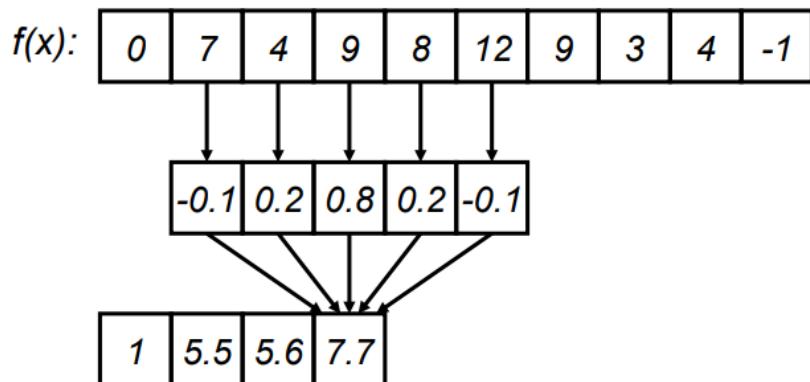
Falte mit Hilfe der Filtermaske



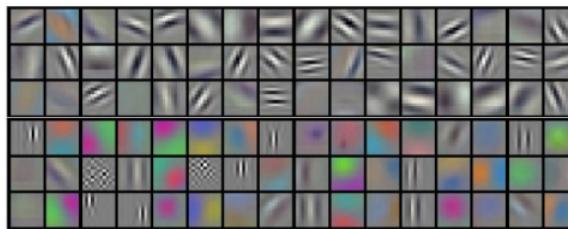
→ keine Zelle fällt weg, weil außen rum 0 (Schwarz) gesetzt wurde.

Filterung:

Falte mit Hilfe der Filtermaske

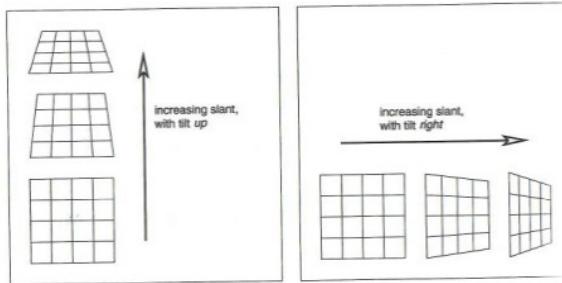


Was für Filter kommen raus? → Gabor Patches



6.4. Oberflächen Orientierung

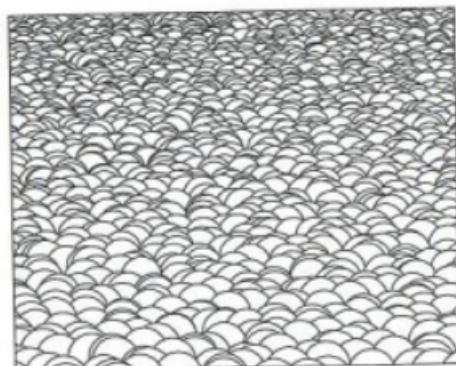
- Materie neigt dazu, sich zu bündeln, was zu sich wiederholende Muster führt.
- Aufgrund der Perspektive variiert die Größe mit der Entfernung
 - werden die Muster kleiner
 - die Linien des Musters konvergieren.
- Wenn eine Fläche aus der Sichtlinie geneigt ist der Sichtlinie geneigt ist, werden die Flecken kleiner und konvergieren.
 - **Foreshortening**



Die Linien konvergieren und wirken dadurch weiter weg. Da die Orientierung auch die Textur beeinflusst, kann man die Orientierung aus der Textur decodieren.



Die Abstände zwischen den Kugeln wird kleiner, die relative Dichte bleibt jedoch erhalten. Da die Kugeln aber kleiner werden, sind insgesamt mehr dort.



Kugeln, die weiter weg sind, sind dunkler.

7. Beleuchtung und Schatten

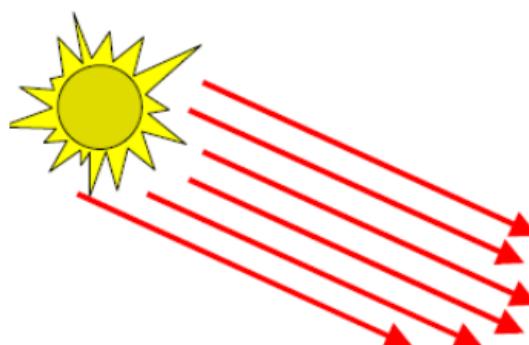
Die durch Oberflächen verursachte Struktur im Lichtfeld sagt uns etwas über diese Oberflächen...

...aber welchen Einfluss hat die Lichtquelle auf die Struktur des Lichtfelds?
→ das gleiche Objekt sieht unter verschiedenen Lichtverhältnissen anders aus:



7.1. Kumulierte Licht

Beim kumulierten Licht (Gerichtete Beleuchtung) gehen alle Lichtstrahlen in die selbe Richtung und sind parallel (Natur z.B. Sonne ünendlich weit weg)



7.2. Ambientes Licht

Beim ambienten Licht (Hintergrundbeleuchtung) kommt das Licht aus dem Hintergrund gleichmäßig aus allen Richtungen.

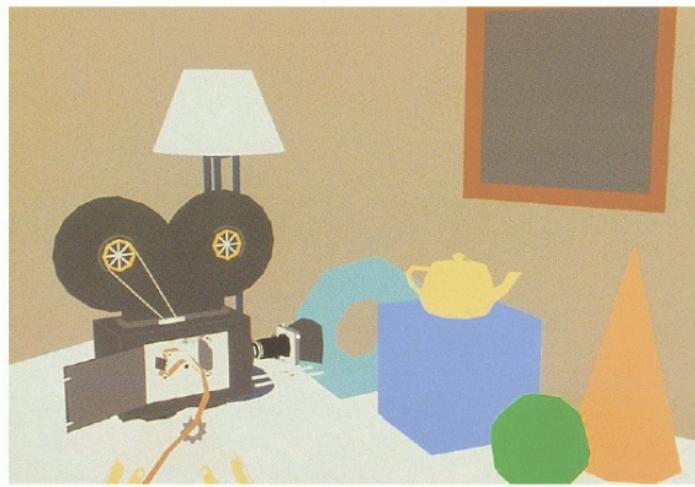
Da das Licht aus allen Richtungen kommt, ist die Helligkeit der Oberfläche unabhängig von so ziemlich allem, außer der „Oberflächen Reflexion“ (Albedo):

$$(I_{Ges}) = I = I_a \cdot k_a$$

k_a : Oberflächen Konstante des Ambienten Licht

Weil überall aus dem Hintergrund Licht kommt, gibt es keine Schatten

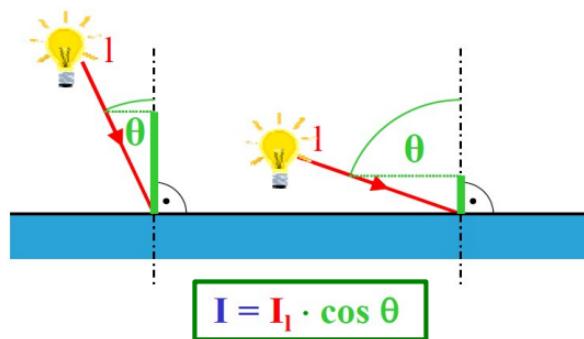
und konvexe Oberflächen haben trotzdem eine einheitliche Beleuchtung, was unnatürlich aussieht.



