

## 12. 3D-Bildanalyse

Extraktion von 3D-Information aus einem Bild / aus einer Bildsequenz wichtig für

- Szenenrekonstruktion
- wissensbasierte Bildverarbeitung und Bildverstehen
- Robotik-Anwendungen
- ...

Grundproblem:

*Projektion von 3D in 2D bewirkt Verlust der Tiefeninformation.*

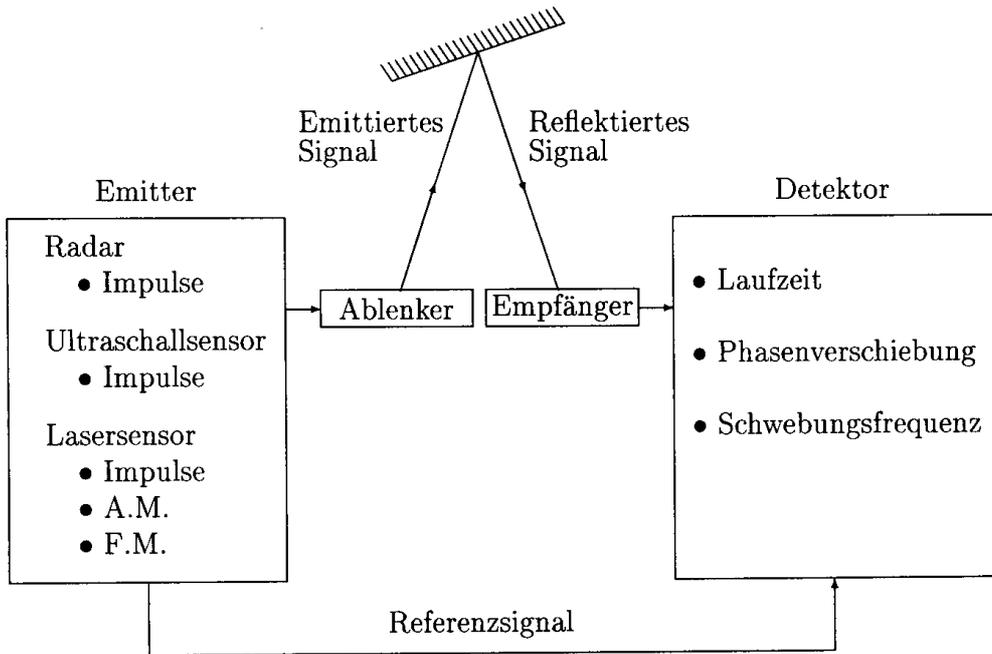
Abhilfe:

- Wiedergewinnung von Tiefeninformation durch Tiefensensoren (zusätzliche Messung),
- aktives Sehen (gezieltes Ausleuchten der Szene, verbunden mit 1 oder mehreren Aufnahmen),
- Aufnahmen von 2 oder mehr Kamerapositionen aus (vgl. Funktion der 2 menschl. Augen: Stereosehen, *Stereopsis*),
- Analyse einer monokularen Bildsequenz, die Bewegung enthält (unter schwachen Annahmen – "Shape from motion"),
- Ausnutzung von Informationen aus dem Einzelbild ("monokulares Stereosehen") unter starken Annahmen ("Shape from texture", "shape from shading"...) )

*Aktive Tiefengewinnung durch Tiefensensoren:*

Laufzeitverfahren

- Ultraschall, Radar, Laser
- direkte Messung der Laufzeit vs. indirekte Messung durch Bestimmung der Phasenverschiebung bzw. Schwebungsfrequenz

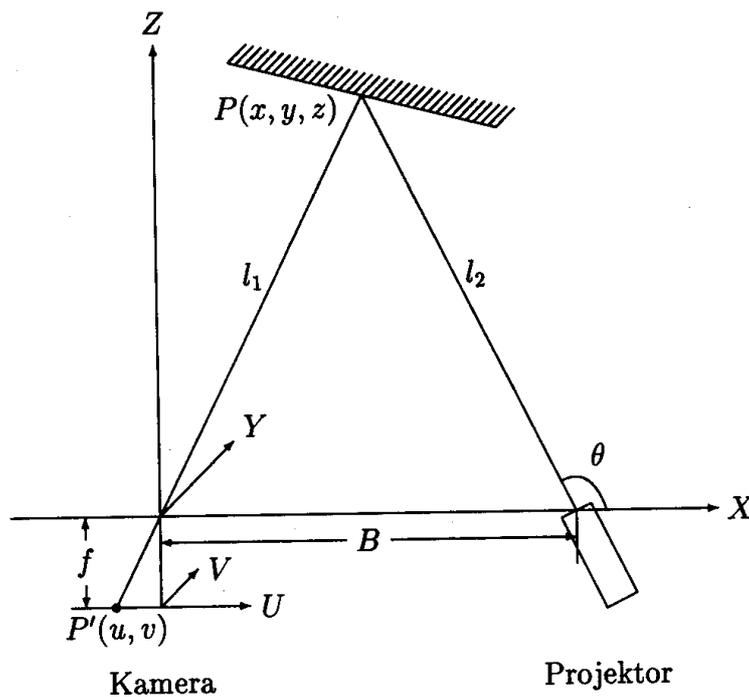


Prinzip der aktiven Tiefengewinnung durch Laufzeitmessung  
(aus Jiang & Bunke 1997)

*Aktive Tiefengewinnung durch Triangulation:*

einfachstes Verfahren: Projektor mit bekannter Position und Orientierung wirft Lichtstrahl in die Szene

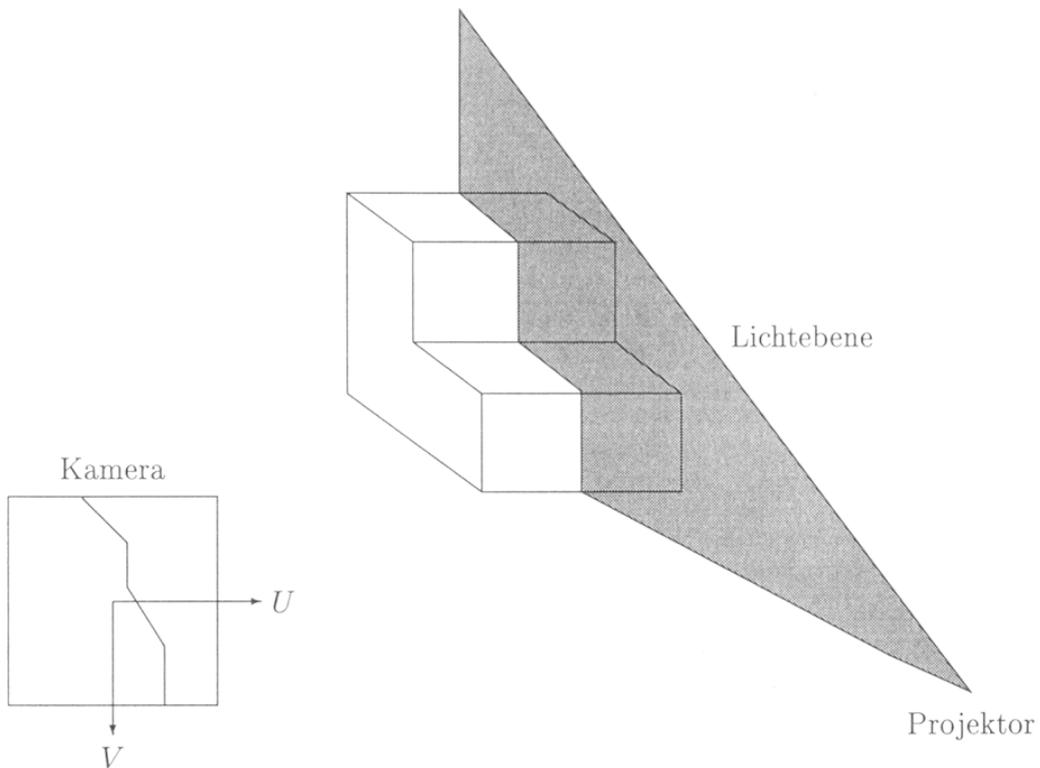
- Dreiecksgeometrie liefert den Abstand (siehe Übung)



