Seminar "Humanoide Roboter"
Aufmerksamkeitssteuerung
WS04-05
Andrea Schürmann, Sven Pohl
aschuerm@techfak.uni-bielefeld.de
spohl@techfak.uni-bielefeld.de

# 1 Zusammenfassung

Der Schwerpunkt bei bottom-up Modellen, die sich mit der Aufmerksamkeitssteuerung befassen liegt in einer Saliency-Map, die die Auffälligkeit der Stimuli kodiert. Die Saliency-Map erhält Eingangssignale von verschiedenen Filtern, die das Bild angelehnt an neuronale Netzwerke im menschlichen Gehirn nach einfachen visuellen Merkmalen vorverarbeiten. Mit Hilfe von Hemmung der aktuell betrachteten Position lassen sich Blicktrajektorien aufbauen, die der Auffälligkeit folgen. Zur Berücksichtigung von Szenen und Objektverständnis kann zusätzlich toirgendeineirgendeinep-down Einfluß eingearbeitet werden.

# 2 Einleitung

Visuelle Aufmerksamkeit spielt eine große Rolle bei der Erkennung von Objekten innerhalb des Blickfeldes. Sie kann sich auf bestimmte Regionen oder Merkmale konzentrieren und verbessert die kortikale Repräsentation eines Objektes und unterdrückt die Verarbeitung uninteressanter Stimuli.
Sie ist daher so wichtig, da es dem menschlichen Gehirn aufgrund seiner endlichen Kapazität nicht möglich ist alle Objekte in verschiedenen Orientierungen, Positionen und Größen parallel zu identifizieren [1].
Die Verarbeitung der Information verläuft im Gehirn entlang neuronaler Pfade (3). Sie kann nur dann parallel stattfinden, wenn sich diese Pfade nicht überlappen, da sonst Interferenzen zwischen den Stimuli auftreten. Die zu verarbeitende Information wird durch visuelle Aufmerksamkeit auf einzelne Objekte aufgespalten, um jene Inferenzen einzuschränken.
Dabei ist schon allein aus evolutionärer Sicht wichtig, dass die visuelle Aufmerksamkeit den Blick zuerst auf relevante Objekte richtet, z.B. Raubtier und Beute [2]. Relevante Objekte zeichnen sich oft dadurch aus, dass sie sich vom Rest des Blickfeldes abheben, auf irgendeine Eigenschaft hin einzigartig (engl.: salient) sind.
Unterschieden wird dabei zwischen in einem Kontext an sich auffälligen Stimuli, die sofort unbewusste Aufmerksamkeit auf sich ziehen und einer aufgabenstellungabhängigen Aufmerksamkeit, die allerdings erst ein paar hundert Millisekunden später einsetzt. Am häufigsten wird erstere modelliert, da sich die "saliency" allein aus den Bildmerkmalen berechnen lässt.

Im Späteren folgt ein Überblick vom allgemeinem Aufbau von bottom-up basierten Modellen (4.1) und Modellen, die den Einfluss der Aufgabenstellung mit einbeziehen (4.2).

# 3 Das menschliche visuelle System

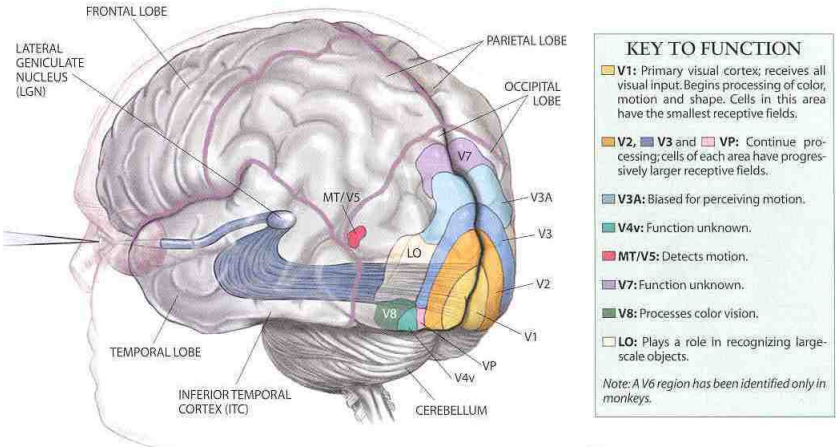



Abb.:1 menschlicher visueller Kortex [4]

Das Eingangssignal vom Auge gelangt über den *Nucleus geniculatum lateralis* fast vollständig in den primären visuellen Kortex. In diesem werden durch einfache Transformationen Oberflächen grob nach Farbe, Orientierung, etc. analysiert. Die gewonnene Information wird daraufhin parallel in zwei Richtungen weitergeleitet: den dorsalen Weg Richtung parietalen Lappen, in dem die Ortsbestimmung des betrachteten Objekts sowie der Lenkung der Aufmerksamkeit stattfindet , und ventral zum temporalen Lappen, in dem das Objekt indentifiziert wird.
Letzterer speichert Prototypen verschiedener Objekte, so dass diese unabhängig von Beleuchtung, Orientierung und Größe erkannt werden können. Neuronen in weiter hinten liegende Regionen erhalten von ebenfalls komplexeren Neuronen aus dem pariteal Lappen Transformationen des Bildes auf der Retina, und vergleichen es mit den Prototypen [3].
Modelle sind oft an die biologischen Eigenschaften von Neuronen angelehnt. So werden Neuronen, die auf Intensitätskontrast reagieren, durch "Mexicanhat"-Filter modelliert. Neuronen, die auf bestimmte Orientierungen ansprechen, kann man durch Gaborfilter simulieren. Modelle, die so eine detaillierte Merkmalsberechnung enthalten, lassen sich auch auf natürliche Umgebungen anwenden.