7.3. Zusammenhang Helligkeit Winkel

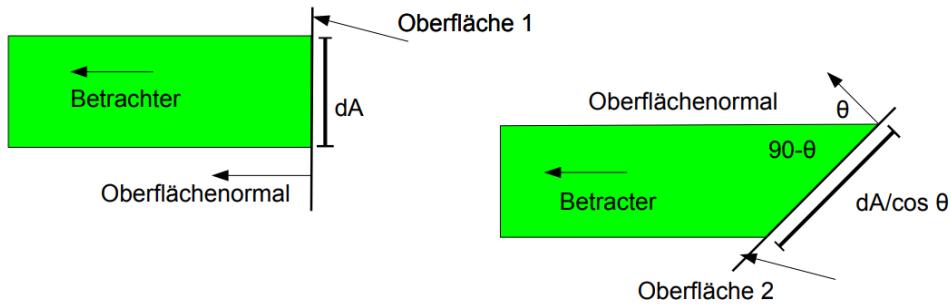
Helligkeit abhängig vom Winkel θ zwischen

- Vektor L zur Lichtquelle
- Oberflächennormalenvektor N



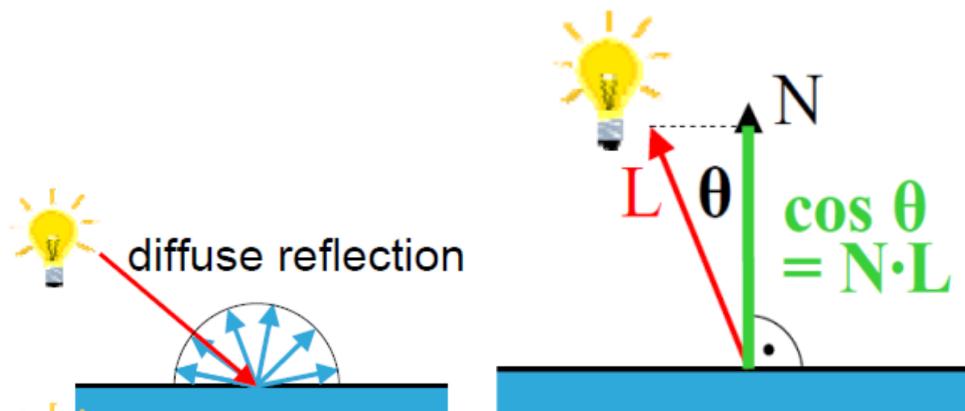
7.4. Lambertsches Gesetz für Reflektion

Menge an Licht, die von einem infinitesimal kleinen Flächenstück (dA) zum Betrachter ausgesandt wird, ist direkt proportional zum Kosinus des Winkels (θ) zwischen Betrachtungsrichtung und Oberflächennormal



7.5. Diffuse Beleuchtung

Bei einer lambertschen Fläche (keine Textur, keine Glitzer-Reflektion) wird Licht in alle Richtungen gleichmäßig reflektiert \Rightarrow **Diffuse Beleuchtung**



$$\begin{aligned} I_{l,diff} &= I_l \cdot k_d \cdot \cos(\theta) \\ &= I_l \cdot k_d \cdot (N \cdot L) \end{aligned}$$

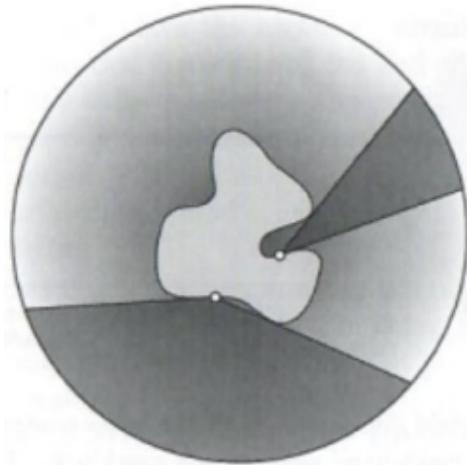
L: Vektor zur Lichtquelle, N: Normale der Oberfläche, k_d =Oberflächen Konstante für Diffuses Licht, $k_d=1$ \rightarrow perfekt in alle Richtungen reflektiert

$N \cdot L$ werden normiert (Länge 1) $\rightarrow N \cdot L = \cos(\theta)$

Orientierung bekommen: Richtung des Lichts (I_l) und Intensität des reflektierten Lichts (I) und nimmt den Arkuskosinus

7.6. Vigneting

Bei Vignetting/Self-Occlusion/Self-Shadowing verursacht ein Objekt bei sich selbst Schatten aufgrund der eigenen Geometrie. So wirkt eine Höhle dunkler, weil weniger indirekte Beleuchtung vorliegt, bedingt aber durch die Geometrie und nicht die Orientierung



7.7. Hemispherical Diffuse Illumination

Bei der modifizierten diffusen Beleuchtung wird mittels eines erweiterten Kosinus werden harte Schatten und komplett schwarze Bereiche vermieden



7.8. Realität

Die Realität ist wesentlicher komplexer, weil es

- Mehrere Lichtquellen.

-
- Spektrale Zusammensetzung
 - Wenige „lambertsche“ Oberflächen
- gibt

7.9. Shape From Shading

Shape From Shading beschreibt, dass man über den Schatten auf die Form eines Objektes schließen kann

7.9.1. Computer Vision

In der Computer Vision nimmt an:

- Lambertsche Oberflächen (keine Textur, kein Glanz).
- Keine Transparenz
- Keine Schatten
- Keine Reflektionen von anderen Objekten (nur direkte Beleuchtung)
- Richtung des Lichts ist bekannt

Die Intensität der Oberfläche an einem beliebigen Punkt steht in direktem Zusammenhang mit der Menge der verfügbaren Lichtmenge und der Ausrichtung der Oberfläche. Wenn man von einer konstanten Beleuchtung ausgeht, dann ist die Intensität in einem Bild nur eine Funktion der Ausrichtung!

7.9.2. Human Vision

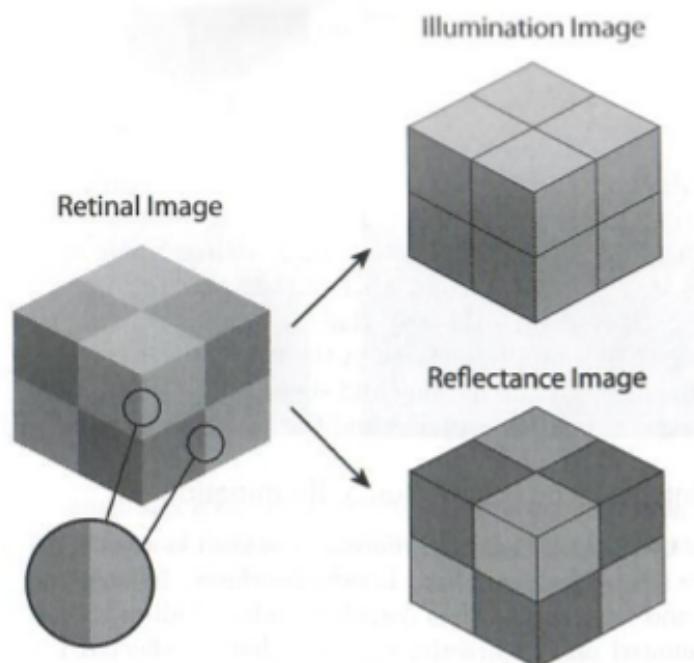
Beim menschlichen Sehen werden keine Einschränkungen, aber es wird trotzdem Shape From Shading umgesetzt. Keiner weiß wie, man geht davon aus dass wir Textur (Foreshortening) und Vignettierung (vigneting) nutzen, um das Shape From Shading Modell aufzubauen

7.10. Estimating Lighting

- Die Intensität einer Oberfläche hängt direkt von der verfügbaren Lichtmenge, der Orientierung der Oberfläche und der Oberflächenreflexion (oder Albedo) ab.
- Da wir die tatsächliche Lichtmenge nicht kennen, müssen wir anders vorgehen.
- Gilchrist schlug vor, das Bild in kleinere Bereiche (Framework) zu unterteilen und anzunehmen, dass die Beleuchtung in jedem Bereich konstant ist.
- Wir konzentrieren uns also nur auf die relative Lichtmenge, die reflektiert wird, und haben somit die relative Reflexion.
- Um zu einer absoluten Wahrnehmung von Helligkeit zu gelangen, wird einfach die höchste Intensität in einem Rahmen oder Bereich als "weiß" verankert (Anchor) und von dort aus skaliert.

