CSCI 4830 / 5722
Computer Vision

University of Colorado **Boulder**

CSCI 4830 / 5722
Computer Vision

Dr. Ioana Fleming

Spring 2019
Lecture 21

University of Colorado **Boulder**

Reminders

Submissions:

- Homework 4: Wed 3/20 at 11 pm

Readings:

- Szeliski:
  - chapter 11 (Stereo correspondence)
- P&F:
  - chapter 7 (Stereopsis)

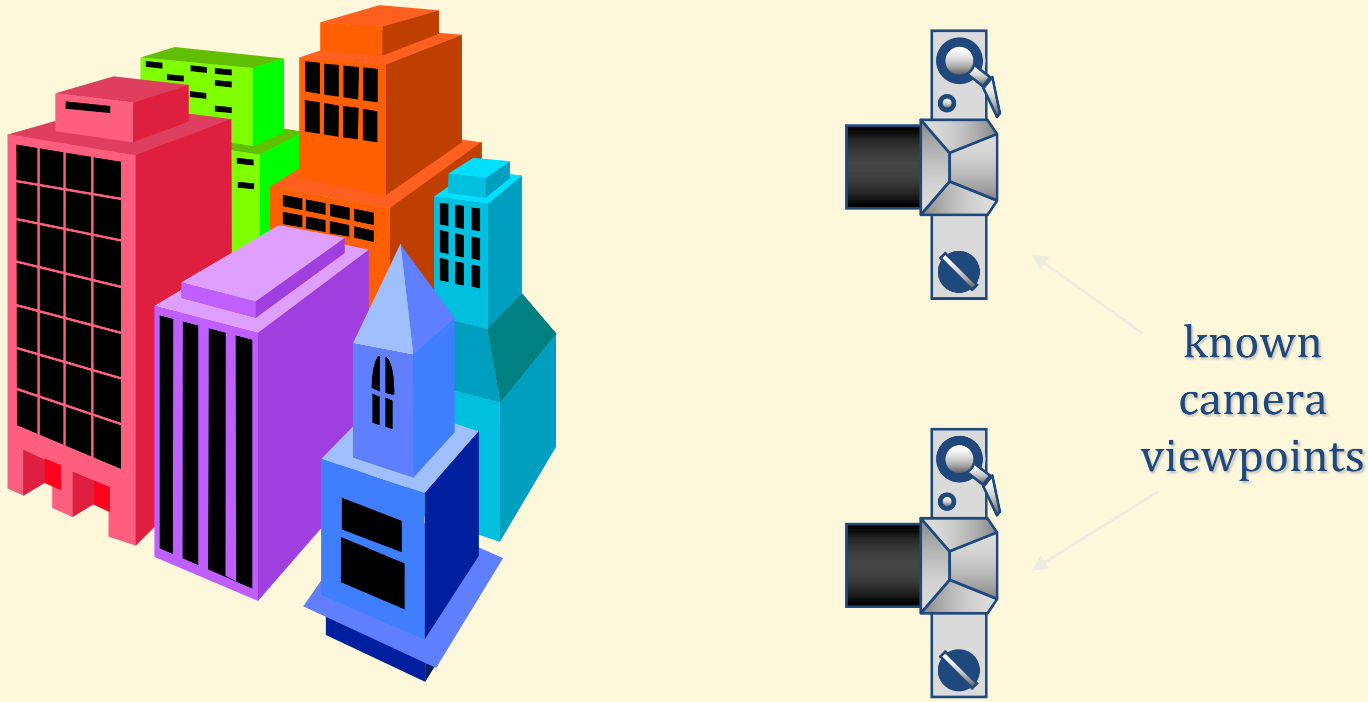
Today

- Fundamental Matrix
- Week calibration
- Stereo vision applications

…with a lot of slides stolen from Steve Seitz, Fei Fei Li, Alexei Efros

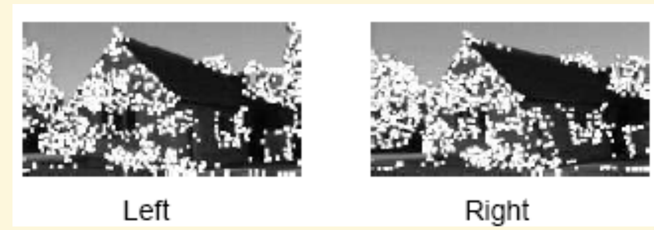
Stereo Reconstruction

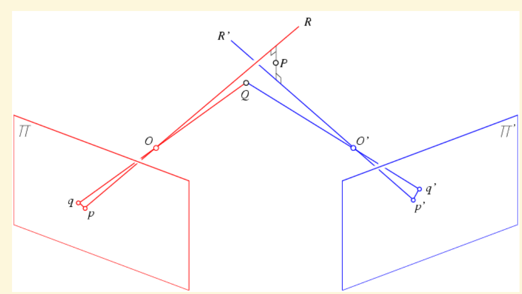
- The Stereo Problem
  - Shape from two (or more) images
  - Biological motivation



Recap: stereo with calibrated cameras

- Given image pair, **R, T**
- Detect some features
- Compute essential matrix **E**
- Match features using the epipolar and other constraints
- Triangulate for 3d structure





Stereo reconstruction pipeline

What will cause errors?

- Camera calibration errors
- Poor image resolution
- Occlusions
- Violations of brightness constancy (specular reflections)
- Large motions
- Low-contrast image regions (textureless)

Uncalibrated case

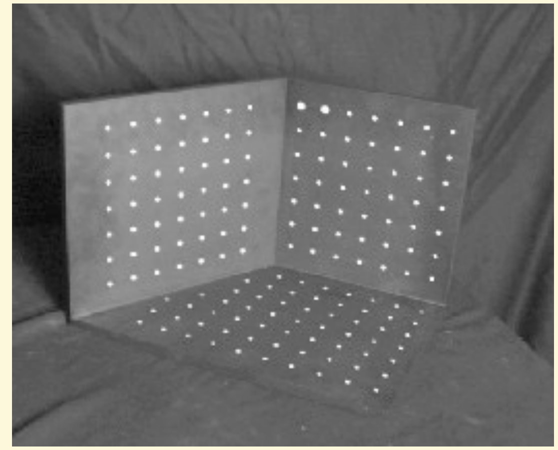
- What if we don't know the camera parameters?

