

From the Human Eye to Computer Vision

Einführung in das wissenschaftliche Arbeiten

Betreuer Dipl.-Ing. Mauthner Thomas

Sprache: Deutsch

Lukas Prokop # 1031367

Andreas Fauler # 1031225

{lukas.prokop,fauler} @ student.tugraz.at

13. Dezember 2010

Inhaltsverzeichnis

1	Biologische Aspekte des Auges	2
1.1	Der anatomische Aufbau des Auges	3
1.2	Die Retina als bildgebender Sensor	4
1.3	Das neurowissenschaftliche Modell des Retinaaufbaus	4
2	Von der Retina zur Bildverarbeitung	6
2.1	Das Brodmann-Areal 17, die Area striata oder der V1	6
2.2	Die Reaktionen des V1 auf visuelle Stimulation am Beispiel eines Experiments	9
3	Die Bionische Umsetzung	10
3.1	Offene Fragen	10
3.2	Verarbeitungsgeschwindigkeit	10
3.3	Die technische Implementierung in Bildsensoren	11
3.4	Der Ausblick	12

Zusammenfassung

In dieser Arbeit setzen wir uns mit dem Auge als Medium für die visuelle Wahrnehmung auseinander. Der Mensch nimmt visuelle Eindrücke auf, vergleicht sie unwillkürlich mit Erfahrungswerten und trifft Entscheidungen auf Basis derer. Der Bereich der *computer vision* ist daran interessiert das Auge technisch nachzubauen, um die digital-visuelle Wahrnehmung zu ermöglichen. Dieser Bericht setzt sich mit der Funktionsweise der Bildverarbeitung im menschlichen Körper von der Hornhaut über die Netzhaut bis zum Visuellen Cortex auseinander, um Ausblick auf technische Nachbauten zu geben. Kameratechnologien und *smartsensors* sind heute Teil unseres technischen Alltags und implementierte Bildverarbeitungsalgorithmen und Sensortechniken basieren auf dem natürlichen Vorbild.

1 Biologische Aspekte des Auges

Das Auge ist jenes Sinnesorgan der Lebewesen, welches für die visuelle Wahrnehmung und damit Interpretation von Lichtsignalen zuständig ist. Das menschliche Auge kann Lichtwellen im Bereich von 350 nm bis zu 750 nm wahrnehmen. Diese Signale werden an der Netzhaut gespiegelt abgebildet, von Stäbchen und Zapfen in elektromagnetische Reize umgewandelt und im Gehirn verarbeitet. Diese Informationen nutzt der Mensch, um seine Umgebung zu beobachten und damit überlebensfähig zu bleiben [2].

Die Augen liegen symmetrisch in den Augenhöhlen beider Gesichtshälften, wodurch das menschliche Augenpaar frontal angeordnet ist, wie es bei allen Jägern in der Natur vorkommt. Es ist an die verschiedenen natürlichen Lebensweisen des Menschen angepasst. So kann ein gesunder Mensch sehr gut Farbsehen, bewegliche Objekte verfolgen und durch kontrastreiches Sehen Konturen und Helligkeiten unterscheiden. Durch die Elastizität der Linse und das Anspannen der *Corpus ciliare* (Zonulafasern) kann er nahe und ferne Objekte fokussieren, um sie scharf erkennen zu können. Diese Anpassung der Linse an die Entfernung nennt sich *Akkommodation*. Jedoch ist der Mensch nicht in der Lage (wie es bei Vögeln der Fall ist), Objekte in einem beschränkten Bereich der *Retina* (Netzhaut) stark vergrößert zu sehen. Das Auge des Menschen ist wie die meisten Wirbeltiere als physiologisch hoch entwickelt einzustufen [1, S. 58].

Diese biologischen Gegebenheiten sind jedoch nicht mehr an unsere heutigen Lebensgewohnheiten angepasst. Wir beschränken uns im Alltag auf das Fokussieren von nahen Objekten (Zeitung lesen, Arbeit mit Bildschirmen), weshalb es oft zur Volkskrankheit Sehschwäche in Form von Kurz- oder Weitsichtigkeit kommt.

