Die Welt in unseren Köpfen

Heinrich H. Bülthoff

Wie nimmt unser Gehirn die Welt wahr? Wie erfassen wir dreidimensionale Gegenstände? Wie erkennen wir ein menschliches Gesicht? Das sind einige der Fragen, mit denen wir Wahrnehmungsforscher und Psychophysiker uns beschäftigen. So alt und so komplex diese Fragen sind, so haben wir doch bei ihrer Beantwortung Fortschritte gemacht. Dieses neu erworbene Wissen befriedigt nicht nur unsere wissenschaftliche Neugier. Es lässt sich auch in technische Systeme umsetzen, die vielleicht einmal Erkennungsleistungen vollbringen können, die genau so gut sind wie die des Menschen – oder sogar besser.

Wahrnehmung ist Rekonstruktion

Von den fünf Sinnen des Menschen ist das Sehsystem am besten erforscht. Das visuelle System ist dasjenige Sinnessystem, das ungefähr die Hälfte unseres Gehirns beansprucht.
Nach David Marr, dem Vater der modernen Sehforschung, kann man die visuelle Informationsverarbeitung des Menschen in drei Ebenen[1] unterteilen (Marr 1982):

- Extraktion:
  Auf der untersten Ebene werden aus dem Bild, welches durch die Linse auf die Netzhaut im Auge geworfen wird, verschiedene Informationsgehalte extrahiert, zum Beispiel Helligkeit, Farbe, Bewegung, aber auch Tiefeninformation.
- Integration:
  Auf der mittleren Ebene werden diese verschiedenen Informationsquellen zusammengefasst, um eine möglichst fehlerfreie Interpretation der Welt zu erreichen. Unser Gehirn verlässt sich nicht nur auf eine Quelle, sondern nutzt sämtliche Informationen, die das Auge zur Verfügung stellt. Wenn unser Gehirn nicht genügend Informa-

[1] Diese Ebenen sind logisch zu verstehen, nicht räumlich. Sie entsprechen nicht drei Schichten im Gehirn. Das Gehirn nutzt für die Aufgaben aller drei Ebenen eine Vielzahl spezialisierter Regionen, die „visuellen Zentren" im Zwischenhirn und im Großhirn.

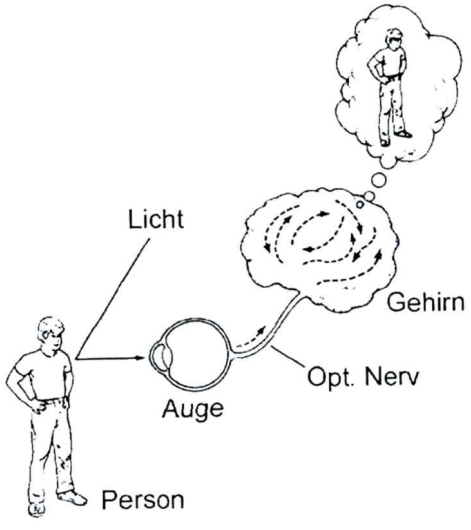


Abb. 1: Vom Auge ins Gehirn. Lichtstrahlen fallen auf ein Objekt (hier eine Person) und werden von diesem reflektiert. Die Linse im Auge des Betrachters sammelt die Lichtstrahlen und projiziert ein Bild auf die Netzhaut des Auges. Die Lichtenergie wird durch die Photorezeptoren in der Netzhaut in elektrische Signale umgewandelt und durch den optischen Nerv ins Gehirn geleitet. Durch den Vergleich mit gespeicherten Repräsentationen von vorher gesehenen Objekten können diese wieder erkannt werden und eine Vorstellung im Kopf des Betrachters entstehen.

tion zur Verfügung hat, entstehen optische Täuschungen oder Illusionen, wie wir später sehen werden.

- Repräsentation:
  Auf der höchsten Verarbeitungsebene werden diese Informationen dann mit einer internen Repräsentation in unserem Gedächtnis verglichen. Erst dieser Vergleich erlaubt es uns, etwas wieder zu erkennen („Re-Cognition"), das heißt „wahrzunehmen".

Nach dieser These erfolgt unsere Wahrnehmung der wirklichen Welt also nicht unmittelbar, sondern sie ist eine Rekonstruktion der Wirklichkeit. Diese Idee ist nicht neu. Auch der griechische Philosoph Plato hat sie in seinem berühmten „Höhlengleichnis" vertreten:[2]

> Gefesselt, mit dem Rücken gegen den Höhleneingang, erblickt der Mensch nur die Schatten der Dinge, die er für die alleinige Wirklichkeit hält. Löste man seine Fesseln und führte ihn aus der Höhle in die lichte Welt mit ihren

[2] Zitiert nach Theodor Ballauff: Die Idee der Paideia: eine Studie zu Platons „Höhlengleichnis" und Parmenides' „Lehrgedicht". Westkulturverlag, Meisenheim: 1952.

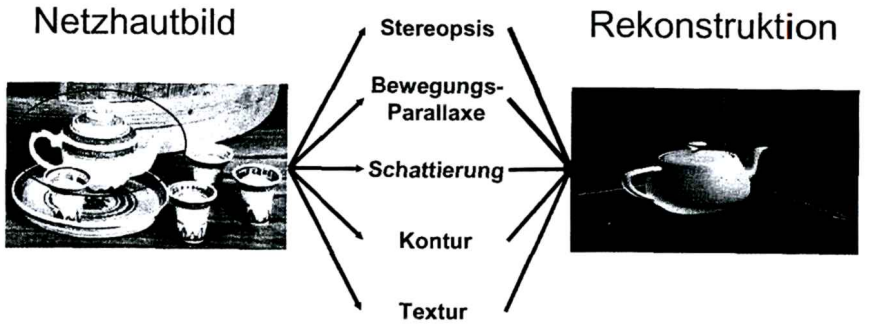


Abb. 2: Module des Sehens. Unser visuelles System extrahiert verschiedene Informationsgehalte aus den Signalen der Photorezeptoren. Im Bild sind die Verarbeitungsmodule benannt, die für das räumliche Sehen und die Formerkennung wichtig sind: Stereopsis (durch Vergleich der leicht unterschiedlichen Netzhautbilder der beiden Augen kann die Distanz von Merkmalen in der Szene berechnet werden), Bewegungsparallaxe (ähnlich zur Stereopsis erlaubt die zeitliche Veränderungen des Netzhautbildes durch Bewegung, Form und Distanz von Objekten zu berechnen), Schattierung, Kontur (Umriss) und Textur (Oberflächenbeschaffenheit).