# 4 Modellierung

Modelle von der selektiven Aufmerksamkeit sind nicht nur in der Steuerung von Robotern von Nutzen (5.1). Sie bieten auch in der Psychologie die ideale Grundlage empirische Vorhersagen zu treffen. Erkenntnisse aus Fallstudien können mit genau spezifizierten Hypothesen in ein Modell gebracht werden. Jedes Experiment kann mit beliebiger Stimulusveränderung wiederholt werden, und Hypothesen somit überprüft werden. Auch unbekannte Interaktionen einzelner Komponenten könnten so entdeckt werden [1].
Ein paar erfolgreich angewandte Modelle werden später noch vorgestellt(4.1,4.2).

## 4.1 Bottom-up Modelle

Viele bottom-up Modelle für visuelle Aufmerksamkeit lehnen sich an das 1985 vorgestellte Modell von Koch und Ullman an (5). Dessen Hauptbestandteil ist eine "saliency map", welche die Auffälligkeit aller Bildmerkmale skalar repräsentiert. Bevor so eine saliency map berechnet werden kann, wird das visuelle Eingangssignal durch Filter, die auf einfachere visuelle Merkmale reagieren, vorverarbeitet. Der Blick folgt dann den Regionen, die sich auf der saliency map als am hervorstechensten erweisen. Jedes Bottom-up Modell setzt sich zusammen aus: der Vorberechnung visueller Merkmale, deren Zusammenführung in eine Saliency Map, auf der dann die Blickrichtung berechnet wird, und gegebenenfalls top-down Einfluss zum Verständnis der Szene.

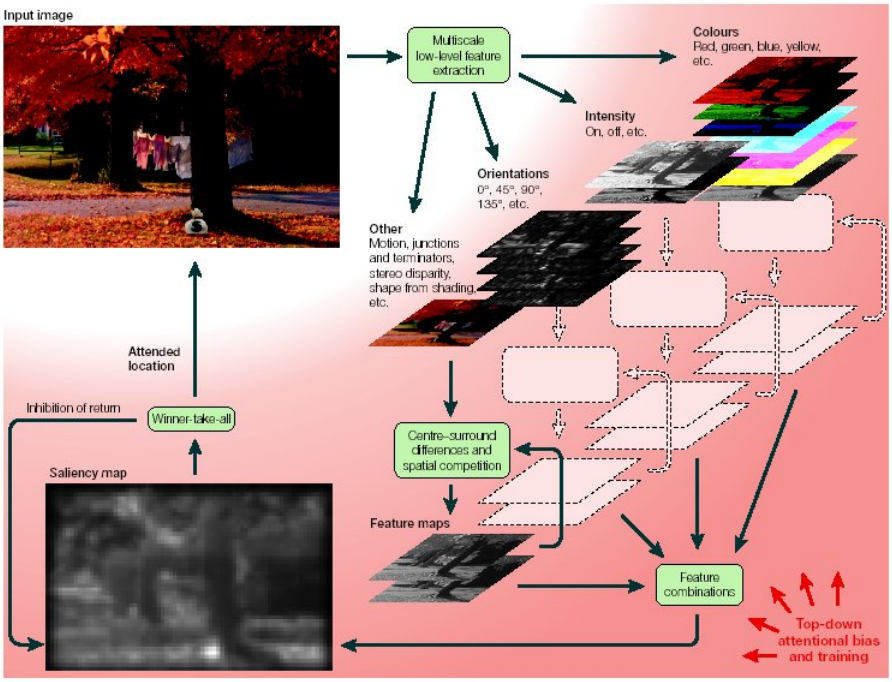



Abb.:2 Typisches Modell der bottom-up Attention[2]

### 4.1.1 Vorverarbeitung

Angelehnt an die Neuronen des menschlichen Gehirns, die auf Kontraste in Farbe, Helligkeit und Orientierung reagieren wird das Bild parallel durch verschiedene Filter in sogenannten "feature maps" vorverarbeitet. Neuronen in jeder dieser einzelnen feature maps konkurrieren um Einzigartigkeit miteinander, so dass viele mehr oder weniger hervorstechende Stellen auf möglichst wenige Cluster eingeschränkt werden.
Weiterhin kann man falls gewünscht die feature maps unterschiedlich gewichten, so dass bestimmte Eigenschaften auf die "saliency map" einen höheren Einfluss haben als andere. Während Neuronen stark auf verschiedene Ausprägungen eines Merkmals reagieren, besteht ähnlich zum menschlichen Gehirn zwischen verschiedenen feature maps keinerlei Interaktion. So fällt es dem Menschen schwer parallel nach zwei verschiedenen Merkmalen zu suchen [2].

