Universität Ulm

Abteilung Neuroinformatik

Prof. Dr. Heiko Neumann

Seminar Sehen und Hören

[Vortrag am 23.6.2004]

Theorien der visuellen Aufmerksamkeit und Suche

Ausarbeitung von

Tobias Rittel

tobias.rittel@informatik.uni-ulm.de

Inhaltsverzeichnis

		Seite
1.	Einleitung	3
2.	Pop-Out-Effekt bei der visuellen Suche	3
3.	„A Feature-Integration Theory of Attention“ (FIT)	5
3.1	Mehrere Phasen der Wahrnehmung	5
3.1.1	Merkmaldetektion	5
3.1.1.1	„Feature-Maps“	6
3.1.2	Objekterkennung	6
3.1.2.1	Gerichtete Aufmerksamkeit	6
3.1.2.2	Objektspeicherung	7
3.1.2.3	Objektidentifizierung	7
3.2	Darstellung der FIT	7
3.3	Beweis der Theorie	7
3.3.1	Visuelle Suche	7
3.3.2	Texturbereichstrennung	8
3.3.3	Illusionäre Verbindungen	9
3.3.4	Identifizierung und Ortsbestimmung	9
4.	„Guided Search“	9
4.1	Abweichende Testergebnisse	9
4.2	Modifikation der FIT	10
4.2.1	Verbesserte gerichtete Aufmerksamkeit	10
4.3	Weitere Modifikation	10
4.4	Guided-Search Modell	11
5.	Computermodelle zur visuellen Wahrnehmung	11
5.1	Trennung Bottom-Up und Top-Down Modell	11
5.2	Salienzkarte	12
5.3	Fünf Rahmenpunkte für Computermodelle der Aufmerksamkeit	12
5.3.1	Punkt 1: pre-attentive Aspekt	12
5.3.2	Punkt 2: Salienzkarte	12
5.3.3	Punkt 3: „Attentional Scanpath“ und „IOR“	12
5.3.4	Punkt 4: Interaktion von offensichtlicher und verborgener	13
5.3.5	Punkt 5: Gerichtete Aufmerksamkeit durch Szenenverständnis	13

1. Einleitung

Würde man auf einmal im Zentrum einer unbekannten Stadt aufwachen, wäre der erste Eindruck den man hat, erkennbare Objekte zu sehen. Man sieht Häuser, Leute, Autos und Bäume. Man ist sich nicht bewusst, Farben, Ecken, Bewegungen und Entfernungen zu erkennen und diese zu Einheiten zusammenzusetzen, für die man aus dem Gedächtnis Bezeichnungen abrufen kann [1].
Aber welche Vorgänge geschehen hier im visuellen System des Menschen? Wie gelingt es uns, die erkannten Farben, Ecken, Bewegungen und Entfernungen zu erkennbaren Objekten zusammenzusetzen? Ist hierfür Aufmerksamkeit notwendig oder läuft dies automatisch ab? Kann man diesen Vorgang auch im Computer simulieren?
Diese und ähnliche Fragestellungen sind die Grundmotivation, das visuelle System des Menschen genauer zu untersuchen und herauszufinden, welche Prozesse im Gehirn zur Erkennung von Objekten ablaufen!

2. *Pop-Out*-Effekt bei der visuellen Suche

Die visuelle Suche ist eine häufige Aufgabe bei visuellen Wahrnehmungstests. Versuchspersonen müssen hierbei ein vorher bekannt gegebenes Zielobjekt („Target“[1]) unter mehreren Ablenkobjekten (Distraktoren) finden. Gemessen wird die Reaktionszeit, welche die Versuchsperson benötigt, um eine Entscheidung zu treffen. Entweder ob das *Target* vorhanden ist oder nicht.

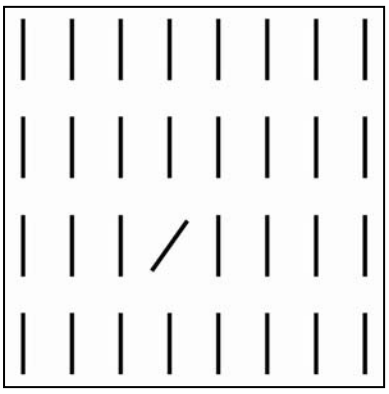

Abb. 1: Beispiel einer visuellen Suchaufgabe

Eine Versuchsanordnung könnte zum Beispiel sein, eine diagonale Linie unter einer Menge von geraden Linien zu finden (Vgl. Abb. 1). Als Aufgabe würde hierbei der Testperson die Frage gestellt, ob eine diagonale Linie im Bild vorhanden ist?
Betrachtet man Abbildung 1, so springt einem die diagonale Linie ja förmlich ins Auge. Dieser Effekt des ins Auge springen wird in der Fachsprache auch *Pop-Out*-Effekt genannt. Die Versuchsperson wird vermutlich sehr schnell eine Antwort geben können, wenn die diagonale Linie vorhanden ist.

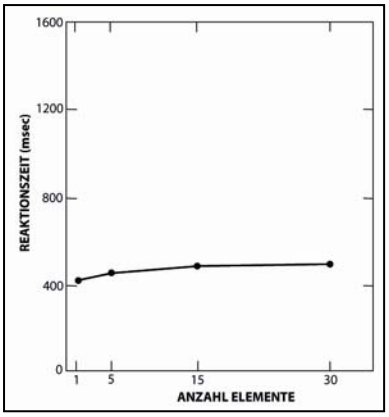

Abb. 2 [2]: Ergebnisgraph zu Abb. 1

Vergleichbare, professionell durchgeführte Tests [2] können diesen Effekt experimentell belegen. Hier wurde festgestellt, dass die Reaktionszeit zum Einen sehr kurz ist und zum Anderen auch unabhängig von der Anzahl der Distraktoren im Bild. Es spielt also keine Rolle, ob fünf oder 30 Linien im Bild gezeigt werden, die Reaktionszeit bleibt gleich.
Abbildung 2 zeigt den Ergebnisgraph einer vergleichbaren Testanordnung. Hier sieht man, dass die Reaktionszeit (Y-Achse) unabhängig von der Anzahl der Elemente (X-Achse) ist, da sie fast konstant bleibt!

[1] Im weiteren Text sind englische Ausdrücke *kursiv* gekennzeichnet.