Two possibilities:

1. Calibrate with a calibration object
2. Weak calibration

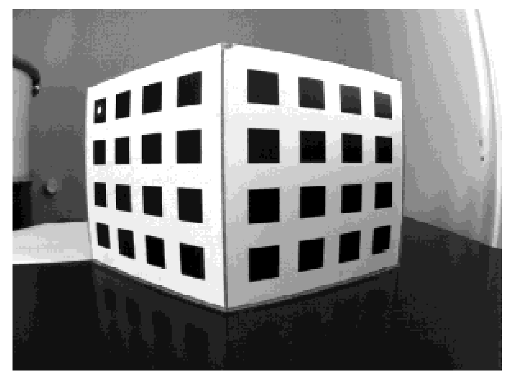
Calibrating a camera

- Compute intrinsic and extrinsic parameters using observed camera data



Main idea

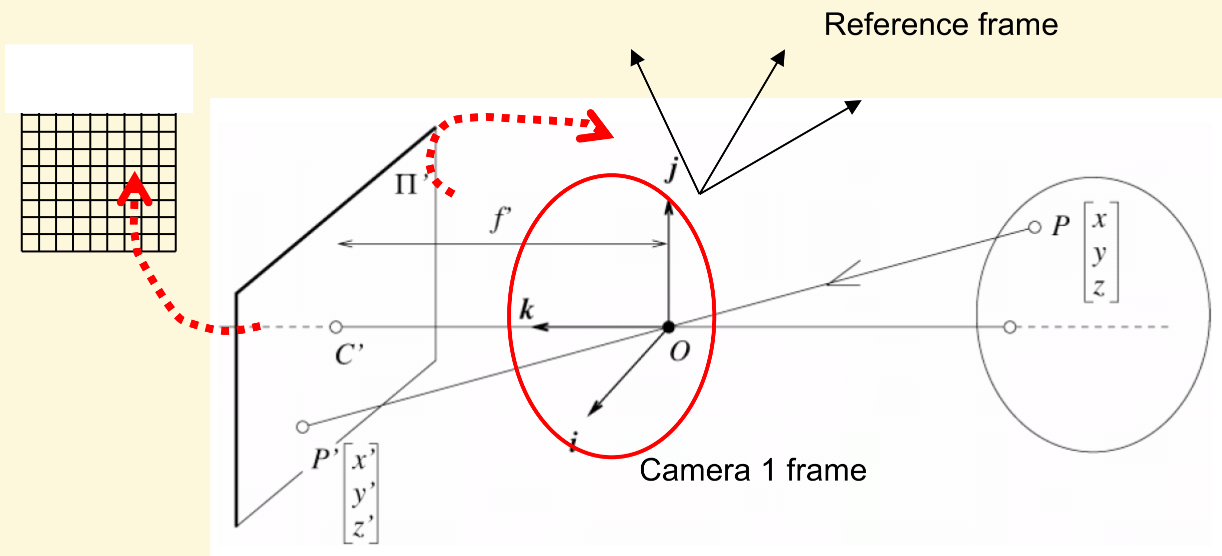
- Place "calibration object" with known geometry in the scene
- Get correspondences
- Solve for mapping from scene to image



The Opti-CAL Calibration Target Image

Camera parameters

- **Extrinsic: location and orientation of camera frame with respect to reference frame**
- Intrinsic: how to map pixel coordinates to image plane coordinates



Extrinsic camera parameters

$$\mathbf{P}_c = \mathbf{R}(\mathbf{P}_w - \mathbf{T})$$

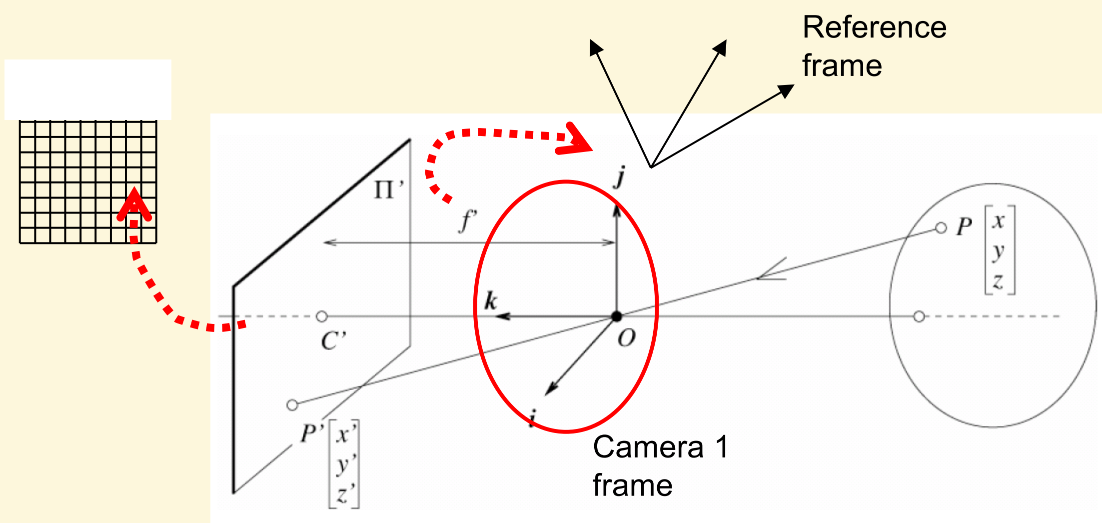
Camera reference frame

World reference frame

$$\mathbf{P}_c = (X, Y, Z)^T$$

Camera parameters

- Extrinsic: location and orientation of camera frame with respect to reference frame
- **Intrinsic: how to map pixel coordinates to image plane coordinates**



Intrinsic camera parameters

- Ignoring any geometric distortions from optics, we can describe them by:

$$x = -(x_{im} - o_x)s_x$$

$$y = -(y_{im} - o_y)s_y$$

Coordinates of projected point in camera reference frame

Coordinates of image point in pixel units

Coordinates of image center in pixel units

Effective size of a pixel (mm)

Camera parameters

- We know that in terms of camera reference frame:

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$

and

$$\mathbf{P}_c = \mathbf{R}(\mathbf{P}_w - \mathbf{T})$$
$$\mathbf{P}_c = (X, Y, Z)^T$$

- Substituting previous eqns describing intrinsic and extrinsic parameters, can relate *pixels coordinates* to *world points:*

$$-(x_{im} - o_x)s_x = f\frac{\mathbf{R}_1 \cdot (\mathbf{P}_w - \mathbf{T})}{\mathbf{R}_3 \cdot (\mathbf{P}_w - \mathbf{T})}$$

$$-(y_{im} - o_y)s_y = f\frac{\mathbf{R}_2 \cdot (\mathbf{P}_w - \mathbf{T})}{\mathbf{R}_3 \cdot (\mathbf{P}_w - \mathbf{T})}$$

$\mathbf{R}_i$ = Row i of rotation matrix

Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates, where:

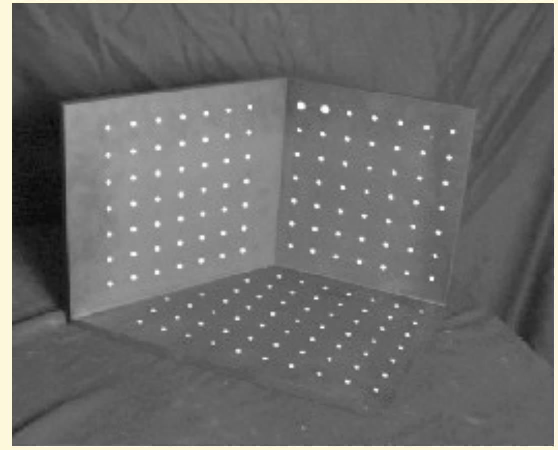
$$\begin{bmatrix} wx_{im} \\ wy_{im} \\ w \end{bmatrix} = \mathbf{M}P_w$$

$$\mathbf{M}_{\text{int}} = \begin{bmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{M}_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -\mathbf{R}_1{}^{\mathrm{T}}\mathbf{T} \\ r_{21} & r_{22} & r_{23} & -\mathbf{R}_2{}^{\mathrm{T}}\mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3{}^{\mathrm{T}}\mathbf{T} \end{bmatrix}$$