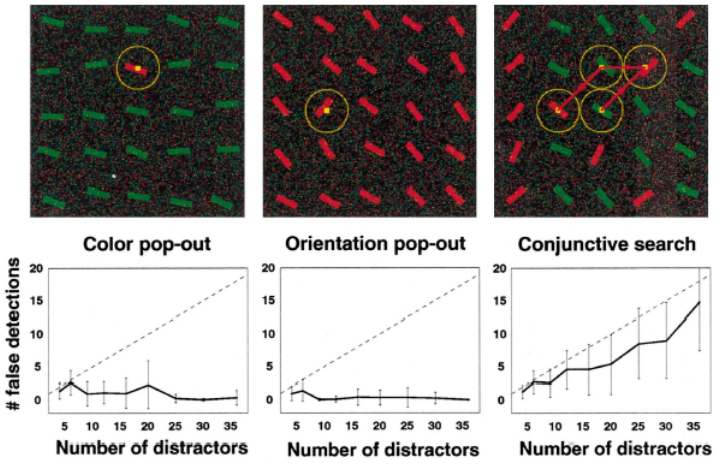



Abb.:3 Blickbewegungen eines Menschen bei konjunktiver Suche[5]

### 4.1.2 Saliency Map

Die Saliency-Map ist das Herzstück vieler Bottom-Up Modelle. Sie vereint die Antworten aus den feature maps. Es spielt zur Berechenung der Aufmerksamkeitssteuerung keine Rolle mehr, welches Merkmal eine Postion einzigartig gemacht hat. Eine Salieny Map ist skalar und entspricht topographisch dem Eingangsbild. Ihre Werte entsprechen der Einzigartigkeit eines Bildpunktes. Das heißt, je höher der Wert eines Bereichs auf der Saliency Map, desto hervorstechender ist jene Stelle. Mit Hilfe dieser Map reicht es, wenn der Blick sich auf dem höchsten Wert fokusiert, gefolgt von den nächst kleineren Werten. Dabei kann man auf unterschiedliche Weisen den Eingang aus den feature maps auswerten.
Tsotsos et al. haben in ihr neuronales Modell ein zusätzliches feedback eingebaut, das an der Gewinnerpostion nicht beteiligte feature maps aussschließt.
In jeder Stufe ihres hierarchischen Modells, an dessen Spitze die saliency map ist,

sind winner-take-all Netzwerke. So konkurrieren sowohl feature maps als auch weiter oben liegende Verarbeitungsstufen miteinander, und nur die Gewinner haben einen weiteren Einfluß auf die saliency [2].
Ein Modell von Milanese et al., das sich auch für natürliche Umgebungen eignet, besteht aus einem Energieoptimierungsproblem. Es minimiert die Zusammenhanglosigkeit zwischen verschiedenen feature maps, indem Regionen, die in mehreren feature maps salient sind, begünstigt werden. Gleichzeitig begünstigt es Clusterbildung innerhalb einer feature map und beschränkt sogleich die Gesamtaktivität. Nebenbei maximiert es die dynamische Reichweite einer jeden feature map, damit sie nicht uniform werden [2].

Nach der Berechnung der saliency map, muss nun die aktivste Stelle zur Setzung des Fokuses berechnet werden. Neuronal kann man auch das mit einem winner-take-all Netzwerk erreichen mit eingebauter Maximum Erkennung. Damit der Fokus zum nächsten Punkt wechseln kann, muss man die aktuell besichtigte Stelle vorübergehend deaktivieren, z.B. durch Hemmung der Neuronen dort. Dadurch kann das winner-take-all Netzwerk auf die nächste Position springen, und so Blicktrajektorien aufbauen.
Beim Menschen nennt man dieses Verhalten "inhibition of Return" (IOR) Im Unterschied zu einem Computer Modell ist es einem menschlichem Gehirn möglich, auch sich bewegende Objekte zu hemmen, bzw. Objekte vor einem sich bewegenden Betrachter.

## 4.2 Top-down Modelle

Ein einfaches bottom-up Modell reicht zwar für die ersten paar hundert Millisekunden aus, doch dann setzt beim Menschen die bewußte Betrachtung ein, die für die Objekterkennung wichtig ist. Es ist also von Vorteil, z.B. bei der Erkennung von Wörtern, dass ein Modell auch top-down Einflüsse verarbeiten kann.
Das Modell von Schill zur Objekterkennung [2] benutzt einen Wissensbaum, nach dem der Blick als nächstes auf die Stelle fällt, die zur Einordnung des Objektes am informativsten ist. Der Baum wird durch Training aufgebaut. Seine Blätter enthalten identifizierte Objekte und die Verzweigungen „je weiter oben sie liegen, immer allgemeinere Klassen. Auf den Ästen befinden sich Anweisungen zur Lenkung der Blickrichtung auf bestimmte Postionen von erwarteten bottom-up Merkmalen, um die Objektklasse einschränken zu können. Die Erkennung verläuft dem maximalen Informationsgewinn folgend iterativ durch den Baum. Das Modell ist somit in der Lage Objekte effizient in bekannte Klassen einzuteilen.