> wirklichen Dingen, so würden ihm zuerst die Augen wehtun, und er würde seine Schattenwelt für wahr, die wahre Welt für unwirklich halten. Kehrte er aber in die Höhle zurück, um die anderen Menschen aus ihrer Haft zu befreien und von ihrem Wahn zu erlösen, so würden sie ihm nicht glauben, ihm heftig zürnen und ihn vielleicht sogar töten.

Zum Glück müssen wir Wahrnehmungsforscher von heute nicht mehr befürchten, tätlich angegriffen zu werden. Aber im Grunde behaupten wir das Gleiche wie Plato: dass der Mensch sozusagen in einer Schattenwelt lebt. Wir leben in einer Welt, die wir uns aus den Bildern rekonstruieren, die die Linse unseres Auges auf die Netzhaut projiziert.

Nehmen wir zum Beispiel eine Teekanne wie in Abbildung 2. Wie rekonstruiert unser Gehirn aus dem zweidimensionalen Netzhautbild der Teekanne (links) ein „inneres Bild" der Teekanne (rechts)?

Wie wir wissen, liefert die reale Teekanne unserem optischen System viele verschiedene Informationsgehalte oder -module. Darunter sind

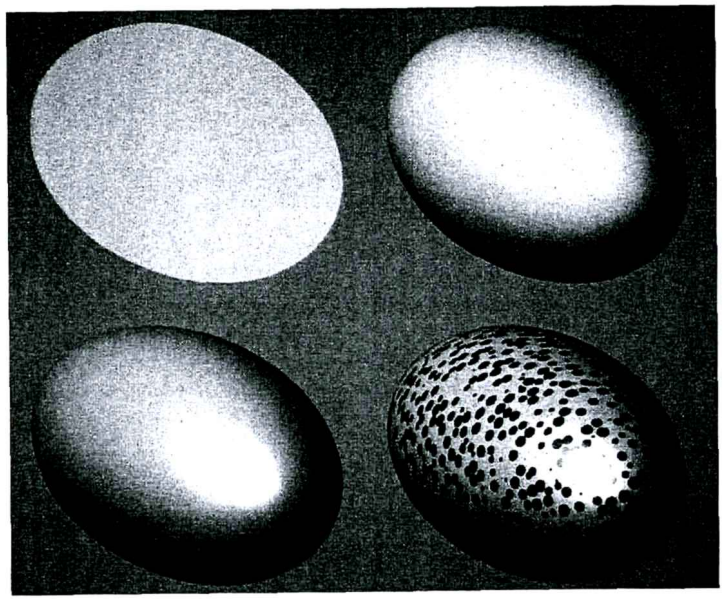


Abb. 3: Gefleckt und schattiert. Die Kontur allein ergibt keinen guten Eindruck von Form und Orientierung dieser eiförmigen Objekte. Die Flecken-Textur und die Lichtreflexe (Schattierung) helfen dem Auge auf die Sprünge. Wenn alle Module zusammenkommen, nehmen wir die Form am besten wahr.

auch Module, die uns das räumliche Sehen ermöglichen: Schattierung, Kontur und Textur tragen zusammen mit dem beidäugigen Sehen (Stereopsis) und den Veränderungen, die beim Bewegen des Gegenstands oder des Betrachters entstehen (Bewegungsparallaxe), dazu bei, dass wir die dreidimensionale Struktur der Teekanne rekonstruieren können.

Wie unser Gehirn diese Informationsmodule einsetzt, haben Wahrnehmungsforscher mit raffinierten Methoden untersucht. Das Prinzip dabei: Die Psychophysiker reduzieren in ihren Experimenten die Welt auf einzelne Module. Solche sehr speziellen visuellen Reize bieten sie dann ihren Versuchspersonen an und messen deren Wahrnehmungsleistungen. In Abbildung 3 sehen Sie, wie die gleiche Form durch Reduzierung auf verschiedene Module der Formwahrnehmung (Kon-

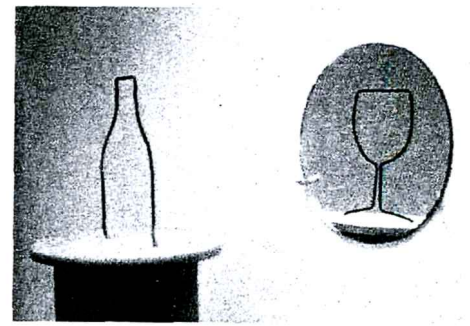


Abb. 4: Wie kann ein und dieselbe Drahtskulptur gleichzeitig wie ein Glas und wie eine Flasche (im Spiegelbild) aussehen? Die Skulptur ist nicht flach sondern so verbogen, dass je nach Blickwinkel ein unterschiedliches Bild der Kontur entsteht. Die Skulptur ist ein Werk des Schweizer Künstlers Markus Raetz (1994).

tur, Schattierung, Texturierung) ganz unterschiedlich wahrgenommen werden kann.

Allerdings wird durch diese Reduktion auf einzelne Module die Wahrnehmung unter Bedingungen untersucht, unter denen das Gehirn normalerweise nicht arbeitet. Unter solchen Bedingungen schlägt die Erkennungsleistung oftmals fehl und es kommt zu Illusionen, welche wiederum Aufschluss über die Grenzen der Wahrnehmung geben.

Markus Raetz, ein Schweizer Künstler, hat solche Illusionen in überraschende Kunstwerke umgesetzt. Wenn unserem Gehirn etwa bei einer Drahtskulptur wie der in Abbildung 4 nur die Kontur-Information zur Verfügung steht, ist die Interpretation nicht eindeutig. So gibt es hier mindestens zwei sinnvolle Interpretationen: Wenn man um die Skulptur herumgeht, kann man bei einer bestimmten Blickrichtung eine Flasche sehen (links) und im Spiegelbild, das eine andere Blickrichtung zeigt, ein Glas (rechts).

