

## 8. Bildverstehen und Wissensrepräsentation

### Begriffsklärungen

#### *Bildverstehen:*

beinhaltet

- Bedeutungserfassung einer Bildstruktur durch die Zuweisung von Namen
- Auffinden wichtiger Details (Bewertung)
- Assoziation mit anderen Bildern
- Begreifen von ursächlichen Zusammenhängen (kausales Verstehen des Bildinhaltes)
- intuitives Erfassen von Zusammenhängen (intuitives Verstehen des Bildinhaltes)
- Ziehen von Schlussfolgerungen aus den gefundenen Zusammenhängen

insbesondere:

- Erkennen von Objekten, Ereignissen, Szenen
- Erkennen von exemplarischen Situationen
- Erkennen von Absichten
- Ableitung von Schlüssen (z.B. hinsichtlich neuer Aktionen)

Beispiele:

A) Luftaufklärung, Fernerkundung: *Erkennen (Interpretation)* von Schiffen.  
Modelle:

- a) Bild: Hafen mit Suchbereich Landestege
- b) Merkmal: ovale Bereiche und Rechtecke
- c) Szene: Hafen (Hafenkarte)

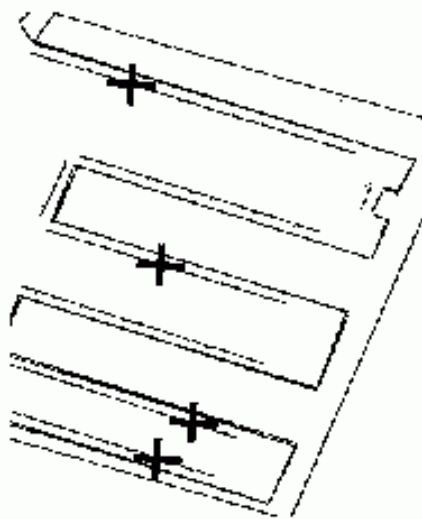
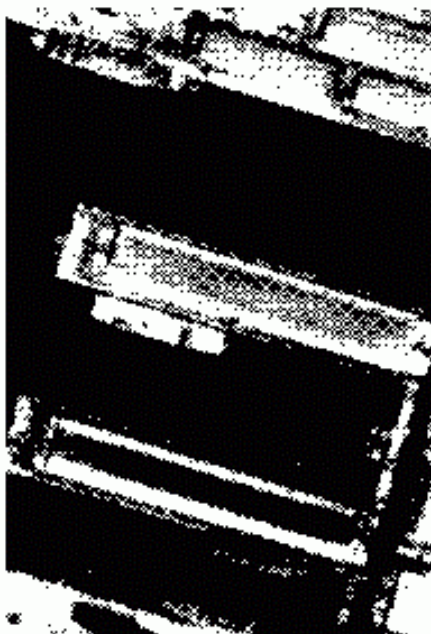
Zielorientierte modell-gestützte Interpretation. Top-Down von c) nach a) als geplantes Vorgehen (prozed. Wissen): Organisation des Interpretationsprozesses (Vorhersagen, Vergleichen).



(a)



(b)



Luftbild einer Hafenanlage; kartografische Darstellung (= Wissensbasis, auf die zurückgegriffen werden kann), Detail des Luftbildes, gefundene Landungsstege im Modell des Hafens

(aus Levi 2002)

B) Tomographie: *Rekonstruktion* zweier dreidimensionaler Nieren und des Rückenmarkes.

Modelle:

- a) Bild: "Schnittbild"
- b) Merkmale: zwei ovale Bereiche getrennt durch ausgedehnten Bereich (Rückenmark)
- c) Objekt: geometrisches Modell der Nieren und des Rückenmarks

Zielorientierte modell-gestützte Suche, die top-down von c) nach a) erfolgt.

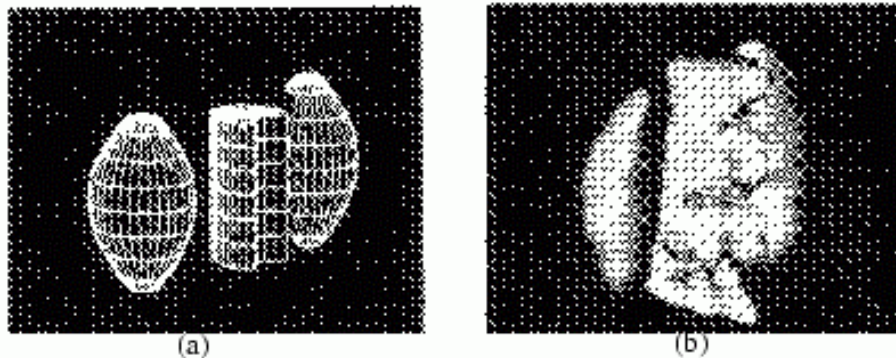


Abb.2.3: Bild der Nieren und des Rückenmarks, (a) als geometrisches Modell, (b) als tomografische Rekonstruktion.

(Levi 2002)

C) Montageroboter: Erkennen (Interpretation) von Bauteilen, um sie zu greifen und zu montieren.

Modelle:

- (a) Bild: Fließbandbereich
- (b) Merkmale: Bohrung, Komponenten, Längen...
- (c) Objekt: geometrisches Modell des Montageteils

Vorgehensweise: zielorientierte Suche von (c) nach (a).

Geplantes Vorgehen: Organisation der Suche, z.B. Verhalten bei Bildstörungen, bei defekten oder übereinanderliegenden Teilen, wenn Menschen dazwischenkommen...

## D) Verstehen von Karikaturen

2 Beispiele:



1. Virtueller Grüner Parteitag

- Erkennen von Personen
- Erkennen von (stark vereinfachten, symbolischen) Gegenständen (Bombe)
- Metaphern (Schatten)
- Kombination von Schrift und Bild
- Erkennen von Stereotypen (typischer "Grüner"...)
- Einbeziehen von Hintergrundwissen über politische Themen (hier: Debatte um J. Fischers militante Vergangenheit)

in jedem Fall notwendig: Verwendung von *Modellen*

Modell: Darstellung prototypischer Eigenschaften, Strukturen und Methoden einer Klasse von Bildprimitiven, Bildkomponenten, -objekten, Szenen oder Situationen oder eines idealen Individuums einer Klasse (Levi 2002).

*ikonisches Modell*: analog zum geometrischen und topologischen Bildinhalt aufgebaut. (Verwendet werden üblicherweise Matrix-, Vektor- oder Stringdarstellungen.)

*symbolisches Modell*: keine direkte Analogie zur Geometrie und Topologie des Bildinhaltes vorhanden. (Darstellung erfolgt üblicherweise relational, prädikatenlogisch, oder durch Graphen oder Grammatiken.)

weitere Definitionen:

*Bildprimitiv*: erste Aggregationsstufe klassifizierter Pixel (z.B. Kantenpunkte). Beispiele sind Kanten und Regionen.

*Bildkomponente*: Komponente eines Bildobjektes. Objekt hier: ein realer Gegenstand, bzw. die Region(en), die ihm im Bild entsprechen. (z.B. Hände, Augen, Nase...)

*Bildstruktur*: Struktur, die aus Bildprimitiven oder Bildkomponenten zusammengesetzt ist und die auch selbst wieder Teil komplexer Bildelemente (z.B. einer Bildszene) sein kann. Beispiele: Menschen, Roboter...

*Bildszene*: Struktur, die aus Bildstrukturen zusammengesetzt ist. Beispiele: Hörsaal, Kreuzung, Autobahn, Hafen, Bahnhof.

*Bildsituation*: semantische Beschreibung von Abläufen, die einer Bildszene zugeordnet werden können. Beispiele: zuhörende Studenten, Linksabbieger, Überholer, weinende und zuwinkende Menschen (Abschiedssituation).

Von der Stufe der "Bildkomponente" aufwärts erfolgt die Darstellung in der Regel symbolisch.

Unterscheidung von *high level* und *low level*-Bildanalyse.

## Schichtenmodell der Bildanalyse (aus Niemann & Bunke 1987):

von unten nach oben zu lesen

|   |   |                       |
|---|---|-----------------------|
| 4. Schlußfolgerungen  |   | h<br>i<br>g<br>h      |
| Zustandsbeschreibung:<br>aktueller Zustand<br>gewünschter Zustand<br>Abweichungen   | Aktionen:<br>Inferenzen<br>Plan generieren<br>planmäßig agieren   |                       |
| 3. Analyse  |   | l<br>e<br>v<br>e<br>l |
| <u>Objekte:</u><br>2-D und 3-D Linien<br>2-D und 3-D Flächen<br>Volumina, Projektionen von<br>Volumina<br>2-D und 3-D physikalische<br>Objekte, Objektgruppen,<br>Ereignisse  | <u>Eigenschaften:</u><br>symbolische Objektnamen<br>und Objektrelationen<br>(zeitliche und räumliche Bezüge)<br>Interpretation von<br>Bedeutungen |                       |
| 2. Segmentierung  |   | l<br>o<br>w           |
| <u>Objekte:</u><br>Linienelemente, Linien<br>Knotenpunkte<br>Regionen   | <u>Eigenschaften:</u><br>Grauwert (änderung)<br>Farbe, Form, Textur, Tiefe,<br>Bewegung, relat. und absol. Lage                                   |                       |
| 1. Vorverarbeitung  |   | l<br>e<br>v<br>e<br>l |
| <u>Operationen:</u><br>Codierung, Schwellwertoperationen, Filterung (z.B. Störungsreduktion,<br>Kontrastverstärkung), Restauration, Normierung (z.B. Histogramm,<br>Energie, Farbwerte, geometrische Entzerrung)<br><u>Objekte:</u> Abtastwerte |   |                       |
| 0. Abtastwerte eines Bildes   |   | l                     |

## Beispiel aus der Medizin (aus Niemann & Bunke 1987):

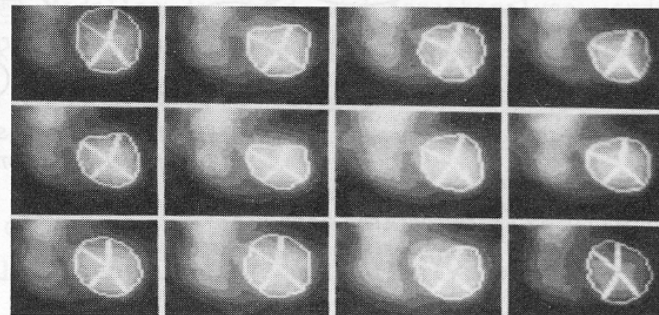
Ausgabe: diagnostische Beschreibung zum Beispiel:  
Das IA-Segment zeigt akinetisches Bewegungsverhalten, da der Stagnationsanteil mit 58 % während eines Zyklus einem akinetischen Verhalten entspricht, ebenso wie die EF mit 28 % im Normbereich für akinetisches Verhalten liegt. Über die EF von 22,7 % des LV als Ganzem wurde die Diagnose "fast-bewegungslos" mit Bewertung 0.98 erstellt. Da das IA- bzw. PL-Segment mit Bewertung 0.3 bzw. 0.14 ausgeweitet sind, ist der Verdacht auf ein Aneurysma mit Bewertung 0.14 nicht auszuschließen. (Bewertungen auf einer Skala von 0.0 bis 1.0)

Diagnosen: Angaben wie Hypokinesie oder Diskinesie in einem der 4 Segmente und Gesamtbeurteilung wie normales Herz oder Aneurysma

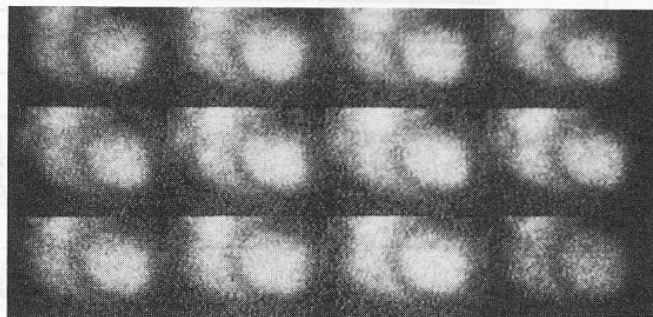
Bewegungen: anatomische Bewegungsphasen (Zyklus, Systole, Diastole, preejection phase, ejection phase, ...), längere Bewegungsphasen über mehrere Bilder, von Bild i nach Bild i+1 Angaben über Kontraktion, Stagnation, Expansion des linken Ventrikels

Objekte und Parameter je Bild: Konturen von Herz und linkem Ventrikel LV, 4 Segmente (inferioapikal IA, posterolateral PL, basal B, septal S), 12 Sektoren, Form der Objekte, Fläche, Schwerpunkt

