

# Visuelle Kognition A. Schwaninger

## Zusammenfassung

Kosslyn, S.M. (1994). Image and Brain. The resolution of the imagery debate. Cambridge, MA: MIT Press.

## 1. Einführung

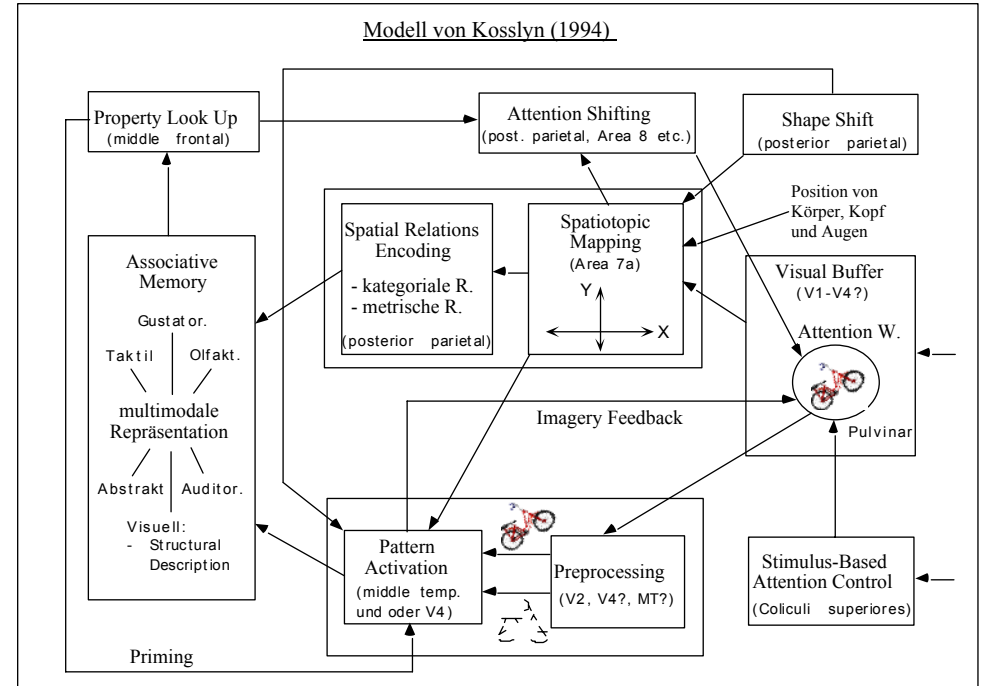
In seinem Buch entwickelt Kosslyn eine Theorie der High-Level Vision. Basierend auf der bereits von William James im letzten Jahrhundert formulierten Annahme der Gemeinsamkeit von Wahrnehmungs- und Vorstellungsprozessen im Bezug auf zugrundeliegende neuronale Systeme, versucht Kosslyn neben der Objekterkennung auch die Imagery zu erklären. In "Image and Brain" entwickelt Kosslyn ein Modell für beide Leistungen, indem er sich auf Befunde aus der Wahrnehmungspsychologie, der Psychophysik, sowie aus den Neurowissenschaften stützt. Zentral ist ausserdem der computationale Ansatz. Ausgehend von der Überlegung, dass man ein System am ehesten versteht, indem man es auf einem bestimmten Abstraktionsniveau nachbaut, bezieht Kosslyn bei der Modellentwicklung Kenntnisse aus der Bildererkennung mit Computer mit ein.

### 1.1. Wiedererkennung (Recognition) und Identifikation

Kosslyn unterscheidet den Begriff der Wiedererkennung (Recognition) vom Begriff der Identifikation (S.72). Wiedererkennung findet dann statt, wenn ein Input mit einer visuellen Gedächtnisrepräsentation hinreichend übereinstimmt und man daher ein Objekt als bekannt erlebt. Die Identifikation tritt dann ein, wenn ein Input eine multimodale Gedächtnisrepräsentation im Assoziativen Gedächtnis aktiviert und man dadurch Zugang zu weiteren multimodalen und konzeptuellen Informationen über das Objekt hat (seinen Namen, Verwendungszweck, Teile und Relationen, mit dem Objekt assoziierte Klänge usw.).

