CSCI 497P/597P: Computer Vision

Lecture 25

Epipolar Geometry

Structure From Motion

Multiview Stereo

Announcements

- P3 is out!
- Deadline for grads to opt in for P/NP is Friday
  - Undergrads have until June 5th

Goals

- Understand some of the properties of the fundamental matrix:
  - rank deficiency
  - epipolar lines; epipoles
- Understand the general idea of how Structure From Motion is solved.

(a tiny bit more whiteboard)

8-point algorithm

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

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

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

Each match yields **one** 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

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

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

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

Each match yields equation:

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

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

8-point algorithm

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

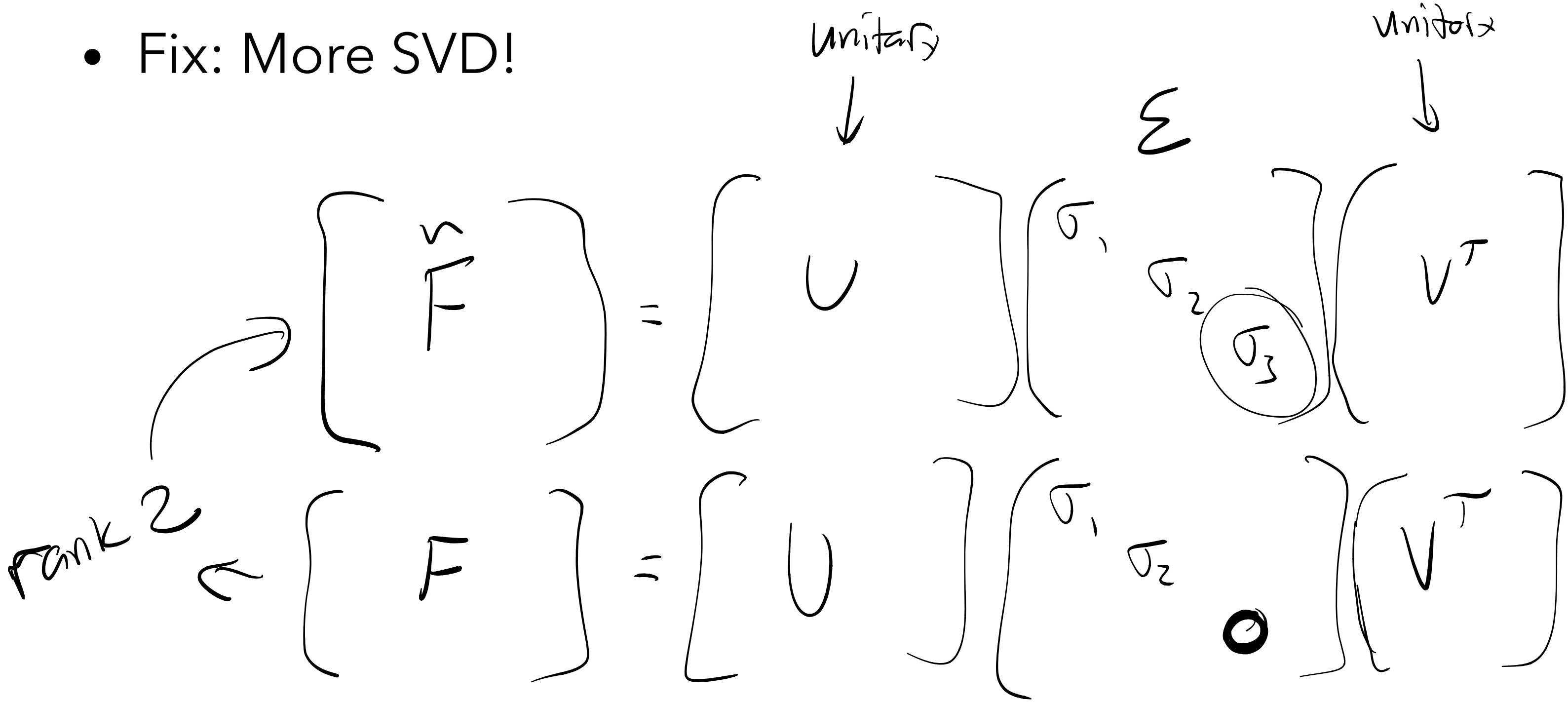
As with homographies, this is has the form Ax = 0

Solve homogeneous system using the SVD.