Eine weitere Versuchsanordnung eines visuellen Suchtests ist in Abbildung 3 dargestellt. Die Testpersonen müssen hier entscheiden, ob ein horizontaler, schwarzer Balken vorhanden ist. Betrachtet man Abbildung 3, so benötigt man eine gewisse Zeit, bis man das *Target* gefunden hat. Der schwarze, horizontale Balken springt einem nicht so ins Auge wie dies vergleichsweise bei voriger Versuchsanordnung (Vgl. Abb. 1) der Fall war. Ebenfalls professionell durchgeführte Tests mit vergleichbarer Versuchsanordnung [2] zeigen denselben Effekt. Versuchspersonen benötigen eine längere Reaktionszeit und diese steigt mit der Anzahl der Elemente auch linear an!

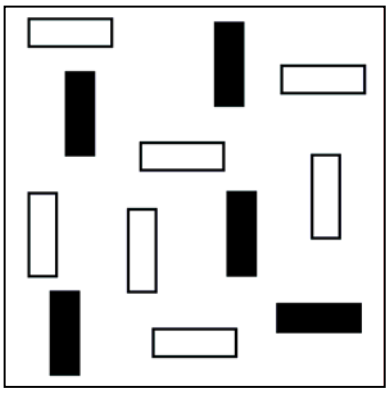

Abb. 3: Zweites Beispiel einer visuellen Suchaufgabe

In Abbildung 4 ist der Ergebnisgraph eines vergleichbaren Tests dargestellt. Die Reaktionszeit (Y-Achse) steigt linear mit der Anzahl der Elemente (X-Achse) im Bild!

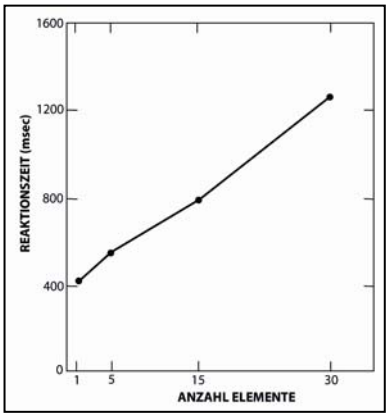

Abb. 4 [2]: Ergebnisgraph zu Abb. 3

Wo liegt nun aber der Unterschied zwischen beiden Tests? Betrachten wir Abbildung 1, so ist das einzige Merkmal („Feature“), welches das *Target* von den Distraktoren unterscheidet, die Orientierung. Ansonsten sind beide Elemente identisch. In Abbildung 3 hingegen unterscheiden zwei *Features* das *Target* von den Distraktoren: zum Einen die Farbe und zum Anderen die Orientierung. Hier kommt es zu keinem *Pop-Out*-Effekt. Die Versuchsperson muss ein Element nach dem anderen betrachten um das *Target* zu finden, führt also eine serielle Suche durch.

An dieser Stelle könnte man annehmen, dass die visuelle Suche nach nur einem *Feature* zu einem *Pop-Out*-Effekt führt, wohingegen die Suche nach einer Verbindung („Conjunction“) von *Features* eine serielle Suche impliziert. Anne Treisman verfolgte diesen Aspekt weiter und führte Tests [1] durch um zu sehen, ob die *Feature*-Suche immer zu einem *Pop-Out*-Effekt führt. Abbildung 5 zeigt beide Versuchsanordnungen. Bei der rot umrandeten Versuchsanordnung muss die Testperson den Kreis ohne Liniensegment finden. Bei grün umrandeter Versuchsanordnung ist dies genau umgekehrt. Hier soll der Kreis mit Liniensegment gefunden werden. Jeweiliger Unterschied zwischen *Target* und Distraktoren ist nur das Liniensegment, also ein *Feature*. Bemerkenswert sind die Ergebnisse der Tests (Vgl. Abb. 6). Besitzt das *Target* das *Feature*, kommt es zu einem *Pop-Out*-Effekt und die Suchzeit ist unabhängig von der Anzahl der Elemente im Bild (Vgl. Abb. 6: grüne Linie). Fehlt dem *Target* das *Feature*, steigt die Reaktionszeit linear mit der Anzahl der Elemente im Bild (Vgl. Abb. 6: rote Linie). Wie man anschaulich feststellen kann, werden hier die Elemente im Bild (wie bei Abb. 3) einer seriellen Suche unterworfen [1].

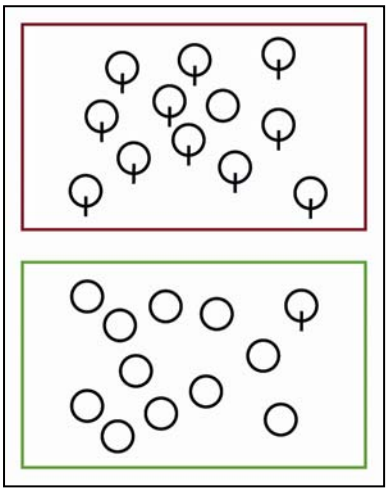

Abb. 5 [1]: visueller Suchtest

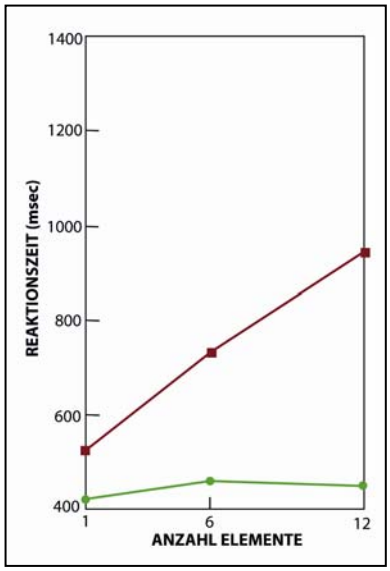

Abb. 6 [1]: Ergebnisgraph zu Abb. 5

Die *Feature*-Suche allein führt also zu keinem *Pop-Out*-Effekt. Entscheidend ist, ob das *Target* das *Feature* besitzt oder nicht!

Zusammengefasst wissen wir bis hier, dass die *Feature*-Suche zu einem *Pop-Out*-Effekt führt, wenn das Zielobjekt das *Feature* besitzt. Fehlt ihm das *Feature*, oder führen wir eine *Conjunction*-Suche durch, führt die Testperson eine serielle Suche durch und die Reaktionszeit steigt etwa linear mit der Anzahl der Elemente im Bild an!

3. „Feature-Integration Theory of Attention" (FIT)

Das derzeit wohl bekannteste Modell der visuellen Aufmerksamkeit ist das „*Feature-Integration Model*" von Anne Treisman [2]. Dieses Modell liefert eine schematische Darstellung, in welchen Phasen und Schritten der visuelle Wahrnehmungsprozess abläuft. Interessant ist die Art und Weise, wie Treisman auf dieses Modell kommt. Ausgehend von bereits bewiesenen Aussagen der Physio- und Psychologie (z.B. dass bestimmte Regionen im Gehirn auf bestimmte Eigenschaften wie Orientierung oder Bewegung besser reagieren), stellt sie logisch gefolgerte Hypothesen (ihr Modell sozusagen) zum Wahrnehmungsprozess auf, welche dann durch Experimente bewiesen werden.