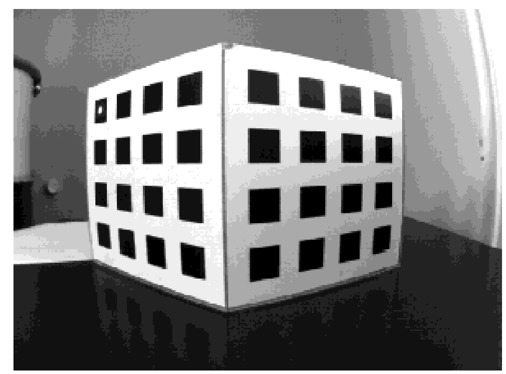
Calibrating a camera

- Compute intrinsic and extrinsic parameters using observed camera data



Main idea

- Place “calibration object” with known geometry in the scene
- Get correspondences
- Solve for mapping from scene to image: estimate $\mathbf{M}=\mathbf{M}_{int}\mathbf{M}_{ext}$



The Opti-CAL Calibration Target Image

When would we calibrate this way?

- Makes sense when geometry of system is not going to change over time

Weak calibration

- Want to estimate world geometry without requiring calibrated cameras
  - Archival videos
  - Photos from multiple unrelated users
  - Dynamic camera system

- **Main idea**:
  - Estimate epipolar geometry from a (redundant) set of point correspondences between two uncalibrated cameras

From before: Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates:

$$\begin{bmatrix} wx_{im} \\ wy_{im} \\ w \end{bmatrix} = \mathbf{M}_{\text{int}}\mathbf{M}_{ext}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$\mathbf{M}_{\text{int}} = \begin{bmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{M}_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -\mathbf{R}_1{}^{\mathrm{T}}\mathbf{T} \\ r_{21} & r_{22} & r_{23} & -\mathbf{R}_2{}^{\mathrm{T}}\mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3{}^{\mathrm{T}}\mathbf{T} \end{bmatrix}$$

From before: Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates:

$$\begin{bmatrix} wx_{im} \\ wy_{im} \\ w \end{bmatrix} = \mathbf{M}_{int}\mathbf{M}_{ext}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$\mathbf{p}_{im} = \mathbf{M}_{int}\underbrace{\mathbf{M}_{ext}\mathbf{P}_w}_{\mathbf{p}_c}$$

$$\mathbf{p}_{im} = \mathbf{M}_{int}\mathbf{p}_c$$

Uncalibrated case

For a given camera:

$$\mathbf{p}_{im} = \mathbf{M}_{int}\mathbf{p}_{c}$$

So, for **two** cameras (left and right):

$$\mathbf{p}_{c,left} = \mathbf{M}_{int,left}^{-1}\mathbf{p}_{im,left}$$

$$\mathbf{p}_{c,right} = \underbrace{\mathbf{M}_{int,right}^{-1}}_{\text{Internal calibration matrices, one per camera}}\mathbf{p}_{im,right}$$

$$\mathbf{p}_{c,left} = \mathbf{M}_{int,left}^{-1}\mathbf{p}_{im,left}$$
$$\mathbf{p}_{c,right} = \mathbf{M}_{int,right}^{-1}\mathbf{p}_{im,right}$$

Uncalibrated case

$$\boxed{\mathbf{p}_{c,right}{}^{\mathrm{T}}\mathbf{E}\mathbf{p}_{c,left} = 0}$$

From before, the **essential** matrix **E.**

$$\left(\mathbf{M}_{int,right}^{-1}\mathbf{p}_{im,right}\right)^{\mathrm{T}}\mathbf{E}\left(\mathbf{M}_{int,left}^{-1}\mathbf{p}_{im,left}\right) = 0$$

$$\mathbf{p}_{im,right}^{\mathrm{T}}\underbrace{\left(\mathbf{M}_{int,right}^{-\mathrm{T}}\mathbf{E}\mathbf{M}_{int,left}^{-1}\right)}_{\mathbf{F}}\mathbf{p}_{im,left} = 0$$

“Fundamental matrix”

$$\boxed{\mathbf{p}_{im,right}^{\mathrm{T}}\mathbf{F}\mathbf{p}_{im,left} = 0}$$

Computing F from correspondences

Each point correspondence generates one constraint on F

$$\mathbf{p}^{\mathrm{T}}_{im,right}\mathbf{F}\mathbf{p}_{im,left} = 0$$

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

Collect n of these constraints

$$\begin{bmatrix} u'_1u_1 & u'_1v_1 & u'_1 & v'_1u_1 & v'_1v_1 & v'_1 & u_1 & v_1 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \mathbf{0}$$

Solve for f , vector of parameters.

Computing F from correspondences

$$\mathbf{W} \begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1 \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' & 1 \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' & 1 \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' & 1 \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' & 1 \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' & 1 \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' & 1 \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' & 1 \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} \mathbf{f} = 0$$

- Homogeneous system $\mathbf{W}\,\mathbf{f} = 0$
- Rank 8 ⟶ A non-zero solution exists (unique)
- If N>8 ⟶ Lsq. solution by SVD! ⟶ $\hat{\mathrm{F}}$ $\|\mathbf{f}\| = 1$

Fundamental matrix

- Relates **pixel coordinates** in the two views
- More general form than essential matrix: we remove need to know intrinsic parameters

- If we estimate fundamental matrix from correspondences in *pixel coordinates*, can reconstruct epipolar geometry without intrinsic or extrinsic parameters.