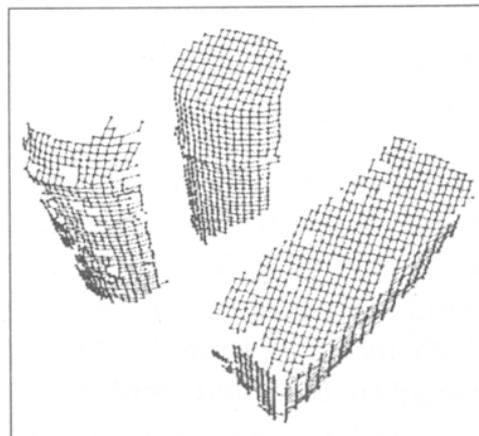
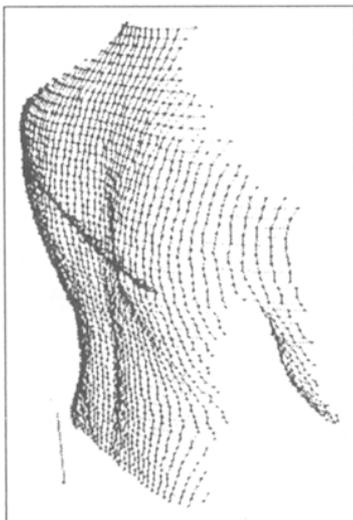
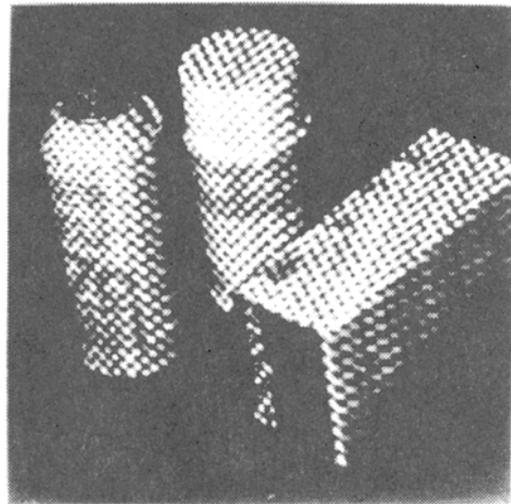
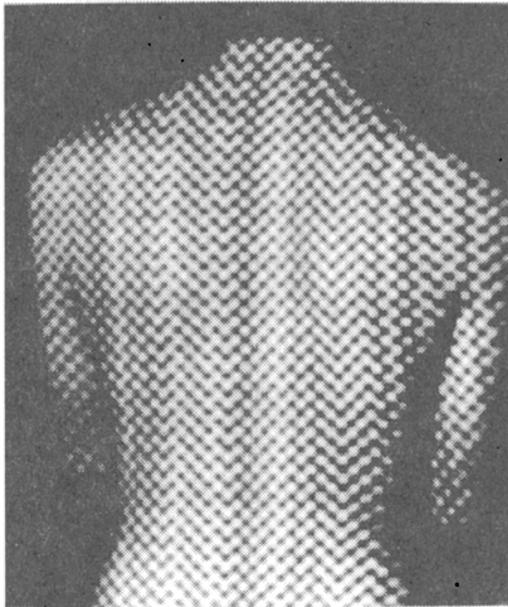
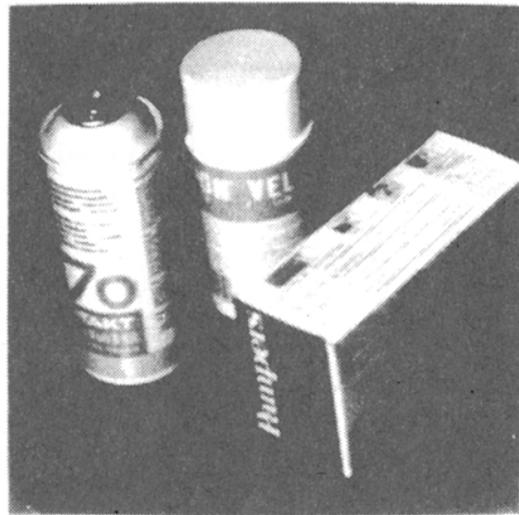
Nachteil der punkweisen Vermessung mit diesem Verfahren: 2-dimensionale Ablenkung des Mess-Strahls, lange Aufnahmezeiten

Abhilfe: Lichtebene statt Lichtstrahl, dann nur 1-dim. Ablenkung nötig

- resultierende "Lichtlinie" als Profillinie im Kamerabild (z.B. Detektion als hellster Punkt in jeder Bildzeile, oder mit spezieller Spektralcharakteristik)



*Aktive Tiefengewinnung durch codierte Projektionsmuster*  
z.B. binäre Muster



## *Kamerakalibrierung*

Bestimmung der intrinsischen (Brennweite, Linsenverzerrung...) und extrinsischen (Position, Orientierung) Kamera-Parameter aus einem oder mehreren Bildern für bestimmte Kamera-Szenen-Konfigurationen.

Notwendig für

- Messung von Objekten aus dem Bild (Fotogrammetrie)
- Navigation als sich bewogender Beobachter in der Szene
- Stereobild-Auswertung
- Kompensation von Kamera-Verzerrungen

verschiedene Fälle:

- Szene enthält Punkte mit bekannter Position (Passpunkte)
- unbekannte Szene ohne Passpunkte: mehrere Ansichten sind erforderlich
  - bekannte Kamerabewegung
  - unbekannte Kamerabewegung ("Kamera-Selbstkalibrierung")

Prinzip:

- aus der Projektionsgleichung (vgl. Vorlesung Computergrafik) wird ein lineares Gleichungssystem hergeleitet
- gewöhnlich > 6 Passpunkte erforderlich
- Gleichungssystem meist überbestimmt  $\Rightarrow$  beste Näherungslösung im Sinne kleinster Fehlerquadratsumme bestimmen (Technik: QR-Zerlegung von Matrizen; Einzelheiten siehe Sonka et al. 1999, S. 448 ff.).