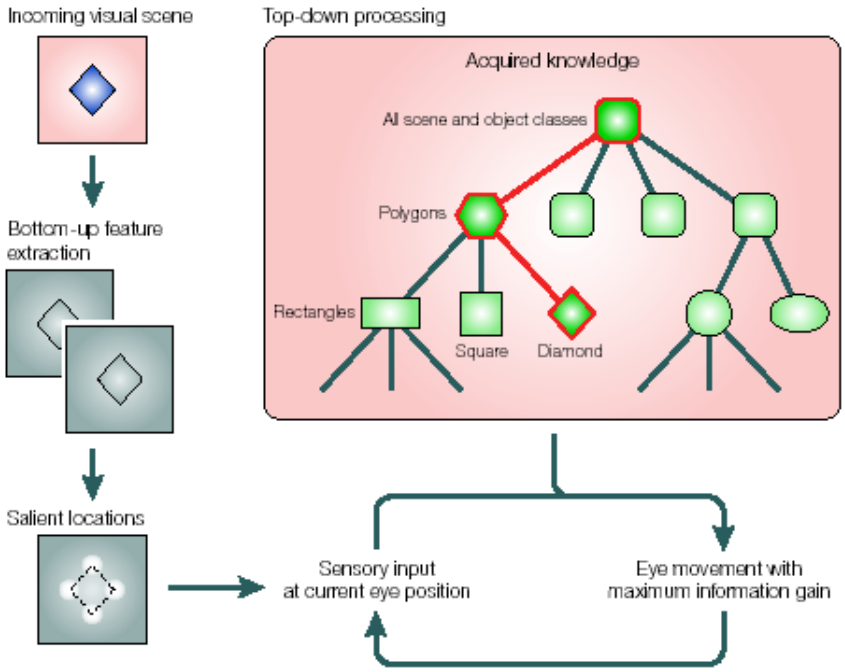



Abb.:3 Modell von Schill zur Aufmerksamsteuerung und Objekterkennung[2]

Das Modell von Rybak et al. ist mehr an die Biologie angelehnt und funktioniert unabhängig von Rotation und Skalierung vom Objekt. Es werden Scanpfade trainiert, dabei geschieht die Speicherung der Blickanweisungen in einem "where-memory", davon getrennt werden im "what-memory" die erwarteten Merkmale gespeichert. Durch Vergleich der vorhandenen Bildmerkmalen mit den gespeicherten und abwechselnder Lenkung der Blickrichtung nach den Anweisungen aus dem "where-memory" werden neue Bilder erkannt.

# 5 Implementation nach Itti,Koch und Niebur

Sven's part

## 5.1 Motivation und Anwendungen

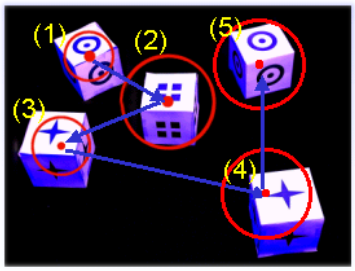
Das mehrschichtige System zur Modellierung dynamischer Aufmerksamkeitssteuerung ist für den Einsatz in Echtzeitanwendungen ausgelegt. Die unvorhersehbaren und vielfältigen Szenen im Strassenverkehr dienen für das vorgestellte Modell als geeignete Testumgebung. Hierbei liegen keine Einzelbilder mehr vor sondern eine ununterbrochene Videosequenz aus der Sicht des Fahrzeugs. Ziel ist die augenblickliche Detektion herausstechender Objekte wie z.B. Verkehrsschilder, Ampeln, Personen und anderer Fahrzeuge zur Unterstützung von höherliegenden Sicherheitssystemen. Auf dieser Ebene geht es noch nicht um konkrete Objekt-Klassifikation sondern zunächst um die Eingrenzung des Merkmalsraums auf die signifikanten Bereiche des Blickfeldes. Diese grenzen sich von der Umgebung ab und sind somit potentiell von grösserem Interesse.

Eine derartige Echtzeit-Aufmerksamkeitssteuerung kann als Basis für Fahrzeug-Sicherheitssysteme eingesetzt werden. Diese können dann den Fahrer auf erhöhte Geschwindigkeit oder Haltesignale aufmerksam machen und dabei helfen, Kollisionen mit plötzlich auftretenden Objekten oder Personen zu vermeiden. Darüberhinaus ist das Aufmerksamkeitssystem auch als Vorbereitung zur allgemeinen Objektklassifikation anwendbar und wird auch in experimentellen Robotik - Anwendungen eingesetzt. Insbesondere für humanoide Roboter ist die schnelle Detektion relevanter Reize umbedingt erforderlich, da diese in Echtzeitumgebungen agieren und sofort bedeutsame Objekte oder Personen von weniger bedeutsamen unterscheiden müssen. Aufmerksamkeitssysteme haben ebenfalls Anwendungen in der Gebäudeüberwachung, automatisierten Suche in Bilddatenbanken sowie in einigen militärischen Bereichen.