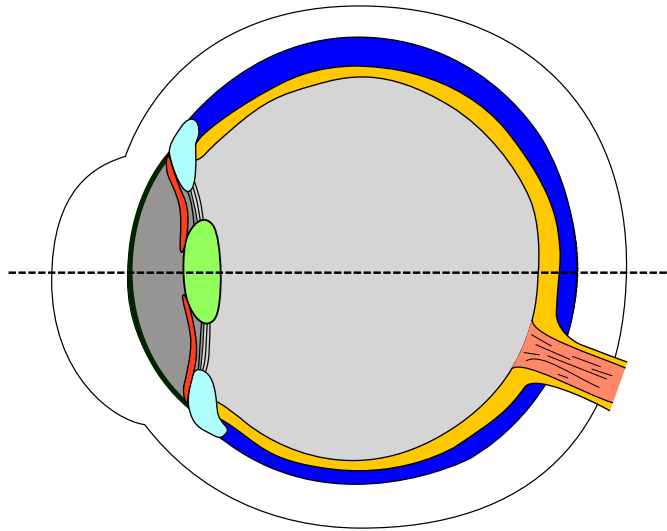


Abbildung 1: *Aufbau des menschlichen Auges*. Bestehend aus Lederhaut (weiß), Aderhaut (blau), Netzhaut (orange), Sehnerv (salmon), Ziliarmuskeln (türkis), Zonulafasern (schwarze Linien), Linse (grün), Iris (rot), Hornhaut (dunkelgrün), hintere Augenkammer (hellgrau), vordere Augenkammer (dunkelgrau)

1.1 Der anatomische Aufbau des Auges

Das Auge befindet sich in der Augenhöhle des Skeletts und wird von einer undurchsichtigen, weißlichen Lederhaut umgeben. Das Licht dringt entlang der Sehachse durch die Hornhaut. Zwischen Hornhaut und Linse befindet sich die vordere Augenkammer. Die Linse wird abgedeckt durch die Regenbogenhaut (Iris) und die Pupille. Die Linse funktioniert wie ein Fotoapparat und bricht die Lichtstrahlen, sodass sie vertikal gespiegelt auf der Netzhaut abgebildet werden.

Die Sehachse trifft an der Netzhaut auf den Gelben Fleck; der ein Zentrum scharfen Sehens bildet. Umgekehrt wird der Bereich an dem der Sehnerv auf die Netzhaut trifft, als Blinder Fleck bezeichnet, da der Mensch an dieser Stelle nichts erkennen kann. Allerdings bemerkt ein gesunder Mensch diesen Fleck nicht, da das zweite Auge diesen Fleck ausgleicht und das Gehirn fehlende Flächen im Bild mit künstlichen Farben ergänzt.

Schützen kann sich das Auge durch das Zusammenziehen der Iris. Dadurch fällt weniger Licht in den Glaskörper (Innere des Auges). Gegen physikalische Gefahren können sich die Lider ausfallen und damit die gesamte Augenoberfläche abdecken. Die Lider haben auch die Funktion die Hornhaut feucht zu halten, damit das Auge durch die großen Augenmuskeln beliebig gedreht werden kann ohne Reibungserscheinungen auftreten zu lassen. Die Tränenflüssigkeit wird in den Tränenrüsen gebildet und reinigt das Auge, wenn Partikel unter das Augenlid kommen [1].

1.2 Die Retina als bildgebender Sensor

Die Retina (Netzhaut) besteht grob unterteilt aus drei Schichten von Nervenzellen. In der obersten Schicht befinden sich lichtempfindliche Zellen, sg. Fotorezeptoren [6].

Trifft das Licht auf die Netzhaut, wird es von diesen Rezeptoren interpretiert. Die fotorezeptiven Zellen umfassen die Zäpfchen und Stäbchen sowie weitere Zellarten. Interessant sind besonders die kegelförmigen Zäpfchen und zylinderähnlichen Stäbchen. Während erstere für das Farb- und das detailreiche, scharfe Sehen zuständig sind, empfangen zweitere insbesondere schwache, reduzierte Signale. So sind die Stäbchen für das Sehen bei geringer Helligkeit und in den Randlagen der Netzhaut zuständig. Die Zäpfchen umfassen drei verschiedene Zapfenarten: rot-, grün und blauempfindliche. Folglich gibt es Zäpfchen die auf Licht verschiedener Wellenlängen reagieren (rot repräsentiert sehr langwelliges Licht, während blauviolett kurzwellig ist). Durch das Vorhandensein dieser drei Typen kann das komplette Farbspektrum abgebildet werden (RGB-Farbmodell, Abb. 2). Die Stäbchen & Zäpfchen verursachen eine Teilung der Informationen Farben und Helligkeit. Dabei handelt es sich um ein Konzept, welches 1:1 in die Bionik übernommen wurde.