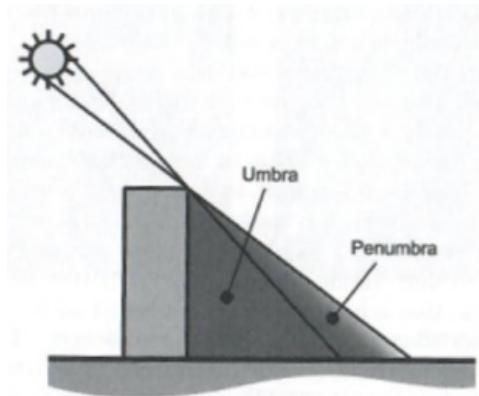
7.11. Intrinsic Image Analysis

Welche Wechsel der Intensität sind auf Belichtung oder Texture zurückzuführen? → Intrinsic Image Analysis



7.12. Schatten

Penumbra: Schattenkante



Je weiter der Schatten weg/sichtbar ist, desto weiter ist das eigentliche Objekt von der Fläche entfernt:



8. Material Eigenschaften

8.1. Surface Appearance

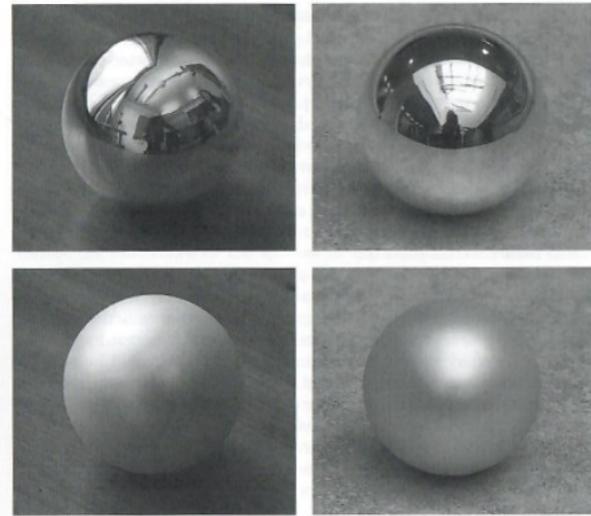
Textur kann auch ohne Objekte erkannt werden:



Wie wird die Oberfläche wahrgenommen? keine genaue Idee, aber Fragen:

- Welche physikalischen Größen geben der Oberfläche ihre Erscheinung?
- Welche visuellen Features sind für die physikalischen Größen verantwortlich?
- Welche Rolle spielt Kontext?
- Wie geht man vom Feature zur Eigenschaft, besonders wenn Features mehrdeutig sind?

Der Wert eines Pixels beruht sowohl auf Kontext als auch dem Material, weshalb man nicht einfach Pattern Matching über die Pixelwerte machen kann:



8.2. Source Separation

Wenn ein Feature ein Specular Highlight ist, dann ist das Material glänzend

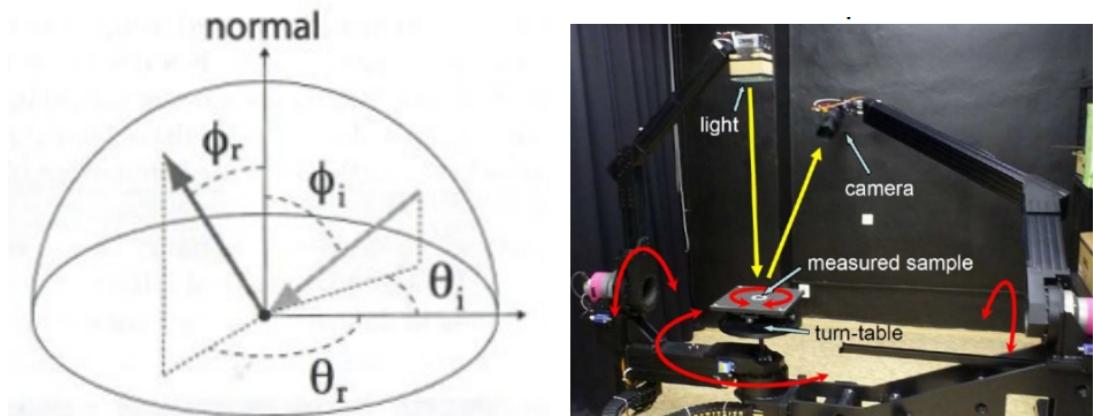
8.3. Approaches

- Inverse Optics
 - Visuelles System „läuft die Physik rückwärts“, um die gesamte Szene zu modellieren.
 - Sehr viele Ressourcen erforderlich.
 - Langsam.
 - Liefert mehr Informationen, als man braucht
 - Unterbestimmt, benötigt also Annahmen (so genannte „Priors“), wie z. B. dass Oberflächen lambert'sch sind, das Licht von oben kommt, usw.
- Image Statistics
 - Finden eines higher-order Feature im Bild
 - Prüfen, ob es in unterschiedlichen Szenen gleich bleibt
 - Prüfen, ob es mit dem Vorhandensein der gewünschten Oberflächeneigenschaft korreliert

- Erfordert Annahmen (über „typische Betrachtungsbedingungen“)

8.4. BRDF

Die **Bidirectional Reflectance Distribution Function** beschreibt ein Reflektion Szene. Für reflektiertes Licht spielen die drei Eigenschaften Eingangswinkel, Oberflächeneigenschaft und Betrachtungswinkel eine Rolle, weshalb für eine kleine Fläche alle Winkel erfasst werden (2 Eingangs- und 2 Ausgangswinkel vom Licht). Ein lokaler Lichtpunkt kann dabei reflektierten oder absorbieren.



8.5. Lightness

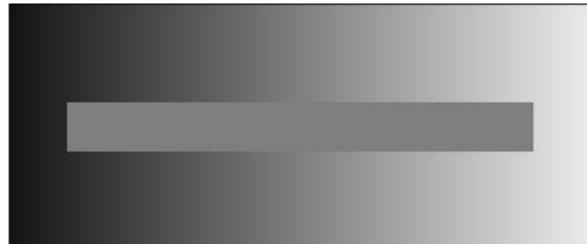
Lightness bezieht sich auf die subjektive Wahrnehmung, dass eine Oberfläche eine bestimmte Graustufe hat, unabhängig von der Beleuchtung. Das bedeutet, dass wir dieselben Oberflächen als gleich hell wahrnehmen, auch wenn sie im Schatten liegen oder von direktem Licht beleuchtet werden.

8.6. Brightness

Brightness ist die subjektive Wahrnehmung der Intensität des Lichts, das von einer Oberfläche reflektiert wird. Hierbei wird die absolute Lichtmenge berücksichtigt, die unsere Augen erreicht. Wenn eine Oberfläche beispielsweise stärker beleuchtet wird, erscheint sie heller, selbst wenn sich die tatsächlichen Materialeigenschaften nicht ändern.

8.7. Simulaneous Contrast

Man nimmt im mittleren Balken einen Kontrast wahr, wobei sich nur außenrum die Farbe ändert, ähnlich zu den Mach Bands

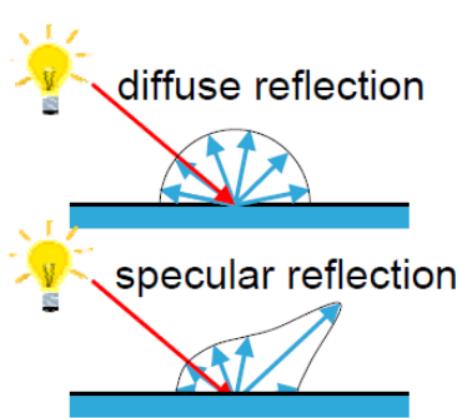


8.8. Gelb-Effekt

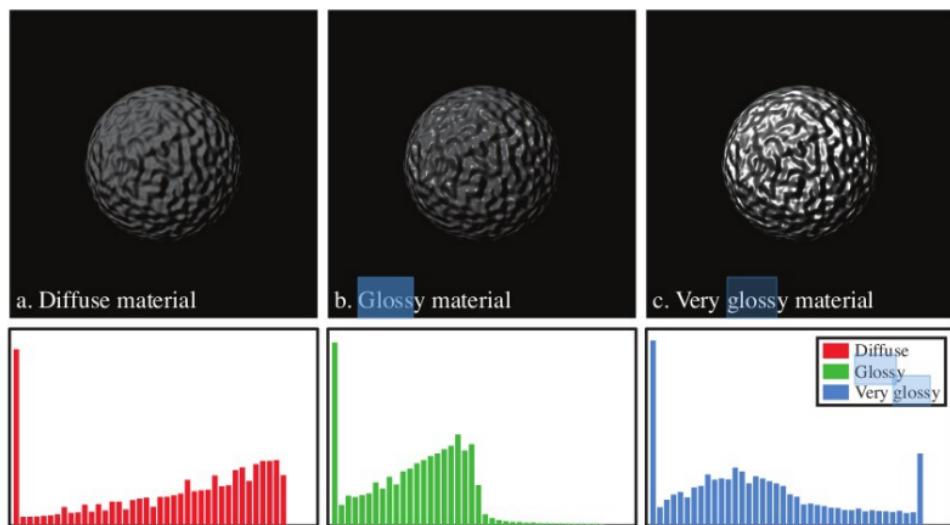
Wenn alles Schwarz (abgedunkelt) ist, dann ist der Punkt mit höchsten Reflektion weiß, und alles wir daran skaliert