*Anwendung für eine Aufmerksamkeitssteuerung*

## 5.2 Modell-Architektur

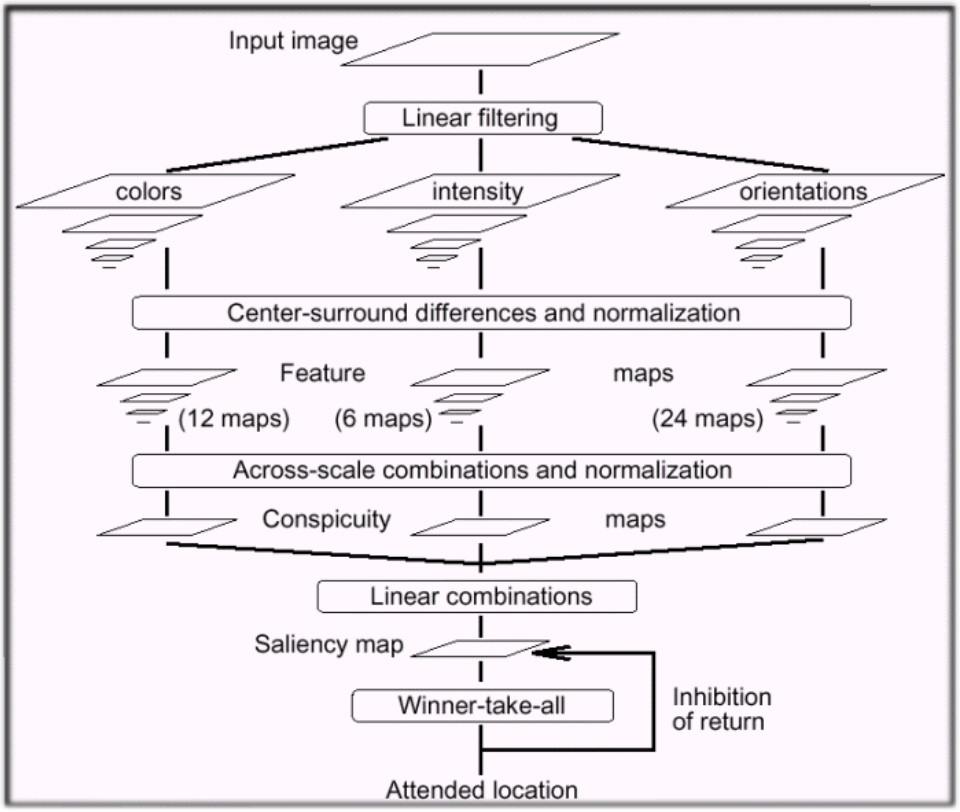
An ein Echtzeitsystem für Aufmerksamkeit werden hohe Anforderungen gestellt. Es soll ein bedeutsames Objekte innerhalb einer komplexe Szene sofort erkennen und danach augenblicklich das nächste potentiell bedeutsame Objekt fixieren. Dabei sollen die resultierenden Fixationen denen einer betrachtenden Person möglichst nahe kommen und realistische Blicktrajektorien erzeugen. Adaptiv sollten häufige auftauchende Objekte mit geringer Bedeutung nach und nach weniger fokusiert werden als neue bislang unbekannte Ziele. Echtzeitdaten haben die Eigenschaft in sämtlichen Beläuchtungssituationen und in verrauschter und schlechter Bildqualität aufzutreten. Das konkrete Anwendungsgebiet soll möglichst universell sein.




*Eine typische Blicktrajektorie eines Menschen*

Zusammenfassend soll das System eine komplexe Szene ähnlich betrachten wie ein Mensch. Das visuelle System wird im wesentlichen von zwei antagonistischen Prinzipien beeinflusst: einer einfachen bottom-up und einer komlexeren top-down Strategie. Da die top-down Strategie Weltwissen erfordert, das dem System hier nicht zur Verfügung steht, begnügt es sich mit der Modellierung der simpleren bottom-up Strategie. Die biologischen Grundlagen stammen hauptsächlich aus Forschungen an Primaten, deren visuelles System dem menschlichen sehr ähnlich ist, wobei aber der bottom-up Aspekt stärker ausgeprägt ist [6]. Die bottom-up Mechanismen sind im Mittelhirn untergebracht und daher entwicklungsgeschichtlich älter als die jüngeren cortikalen top-down Mechanismen. Das bottom-up System konnte sich evolutionär behaupten und ist trotz seines relativ einfachen Aufbaus sehr effektiv.

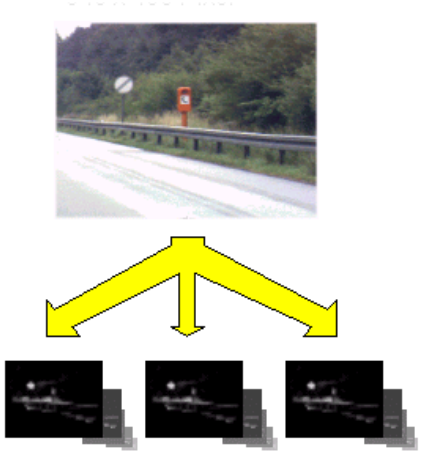
Eine weitere Eigenschaft des visuellen Systems ist die parallele Verarbeitung der verschieden Bildmerkmale Farbe, Intensität und Orientierung. In Anlehung an das biologische Vorbild werden diese Bildmerkmale zunächst getrennt vorverarbeitet und nach einer Normalisierung zu einer 'saliency-map', einer Karte herausstechender Objekte zusammengefügt. Im Anschluss wird diese Karte einem einfachen neuronalen Netz präsentiert, einem selbstinhibieren Winner-take-all (WTA)-Netzwerk. Dieses arbeitet letztendlich als oberster Entscheider, welche Region in einer Szene den Fokus erhält. Dieses Netz ermöglicht durch seinen adaptiven Charakter den Einsatz im Echtzeit-Szenario.




*Schema des Systems nach Itti und Koch*

Der Video-Datenstrom wird in zunächst in Einzelframes zerlegt, die in das Aufmerksamkeitssystem kontinuierlich eingespeist werden. Jedes Frame hat dabei eine Auflösung von 640 mal 480 Pixeln. Anschliessend findet eine Feature-Extraktion statt, die Farbe, Intensität und Orientierung getrennt verarbeitet. So entstehen insgesamt 42 einzelne Karten aller 3 Merkmale.