3.1 Mehrere Phasen der visuellen Wahrnehmung

Die Grundannahme ihres Modells deckt sich mit der Aussage von Psychologen, Physiologen und Informatikern, von mehreren Phasen der visuellen Wahrnehmung zu sprechen [1].
In der ersten Phase, dem so genannten „pre-attentive level", kommt es zur Detektion spezieller *Features*. In den folgenden Phasen kommt es dann durch die Kombination dieser Merkmale zur eigentlichen Objekterkennung.

3.1.1 Merkmaldetektion

Die Erkenntnis aus Punkt 2, dass die Präsenz eines *Features* beim *Target* zu einem *Pop-Out*-Effekt führen kann, erklärt sich Anne Treisman folgendermaßen: bestimmte Merkmale werden in der ersten Phase unseres Wahrnehmungsprozesses, im pre-attentive level, automatisch und parallel detektiert. Ein *Target* mit *Feature* springt uns ins Auge und wir müssen keine Leistung erbringen, das *Target* zu finden. Um herauszufinden, bei welchen *Features* dieser *Pop-Out*-Effekt auftritt, führte sie eine Vielzahl an Experimenten durch [1] und kam dabei zu dem Ergebnis, dass nur eine kleine Anzahl an *Features* in dieser Phase von unserem visuellen System detektiert wird [1].

3.1.1.1 „Feature-Maps“

Identifiziert hat Treisman dabei *Farbe*, *Orientierung*, *Größe* und *Distanz* [1]. Bei diesen Merkmalen kann es zu *Pop-Out*-Effekten kommen. Sie werden von unserem visuellen System automatisch und parallel detektiert, ohne dass wir hierfür aktiv eine Leistung erbringen müssen. Abbildung 7 zeigt die schematische Darstellung der FIT bis einschließlich der Phase Merkmaldetektion.

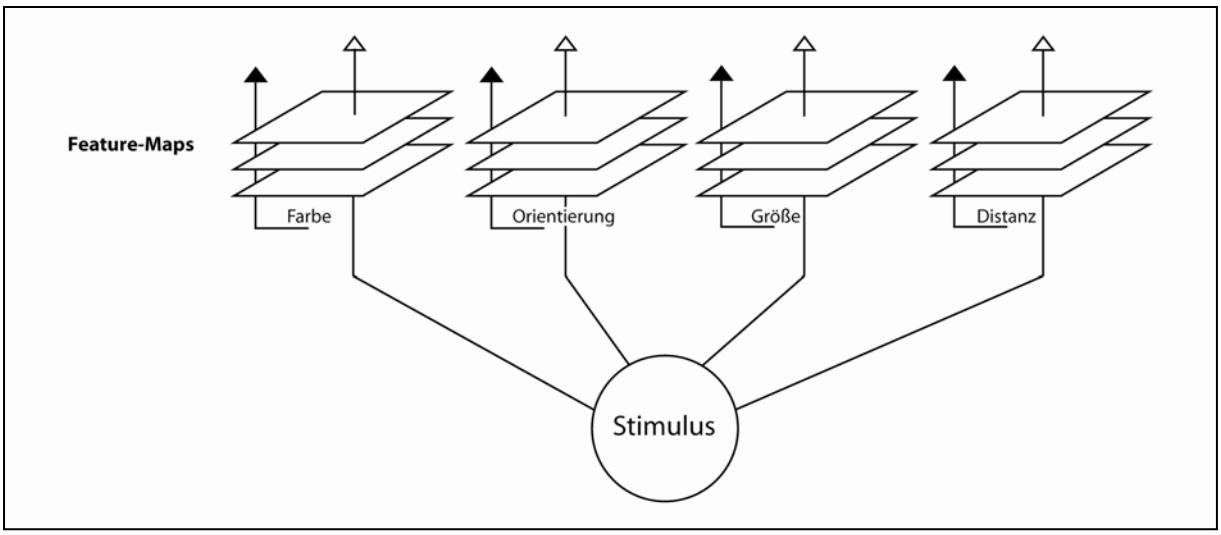

Abb. 7 [1]: Feature-Maps bei der FIT

Ausgehend von einem Bild / Stimulus werden die vier *Feature-Maps* parallel detektiert. Dabei werden Merkmale wie rot oder grün von der *Feature-Map* „Farbe“ detektiert, Winkel wie 0° oder 45° von der *Feature-Map* für Orientierung und so fort.

3.1.2 Objekterkennung

Nach der ersten Phase des visuellen Wahrnehmungsprozesses haben wir *Features* detektiert, können aber noch keine Aussage über das Objekt selbst machen. Bis zur endgültigen Objekterkennung sind noch mehrere Schritte notwendig.

3.1.2.1 Gerichtete Aufmerksamkeit

Bei dem in Punkt 2 genannten Versuchsbeispiel zur *Feature*-Suche haben wir zwar die diagonale Linie schnell erkannt, können wir aber bereits Aussagen über deren Lokation machen? Bei der *Conjunction*-Suche hingegen, die ja seriell abläuft, liegt die Vermutung nahe, auch Aussagen über den Ort des gefundenen *Targets* machen zu können. Treisman führte genau zu diesem Aspekt einen Versuch durch [1] und konnte damit genau diese Aussage stützen! Bei der *Feature*-Suche haben wir keine Information über den Ort des erkannten Merkmals. Bei der *Conjunction*-Suche allerdings, wo wir ein Element nach dem anderen aufmerksam betrachten, können wir den Ort des gefundenen *Targets* in den meisten Fällen richtig bestimmen!

Für die *Conjunction*-Suche ist somit Aufmerksamkeit, oder genauer gesagt, gerichtete Aufmerksamkeit nötig! Erst mit deren Hilfe lassen sich die *Features* an einem bestimmten Ort richtig zusammensetzen. Diese gerichtete Aufmerksamkeit kann man sich auch wie einen Scheinwerfer vorstellen, der sich nach und nach über den Stimulus hinweg bewegt und in seinem Fokus erkannte Merkmale kombiniert.

Abbildung 8 zeigt die schematische Darstellung der gerichteten Aufmerksamkeit im Modell der FIT.

