Computer Vision & Image Processing
CSE 473 / 573

Instructor - Kevin R. Keane, PhD

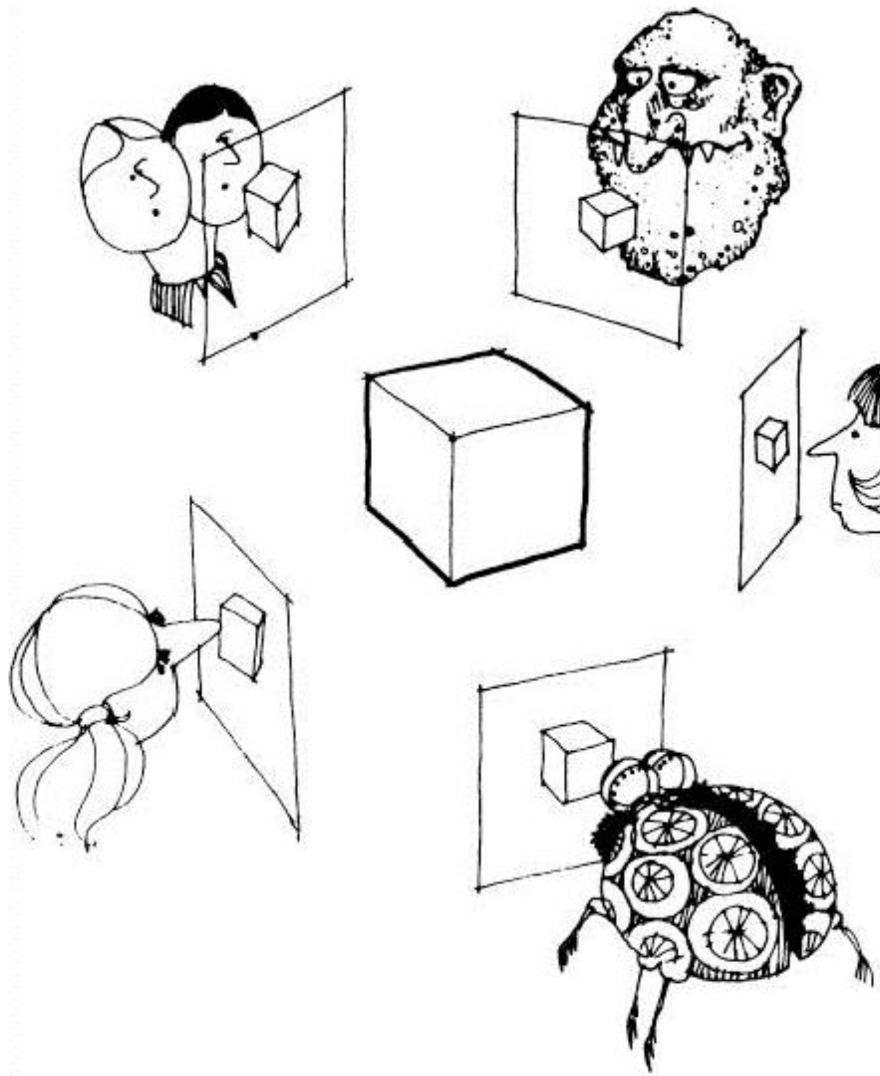
TAs - Radhakrishna Dasari, Yuhao Du, Niyazi Sorkunlu

Lecture 18

October 11, 2017

Multi-view stereo

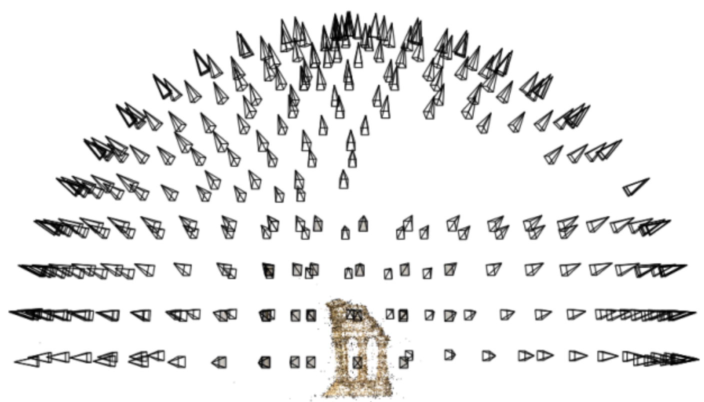
Multi-view stereo



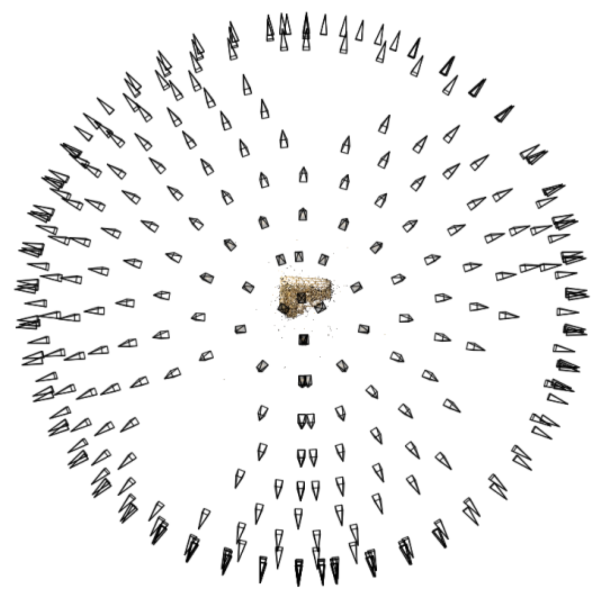
Many slides adapted from S. Seitz

Multi-view stereo

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape



Reconstruction (side)



(top)