8.9. Specular Reflection

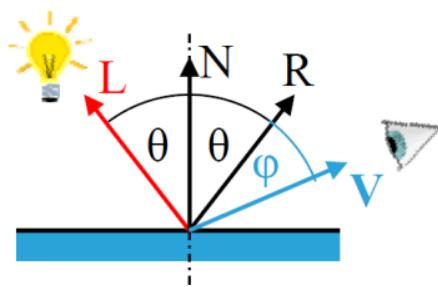
Wenn Licht auf eine nicht lambertsche Fläche trifft, dann wird es nicht in alle Richtungen gleichmäßig reflektiert (diffus), sondern in einen Winkel am stärksten. Man sieht ein **Specular Highlight**, welche die Farbe des einfallenden Lichts hat und vom Betrachtungswinkel abhängt.



Wie glänzend eine Fläche ist, kann man an der Verteilung der Intensitäten feststellen:

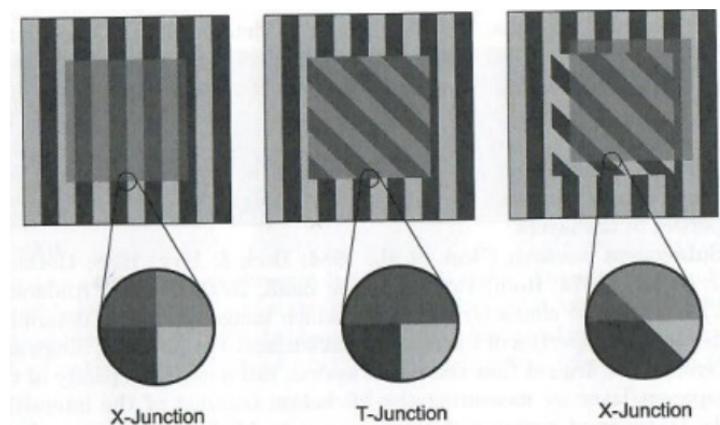


Je kleiner der Winkel ϕ rund um die Reflektionsrichtung R ist, desto glänzender die Fläche. Beim einem perfekten Spiegel (real unmöglich) wird nur in Reflektionsrichtung R reflektiert:



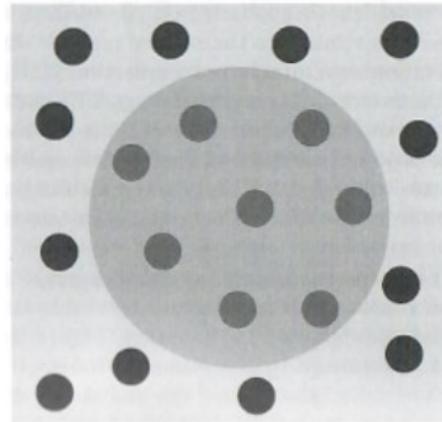
8.10. Transparency

Transparenz beeinflusst nur die Intensität, aber unten drunter geht die gleiche Textur weiter, weshalb X-Junctions entstehen:



Wenn eine T-Junction entsteht, ist die Fläche nicht transparent, sondern eine neue Fläche mit weniger Intensität.

Aber man kann Transparency auch ohne Junctions sehen:

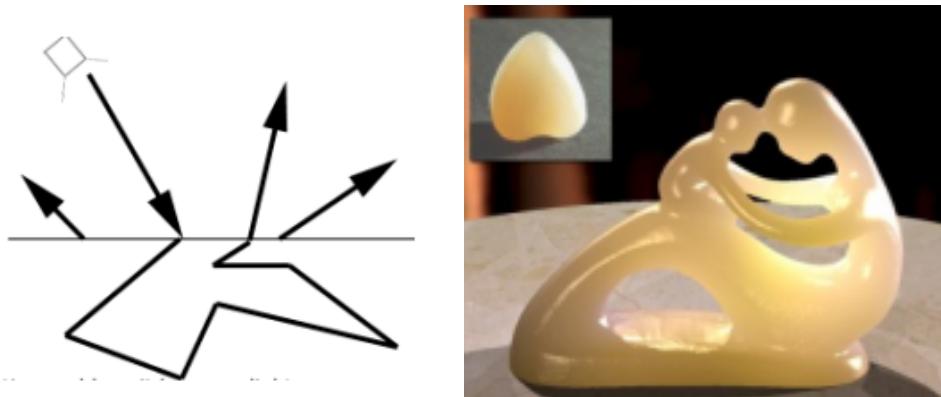


8.11. Subsurface Scattering

Bei Objekten tritt Subsurface Scattering auf, besonders bei transparenten (z.B. Haut, Gewebe). Dabei dringen Lichtstrahlen unter die Oberfläche und werden innen von tiefen Schichten abgelenkt, sodass diese abgeschwächt wieder aus der Oberfläche austreten. Dadurch entsteht der Eindruck, dass die Oberfläche leicht Licht emittiert.

Da dieser Effekt auftritt, erkennt man auch daran, wenn Licht durch das Objekt erkannt werden kann.

Da BRDF diesen Effekt vernachlässigt, weshalb es BSSRDF gibt.



9. Aufmerksamkeit

Aufmerksamkeit: Reize (bewusst) kategorisieren
Focal Attention: endogene/gesteuerte Aufmerksamkeit (eine Ausprägung)

9.1. Experimental Design

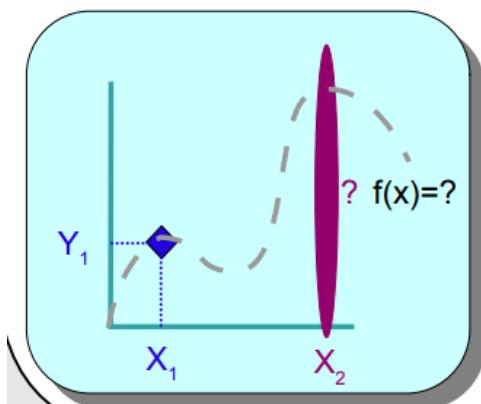
Angenommen man hat ein Orakel, dem man Input sowie jeweils den Output vorgibt, und dieses Orakel gibt die zugrundeliegende Funktion aus (Universal Function Approximater).

Wenn man die Funktion nur oft genug abtastet, kann man die Form der unbekannten Funktion einfach herausfinden, aber:

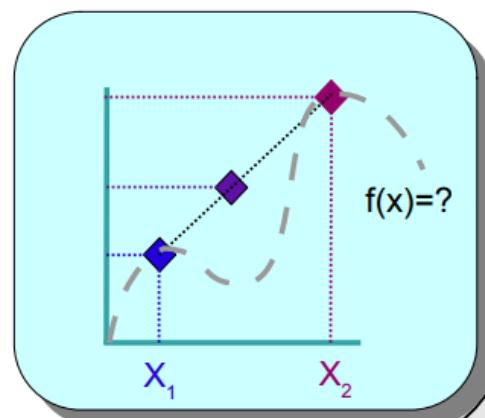
- Was wenn das Abtasten der Funktion jeweils kostet?
- Was wenn es unendlich Inputs gibt, aber nur manche einen Output liefern?
- Welche haben tatsächlich Einfluss?
- In welcher Domäne müssen die Werte liegen?
- Was wenn der Output etwas falsch ist (Rauschen)?
- Was wenn wir nicht wissen was den Rausch verursacht?

Warum ist es so schwer? → wir wissen nicht wie viele Parameter, und welche entscheidend sind. Es kostet was und der Output hat Rauschen.

Wenn man spezifisch misst, kann man nur über diesen Bereich Aussagen treffen.



