Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MIN-Fakultät
Fachbereich Informatik
Arbeitsbereich SAV/BV (KOGS)

# Image Processing 1 (IP1)
# Bildverarbeitung 1

## Lecture 19 – Camera Geometry and 3D Image Analysis

Winter Semester 2014/15

Dr. Benjamin Seppke
Prof. Siegfried Stiehl

# Camera Calibration

Determine intrinsic and/or extrinsic camera parameters for a specific camera-scene configuration. Prior calibration may be needed

- to measure unknown objects
- to navigate as a moving observer
- to perform stereo analysis
- to compensate for camera distortions

Important cases:

- **Known scene**

  Each image point corresponding to a known scene point provides an equation $\vec{v}_p = M\,\vec{v}$

- **Unknown scene**

  Several views are needed, differing by rotation and/or translation

  a. Known camera motion

  b. Unknown camera motion ("camera self-calibration")

# Calibration of One Camera from a Known Scene

- "Known scene" = scene with prominent points, whose scene coordinates are known
- Prominent points must be non-coplanar to avoid degeneracy

Projection equation $\vec{v}_p = M\,\vec{v}$ provides 2 linear equations for unknown coefficients of $M$:

$$x_p\left(m_{31}x + m_{32}y + m_{33}z + m_{34}\right) = m_{11}x + m_{12}y + m_{13}z + m_{14}$$

$$y_p\left(m_{31}x + m_{32}y + m_{33}z + m_{34}\right) = m_{21}x + m_{22}y + m_{23}z + m_{24}$$

Taking $N$ points, $N > 6$, $M$ can be estimated with a least-square method from an overdetermined system of $2N$ linear equations.

From $M = \left(KR\ \ K\vec{t}\right) = \left(A\ \ \vec{b}\right)$, one gets $K$ and $R$ by Principle Component Analysis (PCA) of $A$ and $\vec{t}$ from $\vec{t} = K^{-1}\vec{b}$

# Fundamental Matrix

The fundamental matrix $F$ generalizes the essential matrix $E$ by incorporating the intrinsic camera parameters of two (possibly different) cameras.

Essential matrix constraint for 2 views of a point:

$$\vec{n}^T E \ \vec{n}' = 0$$

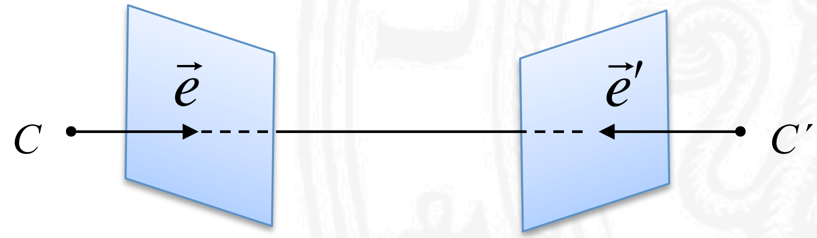
From $\vec{v}_p = K \vec{a} \vec{n}$ and $\vec{v}'_p = K' \vec{a}' \vec{n}'$ we get:

$$\vec{v}_p \left(K^{-1}\right)^T E \left(K'\right)^{-1} \vec{v}'_p = \vec{v}_p F \vec{v}'_p = 0$$

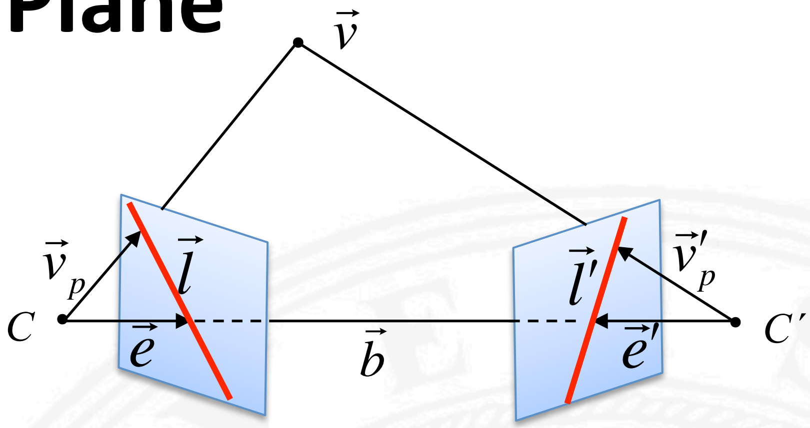
$$K = \begin{pmatrix} fa & fb & x_{p_0} \\ 0 & fc & y_{p_0} \\ 0 & 0 & 1 \end{pmatrix}$$

Note that $E$ and hence $F$ have rank 2.

For each epipole of a 2-camera configuration we have $\vec{e}^T F = 0$ and $F \vec{e}' = 0$



# Epipolar Plane

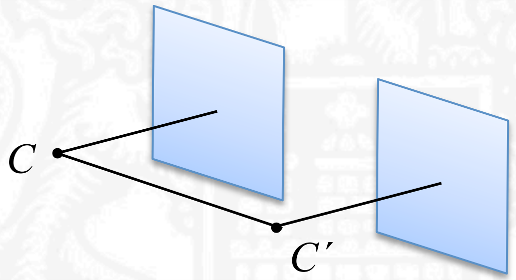



The epipolar plane is spanned by the projection rays of a point $\vec{v}$ and the baseline $\vec{b} = CC'$ of a stereo camera configuration.