Konturbilder aus der Vorverarbeitung und Segmentierung

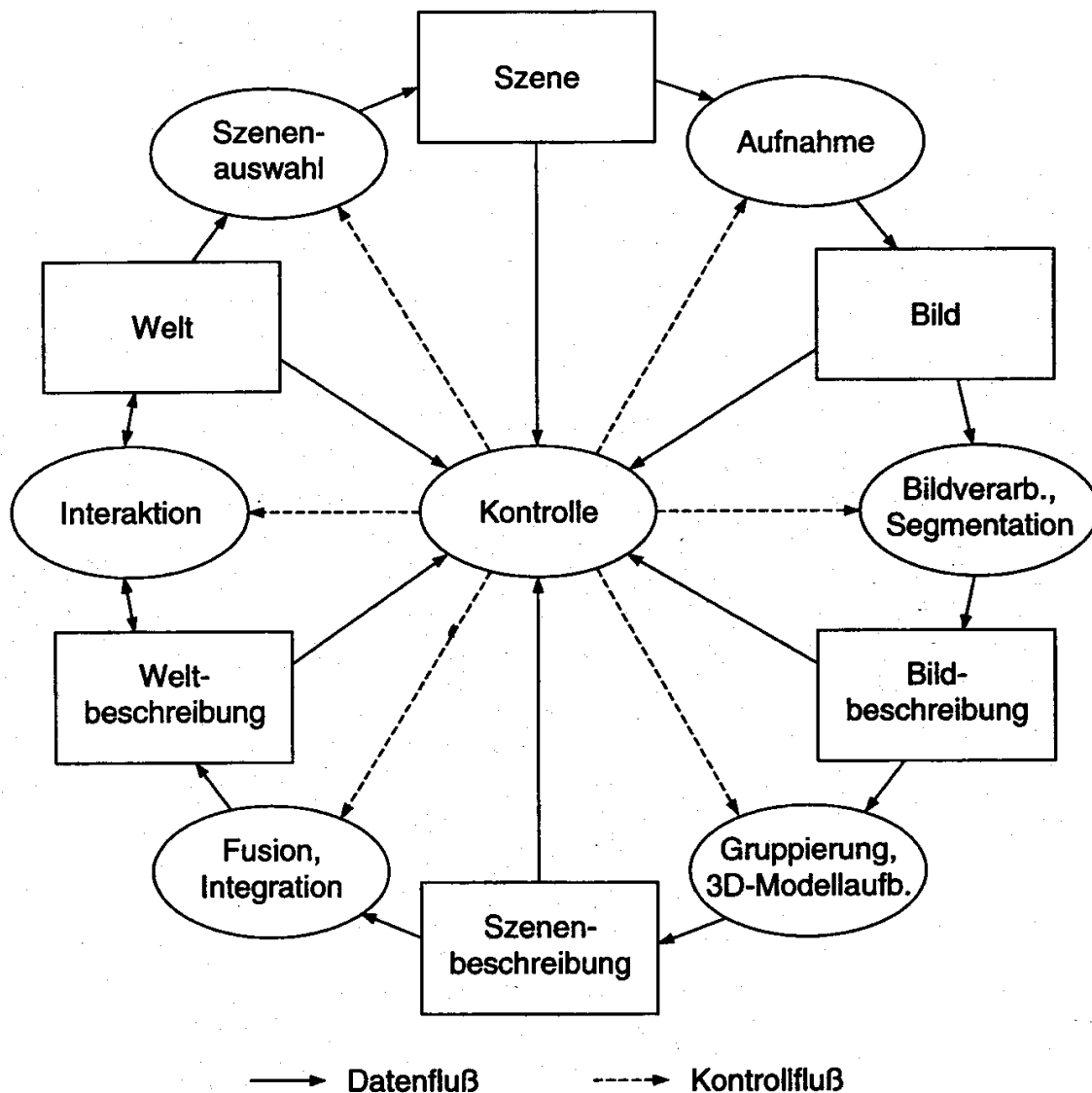


Eingabe: unverarbeitete Bildfolge



## zyklisches Modell des Bildverstehens nach Pinz 1994

(bezieht die reale Welt und ihre Veränderung durch IT-Anwendung mit ein):

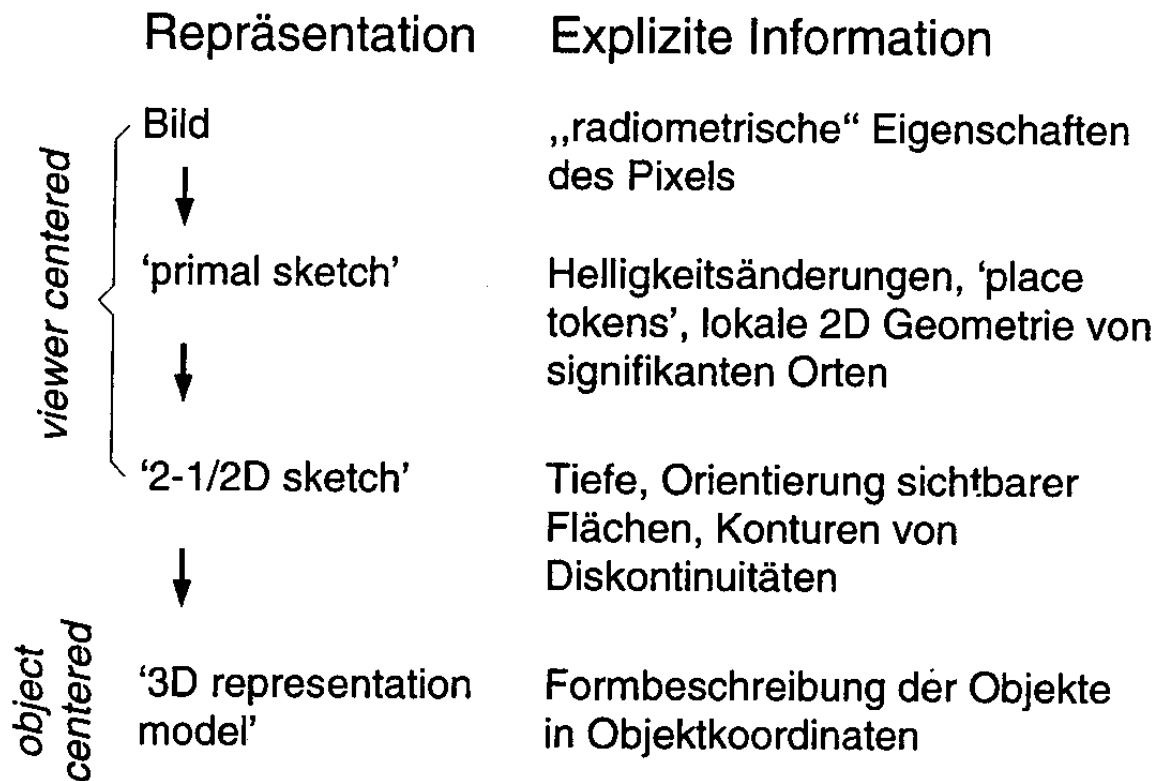


je nach Aufgabenstellung sind im high level-Bereich (hier: Gruppierung, 3D-Modellaufbau, Fusion, Integration) spezifische Probleme zu lösen:

- Mustererkennung
- Ableitung von 3D-Informationen aus 2D-Bildern
- konsistente Kombination mehrerer 2D-Bilder
- Rückschluss auf physikalische Bedingungen...



Zwischenschritte bei der Ableitung von 3D-Modellen können sein (nach Marr 1982, zit. bei Pinz 1994):



*2 1/2 D-Darstellung*: beinhaltet Tiefeninformation, Konturen, Unterscheidung von Knick- und Sprungkanten (bei letzteren ändert sich die Entfernung sprunghaft), dies aber nur für die sichtbaren Oberflächen.

Volles 3D-Modell: macht auch Aussagen über verdeckte Teile.

Strategien des Bildverstehens:

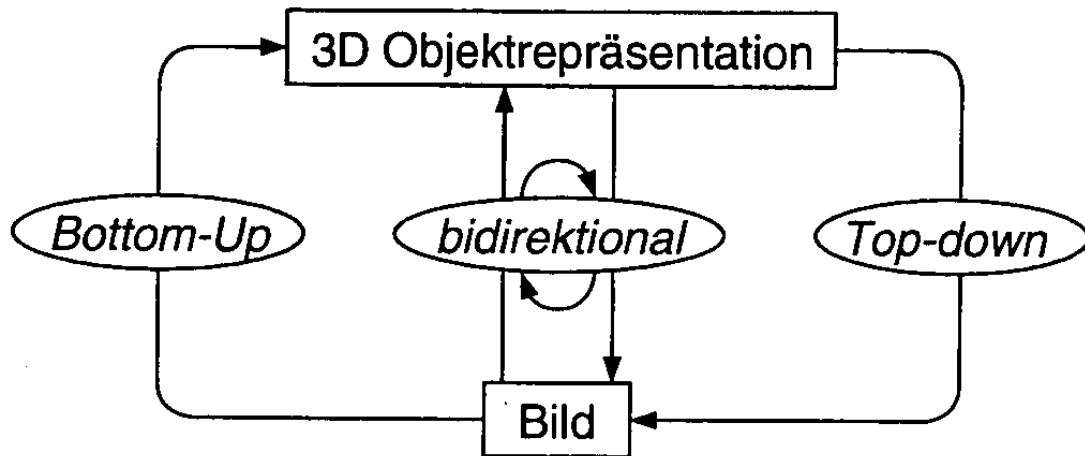
*bottom-up*:

- geht vom Bild aus, leitet Repräsentationen von Objekten ab
- versucht, diese mit in einer Wissensbasis gespeicherten Repräsentationen für bekannte Objekte zu vergleichen

*top-down*:

- geht aus von hoher Repräsentationsebene (z.B. "finde eine bestimmte Objektkategorie"), legt dann einen Bildverarbeitungsverlauf fest

*bidirektionale Strategien*: kombinieren bottom-up und top-down



2 Schulen des Bildverstehens nach Aloimonos (1989):

*reconstruction school*

- versucht, physikal. Parameter der realen Welt zu rekonstruieren
- wird mit top-down Vorgehensweise assoziiert

*recognition school*

- versucht, Objekte zu erkennen oder zu beschreiben und ggf. Aktionen herzuleiten
- wird mit bottom-up Vorgehensweise assoziiert

Reconstruction school:

Rekonstruktion physikalischer Parameter der Außenwelt  
 Finde eine spezifische Lösung für ein allgemeines Problem

**Marr-Paradigma:**

Top-Down  
 ↓

Computational theory  
 Representation and algorithm  
 Stability analysis  
 Hardware implementation

↑  
 Bottom Up

Recognition school:

Erkennen oder Beschreiben von Objekten  
 Finde eine allgemeine Lösung für ein spezifisches Problem

"High level"-Bildverstehen erfordert "höhere" Fähigkeiten,  
Wissensbasis, kognitive Prozesse...  
was ist gemeint?

*Kognition:*

Wortstamm (griech.) gignoskein = erkennen, wahrnehmen,  
wissen

Unter "Kognition" fasst man alle Fähigkeiten und Funktionen  
zusammen, die die Wahrnehmung mit der Sensomotorik  
verknüpfen. Dazu gehören u.a. die folgenden kognitiven  
Prozesse:

- Denken und Problemlösen, Intelligenzleistungen
- Bewusstsein entwickeln
- Zukunft antizipieren, Phantasie
- Emotionen haben
- Lernen
- Erinnern (Gedächtnis)
- Sehen, Hören, Riechen, Schmecken, Fühlen
- Sprechen
- Bewegen (steuern)

(in Anlehnung an Levi 2002)

*Kognitionswissenschaft:*

Ziel der K. ist die Erforschung der Kognition. Dabei stehen die  
Fähigkeiten natürlicher und technischer kognitiver Systeme zur  
Informationsverarbeitung im Mittelpunkt des Interesses.

Paradigma der KI-Forschung:

Alle kognitiven Prozesse können auf Berechnungen  
zurückgeführt werden.

Kognition ist berechenbare Informationsverarbeitung.

Beitrag der Informatik zur Kognitionswissenschaft (neben  
Neurobiologie, Kybernetik, Linguistik, Psychologie, Philosophie,  
Soziobiologie, Medizin...):

KI (Künstliche Intelligenz), KL (Künstliches Leben) (bzw. AI,  
AL); Bildverstehen, Sprachverstehen, Wissensrepräsentations-  
formen, Algorithmenentwicklung, Systemarchitektur.

### *Intelligenz:*

beinhaltet die Fähigkeiten der Auffassungsgabe, des Begreifens und des Urteilens (logisches Schließen). Ferner gehört dazu die Anpassungsfähigkeit an modifizierte und neue Aufgaben.

### *Künstliche Intelligenz:*

Teilbereich der Informatik, die mit Rechnern solche Probleme zu lösen versucht, die Intelligenzleistungen voraussetzen. Diese Probleme sind nicht mit überschaubaren (einfachen) Algorithmen zu lösen. Mit der KI soll intelligentes Verhalten von Menschen bzw. von Lebewesen erfasst und nachvollzogen werden.

*alternative Def.:* KI ist die Untersuchung und die Modellierung von Systemen, die rational schlussfolgern und handeln mit Hilfe von Rechnern und Maschinen.

(Levi 2002)

### *Verteilte Künstliche Intelligenz:*

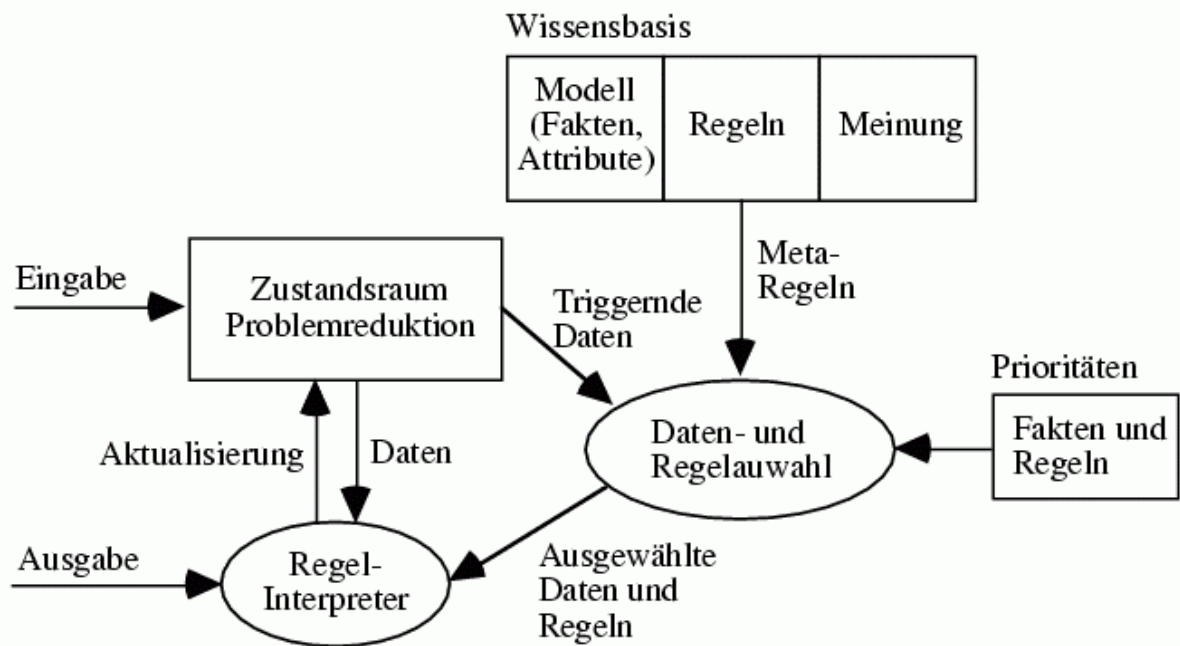
ursprünglich der Bereich der KI, der die Darstellungen und Verfahren der KI verteilt implementiert (verteiltes Problemlösen). Heute (auch) der Ansatz, Aufgaben und Methoden, die in soziologischen Gruppen (Gesellschaften) vorhanden sind, auf Rechnernetze zu übertragen und dabei neue, wesentliche Aspekte des gruppenbezogenen Verhaltens (Entscheidens) und Handelns zu untersuchen (Multiagentensysteme).

### *Wissen:*

"Wissen ist vorwiegend die Kenntnis richtiger (wahrer) Fakten (Tatsachenwissen, deklaratives Wissen) und Verfahren (Verfahrenswissen, prozedurales Wissen)" (Levi 2002)

### *wissensbasiertes System:*

stellt Wissen dar und nutzt es.