Stereo pipeline with weak calibration

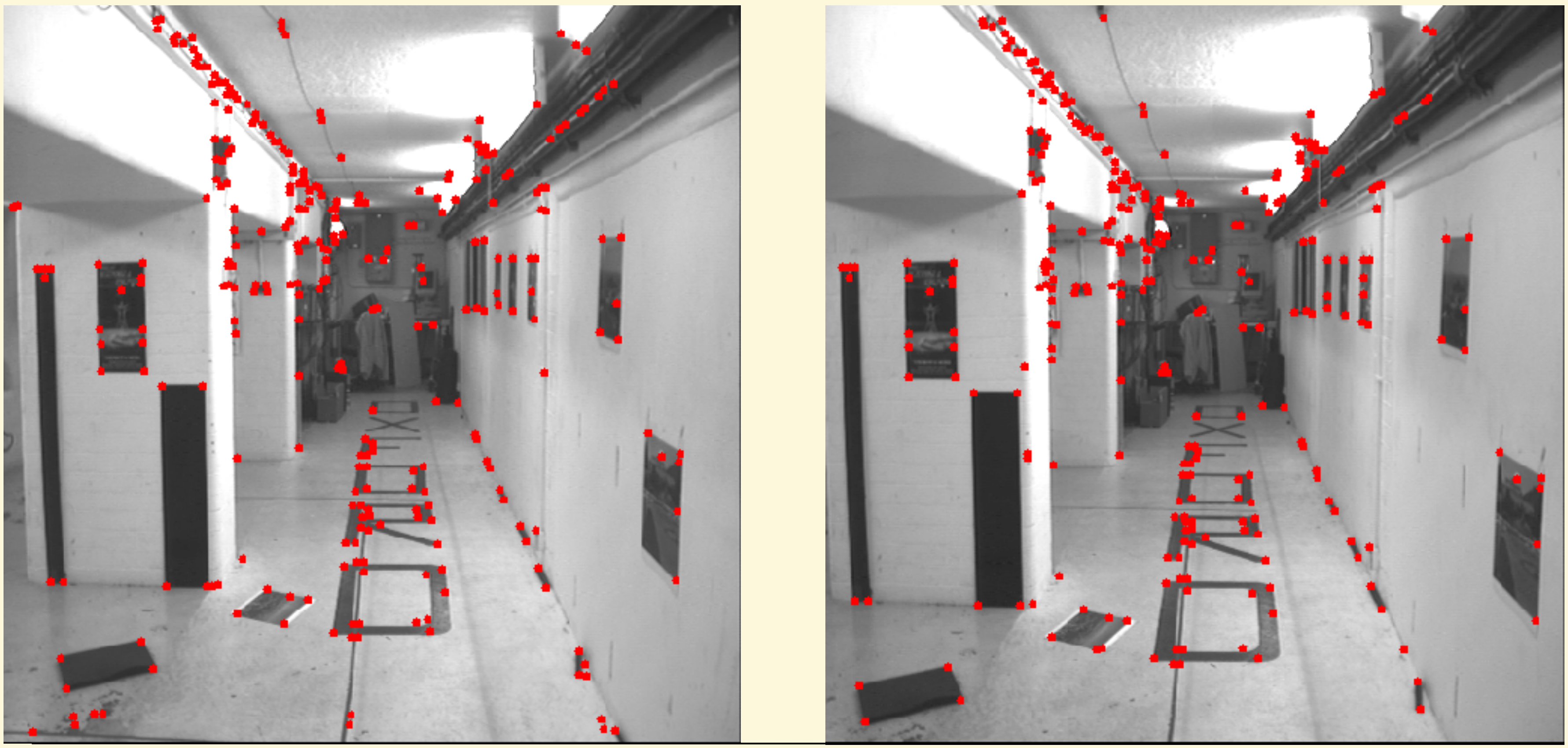
- So, where to start with uncalibrated cameras?
  - Need to find fundamental matrix F **and** the correspondences (pairs of points (u',v') ↔ (u,v)).

1) Find interest points in image
2) Compute correspondences
3) Compute epipolar geometry
4) Refine

