HANDBUCH DER PSYCHOLOGIE

Handbuch der Allgemeinen Psychologie – Kognition

herausgegeben von

Joachim Funke und Peter A. Frensch

HOGREFE

GÖTTINGEN · BERN · WIEN
TORONTO · SEATTLE · OXFORD · PRAG

Objektwahrnehmung
Object Recognition

Isabelle Bülthoff & Heinrich Bülthoff

1 Einführung

Wenn wir ein Zimmer betreten, können wir in der Regel sofort alle Objekte erkennen und als Stühle, Tische, Lampen usw. klassifizieren. Objekterkennung ist für uns scheinbar eine ganz einfache Aufgabe. Dass es allerdings viel schwieriger ist als der Laie annimmt, zeigt sich darin, dass es bis heute noch kein automatisches Erkennungssystem gibt, welches die für uns triviale Aufgabe der Objekterkennung lösen kann.

Diese Schwierigkeit deutet auf die tatsächliche Komplexität der Objekterkennung hin. In der Tat hat sich die Objekterkennung als eines der schwierigsten und wichtigsten Probleme in der kognitiven Neurobiologie herausgestellt.

In Abbildung 1 veranschaulichen wir einige Probleme, die unser visuelles System löst, wenn es Stühle erkennt. Wir können mühelos alle Stühle in Abbildung 1 erkennen, obgleich die Muster auf unserer Netzhaut durch unterschiedliche Orientierung oder Beleuchtung der Stühle ganz verschieden sind. Stühle können aber auch ganz unterschiedliche Formen haben oder auch nur teilweise sichtbar sein (etwa der Stuhl hinter dem Schreibtisch). Wenn die Aufgabe lautet, Stühle zu finden, auf denen wir wirklich sitzen können, würden wir z. B. den Stuhl auf dem Schreibtisch nicht in diese Kategorie einschließen, obwohl sein Bild die gleiche Größe hat wie die des Stuhls im Hinterzimmer.

Bevor wir einen Stuhl als Stuhl bezeichnen können, müssen wir vorher das Objekt vom Hintergrund trennen. Obwohl wir in diesem Kapitel nur über die Erkennung von bereits isolierten Objekten sprechen werden, sollte es dem Leser bewusst sein, dass die Trennung des Objekts vom Hintergrund kein triviales Problem ist (Peterson, 1994).

Der Eindruck, dass wir Gegenstände mühelos aus jedem Blickwinkel erkennen können, trügt. Unter allen Ansichten eines Objekts erlauben einige (kanonische Ansichten) eine schnellere Erkennung der Objekte (Palmer, Rosch & Chase, 1981). Die Existenz solcher „besserer" (kanonischer) Ansichten hat wichtige Implikationen für die beiden wichtigsten Modellvorstellungen zur Objekterkennung, die wir in diesem Kapitel beschreiben werden.