Exemplarische Architektur eines wissensbasierten Systems (WBS)

– in den Anfängen der KI wurden WBS vorwiegend als Symbolsysteme verstanden

inzwischen wird auch der *Konnektionismus* einbezogen:

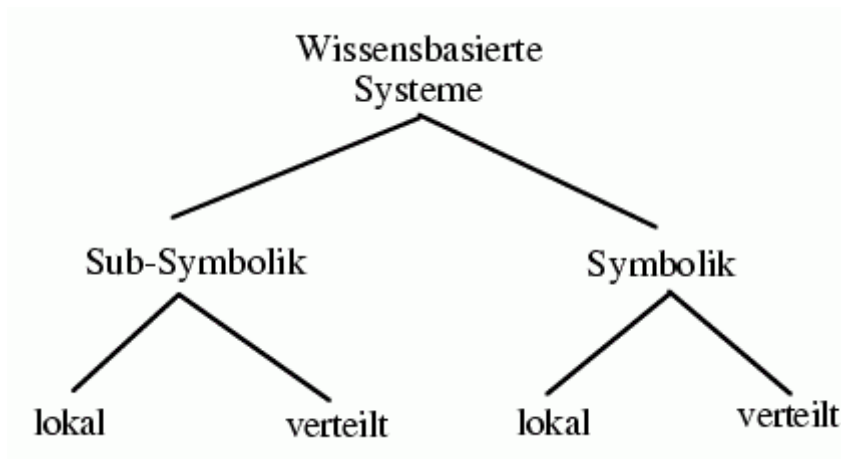
Der *Konnektionismus* beschäftigt sich mit informationsverarbeitenden Systemen, die aus sehr vielen einfachen Verarbeitungseinheiten bestehen, welche miteinander verbunden sind und gegenseitig Information (Nachrichten) austauschen.

Typischer Vertreter: *Neuronale Netze*.

"*Sub-Symbolik*":

ursprünglich eine Repräsentationsform des Konnektionismus, die sich mit der Wissensbildung vor dem symbolischen Bereich befasst. Inzwischen meist synonym zu "Konnektionismus" gebraucht.

Wissensbasierte Systeme = Sub-Symbolik + Symbolik.



Gegenüberstellung der entsprechenden Methoden  
(nach Levi 2002):

| Sub-Symbolik            | Symbolik                            |
|-------------------------|-------------------------------------|
| Regeln / Steuern        | Planen / Verhandeln                 |
| Bildverarbeitung        | Bildverstehen                       |
| Sprachverarbeitung      | Sprachverstehen                     |
| Sensomotorisches Lernen | taktisches und strategisches Lernen |

weitere Def. des Bildverstehens:

*Bildverstehen* wird def. durch die Aufgabe, Objekte, Szenen oder Situationen zu erkennen, wobei für die Bildelemente und Modelle sowohl in der Extraktions-, als auch in den Vergleichsoperationen vorwiegend symbolische Darstellungen verwendet werden.

weitere Begriffe:

### *Mustererkennung*

- Prozess der Identifikation von Objekten (anhand von bereits gelernten Beispielen)
- Menschen lernen dies anhand von gemeinsamen und unterschiedlichen Merkmalen von Objekten oder anhand von formalen Regeln – dazu ist Intelligenz nötig.

### *Muster:*

- Texturen, spezielle Objekte (Prototypen), geometrische oder topologische Strukturen, Schlüsselreize (z.B. menschliches Gesicht; Kindchenschema)
- auch: Verhaltensmuster, Melodie, grammatikalische Formen

### *Klassen in der Bildanalyse:*

Zusammenfassung verschiedener Objekte unter gemeinsamen Eigenschaften, ausgewählt aus der Menge der Eigenschaften, die ein Objekt beschreiben  
(Beispiele: Wolken, Blätter, Personen...)

### *Identifikation:*

Akt der Zuordnung eines (Bild-) Objektes zu einer Klasse (von Mustern), zu der das Objekt gehört.

### *Erkennung (recognition):*

(Wieder-) Erkennen von etwas, das man bereits einmal gesehen hat, ohne es notwendigerweise einer Klasse zuzuordnen.

### *Klassifikation (Klassifizierung):*

Sortierung von Objekten verschiedener Klassen in Gruppen, so dass jede Gruppe nur Objekte einer Klasse enthält.

Beispiel: Klassifikation von Pilzen nach der (botanischen) Art anhand von Größe, Farbe, Form, Textur.

Klassifikationskriterien sind nicht fest, sondern hängen stark von der Aufgabenstellung ab.

*automatische Mustererkennung:*

Entwicklung von Maschinen und / oder Algorithmen, die (neue) Objekte anhand ihrer *Merkmale* (bekannten) Klassen zuordnen können.

*Merkmale* (Attribute) können qualitativ oder quantitativ, kontinuierlich oder diskret sein.

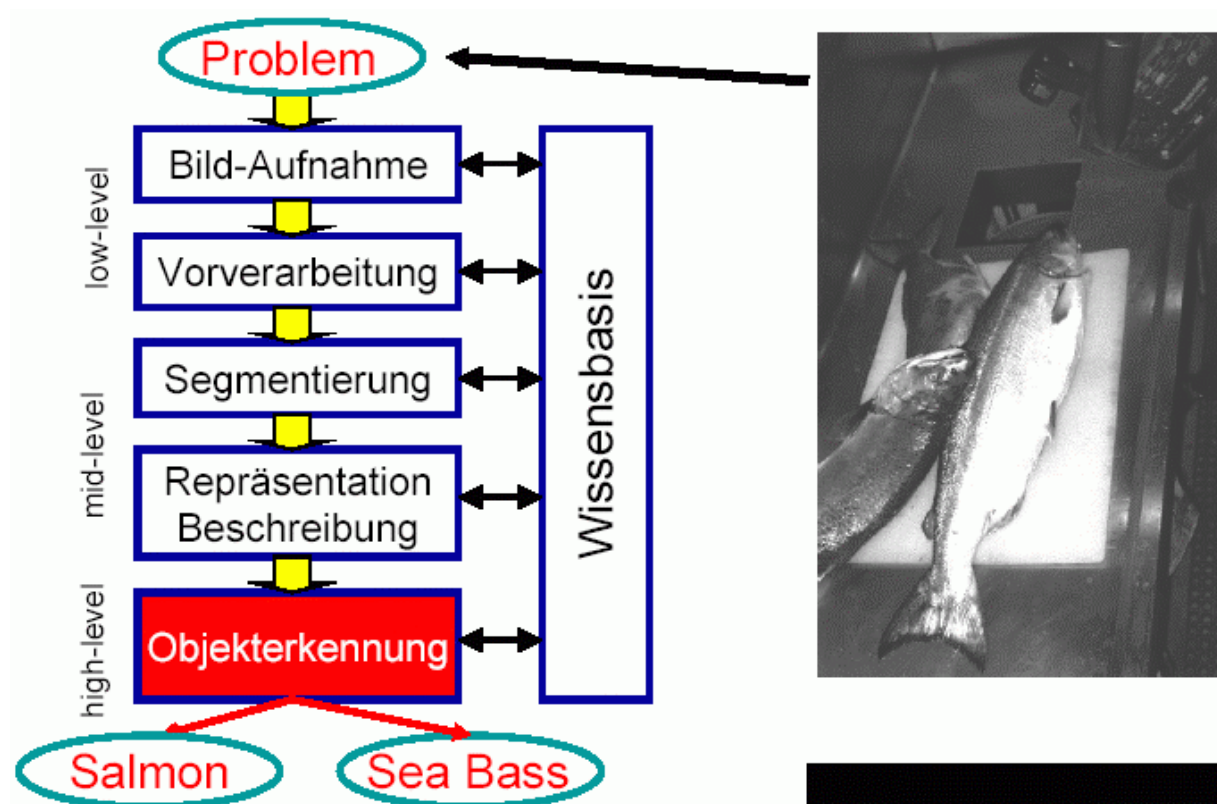
- Aus der Statistik bekannte Einteilung von Merkmalen nach der "Skalenebene":  
nominal, ordinal, metrisch, intervallskaliert.

Für Klassifikation nach metrischen / intervallskalierten Merkmalen gibt es eine Reihe von Standardverfahren (siehe späteres Kapitel).

*Erstes Beispiel:*

Klassifikation von Fischen

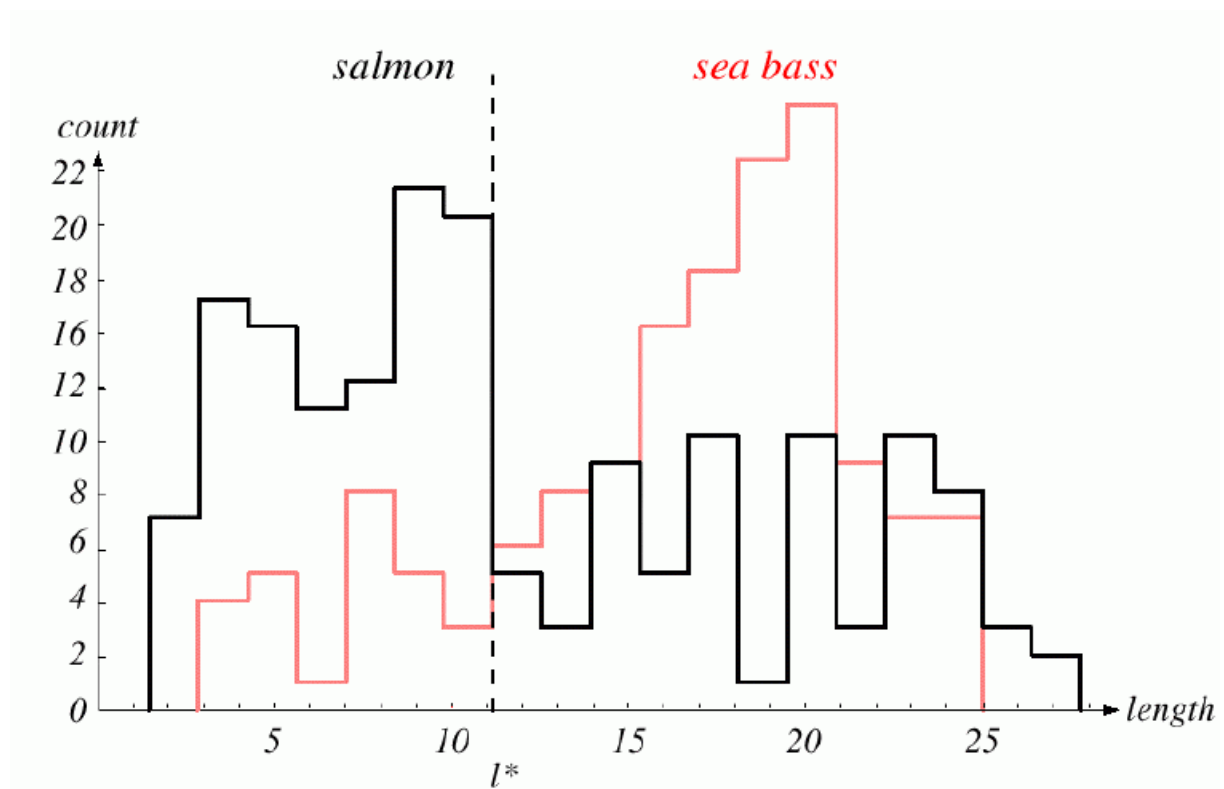
(aus Beichel 2002)