## 2. Modell

Im folgenden wird das von Kosslyn postulierte Modell in einer knappen Übersicht dargestellt. (Aus Gründen der Übersichtlichkeit wurde auf die weitere Unterteilung einzelner Subsysteme verzichtet. Für das vollständige Modell siehe Kosslyn, 1994, S. 383.)



### 2.1. Visual Buffer

Diejenigen zentralnervösen Strukturen, welche die Figur-Grund-Unterscheidung vornehmen, fasst Kosslyn als den Visual Buffer zusammen (S.70). Mögliche Areale dafür sind V1, V2, V3a, und evt. V4. Die Repräsentation im Visual Buffer ist topographisch organisiert und entspricht dem 2.5D Sketch nach Marr (S.86). Das bedeutet, dass nur derjenige Teil der dritten Dimension repräsentiert ist, welcher vom Objekt zum Betrachter projiziert (S.70, 86, 169). Im Visual Buffer ist ein Bild gleichzeitig in mehreren Auflösungen repräsentiert, was durch Konvergenz und Mittelung von Signalen erreicht wird (S.95).

### 2.2. Attention Window

Das Attention Window wird definiert als derjenige Teil des Visual Buffers, welcher an höhere Areale für die Objekterkennung weitergeleitet wird (S.87, 89). Die Selektion der zu verarbeitenden Information könnte durch den Pulvinar gesteuert werden (Projektionen in Schicht 1, 2, 3 von V1 und in Schicht 1, 3, 4 von V2 und V4). Die top-down kontrollierte Steuerung der Auswahl aus dem Visual Buffer könnte über reziproke Verbindungen des Pulvinar mit IT und V4 vor sich gehen (S.91).

### 2.3. Stimulus-Based Attention Control

Kosslyn postuliert ein sogenanntes Stimulus-Based Attention Control Subsystem, welches auf saliente Stimuli im gesamten Visual Buffer reagiert, indem es Körper, Kopf, Augen und das Attention Window so steuert, dass die Repräsentation des Stimulus im Visual Buffer vom Attention Window umschlossen wird (S.93). Diese Steuerung ist reflexiv und der bewussten Aufmerksamkeitssteuerung vorgeordnet (präattentiv). Am ehesten kommen dafür die Colliculi Superiores in Frage (S.93, 380). Die Verbindung dieser Kerne mit dem Pulvinar legt eine Bedeutung bei der präattentiven Aufmerksamkeitssteuerung nahe (tectopulvinare Projektionsbahn).

### 2.4. Attention-Shifting Subsystem

Neben der präattentiven Steuerung kann die Aufmerksamkeit auch durch höhere Areale des Cortex beeinflusst werden. Jene werden als Attention Shifting Subsystem zusammengefasst. Dazu gehören Areale im posterioren Parietalcortex, insbesondere Area 7b und LIP, sowie die frontalen Augenfelder (Area 8). Möglicherweise ist auch das anteriore Cingulum und das für Augenfolgebewegungen wichtige Areal MST (medio- superiorer temporaler Cortex) beteiligt. Die verschiedenen Areale haben unterschiedliche Aufgaben, welche sich aus elektrophysiologischen Stimulationsexperimenten und Läsionsexperimenten bei Tieren folgern lassen (S.237). Aus Platzgründen wird darauf nicht näher eingegangen. Es soll aber darauf hingewiesen werden, dass durch das Attention Shifting Subsystem die Aufmerksamkeit auch bewusst gesteuert werden kann, ein Beispiel hierzu wäre die willkürliche Unterdrückbarkeit von Sakkaden.

