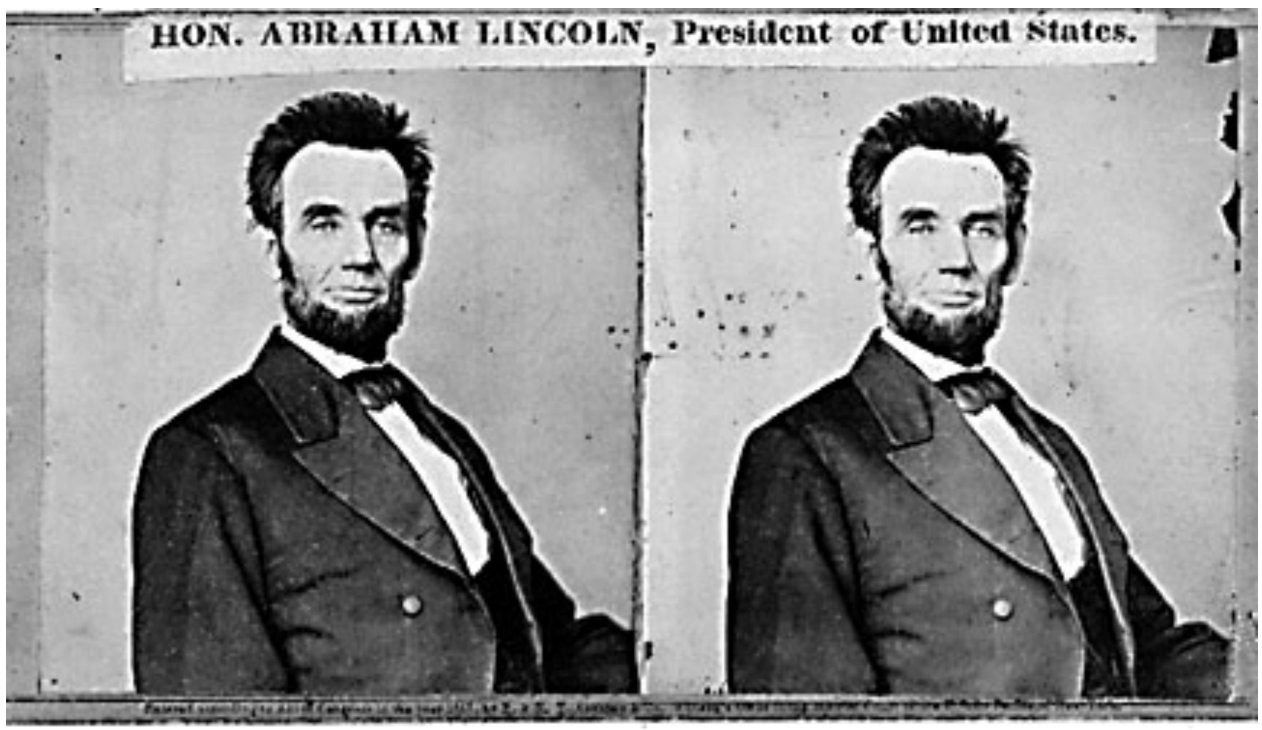
CS4670 / 5670: Computer Vision

Kavita Bala

Lec 27: Stereo



Announcements

- PA 3 due on Monday 1pm
- Regrades etc.
- PA 4: Photometric stereo

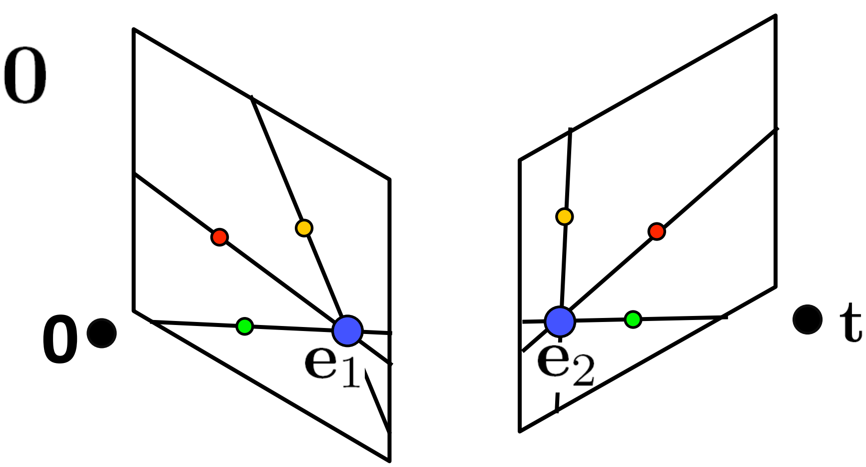
Fundamental matrix result

$$\mathbf{q}^T \mathbf{F} \mathbf{p} = 0$$

(Longuet-Higgins, 1981)

Properties of the Fundamental Matrix

- $\mathbf{F}\mathbf{p}$ is the epipolar line associated with $\mathbf{p}$
- $\mathbf{F}^T\mathbf{q}$ is the epipolar line associated with $\mathbf{q}$
- $\mathbf{F}\mathbf{e}_1 = \mathbf{0}$ and $\mathbf{F}^T\mathbf{e}_2 = \mathbf{0}$
- $\mathbf{F}$ is rank 2



Estimating **F**

- If we don't know $\mathbf{K}_1$, $\mathbf{K}_2$, **R**, or **t**, can we estimate **F** for two images?
- Yes, given enough correspondences

Estimating F – 8-point algorithm

- The fundamental matrix F is defined by

$$\mathbf{x}'^{\mathrm{T}}\mathbf{F}\mathbf{x} = 0$$

for any pair of matches x and x' in two images.

- Let x=$(u,v,1)^{\mathrm{T}}$ and x'=$(u',v',1)^{\mathrm{T}}$,

$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

each match gives a linear equation

$$uu'f_{11} + vu'f_{12} + u'f_{13} + uv'f_{21} + vv'f_{22} + v'f_{23} + uf_{31} + vf_{32} + f_{33} = 0$$

8-point algorithm

$$\begin{bmatrix} u_1u_1' & v_1u_1' & u_1' & u_1v_1' & v_1v_1' & v_1' & u_1 & v_1 & 1 \\ u_2u_2' & v_2u_2' & u_2' & u_2v_2' & v_2v_2' & v_2' & u_2 & v_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_n' & v_nu_n' & u_n' & u_nv_n' & v_nv_n' & v_n' & u_n & v_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

- In reality, instead of solving $\mathbf{Af} = 0$, we seek $\mathbf{f}$ to minimize $\|\mathbf{Af}\|$, least eigenvector of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

8-point algorithm – Problem?

- **F** should have rank 2
- To enforce that **F** is of rank 2, F is replaced by F' that minimizes $\|\mathbf{F}-\mathbf{F}'\|$ subject to the rank constraint.

- This is achieved by SVD. Let $\mathbf{F}=\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}$, where

$$\Sigma=\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & \sigma_3\end{bmatrix}, \text{ let } \Sigma'=\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & 0\end{bmatrix}$$

then $\mathbf{F}'=\mathbf{U}\Sigma'\mathbf{V}^{\mathrm{T}}$ is the solution.