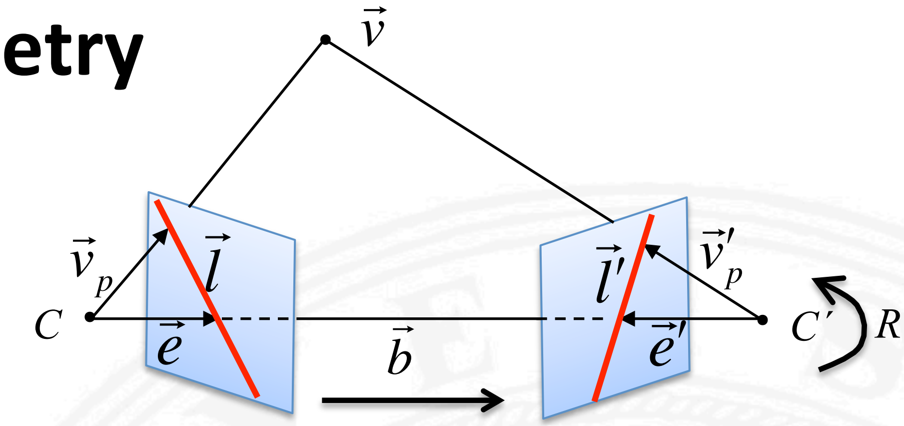
The epipoles $\vec{e}$ and $\vec{e}'$ are the intersection points of the baseline with the image planes. The epipolar lines $\vec{l}$ and $\vec{l}'$ mark the intersections of the epipolar plane in the left and right image, respectively.

**Search for corresponding points in stereo images may be restricted to the epipolar lines.**

In a canonical stereo configuration (optical axes parallel and perpendicular to baseline) all epipolar lines are parallel:



# Algebra of Epipolar Geometry




Observation $\vec{v}'_p$ can be modelled as a second observation after translation $\vec{b}$ and rotation $R$ of the optical system.

Coplanarity of $\vec{v}_p$, $\vec{b}$ and $\vec{v}'_p$ (rotated back into coo-system at $C$) can be expressed as:

$$\vec{v}_p\left(\vec{b}\times R\vec{v}'_p\right)=0=\vec{v}_p\left(\vec{b}\right)R\vec{v}'_p=\vec{v}_p E\vec{v}'_p$$

**essential matrix** (for $E$)

A vector product $\vec{c}\times\vec{d}$ can be written in matrix form:

$$\vec{c}\times\vec{d}=\begin{pmatrix} c_y d_z - c_z d_y \\ c_z d_x - c_x d_z \\ c_x d_y - c_y d_x \end{pmatrix}=\begin{pmatrix} 0 & -c_z & c_y \\ c_z & 0 & -c_x \\ -c_y & c_x & 0 \end{pmatrix}\begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix}$$

# Correspondence Problem Revisited

For multiple-view 3D analysis, it is essential to find corresponding images of a scene point - the correspondence problem.

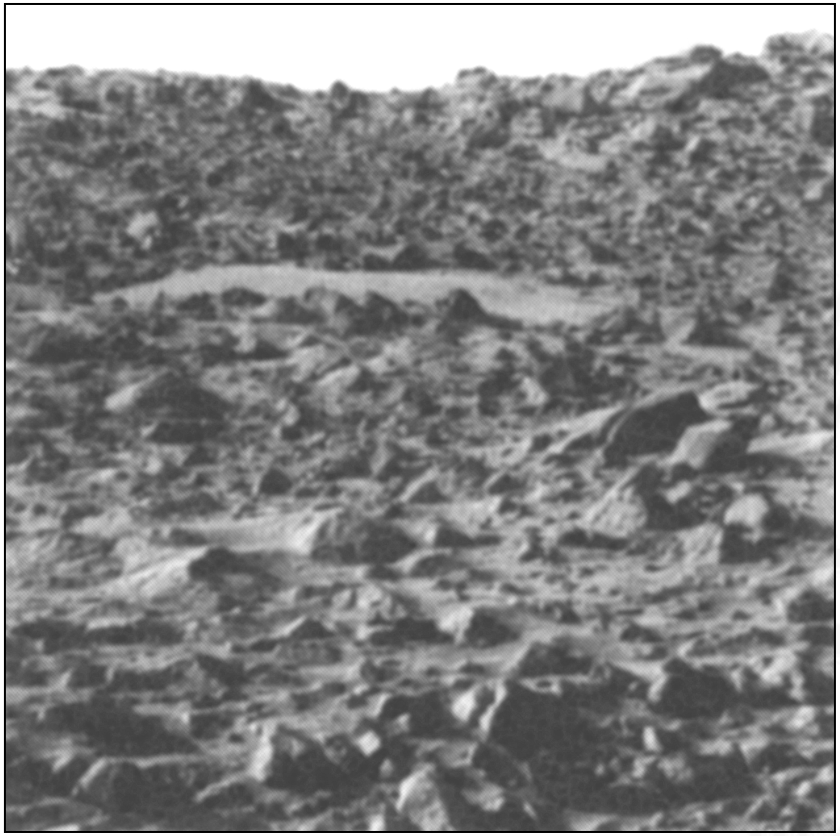
Difficulties:

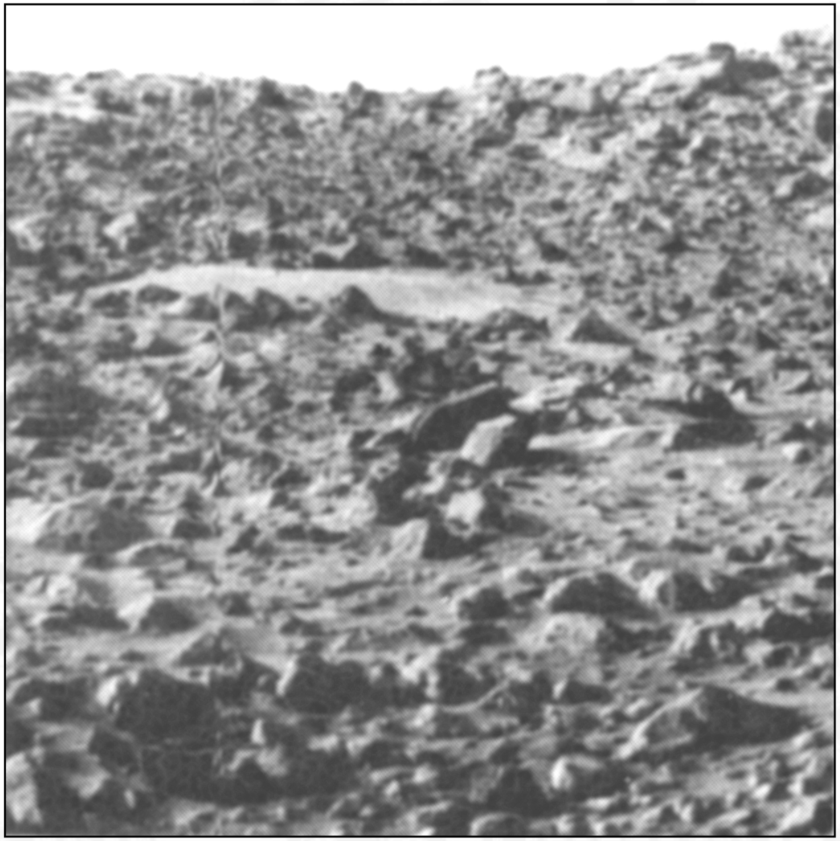
- scene may not offer enough structure to uniquely locate points
- scene may offer too much structure to uniquely locate points
- geometric features may differ strongly between views
- there may be no corresponding point because of occlusion
- photometric features differ strongly between views

Note that difficulties apply to multiple-camera 3D analysis (e.g. binocular stereo) as well as single-camera motion analysis.

# Correspondence Between Two Mars Images

Two images taken from two cameras of the Viking Lander I (1978).
Disparities change rapidly, moving from the horizon to nearby structures.
(From B.K.P. Horn, Robot Vision, 1986)





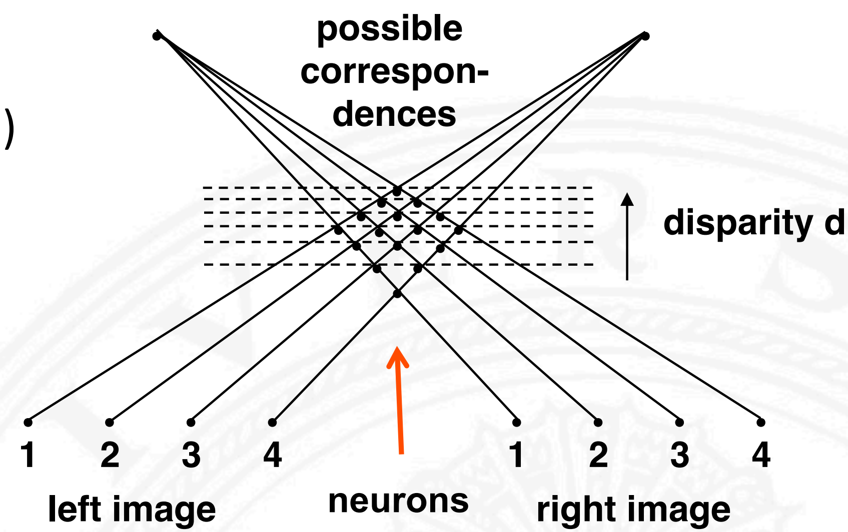
# Constraining Search for Correspondence

The ambiguity of correspondence search may be reduced by several (partly heuristic) constraints.