Der Unterschied zwischen den verschiedenen Aufgabengebieten (scharfes bzw. peripheres Sehen) ergibt sich teilweise durch die Ungleichverteilung der Rezeptorzellen auf der Netzhaut. So befinden sich im Zentrum – beim Gelben Fleck – viele Zäpfchen dicht bei einander, während in der Peripherie die Dichte der Zellen (hauptsächlich Stäbchen) wesentlich kleiner ist. Am Gelben Fleck kann man etwa 50 000 Ganglienzellen pro Quadratmillimeter beobachten, während sich in der Peripherie nur etwa 1000 Zellen pro Quadratmillimeter finden lassen. Unter der Annahme, dass der Radius des Augapfels eines ausgewachsenen Menschen etwa 12mm beträgt und der Gelbe Fleck etwa 2° in Anspruch nimmt, so beträgt die Fläche des Gelben Flecks (unter Vernachlässigung der Krümmung) etwa $1.7013 \cdot 10^{-5} \text{ mm}^2$. So ergibt sich zwar ein kleineres scharfes Sichtfeld, die Menge an eingehender Information und damit der Aufwand bei der Verarbeitung ist aber wesentlich geringer. Aus dem Grund der besseren Sichtbarkeit entlang der Sehachse bewegt der Mensch seinen Kopf jeweils in die für ihn interessante (bzw. aufmerksamkeitsregende) Richtung, wobei das Sehfeld des Menschen etwa 130° vertikal und 180° horizontal umfasst.

1.3 Das neurowissenschaftliche Modell des Retinaaufbaus

Betrachten wir die Retina im Detail, gibt es verschiedene Modelle, um die Interaktion der Synapsen¹ zu beschreiben. Im Folgenden sei das neurowissenschaftliche Modell erklärt, welches sich an der Aufgabenverteilung der Zellen orientiert.

¹Kontaktschnittstelle zwischen 2 Zellen

Die oberste Schicht von Nervenzellen leitet die Signale an die darunterliegende Schicht weiter. In dieser befinden sich bipolare Zellen, horizontale Zellen und auch Amakrin-Zellen (Abb. 3). Aufgrund der – verglichen mit den Fotorezeptorzellen – geringeren Anzahl dieser Zellen ist eine Zelle für mehrere, verschiedene Rezeptorzellen der oberen Ebene "zuständig". Daraus ergibt sich für jede Zelle ein sg. rezeptives Feld, also ein Bereich aus dem diese Zelle Signale erhält. Die bipolaren und horizontalen Zellen haben entgegengesetzte Aufgaben. Während die bipolaren Zellen bei einem Reiz im Zentrum des rezeptiven Feldes Signale weiterleiten, übernehmen die horizontalen Zellen Reize die am Rand ihres zugeordneten Feldes auftreten. Die Amakrin-Zellen [6, S. 2] steuern den Umfang der rezeptiven Felder und die Stärke der Reaktion auf Reize innerhalb dieser. Sie sind wesentlich für Aufgaben wie Bewegungserkennung, Konturenwahrnehmung und Helligkeitsanpassung. Man hat 22 verschiedene Unterarten von Amakrin-Zellen entdeckt. Zwischen all diesen Zellarten gibt es über die Ebenen synaptische Verbindungen, die Datenströme erlauben [4].

Die unterste Schicht bilden Ganglien-Zellen. Sie sind die Schnittstelle zwischen der Retina und dem Gehirn. Hier werden die Signale schlussendlich über den Sehnerv an das Gehirn weitergereicht. Nachdem das Verhältnis von Fotorezeptoren und Ganglienzellen über 100:1 beträgt [6, S. 4] müssen die Informationen vor dem Transport noch "komprimiert" werden, was teilweise schon in der zweiten Schicht geschehen ist. Grundsätzlich unterscheidet man zwei² Typen von Ganglienzellen: P-Zellen (für parvozellulär) und M-Zellen (magnozellulär), [6, S. 4, S. 1] wobei die P-Zellen 90% aller Ganglien-Zellen ausmachen und vor allem im Zentrum der Retina vorkommen. Auch die Ganglien haben rezeptive Felder, das heißt auch hier ist eine Zelle für viele Rezeptoren "zuständig" (das ergibt sich schon aus der großen Überzahl an Rezeptorzellen). Die rezeptiven Felder sind im Gelben Fleck sehr klein, d.h. die Ganglien sind nur für sehr wenige (bzw. einzelne) Rezeptorzellen zuständig, während nach außen hin die rezeptiven Felder groß werden. Außerdem haben M-Zellen stets größere rezeptive Felder als P-Zellen.