8-point algorithm

- Pros: it is linear, easy to implement and fast
- Cons: susceptible to noise
- Degenerate: if points are on same plane

- Normalized 8-point algorithm: Hartley
  - Position origin at centroid of image points
  - Rescale coordinates so that center to farthest point is sqrt (2)

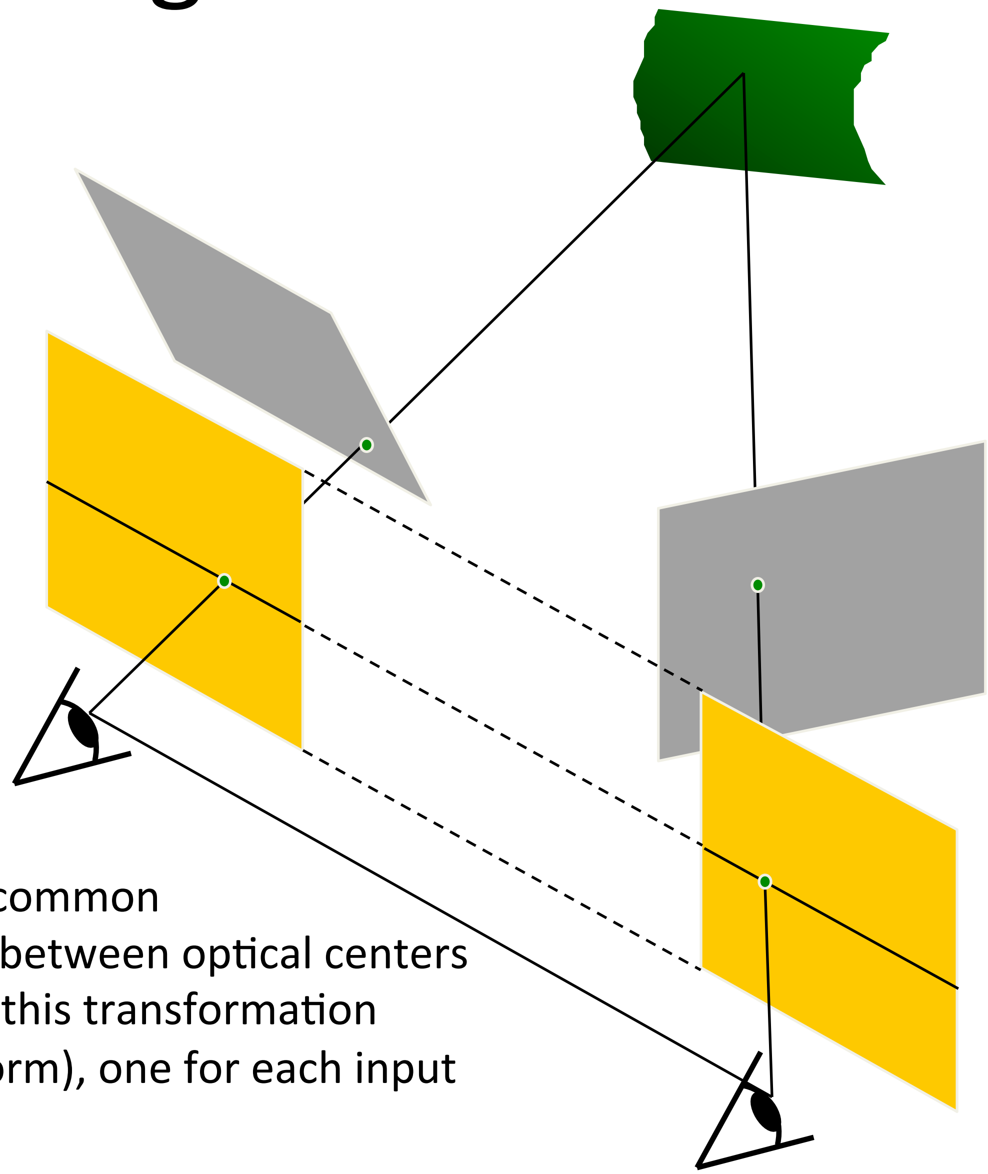
What about more than two views?

- The geometry of three views is described by a 3 x 3 x 3 tensor called the *trifocal tensor*
- The geometry of four views is described by a 3 x 3 x 3 x 3 tensor called the *quadrifocal tensor*
- After this it starts to get complicated…
  - Structure from motion

Stereo reconstruction pipeline

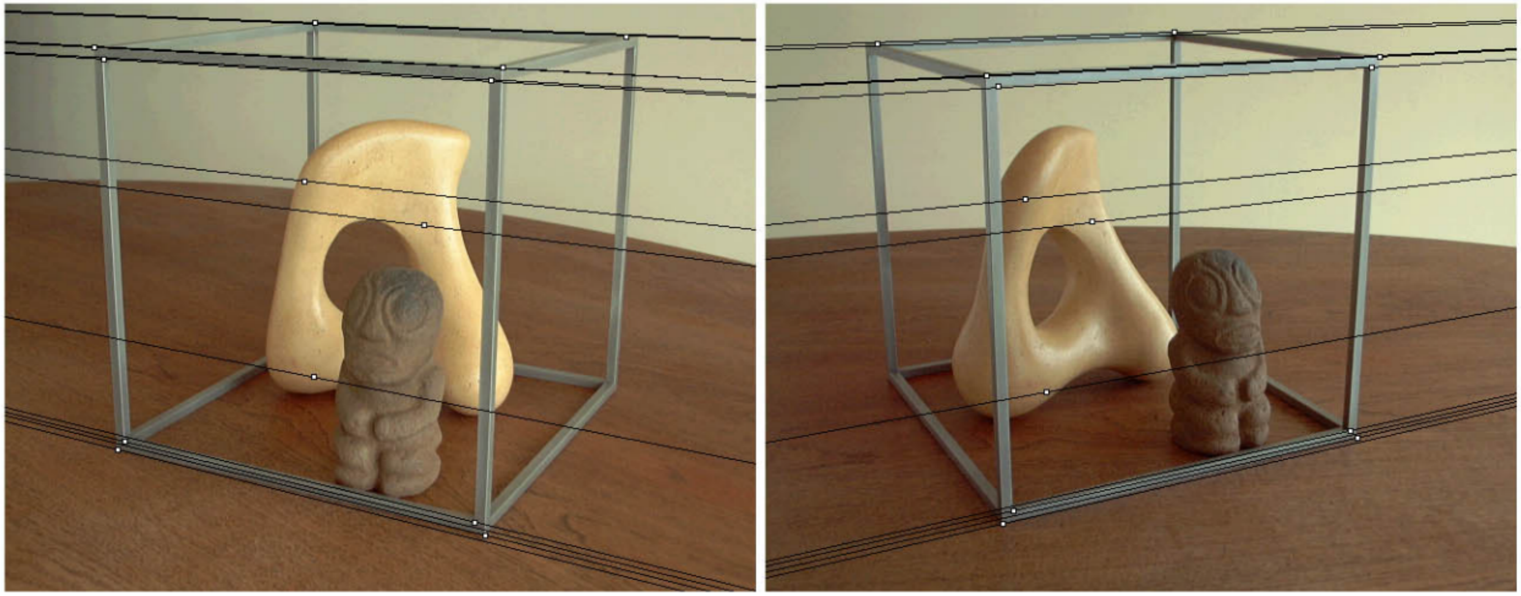
- Steps
  - Compute Fundamental Matrix
  - Calibrate cameras: K1, K2
  - Rectify images
  - Compute correspondence (and hence disparity)
  - Estimate depth