Example from Andrew Zisserman

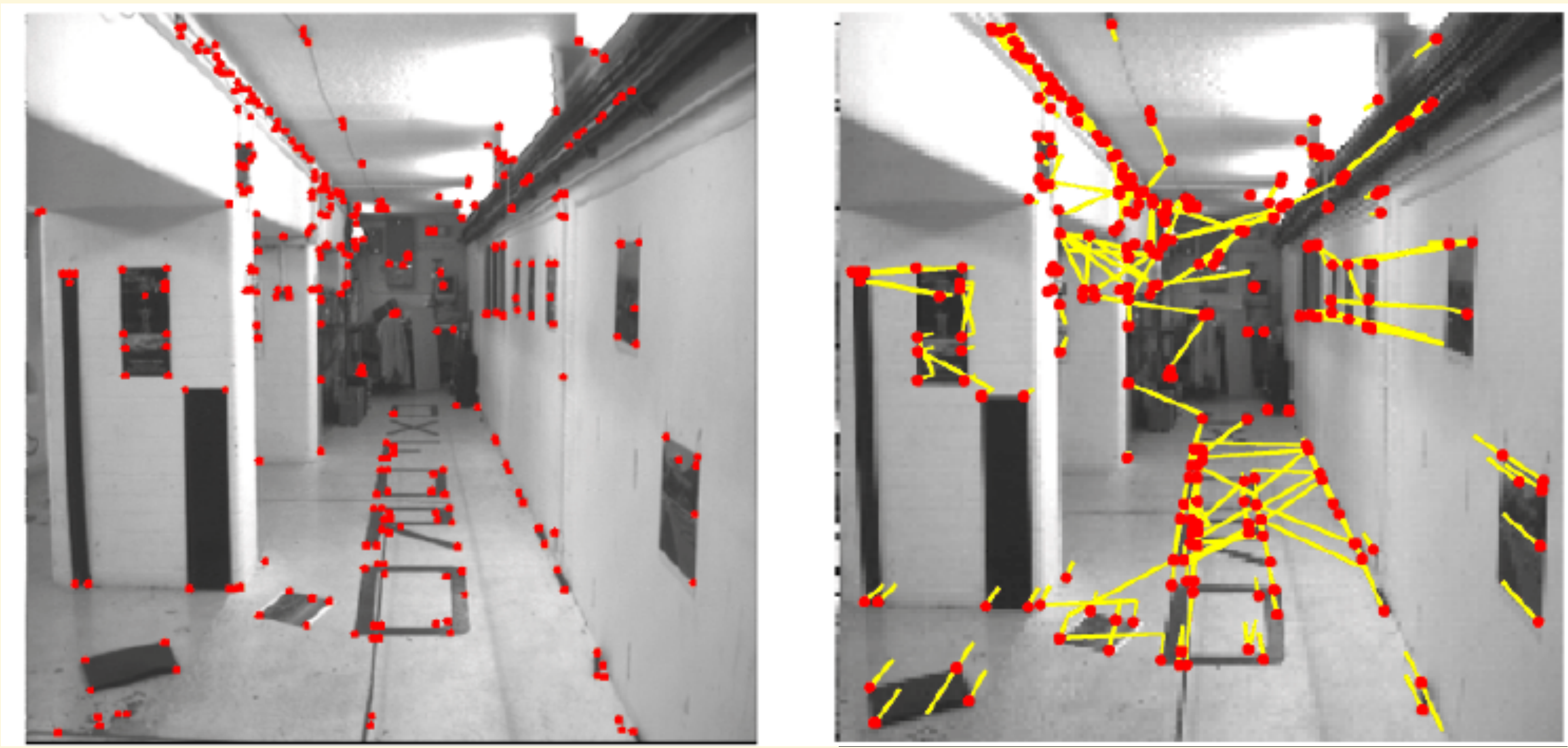
Stereo pipeline with weak calibration

1) Find interest points



Stereo pipeline with weak calibration

2) Match points within proximity to get putative matches



Stereo pipeline with weak calibration

3) Compute epipolar geometry -- robustly with RANSAC

Select random sample of putative correspondences

Compute **F** using them

- determines epipolar constraint

Evaluate amount of support

- inliers within threshold distance of epipolar line

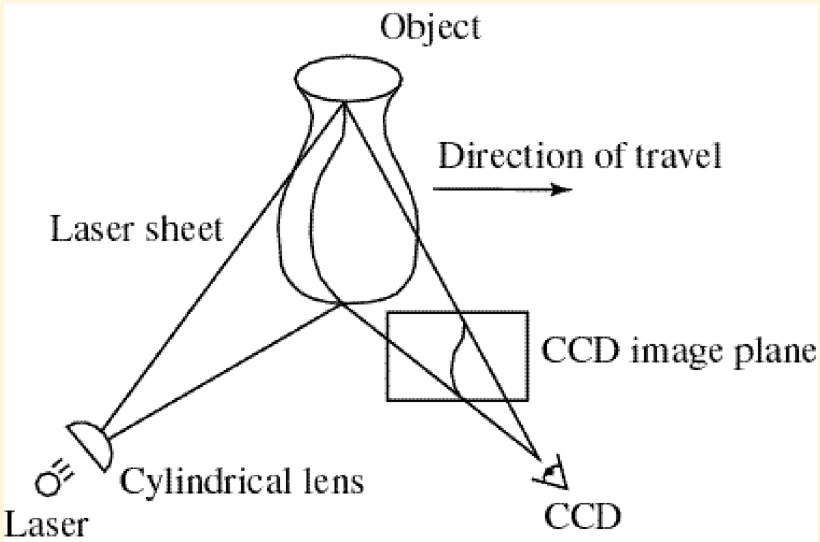
Choose F with most support (inliers)

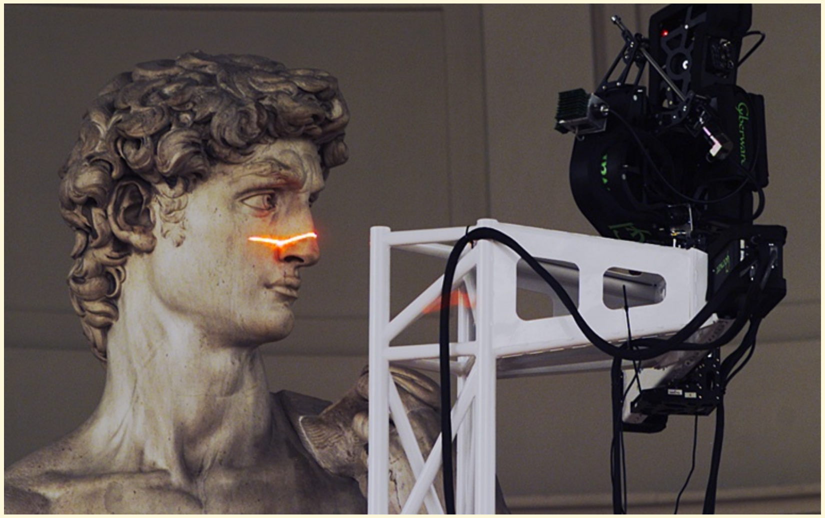
Summary

- **Rectification**: make epipolar lines align with scanlines
- Stereo solutions:
  - **Correspondence**: dense, or at interest points
  - **Non-geometric stereo constraints** (e.g., similarity, order, smoothness)
- Calibration
  - **With calibration object** in scene: relate world coordinates to image coordinates
  - **Weak calibration**: solve for fundamental matrix, relate image coordinates to image coordinates

Stereo Vision Applications

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

Laser scanned models

The Digital Michelangelo Project, Levoy et al.