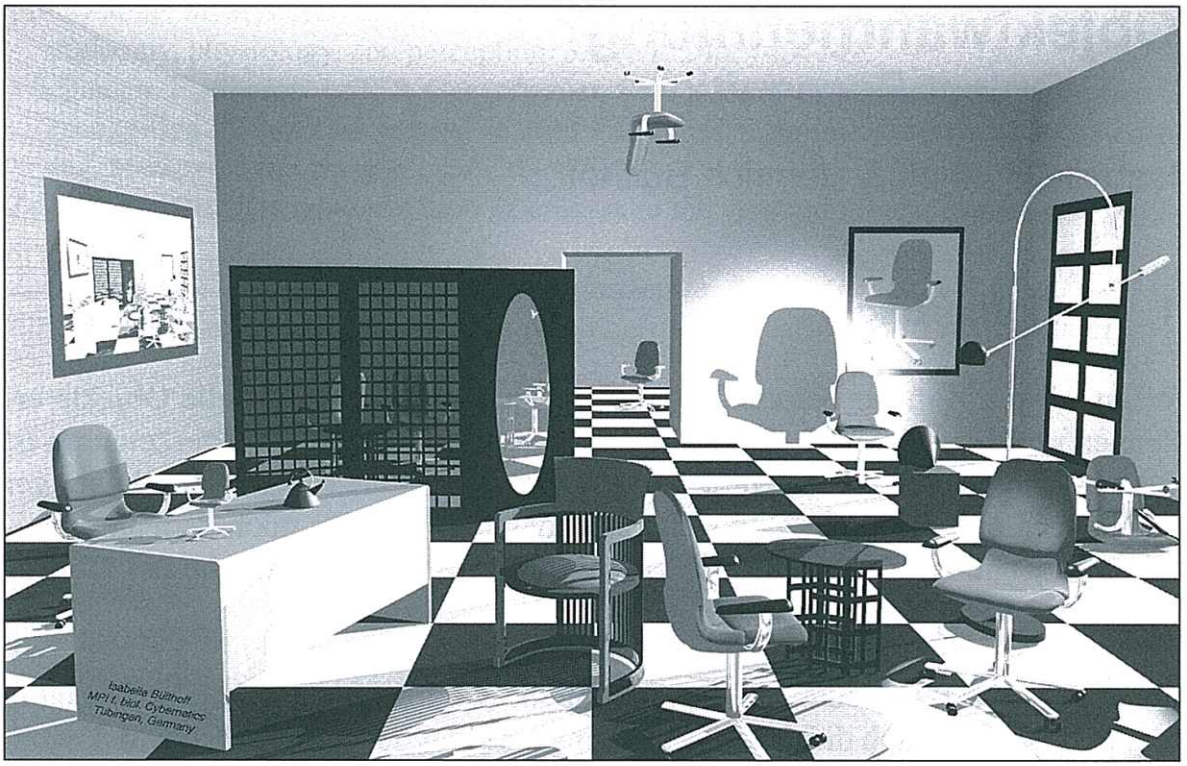

Abbildung 1: Dieses Bild stellt einige der Probleme dar, die unser Gehirn lösen muss, um Stühle zu erkennen: z. B. Orientierungs-, Größen- und Beleuchtungsinvarianz, Verdeckung (durch Schreibtisch oder Paravent), unterschiedliche Stuhlarten. Außerdem sind der Schatten oder das Bild eines Stuhles, ein zerbrochener Stuhl und der Stuhl an der Decke keine Stühle, auf denen man sitzen kann.

Wir werden auch über eine spezielle Objektkategorie – Gesichter – sprechen, nachdem wir kurz erwähnt haben, wo Objekte und Gesichter im Gehirn vorwiegend verarbeitet werden. Aber bevor wir diese Themen aufgreifen, sollten wir zuerst näher definieren, was Objekterkennung ist.

2 Terminologie

Die Bezeichnung Objekterkennung kann unterschiedliche Bedeutungen haben. Wenn wir ein Objekt spontan benennen, nennen wir meistens die *Grundkategorie* (basic level category) dieses Objekts. Wir sprechen dann in diesem Fall von *Kategorisierung*; z. B. kategorisieren wir die Gegenstände in Abbildung 1 als Stühle oder Tische. Die Grundkategorie wird auch am schnellsten genannt (Rosch et al., 1976). Bei der *Identifikation* wird ein Objekt als einzelnes Exemplar erkannt; etwa so: „Hier liegt Bashi, unser Hund."

Wenn wir sagen, dass der Stuhl hinter dem Schreibtisch in Abbildung 1 ein Bürostuhl ist, kategorisieren wir auf einer untergeordneten Ebene (subordinate level)

und sind bei der Benennung etwas langsamer, als wenn wir die Grundkategorie nennen. Wenn wir dagegen die Stühle in Abbildung 1 als Möbel bezeichnen, kategorisieren wir auf einer übergeordneten Ebene. Diese Kategorisierungsleistung ist ebenfalls langsamer.

Eine Grundthese der Objekterkennung ist, dass es keine Erkennung geben kann, ohne dass vorher eine Repräsentation im Gehirn gebildet worden ist. Erkennen heißt eigentlich *Wiedererkennen* oder Vergleichen mit einer internen Repräsentation. Im folgenden Abschnitt beschreiben wir in aller Kürze, was wir über die Repräsentation von Objekten im Kortex wissen.