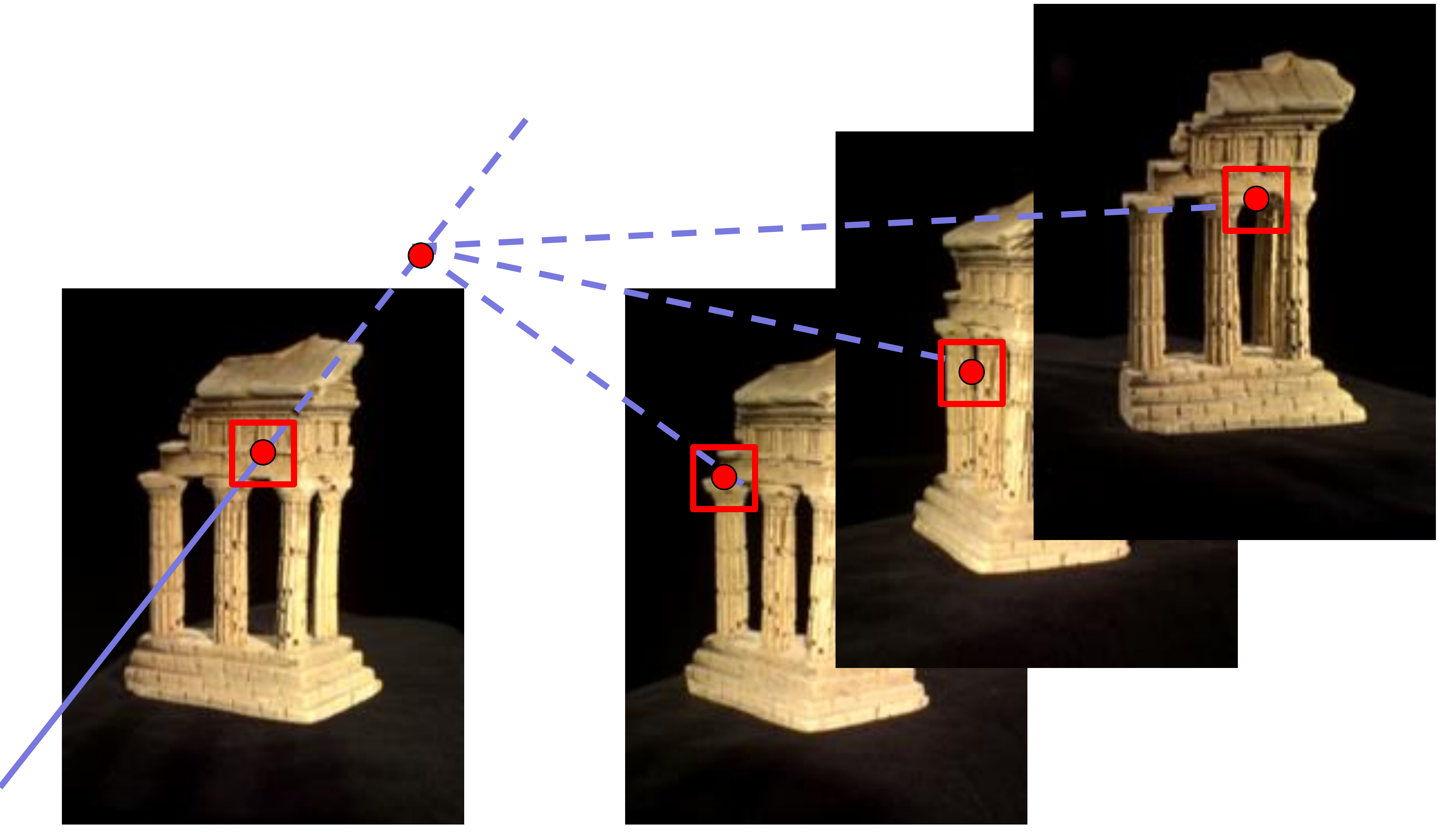
Multi-view stereo

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape
- "Images of the same object or scene"
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (special rig, camera network or video sequence)
  - Calibration may be known or unknown

Multi-view stereo

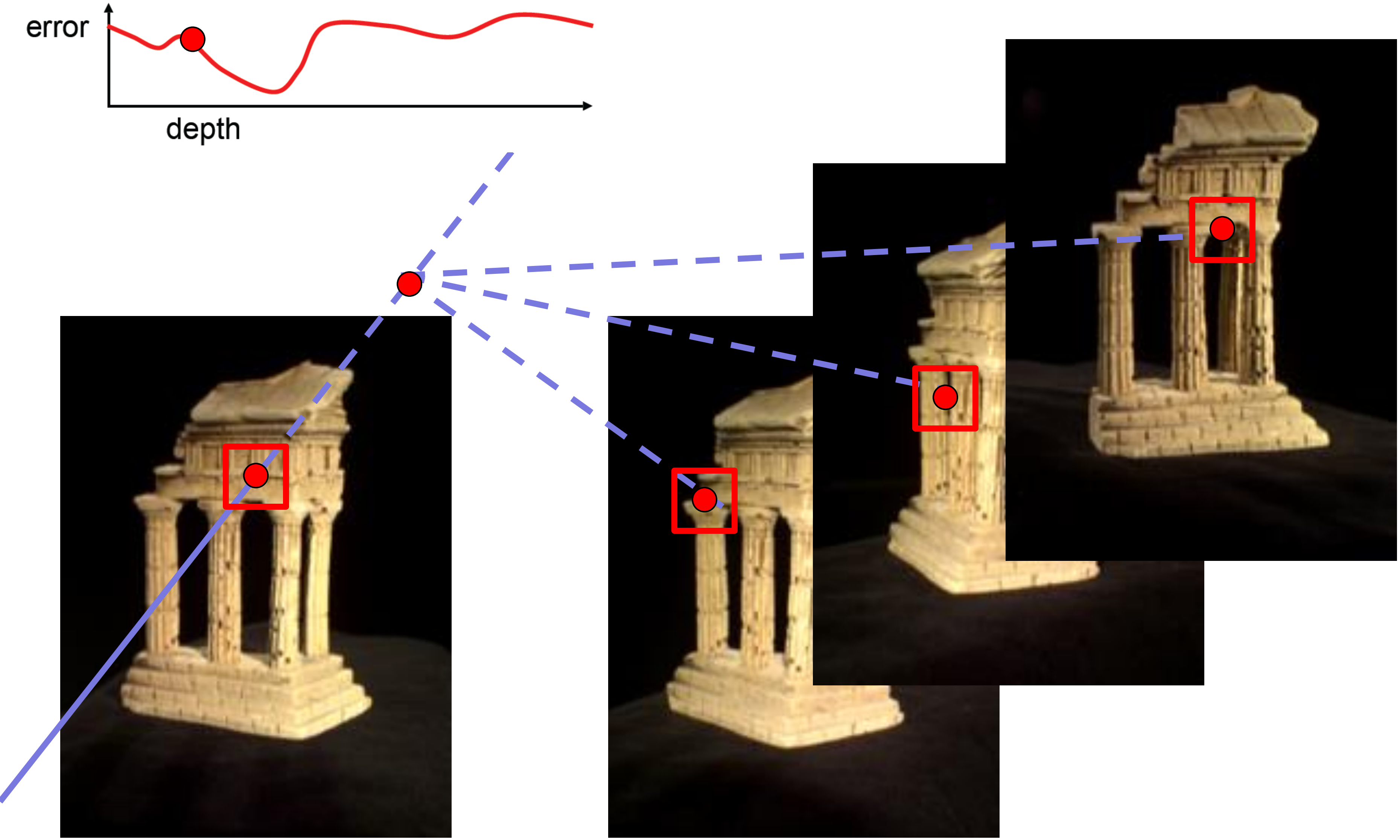
- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape
- “Images of the same object or scene”
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (special rig, camera network or video sequence)
  - Calibration may be known or unknown
- “Representation of 3D shape”
  - Depth maps
  - Meshes
  - Point clouds
  - Patch clouds
  - Volumetric models
  - ….

