

Konrad L. Maul

„Das räumliche Sehen“ physiologische und geometrische Grundlagen zu 3D / Stereo Vision / Stereopsis



1. Einführung

Dem Traum der Menschen von einer naturgetreuen Übertragung und Wiedergabe von Szenen und Ereignissen in die eigenen vier Wände kommen heutige Fernsehgeräte mit neuester Signalverarbeitung und Displaytechnologie schon sehr nahe. Der flache Bildschirm an der Wand ist Realität geworden. Die großformatige, detailgetreue Wiedergabe, angepasst an die Physiologie des menschlichen Gesichtssinnes durch 16:9 und HDTV ermöglicht eine bisher nicht gekannte Einbezogenheit in das Geschehen auf dem Bildschirm (Telepräsenz).

Nun fehlt den hochauflösenden Bildern nur noch die dritte Dimension der realen Welt d.h. die räumliche Darstellung.

In vielen Abhandlungen und Artikeln über das räumliche Sehen wird von einem naiven Verständnis dieses physiologischen Prozesses ausgegangen. Man bietet den Augen einen linken und rechten Bildauszug an und schon entsteht ein räumlicher Bildeindruck. Dies wird der Komplexität des Systems Auge-Gehirn in keiner Weise gerecht. Deswegen möchte ich den Versuch unternehmen zunächst die physiologischen Grundlagen etwas abzuklären.

2. Physiologische Hintergründe / System Auge-Gehirn

Das Auge

Ein horizontaler Querschnitt durch das menschliche Auge ist in Abb. 1 zu sehen.

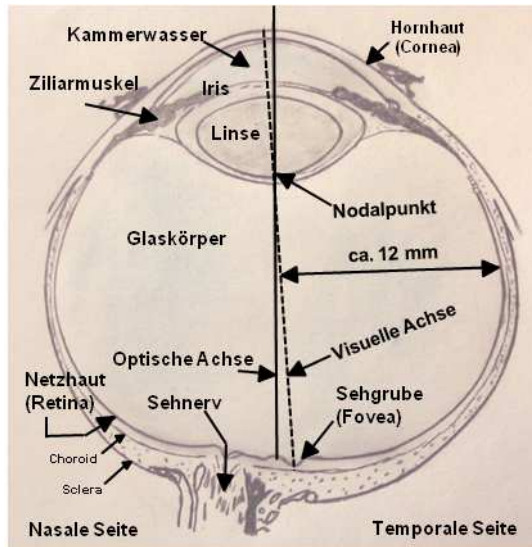


Abb.1: Horizontaler Schnitt durch das rechte menschliche Auge.

(Eigene Zeichnung nach *The vertebrate visual system*, by Polyak, 1957, University of Chicago Press; in Howard I.P. 2002 *Seeing in Depth*)

Das menschliche Auge ist annähernd kugelförmig mit einem Durchmesser von ca. 24 mm. Die Hornhaut (Cornea) hat einen Durchmesser von ungefähr 12 mm und einen Kurvenradius von ca. 8 mm. Es wird oft angenommen, dass die Linse dazu dient, die einfallenden Lichtstrahlen zu bündeln und das Bild auf der Netzhaut (Retina) zu erzeugen. Der Ort an dem beim menschlichen Auge die Lichtstrahlen am stärksten gebrochen werden ist aber die Vorderseite der Cornea. Dies ist darin begründet, dass die Brechkraft einer Linse grundsätzlich von dem Unterschied zwischen den Brechungskoeffizienten des umgebenden Mediums und dem des Linsenmaterials bestimmt wird. Der Brechungskoeffizient der Luft ist niedrig, während der des Kammerwassers hinter der Cornea annähernd so hoch wie der Brechungskoeffizient der Linse des Auges ist. Dadurch hat die Cornea mit ca. 42 Dioptrien ungefähr 70% der Brechkraft des gesamten Auges. Die Linse besitzt einen Durchmesser von ca. 9 mm und ist ungefähr 4 mm dick. Die Brechkraft der Linse kann durch den Ziliarmuskel um ca. 10 Dioptrien verändert werden und sorgt durch Verstellung (Akkommodation) dafür, dass das Netzhautbild bei entfernten und nahen Objekten scharf gestellt werden kann. Damit hat das Gesamtsystem Cornea-Linse eine Brechkraft die von ca. 60 bis zu 70 Dioptrien variiert werden kann. Als visuelle Achse (Mittelpunktstrahl) wird die Verbindungslinie zwischen einem Punkt eines fixierten Gegenstandes und dem Zentrum der Sehgrube (Fovea centralis) bezeichnet. Als optische Achse wird der am besten passende Strahl durch die vier brechenden Oberflächen des Auges bezeichnet. Die optische Achse schneidet die Retina ca. 1,5 mm seitlich entfernt von der Sehgrube und ca. 0,5 mm über der Sehgrube. Sie bildet somit mit der visuellen Achse einen Winkel von ca. 5°. Beide Achsen müssen gemäß den optischen Gesetzen durch den resultierenden Knotenpunkt /Nodalpunkt (nodal point) des optischen Systems Auge gehen. Dieser liegt beim nicht akkommodierten Auge

ca. 17 mm vor der Retina und durch ihn gehen alle Strahlen die eine eins zu eins Abbildung der Punkte des Sehfelds auf der Netzhaut reproduzieren.

Die Iris (Regenbogenhaut) in Verbindung mit den Irismuskeln kontrolliert die Linsenöffnung des Auges (Apertur, Pupillengröße). Oft wird angenommen, dass die Veränderung der Pupillengröße dem Auge erlaubt, ein breites Spektrum von Lichtintensitäten auszunutzen. Da sich ihre Fläche aber nur in einem Verhältnis von 16:1 verändert, kann dies nicht ihre Hauptaufgabe sein, da das Auge einen Helligkeitsbereich von mehr als 1 000 000 : 1 zu nutzen imstande ist. Offensichtlich kontrahiert die Pupille, um die Lichtstrahlen auf den zentralen und optisch effektivsten Teil der Linse zu beschränken und damit Linsenfehler zu minimieren. Die maximale Öffnung wird nur dann eingestellt wenn maximale Empfindlichkeit benötigt wird.

Jedes Auge wird von sechs Muskeln bewegt wie in Abb.2 dargestellt.

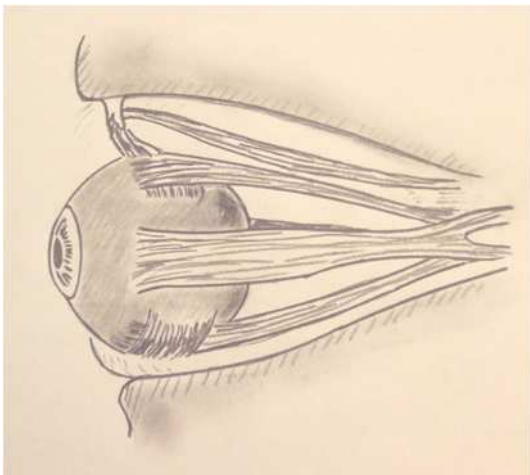


Abb.2: Augenmuskeln des linken menschlichen Auges.

(Eigene Zeichnung nach Cogan 1956 *Neurology of the ocular muscles*, Charles C. Thomas, Publisher, Springfield, Illinois; in Howard I.P. 2002 *Seeing in Depth*)

Dies ist erforderlich, da der Mensch nur in der Sehgrube ein hohes Auflösungsvermögen der Netzhaut besitzt und dadurch die Augen bewegt werden müssen um auszuwählen was scharf gesehen werden soll. Wird ein Objekt betrachtet, so konvergieren beide Augen sodass die Abbildung unabhängig von der Entfernung des Objektes auf der Sehgrube zentriert wird. Die Augenmuskeln schwenken dabei die Augäpfel um den Rotationsmittelpunkt jeweils in Richtung der Nase hin (nasal). Wie wir noch sehen werden wird dieser Konvergenzwinkel auch als einer von vielen Parametern zur Ermittlung des Raumeindrucks vom Gehirn genutzt.

Bildsensor Netzhaut (Retina) - Außenstelle des Gehirns

Die Struktur der Netzhaut ist in Abb. 3 aufgezeigt.