## Stereoverfahren zur Tiefenbestimmung

Prinzip:

- 2 oder mehr Bilder derselben Szene werden von verschiedenen Kamerapositionen aus aufgenommen.
- Kann in jedem der Bilder die Lage eines bestimmten Szenepunktes identifiziert werden, lässt sich seine *räumliche* Position aus den bekannten Parametern der Kameras und der Kamera-Anordnung ermitteln (Triangulation).

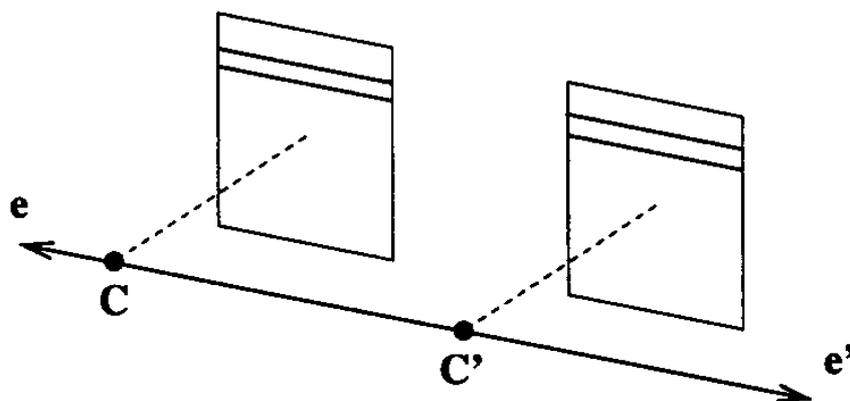
*Korrespondenz-Problem:*

Zuordnung einander entsprechender Punkte (Kanten, Objekte...) in den beiden Stereoaufnahmen.

Standard-Stereogeometrie:

aus Gründen effizienterer Zuordnung wird in der Praxis eine Standard-Stereogeometrie bevorzugt:

- die beiden Bildebenen sind identisch, d.h. sie entsprechen lediglich 2 unterschiedlichen Ausschnitten derselben Ebene
- die Zeilen der beiden Stereobilder liegen parallel zur Verbindungsgeraden zwischen den optischen Zentren der beiden Kameras,
- die Bildkoordinatensysteme der beiden Kameras werden so definiert, dass zueinander kollineare Bildzeilen die gleiche Zeilenkoordinate haben.



Einschränkung für mögliche Zuordnungen:  
*Epipolare Geometrie*

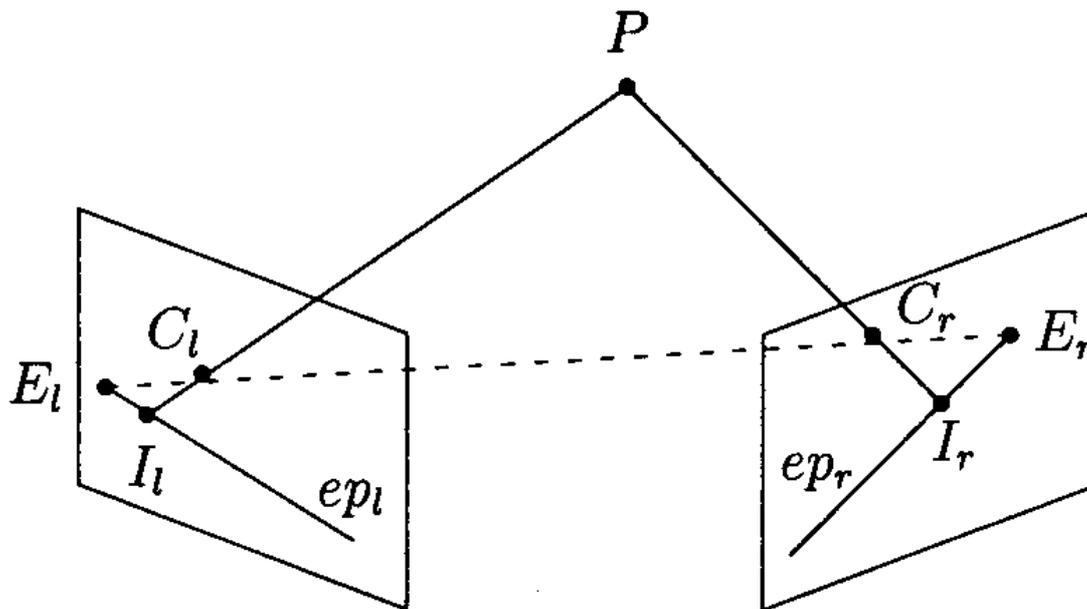
$C_l$  = optisches Zentrum der linken Kamera

$C_r$  = opt. Zentrum der rechten Kamera

$I_l$  = Bild (Image) des Punktes  $P$  in der Bildebene der linken Kamera

$I_r$  = Bild des Punktes  $P$  in der Bildebene der rechten Kamera

Der Punkt  $P$  muss auf der Geraden  $I_l C_l$  liegen, also muss der entsprechende Bildpunkt  $I_r$  bei der rechten Kamera auf der *Abbildung dieser Geraden in die Bildebene der rechten Kamera* liegen: "epipolare Linie"  $ep_r$  (Schnitt der epipolaren Ebene  $I_l C_l C_r$  mit der rechten Bildebene). Analog für die linke Bildebene. (Schnittpunkte der Verbindungslinie  $C_l C_r$  mit den Bildebenen: *Epipole*.)



⇒ bei der Suche nach dem korrespondierenden Punkt im rechten Bild zu einem gegebenen Punkt im linken Bild braucht *nur auf einer Linie* (der Epipolar-Linie des Punktes) gesucht zu werden!

(Reduktion der Dimension des Suchraumes um 1.)

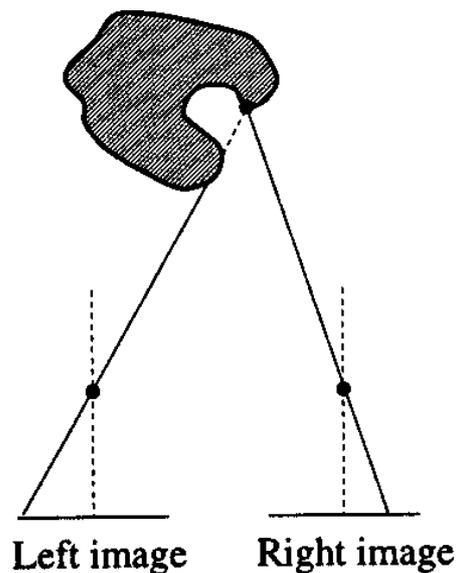
Suchbereich kann weiter eingeschränkt werden auf ein Intervall. (Details bei Jiang & Bunke 1997.)

Lösung des Korrespondenzproblems:  
Zuordnungsverfahren (Matching)

- Korrelationsverfahren
- Relaxationsverfahren
- dynamische Programmierung (Optimierung)

Problem:

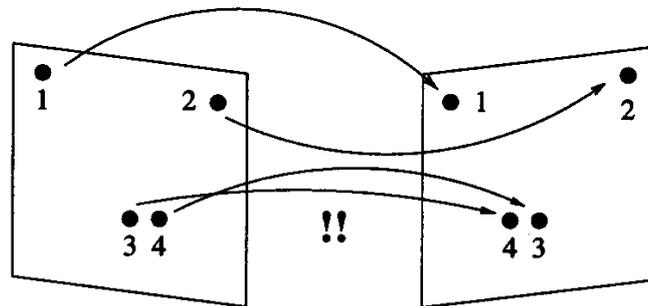
Verdeckung von Punkten in einer der Ansichten kann  
Zuordnung unmöglich machen!