Stereo image rectification

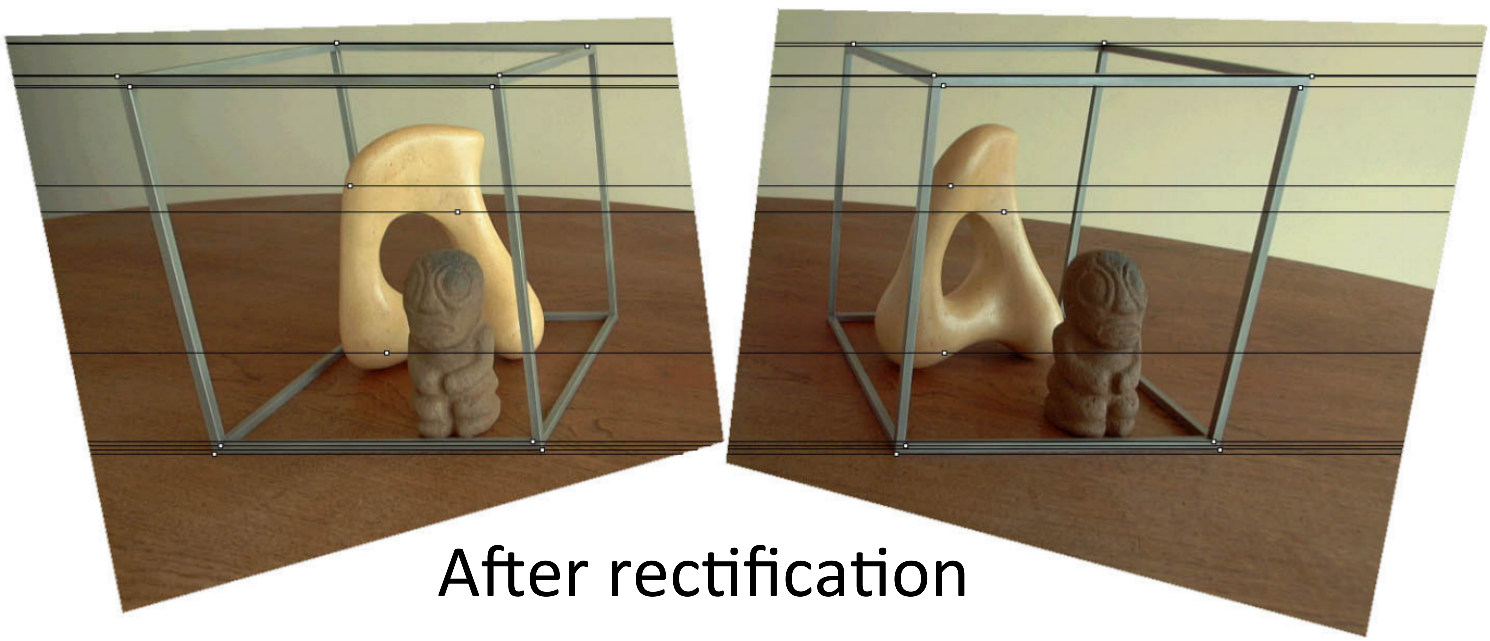


- reproject image planes onto a common
- plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- two homographies (3x3 transform), one for each input image reprojection

➢ C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.



Original stereo pair



After rectification

Correspondence algorithms

Algorithms may be classified into two types:

1. Dense: compute a correspondence at every pixel
2. Sparse: compute correspondences only for features

Example image pair – parallel cameras

First image

Second image

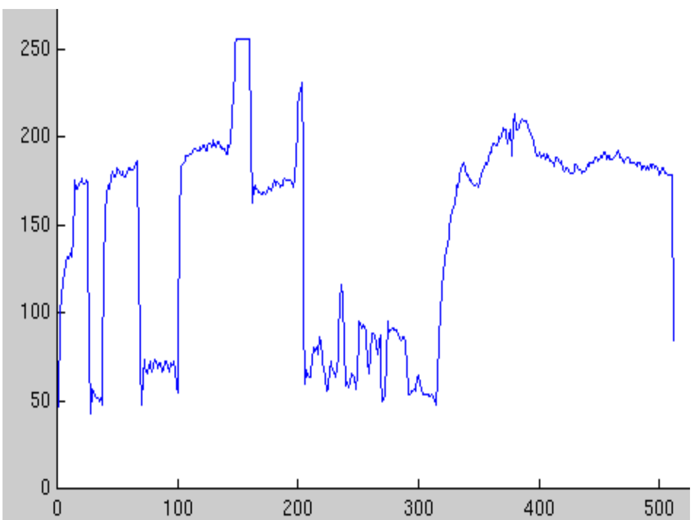
Dense correspondence algorithm

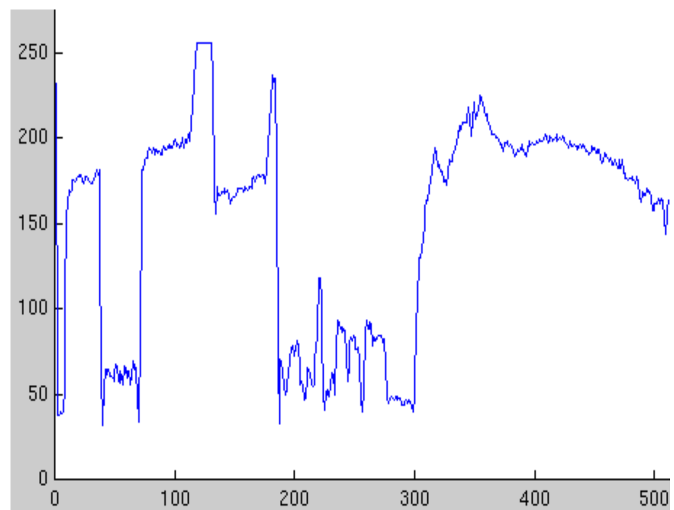
Search problem (geometric constraint): for each point in left image, corresponding point in right image lies on the epipolar line (1D ambiguity)

Disambiguating assumption (photometric constraint): the intensity neighborhood of corresponding points are similar across images

Measure similarity of neighborhood intensity by cross-correlation

Intensity profiles





- Clear correspondence, but also noise and ambiguity

Normalized Cross Correlation

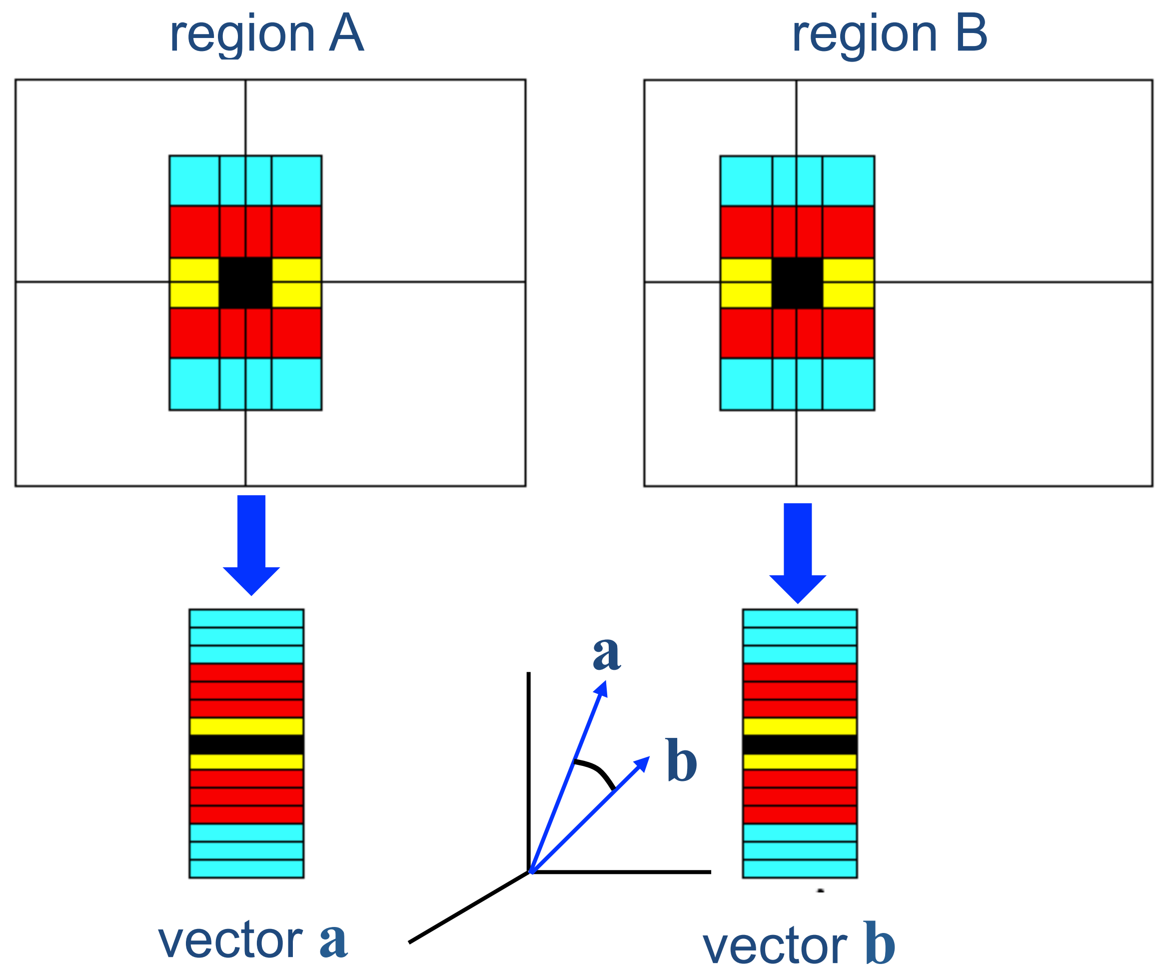
$$\text{NCC} = \frac{\sum_i \sum_j A(i,j)B(i,j)}{\sqrt{\sum_i \sum_j A(i,j)^2}\sqrt{\sum_i \sum_j B(i,j)^2}}$$