### 2.5. Preprocessing Subsystem

Kosslyn postuliert ein Preprocessing-Subsystem, welches rotations- und translations-invariante Objektmerkmale (Nonaccidental Properties, S.109) extrahiert und sie zusammen mit dem Bild innerhalb des Attention Windows an das visuelle Gedächtnis schickt. Kandidaten für die Lokalisierung dieses Subsystems sind intermediäre visuelle Areale, welche v.a. Farb- und Formeigenschaften verarbeiten, insbesondere V2 und V4 (S.31, 113, 116, 140).

### 2.6. Pattern Activation Subsystem

Das visuelle Gedächtnis wird von Kosslyn als Pattern Activation Subsystem bezeichnet. Es enthält visuelle Repräsentationen von Objekten in Form von Feature-Vektoren. Es handelt sich dabei um eine nicht topographische Repräsentation, welche unter anderem die Nonaccidental-Properties und ihre relative Position beinhaltet (S.119, 126). Im Pattern Activation Subsystem werden jeweils mehrere Repräsentationen von Objekten aktiviert. Dieses neuronale Netzwerk ist kompetitiv und funktioniert nach dem "Winner-take-all" - Prinzip, d.h. diejenige Repräsentation, welche am stärksten aktiviert wird, unterdrückt alle

anderen Repräsentationen (S.117, 119). Wie in 1.1. bereits angedeutet wurde, wird ein Objekt dann erkannt, wenn eine visuelle Gedächtnisrepräsentation überschwellig aktiviert wird.

Das Pattern Activation Subsystem könnte im inferioren (S.119) und mittleren (S.144) Temporallappen, bzw. im Gyrus fusiformis (S.144, 254) lokalisiert sein. Diesbezüglich von Interesse ist der superiore temporale Sulcus (STS). Zelleitungen zeigten nämlich, dass dort Neurone existieren, welche z.B. selektiv auf Gesichter oder Hände reagieren (S.128, 139, 143).

Die momentan aktive Repräsentation wird verwendet, um das Bild im Visual Buffer zu vervollständigen (Imagery Feedback). Eine Rückprojektion vom Pattern Activation Subsystem in den Visual Buffer könnte durch die direkte Projektion des Areals TE (anterioren Teil des inferotemporalen Cortex) zu V1 (S.66, 147) und/oder durch Rückprojektionen über MT und V4 zurück zu ihrem Inputgeber V2 realisiert sein (S.146).

### 2.7. Spatiotopic Mapping Subsystem

Ausgehend von Mishkin und Ungerleider (1982) ist ein sog. ventrales System (What-System) für die Erkennung von Objekten zuständig und zwar relativ unabhängig von der Distanz vom Betrachter und der Lokalisation im fokussierten Teil des Gesichtsfelds (S.70, 137, 176). Dies entspricht den Leistungen des Preprocessing Subsystems und des Pattern Activation Subsystems. Das dorsale System (Where-System) liefert dagegen genau die Informationen, von welchen das ventrale System abstrahiert. Das Where-System ist zuständig für die Erkennung von Orientierung, Grösse und Distanzen von Objekten und ihren Teilen.

Das Spatiotopic Mapping Subsystem ermöglicht die Berechnung von Ort, Grösse und Orientierung von Objekten indem die 2.5 D Repräsentation des Visual Buffers (retinope Koordinaten) in dreidimensionale Koordinaten umgerechnet werden (S.168, 170, 192). Für diese Umrechnung muss die Stellung von Körper, Kopf und Augen bekannt sein. Diese Informationen erhält das Spatiotopic Mapping Subsystem durch die entsprechenden propriozeptiven Systeme. Das Spatiotopic Mapping Subsystem könnte im posterioren Parietalcortex lokalisiert sein (S.171, 174).