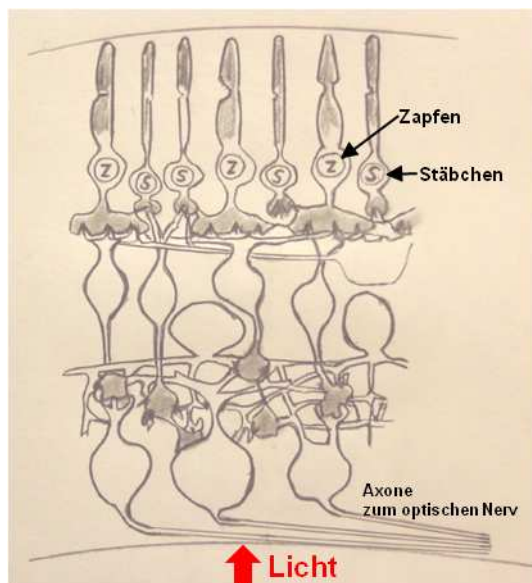


Abb. 3: Die Struktur der Netzhaut.

(Eigene Zeichnung nach Dowling and Boycott 1966; in Howard I.P. 2002 *Seeing in Depth*)

Diese Struktur wurde zuerst von Ramón y Cajal ermittelt und in einer Serie von Abhandlungen zwischen 1888 und 1933 beschrieben. Die Netzhaut ist eine mehrlagige Membran mit einer Fläche von ungefähr 1000 mm². Sie weist in der Sehgrube eine Dicke von etwa 250 µm auf, die in der Peripherie auf ca. 100 µm abnimmt. Die Fotorezeptoren sind dicht gepackt in der äußeren Schicht, also der Schicht die am weitesten von der Lichtquelle entfernt ist. Das bedeutet, dass das Licht vor dem Erreichen der Rezeptoren durch die Blutgefäße und das feine Netzwerk der Nervenfasern, einschließlich dreier Schichten von Zellkörpern und einer Menge Stützgewebe, dringen muss.

Es gibt zwei Arten von Fotorezeptoren. Die Stäbchen haben eine sehr hohe Lichtempfindlichkeit, breite spektrale Abstimmung und sind ausschließlich in der Peripherie, also außerhalb der Sehgrube, verteilt. Sie werden für die Wahrnehmung der Schwarz/Weiß-Helligkeitsempfindung genutzt. Die Zapfen, die für die Farbwahrnehmung genutzt werden, haben eine niedrigere Lichtempfindlichkeit und eine hohe Dichte in der Sehgrube, die in der peripheren Netzhaut abnimmt. Es gibt drei Typen von Zapfen mit jeweils verschiedener spektraler Empfindlichkeit: S-Zapfen mit einer maximalen Empfindlichkeit bei ungefähr 450 nm Lichtwellenlänge (blau). M-Zapfen mit einer Spitzenempfindlichkeit von ca. 535 nm (grün). L-Zapfen mit einem Maximum bei 565 nm (gelb/rot). 5 bis 10% der Zapfen sind S-Zapfen. M- und L-Zapfen sind etwa gleich häufig verteilt.

In dieser Struktur liegt auch die Basis, die unsere technischen Fernsehsysteme bei der Redundanz-Datenreduktion benutzen, bei der keine sichtbaren Verluste entstehen. So kann z.B. die Digitalisierung der Farbinformation mit einer geringeren Datenrate erfolgen als der für die Helligkeitswerte erforderliche Wert.

Der normale Bereich der Lichtempfindlichkeit des menschlichen Auges erstreckt sich über drei logarithmische Einheiten von 0,0000007 cd/m² bis 0,0004 cd/m². Die Helligkeit von äußeren Stimuli variiert um ca. 10 logarithmische Einheiten (log). Spezielle Mechanismen kompensieren den begrenzten dynamischen Bereich des Auges.

Um einen Vergleich der örtlichen Auflösung (Wiedergabe von Details) des menschlichen Auges mit technischen Bildwiedergabesystemen ziehen zu können sind die folgenden Werte des menschlichen Auges wichtig.

Die Retina des erwachsenen Menschen hat zwischen 4 und 6 Millionen Zapfen mit einer Spitzendichte von 100 000 bis 320 000 pro mm² in der Sehgrube. Diese nimmt in der Peripherie stark ab auf z.B. 6000 pro mm² bei 10° Exzentrizität. Die Fovea ist eine Grube von ungefähr 1,5 mm Durchmesser, die ein regelmäßiges hexagonales Mosaik von Zapfen im Abstand von 2 bis 3 µm enthält. Die zentrale Sehgrube (Fovea centralis) hat einen Durchmesser von ca. 0,27 mm, was einem Öffnungswinkel (subtend) von ca. 1° entspricht, und enthält ca. 6000 Zapfen. Die menschliche Retina hat mehr als 100 Millionen Stäbchen, die in der Fovea nicht vorhanden sind. Die Fotorezeptoren der Netzhaut bilden 70% der Rezeptoren des menschlichen Körpers.

Wenn wir den Vergleich zu unserem modernen hochauflösenden Fernsehen ziehen (HDTV) dann besitzt das Display ca. 2 Millionen Bildpunkte, die jeweils aus einem roten, grünen und blauen Subpixel bestehen. D.h. insgesamt sind nur ca. 6 Millionen Bildpunkte vorhanden, die gleichmäßig über die gesamte Displayfläche verteilt sind.

Beim menschlichen Auge sorgen dagegen die 100 Millionen Stäbchen, die insbesondere sehr empfindlich auf bewegte Objekte reagieren, dafür dass die Objekte denen unsere jeweilige Aufmerksamkeit gilt, durch die entsprechenden Augenbewegungen scharf auf der Fovea mit höchster Auflösung abgebildet werden können. Die Natur hat das, was wir heute in der Fernsehtechnik mit Telepräsenz bezeichnen schon in idealer Weise realisiert.

Man hat die Retina als eine Art Ausstülpung des Gehirns beschrieben.

Entwicklungsgeschichtlich spezialisierte sich also ein Teil der Gehirnoberfläche, wurde nach außen verlagert und lichtempfindlich. Die Retina hat bei diesem Prozess typische Gehirnzellen beibehalten, die zwischen den Lichtrezeptoren und den Sehnervenfasern, also in den vorderen Schichten der Netzhaut liegen, und die elektrische Aktivität der Rezeptoren stark modifizieren. D.h. ein Teil der Wahrnehmungssignalverarbeitung findet bereits im Auge statt. Damit wird dieses auch funktional ein Bestandteil des Gehirns. Diese Vorverarbeitung führt die über 100 Millionen Rezeptorsignale auf ca. 1 Million Sehnervenfasern zusammen. Dadurch kann sich die Dicke des Sehnervs stark verringern, was Augenbewegungen erst ermöglicht.

Bilddatenübertragung (Chiasma opticum , primäre Sehrinde)

Das für das Sehen verantwortliche neuronale System beginnt, wie oben gezeigt, schon mit der unglaublich komplexen Struktur der Netzhäute. Jede Netzhaut ist senkrecht in zwei funktionale Hälften geteilt. Die Sehnervenfasern der inneren (nasalen) Hälfte wechseln am Sehnervenkreuz (Chiasma opticum) auf die andere Seite um zur gegenseitigen Hirnhemisphäre zu ziehen (Abb. 4).

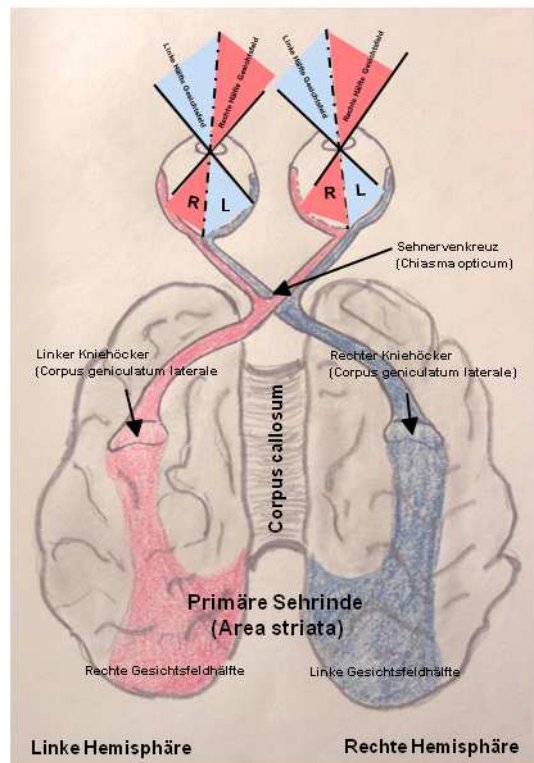


Abb. 4: System Auge/Gehirn von oben gesehen. Schema der Sehbahnen, das die linken und rechten Gesichtsfeldhälften mit den retinalen Bildern und der teilweisen Kreuzung im Sehnervenkreuz zeigt. (Eigene Zeichnung nach Eccles J.C. 1990 *Die Psyche des Menschen*)

Die Fasern der äußeren Netzhauthälfte gelangen zur gleichseitigen Hemisphäre. Am jeweiligen seitlichen Kniehöcker, quasi einer Relaisstation, sind die Sehnerven dann an die jeweilige Hemisphäre angeschlossen. Interessanterweise empfangen die Kniehöcker mehr „absteigende“ Bahnen von höheren Zentren des Gehirns als aufsteigende Bahnen von den Augen. Dies ist die anatomische Basis dafür, dass höhere Zentren des Gehirns die von den Netzhäuten kommenden Signale beeinflussen oder erweitern und damit die Netzhautbilder deuten.