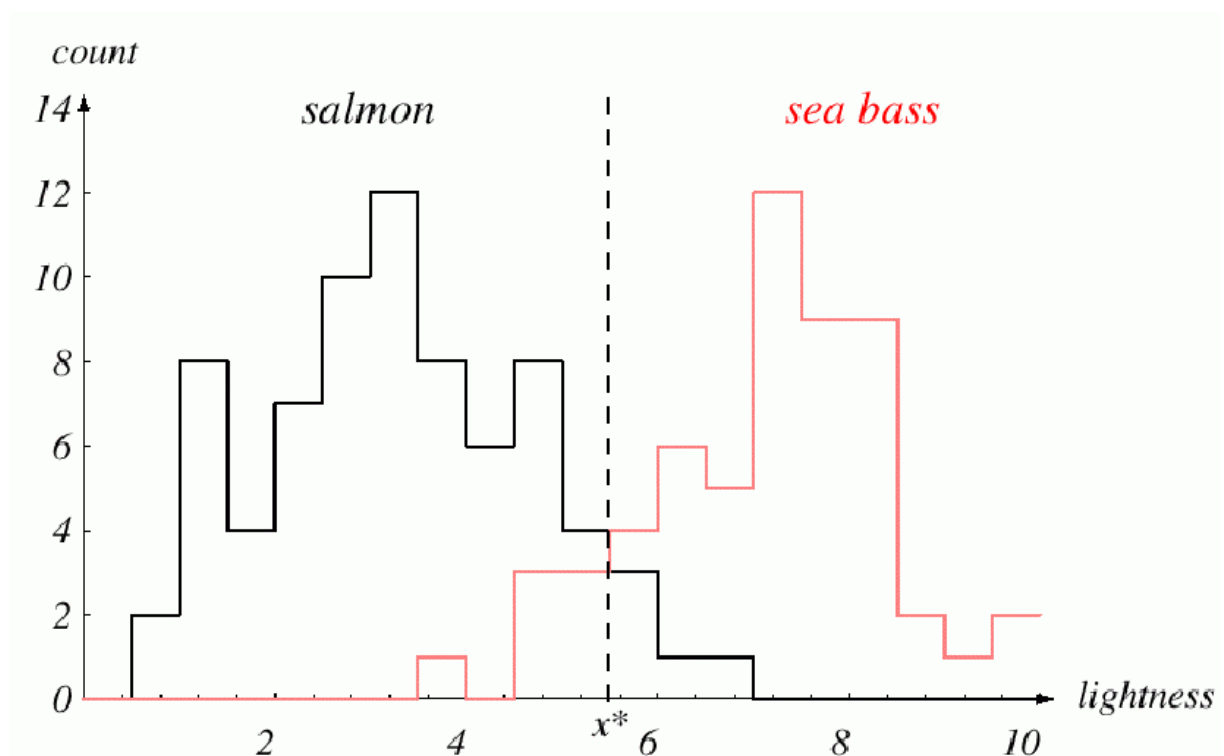


metrische Merkmale: Länge und Helligkeit

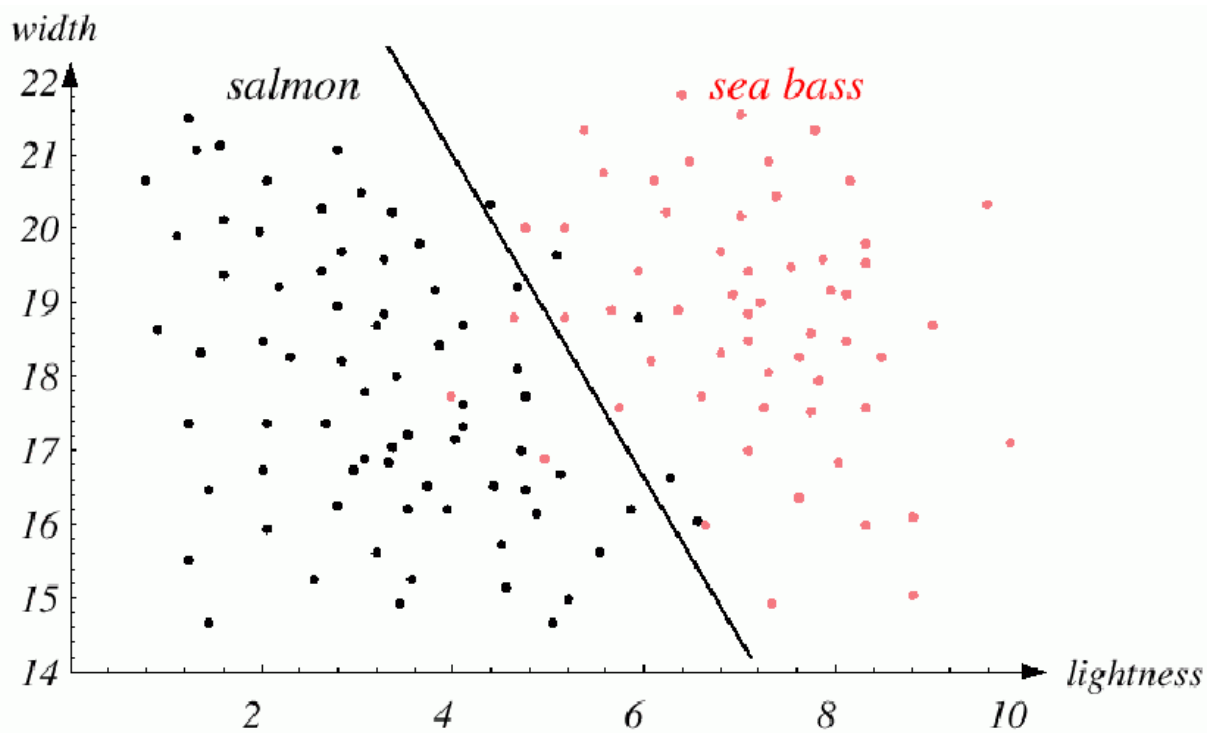
Merkmal Länge: Histogramm



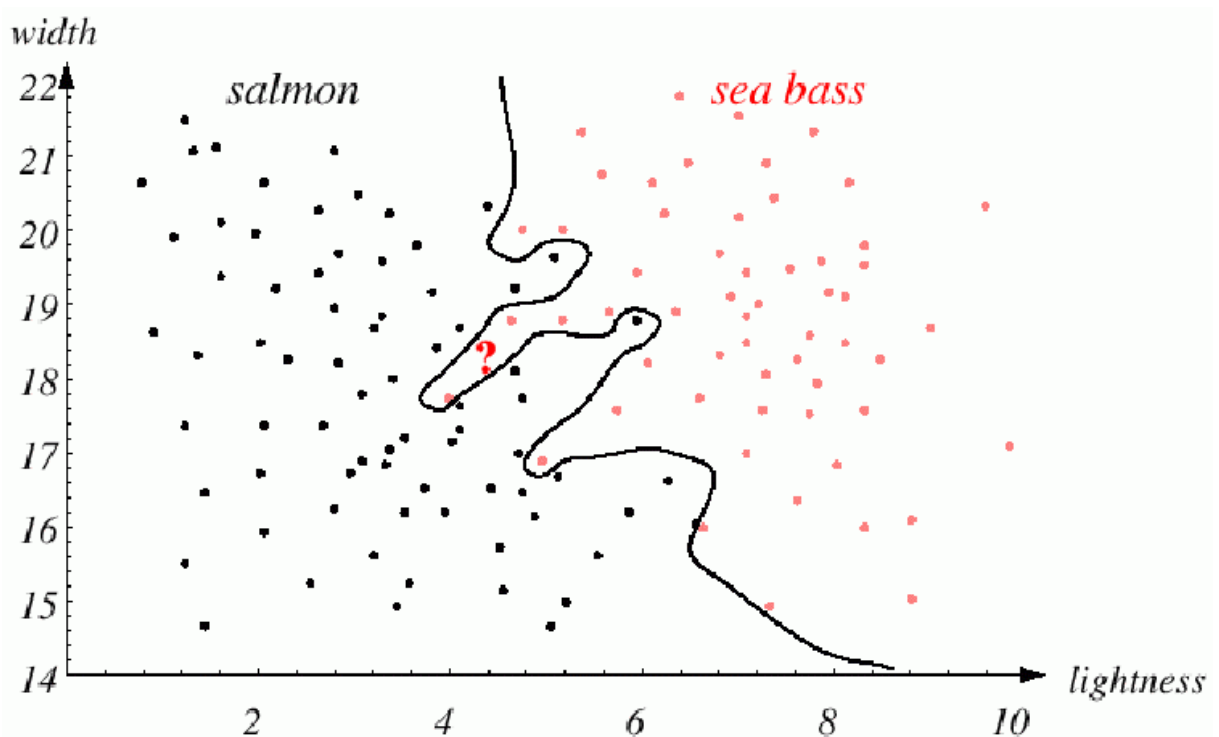
Merkmal Helligkeit: Histogramm



2-dim. Merkmalsraum, Unterscheidung von Klassen:

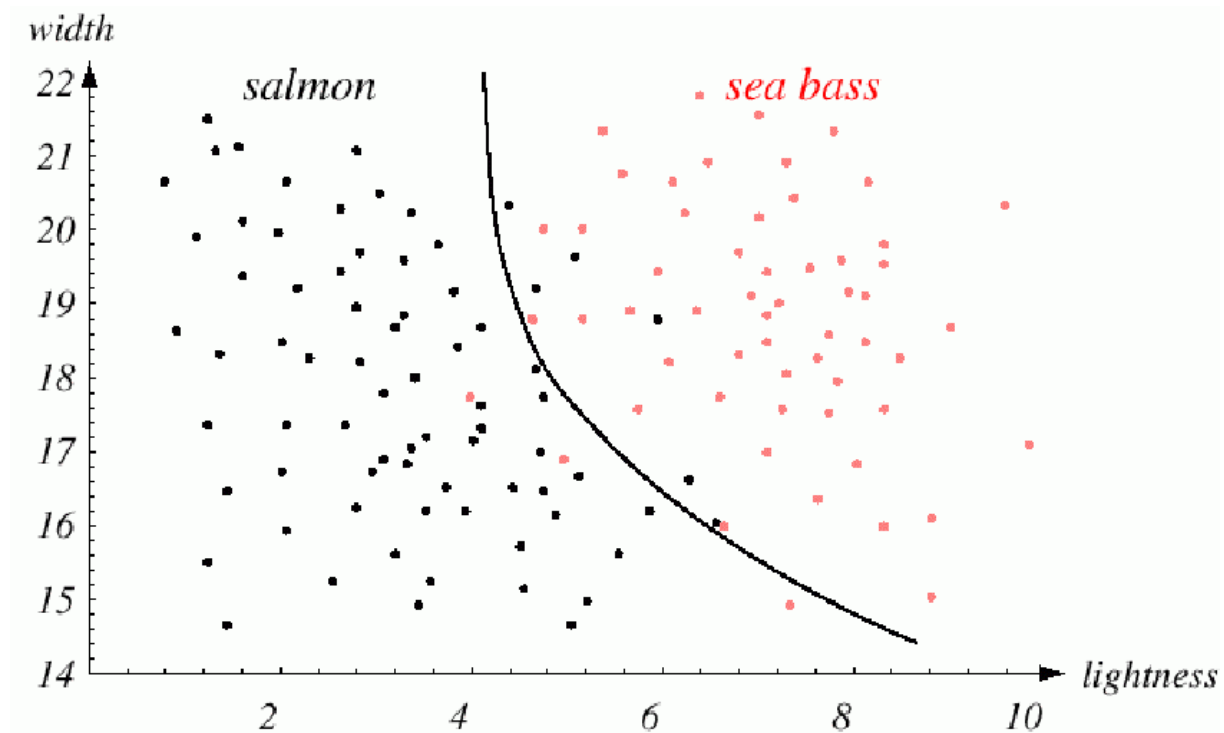


Für die Komplexität des Klassifikators muss ein vernünftiger Kompromiss gefunden werden:



zu komplex  
linearer Klassifikator zu einfach (zu viele Fehlzugeordnungen)

⇒ Kompromiss

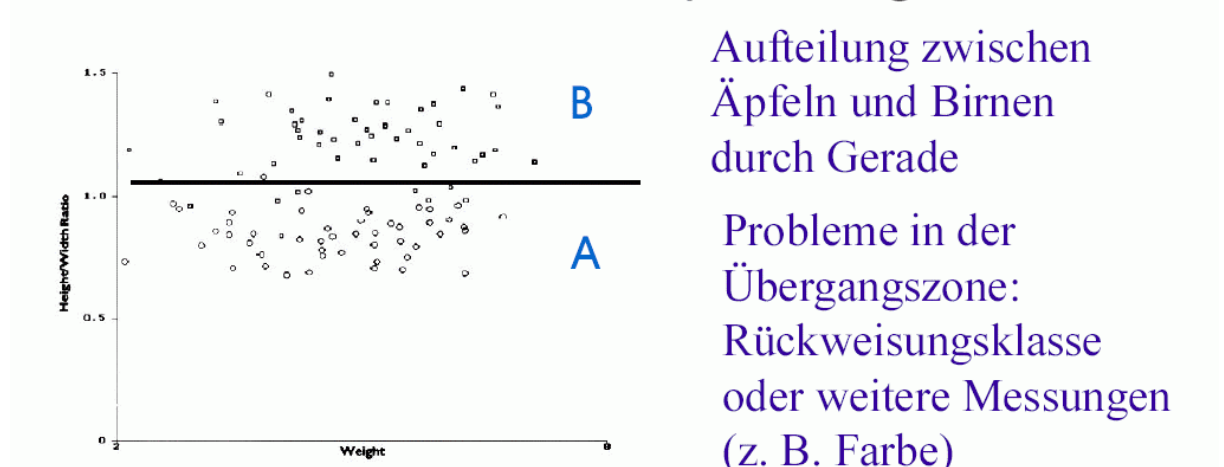


Beispiel mit 3 Merkmalen:

Unterscheidung von Äpfeln und Birnen

3 Messungen: Höhe  $h$ , Breite  $w$ , Gewicht  $W$

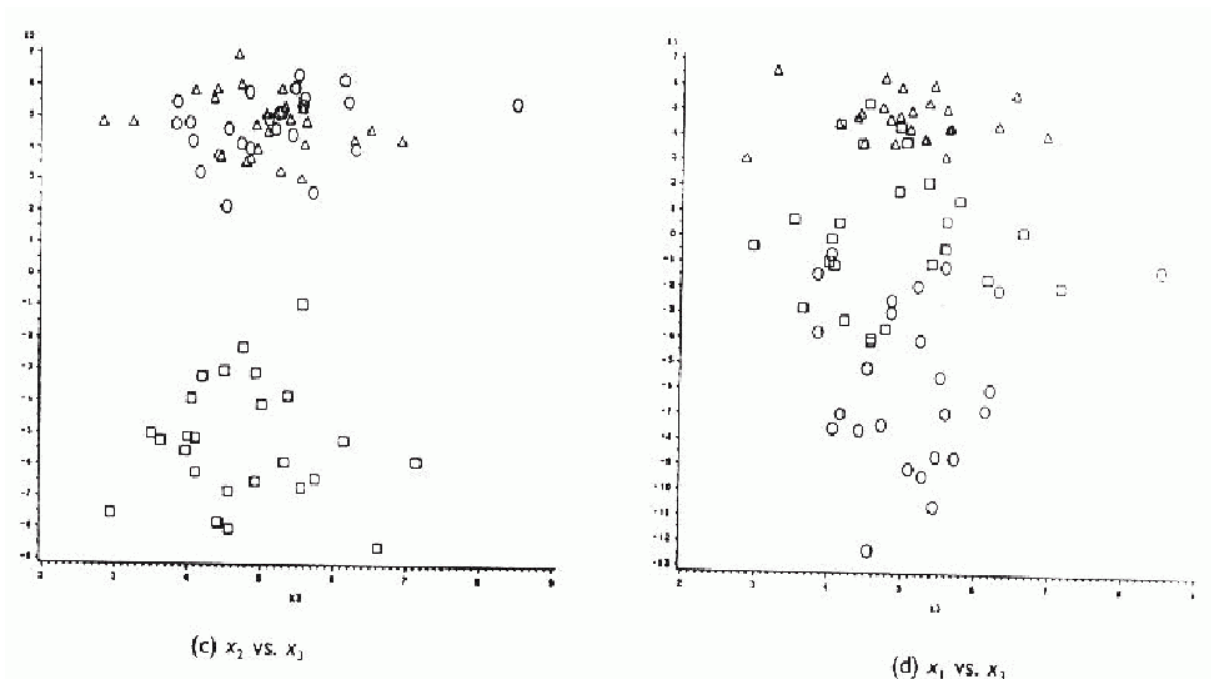
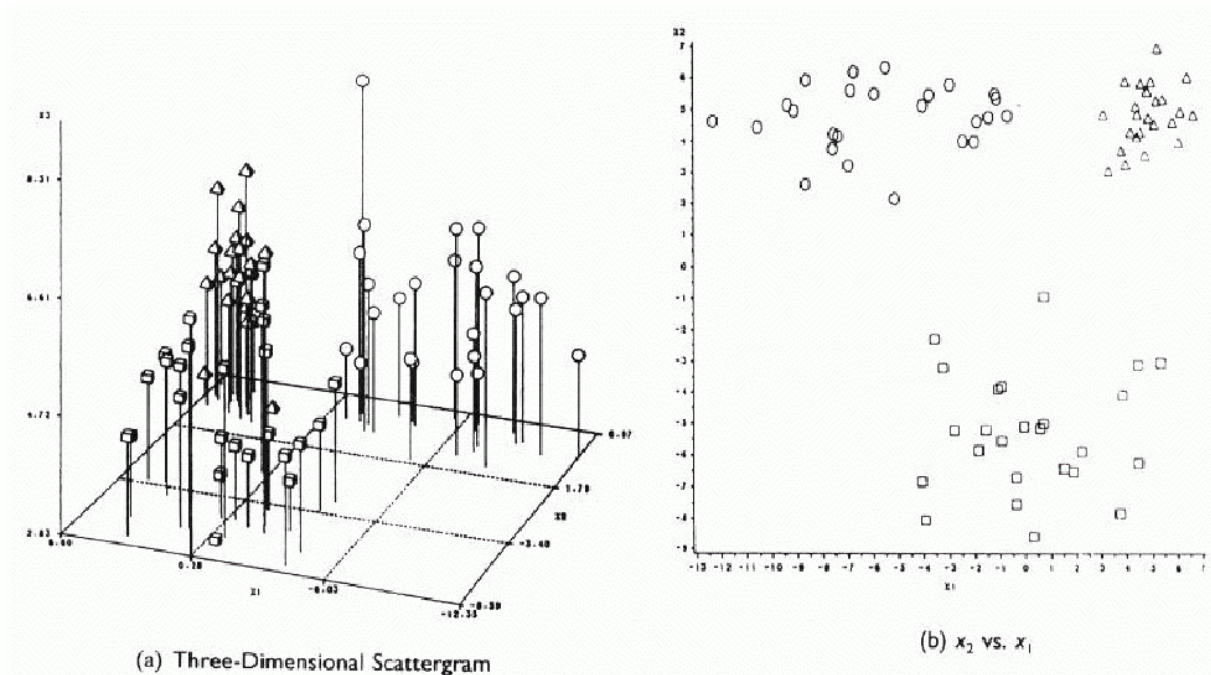
Birnen höher als breit, Äpfel umgekehrt