Von den P- und M-Zellen gehen korrespondierende P- und M-Leitungen aus, über welche die Informationen schließlich über Nervenbahnen in das Gehirn transportiert werden. Im Sinne des neurowissenschaftlichen Modells haben die P-Zellen das Ziel, möglichst verlustfrei zu arbeiten und somit möglichst *viele* Daten an das Gehirn weiterzuleiten. Deshalb kommt es zur besagten Komprimierung der Daten. Umgekehrt haben M-Zellen das Bestreben die Daten möglichst *schnell* an das Gehirn weiterzuleiten [4].

²Man nimmt an, dass es etwa 12-15 verschiedene Subtypen von Ganglienzellen gibt

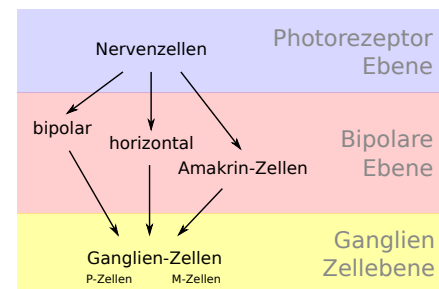


Abbildung 3: Die Ebenen der vertikalen Signalleitung. Symbolisch würden unter der Ganglien-Zellebene Nervenbahnen zum Gehirn liegen.

2 Von der Retina zur Bildverarbeitung

Visuelle Reize werden im hinteren Teil des Großhirns verarbeitet; im sogenannten Okzipitallappen. Dieser besitzt das primäre und sekundäre Sehzentrum, die gemeinsam den Visuellen Cortex (VC) bilden. Während sich das primäre Sehzentrum um die Aufnahme der Informationen (sowohl jeder Bildpunkt wird interpretiert als auch Linien und Konturen aus dem Gesamtbild separat gefiltert) kümmert, so übernimmt das sekundäre Sehzentrum die Aufgabe des Abgleichs mit Erfahrungswerten und Erinnerungen. Allgemein bekannt ist, dass es zu einer kontralokalen Relation kommt: Die *linke* Gesichtshälfte nimmt Informationen für die *rechte* Gehirnhälfte auf.

Hirnforscher teilen den visuellen Cortex in 5 Bereiche auf. Während der Begriff des primären Sehsystems mit dem primären Bereich (V1) übereinstimmt, ist das sekundäre Sehsystem in V2–V5 aufgeteilt.

Erwähnt sei hier noch der LGN. Der *Lateral geniculate nucleus* (Metathalamus) ist ein subkortikaler Vorbereich des Visuellen Cortex, der eine Filterfunktion besitzt und die Aufgabe hat, Informationen an den geeigneten Bereich des VCs weiterzuleiten. Allerdings werden die meisten Daten direkt an den V1 gesandt. Er empfängt die Daten der Retina von der Sehbahn (Nervenbahnen vom primären zum sekundären Sehsystem) und reicht sie an den V1 weiter. Der LGN besteht aus sechsfach geschichteten Zellschichten, wobei M- und P-Zellen im Verhältnis 2:4 auf diesen Schichten aufgetragen sind. Die M- und P-Zellen unterscheiden sich in der oben erwähnten räumlichen Dichte und dem Aufbau ihrer Fortsätze (Dendriten). Die P-Zellen besitzen kleine Dendrite und damit haben sie ein hohes Auflösungsvermögen. Sie erlauben das Farbsehen. Die M-Zellen hingegen bilden wieder einen Gegenpol der Bilddaten von der Retina und verarbeiten mit ihrem wesentlich größeren rezeptiven Feld Helligkeitsstufen [4].

Wir erwähnten, dass die Retina mit einem RGB-Farbmodell arbeitet. Dieses zieht sich bis zum LGN vollständig durch (Abb. 5). Allerdings gibt es auch einen Hemmungsfaktor für das RGB-Farbmodell. Wird eine Zelle durch die Farbe rot erregt, kann diese durch grünes Licht gehemmt werden. Selbiges gilt für die Farben Blau zu Gelb. Welche Relevanz dieses Gegenfarbmuster für die Verarbeitung und Weiterleitung der Sinnesempfindungen hat, ist noch nicht eindeutig geklärt.

Bei einem erheblichen Schaden der M-Zell-Schichten im LGN lassen sich klare Beeinträchtigungen im Bewegungssehen beobachten, während ein Schaden der P-Zell-Schichten eine zerstörte Tiefen-, Struktur- und Farbwahrnehmung hervorrufen [9].

2.1 Das Brodmann-Areal 17, die Area striata oder der V1

Beim V1 handelt es sich um den best erforschtesten Teil des VC. Die *fMRI*-Technologie (functional magnetic resonance imaging) (eine Unterklasse der MRIs) hat hier wesentliche

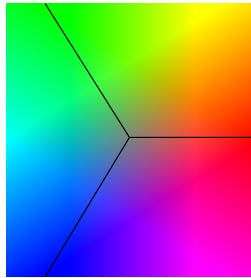


Abbildung 2: Das RGB Farbspektrum. Entlang der schwarzen Linien befinden sich die drei additiven Farben Rot Grün Blau. Weitere Farben ergeben sich durch das Mischen dieser drei Grundfarben.