### 2.8. Spatial Relations Encoding Subsystem

Das Spatiotopic Mapping Subsystem setzt das Bild im Visual Buffer in ein dreidimensionales Koordinatensystem. Diese Information wird nun vom Spatial Relations Encoding Subsystem ausgewertet, um räumliche Relationen zwischen Objekten und ihren Teilen, sowie ihrem Ort, ihrer Grösse und Orientierung im Raum zu spezifizieren (S.193). Diese Relationen sind einerseits metrisch (Distanzen) und andererseits kategorial (z.B. links, rechts, oben, unten etc.). Kosslyn nimmt an, dass dieses System ebenfalls im posterioren Parietallappen lokalisiert ist (S.213).

## 2.9. Shape Shift Subsystem

Wie bereits erwähnt wurde, ist das Pattern Activation Subsystem nicht topographisch organisiert. Durch rückprojizierende Bahnen ausgehend vom Pattern Activation Subsystem wird das Bild im Visual Buffer ergänzt (Imagery Feedback). Die dabei nötige metrische Information erhält das Pattern Activation Subsystem vom Spatiotopic Mapping Subsystem (S.176). Beim Vorgang des Imagery Feedbacks wird durch das Shape Shift Subsystem zunächst das vom Spatiotopic Mapping Subsystem an das Pattern Activation Subsystem geschickte metrische Input verändert und danach die Rückprojektion an den Visual Buffer gemäss dem veränderten metrischen Input modifiziert, so dass das zurückprojizierte Bild in Grösse, Orientierung und Auflösung mit dem Bild im Visual Buffer übereinstimmt (S.121, 145, 351, 353). Klassische Beispiele hierfür sind mentale Rotationsprozesse (S.345ff.), was ein Spezialfall des erwähnten Transformationsprozesses ist, zumal ja lediglich die Orientierungsinformation beim Imagery Feedback modifiziert wird. Das Shape Shift Subsystem könnte im posterioren Parietalcortex lokalisiert sein (S.373).

## 2.10. Assoziatives Gedächtnis

Im Assoziativen Gedächtnis sind die Repräsentationen multimodal, d.h. sie können von verschiedenen Sinnessystemen aktiviert werden (S.216, 242). Im Assoziativen Gedächtnis sind ausserdem abstrakte Konzepte gespeichert (S.215).

Dieser multimodale Informationsspeicher erhält Angaben über die Form von Objekten vom Pattern Activation Subsystem (ventrales System) und Informationen über die Grösse und metrische Relationen seiner Teile vom dorsalen System, in diesem Fall vom Spatial Relations Encoding Subsystem (S.218, 227, 241). Das Assoziative Gedächtnis ist wie das Pattern Activation Subsystem als kompetitives Netzwerk realisiert (S.220). Während die Wiedererkennung (Recognition) im Pattern Activation Subsystem stattfindet, ist die Identifikation an eine Aktivierung einer multimodalen Repräsentation im Assoziativen Gedächtnis gebunden.

Für die Bildung von Verknüpfungen verschiedener Informationen scheint der Hippocampus eine wichtige Rolle zu spielen (S.223). Die Informationen selbst müssen aber an einem anderen Ort gespeichert sein, was durch die klassischen Befunde beim Patienten H.M. bestärkt wird (anterograde Amnesie bei intakter Erinnerung an Ereignisse vor der Operation). Gemäss Kosslyn könnte angenommen werden, dass das Assoziative Gedächtnis im superioren temporalen Cortex und im Übergangsbereich zwischen temporalem, parietalen und occipitalem Cortex lokalisiert ist. Das Areal STP beim Affen (Superior Temporal Polysensory) scheint dabei eine besondere Rolle zu spielen. Es erhält visuelle Afferenzen vom inferioren temporalen Cortex (von IT und STS), auditorisches Input vom superioren temporalen Cortex, sowie Afferenzen von somästhetischen Systemen im posterioren parietalen Cortex (S.223). Die Neurone im STP haben sehr grosse rezeptive Felder ( $> 150^\circ$ ) und zeigen die stärkste Aktivität bei multimodaler Reizung (auditorisch und visuell).