Jede Karte wird pixelweise mit der Umgebung verrechnet. Um die unterschiedlichen Merkmalskarten zu einer Saliency-map zusammenzufügen, werden sie zunächst normalisiert. Anschliessend werden sie linear kombiniert und dem WTA-Netz präsentiert.



*12 x Farbe, 6 x Intensität ,24 x Orientierung*

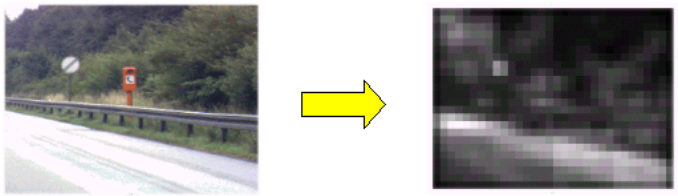
Dieses Architektur bewirkt, dass herausstehende Merkmale nicht nur von einem Feature abhängen sondern als Kombination herausstechender Merkmale betrachtet werden.

## 5.3 Die Merkmals-Karten

### 5.3.1 Intensität

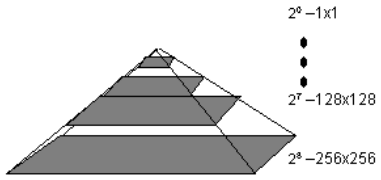
In biologischen System sind Neuronen oft in Form *rezeptiver Felder* organisiert. Sowohl in der Retina als auch im primären visuellen Cortex reagieren Neuronen auf einen hohen Zentrum-Umfeld Kontrast mit besonders hoher Feuerfrequenz. Diese ON/OFF-Felder reagieren besonders auf lokale hell/dunkel-Differenzen in der Umgebung.

*Rezeptive ON/OFF-Felder*



*Transformation des Eingabebildes in ein Intensiätsbild*

Um die Umfeld-Beziehung zu beschreiben, kann eine Gauss-Pyramide verwendet werden. Das Bild wird in Unterblöcke $c$ der Scalierung *scale* unterteilt und pixelweise vom Umfeld $s$ abgezogen. Dadurch entstehen für das Merkmal Intensität 6 Karten für jede Ebene der Gauss-Pyramide.




Gauss-Pyramide $\mathcal{I}\{\sigma\}$ mit $\sigma \in [0..8]$

*Ebenen der Gauss-Pyramide*

Verwendet Skalierung:
scale $0 = 1 : 2^1$... scale $8 = 1 : 2^8$

Zentrum:
$c \in \{2, 3, 4\}$

Umgebungsvariablen:
$s = c + \delta$
$\delta \in \{3, 4\}$

Gauss-Pyramide:
$\mathcal{I}\{\sigma\}$ mit $\sigma \in [0..8]$

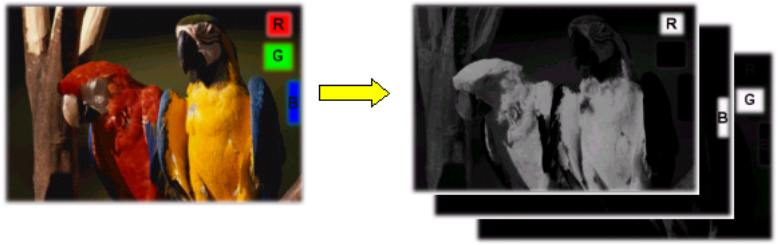
$$\mathcal{I}(c, s) =| \ \mathcal{I}(c) \ \ominus \ \mathcal{I}(c) \ | \longrightarrow \text{ 6 Karten}$$

### 5.3.2 Farbe

Neuronale rezeptive Felder existieren auch für Farben, sogenannte *Blobs*. Diese existieren für die Farbkombinationen Rot/Grün, Grün/Rot, Blau/Gelb und Gelb/Blau im menschlichen Kortex.

Farb-Kanäle:
Rot R=r-(g+b)/2
Grün G=g-(r+b)/2
Gelb Y=(r+b)/2- $| r - g |$ /2

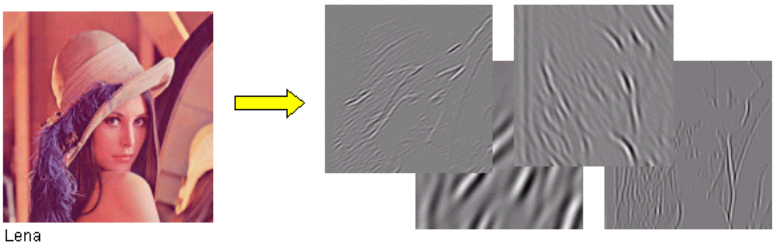
12 Karten:
$$RG(c, s) =| \ (R(c) - G(c)) \ \ominus \ (G(s) - R(s)) \ |$$
$$BY(c, s) =| \ (B(c) - Y(c)) \ \ominus \ (Y(s) - B(s)) \ |$$




*Vorverarbeitung: Zerlegung in Rot-Grün und Blau-Gelb Kanäle*

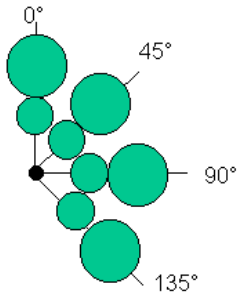
### 5.3.3 Orientierung

Biologische *Simple Zellen* antworten auf Kanten, Ecken und Balken mit hoher Feuerfrequenz. Hierbei können Gabor-Filter verwendet werden, bei denen Bilder und Filter per Fast-Fourier-Transformation in den Frequenzraum übertragen werden. Vergleiche im Frequenzraum zeichnen sich durch eine geringe algorithmische Komplexität aus und können in Echtzeit berechnet werden.