8-point algorithm: Problem

- **Solution is (generally) not rank 2.**

- Fix: More SVD!



8-point algorithm

- **Solution is (generally) not rank 2.**
- Fix: More SVD!

8-point algorithm: Problem 2

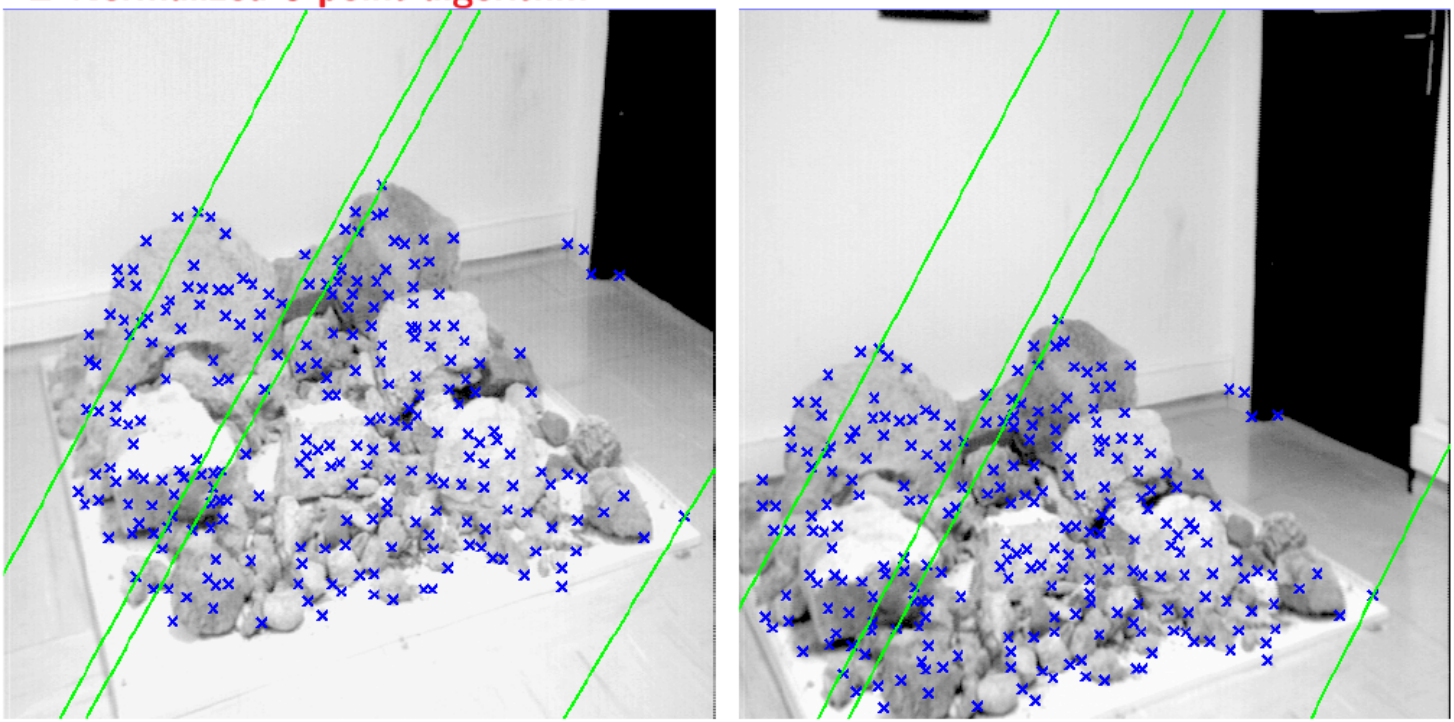
$$\begin{bmatrix} u_1u_1' & v_1u_1' & u_1' & u_1v_1' & v_1v_1' & v_1' & u_1 & v_1 & 1 \\ u_2u_2' & v_2u_2' & u_2' & u_2v_2' & v_2v_2' & v_2' & u_2 & v_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_n' & v_nu_n' & u_n' & u_nv_n' & v_nv_n' & v_n' & u_n &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

~10000 ~10000 ~100 ~10000 ~10000 ~100 ~100 ~100 1

! Orders of magnitude difference between column of data matrix $\rightarrow$ least-squares yields poor results

Fix: scale image positions to the range [0,1], solve, then scale back.