- Epipolar constraint
  reduces search space from 2D to 1D
- Uniqueness constraint
  a pixel in one image can correspond to only one pixel in another image
- Photometric similarity constraint
  intensities of a point in different images may differ only a little
- Geometric similarity constraint
  geometric features of a point in different images may differ only a little
- Disparity smoothness constraint
  disparity varies only slowly almost everywhere in the image
- Physical origin constraint
  points may correspond only if they mark the same physical location
- Disparity limit constraint
  in humans disparity must be smaller than a limit to fuse images
- Ordering constraint
  corresponding points lie in the same order on the epipolar line
- Mutual correspondence constraint
  correspondence search must succeed irrespective of order of images

# Neural Stereo Computation

Neural-network inspired approach to stereo computation devised by Marr and Poggio (1981)




**Exploitation of 3 constraints:**

- depth varies smoothly
- each point in the left image corresponds to only one point in the right image
- similar sensor signal

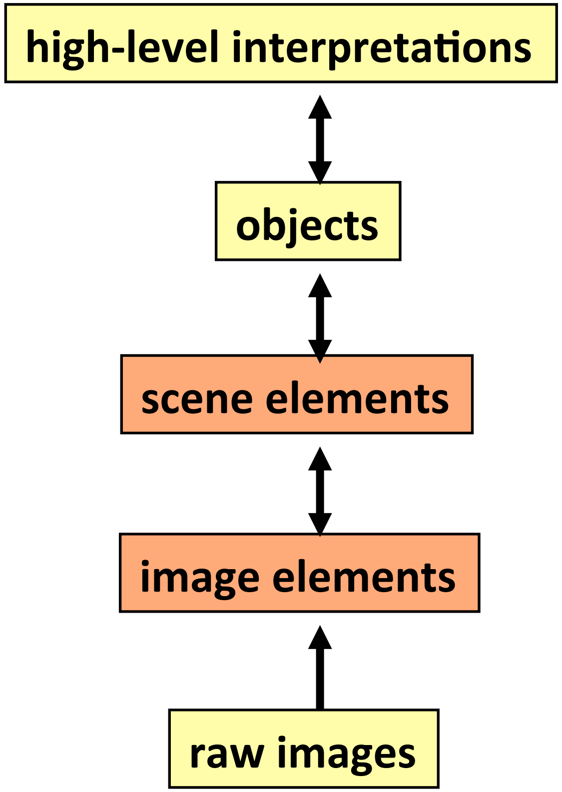
**Relaxation procedure:**

Modify correspondence values $c_k$ interatively until values converge.

$$c_k^{n+1} = w_1 \sum_{i \in S_1} c_i^n + w_2 \sum_{i \in S_2} c_i^n + w_3 \sum_{i \in S_3} c_i^n$$

$S_1$ = { neighbours of $k$ with similar disparity $d$ }
$S_2$ = { neighbours of $k$ on same projection ray }
$S_3$ = { neighbours of $k$ with similar sensor values }

# General Principles of 3D Image Analysis




**Extraction of 3D information from an image (sequence) is important for**

- vision in general (= scene reconstruction)
- many tasks (e.g. robot grasping and navigation, traffic analysis)
- not all tasks (e.g. image retrieval, quality control, monitoring)

**Recovery of 3D information is possible**

- by multiple cameras (e.g. binocular stereo)
- by a monocular image sequence with motion + weak assumptions
- by a single image + strong assumptions or prior knowledge about the scene

# Single Image 3D Analysis

Humans exploit various cues for a tentative (heuristic) depth analysis:

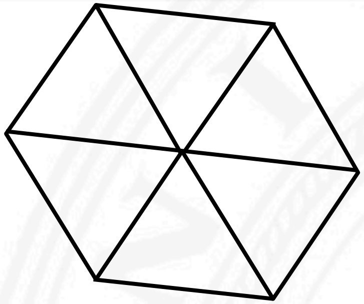
- size of known objects
- texture gradient
- occlusion
- colour intensities
- angle of observation
- continuity assumption
- **generality assumption**

# Generality Assumption

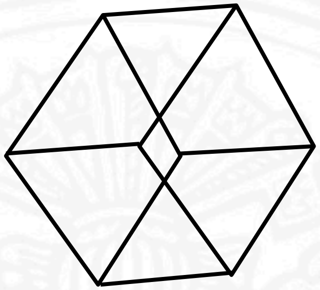
**Assume that**
- **viewpoint**
- **illumination**
- **physical surface properties**

**are all general, i.e. do not produce coincidental structures in the image.**

**Example**:
Do not interpret this figure as a 3D wireframe cube, because this view is not general.



**General view:**



The generality assumption is the basis for several specialized interpretation methods, e.g.

- shape from texture
- shape from shading

...