3 Objekt-Repräsentation im Kortex

Von der primären Sehrinde *(V1)* gelangt die visuelle Information über den *dorsalen Pfad* zu den visuellen Assoziationsarealen im Parietallappen. Zu den visuellen Arealen im Temporallappen gelangt sie über den *ventralen Pfad* (vgl. Abb. 2).

Vereinfacht gesagt wird im dorsalen Pfad der Ort des Objekts berechnet, während die Areale entlang dem ventralen Pfad für die Frage „Was ist dieses Objekt?“ zuständig sind (Ungerleider & Mishkin, 1982).

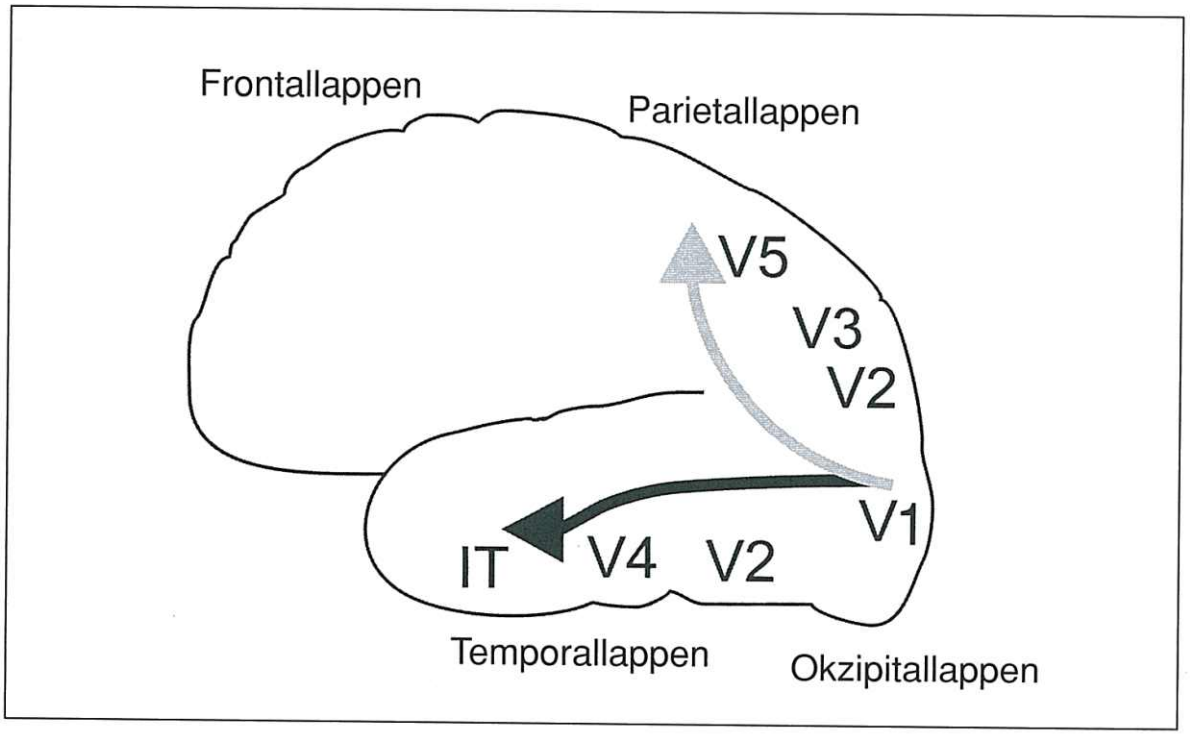

Abbildung 2: Linke Hirnhälfte: Die vier Hauptlappen zusammen mit den Hauptarealen entlang dem ventralen und dorsalen Pfad. Der graue Pfeil beschreibt den dorsalen Pfad, der schwarze Pfeil den ventralen Pfad.