Um diesen Effekt zu erreichen, musste der Künstler den Draht recht trickreich verbiegen, wobei er nicht nur zwei, sondern drei Dimensionen nutzte. Da aber ein verbogener Draht kaum Tiefenhinweise enthält (zumal wenn der Betrachter still steht und die Bewegungsparallaxe fehlt), fällt einem die Verbiegung in der Tiefe kaum auf – man nimmt nur die zweidimensionale Projektion wahr.[3]

[3] Auf der Website des Max-Planck-Instituts für biologische Kybernetik können Sie in einer Computer-Animation einer ganz ähnlichen Drahtfigur besser verstehen, wie ein Draht verbogen werden muss, um so unterschiedliche, aber auch sinnverwandte (Flasche und Glas) Interpretationen zu ermöglichen: www.kyb.mpg.de/links/metamorphosis.html

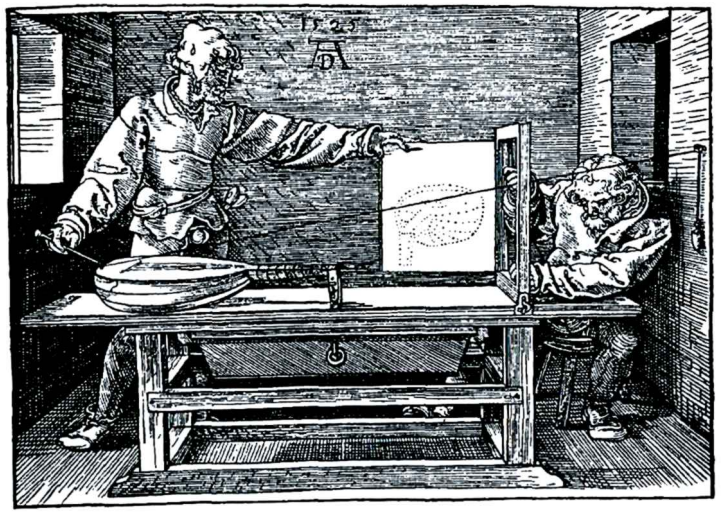


Abb. 5: Aus rund mach flach. Der Künstler in Albrecht Dürers Holzschnitt hat eine vergleichsweise einfache, weil eindeutige Aufgabe zu lösen: Er soll ein dreidimensionales Objekt (eine Laute) zweidimensional zu Papier bringen. Dazu bedient er sich der Projektionsstrahlen – der Strahlen, die von der Laute durch die Zeichenebene auf das Projektionszentrum auf der gegenüberliegenden Wand fallen. So entsteht (genau so wie durch die Linse im Auge) ein zweidimensionales Bild der dreidimensionalen Welt.

Von 3D nach 2D und zurück

Was unser Gehirn bei der Rekonstruktion der Welt zu bewältigen hat, ist ein noch viel schwierigeres Problem als das, welches Dürer in seinem Holzschnitt (Abbildung 5) dargestellt hat: Hier bedient sich ein Künstler der Projektionsstrahlen, um ein dreidimensionales Objekt (eine Laute) auf das zweidimensionale Zeichenblatt abzubilden.
Beim Sehen müssen wir die umgekehrte Leistung vollbringen: nämlich aus der zweidimensionalen Abbildung auf unser Netzhaut eine drei-

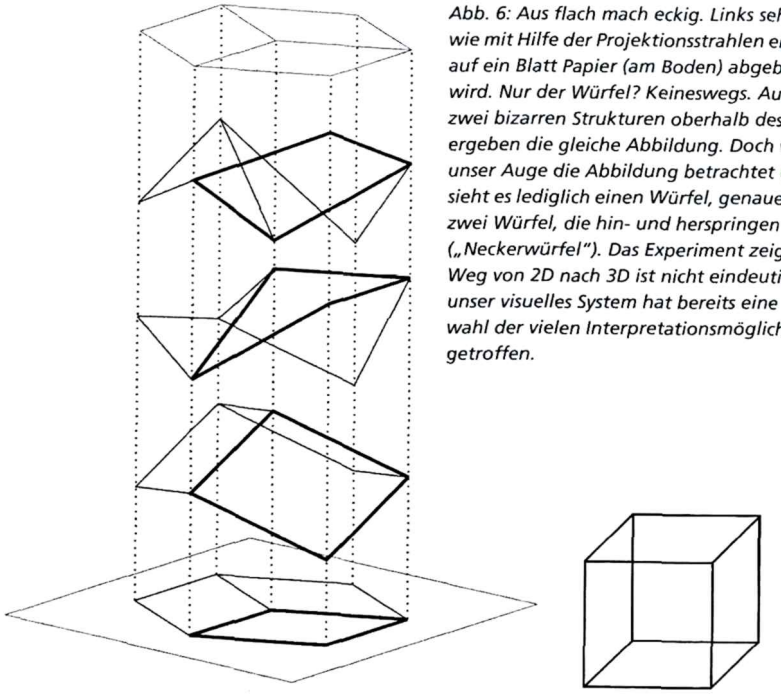


Abb. 6: Aus flach mach eckig. Links sehen Sie, wie mit Hilfe der Projektionsstrahlen ein Würfel auf ein Blatt Papier (am Boden) abgebildet wird. Nur der Würfel? Keineswegs. Auch die zwei bizarren Strukturen oberhalb des Würfels ergeben die gleiche Abbildung. Doch wenn unser Auge die Abbildung betrachtet (rechts), sieht es lediglich einen Würfel, genauer gesagt: zwei Würfel, die hin- und herspringen („Neckerwürfel"). Das Experiment zeigt: Der Weg von 2D nach 3D ist nicht eindeutig – doch unser visuelles System hat bereits eine Vorauswahl der vielen Interpretationsmöglichkeiten getroffen.

dimensionale Welt erschaffen. Wir brauchen eine solche innere Repräsentation, die räumlich ist, damit wir uns in der realen Welt bewegen können, ohne zum Beispiel mit Hindernissen zu kollidieren, aber auch, um Objekte oder Personen zu erkennen.
Die Rekonstruktion der räumlichen Welt aus einer flächenhaften Abbildung ist deshalb so schwierig – eigentlich sogar ein unlösbares Problem –, da es prinzipiell unendlich viele Objekte gibt, die die gleiche Abbildung erzeugen. Wir stehen hier vor einem „unterbestimmten" Problem: einem Problem, das keine eindeutige Lösung hat. Nur durch angeborenes oder erlerntes Vorwissen über die Welt kommen wir überhaupt zu sinnvollen inneren Bildern.
In Abbildung 6 sehen Sie, wie ein Würfel entlang seiner Projektionsstrahlen auf eine Ebene abgebildet wird. Das ist trivial. Sie können nun

aber leicht einzelne Eckpunkte des Würfels entlang der Projektionsstrahlen verschieben, wie es im oberen Teil der Zeichnung gezeigt ist. Dabei kommen Sie zu recht bizarren Gebilden, die überhaupt nicht mehr wie ein Würfel aussehen. Das Interessante dabei: An der zweidimensionalen Abbildung „am Boden" ändert sich gar nichts. Sie ist das Abbild all der Figuren, die Sie gerade konstruiert haben.