Bildsignalverarbeitung im Gehirn (visuelle Merkmale in spezialisierten Modulen)

Zu einer der wichtigsten Entdeckungen bezüglich der visuellen Signalverarbeitung im Gehirn gehört die der beiden amerikanischen Physiologen David Hubel und Torsten Wiesel. Sie registrierten bei der Katze die Aktivität einzelner Zellen der Sehrinde, während den Katzenaugen einfache visuelle Formen, meist Lichtbalken, angeboten wurden. Dies geschah mittels eines Diaprojektors, der diese Muster auf einen Schirm vor der Katze projizierte. Hubel und Wiesel stellten fest, dass einige kortikale Zellen der ersten Verarbeitungsstufe nur dann aktiv waren (feuerten), wenn ein Balken in einem ganz bestimmten Winkel angeboten wurde (Abb. 5).

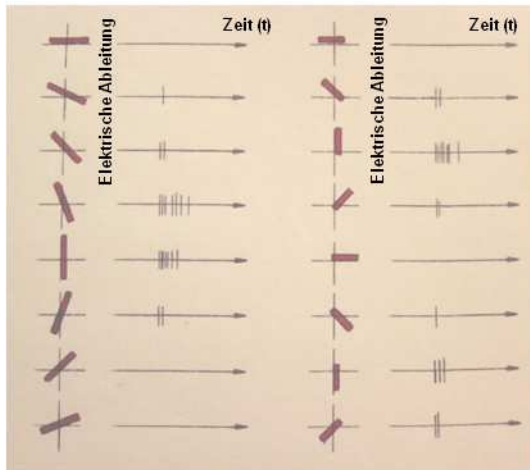


Abb. 5: Hubels und Wiesel's Ableitungen von einzelnen Zellen im visuellen Cortex der Katze. (Eigene Zeichnung nach Gregory R.L. 2001 *Auge und Gehirn; Psychologie des Sehens*)

Bei anderen Winkeln blieb diese Zelle stumm. Die zur Aktivierung erforderliche Neigung des Balkens wechselte von Zelle zu Zelle. Weiterhin stellten Sie fest, dass manche Zellen sich nur durch Bewegung aktivieren ließen oder nur durch Bewegung in eine bestimmte Richtung.

Der Professor für Neuropsychologie Richard L. Gregory schreibt: „Es wird immer deutlicher, dass das Gehirn visuelle Merkmale in spezialisierten „Modulen“ verarbeitet, wobei es unterschiedliche neuronale Kanäle für Form, Bewegung, Tiefenwahrnehmung, Farbe usw. gibt.

Allmählich kristallisiert sich die Organisation der Sehrinde heraus, wenn auch noch immer unklar ist, wie sie mit dem Sehen komplexer Objektformen verknüpft ist.

Vielleicht werden wir das visuelle Gehirn erst dann vollständig verstehen, wenn es uns gelingt, eine Maschine zu entwerfen und zu konstruieren, die über ein komplexes Sehvermögen verfügt.“

Disparationsneuronen (David Hubel und Torsten Wiesel)

Die Tiefenwahrnehmung ist im Prinzip abhängig von Ungleichheiten (Disparität) der beiden Abbildungen auf den beiden Netzhäuten, die durch den 50 bis 70 mm betragenden Augenabstand entstehen. Sie kann auch als eine Täuschung betrachtet werden, weil es auf der Retina nur ein zweidimensionales Bild gibt. Dieses wird aber so interpretiert als ob es aus einem dreidimensionalen Objekt resultieren würde. Mit dem Entstehen dieser Täuschung befasst sich die psychologische und physiologische Forschung der visuellen Wahrnehmung.

Raymòn y Cajal hatte schon 1911 die Theorie entwickelt, dass Eingangssignale von korrespondierenden Bereichen der beiden Netzhäute in den von ihm mit „isodynamic cells“ bezeichneten neuronalen Zellen zusammengeführt würden und damit die Basis für das binokulare Sehen gebildet würde. Diese Idee wurde von David Hubel und Torsten Wiesel (1959; 1962) experimentell verifiziert. Sie berichteten dass Paare von Sehnervenbahnen der Katze in „binokularen“ Zellen in dem Teil der Hirnrinde, der visuelle Signale verarbeitet (visueller Cortex), zusammenlaufen. Und dass die rezeptiven Felder jeder „binokularen“ Zelle korrespondierende Positionen in den beiden Augen besitzen. Wenn eine „binokulare“ Zelle identische rezeptive Felder hätte werden, die in jedem Auge identisch positioniert wären, würde diese optimal auf Netzhautbilder

(Stimuli) mit null Disparität reagieren und keine Tiefeninformation könnte am Ausgang der Zelle gewonnen werden. Das war schon ein Argument das der Physiologe, Physiker und Psychologe Hermann von Helmholtz gegen die Idee von konvergenten visuellen Eingangssignalen angeführt hatte. Helmholtz neigte zu der Auffassung, dass die räumliche Wahrnehmung ein erlernter Prozess sei und sich in den „höheren“ Verarbeitungsschichten des Gehirns abspielen würde.

Der Psychologe Ian P. Howard schreibt dazu in „Seeing in Depth“:

„Das Problem würde gelöst werden können wenn es Zellen gäbe, die speziell darauf abgestimmt wären, auf ähnliche Bilder in leicht unterschiedlichen Positionen der beiden Retinae, zu reagieren. Unterschiedliche Zellen die optimal auf differente Disparitäten abgestimmt sind. Einfach wie diese Idee klingt wurde diese bis 1965 nicht vorgeschlagen. Wahrscheinlich weil die Idee einer auf spezifische Stimuli-Eigenschaften reagierenden kortikalen Zelle nicht „en vogue“ war, bis 1959 Hubel und Wiesel kortikale Zellen entdeckten, die abgestimmt waren auf Ausrichtung und Bewegung zu reagieren. Hubel und Wiesel gelang es nicht Zellen zu finden, die sensitiv auf Disparitäten reagierten. Jack Pettigrew lieferte den ersten Beweis von Disparitäts-Detektoren in einer frühen Stufe der visuellen Signalverarbeitung. Er machte diese Arbeit im Rahmen seiner „undergraduate thesis“, die er 1965 an der Universität von Sydney geschrieben hat.“ In vielen weiteren Arbeiten wurde diese Theorie der „Disparationsneuronen“ inzwischen bestätigt.“

Marr und Poggio haben 1979 basierend auf diesen neuronalen Fakten ein erstes nachrichtentechnisches Signalverarbeitungsmodell (A Computational Theory of Human Stereo Vision) entwickelt. Inwieweit solche Modelle die tatsächlichen Mechanismen der höheren Hirnprozesse abbilden, konnte aufgrund deren ungeheuren Komplexität bisher noch nicht gezeigt werden.

Sehen in der Tiefe (Disparation, Konvergenz, monokulare Informationen)

Nachdem wir mit den Disparationsneuronen quasi die unterste Ebene des Tiefensehens kennengelernt haben wollen wir uns auf der Ebene der Wahrnehmungspsychologie mit dem Sehen in der Tiefe beschäftigen.

Das menschliche horizontale Blickfeld ist in Abb. 6 dargestellt.