Multi-view stereo: Basic idea



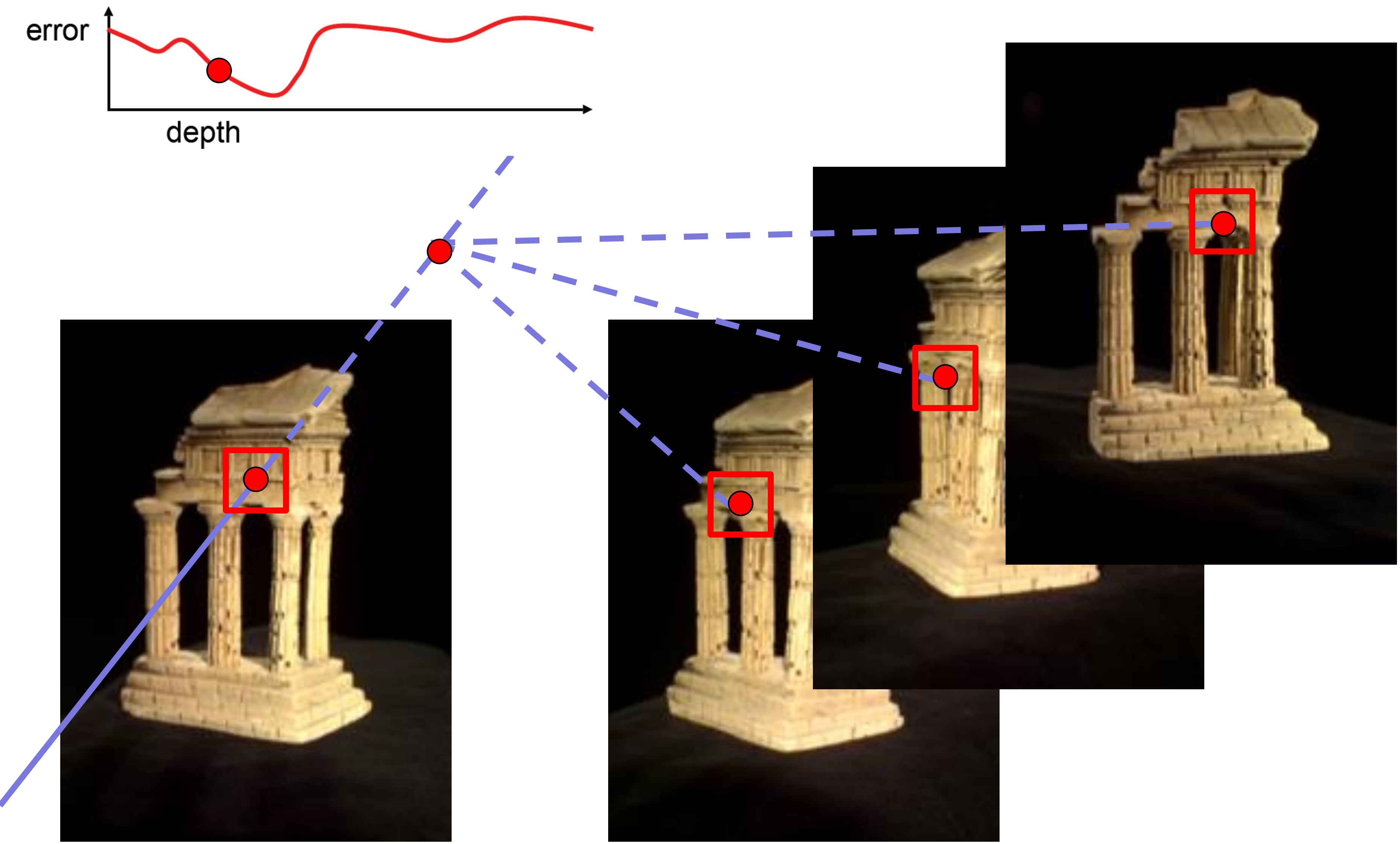
Source: Y. Furukawa

Multi-view stereo: Basic idea



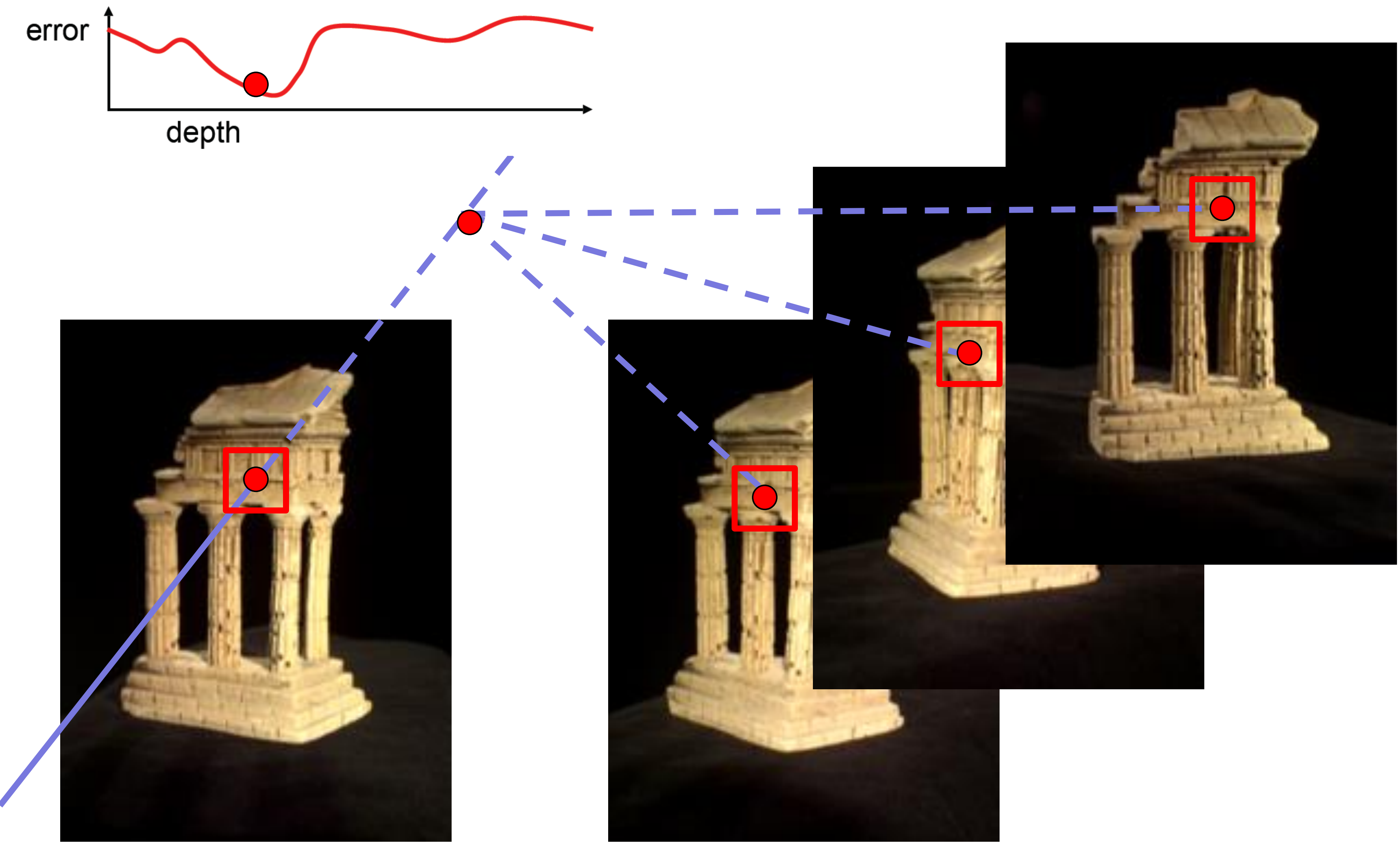
Source: Y. Furukawa

Multi-view stereo: Basic idea



Source: Y. Furukawa

Multi-view stereo: Basic idea

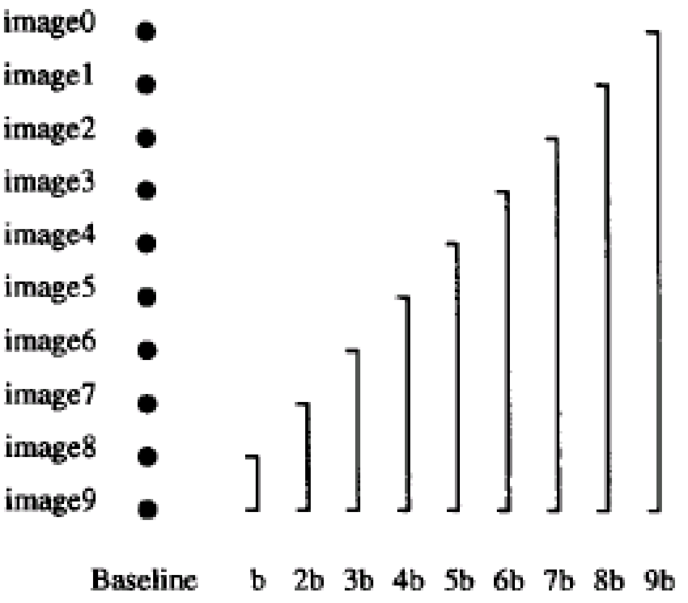


Source: Y. Furukawa

Multiple-baseline stereo

- Pick a reference image, and slide the corresponding window along the corresponding epipolar lines of all other images, using inverse depth relative to the first image as the search parameter

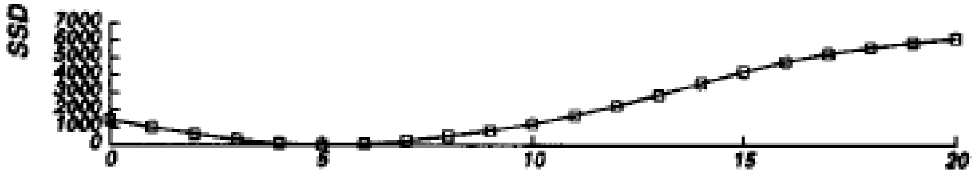
Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.



M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo System,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4):353-363 (1993).