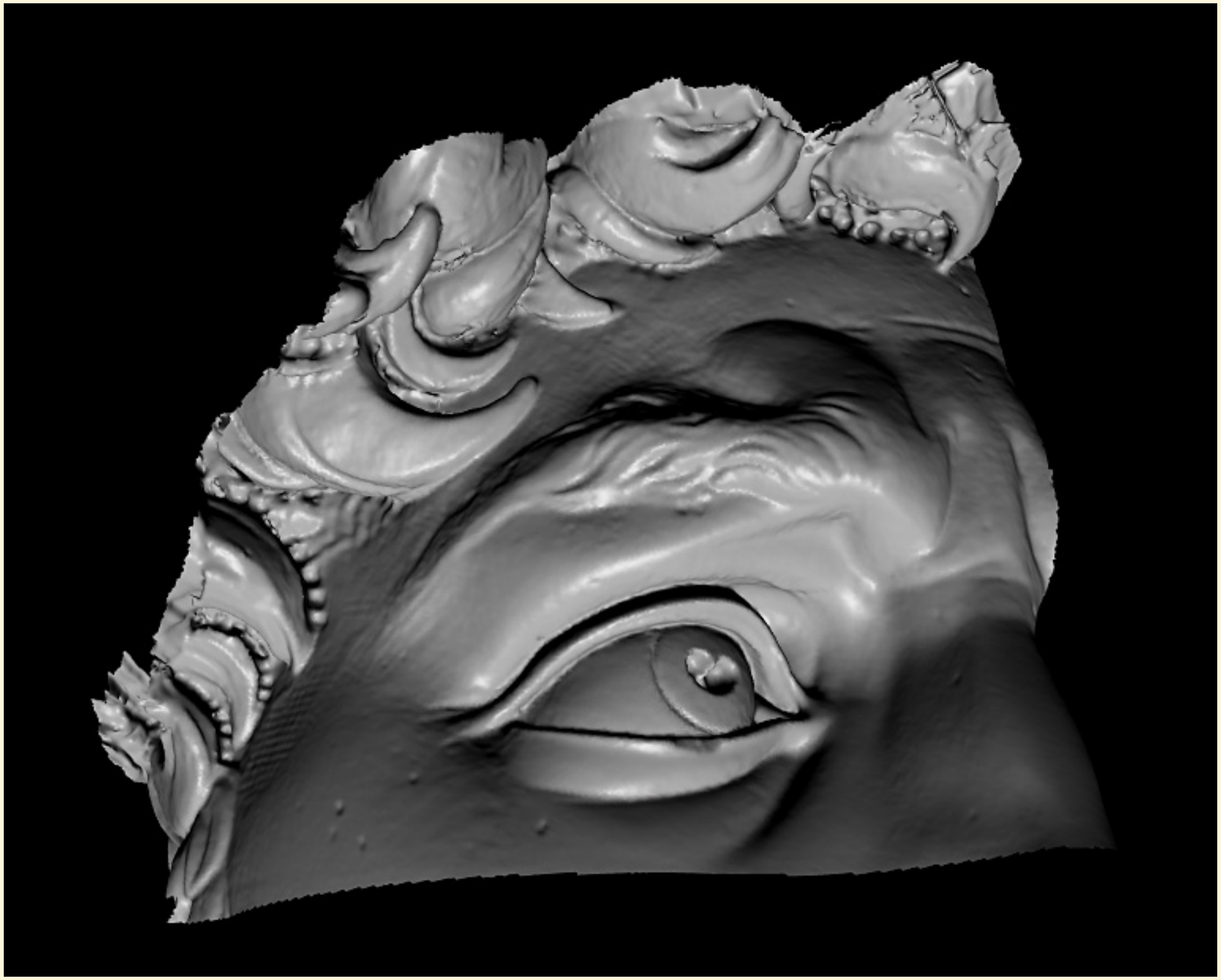
Source: S. Seitz

Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

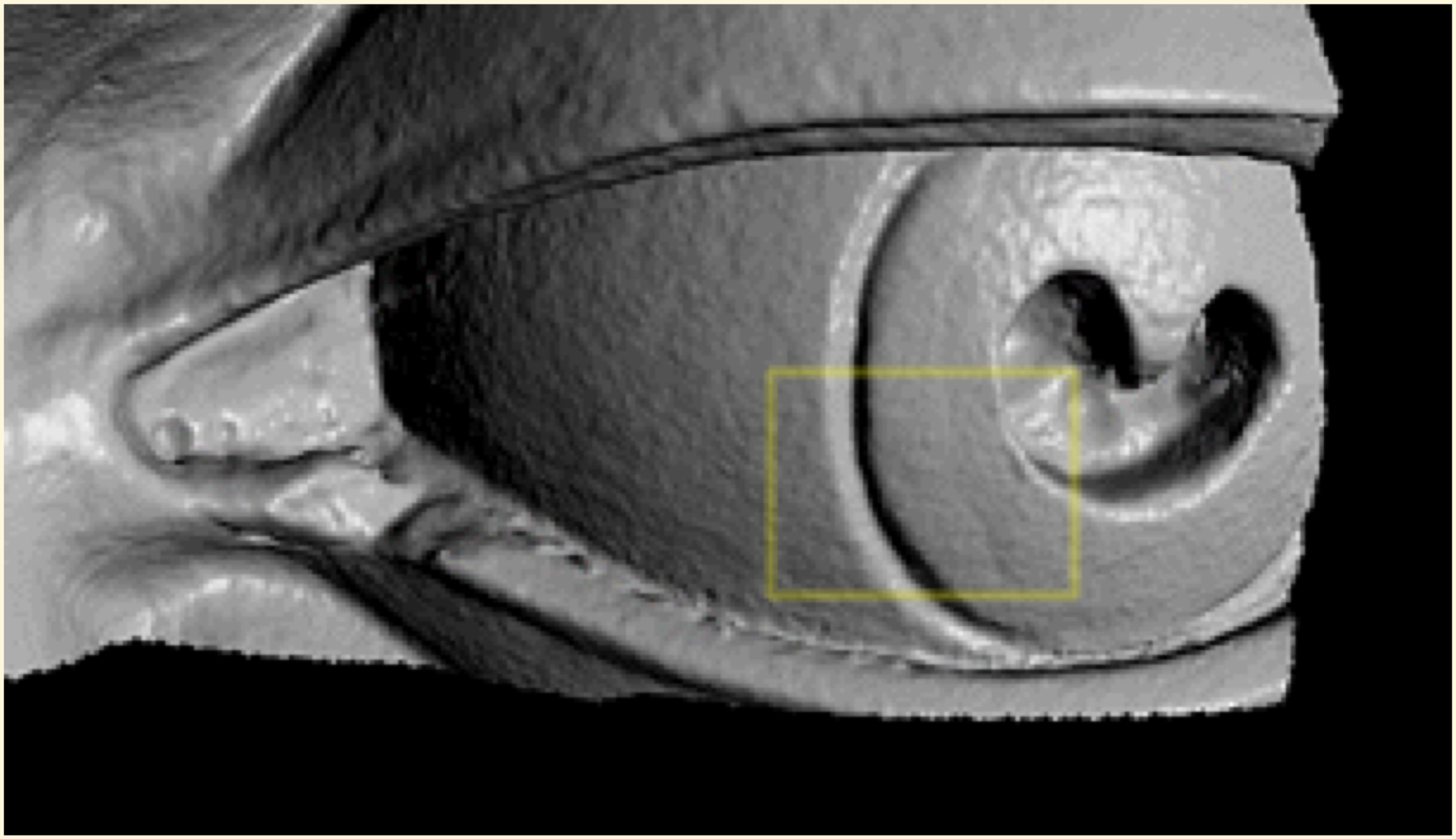
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

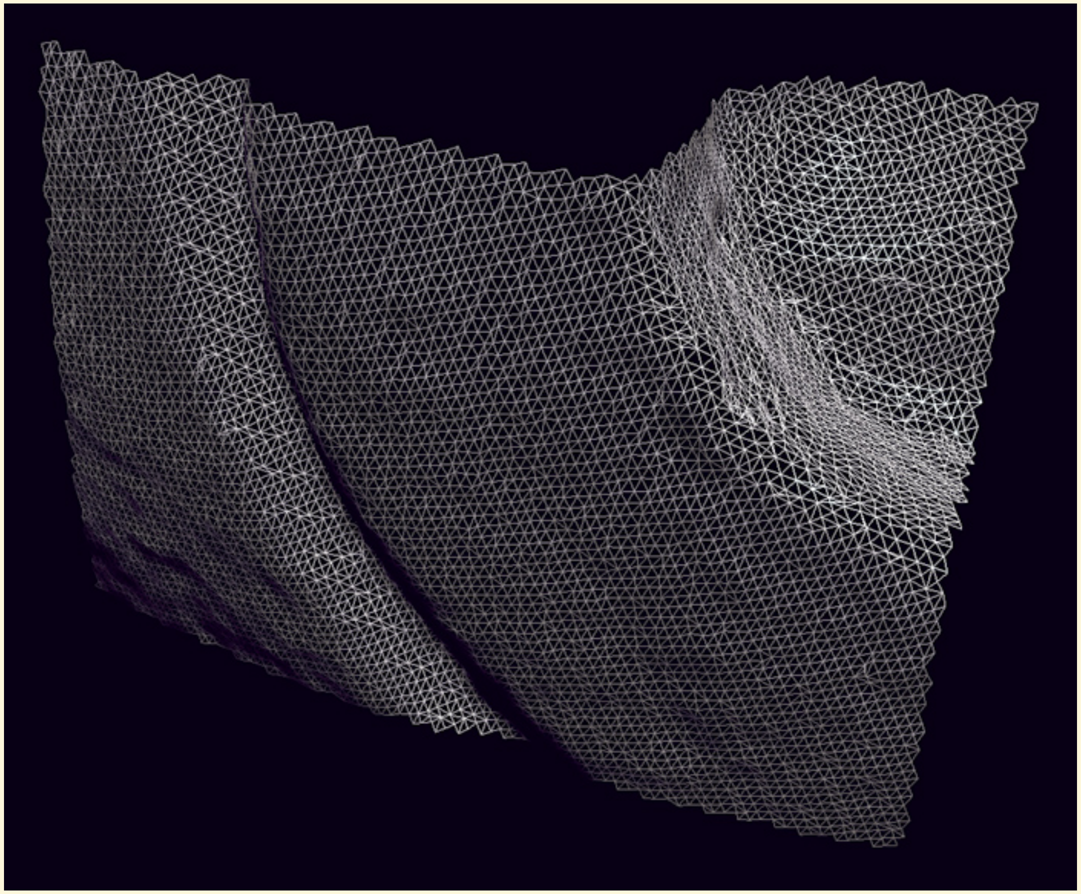


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

1.0 mm resolution (56 million triangles)