*Gabor-Filter am beliebten Beispiel Lena*

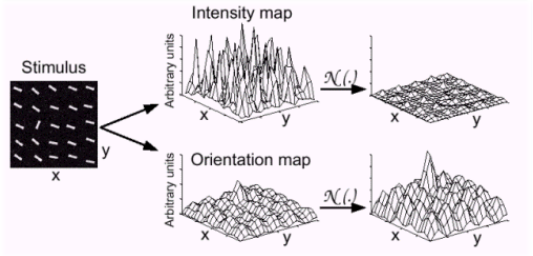
Feature-Map 'Orientierung'
$\Theta \in 0^o, 45^0, 90^o, 135^o$
$O(c,s,\Theta) =\mid O(c,\Theta) \ \ominus \ O(s,\Theta) \mid \longrightarrow$ 24 Karten (6x4)




*4 Gabor-Orientierungen in 2 Skalierungen*

## 5.4 Karten-Normalisierung

Die 3 Merkmale Farbe,Intensität und Orientierung stellen nicht-vergleichbare Modalitäten dar. Es sollen nur lokale Maxima berücksichtigt werden. Alle Merkmale sind gleichberechtigt und voneinander unabhängig, was in den Algorithmus einbezogen wird. Zu diesem Zweck werden die Karten derart normalisiert, dass nur global heraustretende Merkmale hoch gewichtet werden. Außerdem ermöglicht die Normalisierung die lineare Kombination dieser unterschiedlichen Feature.




*Der normalisierungs-Operator N(.)*

Hier der *N(.)*-Operator-Algorithmus in 4 Phasen:

1. Normalisierung aller Werte auf $[0..M]$
2. Globales Maximum $M$ finden
3. Durchschnitt $\overline{m}$ aller anderen Maxima berechnen
4. Karte global mit $(M\text{-}\ \overline{m})^2$ multiplizieren

Nach der *N(.)*-Normalisierung werden die Karten für Intensität, Farbe und Orientierung (I,C und O) aus den Sub-Karten zusammengesetzt.

$$I = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(\mathcal{I}(c,s))$$

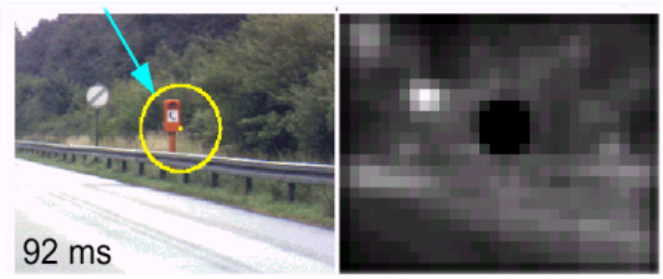
$$C = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [N(RG(c,s)) + N(BY(c,s))]$$

$$O = \sum_{\Theta \in \{0^o, 45^0, 90^o, 135^o\}} N(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(O(c,s,\Theta)))$$

Anschließend werden diese 3 Merkmalskarten linear kombiniert und sind somit vorbereitet zur Präsentation des WTA's:

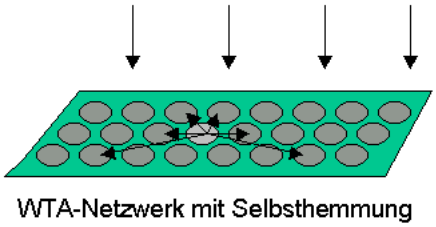
$$S = \tfrac{1}{3}(N(I) + N(C) + N(O))$$

## 5.5 WTA-Saliency Map




*Saliency-Map WTA*

Die erzeugte Saliency-Map erzeugt ein Aktivitätsgebirge aller Feature-Intensitäten. Um die Aufmerksamkeit auf andere herausstechende Bereiche der Karte zu lenken wird sie einem inhibierenden WTA-Netzwerk präsentiert. In der vorliegenden Modellierung wird ein Netzwerk aus *spikenden* Neuronen verwendet. Diese Architektur hat die Eigenschaft, dass benachbarte Objekte zuerst fixiert werden. Zirkulationen zwischen einigen wenigen dominanten Objekten werden unterdrückt. Das Ergebnis ist eine natürlich wirkende Blicktrajektorie auf den herausstechenden Objekten, selbst bei abwechselnden Frames eines Videos.