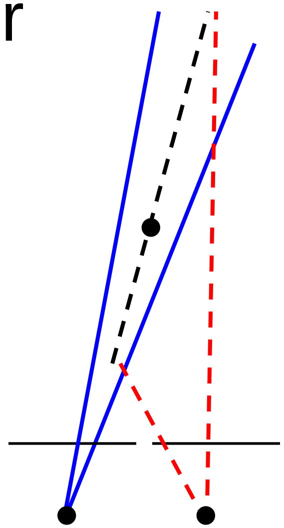
Multiple-baseline stereo

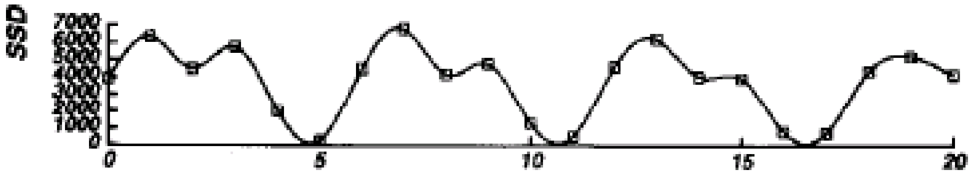
- For larger baselines, must search larger area in second image



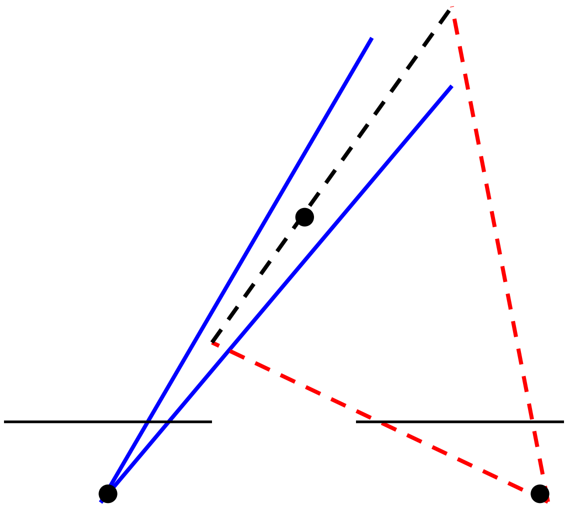
1/z

pixel matching score





1/z



Multiple-baseline stereo

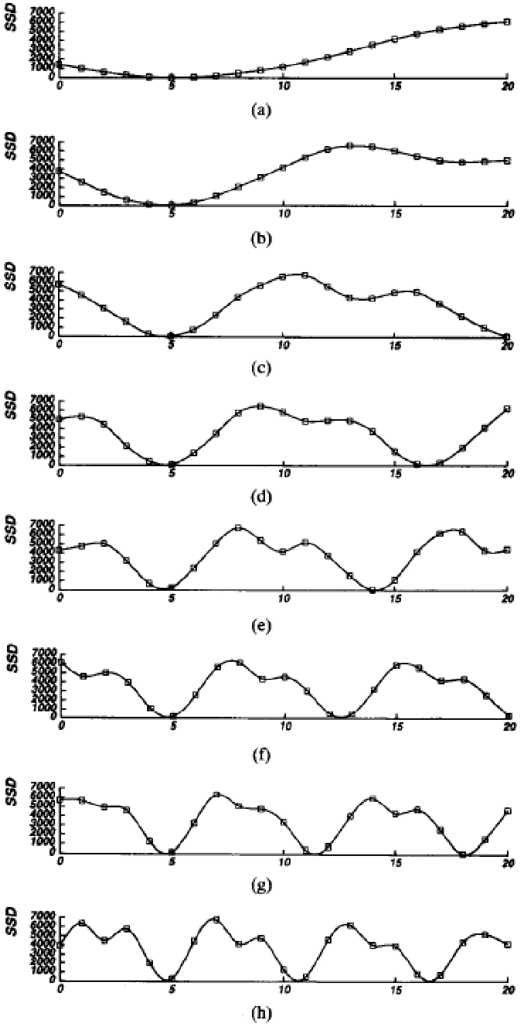


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

Use the sum of SSD scores to rank matches

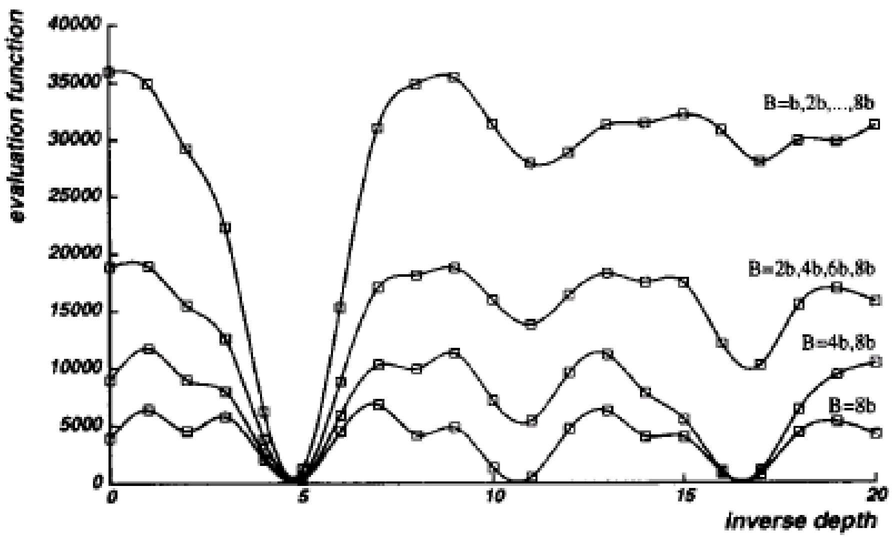


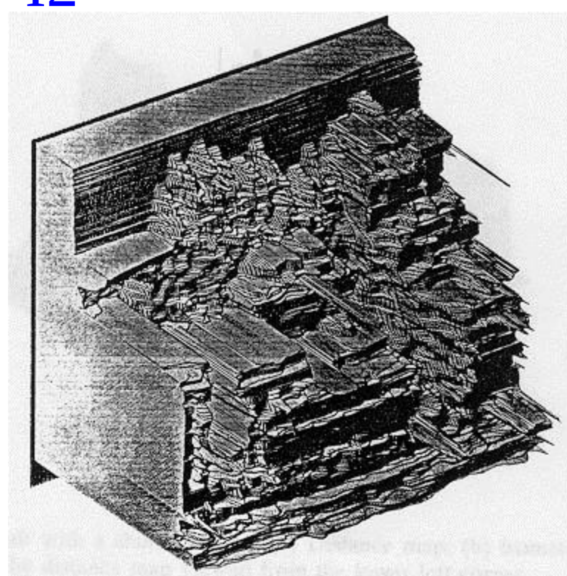
Fig. 7. Combining multiple baseline stereo pairs.