8-point algorithm: Results



What about more than 2 views?

- 2 views: fundamental matrix
- 3 views: trifocal tensor
- 4 views: *quadrifocal* tensor
- more views: ¯_(ツ)_/¯ (it gets complicated…)

Large-scale structure from motion

- https://www.youtube.com/watch?v=sQegEro5Bfo

Structure from Motion

- Given many photos, reconstruct:
  - positions of the cameras
  - positions of 3D points

Chicken/Egg

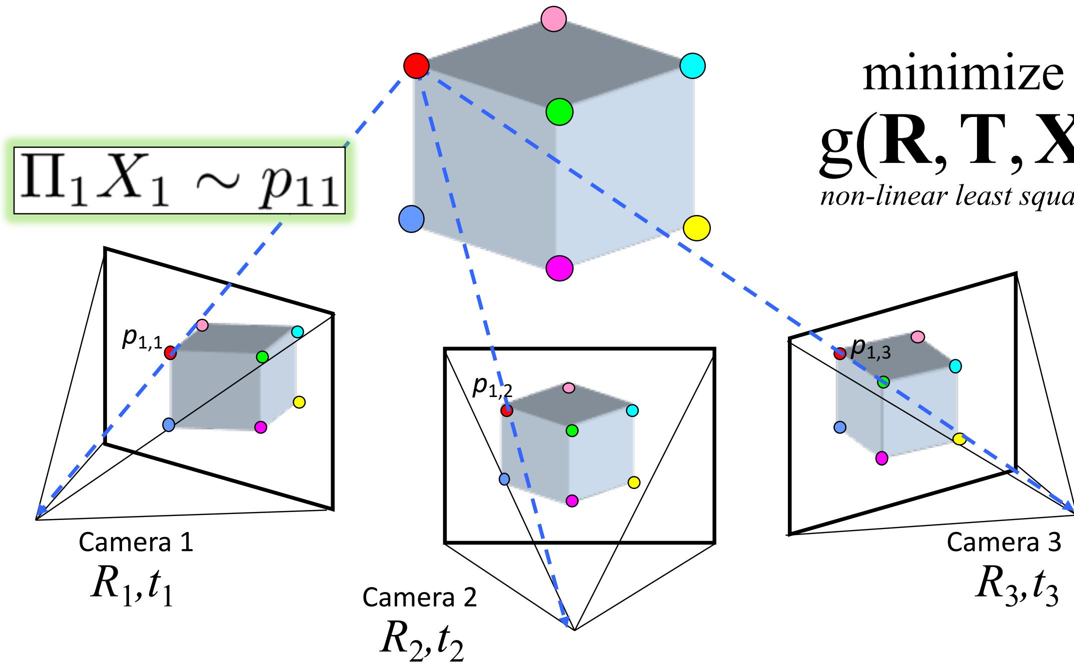
- Step 1: solve for relative pose of pairs (or triples) of cameras using correspondences from feature matching.

- Step 2: alternate between solving:

  - given camera positions, solve for point locations

  - given point locations, solve for camera positions

Structure From Motion



minimize

$$\mathbf{g(R, T, X)}$$

non-linear least squares

$$g(\mathbf{X}, \mathbf{R}, \mathbf{T}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \underbrace{w_{ij}} \cdot \left\| \underbrace{\mathbf{P}(\mathbf{x}_i, \mathbf{R}_j, \mathbf{t}_j)} - \underbrace{\begin{bmatrix} u_{i,j} \\ v_{i,j} \end{bmatrix}} \right\|^2$$

predicted image location

observed image location

indicator variable: is point *i* visible in image *j* ?