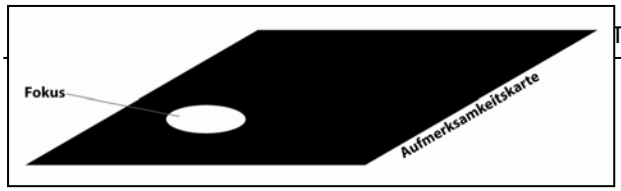

Abb. 8 [1]: gerichtete Aufmerksamkeit

Der Fokus (weißer Kreis) wird nach und nach über den Stimulus hinwegbewegt, der hier durch die Aufmerksamkeitskarte als schwarze Ebene repräsentiert wird.

3.1.2.2 Objektspeicherung

Ab diesem Punkt haben wir somit die Merkmale kombiniert und ein Objekt als Ganzes erkannt. Aber was geschieht, wenn sich das Objekt bewegt? Muss man dann erneut seine Aufmerksamkeit darauf richten und die Merkmale abermals zusammensetzen?
Hierzu stellt Treisman eine Analogie zu einer Datei auf [1]. Wurde ein Objekt erkannt, werden alle erhaltenen Informationen in einer Datei gespeichert. Ändert sich nun der Ort des Objektes, so wird nur diese Information in der Datei aktualisiert, entsprechend einem Update. Die Datei selbst, und somit das Objekt, bleibt weiterhin gespeichert und wird somit immer noch erkannt!

3.1.2.3 Objektidentifizierung

Eine Aussage, um welches Objekt es sich handelt, kann man aber noch nicht treffen!

Hier kommt ein *Top-down* Prozess zum Einsatz. Das erkannte Objekt wird mit bereits bekannten, im Gedächtnis gespeicherten Objekten verglichen und im Falle einer Übereinstimmung als dieses Objekt identifiziert!

3.2 Darstellung der FIT

Fasst man die in Punkt 3.1 genannten Aspekte zur visuellen Wahrnehmung zusammen erhält man die FIT nach Anne Treisman (Vgl. Abb. 9).
Ausgehend von einem Stimulus werden in der ersten Phase der visuellen Wahrnehmung bestimmte *Features* automatisch und parallel erkannt. In den folgenden Phasen kommt es dann zur Objekterkennung. Mit Hilfe der gerichteten Aufmerksamkeit werden die in ihrem Fokus erkannten Merkmale kombiniert und zu einem Objekt zusammengefügt. Anschließend wird dieses Objekt in einer Objektdatei gespeichert und mittels eines *Top-Down* Prozesses mit gespeicherten Objektbeschreibungen verglichen, wodurch das Objekt identifiziert werden kann!

3.3 Beweis der Theorie

Somit wäre die FIT definiert und erklärt. Um das Modell aber plausibel zu belegen, entwickelte Anne Treisman verschiedene Paradigmen [2], welche ihre einzelnen Hypothesen der Theorie testen und somit belegen können.

3.3.1 Visuelle Suche

Diesen Aspekt habe ich bereits in Punkt 2 ausführlich erläutert und auch Ergebnisse von Treisman´s Experimenten präsentiert. Die *Feature*-Suche wird automatisch und parallel durchgeführt, weshalb es zu *Pop-Out*-Effekten kommt. Bei der *Conjunction*-Suche findet ein serieller Prozess statt, der zeitabhängig von der Anzahl der Elemente im Stimulus ist.