Nehmen wir nun die zweidimensionale Abbildung selbst (rechts daneben) als Ausgangspunkt. Was sehen Sie? Sie sehen einen Würfel, der von Zeit zu Zeit „umspringt": mal scheint Ihnen die Fläche links oben, mal die Fläche rechts unten als „Vorderwand" entgegen zu kommen, während die jeweils andere als „Rückwand" zurückweicht. Diese visuelle Illusion heißt „Neckerwürfel" – nach dem Schweizer Kristallographen Louis Albert Necker, der sie zuerst beschrieben hat. Den meisten von Ihnen wird die Zweideutigkeit des Neckerwürfels vertraut sein. Was aber selbst vielen Wahrnehmungsforschern nicht klar ist: Es ist ja geradezu ein Wunder, dass unser Gehirn aus der unendlichen Anzahl möglicher Interpretationen nur zwei auswählt, zwischen denen es sich nicht entscheiden kann. Unser Gehirn wählt die Interpretationen, die in der 3D-Rekonstruktion eine möglichst geringe Variation der Winkel erzeugen, ebene Flächen und eine kompakte Form. Die bizarren Varianten aus der Projektionszeichnung (links) blendet es einfach aus. Dieser Effekt ist unter Gestaltpsychologen als „Gesetz der guten Form" bekannt.

Kommen wir zu einem weniger bekannten Beispiel für die Mehrdeutigkeit in der Interpretation von Bildern (Abb. 7): Der Fußabdruck (links) ist auf einer Postkarte abgebildet, die mir eine Studentin aus dem Urlaub geschickt hat. Und da unsere Studenten fleißig sind und auch im Urlaub arbeiten oder zumindest eifrig nachdenken, hat sie mich darauf hingewiesen, die Karte einmal andersherum zu betrachten (rechts).

Hier haben wir die gleiche Postkarte in zwei verschiedenen Orientierungen. Das heißt: Beide Bilder haben die gleiche Bildinformation. Was sich aber mit der Umkehr der Postkarte geändert hat, ist die Richtung, aus der der Fußabdruck beleuchtet wird. Durch das Umdrehen haben wir quasi eine Beleuchtung von unten eingeführt, was man gut an den Schatten sehen kann, die die Muscheln werfen.

Eine Beleuchtung von unten entspricht allerdings nicht unserer täglichen Erfahrung. Tatsächlich interpretieren wir daher auch das rechte Bild so, als ob das Licht von links oben käme. Folgerichtig sehen wir

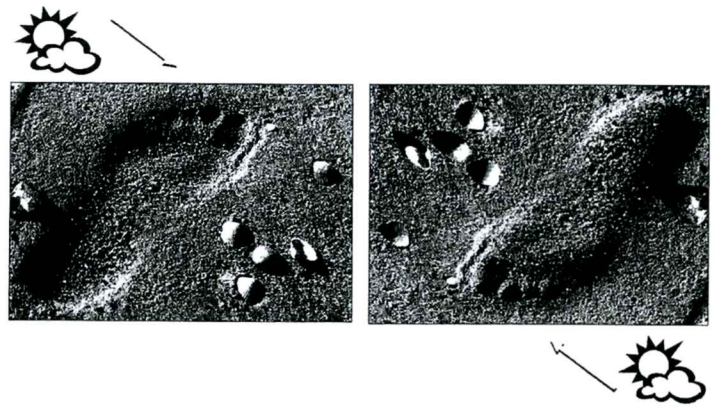


Abb. 7: Ein Fuß steht Kopf. Ob der Fußabdruck ganz normal als Abdruck im Sand (links) oder als Fußskulptur (rechts) gesehen wird, hängt von der Orientierung der Abbildung und damit der Beleuchtung ab. Bei dem auf den Kopf gestellten Foto (rechts) kommt das Licht von unten. Das ist für unser Auge allerdings so ungewohnt, dass es automatisch eine Beleuchtung von oben annimmt („von oben strahlt die Sonne"); unter dieser Annahme ist nur eine Fußskulptur mit der Richtung der Schatten im Bild kompatibel.

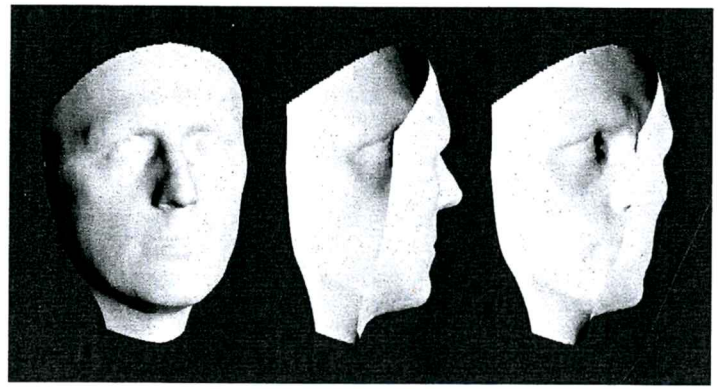


Abb. 8: Ein hohles Gesicht. Wenn eine Hohlmaske langsam rotiert, blicken wir irgendwann in ein Hohlgesicht – doch es will uns nicht gelingen, es so zu sehen! Dass bei einem Gesicht die Nase immer dem Betrachter entgegen ragt und nicht von ihm weg, ist eine Annahme unseres Gehirns, die durch lange Erfahrung gefestigt wurde. Schließlich begegnen uns Hohlgesichter im Leben so gut wie nie!