Die Areale des ventralen Pfades befinden sich im unteren Teil des Temporallappens (IT, inferiorer temporaler Kortex) und sind ausschließlich für die visuelle Verarbeitung zuständig. Danach folgen andere Zentren im Temporal- und Frontallappen, die multimodal sind. Im IT befinden sich alle neuronalen Strukturen, die für die Bildung von Objektbeschreibungen notwendig sind. In dieser Region haben elektrophysiologische Untersuchungen an Primaten Neurone entdeckt, die spezifisch auf Gesichter oder bestimmte Objekte antworten. Dagegen haben vorgeschaltete Neurone eher auf einfachere Muster oder auf Objektattribute wie Farbe oder Bewegung geantwortet.

Funktionelle Bildgebung bei gesunden Probanden hat gezeigt, dass ein Bereich des Gyrus fusiformis auf der Unterseite des Temporalkortex besonders aktiv ist, wenn wir Gesichter sehen (Fusiform Face Area; Kanwisher, McDermott & Chun, 1997). In diesem Areal wird vermutlich die Identität von individuellen Gesichtern bestimmt.

4 Modelle der Objekterkennung

Wie in unserem visuellen System dreidimensionale (3D) Objekte repräsentiert sind, ist eine schwierige und leidenschaftlich debattierte Frage, die sich im Wesentlichen auf zwei Modellvorstellungen zurückführen lässt. 1. Ein Modell, das auf einer strukturellen Beschreibung von 3D-Objekten beruht (strukturelle Repräsentation), und 2. einen bildbasierten Ansatz, der auf einer Interpolation zwischen verschiedenen Ansichten der Objekte beruht (ansichtsbasierte Repräsentation).

4.1 Strukturelle Repräsentation

Der Vorteil einer strukturellen Repräsentation (vgl. Abb. 3 und 4) besteht darin, dass sie wenig Gedächtniskapazität erfordert und blickwinkelunabhängig ist. Eine einzige Beschreibung genügt, um die Erkennung eines Objekts aus fast jedem möglichen Blickwinkel zu erlauben. Alle Ansichten eines gespeicherten Gegenstandes können während des Erkennungsprozesses durch „mentale Rotation" erzeugt und mit dem Bild auf der Netzhaut verglichen werden.

Ein bekanntes Erkennungsmodell, das auf strukturellen Beschreibungen basiert, ist die Recognition-by-components-Theorie von Biederman (1987): Objekte werden in einfache Grundformen (Geone) zerlegt. Das Erkennen besteht dann in der Identifizierung der Grundformen und ihrer räumlichen Beziehungen zueinander.

Diese Theorie ist mit verschiedenen Problemen behaftet: Eines besteht darin, dass bei den meisten natürlichen Objekten die Extraktion der Grundformen nicht ein-

fach ist. (Wie soll man z. B. die Form eines Fisches oder einer Schnecke in einfache Grundformen zerlegen?) Ein anderes Problem ist, dass die Unterscheidung zwischen verschiedenen Objekten nur auf der Ebene der Grundkategorien möglich ist; das heißt, dass geon-basierte Beschreibungen zwischen Tisch und Stuhl unterscheiden können, aber nicht notwendigerweise zwischen verschiedenen Stühlen. Darüber hinaus lässt die Theorie von Biederman erwarten, dass alle Ansichten (außer einigen wenigen, bei denen die Zuordnung der Grundformen nicht eindeutig ist) gleich gut erkannt werden. Diese Annahme ist jedoch nicht vereinbar mit den bereits erwähnten bevorzugten kanonischen Ansichten.