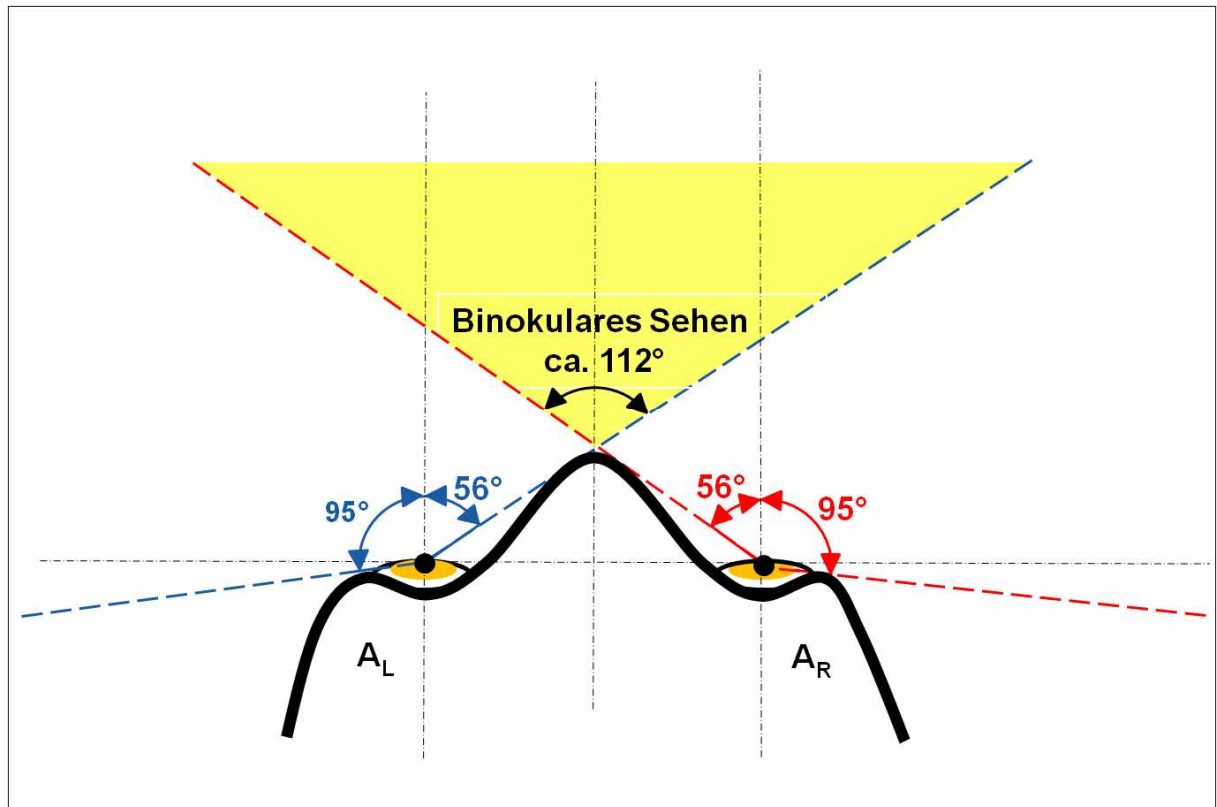


Abb. 6: Das binokulare Gesichtsfeld des Menschen.

Das monokulare Gesichtsfeld des statischen Auges beträgt ungefähr 95° in Richtung der Schläfe (temporal) und ungefähr 56° in Richtung der Nase. Das gesamte Gesichtsfeld ist der volle Winkel der ausgehend von einem Punkt in der Mitte zwischen beiden Augen alle Punkte im Raum einschließt, die entweder einem Auge oder beiden Augen sichtbar sind. Beim Menschen beträgt dieses horizontale Gesichtsfeld (lateral) bei feststehenden Augen ca. 190° und erreicht 290° wenn die Augen seitlich bewegt werden können. Das binokulare Gesichtsfeld ist der Teil des gesamten Gesichtsfeldes innerhalb dessen ein Objekt liegen muss um für beide Augen sichtbar zu sein (bei fester Augenposition). Das binokulare Gesichtsfeld ist von den monokularen Sektoren des linken und rechten Auges flankiert. Jeder monokulare Sektor erstreckt sich ungefähr 37° seitlich (lateral) vom Schläfenring bis zur Grenze des binokularen Sektors. Die rechte und linke Grenze des binokularen Sehens wird durch die Nase geformt. Für binokulares Sehen ergibt sich daraus ein Winkel von ca. 112° , wenn die Augen symmetrisch konvergieren. Betrachten wir zunächst die geometrische Basis. Weil die beiden Augen horizontal um den Augenabstand voneinander getrennt sind, sehen diese dieselbe visuelle Szene von zwei unterschiedlichen Ausgangspunkten. Wenn wir einen speziellen Punkt im Raum, wie z.B. den schwarzen Punkt auf dem Pfeil in Abb. 7A fixieren, sorgen wir dafür dass das Abbild dieses Punktes in die Sehgrube (Fovea) des linken und des rechten Auges fällt, also in die Region der Netzhaut mit der höchsten Auflösung.

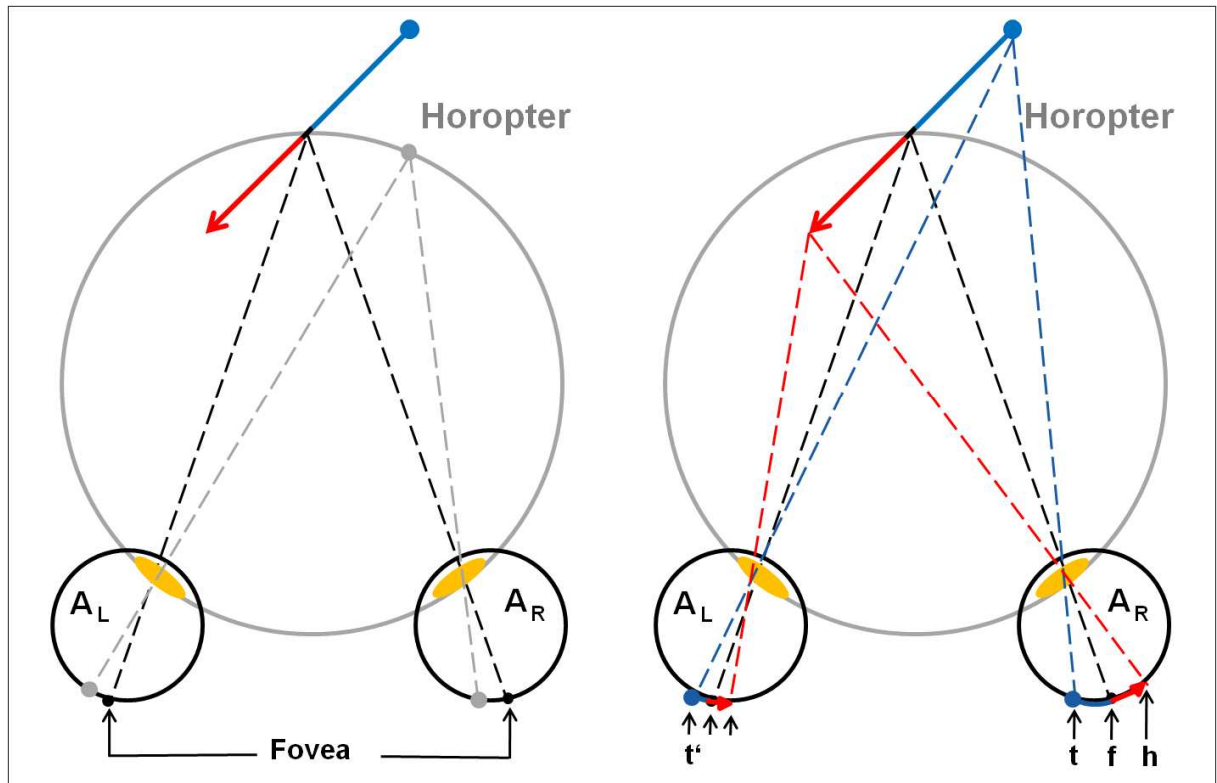


Abb. 7A: Die Geometrie der Stereopsis. Horizontaler Schnitt durch das rechte (A_R) und linke Auge (A_L) von oben gesehen. Auf dem Horoapter liegende Punkte produzieren Abbildungen auf korrespondierenden Punkten der beiden Netzhäute. Als Horoapter wird die kreisförmige horizontale Linie bezeichnet auf der alle Punkte liegen, die bei gegebener Augenstellung mit beiden Augen nur einfach gesehen werden. (Eigene Zeichnung nach Tyler C.W. 2004 *Binocular Vision*; in Ponco R. et.al. 2008 *Stereopsis*)

Abb. 7B: Die Geometrie der Stereopsis. Horizontaler Schnitt durch das rechte (A_R) und linke Augen (A_L) von oben gesehen. Punkte auf dem Pfeil, die nicht auf dem Horoapter liegen, mit verschiedenen Abständen vom Beobachter produzieren Abbildungen mit verschiedenen Abständen von der Fovea auf den zwei Netzhäuten. Diese Differenz wird binokulare Disparität genannt und ist die Basis für die allgemein bekannteste Art der Tiefenwahrnehmung. (Eigene Zeichnung nach Tyler C.W. 2004 *Binocular Vision*; in Ponco R. et.al. 2008 *Stereopsis*)