eine Erhebung, die einen Schatten am rechten Rand wirft. Eine Vertiefung wäre mit einer Beleuchtung von links oben nicht konsistent.
In unsere 3D-Interpretation eines Bildes geht also unsere Erfahrung ein, z. B. Vorwissen über die übliche Richtung der Beleuchtung. Dies ist sogar essentiell, da wir ohne zusätzliche Annahmen keine eindeutige Lösung des Rekonstruktionsproblems von 2D nach 3D erwarten können.
In der Bilderserie[4] in Abbildung 8 sehen Sie ein Beispiel für eine sehr starke Annahme, die unser Gehirn über die Geometrie von Objekten macht. Wie schon am Fußabdruck gezeigt, kann man eine schattierte Oberfläche konvex oder konkav sehen. Wenn wir nun eine rotierende Maske aus größerer Entfernung oder einäugig betrachten (so dass stereoskopische Information über deren Tiefenstruktur keine Rolle spielt), dann haben wir große Schwierigkeiten, im Inneren der Maske ein „Hohlgesicht" zu sehen; dieses „springt" unweigerlich um und wird zu einem gewöhnlichen Gesicht.
Warum? Nun, wir sehen normalerweise nur Gesichter, bei denen die Nase dem Beobachter entgegen ragt und nicht in das Gesicht hinein. Damit ist die a-priori-Wahrscheinlichkeit für eine „Hohlkopf"-Interpretation sehr gering und wird deshalb auch nicht gesehen.
Diese Abwägung kann man auch mathematisch exakt beschreiben. Mit der Wahrscheinlichkeitstheorie von Thomas Bayes (1702–1761)[5] können wir den Schlussmechanismus von 2D-Bildern zu 3D-Szenen formalisieren. Solche Formeln sind für Forscher hilfreich, um komplexe Vorgänge wie das Sehen besser zu verstehen; aber auch, um künstliche Sehsysteme oder Sehhilfen zu bauen.

[4] Einen kurzen Film der rotierenden Maske können Sie auf der MPI-Homepage finden: www.kyb.mpg.de/links/rotating-mask.html

[5] Die Bayes'sche Formel lautet: P(Szene | Bild) = P(Bild | Szene) · P(Szene). Dabei bedeuten die einzelnen Elemente:
P(Szene | Bild): Bedingte Wahrscheinlichkeit einer Szene bei einem gegebenen Bild
P(Bild | Szene): Bedingte Wahrscheinlichkeit eines Bildes bei einer gegebenen Szene (entspricht der Physik der Abbildung)
P(Szene): a-priori-Wahrscheinlichkeit der Szene (Annahmen über Geometrie, Licht, Material).

Abb. 9: Ein Stuhl ist ein Stuhl. Versuchen Sie doch einmal, die Stühle auf diesem Bild zu zählen – fällt Ihnen das schwer? Wahrscheinlich nicht, denn Sie wissen doch, wie ein Stuhl aussieht. Bis heute gibt es kein Computerprogramm, dass in diesem oder ähnlichen Bildern alle Stühle finden kann, denn dazu muss es viele Probleme lösen, die in diesem Bild illustriert sind: Orientierungs- und Größeninvarianz, Verdeckung (durch Schreibtisch oder Paravent), Schatten oder Bild eines Stuhles sind keine Stühle, unterschiedliche Stuhlarten, zerbrochener Stuhl, Stuhl an der Decke.

Wie wir Objekte erkennen

Die Computergraphik Abbildung 9 hängt in meinem Büro an prominenter Stelle, um mich stets daran zu erinnern, dass das Erkennen von Objekten nicht trivial, sondern eine ganz besondere Gabe ist. Wir Menschen haben keine Schwierigkeiten, auf diesem Bild sämtliche Stühle zu identifizieren, ob sie nun groß sind oder klein, verdeckt oder gespiegelt, klassisch oder außergewöhnlich geformt. Ein Computer dagegen wäre mit dieser Aufgabe hoffnungslos überfordert.
Doch wie machen wir das eigentlich? Wie setzen wir unsere Informationen über die sichtbare Welt, die uns physikalisch als elektrische Aktivität in den Photorezeptoren zur Verfügung stehen, in innere Bilder um, mit denen wir neue Eindrücke vergleichen können?

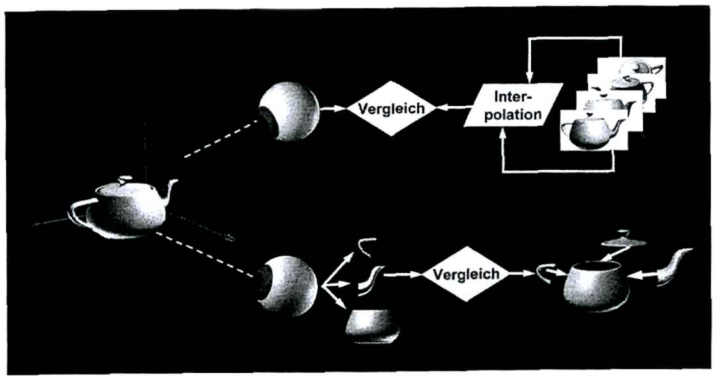


Abb. 10: Wie ist eine Teekanne im Kopf repräsentiert? Wenn wir eine Teekanne (links) wieder erkennen wollen, müssen wir sie mit dem inneren Bild vergleichen, das wir uns von einer Teekanne gebildet haben. Doch wie machen wir das? Haben wir verschiedene Ansichten der Kanne im Kopf (oberer Zweig) – so als hätten wir sie von allen Seiten bildhaft abgespeichert? Oder lagert in unserem Gehirn eine Art Modellbaukasten, dessen Elemente wir drehen und wenden und passend zusammenbauen können, um sie mit der gesehenen Teekanne zu vergleichen?

Es gibt natürlich eine beliebig große Anzahl von Möglichkeiten, wie Objekte im Gehirn abgespeichert sein könnten. Ich möchte hier nur die beiden wichtigsten Modellvorstellungen gegenüberstellen (Abbildung 10).

Beiden Repräsentationen der Teekanne ist gemeinsam, dass eine Vergleichsoperation zwischen einem 2D-Bild auf der Netzhaut und einer internen Repräsentation des zu erkennenden Objektes ausgeführt werden muss:

1. In einer modell-basierten Repräsentation (unterer Zweig) sind die Objekte als Strukturmodelle, ähnlich wie dreidimensionale Architekturmodelle, abgespeichert. Diese Objektrepräsentation besteht aus elementaren geometrischen Formen und deren räumlichen Relationen und ist deshalb weitgehend unempfindlich gegenüber räumlichen Transformationen des Objektes. Dieses Modell sagt voraus, dass die Erkennungsleistung unabhängig von der Blickrichtung ist, solange genügend Elemente des Objekts sichtbar sind. In Erkennungsexperimenten findet man diese Invarianzleistung allerdings nicht.
2. In einer bild-basierten Repräsentation (oberer Zweig) kann man auf solche abstrakten Repräsentationen verzichten: Denn hier ist jedes

Objekt, also jede jemals gesehene Teekanne, schon aus verschiedenen Richtungen bildhaft abgespeichert. Der Vorteil der bildbasierten Repräsentation ist der direktere und damit schnellere Vergleich zwischen Präsentation und Repräsentation. Allerdings verlangt eine bildbasierte Repräsentation eine viel größere Speicherkapazität, da jedes zu erkennende Objekt mit allen Ansichten, aus denen es erkannt werden soll, abgespeichert werden muss.