Multiple-baseline stereo results

I1

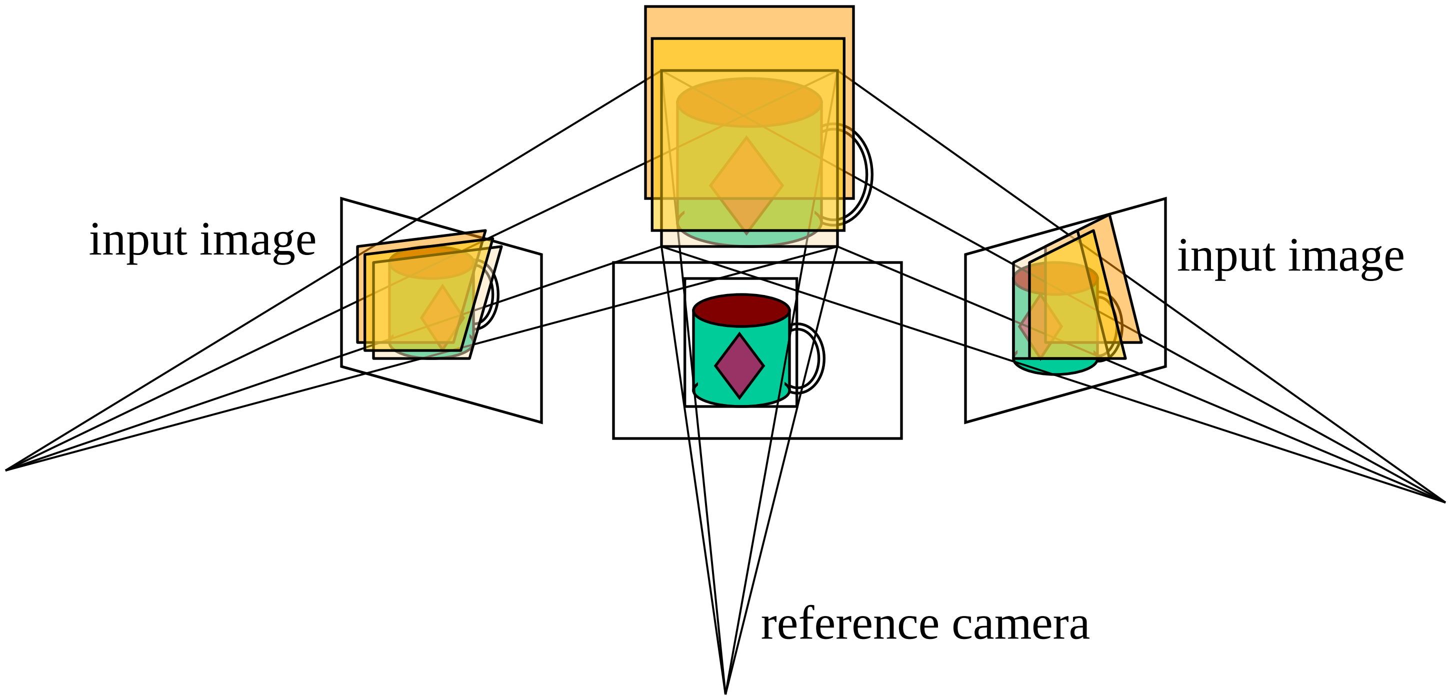
I2

I10



M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo System,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4):353-363 (1993).