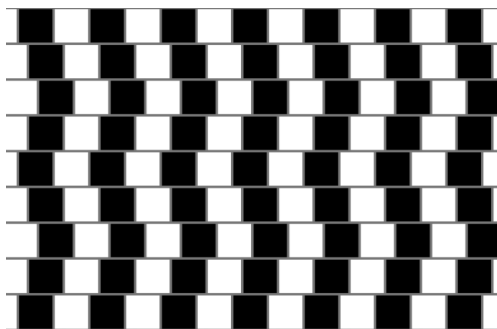


Abbildung 4: Beispiel einer optischen Täuschung. Es kommt zu einer Fehlinterpretation der Daten im sekundären Sehzentrum, da das Gehirn gewohnt ist, versetzte Objekte in eine Tiefenrelation zu setzen.

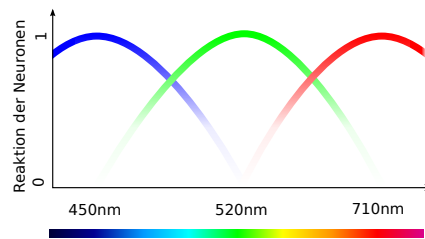


Abbildung 5: Die Grafik illustriert das Reaktionsverhalten der Neuronen auf RGB-Farben. Wird grüne Farbe empfangen, werden nur Neuronen der Farbe Grün aktiviert. Wird Türkis empfangen, werden sowohl grüne als auch blaue Neuronen erregt.

Beiträge geleistet. Sie erlaubt einen visuellen Einblick in die neuronalen Aktivitäten im Verhältnis zum zeitgleichen Blutfluss. Letzterer spielt eine wesentliche Rolle bei der funktionellen Arbeit der Neuronen, da bei neuronaler Reaktion die Zelle sofort Glukose anfordert, um den entsprechenden Energiebedarf zu befriedigen und regt damit auch den Blutstrom an [12].

Im V1 werden die Informationen vom LGN empfangen. Man vermutet, dass die Informationen einer Verarbeitung hierarchischer Ordnung unterliegen. Man kann generell festlegen, dass die Bilddaten von V1 bis V5 weitergereicht werden und stufenweise verarbeitet werden. Dies lässt sich beobachten, weil das rezeptive Feld Ebene für Ebene immer größer wird, sodass die Bildinformationen in den höheren Ebenen kontextsensitiver verarbeitet werden können als im V1. Grundsätzlich gibt es 2 Datenströme, die die Bilddaten in die höheren kortikalen Bereiche bringen: der vorderseitige Datenstrom ($V1 \rightarrow V2 \rightarrow V5 \rightarrow$ Parietallappen) und der rückseitige Strom ($V1 \rightarrow V2 \rightarrow V3 \rightarrow V4 \rightarrow$ Temporallappen). Es ist allerdings sehr schwer diese beiden Ströme entlang des Visuellen Cortex zu verfolgen, da Daten oftmals über getrennte, kortikale Kanäle übertragen werden (als Beispiel sei hier die Farb- & Helligkeitsseparation erwähnt). Zu diesem Ergebnis kamen auch Yoshinobu Goto, Takao Yamasaki und Shozo Tobimatsu in ihrer Arbeit [6] in der sie zusammenfassend erwähnen, dass die physikalische Natur von visuellen Stimulationen von mehreren, separaten und parallelen Leitungen auf mehreren Seiten im Sehsystem getragen wird.

Gewisse Neuronenklassen im V1 reagieren bevorzugt auf einfache Reize wie Linien einheitlicher Richtung in einem beschränkten Gesichtsfeld. Umgekehrt reagieren Neuronenklassen des V5 auf Gesamtbilder wie Gesichter und Hände unabhängig von ihrer Position im Gesichtsfeld und ihrer Drehung.

Die Neuronen im LGN und V1 sind anatomisch und funktional in 2 Bereiche getrennt, die separat Form (Objekterkennungsaufgaben) und Bewegung (Objektortung – Bewegungsrichtung und räumliche Position) verarbeiten, obwohl jeweils Verbindungsstellen zwischen den beiden Bereichen existieren [7, S. 2]. Im V1 konnte 3 verschiedene Arten von Neuronen nachgewiesen werden: einfache Neuronen, komplexe Neuronen und hyperkomplexe Neuronen [5], die sich alle im Detail für unterschiedliche Bilddaten interessieren. Genauso kann man aber

eine Einteilung der Neuronen in die 3 Gruppen vornehmen, die sensitiv für Farben, die Position des Objekts oder eine Mischung aus beiden sind.