Wenn man zu generell misst, dann hat zwischen den Messpunkten zu große Lücken, muss interpolieren, und die Abtastung ist damit auch fehlerbehaftet



In der psychologischen Forschung versucht man daher, möglichst viel abzutasten, und möglichst viele Parameter konstant zu halten, und die Veränderungen zu beobachten. Da man aber nicht alles konstant halten kann, bleibt das Ergebnis nur eine Annahme.

9.2. Reizinduzierte vs. Zielgesteuerte Auswahl

Bei der **zielgesteuerten Wahl** richten wir unsere Aufmerksamkeit **willentlich** auf einen Reiz. Dies ist der Fall, wenn wir etwas besser erkennen oder identifizieren möchten. Die Aufmerksamkeit wird **endogen** (von innen) durch den Willen **gesteuert**.

Bei der **reizinduzierten Wahl** wird unsere Aufmerksamkeit automatisch und **unwillkürlich** auf den Reiz gelenkt. Dies geschieht meist, wenn ein Reiz plötzlich erscheint oder eine besonders hohe Intensität hat (z.B. durch einen lauten Knall oder einen hellen Lichtblitz). Die Aufmerksamkeit wird **exogen** (von außen) durch den Reiz selber angezogen.

9.3. menschliche Informationsverarbeitung

Unsere menschliche Informationsverarbeitung ist **massiv parallel**, weil simultan/gleichzeitig Informationen von allen Sinnen (Sehen, Hören,...) aufgenommen werden.

Im Konflikt dazu steht der Fakt, dass wir im Alltag bei Multi-Tasking versagen, weil einige Tätigkeiten unweigerlich unter mangelnder Aufmerksamkeit leiden.

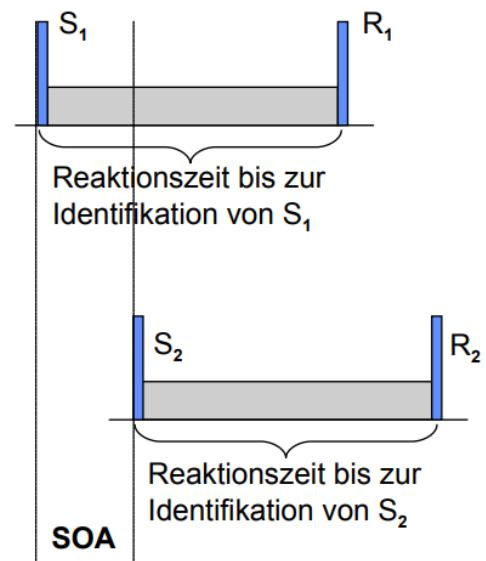
→ Konsequenz: Es gibt irgendwann in der Informationsverarbeitung einen **Übergang von massiv parallel zu seriell**, es existiert eine Art **Engpass/Flaschenhals** (Welford (1952)).

9.4. Psychologische Refraktärperiode nach Welford (1952)

Versuchsaufbau:

Es werden zwei akustische Reize präsentiert und deren Reaktion abgewartet.

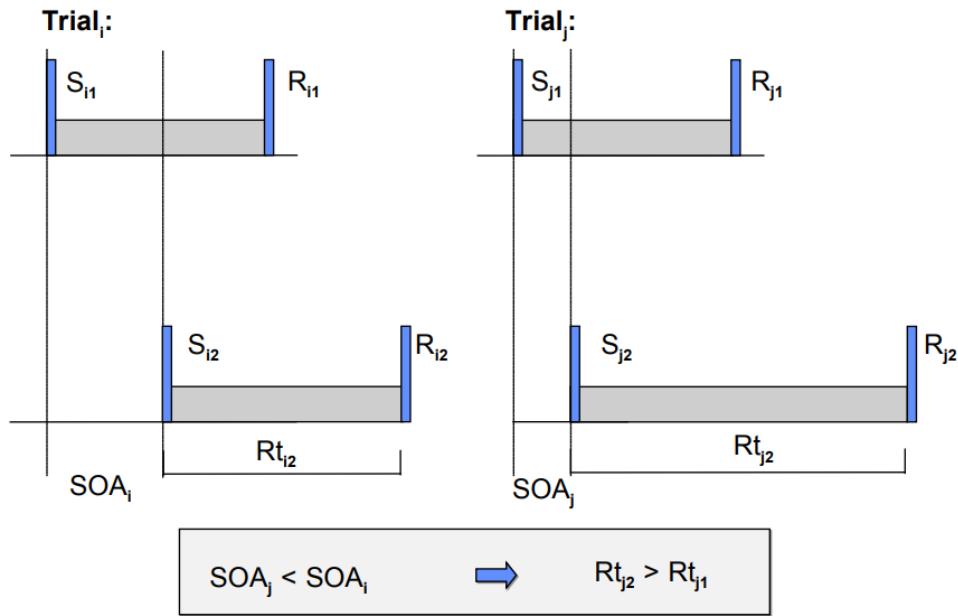
Die beiden Reize werden asynchron präsentiert (SOA / Stimulus Onset Asynchrony), und der zweite Reiz (S2) wird noch vor der ersten Reaktion (R1) präsentiert.



Beim eigentlich Experiment wurden nun die Abstände zwischen den Reizen (SOA) verändert und die Wirkung auf die Reaktionszeiten analysiert.

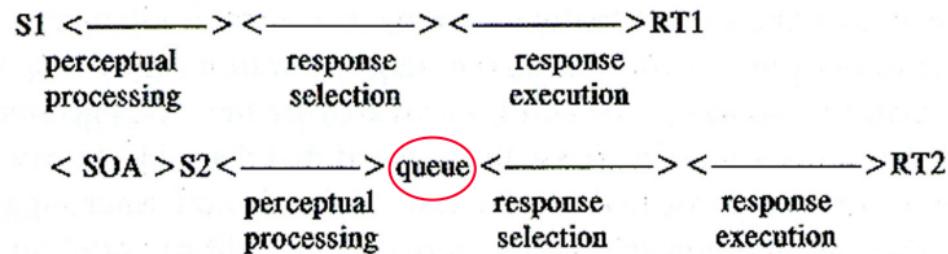
Ergebnis:

die erste Reaktionszeit konstant blieb, und die zweite Reaktionszeit wurde kürzer, wenn der Abstand (SOA) länger wurde und umgekehrt:



Erklärung:

Die Verzögerung der zweiten Reaktion wird in der Psychologie mit der **Refraktärzeit** (PRP/Psychological Refractory Period) erklärt. Diese beschreibt die Periode erzwungener Ruhe nach einem Aktionspotential, in der ein Neuron keine Erregung weiterleiten kann. Die Refraktärzeit (**PRP**) wird vom Abstand (**SOA**) beeinflusst und beide hängen **indirekt proportional** zusammen.



Die **Kapazität** der menschlichen Informationsverarbeitung ist **begrenzt**, so dass Reize nur nacheinander (**serial**) verarbeitet werden können.

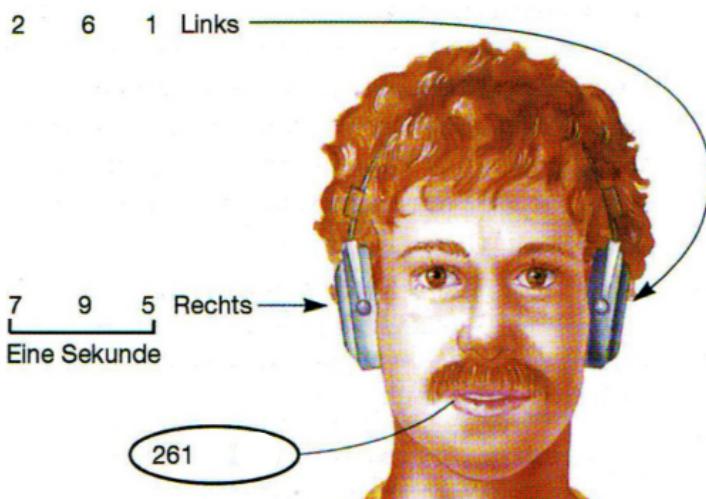
Dieser **Engpass** („bottleneck“) entsteht dadurch, dass die **Auswahl** der Reaktion auf den ersten Reiz erst **abgeschlossen** sein muss, ehe die reaktionsauswahl auf den zweiten Reiz erfolgen kann.

Die zweite Reaktion muss „in eine Warteschleife“ (”**queue**“), was zur Verlängerung der Reaktionszeit auf den zweiten Reiz führt.