Applications

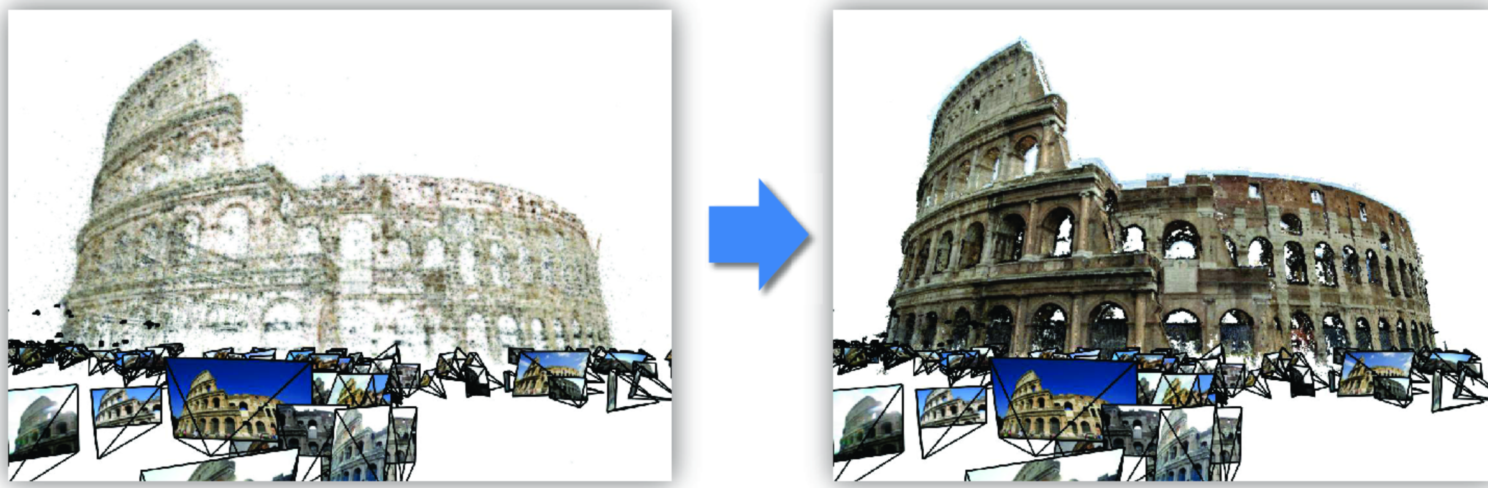
- Hyperlapse https://www.youtube.com/watch?v=SOpwHaQnRSY

- SLAM: https://medium.com/scape-technologies/building-the-ar-cloud-part-three-3d-maps-the-digital-scaffolding-of-the-21st-century-465fa55782dd

- Graphics, movies, games, self-driving cars, robots, ...

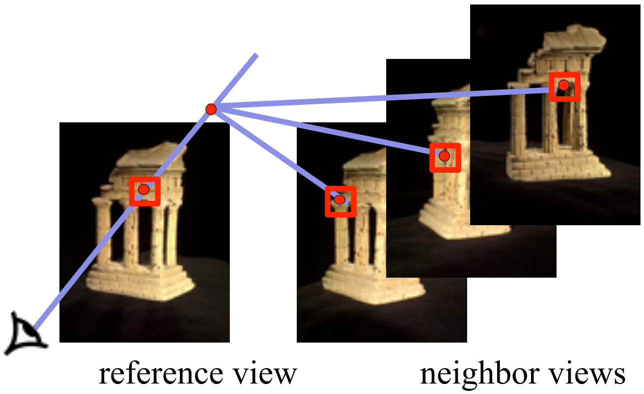
Multiview Stereo

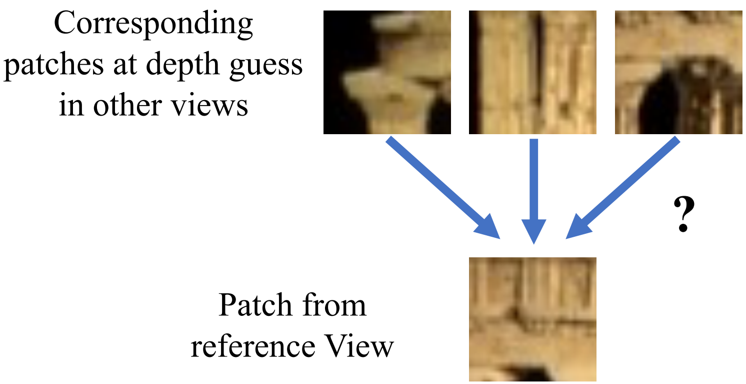
- Once you've solved for all those camera positions, how good a 3D model can you create?