Der damit anwachsende Speicherbedarf ist ein beliebtes Argument gegen eine solche Repräsentation. Aber müssen wirklich alle Bilder und alle Bildpunkte in allen Ansichten abgespeichert werden, oder kann unser visuelles System eine Interpolation der gespeicherten Ansichten durchführen?

Die Bewegungsstudien des schwedischen Forschers Gunnar Johansson zeigen, dass wir auch nicht alle Bildpunkte abspeichern müssen. Johansson hat Personen in schwarzen Anzügen, an denen kleine Glühbirnen befestigt waren, vor einer schwarzen Leinwand verschiedene Bewegungen ausführen lassen. Seine Aufnahmen, in denen nur bewegte Lichtpunkte zu sehen sind, zeigen deutlich, dass wir auch mit sehr wenig Bildinformation Personen (oder auch Tiere) erkennen können, sobald sich diese Punkte in einer natürlichen Art und Weise bewegen.[6]

In unserem Labor haben wir viele Objekterkennungsexperimente durchgeführt, die für eine bildbasierte Repräsentation sprechen: Dabei bekamen Versuchspersonen Gegenstände, die sie vorher noch nie gesehen hatten, in einer bestimmten Ansicht gezeigt. Danach sollten sie die Objekte aus einer anderen Blickrichtung wieder erkennen – doch das gelang ihnen nicht (siehe Abbildung 11). Dabei stand den Versuchspersonen (objektiv gesehen) genügend Information zur Verfügung, dass sie aus einer Stereoansicht ein räumliches Strukturmodell des Gegenstands hätten ableiten können. Doch solche Rechenoperationen, bei denen eine mentale Rotation durchgeführt werden muss, fallen uns Menschen offensichtlich schwer oder dauern zu lange, um eine schnelle Objekterkennung zu ermöglichen. Diese Befunde aus der Psychophysik sprechen also gegen strukturelle Modelle (Bülthoff und Edelmann 1992).

[6] Vergleiche www.kyb.mpg.de/links/biological-motion.html

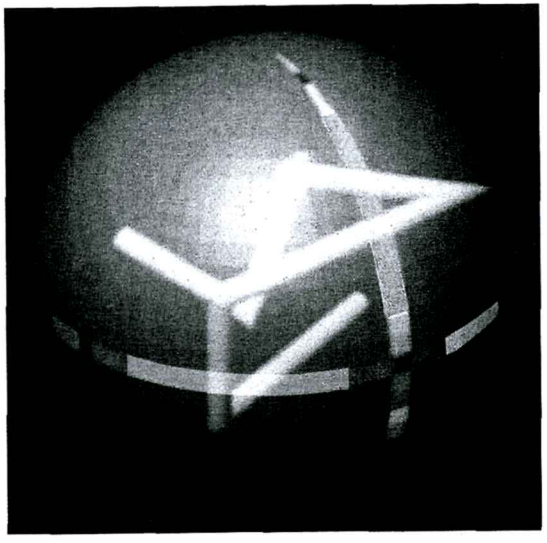


Abb. 11: Ansichtssache. Ein merkwürdig geformtes weißes Objekt, das einer verbogenen Büroklammer gleicht – wie sieht es wohl von hinten aus? Sie haben Schwierigkeiten, sich das vorzustellen? Kein Wunder. Bei einem Experiment bekamen Versuchspersonen das Objekt (das sie vorher nie gesehen hatten) aus zwei Blickwinkeln (rote Bereiche) zu sehen. Das war die Trainingsphase. Danach wurde ihnen das Objekt aus anderen Winkeln gezeigt. Ergebnis: Im gesamten grünen und blauen Bereich versagten die Versuchspersonen – sie konnten nicht sagen, ob sie die gleiche Büroklammer sahen wie im Training oder eine anders verbogene Büroklammer. Nur wenn der Blickwinkel zwischen den Trainingswinkeln lag (orangefarbener Bereich), waren sie sich sicher.

Aber auch Neurophysiologen haben gute Argumente für eine bildbasierte Repräsentation beizutragen: Sie haben nämlich im Affengehirn Nervenzellen gefunden, die nur dann antworten, wenn ein Objekt aus einer bestimmten Richtung gezeigt wurde (Logothetis, Pauls und Poggio 1995). So hat man zum Beispiel Zellen gefunden, die nur aktiv sind, wenn der Affe seinen Pfleger im Profil sieht. Sieht er ihn von vorn, sind sie inaktiv. Informatiker haben die Eigenschaften solcher ansichtsspezifischer Zellen im Computer simuliert und nachgewiesen, dass man aus ihnen Erkennungssysteme bauen kann, wenn man sie in so genannten „neuronalen Netzen" interagieren und Bild-Interpolationen ausführen lässt.
Zusammenfassend kann man sagen, dass wir auf drei Untersuchungsebenen – der Psychophysik, der Neurophysiologie und der theoretischen Modellierung – gute Hinweise für eine bildbasierte Repräsentation gefunden haben. Das heißt: Wir können Objekte wieder erkennen, ohne dass wir ihre dreidimensionale Struktur explizit im Gehirn abgespeichert haben. Die mentale Rotation eines 3D-Modells ist nicht notwendig.