→ Welford (1952) fand diese Queue, diese wurde auch nie hinterfragt, die späteren Theorien setzen sie an eine andere Position im Prozess!

9.5. Shadowing

Beim Shadowing (Beschatten/Verfolgen) bekommt ein Proband auf zwei Kanälen Reizen und muss einen verfolgen(beschatten) und dabei nachsprechen. Am Ende kann der Proband die Informationen vom beschatteten Kanal wiedergeben, die vom anderen Kanal gehen hingegen verloren.



9.6. Broadbents Filtertheorie (1958)

Nach Broadbent muss die Informationsverarbeitung wie ein Fluss auf einem **begrenzten Kanal** betrachtet werden. Dabei bedingt die begrenzte Kapazität eine **limitierte Informationsrate**, weshalb eingehende Informationen **gefiltert** werden. dadurch werden unnötige Daten abgeblockt und der Kanal vor Überlastung geschützt.

Die Filterung/Selektion beruht nach Broadbent auf **physikalischen Merkmalen** wie Lokation (linke o. rechte Seite) oder Stimmenlage (männlich o. weiblich).

9.6.1. Kritik

Nach Broadbent sind für die Filterung nur **physikalische Merkmale** und **keine semantischen Eigenschaften** wichtig. Demnach findet die **semantische Verarbeitung** erst statt, **nachdem** die Information die den Filter passiert hat. Da die selektive Aufmerksamkeit (Filter) vor der

semantischen Verarbeitung liegt, nennt man die Theorie **early selection model**.

Kritik: Experimente haben gezeigt, dass die Reizauswahl auch auf semantischen Merkmalen beruht! (**eigenen Namen** auf unbeschatteten Kanal gehört, **Moray (1959)**)

9.7. Shadowing-Experiment von Gray & Wedderburn (1960)

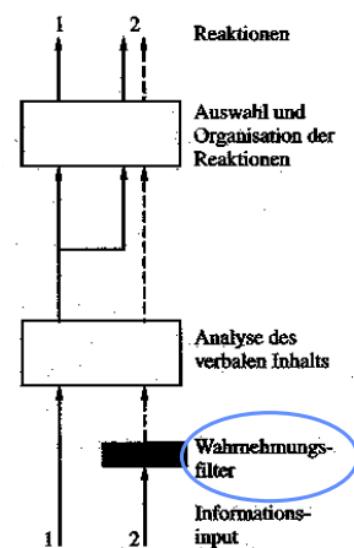
Beim Experiment wurde eine bedeutungsvolle Nachricht auf 2 Kanälen abwechselnd verteilt, währenddessen Störsignale (keine Bedeutung für die Nachricht) rein gestreut wurden.

Ergebnis: die Probanden waren in der Lage, der **Bedeutung zu folgen**, und die Aufmerksamkeit zwischen den Kanälen zu wechseln.



9.8. Dämpfungstheorie nach Treisman (1964)

- Wahrnehmungsfilter
- selektiert früh (**early selection model**)
- blockt irrelevante Reize nicht ab, sondern **dämpft**/schwächt Intensität
- Wahrnehmung → Filter → semantische Analyse → Auswahl der Reaktion



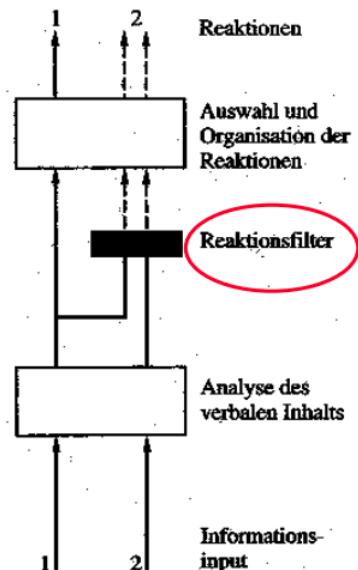
Moray (1959) → eigenen Namen (semantisch) auf dem nicht beachteten

Ohr erkannt → einige Informationen durchbrechen den Filter → Erklärung innerhalb Dämpfungstheorie:

- Reiz bewirkt **Aktivierung** der mentalen Repräsentation des Reiz
- die Aktivierung muss die **Aktivierungsschwelle** der mentalen Repräsentation überschreiten, damit der Reiz erkannt wird
- die **Höhe** der Schwelle hängt von der **Wichtigkeit** ab, wichtige Dinge haben eine vergleichsweise **niedrige Schwelle** (z.B. eigener Name)
- durch Filter abgeschwächter Reiz kann immer noch niedrige Schwelle überwinden

9.9. Late Selection Theorie nach Deutsch & Deutsch (1963)

- **Reaktionsfilter**
- selektiert spät (**late selection model**)
- auf beide Kanäle werden **alle Attribute** (physikalisch, semantisch) **vollständig** und parallel analysiert (→ viel **Kapazität**)
- Wahrnehmung → semantische Analyse → Filter → Auswahl der Reaktion
- jeder Reiz erhält Relevanzwert, und der Filter selektiert die relevanteste Reaktion, und unangemessene Reaktionen werden abgeblockt



9.10. Experiment von Treisman & Geffen (1967)

Ziel: Entscheidung zwischen Dämpfungstheorie (Treisman) und Late Selection Theorie (Deutsch & Deutsch)

Aufgabe: Shadowing + beiden Kanälen **Zielwort** erkennen

Vorhersagen bezüglich Erkennungsrate:

- Dämpfungstheorie: Intensität der Nachricht auf dem unbeschatteten Kanal wird reduziert, so dass das Zielwort auf diesem Kanal schwieriger erkannt wird → Erkennungsrate des beschatteten Kanals höher
- Late Selection Theorie: das Zielwort kommt auf beiden Kanälen gleich gut durch → Erkennungsrate auf beiden Kanälen gleich

Ergebnis: beschattetes Ohr 87%, unbeschattetes Ohr 8% → unterstützt Dämpfungstheorie

9.11. Aktuelle Erklärung

Aufgrund der Ergebnisse aktueller Experimente geht man davon aus:

- die Selektion ist **flexibel** und auf die Anforderung der Situation abgestimmt:
 - ein **frühe** Selektion erfolgt bei **schwierigeren** Aufgaben, welche **viel Verarbeitungskapazität** benötigt
 - eine **späte** Selektion erfolgt bei Aufgaben, die **wenig Ressourcen** binden. **Freie Kapazitäten** werden dabei für die Verarbeitung **anderer Reize** genutzt, was die endgültige Reaktion mitbeeinflussen kann
- Aufmerksamkeit fungiert dabei als **limitierte Ressource**

9.12. Visuelle Aufmerksamkeit

9.12.1. Visuelle Aufmerksamkeit im visuellen Feld

Im Fokus der visuellen Aufmerksamkeit liegt die Fovea centralis, das heißt die meiste Verarbeitungskapazität wird dem Bereich zugewiesen, den wir per Blick fixieren. Aber Reize außerhalb des Zentrum des Fokus werden

auch verarbeitet, nur mit weniger Kapazitäten. Dadurch können wir die Aufmerksamkeit auch außerhalb des Blickfokus richten.

9.12.2. Spatial-Cueing-Experiment von Posner (1980)

Fragestellung: Können Personen ihre Aufmerksamkeit auf Reize außerhalb der Fixation richten, und wie geschieht dies?

Versuchsablauf: Probanden gucken auf einen **Fixationspunkt** (keine Blickbewegung) und bekommen im Sichtfeld links oder rechts **Cues** (Hinweisreiz, Turbo) präsentiert. Anschließend wird der Testreiz präsentiert, der erkannt werden soll.

9.12.2.1. Cues

bei den Cues wird unterschieden in:

- **Zentraler Cue:** Pfeil am zentralen Fixationspunkt, der nach links oder rechts weist; erfordert Analyse durch zentrale kognitive Prozesse (Verarbeitung eines Symbols).
- **Peripherer Cue:** kurze Beleuchtung der linken oder rechten Hälfte des visuellen Feldes als Indikator, in welcher Hälfte der Stimulus zu erwarten ist.
- **Valider Cue:** Hinweis zeigt korrekt an, wo der Stimulus erscheint.
- **Nicht-valider Cue:** Hinweis ist falsch, und der Stimulus erscheint an einer anderen Stelle.
- **Neutraler Cue:** Zentrales Kreuz ohne Ortshinweis, das lediglich auf den Stimulus vorbereitet.