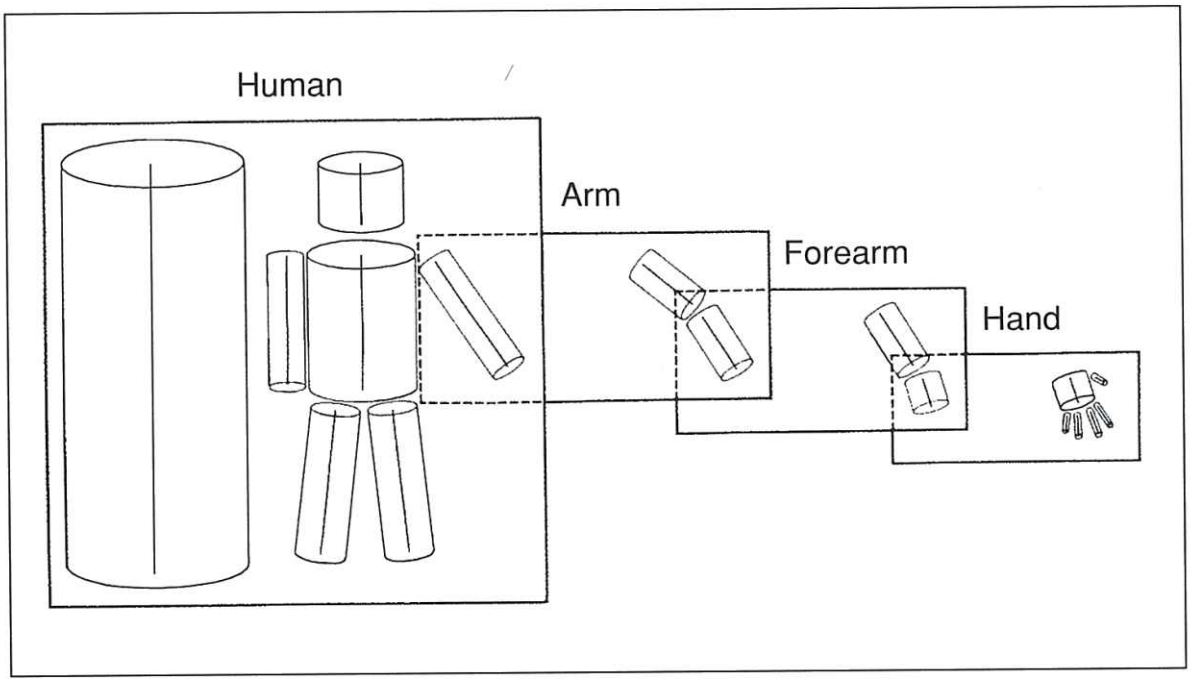

Abbildung 3: Strukturelle Objektbeschreibung nach Marr und Nishihara, (Abb. 3, S. 278: in Marr, D. & Nishihara, H. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society London Series B, 200, 269–291, abgedruckt mit Erlaubnis). Dieses 3-D-Modell beruht auf Zylindern, deren Lage durch eine Hauptachse bestimmt ist. Es wird angenommen, dass die Hauptachse leicht aus jedem Blickwinkel ermittelt werden kann.

4.2 Ansichtsbasierte Repräsentation

Ansichtsbasierte Erkennungsmodelle setzen voraus, dass jedes gesehene Objekt mehrfach abgespeichert ist, so als hätte man es aus verschiedenen Blickwinkeln fotografiert (Bülthoff & Edelmann, 1992; Tarr & Bülthoff, 1999). Allerdings wird die Bildinformation sicherlich nicht im Gehirn wie eine Fotografie abgespeichert, sondern unter „Ansicht" soll hier die Information verstanden werden, die in einem