Aufteilung zwischen Äpfeln und Birnen durch Gerade

Probleme in der Übergangszone: Rückweisungsklasse oder weitere Messungen (z. B. Farbe)

### 3-dimensionaler Merkmalsraum, 3D-Darstellung und Projektionen:



⇒ aus der 3D-Darstellung ist nicht sofort ersichtlich, ob sich die Objekte trennen lassen

- Projektion auf  $x_1x_2$ -Ebene zeigt Trennbarkeit, nicht aber Projektion auf die anderen beiden Koordinatenebenen
- ggf. verschiedene Ansichten testen!

### *Auswahl der Merkmale:*

die Merkmale sollten

- möglichst invariant sein bzgl. Translation, Skalierung, Rotation etc.
- relevant sein: Hervorhebung von Unterschieden zwischen verschiedenen Klassen (*discrimination*)
- effektiv bestimmbar sein

Wie repräsentieren wir

- qualitative (nominale) Merkmale,
- Faktenwissen zur Unterstützung der Klassifikation,
- prozedurales Wissen?

Wissensrepräsentationsmethoden der KI

*Produktionssysteme (Regelsysteme)*

- häufig in Expertensystemen benutzt
- besonders geeignet zur Darstellung von prozeduralem Wissen

Ein KI-Produktionssystem besteht aus den Hauptkomponenten:

- Daten
- Menge von Produktionsregeln (Operationen)
- Kontrollsystem.

Produktionsregel: hat die allgemeine Form

*if Prämisse then Conclusio*

oder, bei prozeduraler Interpretation:

*if Bedingung X ist erfüllt then Aktion Y ist ausführbar*

Das Kontrollsystem analysiert, welche Prämissen aufgrund der aktuellen Daten erfüllt sind, und entscheidet, welche der somit anwendbaren Regeln ausgeführt werden sollen.

Die Conclusio (Aktion) kann dann eine Änderung der Daten bewirken.

Als Programmier-Paradigma (Fallregel- oder van Wijngaarden-Paradigma):

typische Unterschiede zum klassischen prozeduralen (von Neumann-) Paradigma

- keine lokalen Daten
- keine starre Sequenz von Programmbefehlen
- keine hierarchische Organisation
- kein Aufruf einer Regel von einer anderen Regel

Beispiel:

Wissensbasis:

|       |          |      |               |
|-------|----------|------|---------------|
| Daten | object A | is_a | ball          |
|       | object B | is_a | ball          |
|       | object C | is_a | shoe          |
| Regel | if       | ball | then circular |

Die Regel kann auf die Daten zu Objekt A und B angewandt werden; es kann z.B. die Information abgeleitet werden, dass 2 der Objekte "circular" sind.

Mit einer rein enumerativen Wissensrepräsentation müsste zur Gewinnung dieser Information eine Auflistung der Objekte etwa so aussehen:

```
object A is_a (ball, circular)
etc.
```

was wesentlich speicherplatzintensiver wäre.

*Strategien der Problemlösung* mit Produktionssystemen:

Ausgehend von einem Startzustand (gegebene Menge von Daten) soll ein Zielzustand durch Regelanwendungen erreicht werden.

- vollst. Graph eines Produktionssystems: verbindet jeden Zustand mit seinen Nachfolgezuständen bzgl. Regelanwendung.

Aufgabe der Kontrollstrategie: Finde einen Weg vom Start- zum Zielzustand.

- *unwiderrufliche* (irrevocable) Kontrollstrategien
- *versuchende* (tentative) Kontrollstrategien (z.B. Backtracking; Graphensuchalgorithmen)
- *Forward-Produktionssysteme*: gehen vom Startzustand aus; versuchen, durch Anwendung von Regeln (F-rules) das Ziel zu erreichen
- *Backward-Produktionssysteme*: gehen vom Ziel aus; Anwendung von B-rules, um zum Startzustand zu kommen
- *Bidirectional PS*: Kombination beider Strategien, gehen gleichzeitig von Start- und Zielzustand aus

## *Prädikatenlogik erster Ordnung*

- Darstellung von Faktenwissen und Regelwissen (Inferenzen)
- sehr einheitliche Repräsentationsform mit etabliertem Theorie-Hintergrund
- Programmiersprache PROLOG
- Nachteil: keine Darstellung von unvollständigem oder unscharfem Wissen.

### Bausteine: *Formeln*

Syntaktische Grundbausteine der Formeln:

- Prädikatensymbole
- Konstantensymbole
- Variablensymbole
- Funktionssymbole

jedes Prädikaten- und jedes Funktionssymbol besitzt eine nichtverschwindende Anzahl von Argumenten (übl. Klammer-schreibweise)

*Term* : Konstantensymbol | Variablensymbol | Funktionssymbol mit Termen als Argumenten

*atomare Formel* : Prädikatensymbol mit Termen als Argumenten

Syntaktisch korrekte *Formeln* werden aus atomaren Formeln mittels Konnektoren (und, oder, nicht, Implikation, Äquiv.) zusammengesetzt

Erweiterung: Zulassen von Quantoren ( $\forall$ ,  $\exists$ ).

### *Interpretation* einer Formel:

stellt die Verbindung zwischen den Symbolen und einem gegebenen Problembereich (z.B. Bildinformationen) her

- jedem Konstantensymbol wird ein Objekt, jedem Prädikatensymbol eine Relation und jedem Funktionssymbol eine Funktion des jeweiligen Problemkreises zugeordnet



Beispiel: kollinear( $x, y, z$ )

die Belegung der Variablen  $x, y, z$  mit den Konstantensymbolen  $A, B, C$  liefert für die Formel den Wert TRUE, wenn die  $A, B, C$  entsprechenden Punkte auf einer Geraden liegen.

Variablen in einer Formel im Wirkungsbereich eines Quantors heißen *gebundene Variablen*.

*freie Variablen*: außerhalb d. Wirkungsbereichs eines Quantors.  
Ein *Satz* ist eine Formel ohne freie Variablen.

- Subkalkül des Prädikatenkalküls 1. Ordnung:  
*Propositionenkalkül*, verzichtet auf Variablen und Quantoren
- Erweiterung des Prädikatenkalküls 1. Ordnung:  
*Prädikatenkalkül 2. Ordnung* – Variablen und Quantoren nicht nur für Objekte, sondern auch für Relationen zugelassen ( $\Rightarrow$  Gewinn an Mächtigkeit, aber Verlust an Entscheidbarkeit)

Zwei Formeln heißen *äquivalent*, wenn sie für jede gemeinsame Interpretation den selben Wahrheitswert liefern.

Ist  $M$  eine Menge von Formeln und  $I$  eine Interpretation, so heißt  $I$  *Modell* für  $M$ , wenn sich für jede Formel in  $M$  unter  $I$  der Wahrheitswert TRUE ergibt.

$M$  heißt *erfüllbar*, wenn für  $M$  ein Modell existiert.

Eine Formel  $f$  *folgt logisch aus*  $M$ , wenn jedes Modell für  $M$  auch Modell für  $f$  ist.

Eine *Inferenzregel* ist eine Vorschrift, um aus einer gegebenen Menge von Formeln neue Formeln abzuleiten.

Beispiel: *modus ponens*

$$\frac{f \quad f \rightarrow g}{g}$$

*reductio ad absurdum*

$$\frac{f \rightarrow g \quad f \rightarrow \neg g}{\neg f}$$

Resolutionsregel (wichtig in PROLOG):

$$\frac{f \vee g \quad \neg g \vee h}{f \vee h}$$

- Eine Menge  $R$  von Inferenzregeln heißt *widerspruchsfrei*, wenn jede Formel, die sich mittels  $R$  aus einer Menge  $M$  von Formeln ableiten lässt, auch logisch aus  $M$  folgt.
- $R$  heißt *vollständig*, wenn jede Formel, welche logisch aus einer Menge  $M$  von Formeln folgt, auch mittels Regeln aus  $R$  aus  $M$  gewonnen werden kann.

(Anwendung: automatisches Theorem-Beweisen)

Im Prädikatenkalkül 1. Ordnung gibt es vollständige und widerspruchsfreie Inferenzregeln. (Speziell: Beweisverfahren mit Hilfe der Resolutionsregel, Verwendung in PROLOG.)

*Partielle Entscheidbarkeit* des Prädikatenkalküls 1. Ordnung:

- Wenn ein Theorem aus den Axiomen folgt, kann immer in endlich vielen Schritten durch Anwendung der Inferenzregeln ein Beweis gefunden werden.
- Aber: wenn eine Formel *nicht* aus den Axiomen folgt, ist die Termination des Ableitungsprozesses mittels der Inferenzregeln nicht sichergestellt.

*ein Logisches Programm:*

Darstellung von Wissen in Form von Formeln der Prädikatenlogik 1. Ordnung, evtl. in spezieller Form (z.B. Horn-Klauseln; PROLOG).

*Beispiel:*

- (1) Jedes Bild enthält mindestens ein Objekt.
- (2) Aufnahme 1 ist ein Bild und enthält nur Würfel.
- (3) Aufnahme 1 enthält mindestens einen Würfel.

Übersetzung in Formeln:

- (1)  $\forall x (\text{BILD}(x) \rightarrow \exists y (\text{IN}(y, x)))$
- (2)  $\text{BILD}(\text{AUFNAHME1})$
- (2a)  $\forall x (\text{IN}(x, \text{AUFNAHME1}) \rightarrow \text{WÜRFEL}(x))$
- (3)  $\exists x (\text{WÜRFEL}(x) \wedge \text{IN}(x, \text{AUFNAHME1}))$

Beachte: Formel (3) folgt logisch aus (1), (2) und (2a).

Verbindung zu Produktionssystemen:

Anwendung der Inferenzregeln im Prädikatenkalkül lässt sich als Anwendung eines Produktionssystems auffassen.

speziell in der PROLOG-Variante: tentatives Backward-Produktionssystem mit Backtracking.

## Wissensdarstellung mit relationalen Strukturen

### Definition Relationalstruktur RS

$$RS = (N, R)$$

$N = \{n_1, n_2, \dots, n_l\}$  = Knotenmenge,  $n_i$  beschreiben häufig Objektkomponenten.  
ten.  $R = \{R_1, \dots, R_m\}$  = Relationenmenge;

$$R_i \subseteq \underbrace{N \times \dots \times N}_{s_i} = N^{s_i}$$

$R_i$ :  $s_i$ -stellige Relation auf der Knotenmenge  $N$ ;  $i=1, \dots, m$   $s_i \geq 1$ .

Spezialfall: p-partiale Relationalstruktur

$$N = N_1 \cup N_2 \cup \dots \cup N_p, \quad N_i \cap N_j = \emptyset \\ i \neq j, \quad i, j = 1, \dots, p$$

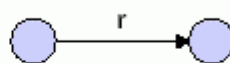
$$R_i \subseteq \underbrace{N_{i_1} \times N_{i_2} \times \dots \times N_{i_{s_i}}}_{s_i}$$

Spezialfall: (gerichtete) *Graphen* ( $s_i = 2$ )

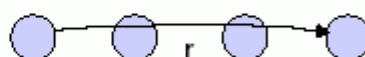
grafische Darstellung von Relationen:

#### Graphical representation

binary relation:



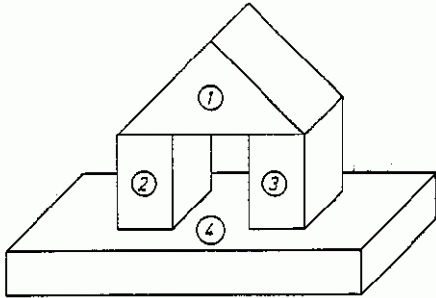
n-ary relation:



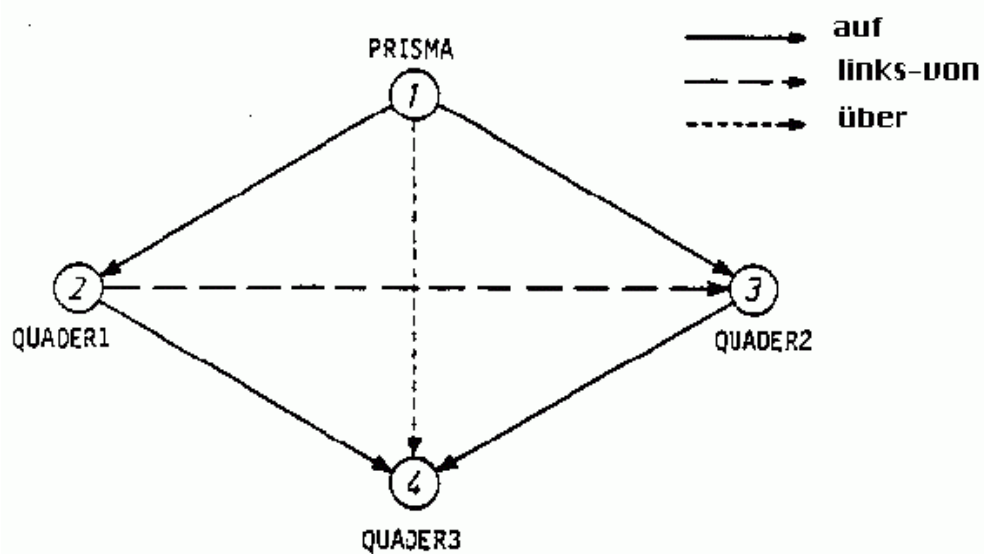
"hypergraph"