Ausnutzung zusätzlicher Annahmen bei der  
Korrespondenzsuche:

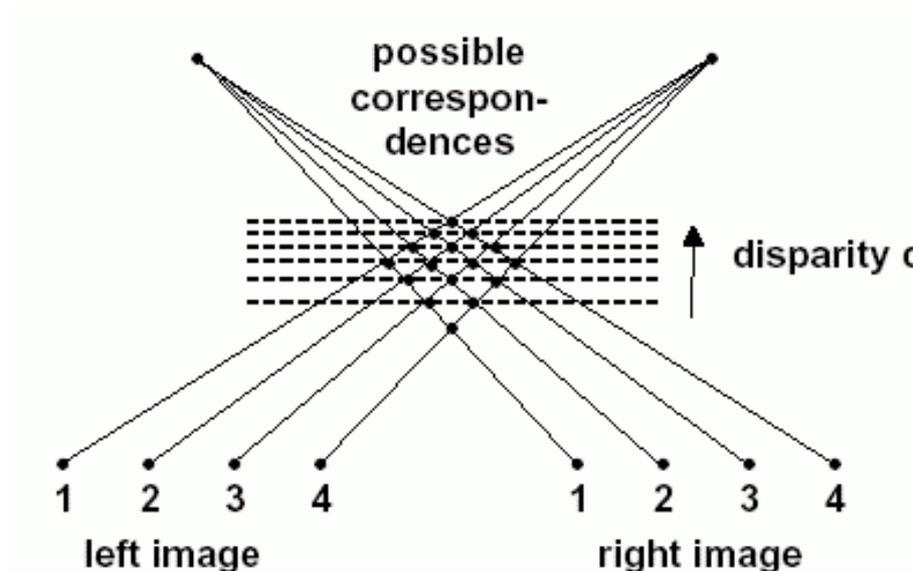
- Eindeutigkeit der Zuordnung (Ausnahme: wenn Punkte in einer Ansicht durch spezielle Lage genau zusammenfallen)
- Symmetrie der Zuordnung
- photometrische Kompatibilität (nur kleine Unterschiede in den Lichtintensitäten entsprechender Punkte)
- geometrische Ähnlichkeit von Objekten in beiden Ansichten
- *Feature compatibility*: z.B. Kanten in einem Bild müssen Kanten im anderen Bild entsprechen
- *Ordering constraint*: Anordnung von Punkten auf Kurven muss erhalten bleiben

es kann trotz dieser Regeln zu Fehlzusordnungen kommen:



⇒ ggf. muss von mehreren lokalen Optima der Zuordnung ein globales Optimum bestimmt werden.

weitere Annahme: Tiefe variiert in der Szene nur allmählich  
⇒ ein "Disparitäts-Index" kann benutzt werden, um die Qualität von Zuordnungen zu bewerten und Optimierung durchzuführen.



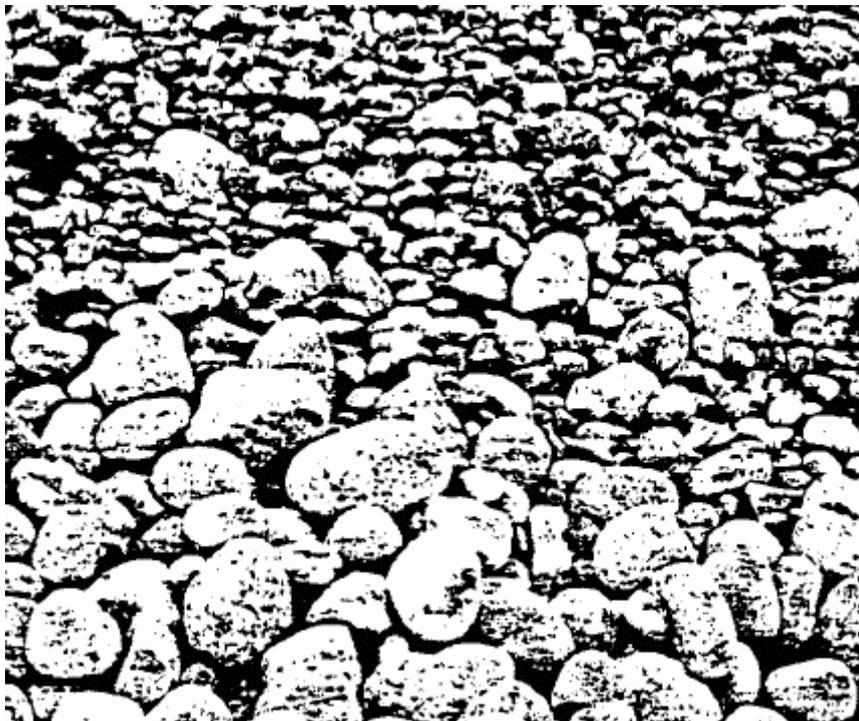
## *Auswertung monokularer Tiefenhinweise*

Auch das menschliche visuelle System gewinnt Tiefeninformation nicht allein aus dem Stereosehen mit beiden Augen.

Verschiedene Hinweise für eine heuristische Tiefenbestimmung werden genutzt:

- Größen bekannter Objekte
- Texturgradienten
- partielle Verdeckung bekannter Objekte
- Farbintensitäten, Dunst
- Stetigkeitsannahmen
- Wissen über "normale" Lage und normales Verhalten von Objekten
- etc.