Die Daten werden im V1 *retinotop* abgebildet. Die Retinotopie bezeichnet den Effekt der neuronalen Reaktion beim Betrachten einer Form: Betrachtet man eine Form, wird eine Neuronenaktivität im VC hervorgerufen, die die selbe Form hat. Retinotopie in dieser Art kommt bei allen Primaten vor.

Die neuronalen Netze im V1 unterscheiden sich durch Erwartungswerte bezüglich der Eingabeparameter, ihrer internen Verschaltungen (ähnlich Logikgattern) und ihrem Antwortverhalten. Zum Beispiel kann es Unterschiede geben wie Antworten aufgebaut sind (fixe Größen gegenüber einem Datenstrom). Das selbe Verhalten wurde 1:1 für *FIR*-Filter (finite impulse response) übernommen.

Weil sich die enorme Dichte an Informationen an der enormen Dichte an neuronalen Netzen widerspiegelt, wird der V1 illustrativ auch als "*Sortierbüro*" (sorting office) bezeichnet.

2.2 Die Reaktionen des V1 auf visuelle Stimulation am Beispiel eines Experiments

2004 wurden Experimente an den japanischen Universitäten Tokai University und Waseda University durchgeführt, wie das Gehirn auf Reize reagiert. Die Reaktionen des V1 wurden mit der fMRI-Technik aufgezeichnet [12].

2.2.1 Ein Quadrat als Reiz

Die Versuchsperson befindet sich in einem abgeschotteten Raum und wird mit visuellen Reizen auseinander gesetzt. Sie sieht frontal auf einen Bildschirm. Auch wenn die Forscher auch mit auditiven Reizen experimentieren, beschäftigt sich das erste Experiment nur mit Eindrücke im Sehsystem. Ein quadratisches Muster wächst um die Größe von 110 % alle 750 ms an. Diese Größenveränderung geschieht 2mal pro Versuchsdurchlauf.

Ergebnis: Der Visuelle Cortex wurde bei allen Versuchspersonen vollständig aktiviert. Wie erwartet konnte die Größe des Musters direkt an der neuronalen Aktivität abgelesen werden (die Neuronenaktivität ist direkt proportional zur Größe des Musters). Die Bereiche des Temporallappens wurden ebenso von den visuellen Eindrücken stimuliert. Allerdings war der VC wesentlich stärker belastet als der Temporallappen. Bereiche des nahegelegenen Hörsystems wurden nicht genutzt.

Die untersuchenden Physiologen kommen zu dem Schluss, dass dynamische Reize (sowohl sich visuell verändernde Reize als auch visuelle und Audioreize, die versetzt auftreten) die einzelnen Cortex stärker aktivieren als statische Wahrnehmungen. Aus diesem Experiment geht hervor, dass das Gehirn für unveränderte Informationen nur geringfügige Kapazitäten

vergibt.

Des Weiteren interessiert die Wissenschaftler die Korrelation zwischen verschiedenen Reizen unterschiedlicher Quellen. Als Beispiel sei erwähnt, dass wir auf Basis von einer gesehenen Form und einem erfahrenen Geschmack diese Form auf den Geschmack abbilden können und diese beiden Reize zukünftig assoziieren. Welche physiologischen und psychologischen Modelle darauf gründen, ist noch ungewiss [12].

3 Die Bionische Umsetzung

3.1 Offene Fragen

Es sind noch offene Fragen in der Physiologie des menschlichen Auges vorhanden. So gilt der V1 als gut erforscht, doch über die höheren Ebenen des Visuellen Cortex ist wenig bekannt. Genauso wurde in dieser Arbeit verschiedene doppelseitige Datenströme vorgestellt (Helligkeitsfaktor und RGB-Farben, vorderseitiger und rückseitiger Datenstrom, ...). Es ist allerdings unklar, welche Relevanz diese Ströme wirklich besitzen und in welcher Form die Daten folglich im Gehirn gespeichert werden. Es kann kein eindeutiger Schluss auf die Datenverarbeitung gemacht werden und die technischen Implementierungen nutzen meist eigene Konzepte, die sich an technischen Gegebenheiten orientieren [7].