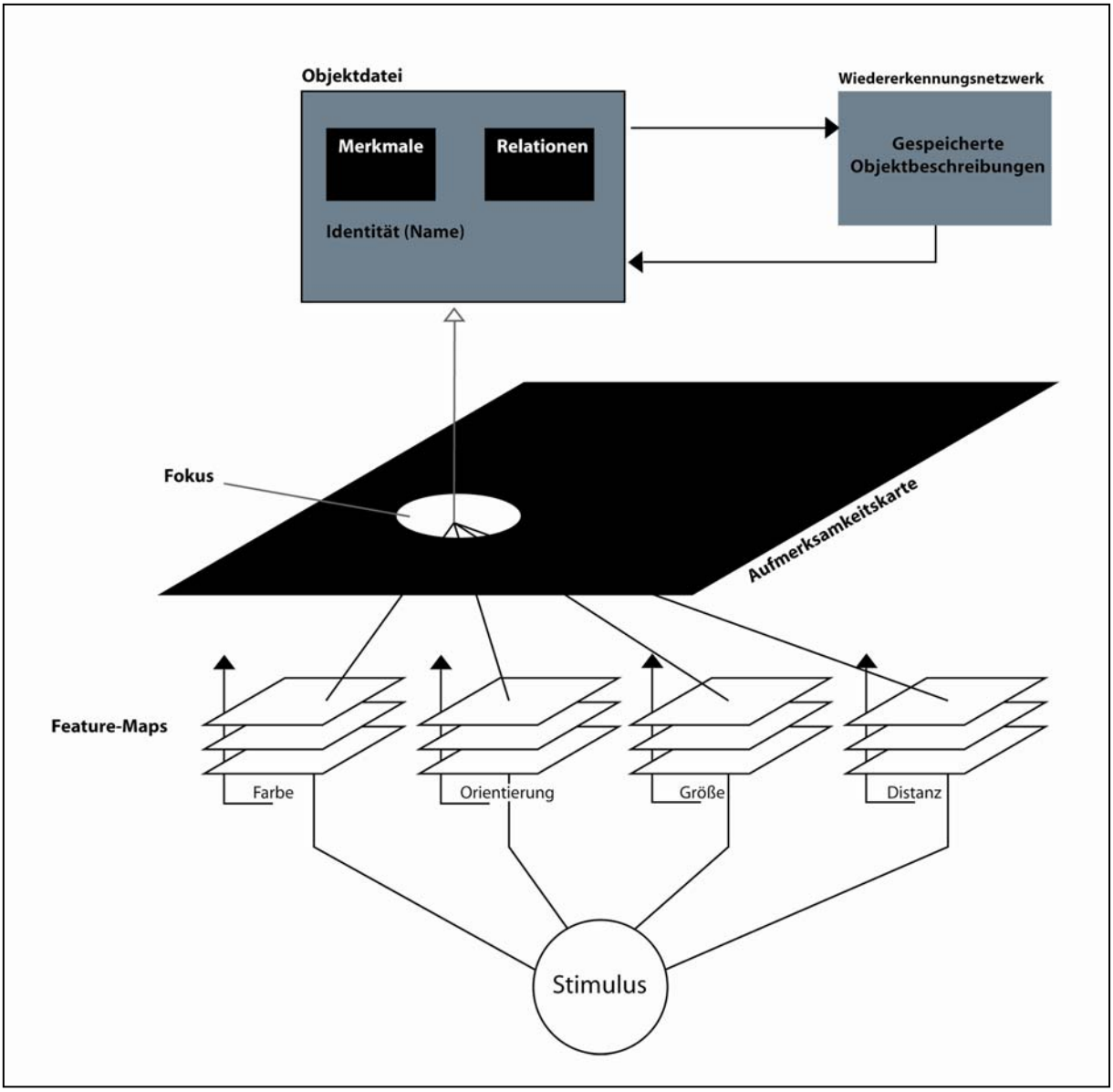

Abb. 9 [1]: Darstellung der FIT

3.3.2 Texturbereichstrennung

Ist Texturbereichstrennung ebenfalls ein paralleler Prozess der visuellen Wahrnehmung, so dürfte seine Bestimmung nur durch räumliche Trennung von Gruppen mit denselben *Features* erreicht werden [2]. Anders ausgedrückt heißt dies, dass ein *Pop-Out*-Effekt hinsichtlich der Textur nur auftreten kann, wenn die jeweils räumlich getrennten Gruppen aus keinen *Conjunctions* bestehen. Abbildung 10 und 11 soll dieses Paradigma verdeutlichen.

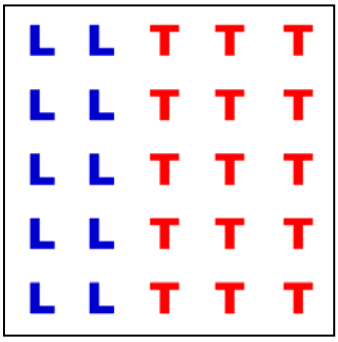

Abb. 10: Pop-Out bei Texturbereichstrennung

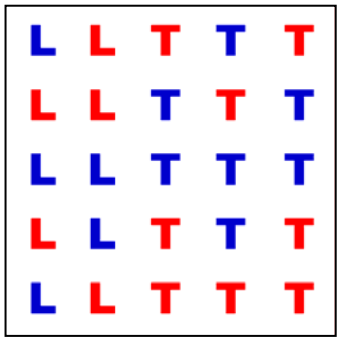

Abb. 11: kein Pop-Out-Effekt

Bei Abbildung 10 springen einem zwei Gruppen ins Auge. Zum Einen die Gruppe mit den blauen L´s und zum Anderen die Gruppe mit den roten T´s. Es kommt zu einem *Pop-Out*-Effekt. Abbildung 11 hingegen führt zu keinem *Pop-Out*-Effekt. Es ist zwar möglich, die L´s sowie die T´s zu Gruppen zusammenzufügen, hierfür ist aber Aufmerksamkeit notwendig.
Betrachtet man den Unterschied zwischen beiden Abbildungen, so deckt sich dieser mit der vorhergesagten Hypothese von Treisman. Bei Abbildung 10 bestehen die Gruppen aus keinen *Conjunctions*, es kommt zum *Pop-Out*. Abbildung 11 hingegen verwendet zusätzlich, innerhalb einer Gruppe, noch die Farbe als Unterscheidungsmerkmal, wodurch innerhalb einer Gruppe *Conjunctions* auftreten und es somit, wie von Treisman vorhergesagt, zu keinem *Pop-Out*-Effekt kommt!

3.3.3 Illusionäre Verbindungen

In der ersten Phase des Wahrnehmungsprozesses kommt es zur Merkmaldetektion. Erst in späteren Phasen werden diese detektierten *Features* kombiniert und zu einem Objekt zusammengefügt. Wird nun gerichtete Aufmerksamkeit verhindert, müssten auch falsche Kombinationen der *Features* und folglich falsch erkannte Objekte möglich sein [2]! Abbildung 12 repräsentiert eine Versuchsanordnung nach Treisman [1], die zum Testen dieses Paradigmas verwendet werden kann.

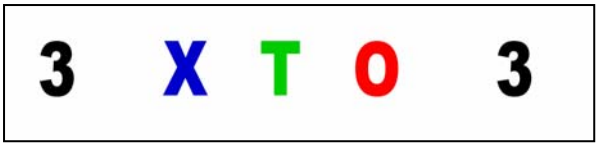

Abb. 12: Versuchsanordnung zum Testen illusionärer Verbindungen

Der Stimulus (Abb. 12) wurde den Versuchspersonen 200msec präsentiert. Aufgabe der Testperson war es dabei, nach Präsentation des Stimulus die gezeigte Ziffer zu nennen und danach die jeweiligen Buchstaben, inklusive deren Farbe. Durch die Ablenkung mit den Ziffern auf jeder Seite und der damit zusätzlich gestellten Aufgabe, sowie aufgrund der kurzen Präsentationsdauer konnte Treisman gerichtete Aufmerksamkeit auf die Buchstaben ausschließen.
Das Ergebnis deckt sich mit der Aussage, dass es zu illusionären *Conjunctions* kommen kann. Die Testpersonen erkannten in einem von drei Fällen die falsche Kombination aus Buchstabe und Farbe, also zum Beispiel ein rotes X, obwohl diese nicht im Stimulus präsent war. Die Versuchspersonen haben also die erkannten Features „rot“ und „X“ falsch kombiniert!

3.3.4 Identifizierung und Ortsbestimmung

Bei der frühen Merkmaldetektion wird noch keine räumliche Information ausgewertet. Diese erhält man erst durch gerichtete Aufmerksamkeit über den Stimulus hinweg! Da aber, laut Theorie, *Conjunctions* durch serielle, der gerichteten Aufmerksamkeit folgende, Wahrnehmung erkannt werden, sollte hier die Ortsinformation bereits gespeichert sein! Dieser Aspekt deckt sich mit dem in Punkt 3.1.2.1 genannten Experiment von Treisman und wurde von ihr somit bewiesen.

4. „Guided Search“

4.1 Abweichende Testergebnisse

Nach der FIT ist die Suche nach *Conjunctions* eine serielle Suche und zeitabhängig von der Anzahl der Objekte im Bild.

Jeremy Wolfe führte mit seinen Kollegen, analog zu Treisman´s Experimenten, Testreihen zur *Conjunction*-Suche durch, bei denen Testpersonen visuelle Suchaufgaben nach Farbe x Form, Farbe x Orientierung sowie Farbe x Größe lösen mussten [3]. Die Ergebnisse sind verwirrend, denn die Reaktionszeit zum Finden des Zielobjektes in Abhängigkeit von der Anzahl der Objekte im Bild weist eine ähnliche Steigung wie bei der Suche nach nur einem *Feature* auf, ist somit unabhängig von der Anzahl der Objekte im Bild! Eigentlich müsste hier, nach Treisman, eine serielle Suche stattfinden, bei der die Kurve mit zunehmender Anzahl der Objekte steigt.
Diese abweichenden Ergebnisse sind allerdings nicht bei allen *Conjunction*-Suchen zu finden. Treisman führte ja bereits visuelle Suchaufgaben nach *Conjunctions* durch (siehe 2.), bei denen die Suchzeit von der Anzahl der Objekte im Bild abhängt!