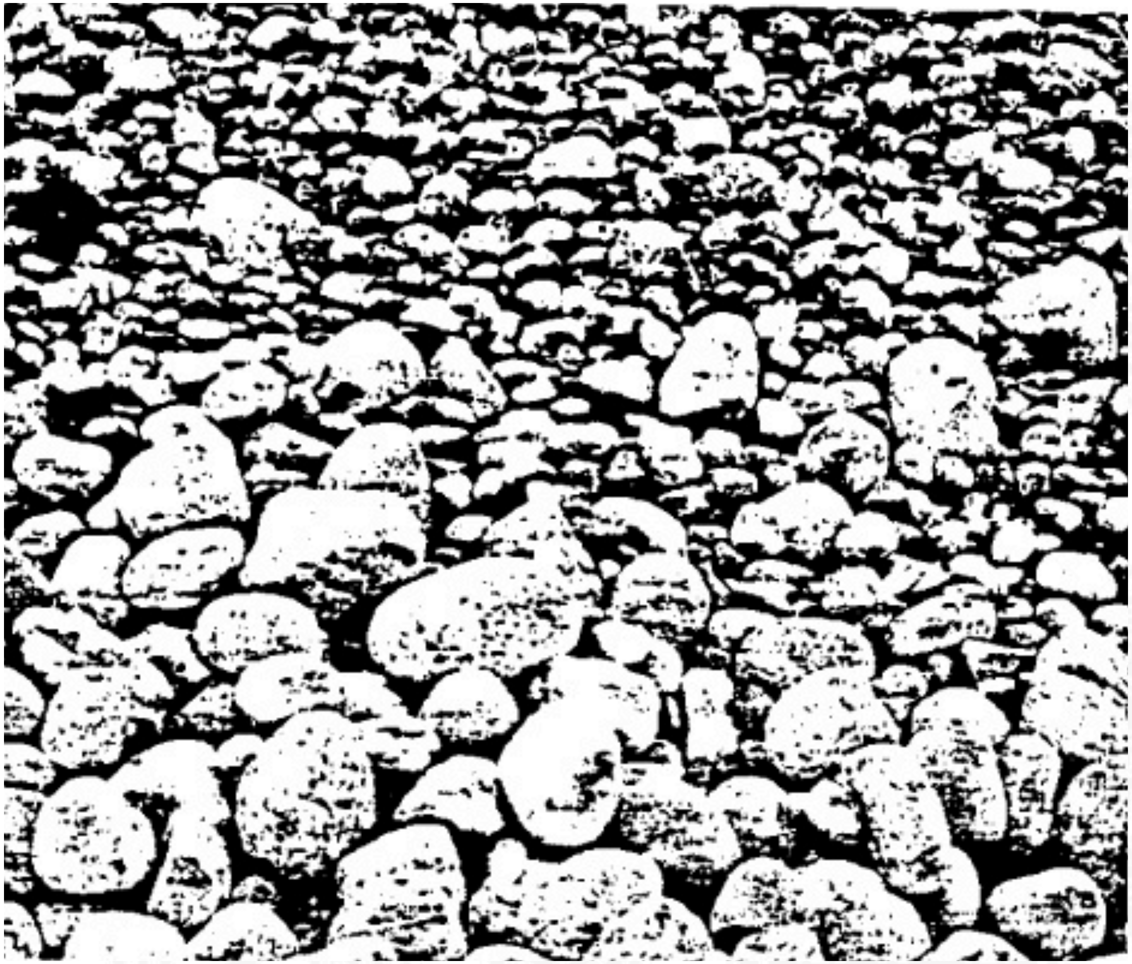
- "shape from X"

# Texture Gradient

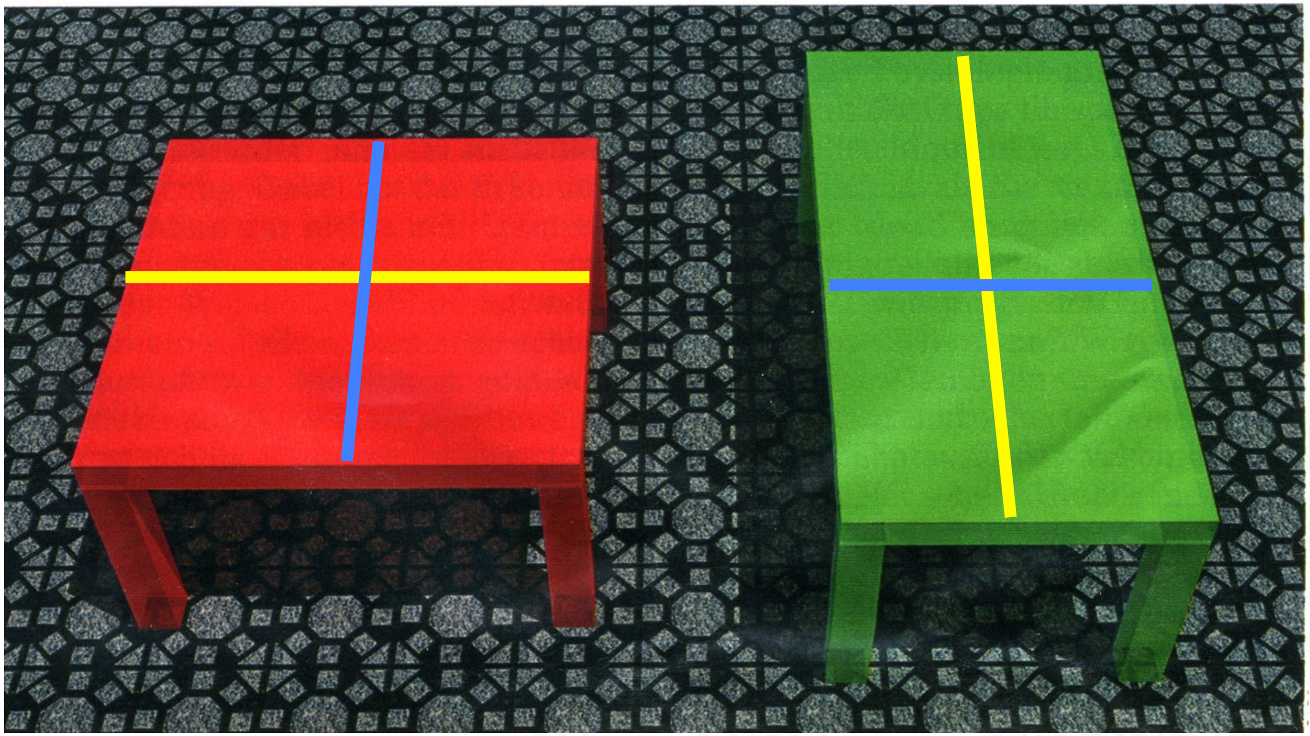
Assume that texture does not mimick projective effects