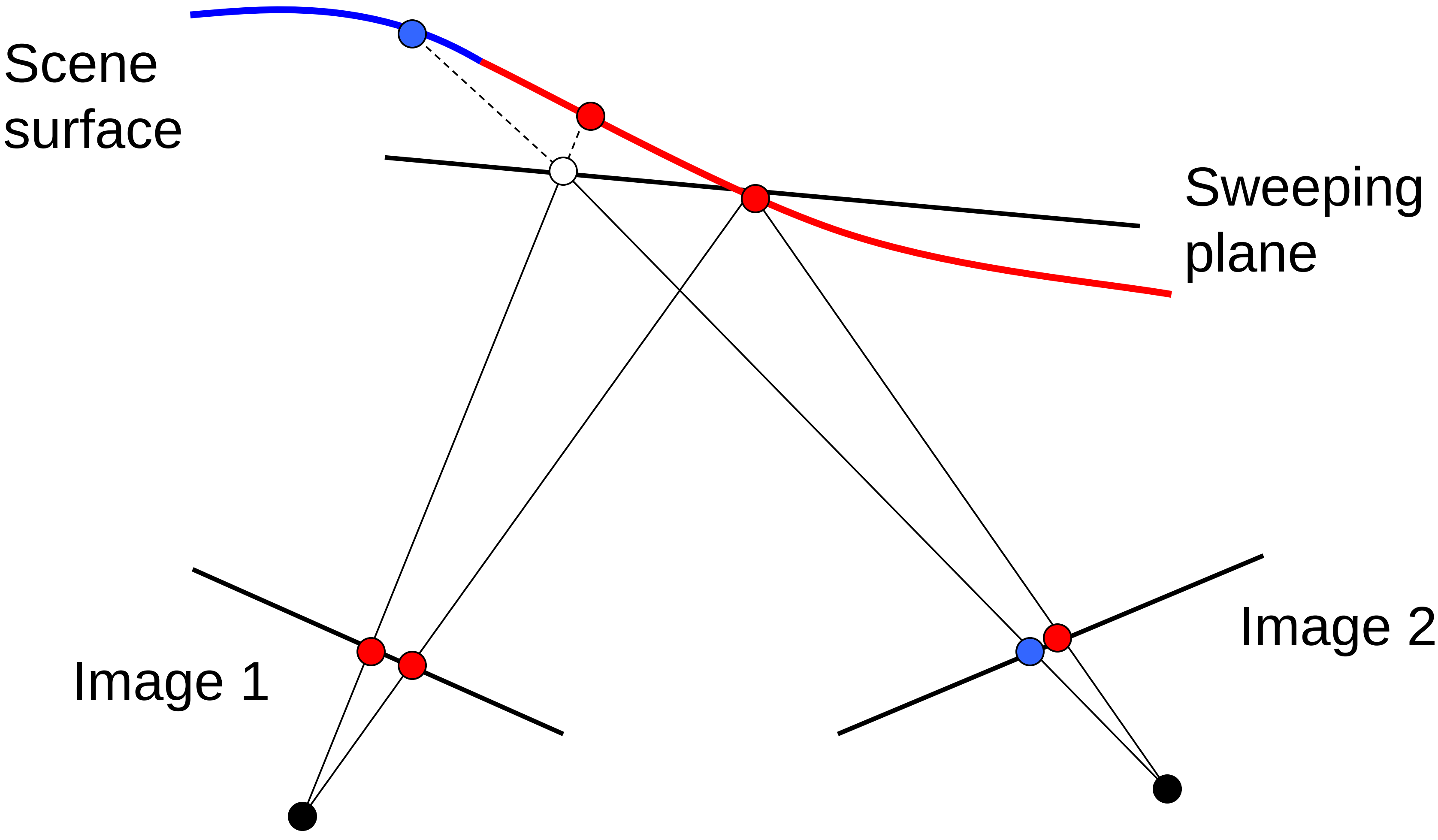
Plane Sweep Stereo



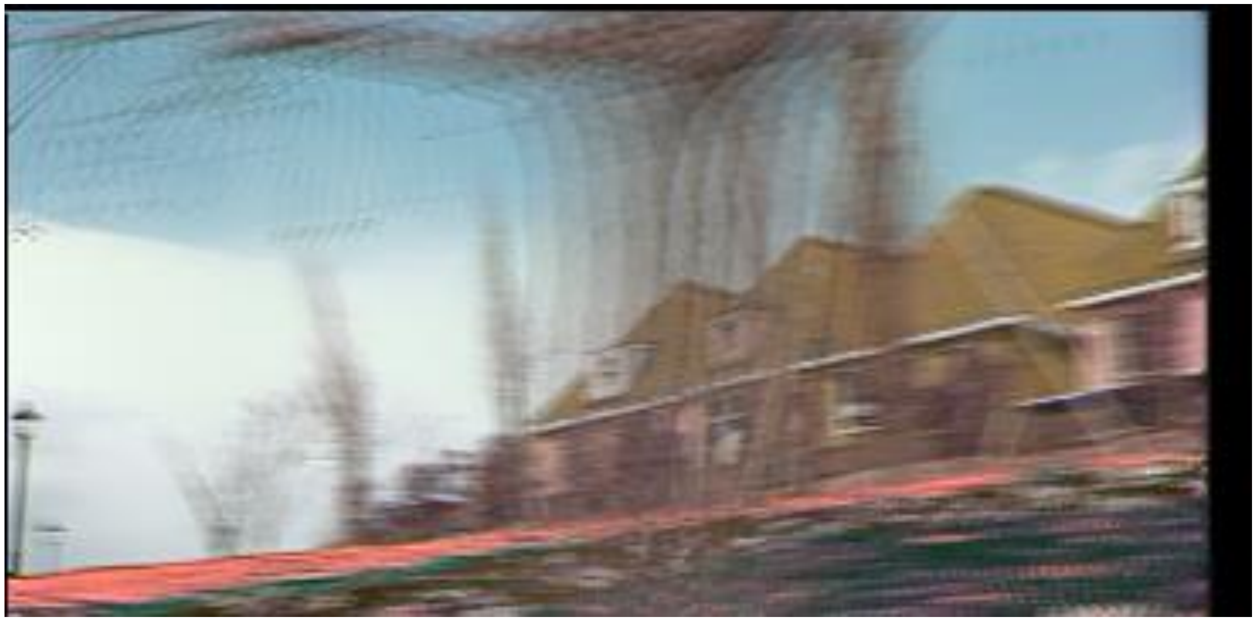
- Sweep family of planes at different depths w.r.t. a reference camera
- For each depth, project each input image onto that plane
- This is equivalent to a homography warping each input image into the reference view
- What can we say about the scene points that are at the right depth?