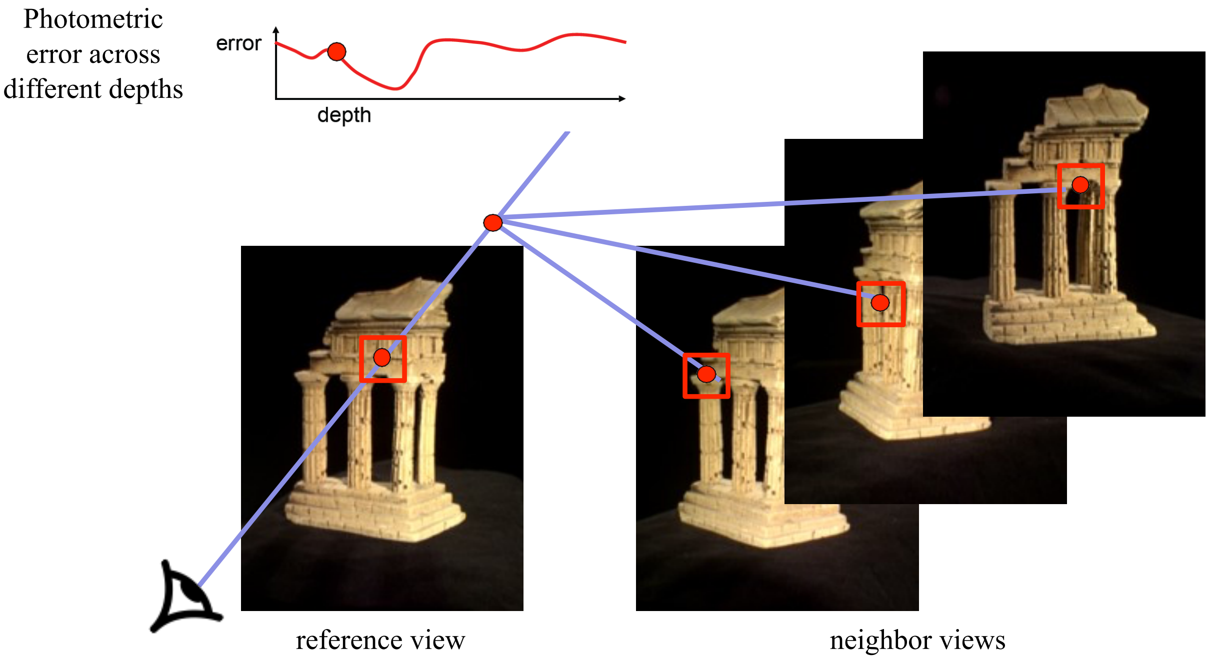
Multiview Stereo: Basic Idea

Evaluate the likelihood of geometry at a particular depth for a particular reference patch:

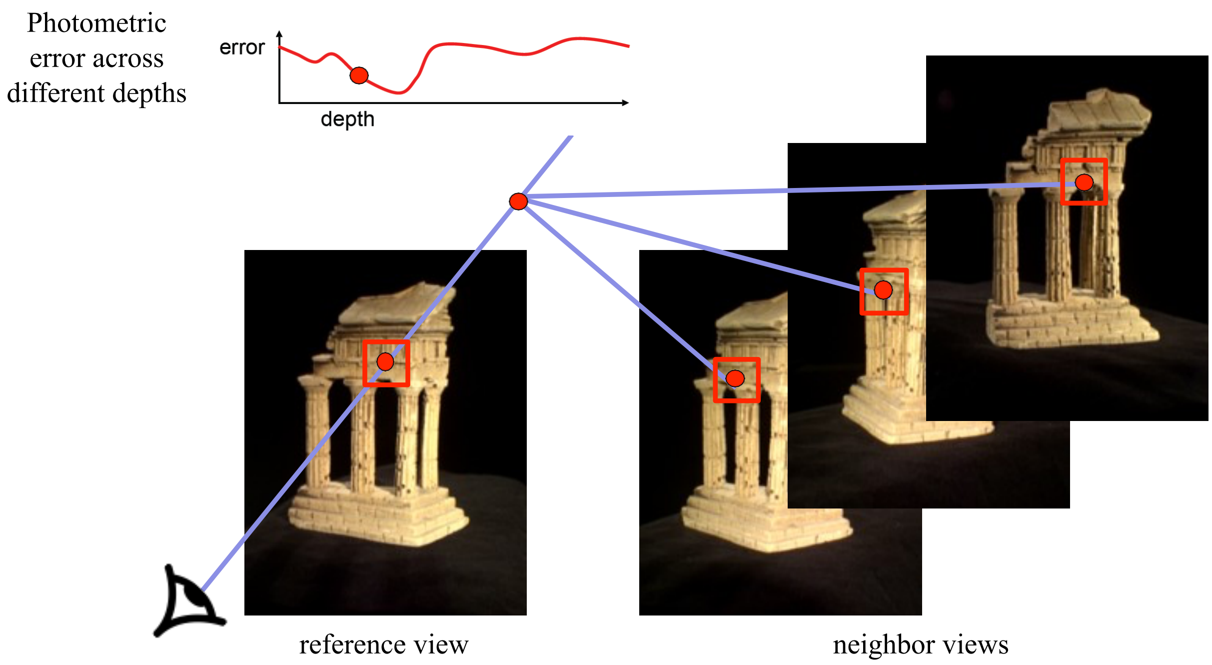




Source: Y. Furukawa

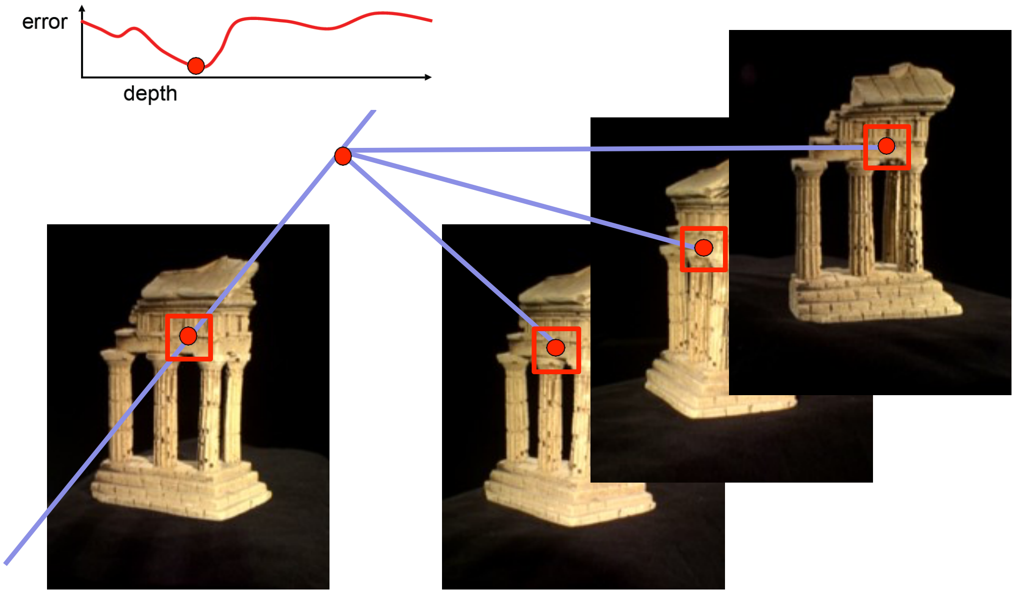


Source: Y. Furukawa



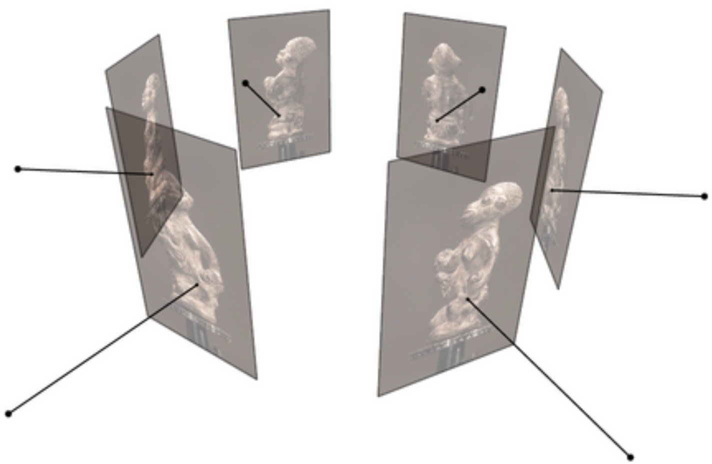
Source: Y. Furukawa

Photometric
error across
different depths



Depth map fusion

- Compute depth maps for multiple cameras, then fuse them into a 3D model



Figures by Carlos Hernandez

Result

- https://www.youtube.com/watch?v=N6Douyfa7l8
-