Ergebnis:

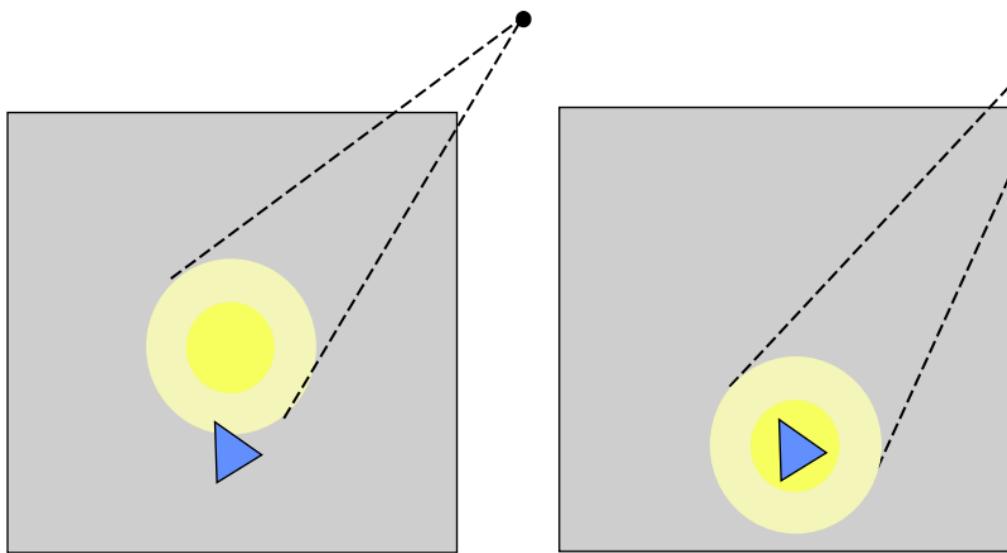
- Bei *validen Cues* war die Reaktionszeit kürzer als bei neutral, weil die Aufmerksamkeit bereits in die richtige Hälfte gelenkt wurde, wohingegen bei *nicht-validen Cues* die Reaktionszeit länger, da sich erst umorientiert werden musste.
- Bei den *zentralen Cues* gab es **keine Unterschiede** bei den Reaktionszeiten zwischen valider und nicht-valider Cue, bei Peripheren Cues hingegen schon → zentrale Cues können ignoriert werden, weil ihre Verarbeitung zu lange dauert und keinen Vorteil/Nachteil für die Reaktionszeit bringt.

-
- ⇒ der größte Teil der Aufmerksamkeit liegt im visuellen Fokus, dort wo wir gezielt hinsehen (willentlich, endogen), aber ein Teil reagiert auch unwillkürlich auf Reize außerhalb des Fokus (exogen)

9.12.3. Metapher des "Attentional Spotlight"

Visuelle Aufmerksamkeit ist wie der Lichtkegel eines Scheinwerfers:

Lichtkegel	↔	Intensität
Helligkeitsunterschiede	↔	Selektion (Zentrum / Peripherie)
Schwenk des Scheinwerfers	↔	Wechsel des Aufmerksamkeitsfokus



Die Reaktion ist am schnellsten, wenn der Reiz im fovealen Bereich liegt, andernfalls sorgt Umorientierung der Aufmerksamkeit im Mittel für eine Verlängerung der Reaktionszeit von 50ms (LaBerge, 1983)

9.12.4. "Zooming In"

Die Größe des Spotlights hängt mit dem Sehwinkel zusammen und kann bis zu 24 Grad von der Fovea abweichen (Posner).

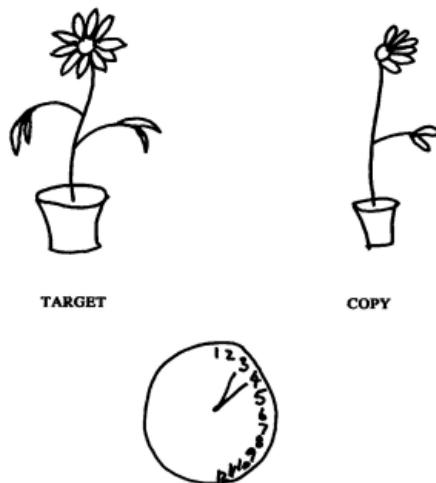
Wenn es keine Hinweise gibt, wird die Aufmerksamkeit auf das visuelle Feld verteilt, wobei die Verarbeitung **obflächlicher** wird, je **ausgedehnter** der

Fokus.

Wenn es einen **Cue** gibt, dann „**zoomt**“ die Aufmerksamkeit auf diesen Punkt, wobei ein **Rest** von Aufmerksamkeit die Peripherie des visuellen Felds überwacht.

9.12.5. Visual Neglect

Visual Neglect ist vorwiegend eine Vernachlässigung der linken Seite des Gesichtsfeldes bei einer Schädigung in der rechten Gehirnhälfte.



Ergebnisse bei Zeichenaufgaben

9.12.6. Raum- und Objektgebundenheit

Aufmerksamkeit kann sich nicht nur auf eine **Region** im visuellen Feld richten, sondern auch auf **Objekt**, die sich bewegen oder ruhen.

9.12.7. Verarbeitung visueller Reize

Visuelle Reize können als unterschiedlich **komplexe Objekte** im visuellen Feld aufgefasst werden. Objekte zeichnen sich dabei durch **Merkmale** aus, die über **Dimensionen** beschrieben werden können:

- Form
- Orientierung
- Farbe
- Position

9.13. Feature Integration Theory

Wie werden die **Merkmale** des Objekts **erkannt**, und wie werden sie im Verarbeitungsprozess **kombiniert**, sodass der **Eindruck eines Objekts** entsteht?

→ **Feature Integration Theory (FIT)**

9.13.1. Annahme

Die Merkmale werden früh im Wahrnehmungsprozesse, **automatisiert** (keine Kontrolle) und **parallel** analysiert. Die Objekte werden erst auf einer späteren Verarbeitungsstufe erkannt. Ihre Identifikation erfolgt bei der Merkmalsintegration **seriell**.

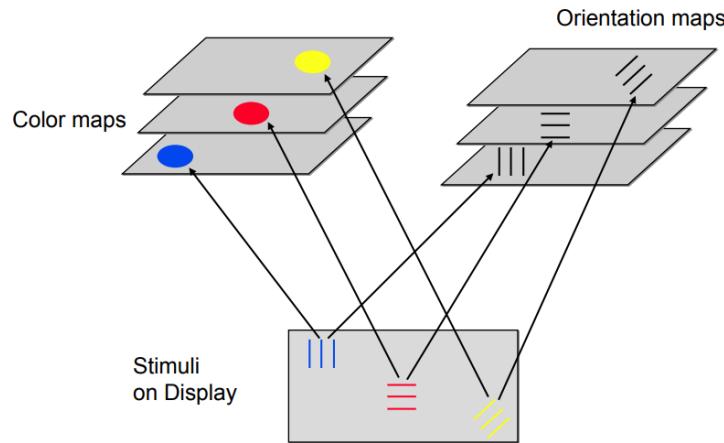
→ Ein **einzelnes Merkmal** kann also recht **schnell** und ohne Aufmerksamkeit identifiziert werden, das **Unterscheiden** von Objekten auf Basis mehrerer Merkmale hingegen ist **aufwendiger** und erfordert Aufmerksamkeit.