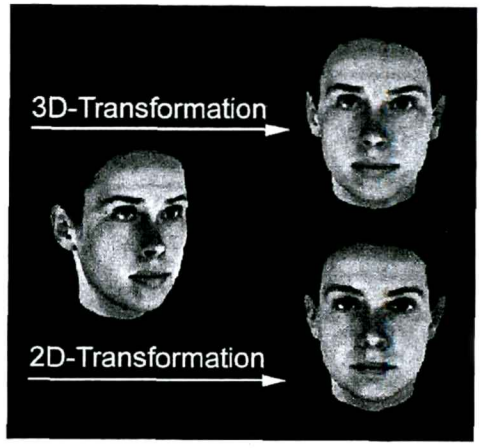


Abb. 12: Synthetische Gesichter. Man nehme das Foto einer Person (links) im Halbprofil. Kann man dann sagen, wie sie von vorne aussieht? Im MPI für biologische Kybernetik haben Volker Blanz und Thomas Vetter ein technisches System entwickelt, das die Vorderansicht „ausrechnen" kann, indem es eine 2D-Transformation ausführt. Allerdings muss Vorwissen in die Operation einfließen, damit etwa ein passendes linkes Ohr konstruiert werden kann. Das Ergebnis (rechts unten) ist jedenfalls der realen Vorderansicht (nach 3D-Rotation des Kopfes, oben) ziemlich ähnlich.

Künstliche Köpfe für Forschung und Technik

Zum Schluss möchte ich Ihnen zeigen, wie sich unser bildbasierter Ansatz in technische Systeme umsetzen lässt. Thomas Vetter und Volker Blanz haben in unserer Arbeitsgruppe ein technisches System entwickelt, das es erlaubt, aus einer einzigen Ansicht eines Gesichts eine neue, rotierte Ansicht zu generieren (Abbildung 12). Aus einem einzigen Foto einer Person (links) kann man somit beliebig viele neue Ansichten dieser Person synthetisieren (Blanz und Velter 1999). Diese synthetisierten Gesichter können zum Beispiel die Erkennungsleistung eines automatischen Gesichtserkennungssystems verbessern.
Von der Leistungsfähigkeit des Systems können Sie sich selbst überzeugen: Die Frontalansicht unten rechts wurde durch eine reine Bildtransformation, also durch eine 2D-Operation, erzeugt. (Wie Sie sehen, ist dieses Bild dem oberen rechten Bild sehr ähnlich, das durch eine echte 3D-Rotation des Kopfes gewonnen wurde.)
Eine solche 2D-Operation ist natürlich nur dann sinnvoll, wenn wir Vorwissen über das Aussehen von Köpfen in das System hineinstekken – nicht über die Form dieses speziellen Kopfes, die wir ja nicht haben, sondern ganz allgemein über das Aussehen von Köpfen. Wir

benutzen dazu eine eigene Gesichterdatenbank[7], in der wir Informationen über Form und Aussehen von vielen Gesichtern abgelegt haben.
Wie ist diese Datenbank aufgebaut? Um zu lernen, wie sich die Bilder von Köpfen verändern, wenn man sie aus unterschiedlichen Richtungen betrachtet, haben wir mit Hilfe eines Laserscanners zunächst die Geometrien und Ansichten von weit über 200 Köpfen gesammelt. In Abbildung 13 sehen Sie, wie ein Kopf eingescannt wird.
Im Gegensatz zu vielen anderen Gesichter-Datenbanken haben wir also nicht nur Fotografien gespeichert. Mit Hilfe des Scanners konnten wir von jedem Kopf zwei Datensätze gewinnen, einen in 3D und einen in 2D: Die Form-Information (3D) meines Kopfes finden Sie im oberen rechten Bild dargestellt und die „abgerollte" Textur (2D) unten. Durch Projektion der Textur-Information auf die 3D-Struktur und Rotation lassen sich nun beliebige Ansichten meines Kopfes im Computer erzeugen – von vorn, von der Seite, von oben oder von schräg unten links. Und mit den anderen 200 Köpfen der Datenbank geht das natürlich genau so.
Mit Hilfe unserer Datenbank kann man aber auch aus vielen einzelnen Köpfen einen neuen Kopf zusammenbauen. Damit alles zusammen passt, wurde ein Verfahren entwickelt, das automatisch in allen Gesichtern „korrespondierende Punkte" findet. Im Gegensatz zu anderen Verfahren, bei denen von Hand korrespondierende Punkte (etwa die Nasenspitze, der linke und der rechte Mundwinkel, die linke und die rechte Pupille) bestimmt werden, kann mit unserer Methode ein dichtes Korrespondenzfeld für alle Punkte der Gesichter festgelegt werden. Beim Zusammenbau eines neuen Kopfes kann man nun die korrespondierenden Teile von verschiedenen Gesichtern unterschiedlich gewichten, etwa 10 Prozent der Nase von Kopf A und 90 Prozent der Nase von Kopf B mischen und an die richtige Stelle eines neuen Kopfes C setzen.
Durch gleiche Gewichtung aller Köpfe lässt sich auch ein Durchschnittskopf berechnen. Diesen können wir als Ausgangspunkt benut-

[7] Mehr dazu unter: faces.kyb.mpg.de/
Aus dieser Datenbank wird zum Beispiel das linke Ohr generiert, das im Ausgangsbild unseres Models verdeckt ist. Dabei wird nach statistischen Kriterien ein Ohr ausgewählt, welches zu dem vorhandenen Bild am besten passt.

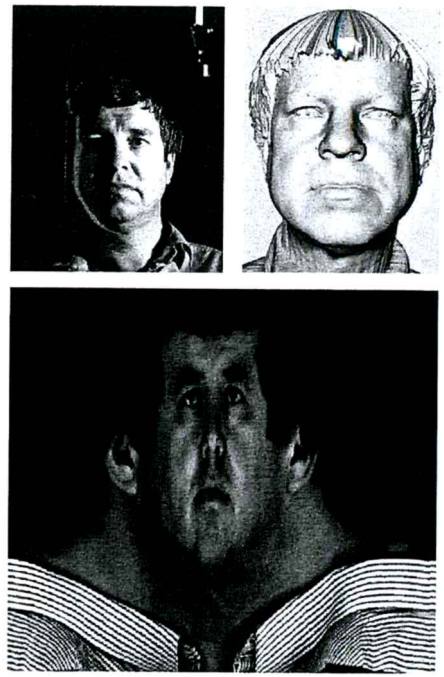


Abb. 13: Ein Kopf wird eingescannt. Hier wird gezeigt, wie die Gesichterdatenbank des MPI erzeugt wurde: Eine Person – hier ist es der Autor dieses Beitrags – sitzt steif auf einem Stuhl, während ein Laserscanner um seinen Kopf herumfährt (links oben). Die Daten, die so gewonnen werden, werden zweifach aufbereitet: als 3D-Struktur (rechts oben) und als 2D-Textur. Ein 3D-Modell des Kopfes entsteht, indem die Textur quasi röhrenförmig um die Kopfskulptur herum gerollt (unten) und die Farbinformation auf die passenden Punkte der Kopfoberfläche projiziert wird.