regions as vectors

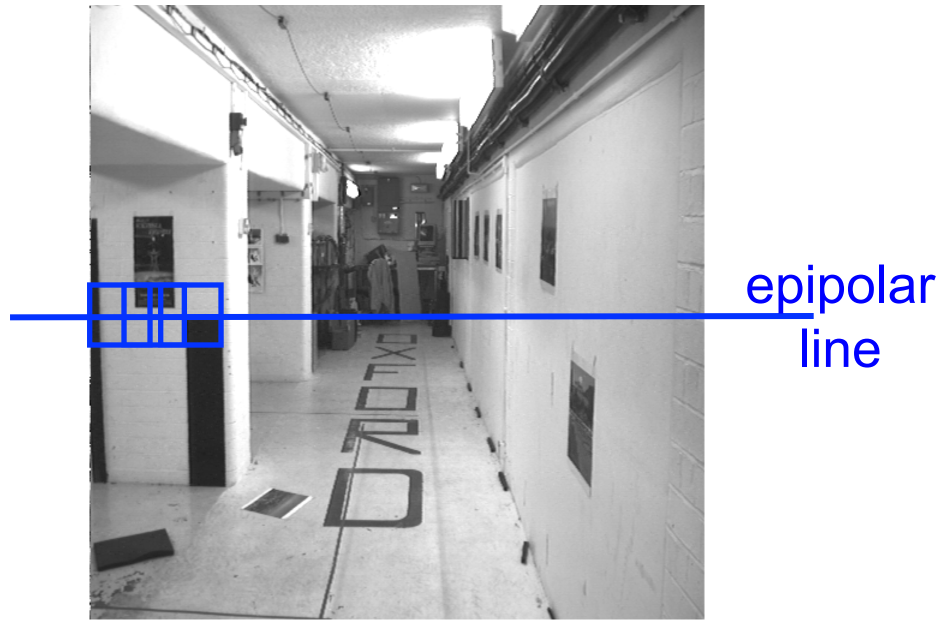
$$\mathtt{A} \rightarrow \mathbf{a}, \;\; \mathtt{B} \rightarrow \mathbf{b}$$

$$\text{NCC} = \frac{\mathbf{a}.\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

$$-1 \leq \text{NCC} \leq 1$$



Cross-correlation of neighborhood



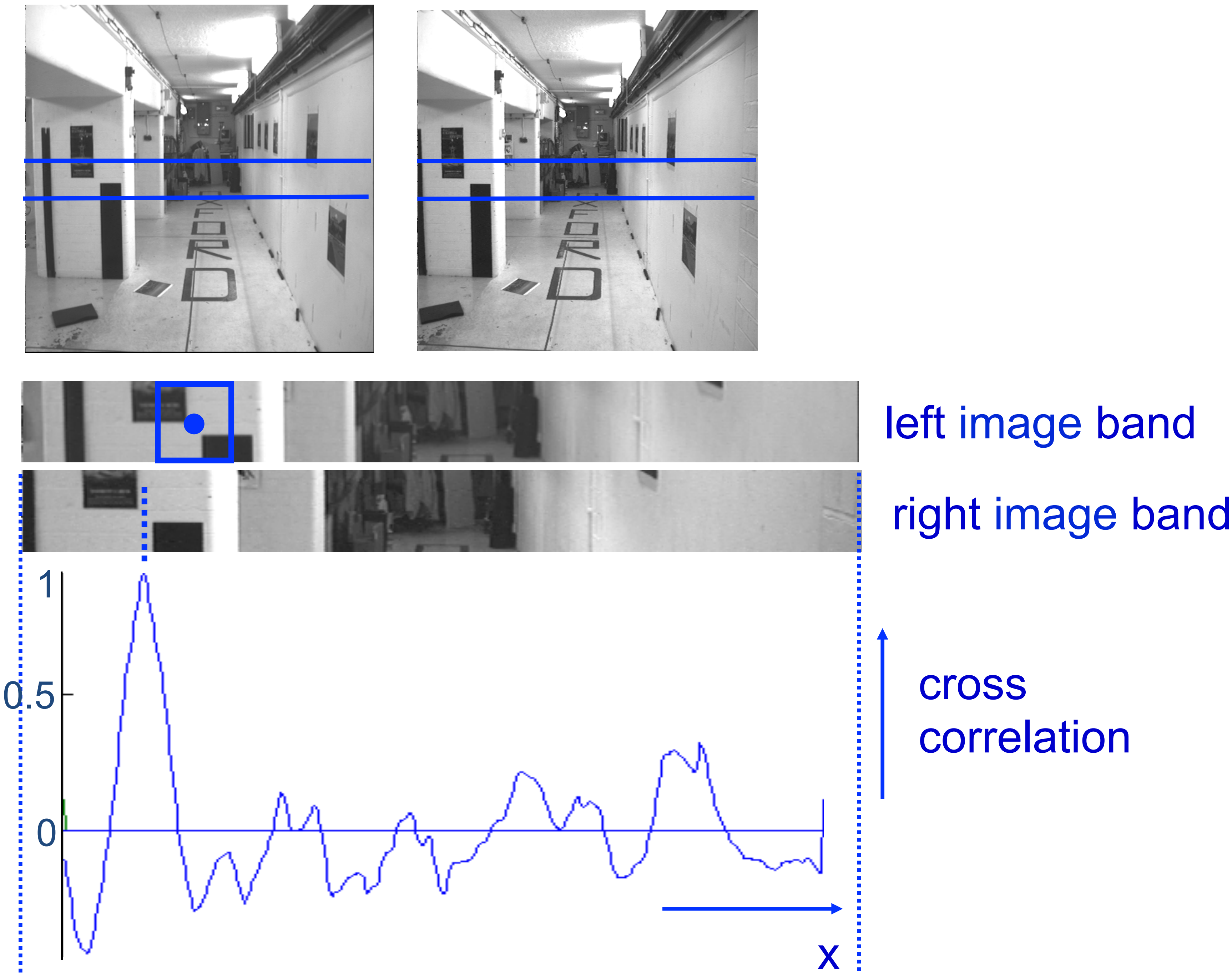
regions $\mathtt{A}, \mathtt{B}$, write as vectors $\mathbf{a}, \mathbf{b}$