4.2 Modifikation der FIT

Diese Erkenntnis, dass die visuelle Suche nach bestimmten *Conjunctions* gleiche Ergebnisse wie die Suche nach nur einem *Feature* liefert, fordert bereits eine Änderung des Modells von Treisman.
Anzumerken ist, dass Wolfe in seinem Artikel leider keine Aufzählung der *Conjunctions* liefert, bei denen solche Ergebnisse zu entdecken sind!

4.2.1 Verbesserte gerichtete Aufmerksamkeit

Wolfe erklärt die gefundenen Ergebnisse damit, dass die bisherige strikte Trennung zwischen der ersten, parallelen Phase und folgenden, seriellen Phasen nicht aufrechterhalten werden kann. Ist es ferner nicht denkbar, dass der parallele Prozess Informationen an den seriellen Prozess weitergeben kann?

An diesem Punkt setzt Wolfe auch mit seiner Modifikation der FIT an. In seinem Guided-Search Modell werden vom ersten, parallelen Prozess Informationen an den seriellen Prozess weitergegeben. Die gerichtete Aufmerksamkeit kann somit vom parallelen Prozess gesteuert (deshalb „Guided-Search“) und der Fokus nur auf die Regionen gerichtet werden, welche bereits in der ersten Phase Ergebnisse lieferten.
Die *Feature-Maps* liefern somit Ortsinformationen weiter und die Aufmerksamkeit wird nur noch auf die gefundenen Regionen gerichtet. Abbildung 13 soll die Kombination zweier solcher weitergegebener Ortsinformationen verdeutlichen und die damit eingeschränkte gerichtete Aufmerksamkeit zeigen.

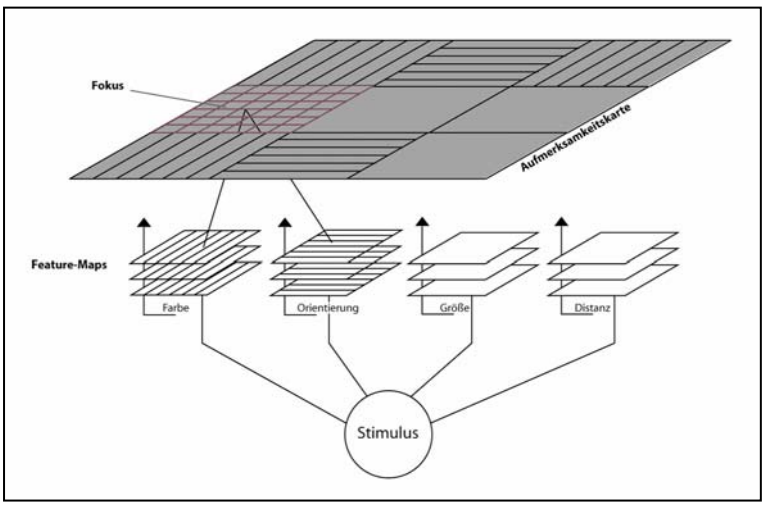

Abb. 13 [3]: Guided-Search Modell zur Conjunction-Suche

Der karierte, rote Bereich wurde sowohl von der *Feature-Map* für Farbe wie auch von der *Feature-Map* für Orientierung im ersten, parallelen Prozess lokalisiert. Der Fokus wird nun nur noch auf diesen Bereich gerichtet, womit die Suche zu einem schnelleren Ergebnis kommt!

4.3 Weitere Modifikation

Das Guided-Search Modell macht Vorhersagen, die qualitativ unterschiedlich von denen der FIT sind. Man könnte effizientere Suchergebnisse erwarten, wenn noch mehr Informationen vom parallelen Prozess im seriellen Prozess verwendet werden könnten [3].

Diesen Aspekt untersuchte Wolfe mit Hilfe eines visuellen Suchexperiments, bei dem 3fach-*Conjunctions* zu finden waren [3]. Die Ergebnisse decken sich mit der Vermutung, noch effizientere Ergebnisse zu erhalten. Die Tatsache, dass drei Informationen vom parallelen an den seriellen Prozess weitergegeben werden können, führt zu schnelleren Suchergebnissen!
Abbildung 14 soll diesen weiteren Aspekt des Guided-Search Modells verdeutlichen.

Analog zu Abbildung 13 werden hier Informationen von drei verschiedenen *Feature-Maps* geliefert und folglich der Fokus noch besser auf eine Region gelenkt.

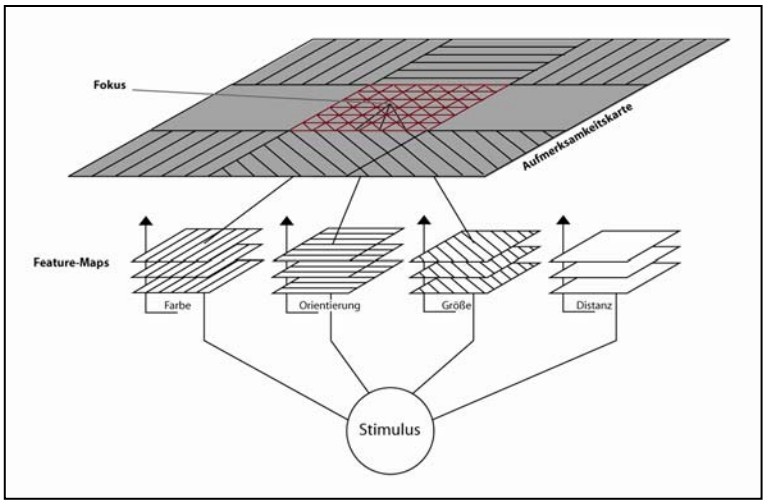

Abb. 14 [3]: Guided-Search Modell zur 3fach-Conjunction-Suche

4.4 *Guided-Search* Modell

Zusammengefasst ist das Guided-Search Modell von Wolfe analog zur Theorie von Treisman, mit der Erweiterung, dass gerichtete Aufmerksamkeit vom ersten parallelen Prozess gesteuert werden kann und somit schneller zu einem Ergebnis führt. Diese Steuerung ist zwar nicht obligatorisch für alle *Conjunctions*, gibt aber eine plausible Erklärung für bisherige, gefundene Ausnahmen (vgl. 4.1) zu Treisman´s Theorie!