Interpret texture gradient as a 3D projection effect

(Witkin 81)



# Optical Illusion from Depth Cues



The left table seems to be square, the right table lengthy.

But their image dimensions are identical, although rotated by $90^0$.

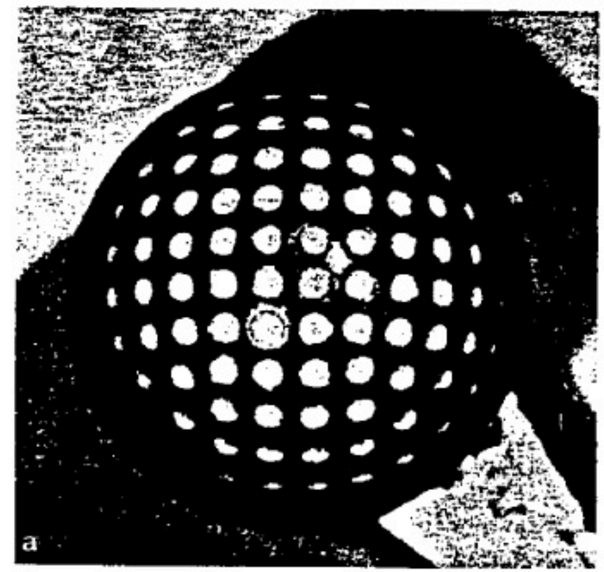
# Shape from Texture

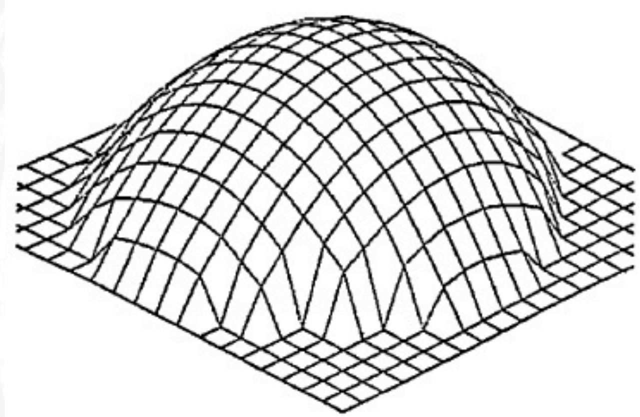
Assume

- homogeneous texture on 3D surface and
- 3D surface continuity

Reconstruct 3D shape from perspective texture variations

(Barrow and Tenenbaum 81)





# Depth Cues from Colour Saturation

Humans interpret regions with less saturated colours as farther away.

**hills in haze nearby Graz**

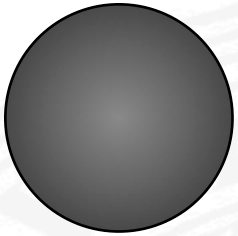
# Surface Shape from Contour

Assume "non-special" illumination and surface properties

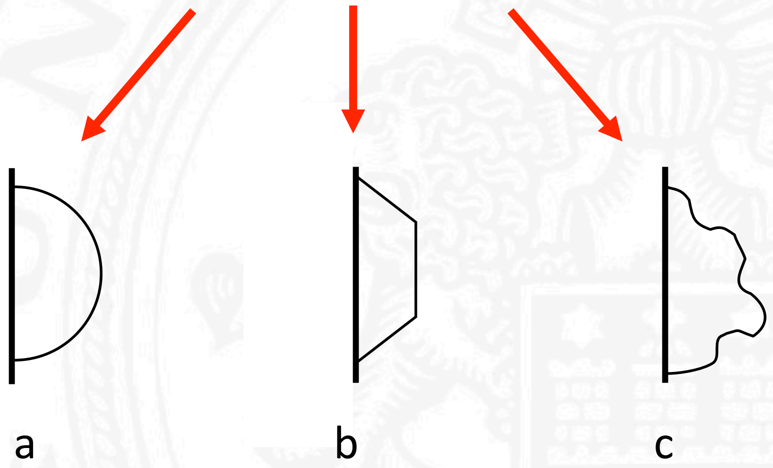
⇩

3D surface shape maximizes probability of observed contours and minimizes probability of additional contours