Bei Aufgaben wird getestet, ob ein einzelnes Merkmal oder eine Konjunktion von Merkmalen zur Lösung benötigt wird (z.B. Suchaufgaben)

9.13.2. Prozesse

9.13.2.1. Merkmalsanalyse

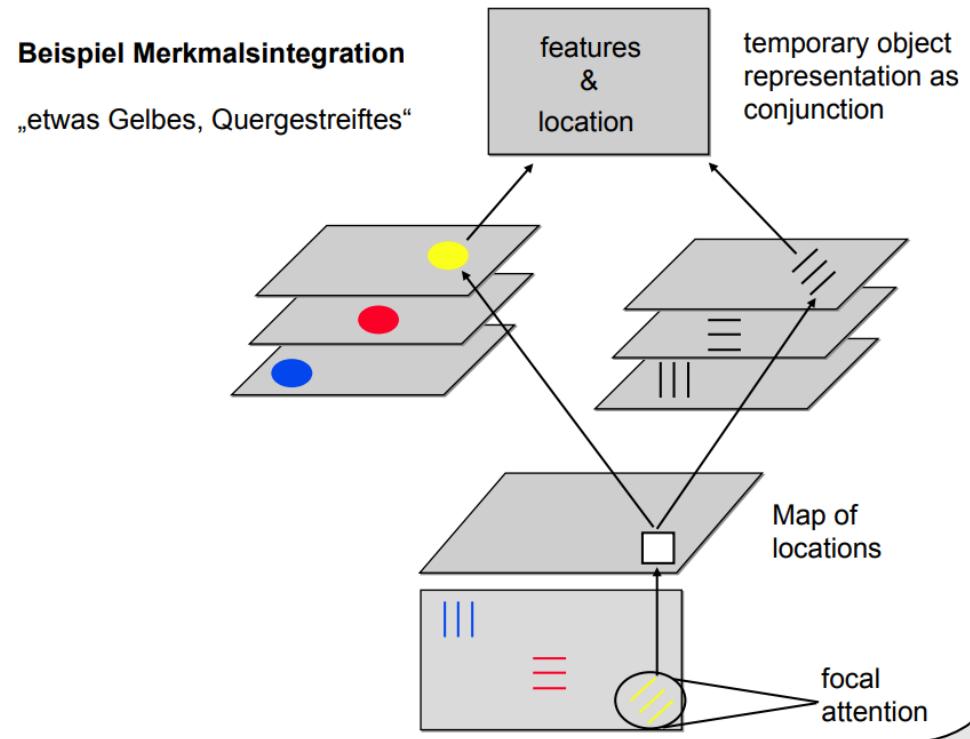
Die Analyse der Merkmal wird **parallel** von **Detektoren** vollzogen, die auf **physikalische Dimensionen** (Form, Orientierung, Farbe) getrimmt sind. Alle Merkmale werden in diesen Dimensionen kodiert und auf **Feature Maps** repräsentiert, die zueinander unabhängig sind.



9.13.2.2. Merkmalsintegration

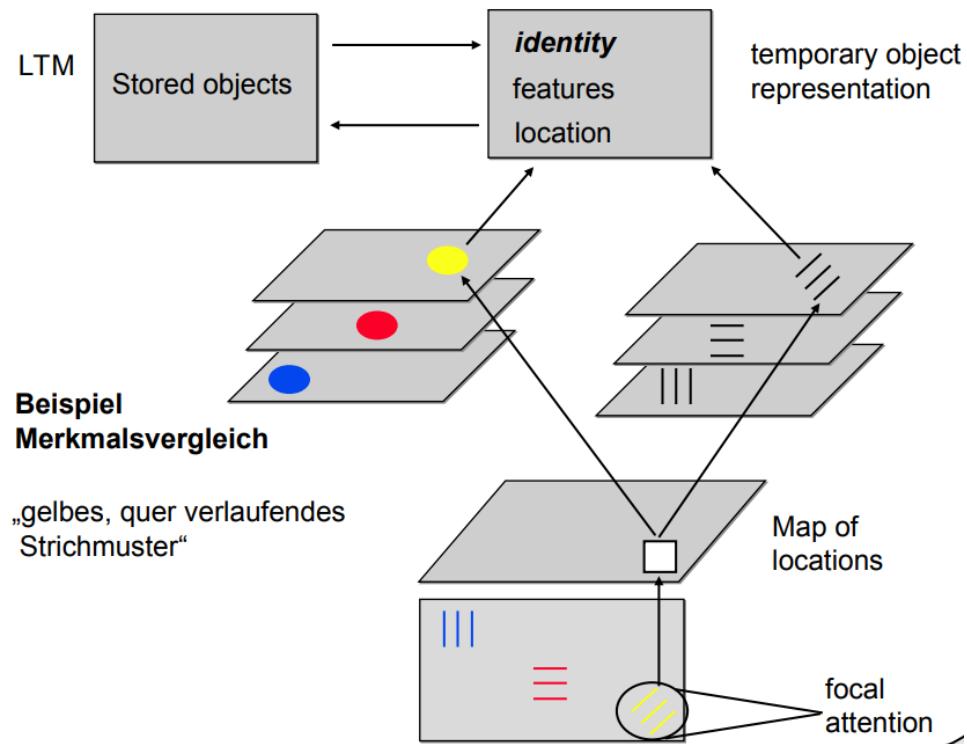
Um die Merkmale zu rekombinieren, werden die **Objektlokationen** (Ort auf dem Bild) nacheinander (seriell) mit Aufmerksamkeit (**Focal Attention**) belegt und alle Merkmale im selben Fokus werden zu einem Objekt kombiniert (**Konjunktion**). Die Merkmalskonjunktion wird nur **temporär gespeichert**.

→ Focal Attention wirkt wie "Kleber"



9.13.2.3. Merkmalsvergleich

Die temporäre Repräsentation der Merkmale wird mit dem Langzeitgedächtnis verglichen. Findet sich eine Übereinstimmung bei diesem Vergleich, so gilt das Objekt als erkannt.



9.13.3. Experiment von Treisman & Gelade

Aus der FIT abgeleitete Annahmen:

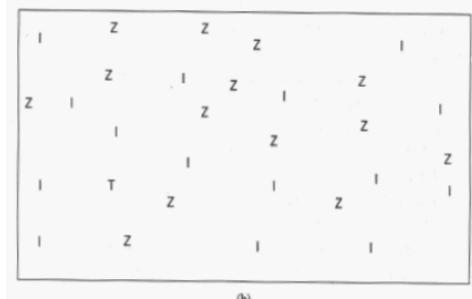
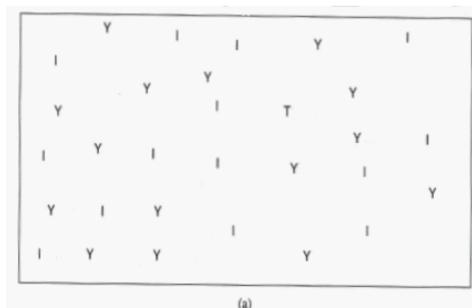
Wenn die Identifikation von Objekten die Kombination von Merkmalen unter Aufmerksamkeit erfordert, dann sollten:

1. Suchaufgaben, die **einzelne Merkmale** erfordern, müssten **schneller** gelöst werden als solche mit kombinierten Merkmalen.
2. Bei Suchaufgaben mit Objekten, wo das komplette visuelle Feld seriell abgesucht werden muss, steigt die Dauer mit der Anzahl der Objekte.

Experiment: Suche nach einem **Zielreize** in einer Menge anderer, ähnlicher Reize (**Distraktoren**) und Messung der **Reaktionszeit**.

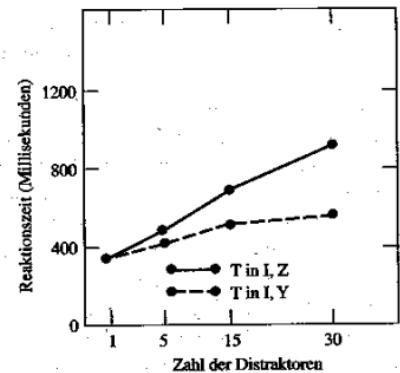
Aufgabe: Suche das T

- (a) ist schneller, weil man nur den oberen Querstrich finden muss (400ms)
- (b) dauert länger, weil man noch bei dem Querstrich differenzieren muss (T oben, Z unten)(800ms)



Ergebnis:

1. (a) 400ms, (b) 800ms, also Suche nach **einzelnen Merkmal schneller ✓**
2. bei Objekten (**seriell**) steigt die Zeit mit der Anzahl der Distraktoren ✓, bei einzelnen Merkmalen (**parallel**) bleibt es annähernd konstant



Fazit:

- einzelne Merkmale wegen paralleler Suche schnell erkannt
- Erkennung von Objekten basiert auf Merkmalskonjunktion und dauert länger, weil es eine serielle Belegung mit Aufmerksamkeit erfordert
- mit der Anzahl von Distraktoren steigt die Anzahl der Lokationen von Objekten, die verarbeitet werden müssen, weshalb die serielle Suche länger dauert.