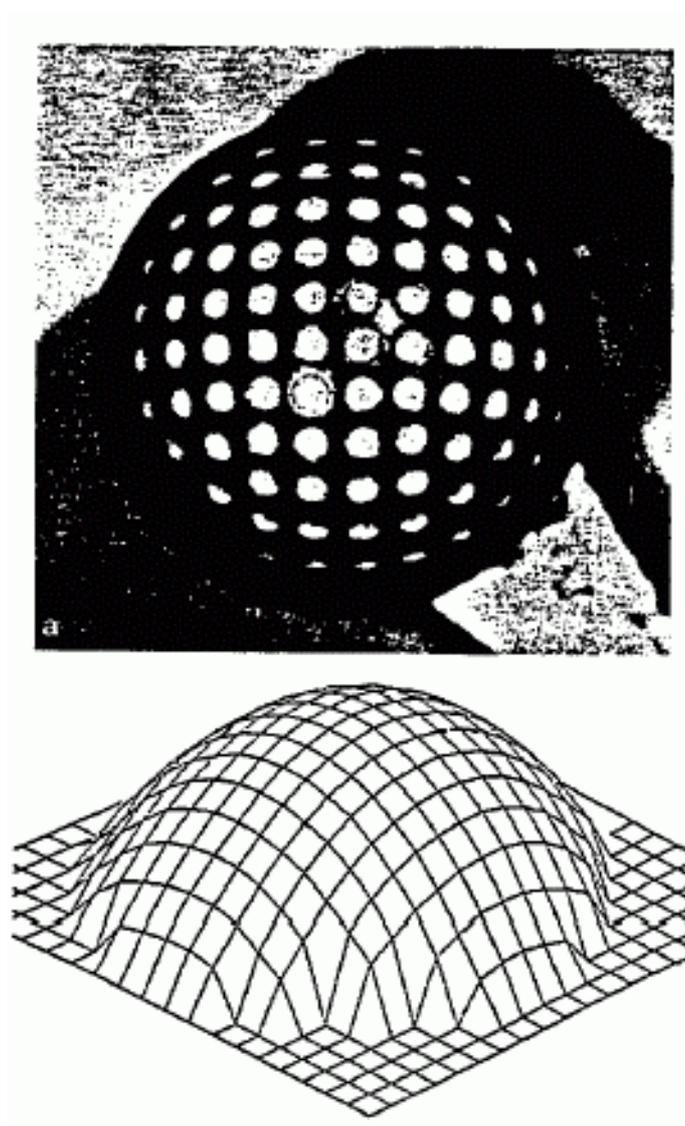
### Texturgradient



Witkin 1981: Interpretation eines Texturgradienten als 3D-Projektions-Effekt

## Shape from texture

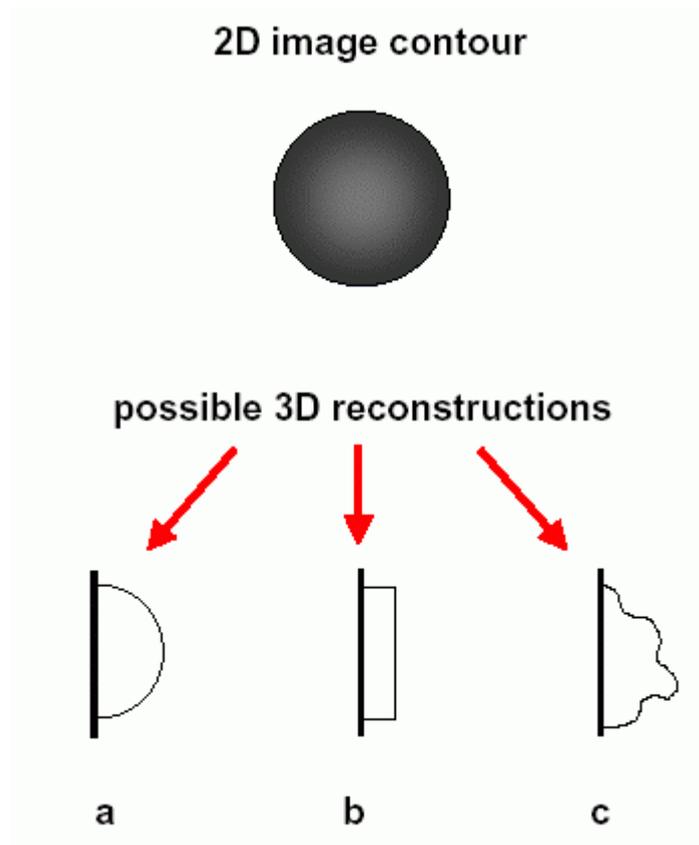
Barrow & Tenenbaum 1981: Annahme homogener Textur auf 3D-Objekten, Stetigkeit der 3D-Oberfläche  $\Rightarrow$  Rekonstruktion der 3D-Gestalt aus perspektivbedingten Texturvariationen



## Shape from contour

Annahme "nicht-spezialer" Beleuchtungssituation und  
Oberflächeneigenschaften

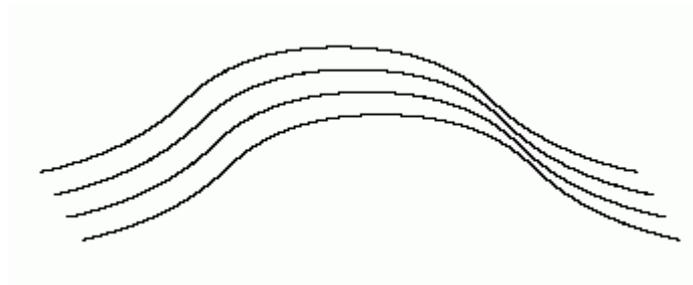
wähle Form der Oberfläche so, dass die Wahrscheinlichkeit der  
beobachteten Konturen maximal ist



⇒ einfachste Interpretation (a) wird gewählt.

## Shape from multiple lines

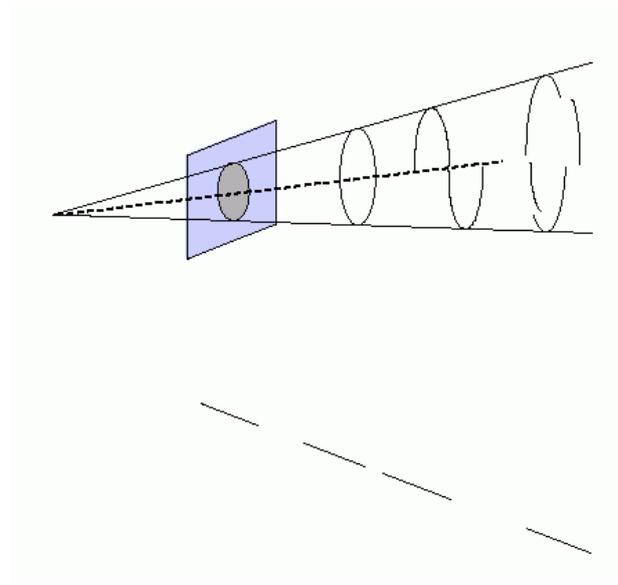
Annahmen: ähnliche Linienformen resultieren aus ähnlichen  
Oberflächenformen, parallele Linien liegen lokal auf Zylinder  
(Stevens 1981)



## Gestalt von 3D-Linien aus 2D-Projektionen

Annahmen:

- Linien, die in 2D verbunden sind, sind auch in 3D verbunden
- Kollineare Linienstücke in 2D sind auch in 3D kollinear
- Rekonstruktion der 3D-Linienform durch *Minimierung der Krümmung und Torsion*

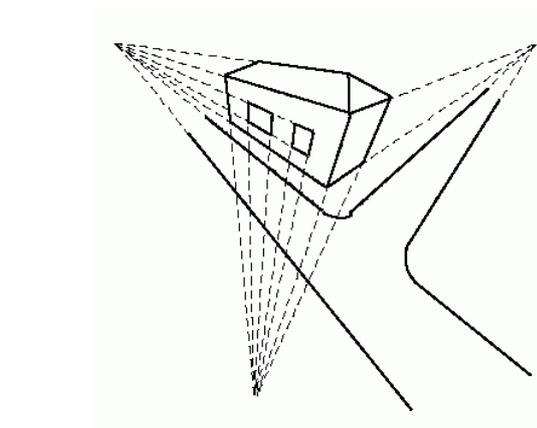


## Detektion von Fluchtpunkten

Annahme: wenn sich mehr als 2 gerade Linien in 1 Punkt schneiden, geschieht dies nicht zufällig

⇒ Annahme, dass ein Fluchtpunkt vorliegt

⇒ Linien sind in 3D parallel



Auffinden von Geradenscharen mit gemeinsamem Schnittpunkt mit modifizierter Hough-Transformation, siehe Übung, Aufgabe U24