Retinamodelle, die direkt von den Erkenntnissen der Neurowissenschaft umgesetzt werden, werden mit Problemen des Energieverbrauchs, der Echtzeitsimulation, der Zuverlässigkeit, der Augapfeldrehung zur Verfolgung von Objekten und der Chipgröße konfrontiert [6]. Mathematische Konzepte zur Sehachsenausrichtung sind in der Arbeit von B.W. Lin und Ou Ma [8] einsehbar.

3.2 Verarbeitungsgeschwindigkeit

Eindrucksvoll konnte das Team rund um Simon Thorpe [11] die Verarbeitungsgeschwindigkeit des V1 feststellen. Es gab viele Fragen bezüglich der Geschwindigkeit, da Daten, die zwischen V1 und V5 übertragen wurden, meist nur wenige Schaltstationen passieren. Danach wurde der Kontext der Bilddaten vollständig aufgelöst und Entscheidungen werden bereits getroffen.

Die Probanden sahen Bilder mit natürlichen Szenen, die mit einer bestimmten Wahrscheinlichkeit Tiere enthielten. Der Proband wurde ersucht eine Taste zu drücken, solange kein Tier gesehen werden konnte. Nach der Auswertung der Versuchsreihe stellte Simon Thorpe et al. fest, dass es bis zum Lösen der Taste durchschnittlich nur 300 Millisekunden dauert. Inbegriffen in dieser Zeit sind selbstverständlich die Zeit der Übertragung der Bildinformation ins Gehirn als auch die motorische Reaktionsfähigkeit. In einem weiteren Versuchsaufbau

wurden die neuronalen Reaktionen mit einem *Elektroenzephalogramm* (EEG) gemessen. Fazit dieser empirischen Untersuchung war es, dass bereits nach 150 Millisekunden das Gehirn sein Urteil gefällt hat, wobei noch etwa 70 Millisekunden für den Transport der Daten in den primären Sehbereich zu subtrahieren sind. Folglich schafft es das Gehirn in etwa 80 Millisekunden ein Urteil zu fällen; eine Zeitspanne, die nach Modellen bisheriger Experten nicht erreicht werden konnte.

3.3 Die technische Implementierung in Bildsensoren

Das dem menschlichen Sehapparat zugrundeliegende Prinzip findet sich im Groben auch in digitalen Kameras. Was für das Auge die Retina ist, sind für digitale Kameras die Bildsensoren. Sie wandeln das einfallende Licht in elektrische Signale um. Es gibt verschiedene Arten von Bildsensoren, am gebräuchlichsten sind CMOS- (Complementary Metal Oxide Semiconductor) und CCD-Chips (Charge-Coupled Device) [10]. Obwohl bei beiden Sensorarten die eintreffenden Photonen (Lichtteilchen) elektrische Ladungen generieren, unterscheiden sie sich in einigen Punkten.

Vereinfacht gesagt, besteht ein CCD-Chip [10, S. 1] unter anderem aus einer zweidimensionalen, rechteckigen Anordnung von Kondensatoren (vorstellbar als Gitter oder Tabelle). Im lichtsensiblen Teil werden die einfallenden Photonen über eine bestimmte Zeit gesammelt und so eine Ladung generiert. Diese Ladungspakete werden dann von Kondensator zu Kondensator weitergegeben. Dabei wird die Schranke zwischen zwei Kondensatoren durchlässig, so dass die Ladungen weiterwandern können. So wandern die Ladungen der obersten Zelle jeder Spalte nach unten. Dort befindet sich ein gemeinsamer Ausgabe-Verstärker für alle Spalten. Diese Konfiguration bedingt, dass die Bildinformationen aller Pixel zur exakt gleichen Zeit gesammelt werden. Außerdem bedingt die Verwendung einer gemeinsamen Ausgabestufe, dass diese bei hoher Pixelanzahl und Bildwiederholraten mit hoher Frequenz arbeiten muss.

Bei CMOS-Chips [10, S. 2] hingegen ist jeder Pixel einzeln ansprechbar. Die jeweilige Sensorzelle beendet, wenn sie elektronisch angesprochen wird, die Sammlung von Photonen und gibt die entsprechende Ladung an einen eigenen Verstärker weiter. Nach dieser "Entleerung" kann die Sensorzelle sofort wieder mit dem Sammeln von Lichtteilchen beginnen. Vor der nächsten Entladung werden alle anderen Sensorzellen entleert. Dieses Funktionsprinzip bedingt, dass die Bildinformationen aller Pixel zu unterschiedlichen Zeiten gesammelt werden. Dadurch, dass jeder Pixel seine eigene Ausgabestufe hat, können diese auch bei hohen Auflösungen und Bildwiederholraten mit geringeren Frequenzen arbeiten.