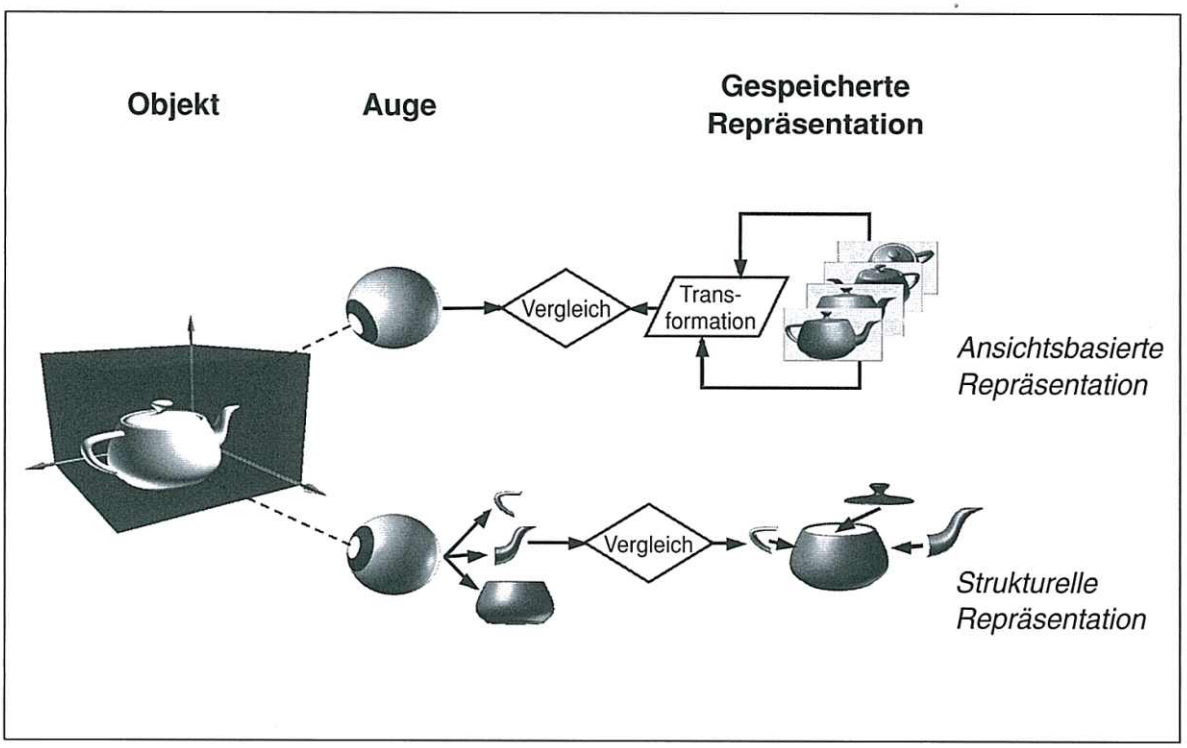

Abbildung 4: Ansichtsbasierte Erkennung im Vergleich zu struktureller Repräsentation.

Abbild eines Objekts oder einer Szene enthalten ist. Dies schließt nicht aus, dass 3D-Information in Form von binokularer Disparität oder anderen Merkmalen für räumliche Tiefe enthalten ist. Während des Erkennungsprozesses wird dann unter den im Gedächtnis (→ Gedächtnis) gespeicherten Ansichten die Ansicht herausgesucht, die dem Netzhautbild am besten entspricht (vgl. Abb. 4).

Ansichtsbasierte Mechanismen zur Objekterkennung verlangen eine große Speicherkapazität, da jede Ansicht, aus der ein Objekt erkannt werden soll, auch vorher abgespeichert sein muss. Dabei sind Schwierigkeit und Genauigkeit der Erkennung bezeichnenderweise von der Vertrautheit der Ansichten abhängig. Eine Methode, mit der das Gehirn das Speicherproblem bewältigen könnte, besteht darin, bei Bedarf eine zusätzliche anpassende Transformation vorzunehmen, so dass ein Objekt auch erkannt wird, wenn es aus einem leicht unterschiedlichen Blickwinkel gesehen wird. Mithilfe dieser Maßnahme könnte die Speichergröße reduziert werden, weil nicht jede Ansicht im Speicher enthalten sein muss. Ähnliche anpassende Transformationen könnten für andere Ansichtsabweichungen (z. B. Änderung der Beleuchtung) vorgesehen sein.

Für Blickwinkelabweichungen sind verschiedene Mechanismen vorgeschlagen worden, um von bekannten zu unbekannten Ansichten generalisieren zu können, unter anderem mentale Rotation (Tarr & Pinker, 1989) und Ansichtsinterpolation (Poggio & Edelman, 1990).