## 2.11. Property Look-Up Subsysteme

Im Assoziativen Gedächtnis sind Objekte auch als sogenannte Structural Descriptions gespeichert (S.216). Eine solche Repräsentation beschreibt das Objekt als eine Hierarchie von untergeordneten Objekten mit räumlichen Relationen (metrische und kategoriale), sowie ihrer Distinktheit. Anhand dieser Informationen kann nun das visuelle System die Aufmerksamkeit gezieht zu Orten im Gesichtsfeld lenken, wo besonders distinkte Merkmale vorhanden sein müssten. Kosslyn postuliert sogenannte Property Look-Up Subsysteme, welche im Assoziativen Gedächtnis besonders kennzeichnende Muster und ihre räumlichen Relationen detektiert (S.226). Das Output dieser Look-Up Subsysteme wird in Form von metrischen, für die Motorik verwendbaren Koordinaten an das Attention-Shifting Subsystem geschickt (S.233). Letzteres verschiebt das Aufmerksamkeitsfenster (Attention Window) an die entsprechenden Stellen des Reizmusters im Visual Buffer. Die Property Look-Up Subsysteme schicken aber auch Informationen an das ventrale System. Das Pattern Activation Subsystem erhält den Code des entsprechenden visuellen Musters. Diese top-down Aktivierung bezeichnet Kosslyn als attentionales Priming (S.236). Es bildet die Grundlage für Imagery. Die Property-Look-Up Subsysteme dürften im dorsolateralen prefrontalen Cortex lokalisiert sein (S.228, 231).

## 3. Objekterkennung unter verschiedenen Reizgegebenheiten

Unser kognitiver Apparat gewährt die Wiedererkennung und Identifikation von Objekten unter einer Vielzahl von Wahrnehmungsbedingungen (S.60ff., 268ff.). Dies ist eine erstaunliche Fähigkeit zumal die Objekte in der Aussenwelt sich in unterschiedlicher Distanz vom Betrachter befinden können, einen unterschiedlichen Ort im Gesichtsfeld einnehmen und um ihre drei Hauptachsen rotiert sein können. Ausserdem sind Objekte oftmals durch andere Objekte teilweise verdeckt, ihre Konturen und Texturen sind nicht vollständig sichtbar oder nur unscharf auszumachen. Wohl am interessantesten ist die Fähigkeit, Objekte auch zu erkennen, wenn sie eigentümlich verdreht sind und die Anordnung ihrer Teile für den Betrachter völlig neuartig erscheint.

### 3.1. Erkennung bei unterschiedlicher Entfernung

Das Preprocessing Subsystem extrahiert dieselben Nonaccidental Properties unabhängig von der Grösse des Objektes im Visual Buffer (Nonaccidental Properties sind ja per definitionem rotations- und translationsinvariant, S.109). Im Pattern Activation Subsystem wird eine visuelle Repräsentation aktiviert. Die metrische Information erhält es vom Spatiotopic Mapping Subsystem. Die Übereinstimmung zwischen der Repräsentation des externen Stimulus (im Attention Window) und der Gedächtnisrepräsentation (im Pattern Activation Subsystem) wird durch zwei Mechanismen erreicht (S.124). Top-Down: Das Shape Shift Subsystem verändert die metrische Information der Gedächtnisrepräsentation (Input vom Spatiotopic Mapping Subsystem) solange, bis die Rückprojektion (Imagery

Feedback) in den Visual Buffer mit dem dortigen Bild optimal übereinstimmt (S.121, 145, 351, 353). Bottom-Up: Die Auflösung im Attention Window wird verändert, wodurch die Grösse des zum Pattern Activation Subsystem geschickten Bildes geändert wird (S.95).