translate so that mean is zero

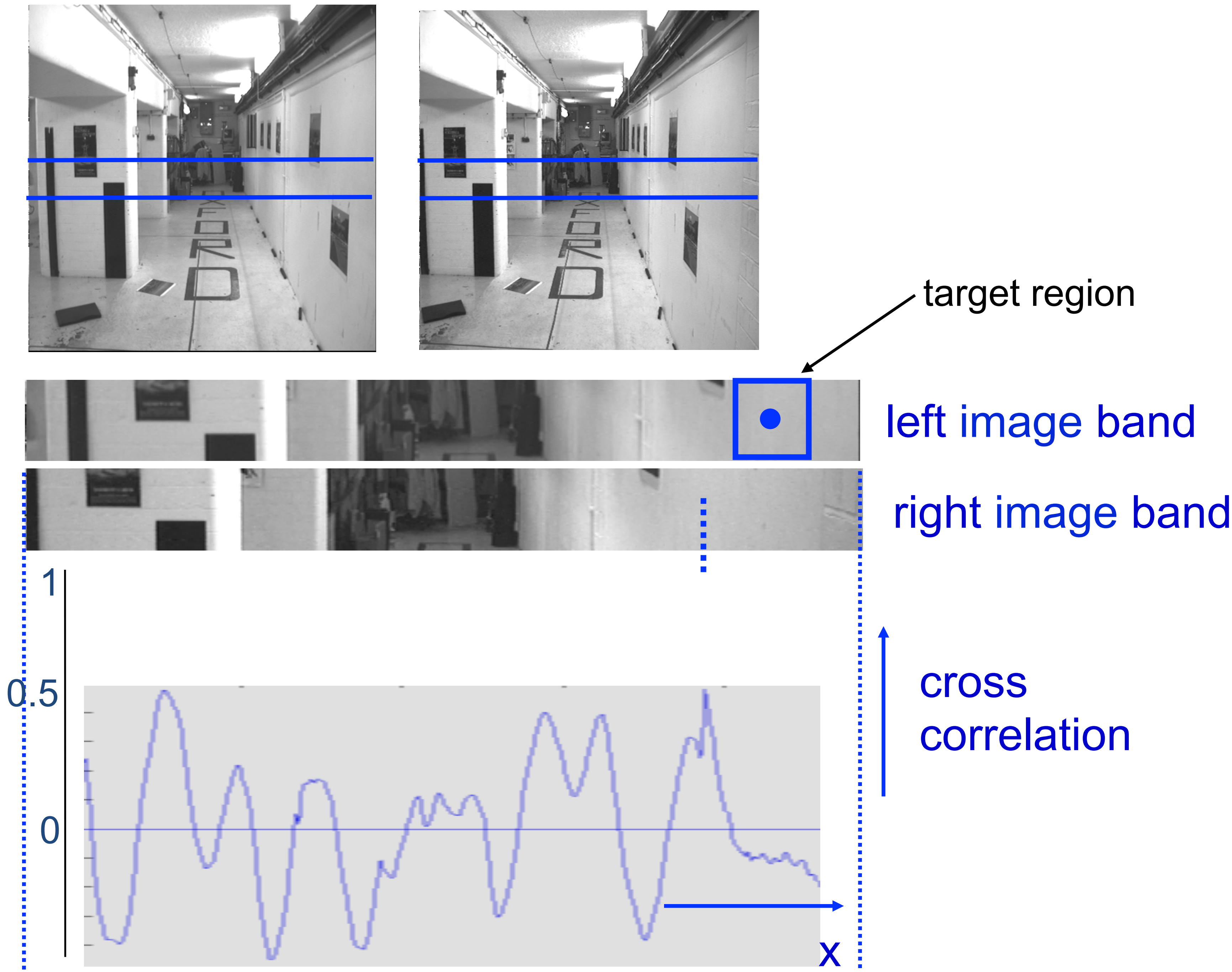
$$\mathtt{a} \rightarrow \mathtt{a} - \langle \mathbf{a} \rangle, \quad \mathtt{b} \rightarrow \mathtt{b} - \langle \mathbf{b} \rangle$$

$$\text{cross correlation} = \frac{\mathbf{a}.\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

Invariant to $I \rightarrow \alpha I + \beta$



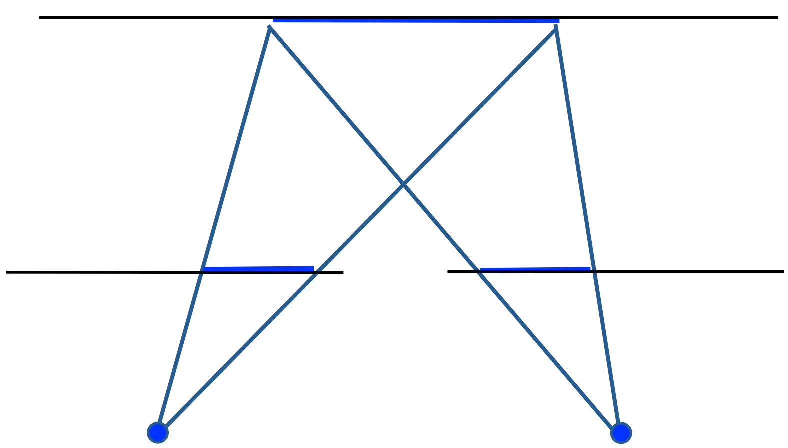
left image band
right image band
1
0.5
0
cross
correlation
x


target region
left image band
right image band
1
0.5
0
x
cross
correlation

Why is cross-correlation not great?

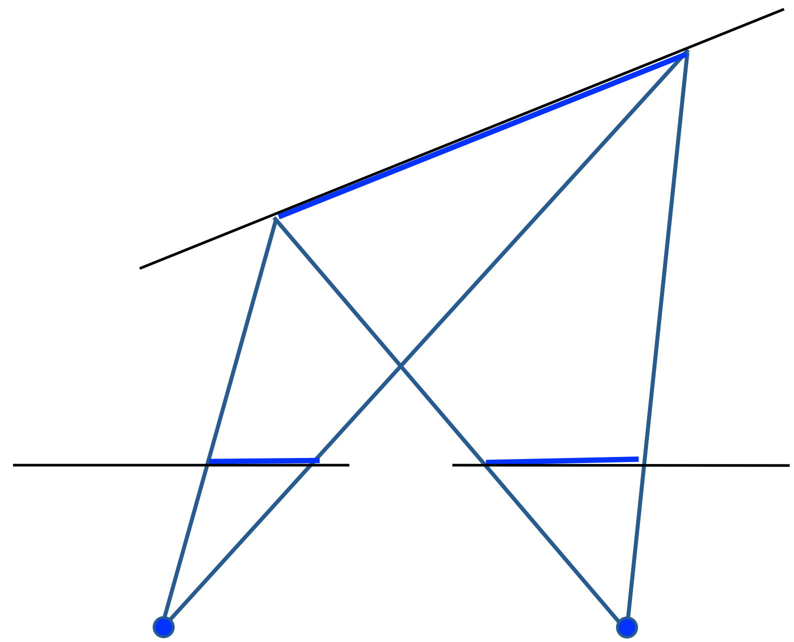
1. The neighborhood region does not have a “distinctive” spatial intensity distribution