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

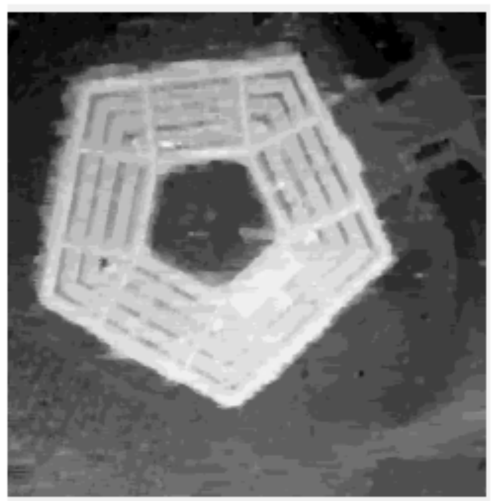
Stereo Vision Applications

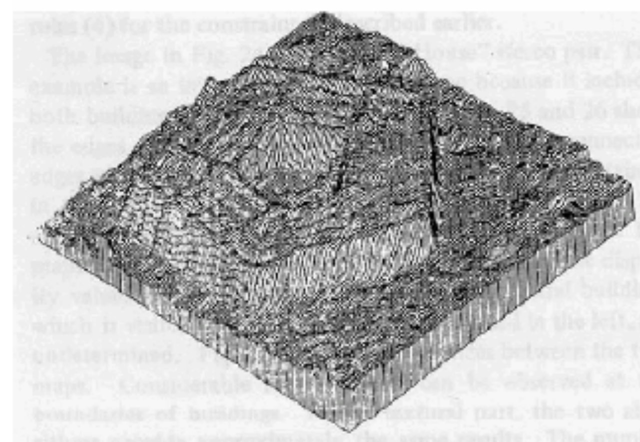
- Example depth maps (pentagon)

left image

right image

range map

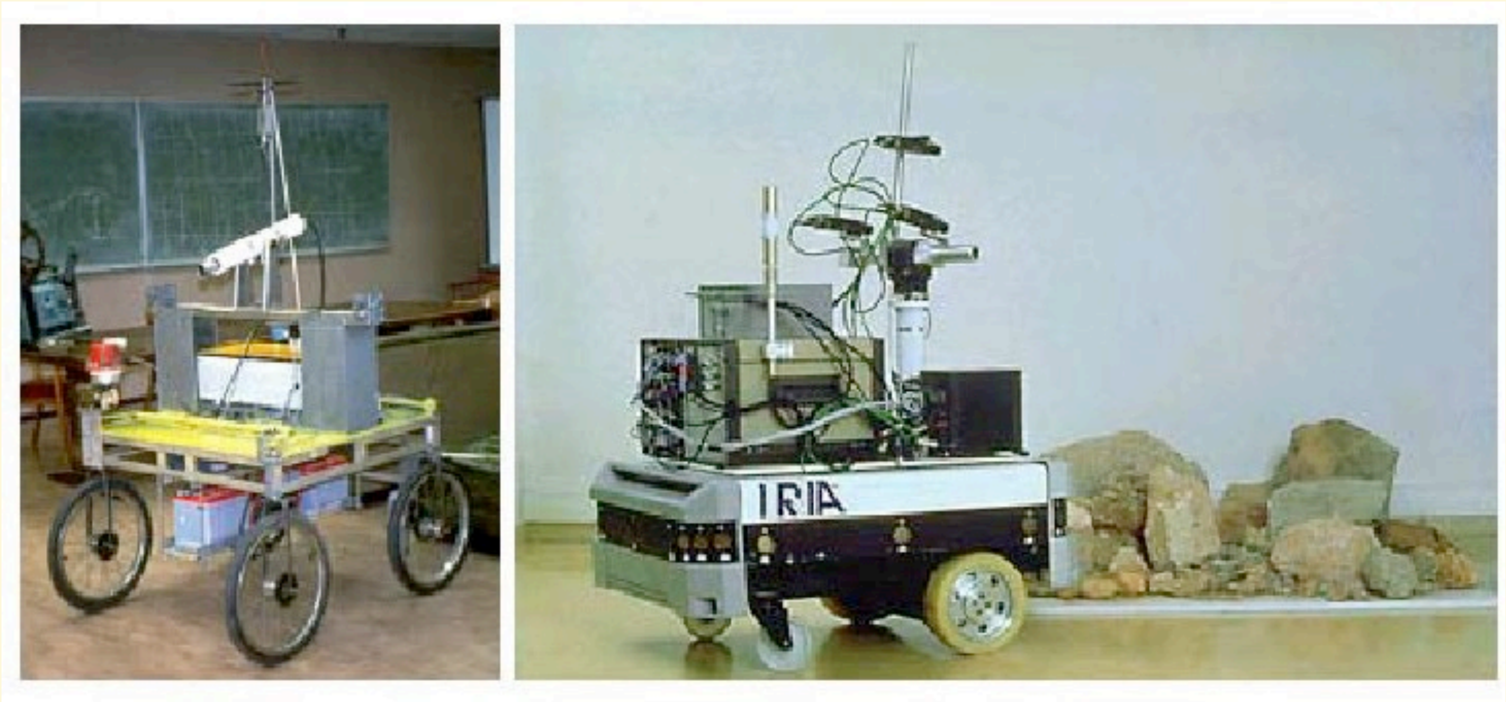




STEREO (Solar Terrestrial Relations Observatory)

STEREO is a solar observation mission. Two nearly identical spacecraft were launched in 2006 into orbits around the Sun that cause them to respectively pull farther ahead of and fall gradually behind the Earth. This enables stereoscopic imaging of the Sun and solar phenomena, such as coronal mass ejections.