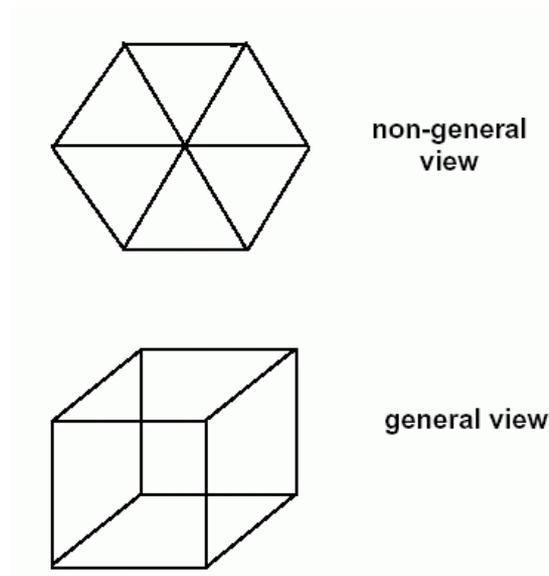
### *Metaprinzip der allgemeinen Ansicht:*

Annahme, dass

- Beobachterposition
- Beleuchtung
- Oberflächeneigenschaften
- relative Lage von Objekten

allgemein sind, d.h. ohne spezielle Koinzidenzen.

Beispiel:

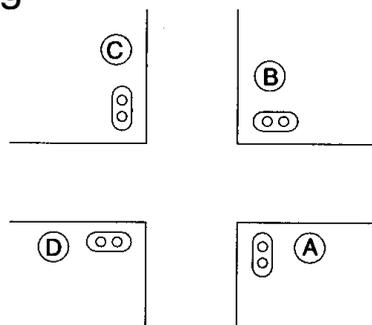


### *Kantenmarkierung zur konsistenten 3D-Interpretation von Objekten*

= Spezialfall der Verwendung von *Constraints*  
(Einschränkungsbedingungen)

einfaches Beispiel für Constraints (nach Bässmann & Kreys 1998):

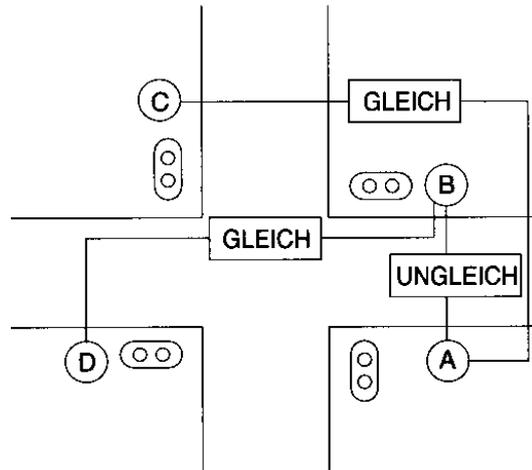
4 Ampeln an Kreuzung



zunächst 16 mögliche Lösungen für die Farben aller 4 Ampeln.  
Aber:

Constraints:

A gleich C, B gleich D, A ungleich B, A ungleich D, C ungleich B, C ungleich D.

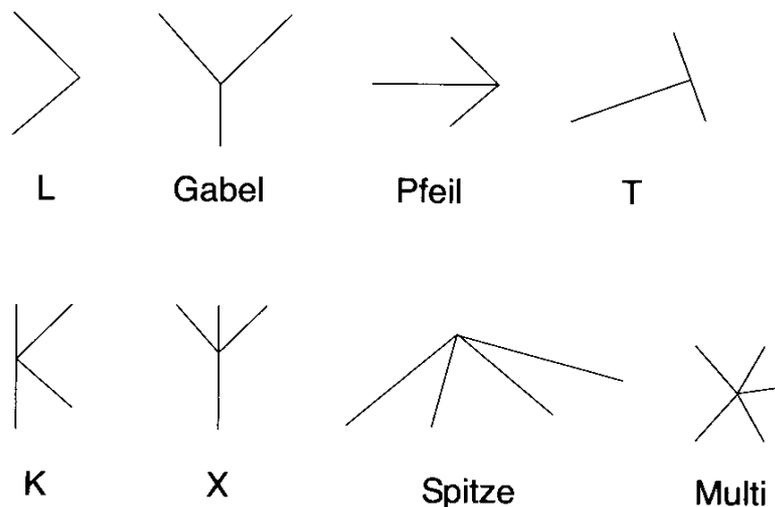


⇒ damit Einschränkung der möglichen Lösungen auf nur noch zwei verschiedene.

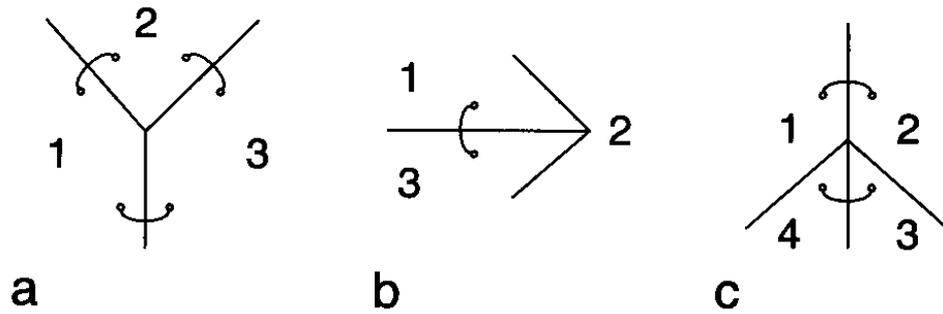
Constraint = Menge von Variablen mit einer Relation, die die Belegung der Variablen durch Werte einschränkt.

Anwendung: 3D-Interpretation von Strichzeichnungen (Roberts 1965, Guzman 1968, Waltz 1972; s. Bässmann & Kreyss 1998, S. 322 ff.).

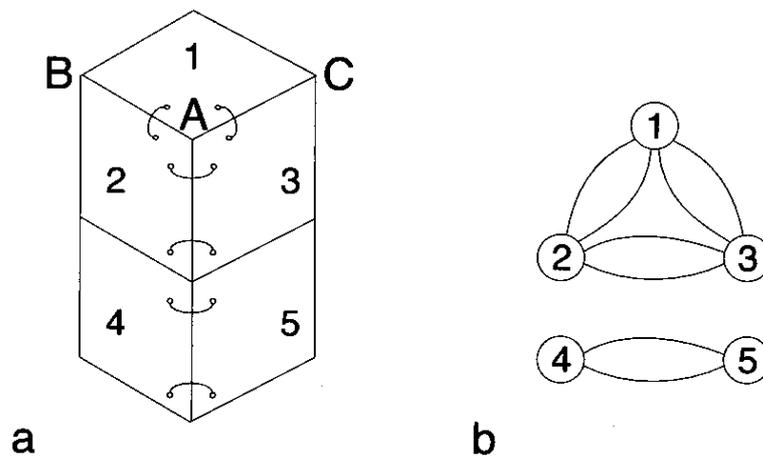
erster Ansatz: Klassifikation von "Kreuzungen" in Bildern von polyedrischen Szenen