Diese unterschiedlichen Funktionsweisen bedingen verschiedene Vor- und Nachteile [10]. Dadurch dass CCD-Sensoren beispielsweise nur einen Ausgabeverstärker benutzen, ist der Anteil an Rauschen im Verhältnis zum Bildsignal (*signal-to-noise-ratio*) wesentlich geringer. Allerdings wirkt sich diese Bauform bei Anwendungen wo Geschwindigkeit bzw. eine hohe Bildwiederholrate ein Kriterium sind (z.B. Video) negativ aus.

Neben der Anwendung bei digitalen Videokameras finden sich CMOS-Chips vor allem im Consumer-Bereich (z.B. in Mobiltelefonen). Das liegt neben dem geringeren Preis bei der Herstellung auch daran, dass CMOS-Chips weniger Strom verbrauchen und geringere Spannungswerte zum Funktionieren benötigen.

Beide Sensorarten haben sich in den letzten Jahren aber in ihren Eigenschaften angeglichen [3, S. 6].

3.4 Der Ausblick

Trotz der offenen Fragen, widersprüchlichen Modelle und technischen Schwierigkeiten kann man von einem wesentlichen Fortschritt der Bionik auf diesem Gebiet sprechen; die letzten Jahre haben die *computer vision* grundlegend verändert. Kameras sind in jedem Handy enthalten und Sensoren entscheiden über unseren Komfort in Autos. Die Zukunft ist gesichert; nicht zuletzt aufgrund des wirtschaftlichen Interesses zuverlässige, digitale Augen zu produzieren und auch des allgemeinen Interesses des Menschen seine eigene Funktionsweise zu verstehen.

Abbildungsverzeichnis

1	Querschnitt des menschlichen Auges	3
3	Die Ebenen der vertikalen Signalleitung	5
2	modifizierter Screenshot der RGB Farbauswahl aus dem Grafikprogramm GIMP	7
4	”Optische Täuschung” unter der GNU FDL lizenziert. Zuletzt abgerufen am 12. Dezember 2010 14:15 http://www.dailycognition.com/content/ image/14/800px-cafe-wall.png	7
5	Reaktionsverhalten der Neuronen auf RGB-Farben	8

Graphiken ohne Lizenz unterstehen der GNU FDL (Free Documentation License) und wurden mit dem freien Grafikprogramm Inkscape erstellt.

Literatur

- [1] Vincente Baixauli and Marcello Negri. *Anatomie des Menschen*. Kaiser, neuauflage 2002 edition, 1997.
- [2] D. Betancourt and C. del Rio. Study of the human eye working principle: An impressive high angular resolution system with simple array detectors. In *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop on*, pages 93–97, July 2006.
- [3] B.S. Carlson. Comparison of modern ccd and cmos image sensor technologies and systems for low resolution imaging. In *Sensors, 2002. Proceedings of IEEE*, volume 1, pages 171–176, January 2002.
- [4] Karl R. Gegenfurtner, Sebastian Walter, and Doris I. Braun. *Visuelle Informationsverarbeitung im Gehirn*, pages 42–45. mentis, 2002.
- [5] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat’s visual cortex. *jphysiol*, 160:106–154, 1962.
- [6] Wei Hui, Guan Xu-Dong, and Zuo Qingsong. Main retina information processing pathways modeling. In *Cognitive Informatics (ICCI), 2010 9th IEEE International Conference on*, pages 318 –324, July 2010.
- [7] Ng Jeffrey, Anil A. Bharath, and Li Zhaoping. A survey of architecture and function of the primary visual cortex (v1). *EURASIP Journal on Advances in Signal Processing*, page 17, 2007.

- [8] B.W. Lin and Ou Ma. Modeling and analysis of human binocular perception of position and velocity in the visual cortex. In *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, pages 973–978, August 2007.
- [9] Peter H. Schiller, Nikos K. Logothetis, and Eliot R. Charles. Functions of the colour-opponent and broad-band channels of the visual system. *letters to nature*, 343:68–70, January 1990.
- [10] A.J.P. Theuwissen. Ccd or cmos image sensors for consumer digital still photography? In *VLSI Technology, Systems, and Applications, 2001. Proceedings of Technical Papers. 2001 International Symposium on*, pages 168–171, January 2001.
- [11] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381(6582):520–522, June 1996.
- [12] K. Yamamoto, Y. Takeuchi, J. Tachibana, K. Uchida, M. Maeno, K. Yanashima, A. Uchiyama, and K. Magatani. Analysis of brain activation by visual and auditory stimulation after visual and auditory simultaneous stimulation using f-mri. In *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, volume 1, pages 1872–1875, September 2004.