Beispiel (aus Levi 2002):

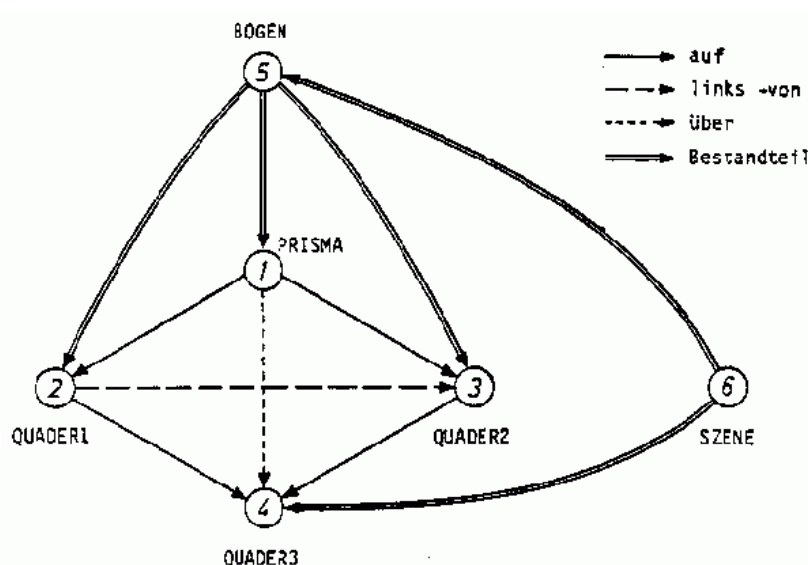
Szene



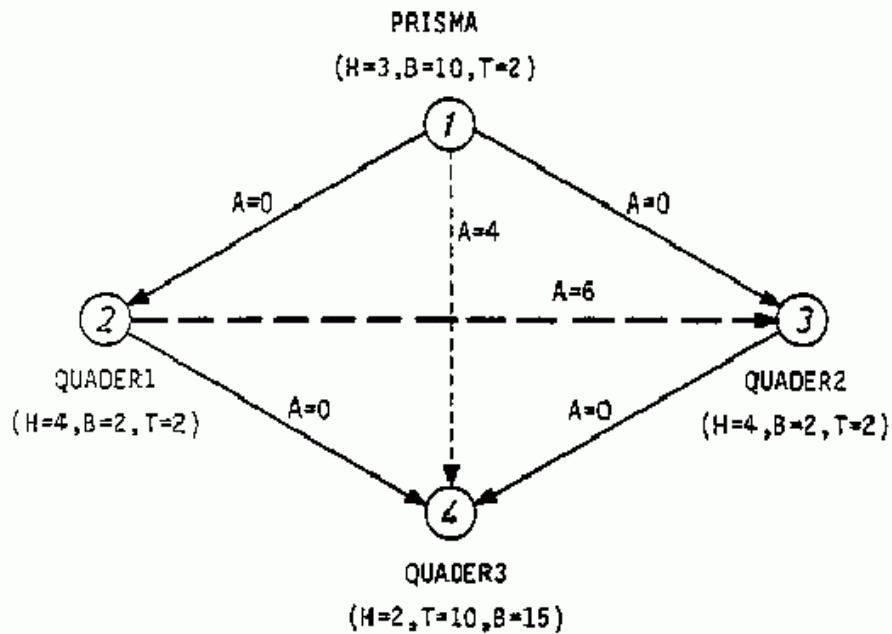
Graph:



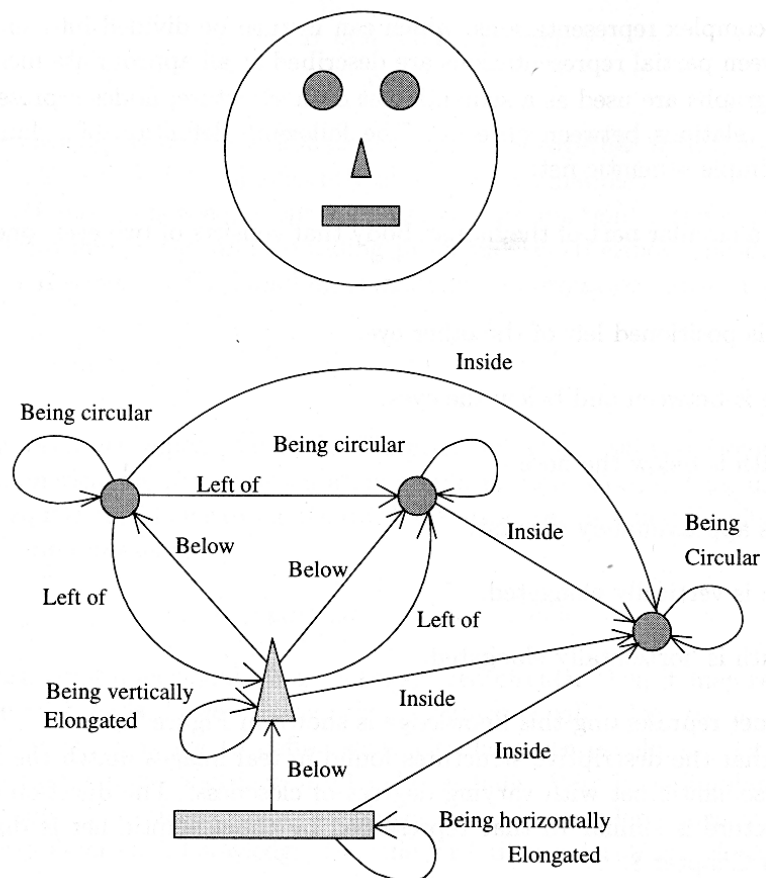
Ergänzung durch hierarchische Szenenbeschreibung (mit zusätzlichem Kantentyp *part-of*):



attributierter Graph (Höhe, Breite, Tiefe, Abstand):

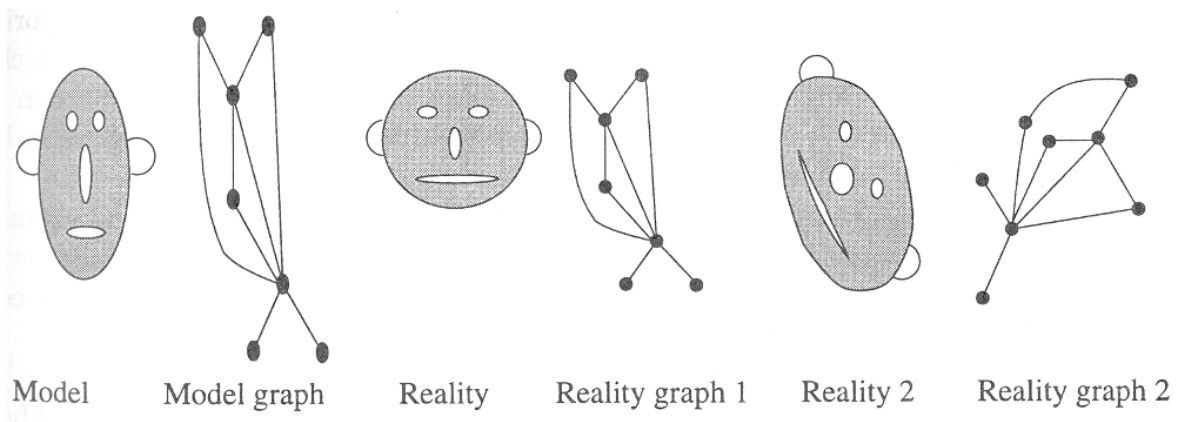


Relationen-Modell eines Gesichts (aus Sonka et al. 1999):

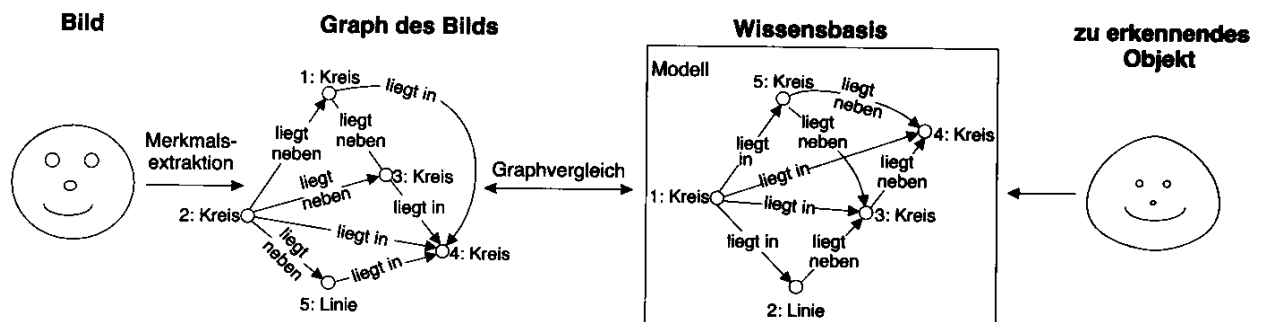


vgl. Szenengraph in VRML und Java 3D (siehe Vorlesungs- und Übungsskript Computergrafik)

⇒ Wiedererkennung von Objekten wird zu einem Graph-Matching-Problem (Auffinden von Graphen-Isomorphismen)

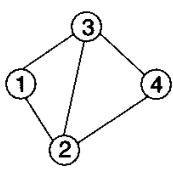


- Die zu erkennenden Objekte müssen vorab als Modelle abgelegt werden (Repräsentation als Graphen).
- Aus einem zu analysierenden Bild müssen die Merkmale extrahiert werden, die man für den Aufbau eines Graphen-Modells benötigt.
- Auf der Basis der extrahierten Merkmale wird der Graph einer Szene erstellt.
- Nun können die Graphen verglichen werden (Suche nach Subgraph-Isomorphismen).

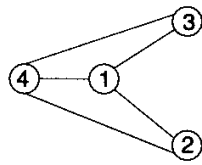


## Beispiele für Graphen-Isomorphien:

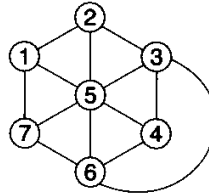
- Graph 1 ist isomorph zu Graph 2 und zu mehreren Subgraphen von Graph 3
- Graph 4 ist auf zweifache Weise Subgraph von Graph 3



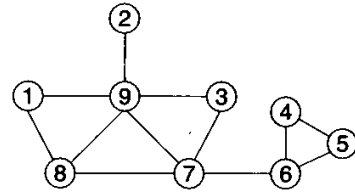
Graph 1



Graph 2



Graph 3

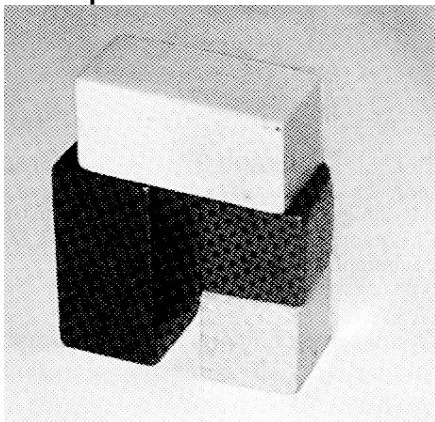


Graph 4

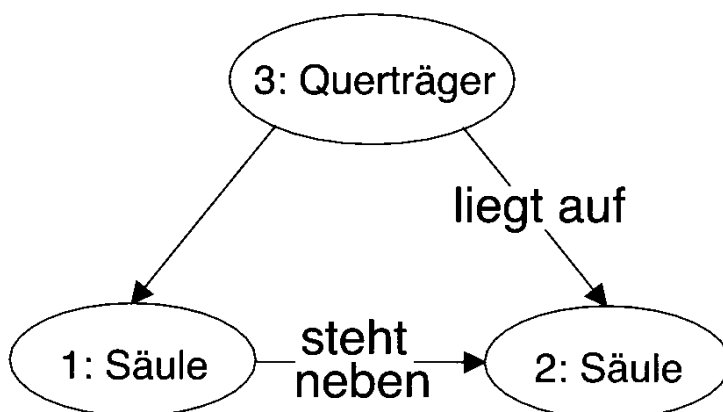
übergeordnete Konzepte können durch bestimmte, attributierte Subgraphen definiert werden

z.B. "Torbogen"

## Beispiel

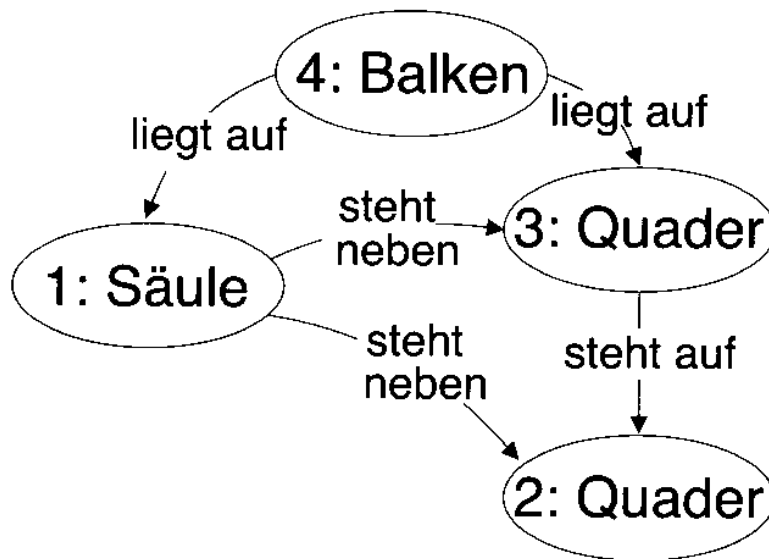


Graphen-Definition eines Torbogens:





Problem: der aus dem obigen Bild extrahierte Graph würde anders aussehen als dieser Prototyp.