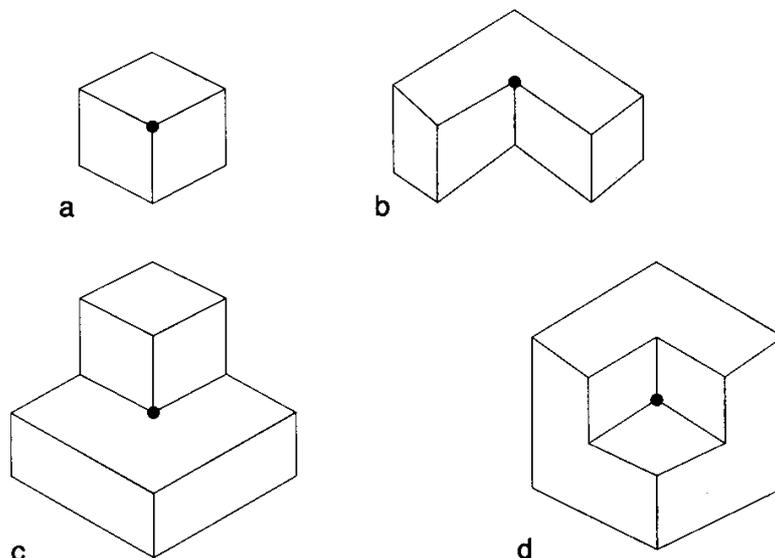
die Kreuzungstypen können Hinweise dafür liefern, welche Flächen zu demselben Objekt gehören:



dadurch Aufbau von Graphen, die Objekte liefern:

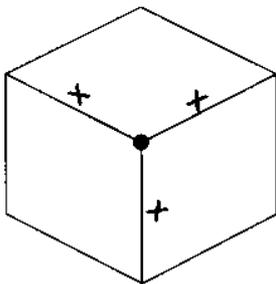


Verfeinerung des Ansatzes: Unterscheidung unterschiedlicher Typen von Scheitelpunkten  
z.B. für 3 planare Flächen:

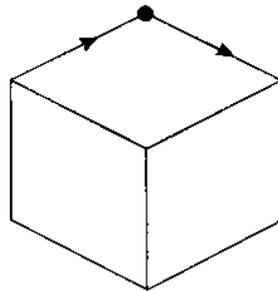


Kanten-Labeling ermöglicht bessere Klassifikation dieser Typen:

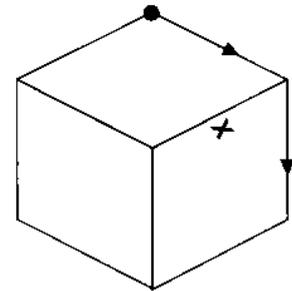
- + konvexe Kante in der Szene
- konkave Kante in der Szene
- > Sprungkante (umrahmende Kante), grenzt Objekt vom Hintergrund ab. Wenn man sich in Richtung des Pfeils bewegt, ist die zum Objekt gehörige Region rechts.



a



b



c

Beachte: die "richtigen" Labels in einer 2D-Darstellung sind zunächst nicht bekannt!

Mögliche Labellings von Ansichten von Scheitelpunkten, die durch das Zusammentreffen von max. 3 Flächen entstehen:



a



b



c



d



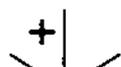
e



f



g



h



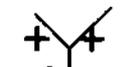
i



j



k



l



m



n



o

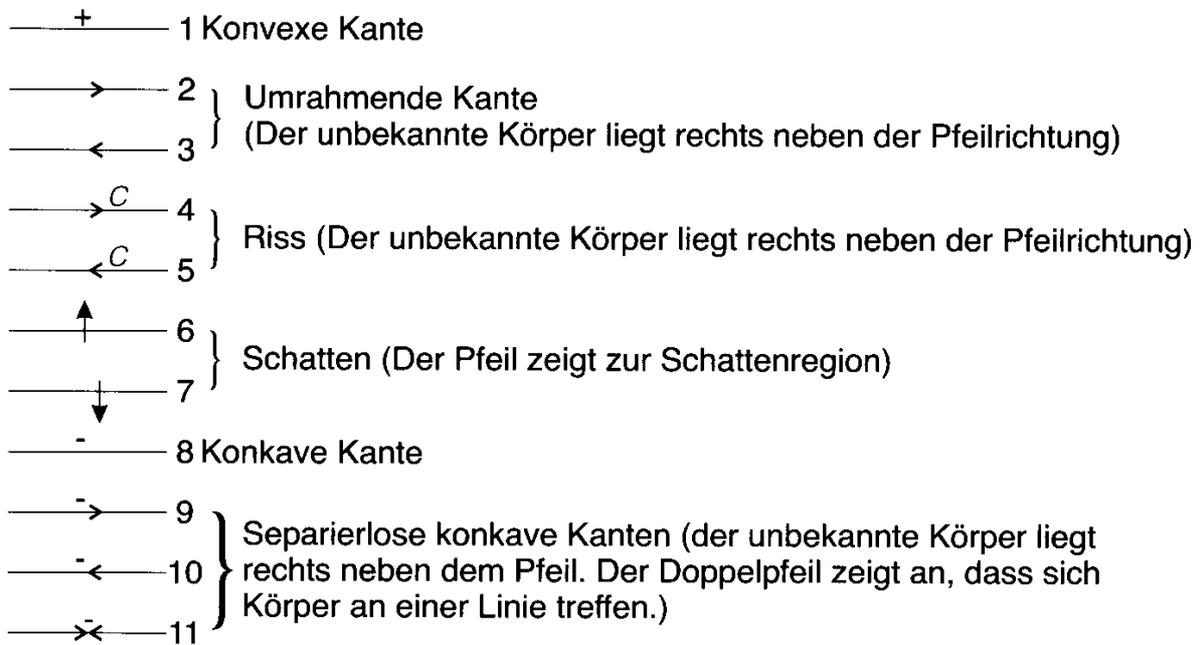


p

= 16 lokal konsistente Interpretationen

gefordert ist nun *globale Konsistenz* beim Labelling:  
lokale Konsistenz als Constraints für globales Labelling