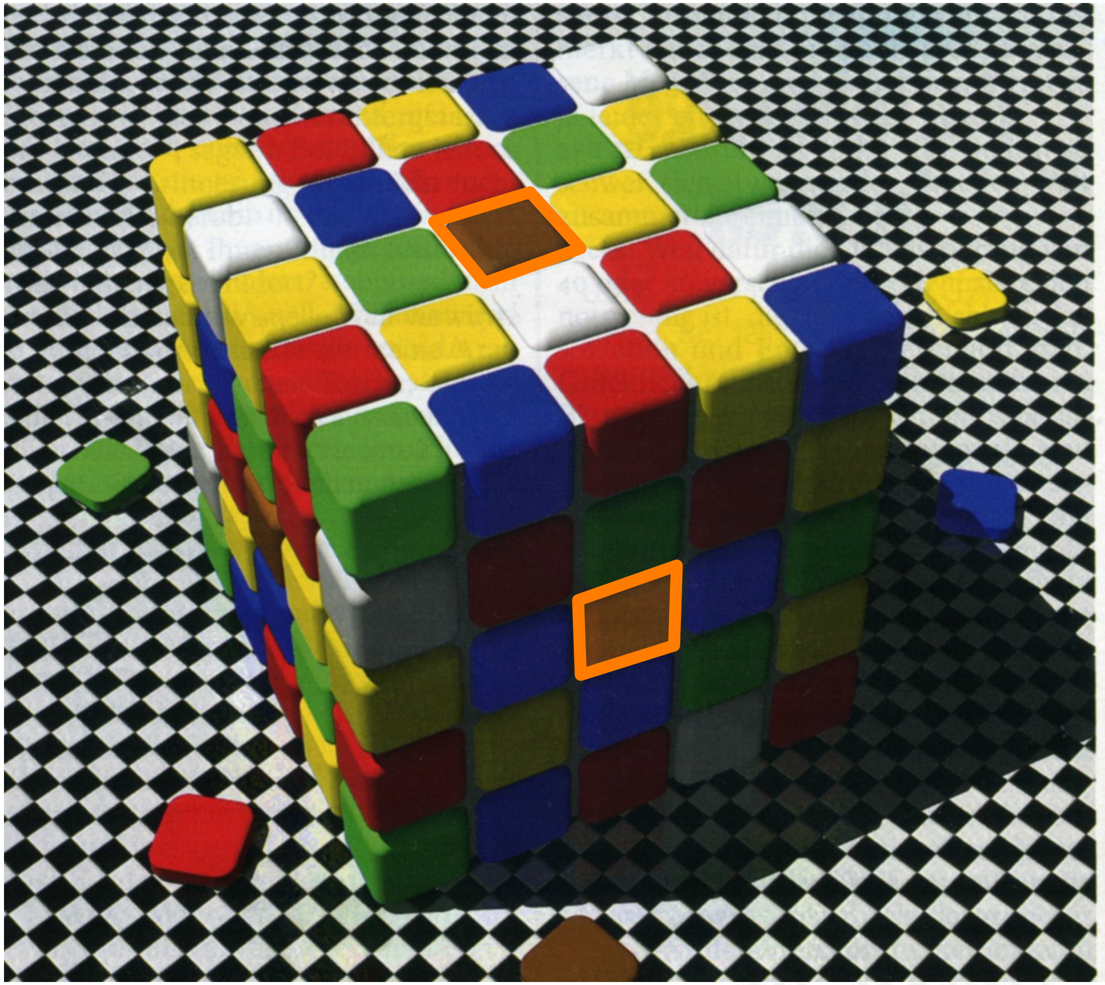
2D image contour



possible 3D reconstructions

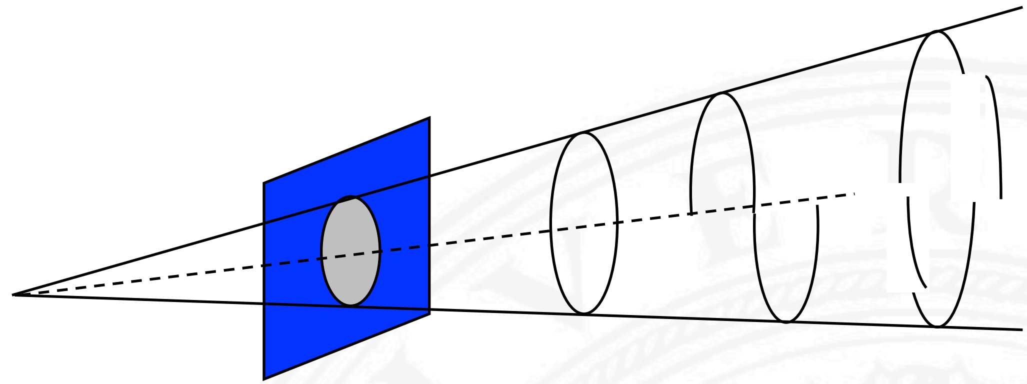


# Colour from Shading Cues



Central squares on top and in front have identical colour,
but shading cues suggest that the front square is brighter.