Für eine ansichtsabhängige Repräsentation sprechen viele Experimente, die gezeigt haben, dass mit Abweichung von der gelernten Ansicht die Erkennungsleistung stark abfällt. Eine zentrale Frage ansichtsbasierter Erkennungsmodelle ist, wie verschiedene Ansichten eines Objekts vereinigt werden, um eine einheitliche und zusammenhängende Objekteinheit zu schaffen. Wenn wir uns einem Objekt nähern, bekommen wir in kurzer Zeitfolge neue Ansichten, wobei sich zeitlich benachbarte Ansichten nur wenig unterscheiden. Wallis und Bülthoff (2001) haben gezeigt, dass unser visuelles System diese zeitliche Nachbarschaft benutzt, um unterschiedliche Ansichten zu einer Objekteinheit zu verknüpfen.

5 Gesichtserkennung

Gesichter sind von großer sozialer Bedeutung. Im Vergleich zu anderen Objektklassen sind sie untereinander alle sehr ähnlich. Trotzdem können wir die meisten Gesichter ohne Mühe voneinander unterscheiden.

Bereits wenige Stunden nach der Geburt interessieren sich Neugeborene für Gesichter oder gesichtsähnliche Muster. Mit zwei Monaten fangen sie an, das Gesicht ihrer Mutter zu erkennen und reale Gesichter von gesichtsähnlichen Mustern zu unterscheiden. Trotzdem verändern sich die Strategien von Kindern und Erwachsenen bei der Gesichtserkennung bis in die Pubertät (Schwarzer & Leder, 2003). Auch wenn wir es nicht bemerken, verlangt die Gesichtserkennung jahrelange Übung.

In vielen Experimenten wurde gezeigt, dass Gesichtserkennung und Objekterkennung auf unterschiedlichen Mechanismen beruhen (Bruce & Humphreys, 1994). Zum Beispiel wird nur die Gesichtserkennung stark beeinträchtigt, wenn die Stimuli auf den Kopf gestellt werden (Inversionseffekt); bei anderen Objekten ist das nicht der Fall. Solche Befunde und das angeborene Interesse von Säuglingen für Gesichter unterstützen die Theorie, dass Gesichter eine besondere Klasse von Objekten bilden.

Neurophysiologische Untersuchungen haben gezeigt, dass ein als *Gyrus fusiformis* bekannter Bereich in der Hirnrinde besonders aktiv ist, wenn Gesichter gezeigt werden (Kanwisher et al., 1997). Außerdem führen Läsionen in dieser Gehirngegend dazu, dass Patienten Gesichter nicht mehr erkennen können *(Prosopagnosie)*, obwohl ihre Fähigkeit, andere Objekte zu erkennen, nicht betroffen ist.

Trotz all dieser Hinweise gibt es eine heftige Kontroverse darüber, ob Gesichter per se für unsere Wahrnehmung etwas Besonderes sind oder ob sie nur Objekte sind, die eine besonders große Expertise fordern, damit sie individuell erkannt werden können (Tarr & Gauthier, 2000) und aus diesem Grund eine eigene Gehirnregion für ihre Erkennung beanspruchen. Diese Streitfrage ist noch nicht gelöst.

Literatur

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Bruce, V. & Humphreys, G. W. (1994). Recognising objects and faces. *Visual Cognition, 1,* 141–180.

Bülthoff, H. H. & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89,* 60–64.

Kanwisher, N., McDermott, J. & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Marr, D. (1982). *Vision*. New York: Freeman.

Palmer, S. E., Rosch, E. & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and Performance, IX* (pp. 135–151). Hillsdale, NJ: Erlbaum.

Peterson, M. A. (1994). Shape recognition can and does occur before figure-ground organization. *Current Directions in Psychological Science, 3,* 105–111.

Poggio, T. & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Schwarzer, G. & Leder, H. (Eds.). (2003). *The development of face processing*. Cambridge, MA: Hogrefe & Huber.

Tarr, M. J. & Bülthoff, H. H. (Eds.). (1999). *Object recognition in man, monkey, and machine (Cognition Special Issues)*. Cambridge, MA: MIT.

Tarr, M. & Gauthier I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3,* 764–769.

Tarr, M. & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233–282.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT.

Wallis, G. M. & Bülthoff, H. H. (2001). Effect of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98,* 4800–4804.