Stereo in machine vision systems



Left : The Stanford cart sports a single camera moving in discrete increments along a straight line and providing multiple snapshots of scenes. Right : The INRIA mobile robot uses three cameras to map its environment

Real-time stereo

Nomad robot searches for meteorites in Antartica
http://www.frc.ri.cmu.edu/projects/meteorobot/index.html

- Used for robot navigation (and other tasks)

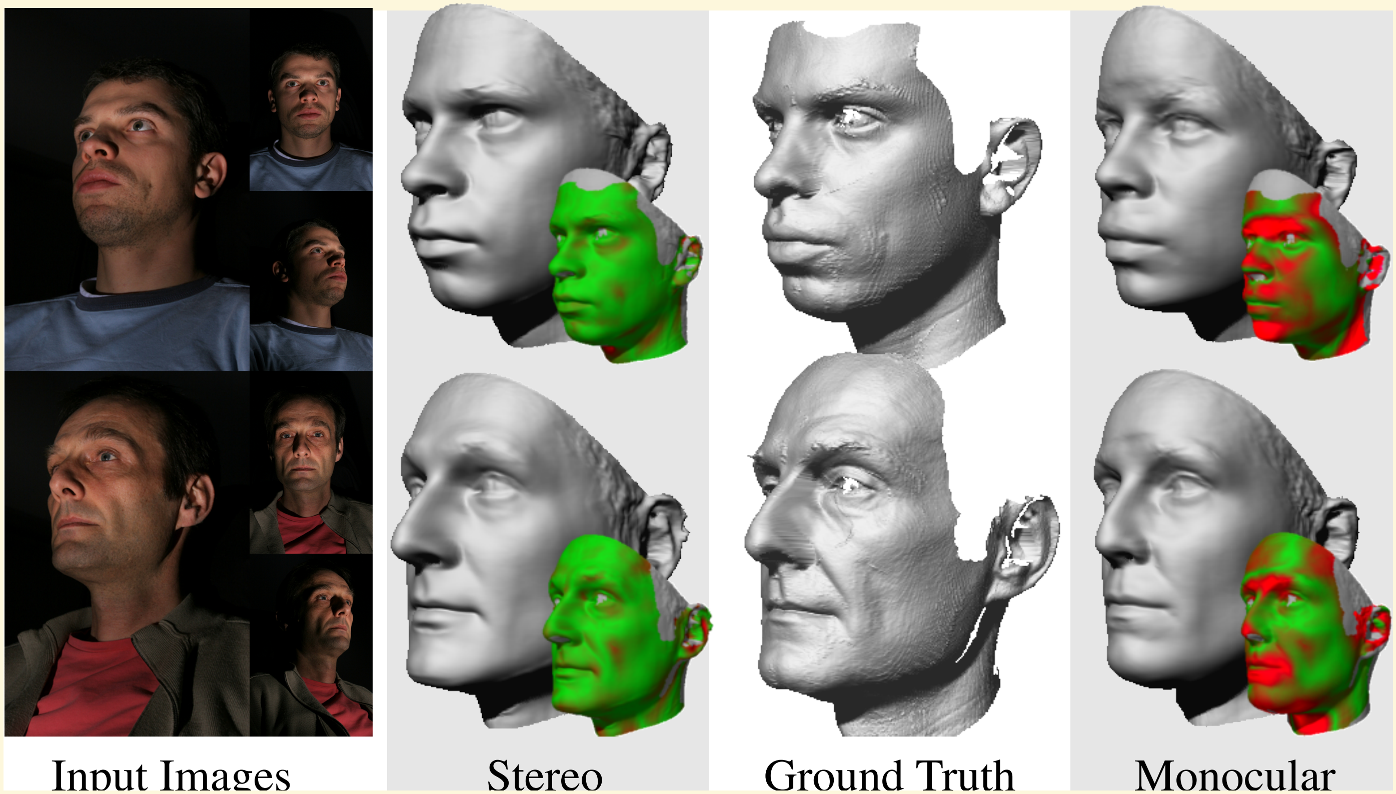
Face modeling

- From one stereo pair to a 3D head model

[Frederic Deverney, INRIA, 1998]

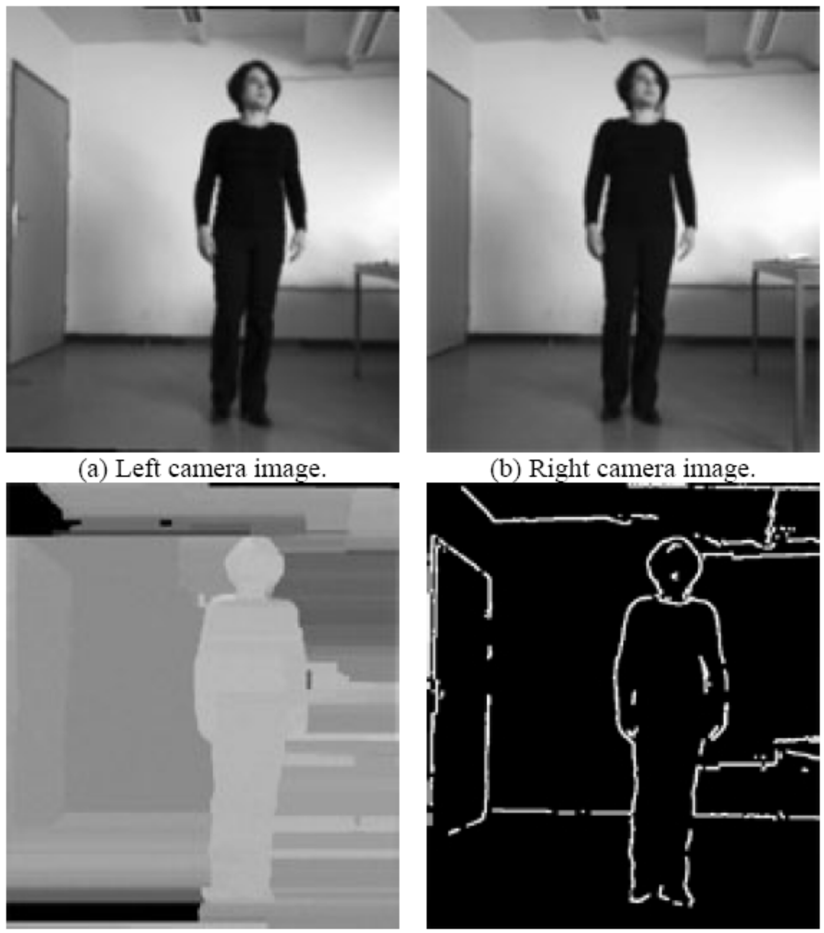
Face modeling

- From multiple stereo pairs to a 3D head model



- [Brian Amberg, Basel, 2005]

Depth for segmentation



(a) Left camera image. (b) Right camera image.

(c) Depth image. (d) Edge combination image.

Figure 3 Stereo video frames with computed depth map and edge combination result.

Edges in disparity in conjunction with image edges enhances contours found

Danijela Markovic and Margrit Gelautz, Interactive Media Systems Group, Vienna University of Technology

Depth for segmentation

(a) Original image with snake initialization.

(b) Final snake on original image.

(c) Final snake on depth image.

(d) Original image with snake from (c) overlaid.

(e) Final snake on edge combination image.

(f) Original image with snake from (e) overlaid.

Danijela Markovic and Margrit Gelautz, Interactive Media Systems Group, Vienna University of Technology