zen, um ein 3D-Kopfmodell an eine Fotografie anzupassen. So ist es möglich, durch entsprechende Gewichtung der Köpfe in der Datenbank das 3D-Modell eines Kopfes zu generieren, dessen eine Ansicht genau der einer gegebenen Fotografie entspricht. So wird aus einem flachen Foto ein runder Kopf.
In unserem Tübinger Labor benutzen wir die künstlichen Köpfe für Wahrnehmungsexperimente, um noch mehr darüber zu erfahren, wie eigentlich die Welt in den Kopf kommt (z. B. Wallis und Bülthoff 1999). Barbara Knappmeyer untersucht beispielsweise, wie wichtig die persönliche Mimik für das Wiedererkennen einer Person ist. Oder wir gehen der Frage nach, wie unser Gehirn weibliche von männlichen Gesichtern unterscheidet. Liegen hier zwei klar getrennte Kategorien vor – oder toleriert unser Gehirn einen fließenden Übergang?
Isabelle Bülthoff hat dazu so genannte „Morphing"-Reihen generiert, bei denen ein Frauengesicht schrittweise in ein Männergesicht über-

Abb. 14: Mann oder Frau – wer weiß es genau? Für ein Kategorisierungs-Experiment hat Isabelle Bülthoff Frauenköpfe aus der Gesichterdatenbank des MPI genommen und sie in kleinen Zwischenschritten in Männerköpfe verwandelt – man nennt das „Morphing". Im Bild sehen Sie eine solche typische Morphing-Reihe. Wenn diese Gesichter nun realen Personen gehörten, die Ihnen auf der Straße entgegen kämen, welche würden Sie für Männer halten, welche für Frauen? Bemerken Sie einen deutlichen Sprung von Mann zu Frau – oder sind sich die Zwischenstufen alle recht ähnlich?

führt wird (Abbildung 14). Diese Reihe hat sie Versuchspersonen gezeigt. Ihr überraschender Befund: Das Geschlecht von Gesichtern wird ohne spezielles Training nicht kategorisch wahrgenommen! Offensichtlich verlassen wir uns im Alltag auch noch auf andere Signale wie Haartracht, Kleidung, Figur, Bewegung und Stimme.
Natürlich ist es denkbar, mit unserem System beispielsweise aus dem Foto eines Schauspielers dessen 3D-Gesicht zu synthetisieren (siehe Blanz und Vetter 1999) und es zu animieren. So kann ein Regisseur

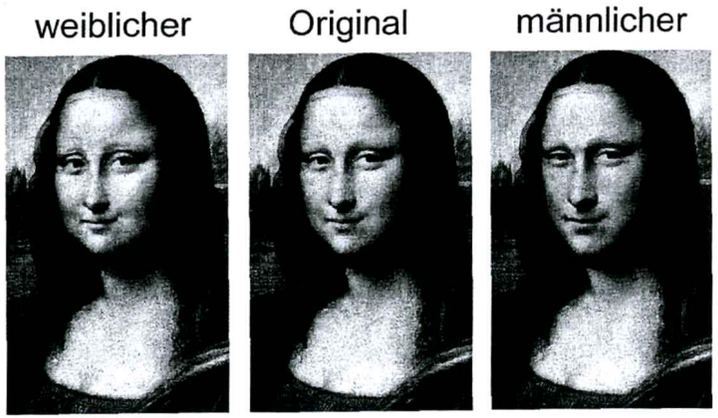
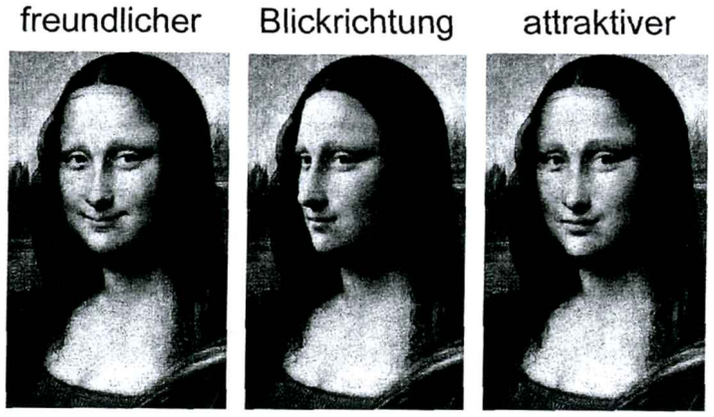


Abb. 15: Die anderen Seiten der Mona Lisa. Die MPI Datenbank von 200 Gesichtern erlaubt es zu lernen, wie sich das Bild eines Gesichtes verändert, wenn man den

seinen Star dicker oder dünner, älter oder jünger, schöner oder hässlicher aussehen lassen, er kann sein Geschlecht oder sein Mienenspiel verändern. Selbst die Mona Lisa erleben wir nun in 3D (Abbildung 14). Solche Manipulationen eröffnen ein großes Anwendungsspektrum: in der Unterhaltungsindustrie, aber auch in der Kosmetik oder der plastischen Chirurgie.

Literatur

Volker Blanz und Thomas Vetter (1999): A Morphable Model for the Synthesis of 3D Faces. SIGGRAPH'99 Conference Proceedings, 187–194.
Heinrich H. Bülthoff und Shimon Edelman (1992): Psychophysical support for a two-dimensional view interpolation theory of object recognition, Proceedings of the National Academy of Sciences U.S.A. 89, 60–64.
Nikos Logothetis, Jon Pauls und Tomaso Poggio (1995): Shape representation in the inferior temporal cortex. Current Biology 5, 552–563.
David E. Marr (1982): Vision. San Francisco: Freeman Publications.
Markus Raetz (1994): Polaroids 1978–1993. Geneva: Musée Rath.
Guy Wallis and Heinrich Bülthoff (1999): Learning to Recognize Objects. Trends in Cognitive Sciences, Vol. 3, No. 1, 22–31.



Kopf dreht oder die Mimik ändert. Dieses Vorwissen hat Volker Blanz benutzt, um ganz neue Ansichten von Mona Lisa aus einer einzigen Bildvorlage zu berechnen.