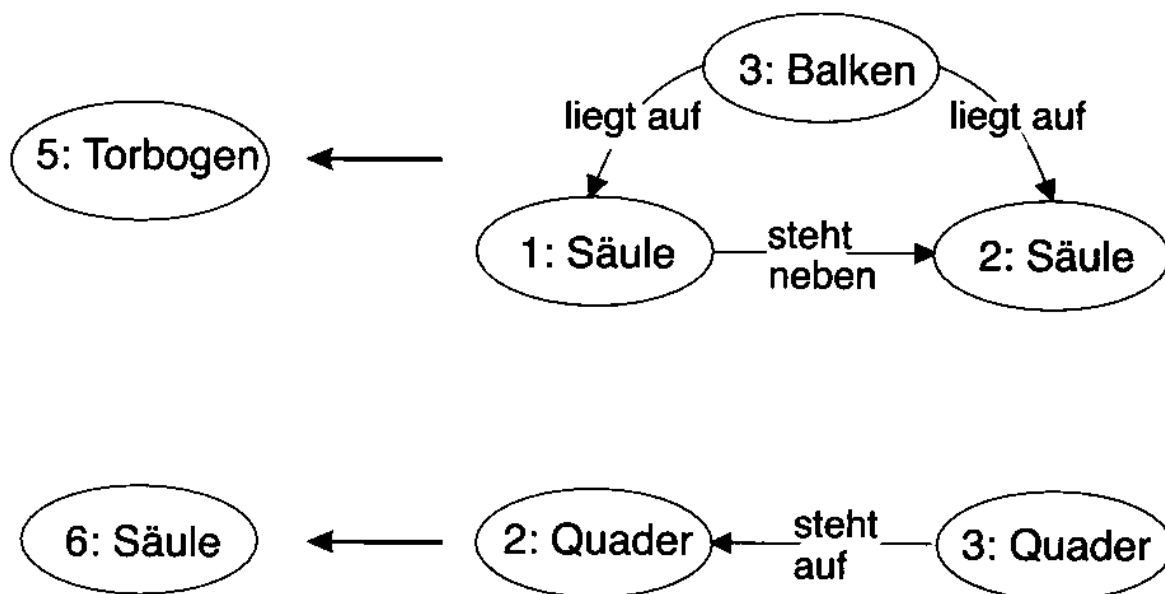


⇒ würde nicht als Torbogen erkannt!

Abhilfe:

- Verwendung von Graph-Grammatiken
- dadurch Kombination (Schachtelung) der Objekterkennungsschritte

einfache Graph-Grammatik für das Torbogen-Beispiel:



einfaches Isomorphie-Matching reicht jetzt nicht mehr für die Objekterkennung

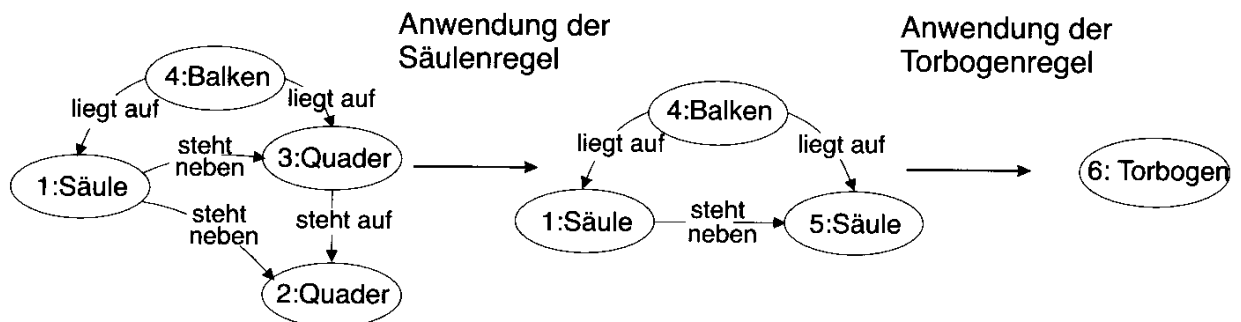
- es muss berücksichtigt werden, wie man in einem Graphen G einen Teilgraphen G', der auf der rechten Seite einer Regel steht, durch den Graphen G'' auf der linken Seite der Regel ersetzt. Es muss dabei gewährleistet sein, dass G'' auf dieselbe Weise in G eingebettet ist wie zuvor G' (Persistenz der Relationen): "korrekte Einbettung" von G''.
- es muss gesteuert werden, in welcher Reihenfolge die Grammatikregeln angewendet und die Graphen ersetzt werden

dafür 2 Strategien:

*datengetriebene Verarbeitung (bottom-up)*

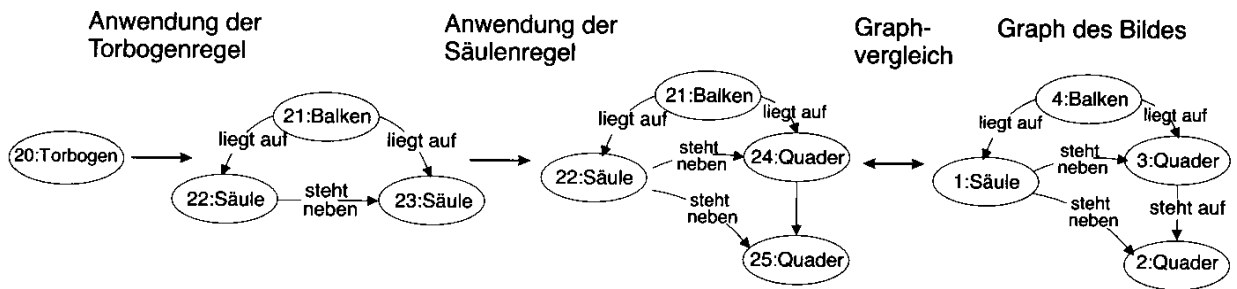
man geht vom Graphen des Eingabebildes aus und ersetzt Subgraphen solange anhand der Regeln, bis man zu einem oder mehreren Objekten gelangt ist, die als Zielobjekte def. sind. (Anwendung der Grammatik-Regeln von links nach rechts)

Im Bsp.: erst "Säule"-Regel, dann "Torbogen"-Regel anwenden



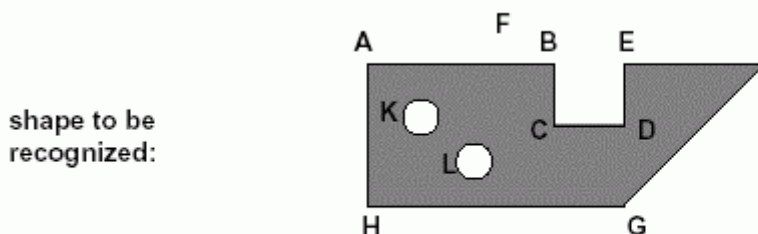
### modellgetriebene Verarbeitung (top-down)

man beginnt mit den Regeln, die Zielobjekte produzieren, und wendet die Regeln sukzessive von links nach rechts an – Synthese neuer Graphen, bis der synthetisierte Graph isomorph zum Graphen des Eingabebildes oder zu einem seiner Subgraphen wird.



Es sind auch Hybridverfahren aus beiden Verarbeitungsstrategien möglich.

Neben der reinen Graphen-Isomorphie müssen bei der Objekterkennung oft die Attribute in Bild und Modell verglichen werden:

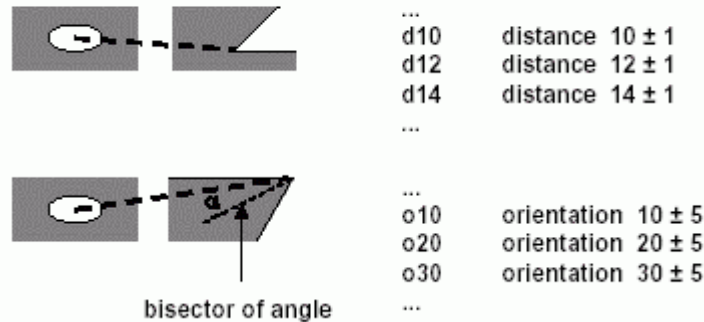


| primitive descriptive elements (nodes) |                 | properties |               |
|--|-----------------|------------|---------------|
|  | hole            | t          | type T1       |
|  | interior corner | f          | area          |
|  | exterior corner | a          | axes relation |
|  |                 | t          | type T2       |
|  |                 | w          | angle         |
|  |                 | t          | type T3       |
|  |                 | w          | angle         |

(aus Neumann 2001)

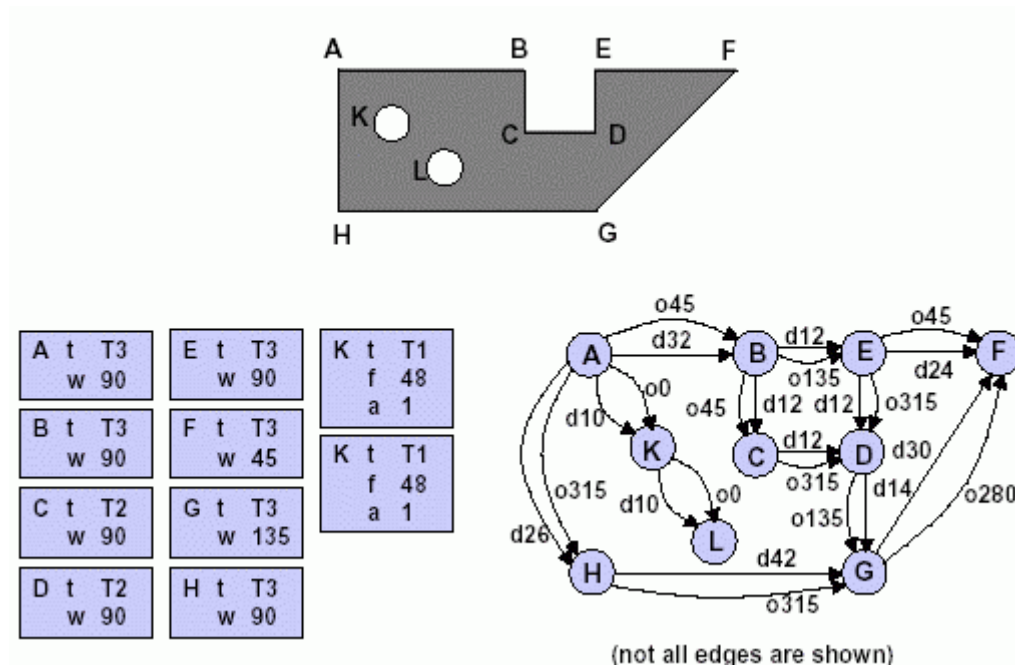
### Knotenattribute

relations between primitive descriptive elements (edges)



Kantenattribute

Darstellung des Modellobjekts als attributierter Graph:

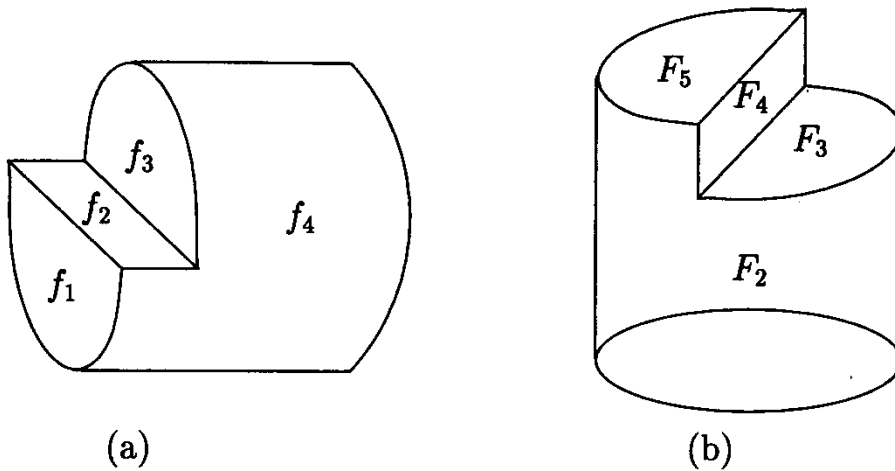


Zur Isomorphie-Feststellung mit einem gegebenen Szenengraphen wird ein neuer Graph definiert, der *Kompatibilitäts-Graph* (auch *Assoziationsgraph*). Jeder möglichen Knotenzuordnung zwischen Modellgraph und Szenengraph entspricht ein Knoten im Kompatibilitäts-Graphen.

- Knoten des Kompatibilitäts-Graphen: Paare mit kompatiblen Knoten-Attributen
- Kanten des Kompatibilitäts-Graphen: konsistente Zuordnungen mit kompatiblen Kanten-Attributen

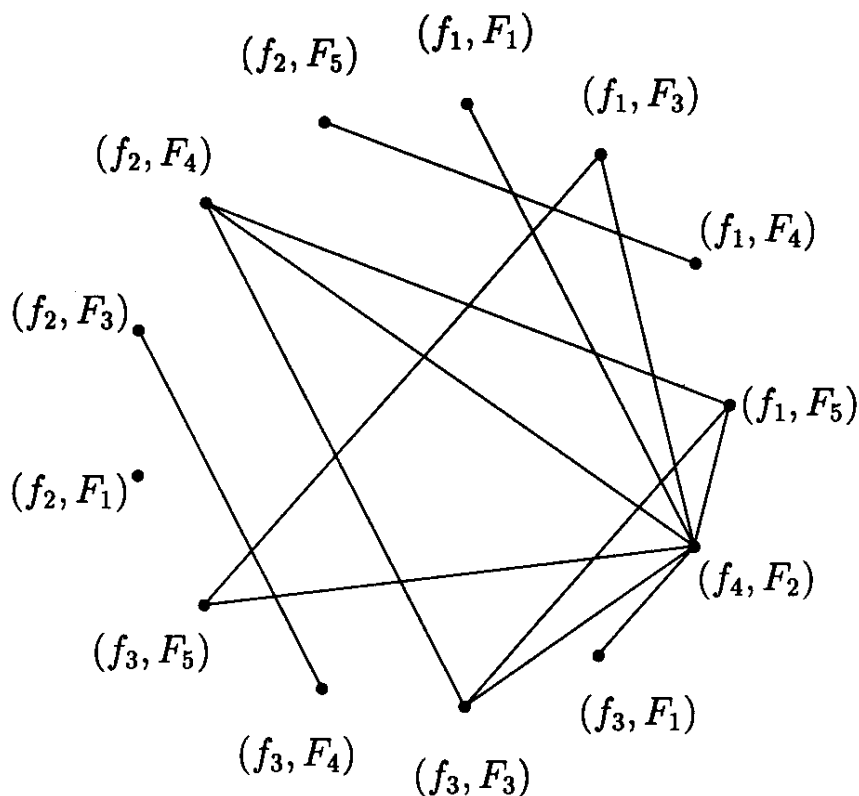
Beispiel:

Szene (a) und Modell (b)



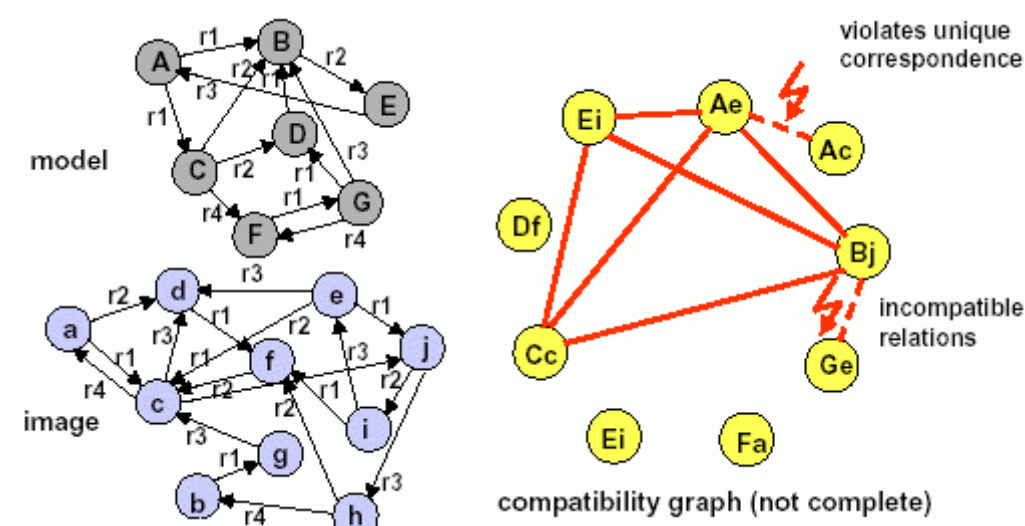
(aus Jiang & Bunke 1997)

zugehöriger Kompatibilitätsgraph:



vollständige Subgraphen (*Cliquen*) im Kompatibilitätsgraphen entsprechen partiellen konsistenten Zuordnungen

- Annahme: "richtige" Zuordnung entspr. *maximaler Clique* (hier die Clique, die  $(f_2, F_4)$  enthält)

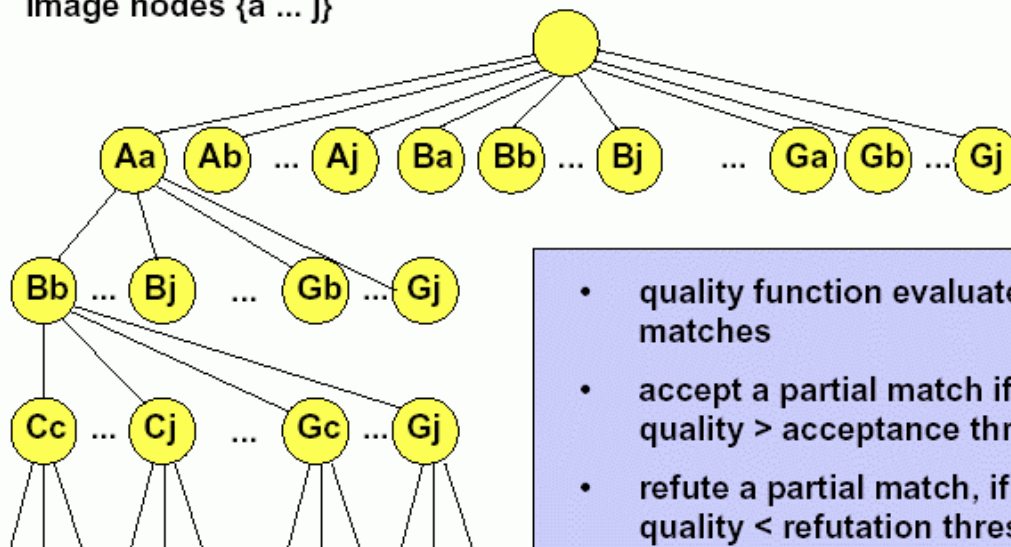


⇒ Matching-Problem wird zurückgeführt auf das Finden maximaler Cliques im Kompatibilitätsgraphen

- in der Literatur gibt es dafür Algorithmen, z.B. Bron & Kerbusch 1973
- aber: exponentieller Aufwand bzgl. Knotenzahl des Kompatibilitätsgraphen
- effiziente, aber suboptimale Lösungen verwenden heuristische Suchstrategien

Beispiel: relationales Matching mit heuristischer Suche

Stepwise correspondence search between model nodes {A ... G} and image nodes {a ... j}



- quality function evaluates partial matches
- accept a partial match if quality > acceptance threshold
- refute a partial match, if quality < refutation threshold

(aus Neumann 2001)

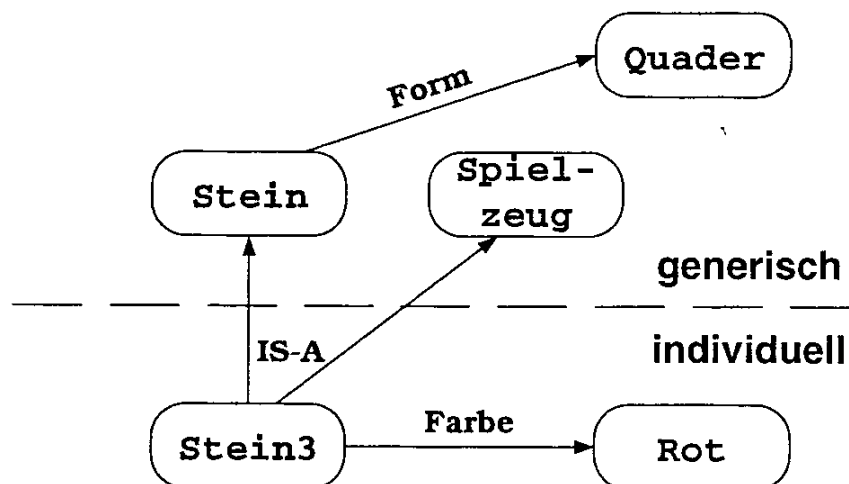
Erweiterung der Objektrepräsentation durch relationale Strukturen:

### *Semantische Netze*

(seit 60er Jahren; 70er: M. Minsky; Grundlage mehrerer KI-Sprachen und Expertensysteme)

semantisches Netz = spezialisierte, attributierte relationale Struktur, die auch Prozeduren enthalten kann, um bestimmte Relationen und Attribute zu berechnen, und die vom Vererbungsprinzip der objektorientierten Programmierung Gebrauch macht (nach Levi 2002)

semantische Netze können individuelle und generische Komponenten enthalten:



Vererbung (*inheritance*) erfolgt entlang der IS-A-Kanten (Instanz "Stein3" erbt Eigenschaften von "Stein").

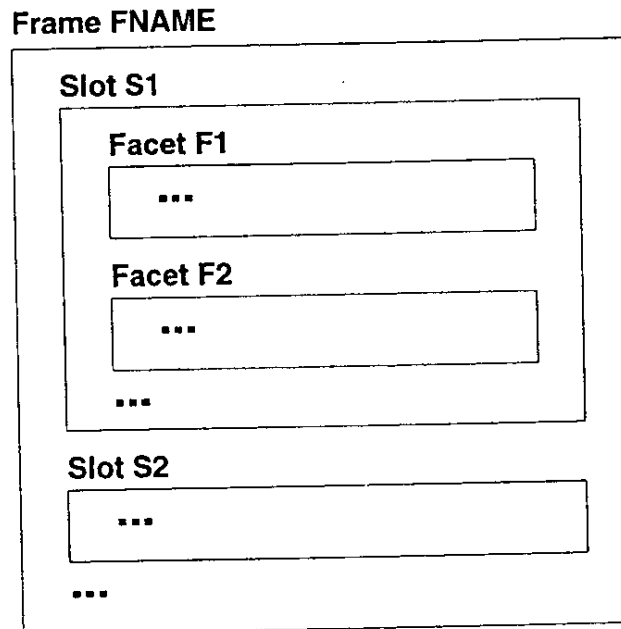
⇒  $\text{Form}(\text{Stein3})$  wird als "Quader" ausgewertet

weitere Eigenschaften:

- Verwendung von Defaults (vgl. OOP)
- *Dämonen* (demons): Prozeduren, die automatisch dann aufgerufen werden, wenn sie benötigt werden
- als Datenstruktur für die Knoten eines semant. Netzes oft verwendet: *Frames*

*Frame*: Datenstruktur, ähnlich der "Klasse" aus der OOP, jedoch für allgemeine Konzepte und spezielle Instanzen gleichermaßen verwendet

- enthält *Slots* (Merkmale, Relationen) mit zugehörigen *Facets* (Werte-Feldern)



spezielle Slots:

- AKO = a kind of: Generalisierungsrelation, verweist auf Superkonzept, invers zu INSTANCE, zeigt auf anderes Frame (Kante eines semant. Netzes)
- INSTANCE: Spezialisierung, verweist auf Subkonzept, invers zu AKO
- CLASSIFY: hat nur die Werte GENERIC oder INDIVIDUAL, vermerkt, ob generische Klasse ("Konzept") oder Instanz vorliegt
- PART OF: verweist auf übergeordnetes Frame im Sinne *räumlicher* Zugehörigkeit

spezielle Facets:

- VALUE: Wert des Slots
- REQUIRE: Bedingungen für Value (z.B. Wertebereich)
- DEFAULT: Default-Wert
- IF-ADDED, IF-REMOVED, IF-NEEDED: Dämonen

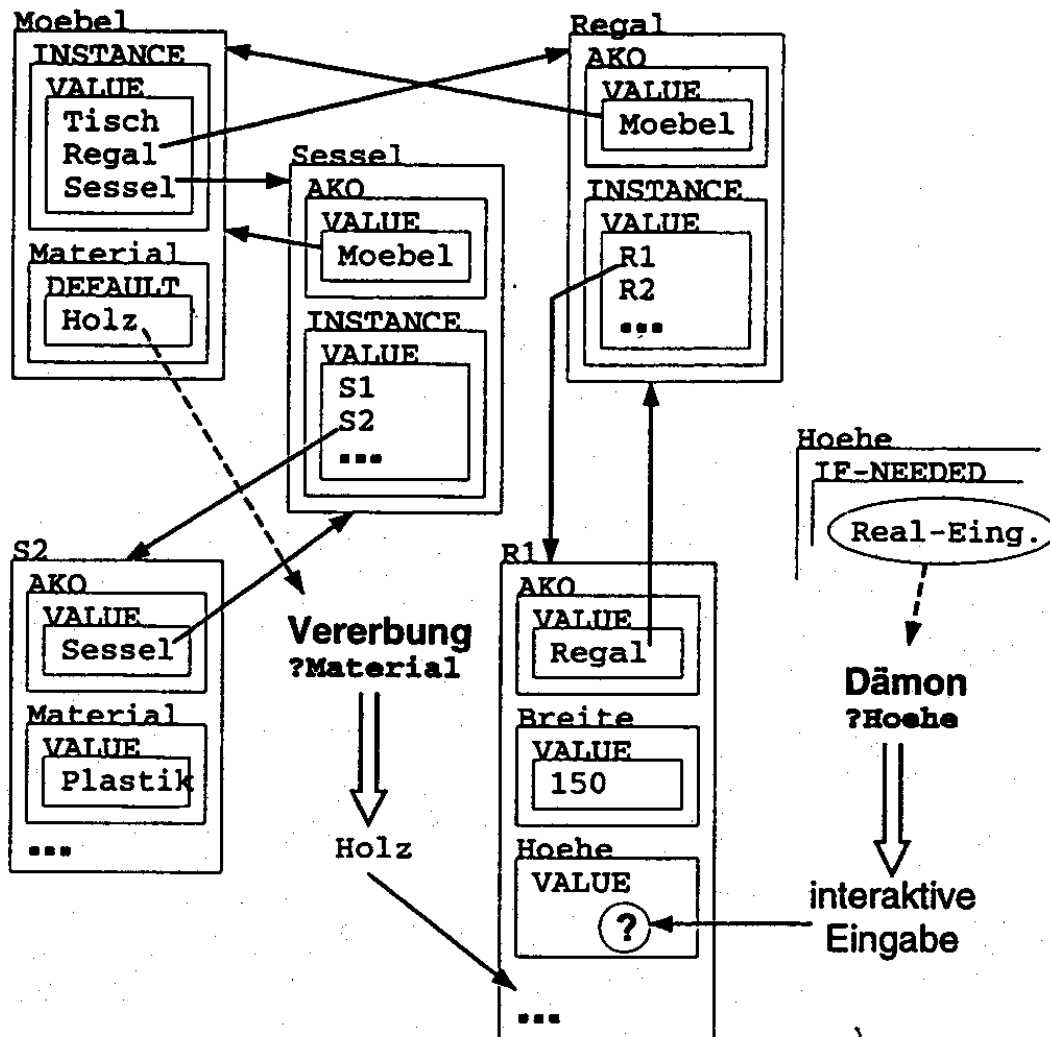


# Beispiel (aus Pinz 1994):

```

(FRAME (Moebel (INSTANCE
              (VALUE (Regal, Sessel, Tisch,...)))
        (Material
          (DEFAULT (Holz)))
        (CLASSY GENERIC)))
(FRAME (Sessel (AKO (VALUE (Moebel)))
        (CLASSY GENERIC)
        (INSTANCE (VALUE (S1, S2,...)))))
(FRAME (Regal (AKO (VALUE (Moebel)))
        (CLASSY GENERIC)
        (INSTANCE (VALUE (R1, R2,\dots )))
        (Hoehe (IF-NEEDED (Real-Eingabe "Hoehe"))))
...
))
(FRAME (S2 (CLASSY INDIVIDUAL)
        (Material (VALUE (Plastik)))
        (AKO (VALUE (Sessel)))))
(FRAME (R1 (CLASSY INDIVIDUAL)
        (AKO (VALUE (Regal)))
        (Breite (VALUE (150))))

```



Verbindung zur Prädikatenlogik:

- individuelle Frames (Instanzen) beinhalten nur Aussagen über konstante Objekte und lassen sich in prädikatenlogische Formeln ohne Variablen und Quantoren übertragen
- generische Frames (Konzepte) legen Klassen individueller Objekte fest und können als Formeln mit Allquantoren dargestellt werden

Nachteil der Wissensrepräsentation mit semantischen Netzen:

großer Speicherplatz- und Suchaufwand, Unübersichtlichkeit bei großen Netzen

- deshalb Kombination mit anderen Formen der Wissensdarstellung