Dazu schwenken die Augenmuskeln die Augen in den jeweils entsprechenden Winkel (Konvergenz) und der Muskel der Augenlinse sorgt durch Akkommodation (Veränderung der Brennweite) für eine scharfe Abbildung. Dieser Akt definiert die horizontale Fixierungskurve. Wenn wir auf diesem schwarzen Punkt des Pfeiles starten und uns eine bestimmte Distanz auf dem grauen Kreis bewegen werden sich die Projektionen des neuen Punktes in den beiden Augen um die exakt gleiche Distanz auf den beiden Netzhäuten bewegen. Diese Projektionspunkte auf den beiden Retinae werden als „korrespondierende Punkte“ bezeichnet. Der „geometrische Horoapter“ ist die Zusammenfassung all der Punkte eines Bildes, die auf „korrespondierende Punkte“ auf den Retinae projizieren. Für einen Beobachter dieser Punkte erscheint es so als ob alle diese Punkte ungefähr in derselben Tiefe wie der Fixierungspunkt liegen würden. Die Menge aller Punkte die exakt in derselben Tiefe erscheinen wird als „empirischer Horoapter“ bezeichnet und ist etwas flacher als der kreisförmige „geometrische Horoapter“. Für prinzipielle Überlegungen kann zur Vereinfachung mit dem „geometrischen Horoapter“ gearbeitet werden. Betrachten wir nun Teile des Objektes die entweder vor oder hinter dem „Geometrischen Horoapter“ liegen, wie zum Beispiel Spitze und Schwanz des Pfeils in Abb. 7B. Diese

Objektpunkte werden nicht auf „korrespondierende Punkte“ der Retinae projizieren, weil die Distanz von f nach h auf der rechten Netzhaut nicht gleich der Distanz von f' nach h' auf der linken Netzhaut ist. Diese Differenz der Netzhaut-Kreisbögen ($hf-h'f'$) wird als „binokulare Disparität“ bezeichnet und meist als Bogenmaß oder Winkel angegeben. Die Projektionslinien für Punkte auf der Netzhaut für nahe Objekte kreuzen sich vor dem Horopter und produzieren was man als „gekreuzte Disparitäten“ bezeichnet und umgekehrt produzieren weit entfernte Objekte, wie z.B. der Schwanz des Pfeils in Abb. 7B, „nicht gekreuzte Disparitäten“.

Wenn die retinalen Disparitäten, die ein Objekt produziert, in einen begrenzten Bereich fallen (ungefähr $0,5^\circ$ in der Nähe der Fovea, etwas mehr bei einer größeren Exzentrizität), wird es als ein einzelnes verschmolzenes Objekt mit Tiefe erscheinen. Außerhalb dieses Bereiches, der als „Panums“ Verschmelzungsbereich bekannt ist, wird das Objekt doppelt erscheinen. Dem Beobachter wird dann jedes der beiden retinalen Bilder bewusst, was dann als „Doppeltsehen“ bezeichnet wird. Trotzdem kann das visuelle System sinnvolle Tiefeninformationen bis zu binokularen Disparitäten von einigen Grad extrahieren, in Abhängigkeit von der Größe des Objektes.

Es ist zu bemerken, die obige Diskussion vorausgesetzt, dass man wissen muss wie zur Abbildung eines Objektes im rechten Auge die passende Abbildung des gleichen Objektes im linken Auge gefunden wird. Dies stellt ein schwieriges Problem dar und ist im Bereich der Bilddatenverarbeitung als „Correspondence Problem“ bekannt. Dieses Problem ist äußerst schwierig, wenn man versucht die einzelnen korrespondierenden Punkte (pixel) auf den zwei retinalen Bildern zu finden (weil es eine große Zahl von Übereinstimmungen gibt). Diese Problemstellung wird leichter handhabbar wenn man lokale Felder von Punkten, z.B. kurze Liniensegmente zum Finden der Übereinstimmungen heranzieht. Zusätzlich ist es wahrscheinlich dass Bildeigenschaften höherer Ordnung, wie z.B. die globale Form eines Objektes, dazu benutzt werden die lokalen Objektübereinstimmungen zu erhärten.

Es gibt noch eine andere Art von Stereopsis bekannt als „da Vinci Stereopsis“ (benannt nach Leonardo da Vinci, der diese zuerst beschrieben hat), die die Benutzung von nicht gepaarten Bereichen der Abbilder in den beiden Retinae der beiden Augen beinhaltet. Dies kommt bei der Überdeckung von Objekten vor. Prinzipiell deswegen, weil das Auge nicht um Objekte herum sehen kann. Ein Objekt in der Form eines Quadrats im Vordergrund wird einen vertikalen Streifen des Hintergrundes auf der linken Seite verursachen der nur durch das linke Auge (A_L) gesehen werden kann. Auf der rechten Seite wird dieses quadratische Objekt einen Streifen verursachen, der nur mit dem rechten Auge (A_R) gesehen werden kann (Abb. 8).

Diese Art von Stereopsis ist ebenfalls für die Tiefenwahrnehmung nützlich, speziell zur Definition der Verdeckung. Wir wollen hier aber nicht näher auf die „da Vinci Stereopsis“ eingehen.

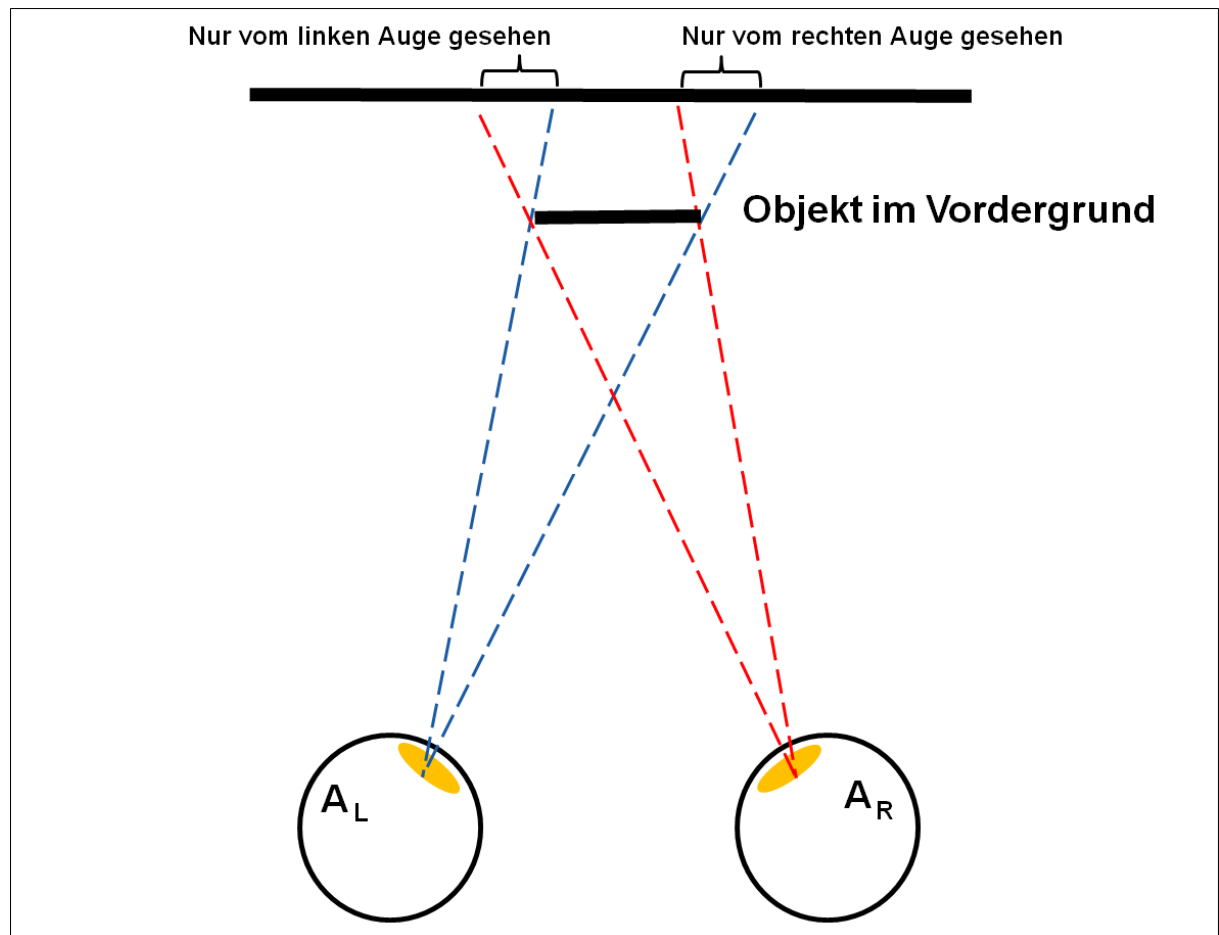


Abb. 8: Die Basis der „da Vinci Stereopsis“

Horizontaler Schnitt durch das rechte (A_R) und linke Auge (A_L) von oben gesehen.