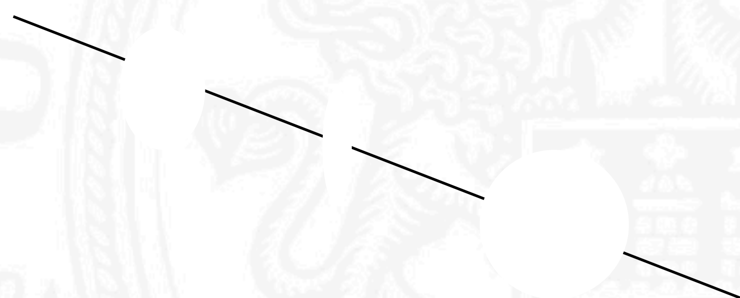
# 3D Line Shape from 2D Projections



Assume that lines connected in 2D are also connected in 3D

Reconstruct 3D line shape by minimizing spatial curvature and torsion

2D collinear lines are also 3D collinear

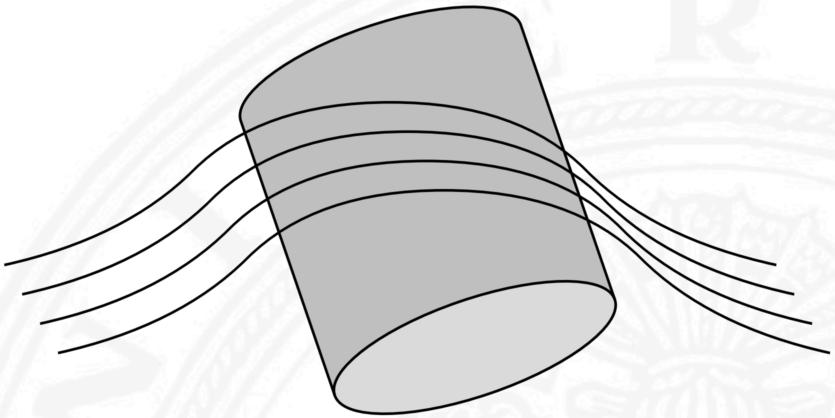


# 3D Shape from Multiple Lines

Assume that similar line shapes result from similar surface shapes

Parallel lines lie locally on a cylinder

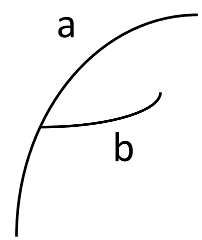
(Stevens 81)



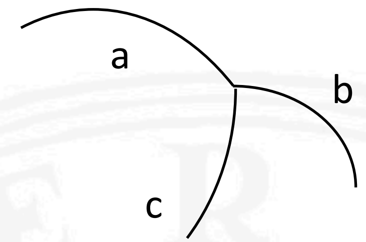
# 3D Junction Interpretation

Rules for junctions of curved lines

(Binford 81)



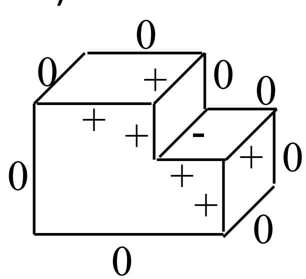

a not behind b



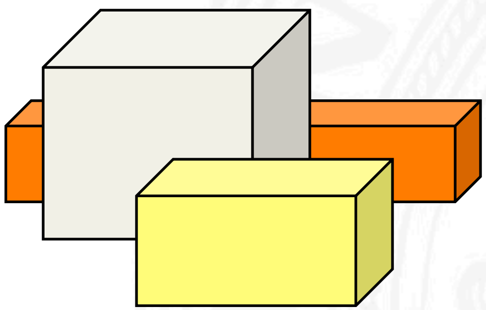

a, b and c meet

Rules for blocks-world junctions

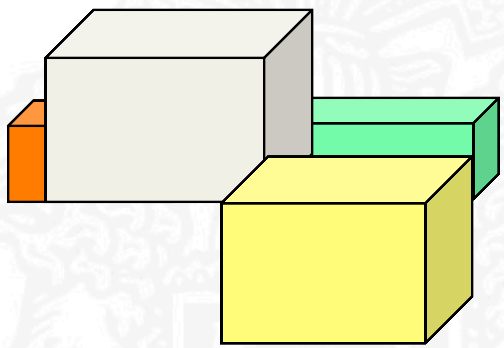
(Waltz 86)




edge labels



"general" ensemble



"special" ensemble

# 3D Line Orientation from Vanishing Points

**From the laws of perspective projection:**

The projections of 3D parallel straight lines intersect in a single point, the vanishing point.

Assume that more than 2 straight lines do not intersect in a single point by coincidence

If more than 2 straight lines intersect, assume that they are parallel in 3D