2. Foreshortening effects



fronto-parallel surface

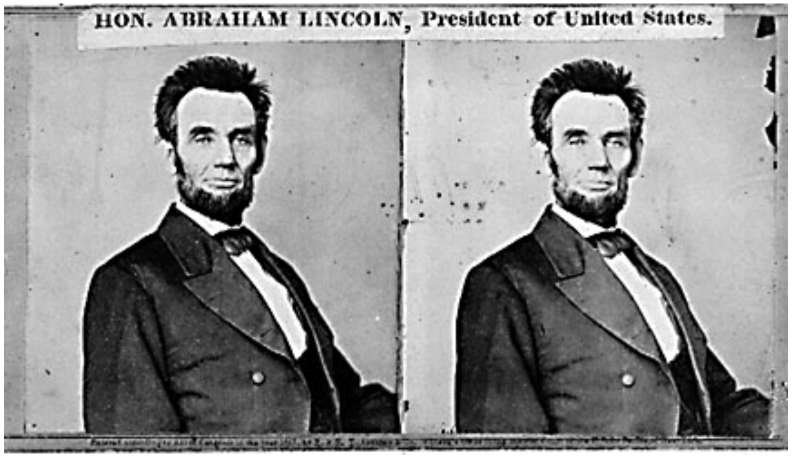
imaged length the same



slanting surface

imaged lengths differ

Limitations of similarity constraint



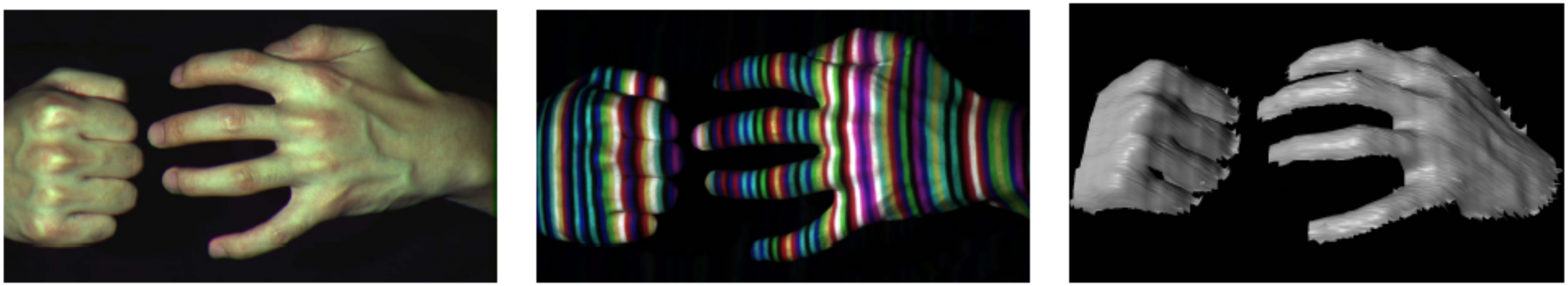
Textureless surfaces

Occlusions, repetition

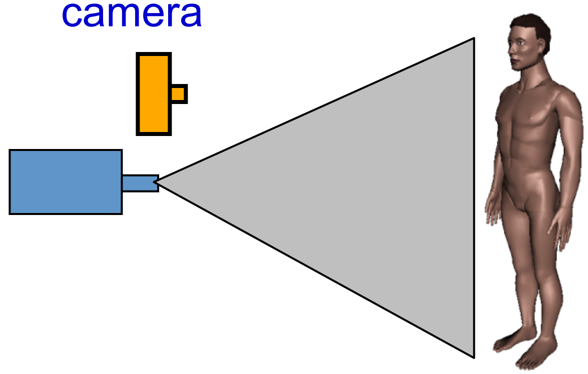
Non-Lambertian surfaces, specularities

Other approaches to obtaining 3D structure

Active stereo with structured light

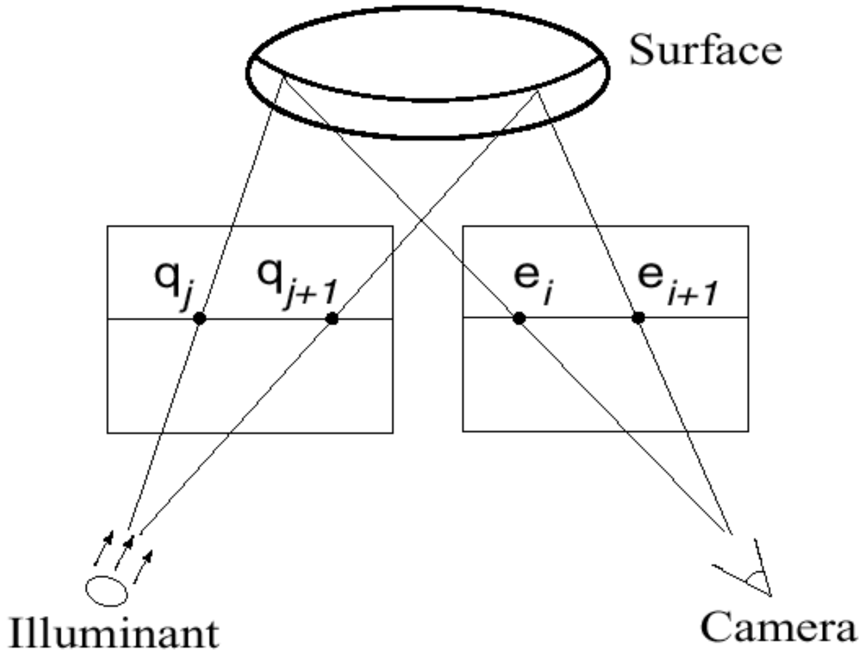


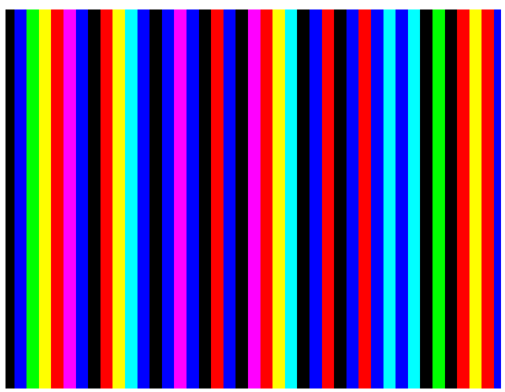
- Project "structured" light patterns onto the object
  - simplifies the correspondence problem
  - Allows us to use only one camera