5. Computermodelle zur visuellen Wahrnehmung

Bis zum heutigen Tag sind keine großen Neuerungen zu den beschriebenen Theorien hervorgekommen. Alle Untersuchungen bestätigen die bisherigen Annahmen und versuchen, den Bereich der gerichteten Aufmerksamkeit genauer zu untersuchen und diesen besser beschreiben zu können.
Basierend auf den gemachten Erkenntnissen war die Entwicklung eines Computermodells, welches beschreibt, wie Aufmerksamkeit in einer gegebenen visuellen Szenerie eingesetzt wird, eine wichtige Herausforderung in der Neuroinformatik. Die mögliche Anwendung solcher Modelle um Aufgaben wie Überwachung, automatische Zielerkennung, Navigationshilfen oder die Kontrolle von Robotern zu lösen, diente hierbei als Motivation [5].

5.1 Trennung *Bottom-Up* und *Top-Down* Modell

Bei der Umsetzung von Computermodellen zur Aufmerksamkeit ist zwischen *Bottom-Up* und *Top-Down* Modellen zu unterscheiden.
Bottom-Up Modelle beschreiben, wie Aufmerksamkeit nach der Präsentation des Stimulus eingesetzt wird. Dies beschränkt sich aber auf die ersten paar hundert Millisekunden nach Präsentation des Stimulus.
Ganzheitlichere Modelle zur Aufmerksamkeitskontrolle müssen einen *Top-Down* Prozess beinhalten. Die Herausforderung zur Umsetzung eines ganzheitlicheren Computermodells liegt dann in der Integration von beiden Prozessen [5].

5.2 Salienzkarte

Ein Aspekt, den fast alle bisherigen *Bottom-Up* Modelle ähnlich umsetzen, ist das Problem, wie gerichtete Aufmerksamkeit nach dem ersten, parallelen Prozess zur Merkmaldetektion geführt werden kann?
Die Modelle verwenden hierzu eine Salienzkarte. Salienz, die Tatsache, sich von anderen Merkmalen / Eigenschaften hervorzuheben, dient hierbei als Wegweiser für die gerichtete Aufmerksamkeit.

Das Guided-Search Modell von Wolfe zeigte uns bereits, dass *Feature-Maps* Ortsinformationen an den späteren Prozess weitergeben können und somit die Aufmerksamkeit zu steuern vermögen. Diesen Aspekt nutzt eine Salienzkarte. Alle erhaltenen Ortsinformationen, wo sich Salienz von *Features* im Stimulus befindet, werden in einer einzelnen Karte gesammelt und zusammengetragen. Eine Salienzkarte enthält demnach Bereiche unterschiedlicher, absteigender Aktivität, nach denen die Aufmerksamkeit gerichtet und gesteuert werden kann.
Anzumerken ist an dieser Stelle, dass inzwischen neuronale Analogien zur Salienzkarte im visuellen System gefunden wurden [5]!

5.3 Fünf Rahmenpunkte für *Bottom-Up* Computermodelle der Aufmerksamkeit

Laurent Itti beschreibt in seinen Ausarbeitungen [4] & [5] verschiedene Ansätze umgesetzter Computermodelle. Dabei haben sich fünf Rahmenpunkte herauskristallisiert, die von einem *Bottom-Up* Computermodell umgesetzt werden müssen und im Folgenden beschrieben sind.

5.3.1 Punkt 1: pre-attentive Aspekt

Analog zur Theorie von Treisman werden zuerst bestimmte *Features* über den gesamten Stimulus hinweg extrahiert.
Die Implementierung dieser Detektion im Computer wird dabei oft durch eine Imitation von biologischen Eigenschaften umgesetzt. So kann zum Beispiel das Antwortverhalten einer Nervenzelle, welche besonders auf einen Helligkeitskontrast reagiert, durch eine Filterung mit einem Difference-of-Gaussian[2] Filter auf dem Helligkeitskanal des Eingangsbild simuliert und berechnet werden [5].

5.3.2 Punkt 2: Salienzkarte

Ausgehend von den detektierten *Features* wird eine Salienzkarte generiert, welche topographisch differenzierte Aktivitäten von Merkmalen enthält.

[2] Das zu untersuchende Bild wird mit zwei Gaussfiltern mit unterschiedlicher Standardabweichung geglättet. Danach wird von den beiden erhaltenen Hilfsbildern die Differenz berechnet.