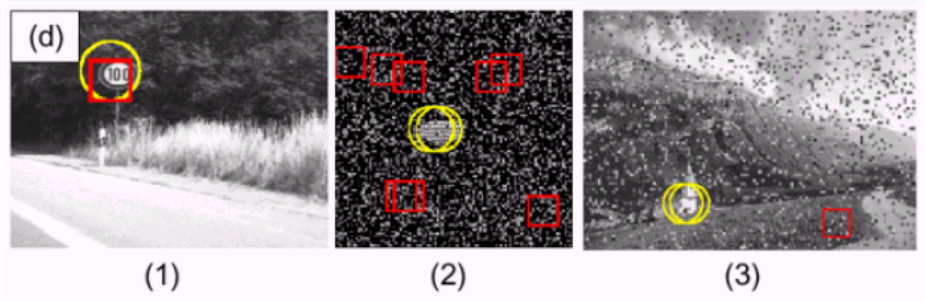



*WTA-Netzwerk mit Selbsthemmung*

# 6 Diskussion

## 6.1 Vergleich mit SFC-Modellen

In vielen Aufmerksamkeitssteuerungen wird eine SFC (spatial frequency content) verwendet. Dabei liefert eine FFT Kooeffizienten, die in ihrer Qualität den Intensitätskarten ähnlich sind. Die SFC benutzt keine globale Normalisierung und operiert nur auf lokalen Intensitätsunterschieden. Bei verrauschten Daten reagiert die SFC weniger robust als die Saliency-Map.




*Vergleich Saliency-Map(gelb) mit SFC(rot)*

## 6.2 Stärken und Schwächen der Saliency-Map

In der Modell-Anwendung des Straßenverkehrs hat sich das System als sehr robust herausgestellt. Da auf 3 verschiedenen Hauptmerkmalen operiert wird, lässt sich das System gut parallelisieren (z.B. mit .Net, Corba oder TCP/IP-basierten Implementationen). Durch den einfachen Aufbau kann in Echtzeit-Szenarien gearbeitet werden. Ein weiterer Vorteil ist die enge Anlehnung an Erkenntnisse der Neurophysiologie und biologisch plausiblen Erweiterungen.

Durch die reine Bottom-Up-Architektur werden alle Objekte, die irgendwie aus ihrem Umfeld herausstechen, fixiert. Das trifft natürlich auch auf vermeindlich unbedeutende Objekte zu. Zusammenhänge zwischen den Features können nur indirekt erkannt werden. Auch bei T-Verbindungen und Linienenden zeigt das System signifikante Schwächen. Insbesondere in bewegten Szenarien sind noch Verbesserungen möglich, z.B. bei der Verfolgung bewegter Objekte, die von Interesse sind.

## 6.3 Aussicht

Systeme zur Aufmersamkeitssteuerung gewinnen zunehmend an Bedeutung. Allein der Einsatz in Fahrzeugsicherheitssystemen stellt immer höhere Anforderungen. Damit drängt sich immer mehr der Top-down Aspekt des visuellen Systems in den Vordergrund. Das System benötigt hierbei mehr 'Weltwissen', um beispielsweise die Aufmerksamkeit nicht wiederholt auf herausstechende Pflanzen oder Bäume am Strassenrand zu richten. Als erster Schritt in diese Richtung werden per Top-Down Inhibition nicht benötigte Feature adaptiv gehemmt und damit in den Hintergrund gedränkt.

Viele Aufmerksamkeitssysteme sollen nicht alle herausstehenden Objekte detektieren sondern ein besonderes, bewegtes Objekt in den Fokus nehmen. Hierzu ist es notwendig, das System in die Lage zu versetzen, frameübergreifend ein Objekt zu markieren und für eine bestimmte Periode zu verfolgen, ohne sich dabei von anderen herausstehenden Objekten ablenken zu lassen.
Als weites Anwendungsgebiet wird in der Robotik experimentell mit autonomen 'Buggys' gearbeitet, die dabei beliebigen Objekten ausweichen. Hierbei verschmelzen Aspekte der Pfadplanung, Aufmerksamkeitssteuerung und Objekterkennung miteinander.

Die NASA testet z.Z. die autonome Steuerung von Fluggeräten auf Basis einer visuellen Aufmerksamkeitssteuerung, eine weitere anspruchsvolle Echtzeitumgebung für Aufmerksamkeitsmodelle.

Bemerkenswert sind auch einige biokybernetische Anwendungen, die sich nicht an Primaten, sondern an Insekten orientieren. Insekten verfügen nur über ein relativ kleines Nervensystem und meistern auf Basis einiger visueller Reflexe nicht nur Bottom-up Sehen sondern wenden auch Top-Down Strategien an, indem sie die für sie bedeutungslosen Objekte gezielt ignorieren.

Aufmerksamkeitssysteme, die auf biologischen Vorbildern basieren, sind wesentliche Bausteine zur Weiterentwicklung humanoider Roboter.

# Literatur


[1] Michael C. Moore, Mark Sitton, *Computational modeling of spatial attention*, Attention (1998) 341-393

[2] Laurent Itti, Christof Koch, *Computational modeling of visual attention*, Macmillan Magazines Ltd (2001) volume 2

[3] E.Guigon et al., *Neural network models of cortical functions based on the computational properties of the cerebral cortex*, J. Physiology Paris (1994) 88, 291-308

[4] *Vision: A Window on Consciousness*,Scientific American, 1999

[5] Laurent Itti, Christof Koch, *A saliency-based search mechanism for overt and covert shifts of visual attention*, Vision Research 40 (2000) 1489-1506

[6] Laurent Itti, Christof Koch,*A Model of Saliency-based Visual Attention for Rapid Scene Analysis (2001)*

[7] Laurent Itti, Christof Koch, *Computional modeling of visual attention.* Nat Rev Neurocience 194-203.

[8] Maik Bollmann (1999), *Entwicklung einer Aufmerksamkeitssteuerung für ein aktives Sehsystem.* Dissertation

[9] Treue S (2001). *Neural correlates of attention in primate visual cortex.* Trends Neuroscience. 295-300