Ein im Vordergrund vorhandenes Objekt hat zur Folge, dass Bereiche im Hintergrund jeweils nur von einem Auge gesehen werden können.

(Eigene Zeichnung nach Nakajama K. et. al. 1990 *Da Vinci Stereopsis*; in Ponco R. et.al. 2008 *Stereopsis*)

Räumliche Auflösung des binokularen Sehens (Stereopsis)

Weil die elementaren Daten mit denen das Gehirn arbeiten muss die binokularen Disparitäten sind, wird die kleinste Tiefendifferenz die unterschieden werden kann durch die kleinste retinale Disparität die aufgelöst werden kann bestimmt. Bei menschlichen Beobachtern, erreicht diese Differenz 5 Bogensekunden (entsprechend $0,0014^\circ$), wenn das beobachtete Objekt auf der Fovea abgebildet wird. Dies ist eine wahrlich bemerkenswerte Präzision, wenn man bedenkt dass der Abstand zwischen den Zapfen auf der Fovea in der Größenordnung von 30 Bogensekunden liegt.

Die binokulare Disparität, die sich bei einer gegebenen Tiefendifferenz ergibt, variiert mit dem inversen Quadrat der Distanz zum Objekt. Das bedeutet, dass bei nahen Objekten äußerst kleine Tiefendifferenzen unterschieden werden können. Das können bis zu $25 \mu\text{m}$ sein, was ungefähr der Dicke eines feinen menschlichen Haares entspricht. Auch bei relativ großen Objektentfernungen kann das binokulare Sehen eine Rolle spielen, vorausgesetzt die Tiefe des Objektes ist groß genug. Für eine Entfernung von 100 m z.B. beträgt die minimale auflösbare Tiefe ungefähr 4 m (Ponce 2008).

Andere Parameter für Tiefenwahrnehmung (monokulare Parameter)

Einem signifikanten Anteil der menschlichen Population (ungefähr 5 bis 10%) fehlt die stereoskopische (binokulare) Tiefenwahrnehmung. Man müsste annehmen, dass diese Personen bezüglich ihrer Sehfähigkeit stark behindert wären. Dies ist aber nicht der Fall. Einige von ihnen haben in Berufen Karriere gemacht für deren Ausübung man eine gute räumliche Einschätzung annehmen muss, wie z.B. Chirurgie, Zahnmedizin und professionelles Skifahren. Der Grund dafür ist wahrscheinlich dass es eine Anzahl von anderen Methoden zur Erhaltung von Tiefeninformation gibt. Eine der nützlichsten Verfahren ist durch Kopfbewegungen quasi mehrere Ansichten über die Zeit zu erhalten und damit eine „Bewegungs-Parallaxe“ zu produzieren. Wenn man seinen Kopf in dieser Art bewegt werden Objekte, die sich näher als der Fixierungspunkt befinden, als sich in die Gegenrichtung der Kopfdrehung bewegend wahrgenommen. Objekte, die weiter als der Fixierungspunkt entfernt sind, werden als sich mit der Kopfbewegung bewegend wahrgenommen. Die Geschwindigkeit der wahrgenommenen Bewegung indiziert die Größe der Tiefendifferenz. Im Vergleichstest bei ähnlichen Bedingungen kann die Tiefenwahrnehmung gewonnen aus der „Bewegungs-Parallaxe“ fast so gute Ergebnisse liefern wie die Stereopsis also das binokulare Sehen.

Zusätzlich zur „Bewegungs-Parallaxe“ gibt es noch eine Reihe von Bildparametern, die einen Hinweis auf Tiefe geben. Dies sind auch die Indikatoren die Künstler benutzen um Tiefe auf einer flachen Leinwand vorzutäuschen (Eccles 1990):

- a) Die Position eines Objektes im Bild. Je entfernter ein Objekt ist, desto höher wird es im Bild angesetzt.
- b) Perspektivische Darstellung.
- c) Abgestufte Maserung, die mit der Entfernung feiner wird.
- d) Schatten, die die relative Stellung im Raum anzeigen.
- e) Überdeckungen und Überschneidungen beweisen unzweifelhaft, wie die Objekte relativ zueinander stehen.
- f) Der Freilichteffekt: Verschwommene und mehr blaue Einfärbung entfernter Objekte. Keiner dieser Hinweise kann die feine Tiefenwahrnehmung des binokularen Sehens erreichen, aber sie liefern einen Gesamteindruck der dreidimensionalen Struktur der Umgebung. Kombiniert mit der „Bewegung-Parallaxe“ erreichen diese monokularen Parameter aber nahezu die Tiefenwahrnehmung des binokularen Tiefensehens.

Ein einfaches Experiment kann uns die feine Tiefenwahrnehmung des monokularen Sehens im wahrsten Sinne vor Augen führen: Wenn man ein Auge schließt oder abdeckt fällt ja die binokulare Information aus. Der räumliche Eindruck verschwindet aber nicht. Es wird nur der kleinere Blickwinkel empfunden und vielleicht ein etwas nicht so prägnanter Tiefeindruck.

Bevor wir uns im nächsten Abschnitt den technischen Umsetzungen und Verfahren des 3D-Sehens zuwenden sei zum Abschluss des Teiles „Physiologische Hintergründe“ Ian P. Howard (2002) zitiert: „Wir haben ein wachsendes Repertoire von psychophysischen, physiologischen und computergestützten Techniken. Es gibt eine schnell wachsende Anzahl von Psychologen, Physiologen, Anatomen, Biochemikern, Physikern, Ingenieuren und Informatikern die alle Aspekte des „Sehens“ erforschen. Kein anderer Wissenschaftsbereich umfasst eine größere Vielfalt von Spezialisierungen. Mit all unseren Fortschritten im Verstehen des Sehvorganges, bleibt doch das schiere Erstaunen bestehen, wenn wir ein Stereogramm beobachten. In unseren Versuchen das visuelle Bewusstsein zu verstehen beziehen wir uns auf neuronale Netzwerke, synchrone

Oszillationen, Zellgruppen (cell assemblies), Impulsfolgen und Quantenphysik, aber wir tapen nur im Dunkeln. Das erstaunliche Geheimnis unserer bewussten Wahrnehmung einer dreidimensionalen Welt, der wir begegnen wenn immer wir unsere Augen öffnen und die uns fortwährend fasziniert und verblüfft, erinnert uns an die wundervolle Maschinerie die jeder von uns in seinem Kopf trägt. Diese Maschinerie brauchte einige Milliarden Jahre um sich zu entwickeln aber in jedem von uns nur ungefähr ein Jahr um heranzuwachsen.“

Wie wir gesehen haben stehen wir bei der Physiologie des Sehens trotz der heute schon vorliegenden Erkenntnisse noch immer vor einem ungeheuren Problem, dessen Lösung, wenn dies prinzipiell überhaupt möglich ist, noch viele Jahre in Anspruch nehmen wird. Vor diesem komplexen Hintergrund werden auch bei technischen Umsetzungen, die den Menschen zur Information, Bildung oder Unterhaltung dreidimensionale Abbilder der realen Welt liefern, noch viele Probleme gelöst werden müssen um damit dann eine möglichst naturgetreue 3D-Wiedergabe erreichen zu können.