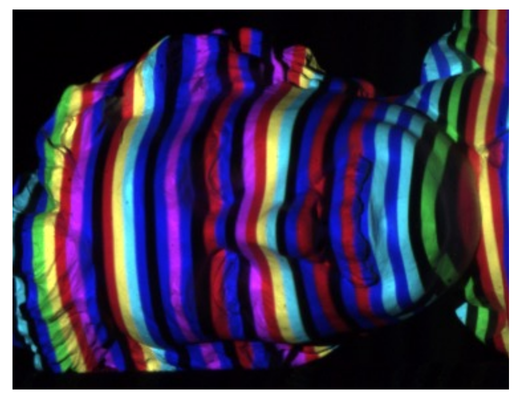


L. Zhang, B. Curless, and S. M. Seitz.
Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.
3DPVT 2002

Active stereo with structured light

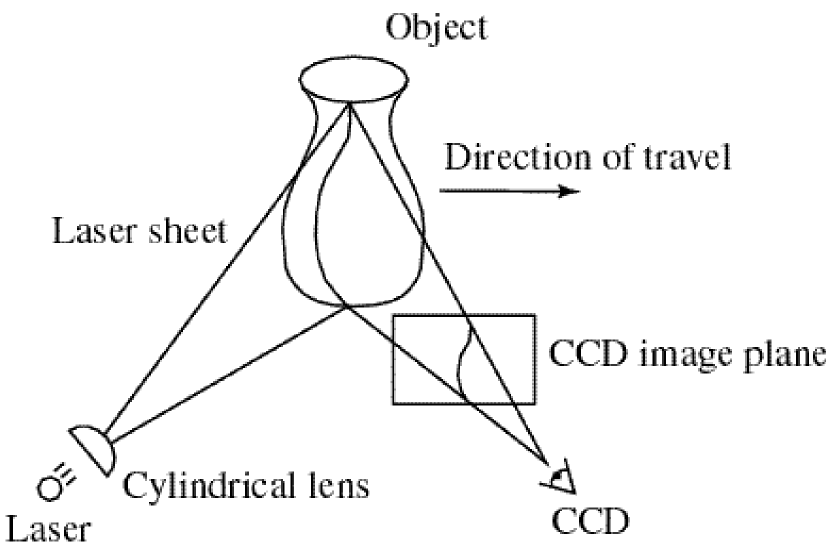


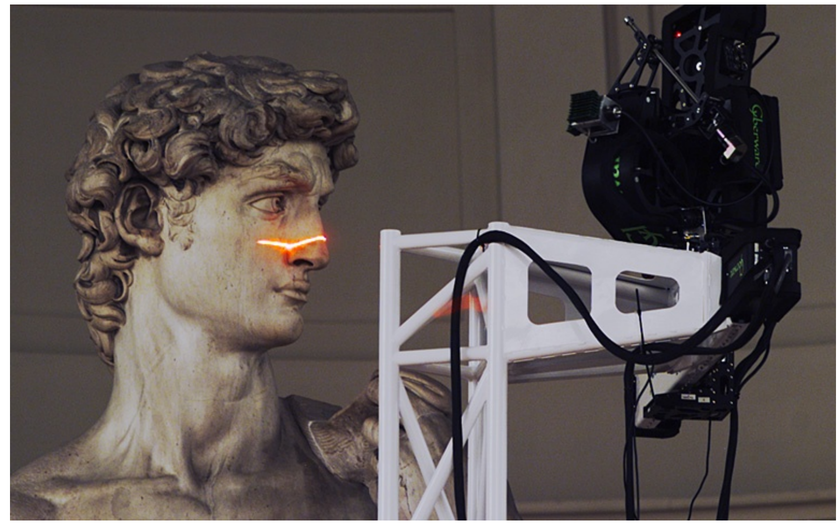




L. Zhang, B. Curless, and S. M. Seitz.
[Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.] *3DPVT* 2002

Laser scanning





Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning

Source: S. Seitz

Laser scanned models

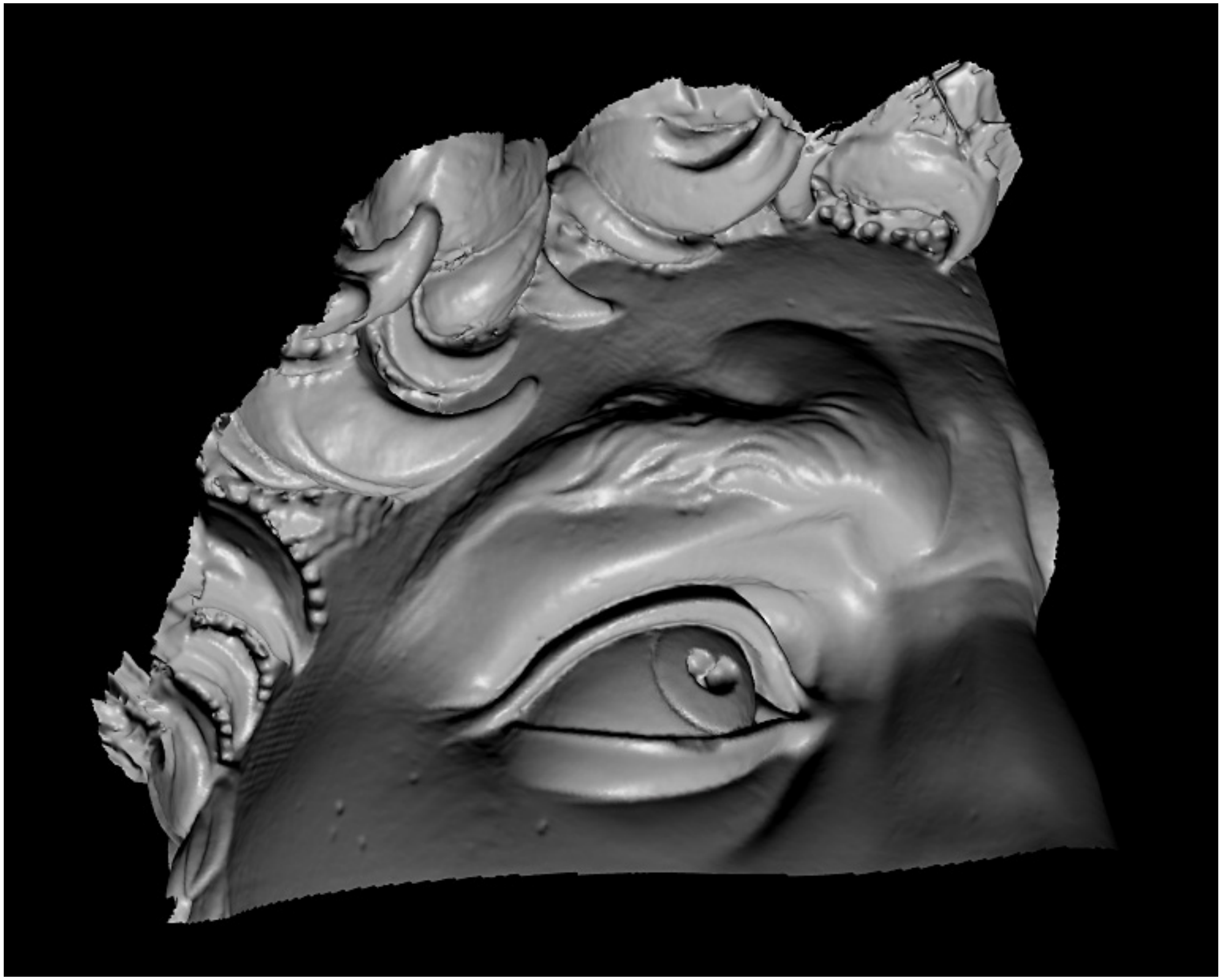
The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

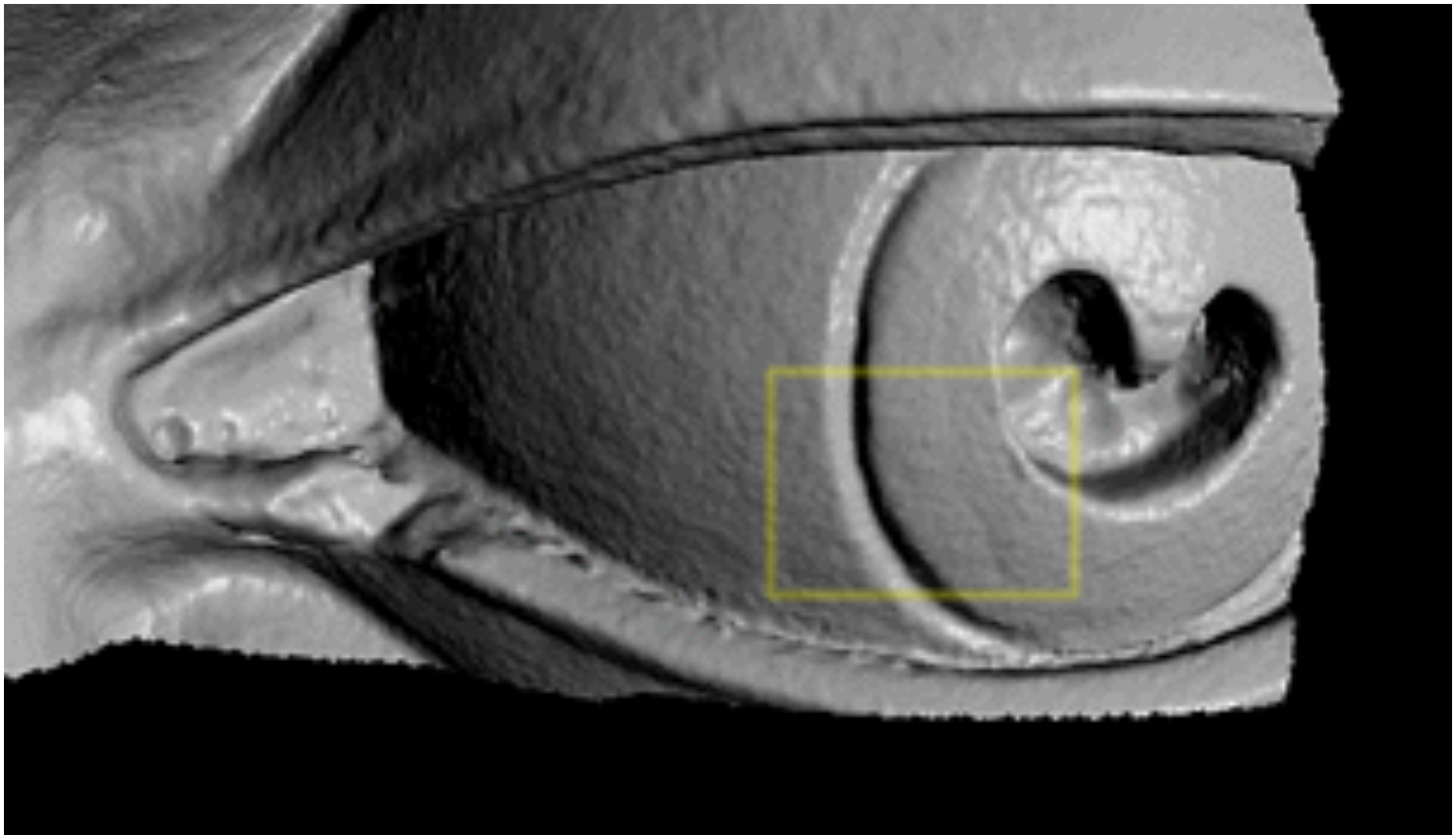
The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz



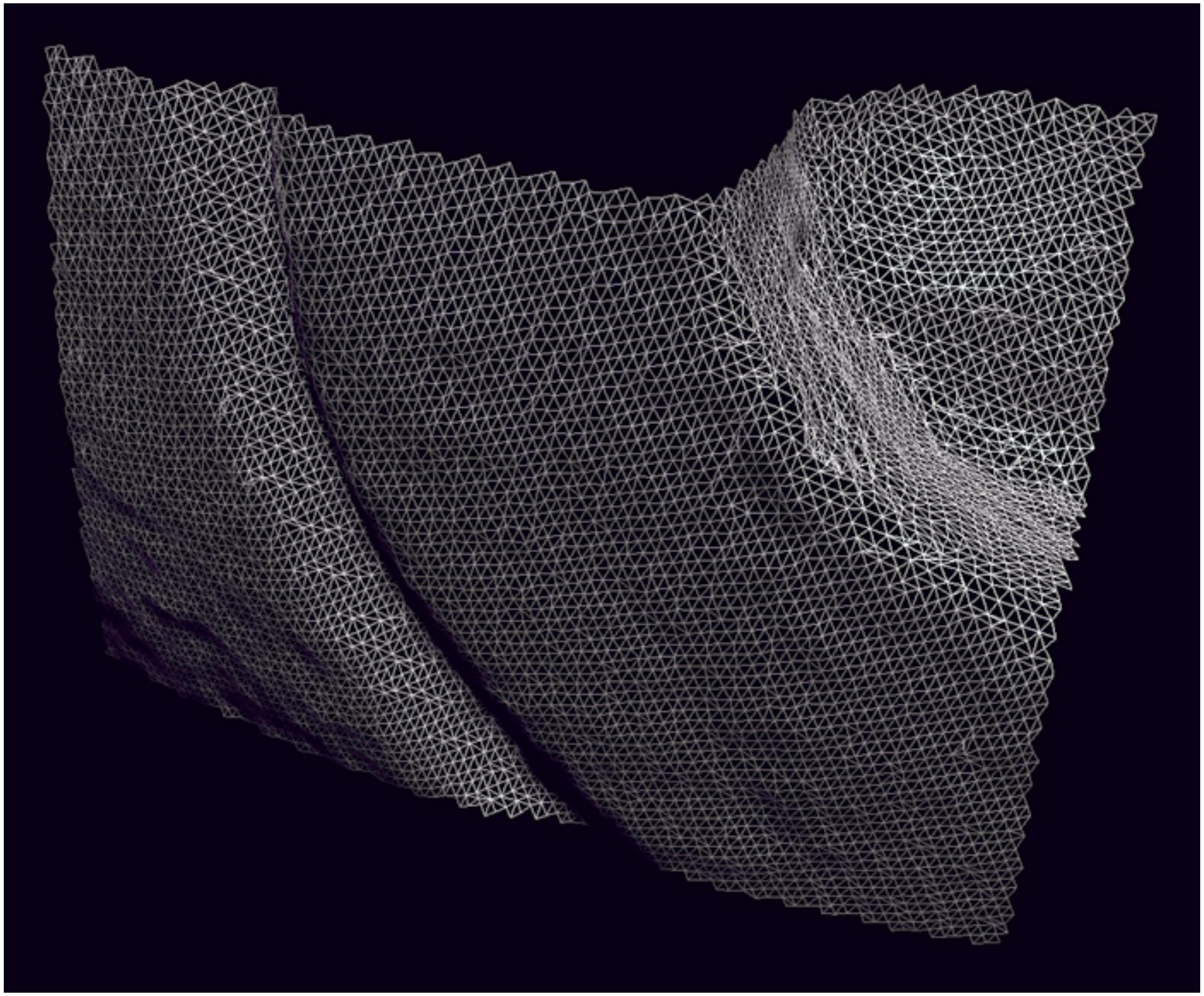
The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz



The Digital Michelangelo Project, Levoy et al.

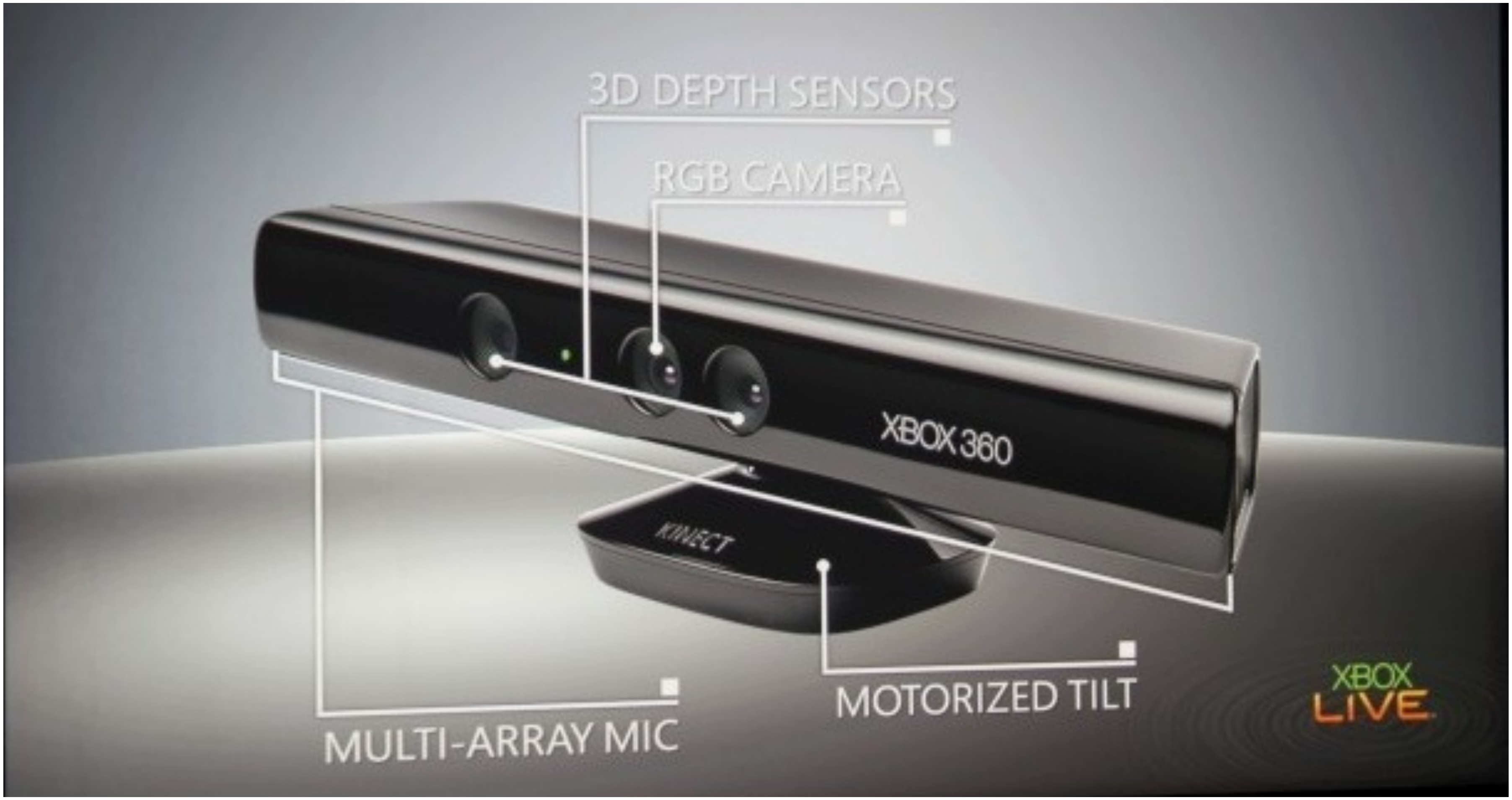
Source: S. Seitz

Aligning range images

- A single range scan is not sufficient to describe a complex surface
- Need techniques to register multiple range images

  - … which brings us to *multi-view stereo*

Microsoft Kinect



Road map

- What we've seen so far:
  - Low-level image processing: filtering, edge detecting, feature detection
  - Geometry: image transformations, panoramas, single-view modeling Fundamental matrices
- What's next:
  - Photometric stereo (PA 4)
  - Finishing up geometry: multi view stereo, structure from motion
  - Recognition

Quiz

Reminder: the cross product of two vectors is defined as:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \times \begin{bmatrix} d \\ e \\ f \end{bmatrix} = \begin{bmatrix} bf - ce \\ cd - af \\ ae - bd \end{bmatrix}$$