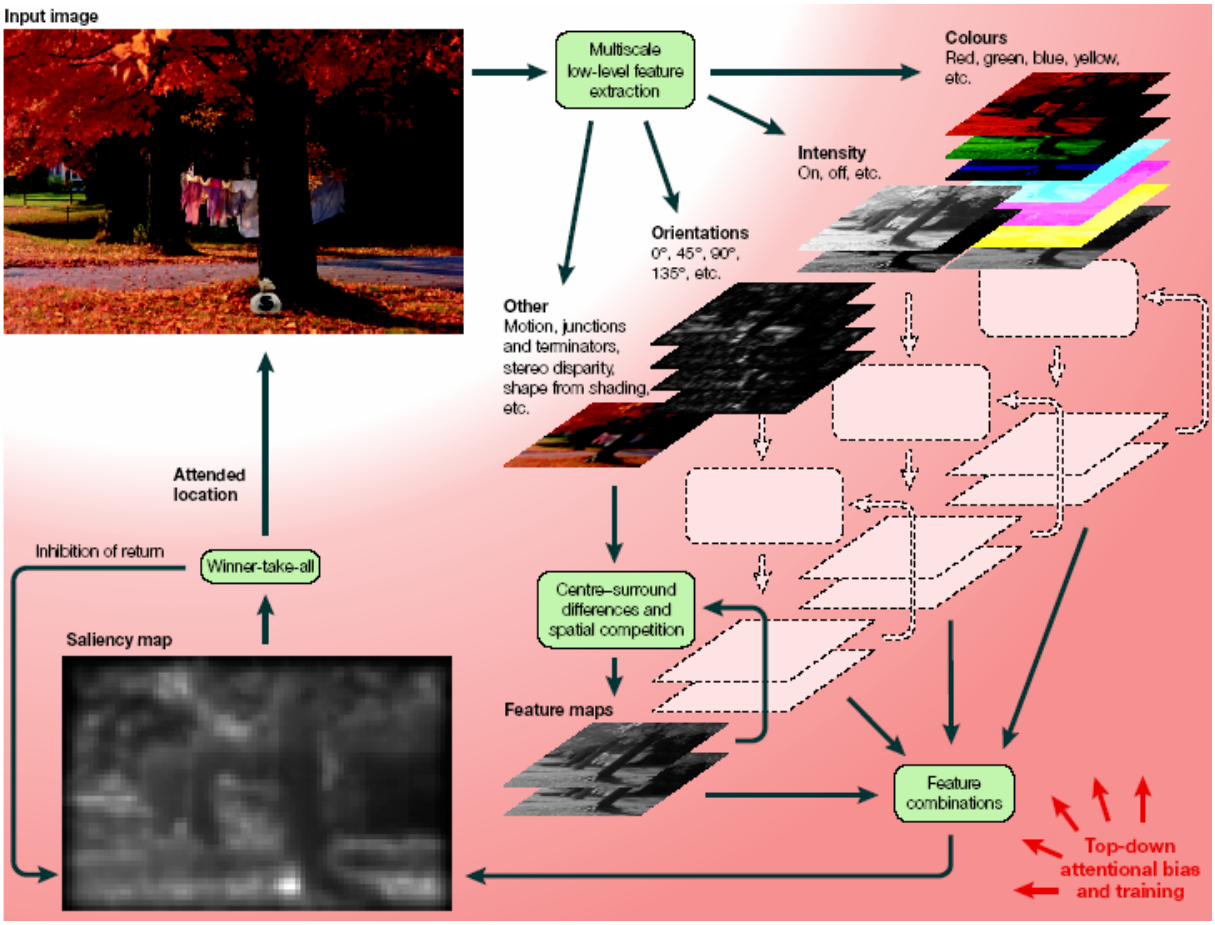

Abb. 15 [5]: Modell eines Bottom-Up-Computermodells

Abbildung 15 soll die Generierung einer Salienzkarte verdeutlichen. Analog zur FIT von Treisman werden bestimmte *Features* detektiert (hier: Colours, Intensity, Orientations, Other). Die damit erhaltenen räumlichen Informationen werden zusammen mit den detektierten Merkmalen in einer einzigen Karte, der Salienzkarte (hier: Saliency Map) zusammengetragen. Die unterschiedlichen Helligkeiten auf dieser Karte sollen Regionen mit unterschiedlich starker Aktivität kennzeichnen.

5.3.3 Punkt 3: „Attentional Scanpath“ und „IOR“

In dieser Stufe wird ein Suchweg für die Aufmerksamkeit generiert. Dabei wird entsprechend der Salienzkarte ein Weg mit absteigender Aktivität an Salienz erzeugt. Das bedeutet, der Ort mit größter Aktivität wird als Erster aufmerksam betrachtet, daraufhin der Ort mit zweitgrößter Aktivität und so fort. Hier liegt auch der Unterschied der meisten Modelle, wie nämlich die erhaltenen Informationen aus dem ersten, parallelen Prozess beschnitten und zu einer gemeinsamen Salienzkarte zusammengefasst werden [5].

Ein Problem tritt in Zusammenhang mit der Anwendung einer Salienzkarte allerdings auf. Wie kann man verhindern, dass immer der Ort mit höchster Aktivität betrachtet wird? Hier wird ein Mechanismus verwendet, der bereits in der Psychophysik[3] erforscht wurde und „Inhibition-Of-Return (IOR)“ genannt wird. Dieser verhindert, dass bereits betrachtete Regionen nochmals betrachtet werden.

[3] Die Psychophysik beschäftigt sich mit den Zusammenhängen zwischen physikalischen Reizen und den Empfindungen, die sie beim Menschen auslösen. Dies kann sich auf alle Wahrnehmungsbereiche beziehen: Psychophysik wird sowohl mit optischen, akustischen, als auch mit Geruchs- und Geschmacksreizen und mit allen Arten der Körperwahrnehmung betrieben. Entsprechend der verschiedenen Wahrnehmungsarten gibt es verschiedene Untergruppen der Psychophysik, wie etwa die Psychoakustik [6].

Auf den Computer angewendet würde dies eine Implementierung eines Kurzzeitspeichers bedeuten, in welchem man sich die bereits betrachteten Lokationen speichert.

5.3.4 Punkt 4: Interaktion von offensichtlicher und verborgener Aufmerksamkeit

Verborgene Aufmerksamkeit, also bewusste Wahrnehmung ohne Augenbewegung, ist bereits experimentell erforscht [5]. Für die Objekterkennung ist gerichtete Aufmerksamkeit in Verbindung mit dieser verborgenen Aufmerksamkeit notwendig. Diesen Aspekt muss auch ein Computermodell umsetzen und somit eine Interaktion dieser beiden Aufmerksamkeitsmechanismen implementieren und simulieren können.

5.3.5 Punkt 5: Gerichtete Aufmerksamkeit durch Szenenverständnis

Aufmerksamkeit kann durch Vorwissen beeinflusst werden. Stellen wir uns ein Bild von einem Flugzeug auf der Landebahn vor, in welchem wir eine Person finden möchten. Haben wir in einem frühen Aufmerksamkeitspunkt (z.B. links, mitte) bereits die Schnauze des Flugzeuges identifiziert, brauchen wir unseren Blick nicht weiter nach rechts schweifen lassen, da wir hier den Rumpf des Flugzeuges vermuten. Vielmehr werden wir unsere Aufmerksamkeit auf Punkte richten, in denen eine Person überhaupt vorkommen kann! Durch unser Vorwissen konnten wir die Person somit schneller finden. Dieses Beispiel soll zeigen, wie eng verbunden die Steuerung der Aufmerksamkeit mit Szenenverständnis und Vorwissen ist und dass deshalb auch ein Computermodell diese Prozesse simulieren können sollte.

Die in Punkt 5 gemachten Aussagen sollten als kurze Aufzählung dienen, welche Aspekte ein Computermodell zur visuellen Wahrnehmung und Aufmerksamkeit beachten und simulieren können sollte.

Literaturverzeichnis

[1] A. Treisman: Features and Objects in Visual Processing (1986)
Scientific American 255, pages 106 – 115

[2] A. Treisman, G. Gelade: A Feature-Integration Theory of Attention (1980)
Cognitive Psychology 12, pages 97 - 136

[3] J. Wolfe, K. Cave, S. Franzel: Guided Search: An Alternative to the Feature Integration Model for Visual Search (1989)
Journal of Experimental Psychology in Human Perception an Performance 15, pages 419 - 433

[4] L. Itti: Visual Attention (2003)
MA Arbib (ed.) The Handbook of Brain Theory and Neural Networks, 2nd Edition. MIT Press, Cambridge, MA, pages 1196 - 1201

[5] L. Itti, Ch. Koch: Computational modelling of visual attention (2001)
Nature Review Neuroscience 2, pages 1 – 11

[6] Wikipedia, die freie Enzyklopädie
http://de.wikipedia.org/wiki/Psychophysik; Download: 15.06.2004