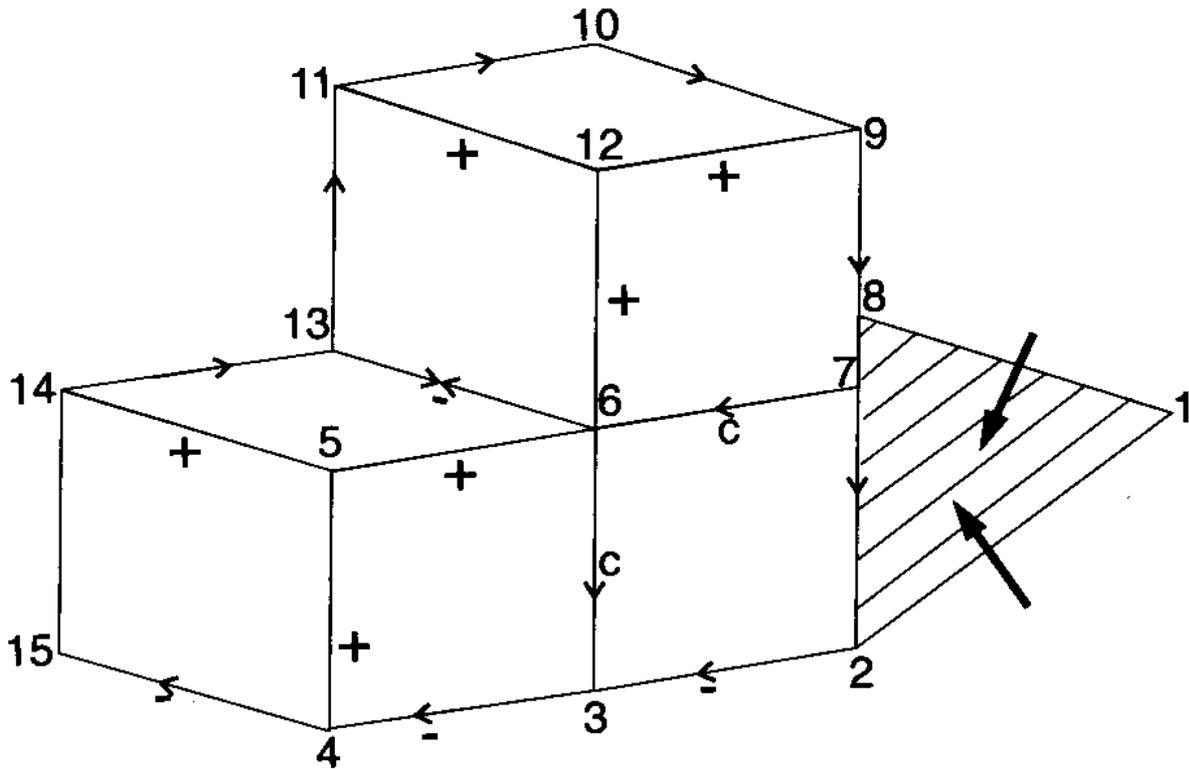
Labels für Linien nach Waltz:



Anzahlen der kombinatorisch und physikalisch möglichen Labellings:

	ungefähre Zahl der kombinatorisch möglichen Labels	ungefähre Zahl der physikalisch möglichen Labels
	3.249	92
	185.000	86
	185.000	826
	185.000	623
	$11 \times 10^6$	10
	$11 \times 10^6$	435
	$11 \times 10^6$	213
	$11 \times 10^6$	128
	$11 \times 10^6$	160
	$11 \times 10^6$	20

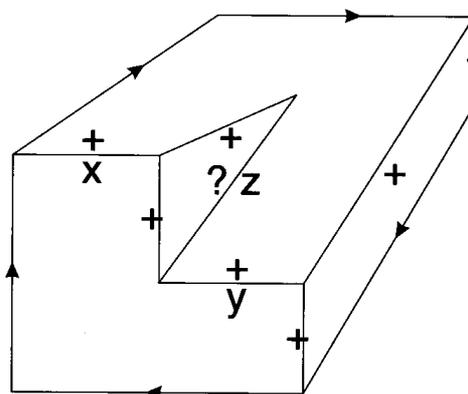
Beispiel für die Anwendung:



- Constraints (lokale Konsistenzforderungen) schränken die Zahl der global möglichen Labellings stark ein
- je mehr Wissen in den Constraints repräsentiert wird, desto stärker die Einschränkung des Lösungsraumes!

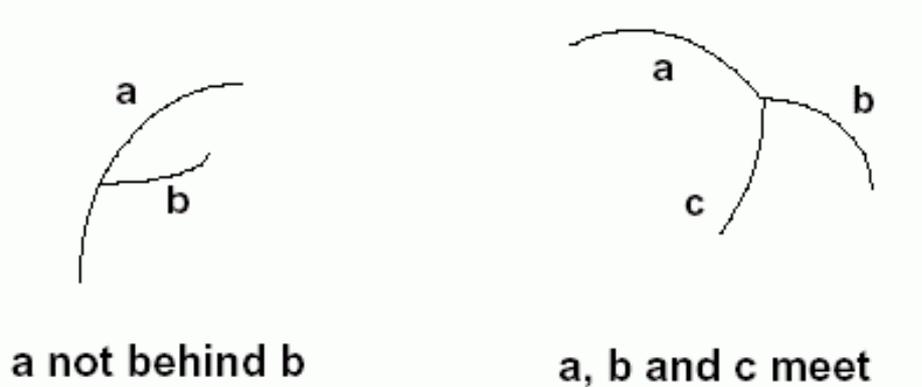
Ansatz von Waltz lässt nur gerade Linienstücke und ebene Flächen zu

Beispiel, wo Waltz nicht anwendbar:



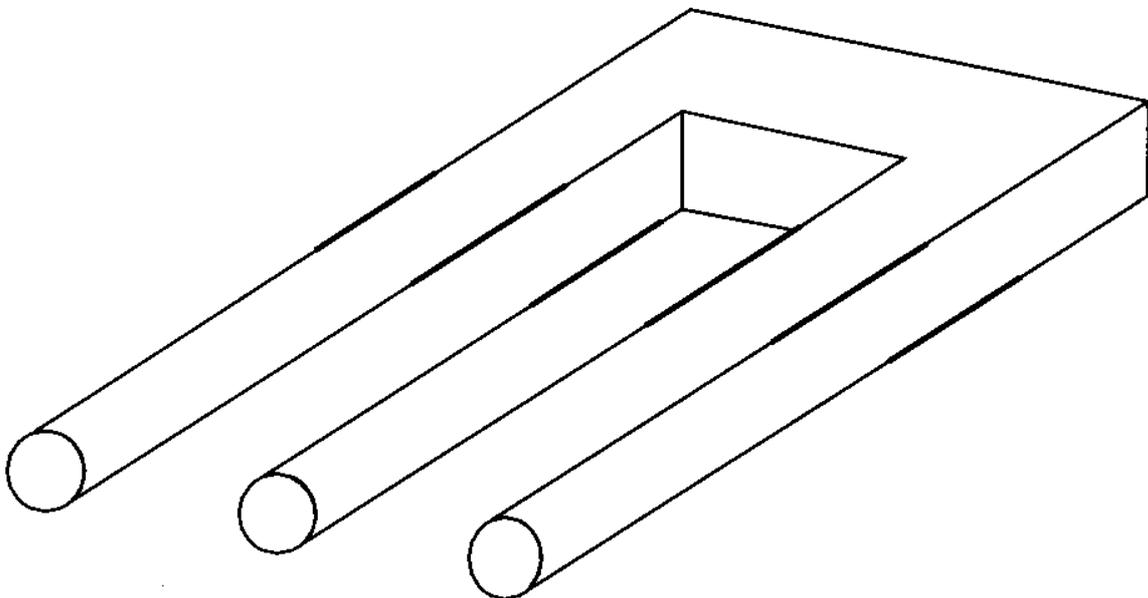
es gibt auch Regelsysteme für gekrümmte Linienstücke:  
Binford 1981

z.B.:



eine Anwendung solcher Constraint-Systeme:  
Detektion von "unmöglichen Figuren", die nicht konsistent  
gelabelt werden können

z.B.



Hinweis: Implementierungshinweise für den Waltz-Algorithmus in C und  
in Prolog bei Bässmann & Kreyss 1998, S. 339–352.