### 3.2. Erkennung bei unterschiedlichem Ort im Gesichtsfeld

Das Stimulus-Based Attention Control Subsystem steuert Körper, Kopf, Augen und Attention Window so, dass die Repräsentation eines Objektes im Visual Buffer vom Attention Window umschlossen wird (S.93). Letzteres ist definiert als derjenige Teil des Visual Buffers, welcher an höhere Areale für die Objekterkennung weitergeleitet wird (S.87, 89). Durch diesen Mechanismus und das dazwischengeschaltete Preprocessing Subsystem ist gewährleistet, dass das Pattern Activation Subsystem unabhängig vom Ort des Objektes in der Frontalebene dasselbe Input erhält<sup>1</sup> (S.114, 259).

### 3.3. Erkennung trotz Rotationen

Wie bereits mehrfach erwähnt wurde, extrahiert das Preprocessing Subsystem rotations- und translationsinvariante Merkmale. Diese und das Bild vom Attention Window werden an das Pattern Activation Subsystem geschickt, wodurch dort eine visuelle Gedächtnisrepräsentation aktiviert wird. Ist das Objekt in einem moderaten Mass rotiert, so reicht die Aktivierung im Pattern Activation Subsystem für die Erkennung aus. Andernfalls wird die Erkennung durch den bereits erklärten Top-Down Mechanismus angestrebt: Das Shape Shift Subsystem verändert die metrische Information der Gedächtnisrepräsentation (Input vom Spatiotopic Mapping Subsystem, S. 176) solange, bis die Rückprojektion (Imagery Feedback) in den Visual Buffer mit dem dortigen Bild optimal sowohl in Grösse, als auch in der Orientierung übereinstimmt (S.121, 145, 351, 353). Dadurch kommt es zur überschwelligeren Aktivierung im Pattern Activation Subsystem, wodurch das Objekt erkannt wird.

### 3.4. Erkennung trotz Verdeckung oder unscharfer Konturierung und Texturierung

Ist das Bild im Visual Buffer unvollständig, gestört oder verzerrt, so wird durch das Imagery Feedback das Bild im Visual Buffer vervollständigt und die Störungen eliminiert, bis die Repräsentation im Pattern Activation Subsystem überschwellig aktiviert wird und es zur Objekterkennung kommt.

### 3.5. Erkennung bei neuartiger Anordnung der Teile

Ist ein Objekt auf eine Weise verdreht, so dass die Anordnung seiner Teile dem Betrachter unbekannt ist, so ist die Übereinstimmung zwischen dem Bild im Visual Buffer und der Gedächtnisrepräsentation oftmals unterschwellig, das Objekt als Ganzes wird also vorerst nicht erkannt (S.216). Das Assoziative Gedächtnis erhält aber Informationen über erkannte Teile vom Pattern Activation Subsystem, sowie metrische Information über die Orientierung und Abstände dieser Teile zueinander vom Spatial Relations Encoding Subsystem (S.218, 227, 241). Dadurch wird eine Structural Description aktiviert, welche ein Objekt als Hierarchie von untergeordneten Teilen mit räumlichen Relationen (metrische und kategoriale) spezifiziert. Diese Aktivierung liefert das Input für die Property Look-Up Subsysteme (S.226ff.), welche die Aufmerksamkeit auf kennzeichnende Bereiche des Objektes richten (S.228). Dieser Prozess wird wiederholt, so dass immer mehr Teile des Objektes erkannt werden und schliesslich das Objekt als Ganzes identifiziert wird (S.241).

---

<sup>1</sup> Eine Ausnahme bildet der Fall, welcher eintritt, wenn das Objekt einen grösseren Schwinkel als das Gesichtsfeld aufweist. Im Alltag wird man seine Distanz zum Objekt vergrössern, um das Problem zu eliminieren. Aber auch wenn dies nicht möglich ist, können Objekte oftmals dennoch erkannt werden. Für zugrundeliegende Mechanismen sei auf S. 239 ff. verwiesen.