3. Erste Technische Umsetzungen

Die Erfindung der Perspektive

Schon immer haben die Menschen versucht auf Zeichnungen und Gemälden die Wirklichkeit so real wie möglich abzubilden also auch auf einer ebenen Fläche Tiefe zu erreichen. Schon im alten Griechenland soll diese Technik bekannt gewesen sein. Es wird jedenfalls berichtet, dass dort ein Maler durch ein Bühnenbild das er für ein Schauspiel des Dichters Aischylos angefertigt hatte, durch eine neue Maltechnik so große Wirklichkeitsnähe erreichte, dass er damit das Publikum in großes Entzücken versetzt haben soll. Doch handwerkliche Kunst galt in der griechischen Polis nicht viel. Dieses Wissen geriet jedenfalls wieder in Vergessenheit. Bis ins Mittelalter wirkten alle Gemälde flach. Tiefe wurde nur durch Verkleinerung von Objekten im Hintergrund und Überdeckung angedeutet. Im Florenz des 15. Jahrhunderts erfolgte dann, wahrscheinlich angeregt durch die entstehende Kartographie, die Wieder- oder Neuentdeckung der Perspektive. Mittels der gedachten Sehlinien, die alle auf den Fluchtpunkt zulaufen und dort konvergieren konnte nun der Eindruck von Tiefe erreicht werden. Und so kam es, dass „La Dolce Prospettiva“ in der Renaissance fast alle Maler, Architekten, Gelehrte und reiche Dilettanten in ihren Bann schlug. Als Beispiel sei ein Maler der Frührenaissance angeführt, Paolo Uccello (1397-1475), den das Thema der Perspektive gefangengenommen hatte und der sich immer wieder intensiv mit der Raumdarstellung – der Darstellung der dritten Dimension auf zweidimensionaler Fläche - auseinandergesetzt hatte (Abb. 9).

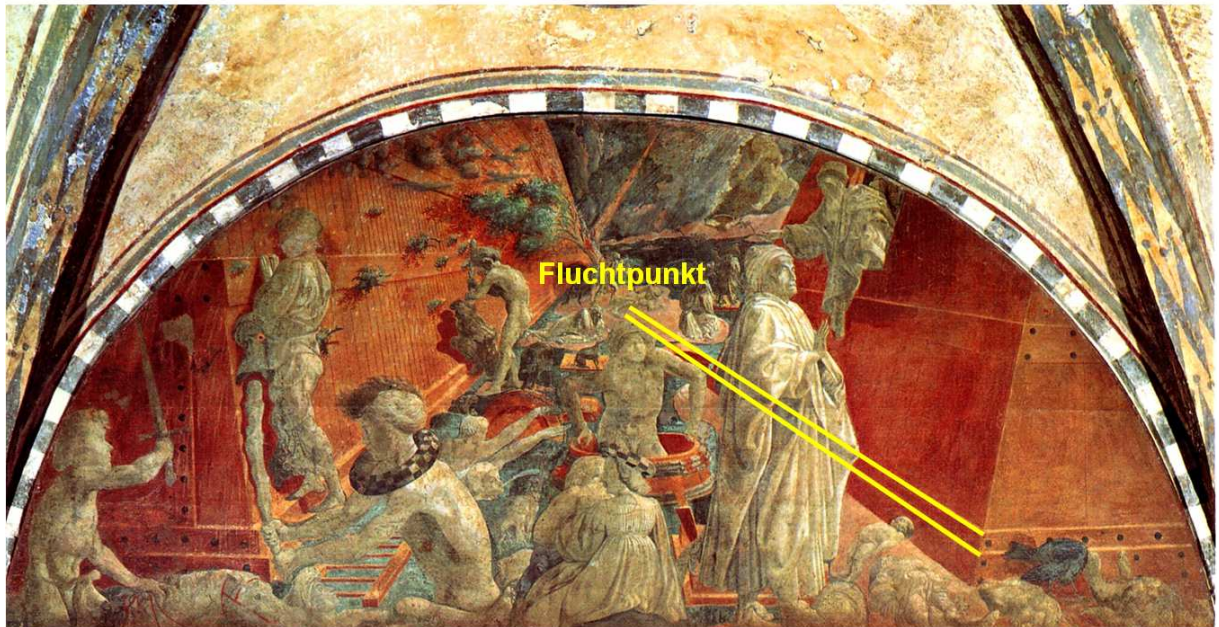


Abb. 9: Paolo Uccello (1397-1475) „Die Flut“ (Visipix.com)

Wheatstones Stereoskop

Es ist erstaunlich wie lange es gedauert hat bis eine Vorrichtung erfunden wurde, mit der das stereoskopische Sehen, basierend auf den leicht unterschiedlichen Bildern, die durch den Augenabstand entstehen, mit künstlichen Bildern nachgebildet und untersucht werden konnte. Die Erfindung eines solchen Gerätes muss Sir Charles Wheatstone zugeschrieben werden. Ende 1832 hatte er zwei solcher Geräte, Wheatstone nannte sie „Stereoscopes“, bei Optikern in London anfertigen lassen. Eines davon war ein Spiegelstereoskop, das andere ein Prismenstereoskop. In Abb. 10 ist die Prinzipskizze des Spiegelstereoskops nach einer Zeichnung von Wheatstone wiedergegeben.

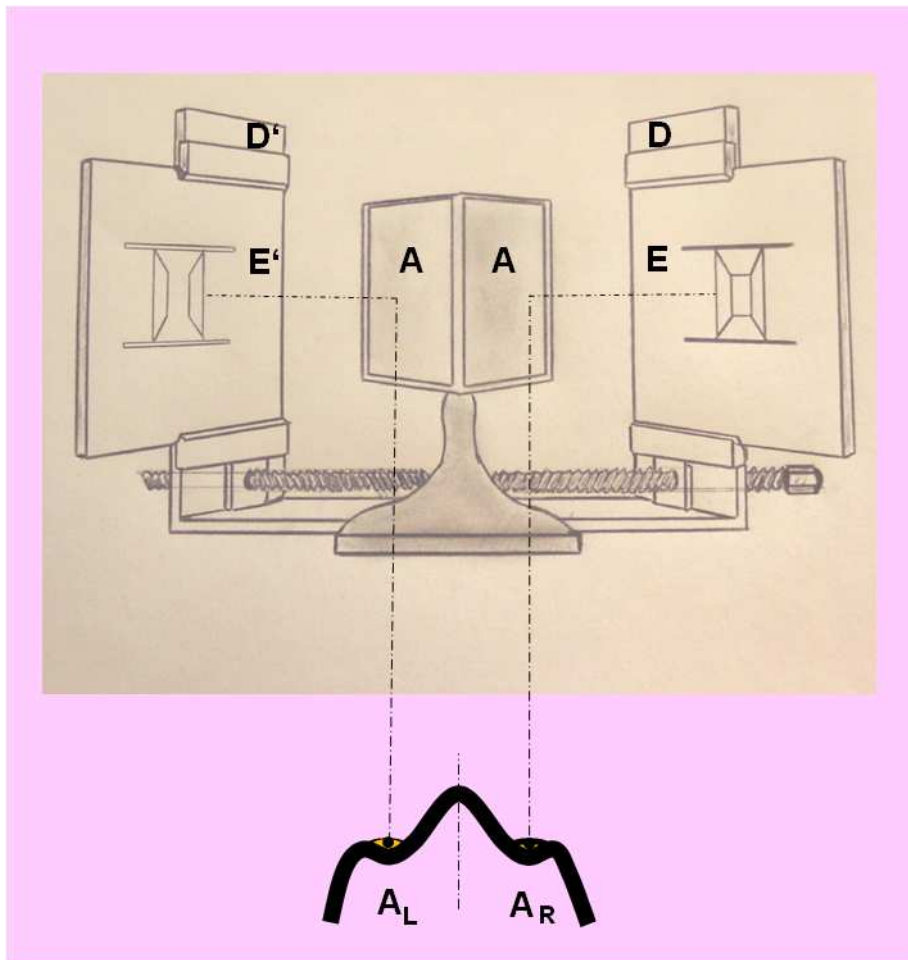


Abb. 10: Wheatstones Stereoskop.
(Eigene Zeichnung nach Gregory R.L. 2001 *Auge und Gehirn; Psychologie des Sehens*)

Es war ein einfaches Instrument welches dem linken und rechten Auge mittels zweier im rechten Winkel angebrachter Spiegel (A in Abb. 10) zwei getrennte 3D-Stereopaarzeichnungen mit entsprechender Disparation (E' und E in Abb. 10) angeboten werden konnten, die in die Bildhalter (D' und D in Abb. 10) eingespannt wurden. Das Wesentliche jedes Stereoskops ist, dass es erlaubt die Bildvorlagen für jedes Auge getrennt zu kontrollieren. Der Experimentator kann dadurch binokulare Variablen isolieren und die entsprechenden Effekte untersuchen. Wheatstone konnte mit diesem neuen Instrument erstmals die Beziehung zwischen binokularer Disparität und Tiefenwahrnehmung aufzeigen.

Louis Daguerre perfektionierte um die Jahre 1837 bis 1839 das Verfahren auf Metallplatten Abbildungen der Umwelt festzuhalten (Daguerreotypien). Fox Talbot stellte 1840 das Prinzip vor, wie mit Hilfe von transparenten Negativen beliebig viele Positive hergestellt werden konnten. Diese Technik war aber zunächst bezüglich der Bildqualität der Daguerreotypie unterlegen, die aber ihrerseits immer nur die Herstellung eines Originals erlaubte.

Diese Entwicklungen führten dazu, dass Wheatstone 1841 zwei Photographen Richard Beard und Henry Collen beauftragen konnte ihm bei der Erstellung der ersten Stereophotographie zu helfen. Diese war eine Porträtaufnahme von Charles Babbage, dem Erfinder der ersten programmierbaren mechanischen Rechenmaschine.

Diese ersten Stereophotographien wurden aufgenommen indem eine einfache Kamera um den Augenabstand verschoben wurde. Der Photograph A. Quinet baute 1853 in Paris die erste binokulare Stereokamera, die er Quinétoskope nannte.

Sir David Brewster brachte das von ihm entwickelte Prismenstereoskop nach Paris und interessierte den Optiker Jules Duboscq, der eine größere Anzahl davon baute. Dazu lieferte er einen Satz von Stereodaguerreotypien. Dieses Stereoskop wurde auf der Weltausstellung 1851 in London gezeigt. Eines davon erhielt Queen Victoria, die großes Interesse an dem Gerät zeigte. Innerhalb von 3 Monaten wurden nahezu 250 000 Prismenstereoskope in London und Paris verkauft. Der erste Markterfolg eines 3D-Wiedergabegerätes war geboren.

Mit der Einführung von illustrierten Magazinen erlosch das breite Interesse an stereoskopischen Photographien.

Stereoskopische Bewegtbildwiedergabe

Nach vielen Vorarbeiten durch Wissenschaftler und Ingenieure haben wohl Louis Le Prince und William Dickson, der im Thomas Edison Labor arbeitete, die ersten praktisch umsetzbaren und noch heute verwendeten Prinzipien der Aufnahme und Wiedergabe von bewegten Bildern basierend auf Mechanik und Photochemie entwickelt. In dieser Stufe glaubte aber Edison nicht an die Zukunftsfähigkeit von Kinoprojektoren, die große Bilder für viele Zuschauer auf eine Projektionswand werfen könnten.

So kam es dass die Gebrüder Louis und Auguste Lumière den ersten funktionsfähigen Kinoprojektor vorstellen konnten. Diese benutzten einen Zelluloidfilm von Eastman. Das erste Lichtspieltheater wurde im Dezember 1895 in Paris eröffnet. 1903 zeigten die Gebrüder Lumière in Frankreich den offiziell ersten Stereofilm von ein paar Sekunden Dauer. Er hatte den Titel „L'arrivée du train“. Er wurde offenbar auf zwei 35 mm Filmstreifen aufgenommen und dann auf einen Filmstreifen in anaglyphischer Zwei-Farbtechnik kopiert. Dies wird aber von manchen Experten, die der Meinung sind, dass zwei Projektoren mit entsprechenden Farbfiltern benutzt wurden, bezweifelt.

1915 brachte die Famous Players Film Company (die spätere Paramount Picture Company) drei kurze anaglyphische Stereofilme heraus, die der amerikanische Filmpionier Edwin Porter produziert hatte. Harry K. Fairall produzierte den ersten kommerziell erfolgreichen Stereofilm „The Power of Love“, der im Ambassador Hotel Theatre in Los Angeles im September 1922 gezeigt wurde. Alle diese Filme benutzten Anaglyphentechnik, d.h. ähnlich wie beim Stereoskop wurde hier mit Farbfilterbrillen dafür gesorgt, dass das linke Auge nur die Aufnahme der linken Kamera, und das rechte Auge nur die Aufnahme der rechten Kamera sehen konnte. Die beiden Filmprojektoren für den Links- und Rechtsauszug waren mit den korrespondierenden Farbfiltern versehen. Die Anaglyphentechnik erlaubt aber aufgrund ihres Prinzips nur eine stark verfälschte Farbwiedergabe.

Edwin Land demonstrierte 1935 einen Stereofilm in Farbe unter Benutzung des Prinzips der Polarisierung von Licht zur Separierung der Links-/Rechtsbilder. Dieser Prozess war in den 30iger Jahren bei Zeiss-Ikon in Deutschland entwickelt worden (Howard 2002). Raymond und Nigel Spottiswoode produzierten für das British Film Institute den ersten 3D-Film ganz in Farbwiedergabe und mit synchronisiertem Stereoton. Dieser wurde 1951 zur Eröffnung der South Bank Exhibition in London gezeigt. Dies kann als der Start des 3D-Booms von 1953 und 1954 angesehen werden.

4. Ausblick

Ich hoffe, dass es mir gelungen ist, die physiologischen und geometrischen Hintergründe des räumlichen Sehens etwas genauer darzulegen. Damit sollte es auch möglich sein die 3D-Verfahren, die sich heute bei 3D-Fernsehgeräten etabliert haben, besser zu verstehen. Mein Technik-Essay „Quo vadis 3D?“ versucht die heute auf dem Markt befindlichen Verfahren verständlich zu beschreiben, sowie deren Vor- und Nachteile herauszuarbeiten. Sie finden es unter <http://www.gisela-und-konrad-maul.de/3D.htm>. Damit sollte es Ihnen auch möglich sein, unabhängig von den üblichen Marketing-Aussagen der Hersteller, das für Sie geeignete 3D-Fernsehgerät zu finden. Bleibt mir noch Ihnen dabei viel Erfolg und letztendlich viel Freude beim 3D-Fernsehen zu wünschen.

Literaturverzeichnis

- Eccles, J.C.** **Die Psyche des Menschen.
Das Gehirn-Geist-Problem in neurologischer Sicht.
Serie Piper, München, 1990**
- Gregory, R.L.** **Auge und Gehirn Psychologie des Sehens
Rowohlt Taschenbuch Verlag, Hamburg, 2001**
- Hayes, R.M.** **3D-Movies.
A History and Filmography of Stereoscopic Cinema
McFarland & Co. Inc., Jefferson U.S., 1989**
- Howard, I.P.** **Seeing in Depth
Volume I Basic Mechanisms
University of Toronto Press, Toronto, 2002**
- Howard, I.P.** **Seeing in Depth
Volume II Depth Perception
University of Toronto Press, Toronto, 2002**
- Pöppel,E. (Ed.)** **Neuronal Mechanisms in Visual Restitution. Human Neurobiology,
Vol. 1, 1982**
- Pöppel,E.** **Lust und Schmerz.
Über den Ursprung der Welt im Gehirn.
Wilhelm Goldmann Verlag, München, 1995**
- Ponce, R.(et al.)** **Stereopsis.
Current Biology
Volume 18 No 18, 2008**
- Zeki S.** **Inner Vision.
An Exploration of Art and the Brain.
Oxford University Press, Oxford, 1999**

Anmerkung: Alle Zitate der englischsprachigen Originale wurden vom Autor ins Deutsche übertragen.



Konrad L. Maul
Dipl.-Ing. (FH)
Certified Counsellor

Konrad L. Maul war 37 Jahre in der Fernsehentwicklung tätig, davon 30 Jahre in leitender Position. Als Gruppenleiter war er für das erste 100 Hertz-TV-Gerät verantwortlich. Von 2001 bis 2008 leitete er die Fernsehentwicklung eines großen deutschen TV-Herstellers.

Damit ist er einer der erfahrensten und profiliertesten TV-Entwickler Deutschlands.

Heute arbeitet er am Aufbau einer Tätigkeit als Berater für Einzelne, Gruppen und Organisationen in technischen, wirtschaftlichen und sozialen Handlungsfeldern.

<http://www.gisela-und-konrad-maul.de>

URHEBERRECHT / COPYRIGHT

Alle Rechte vorbehalten, all rights reserved, 2010, 2012

Konrad L. Maul, Dipl.-Ing. (FH)

90473 Nürnberg

Dieses Technik-Essay, sein Inhalt und die enthaltenen Ideen sind urheberrechtlich geschützt. Sie dürfen nur zu privaten, wissenschaftlichen und nichtgewerblichen Gebrauch zum Zweck der Information kopiert und ausgedruckt werden. Der Autor behält sich das Recht vor, diese Erlaubnis jederzeit zu widerrufen. Ohne vorherige schriftliche Genehmigung des Autors darf dieses Dokument nicht vervielfältigt, archiviert, oder auf einem Server gespeichert, in Webseiten und Newsgruppen einbezogen, in Online-Diensten benutzt oder auf einer CD-ROM gespeichert werden.

Ich erlaube ausdrücklich und begrüße das Zitieren aus diesem Dokument.