R. Collins. [A space-sweep approach to true multi-image matching.]() CVPR 1996.

Plane Sweep Stereo



Plane Sweep Stereo



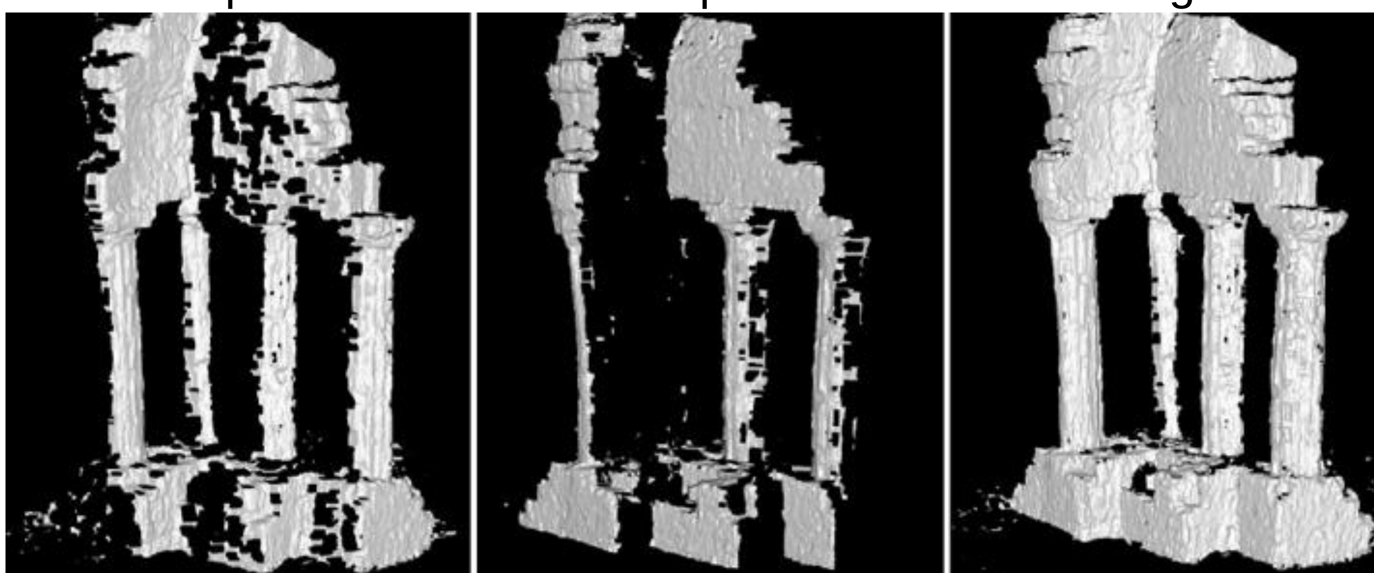
- For each depth plane
  - For each pixel in the composite image stack, compute the variance
- For each pixel, select the depth that gives the lowest variance

- Can be accelerated using graphics hardware

R. Yang and M. Pollefeys. *Multi-Resolution Real-Time Stereo on Commodity Graphics Hardware*, CVPR 2003

Merging depth maps

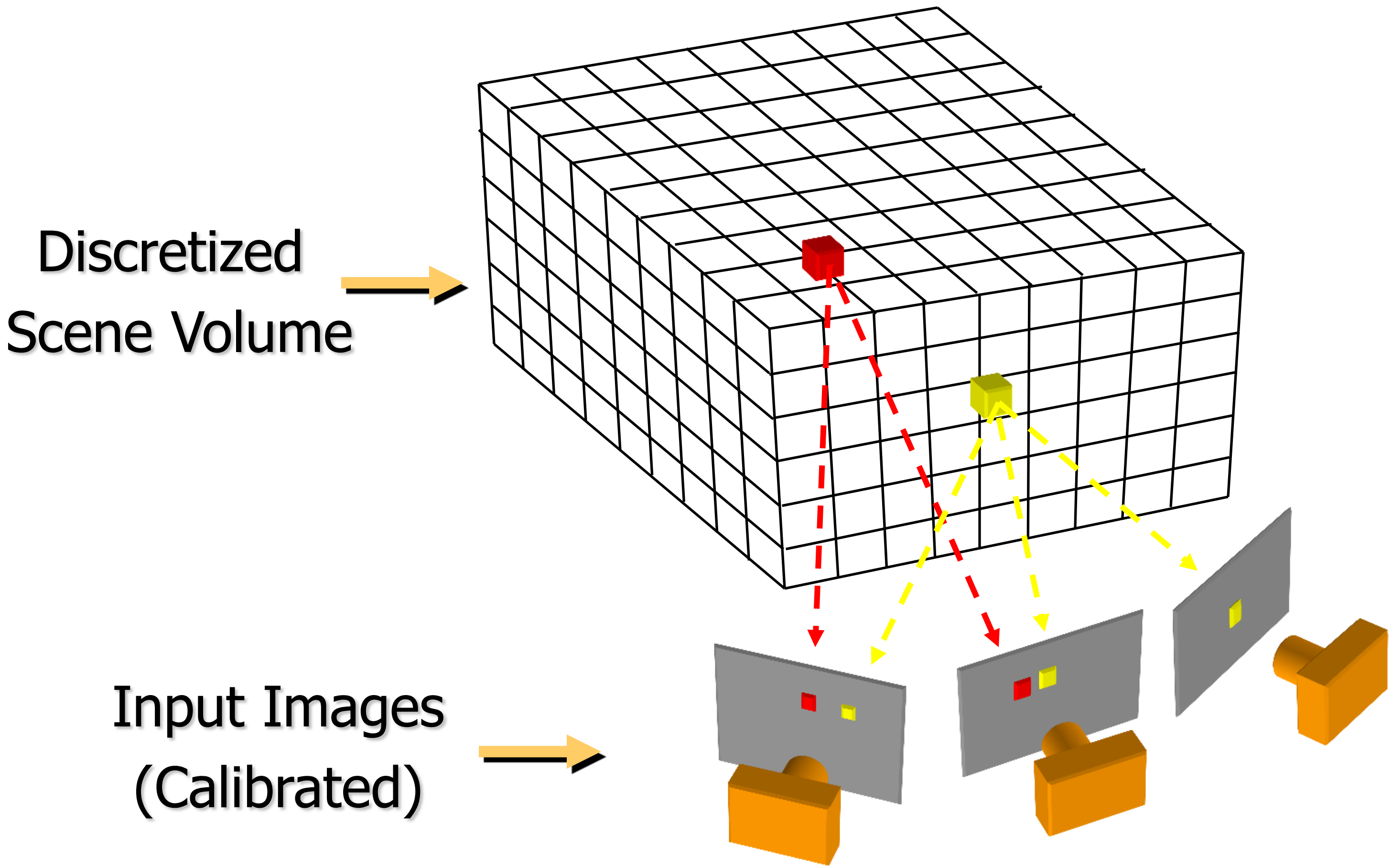
- Given a group of images, choose each one as reference and compute a depth map w.r.t. that view using a multi-baseline approach
- Merge multiple depth maps to a volume or a mesh (see, e.g., Curless and Levoy 96)



Volumetric stereo

- In plane sweep stereo, the sampling of the scene depends on the reference view
- We can use a voxel volume to get a view-independent representation

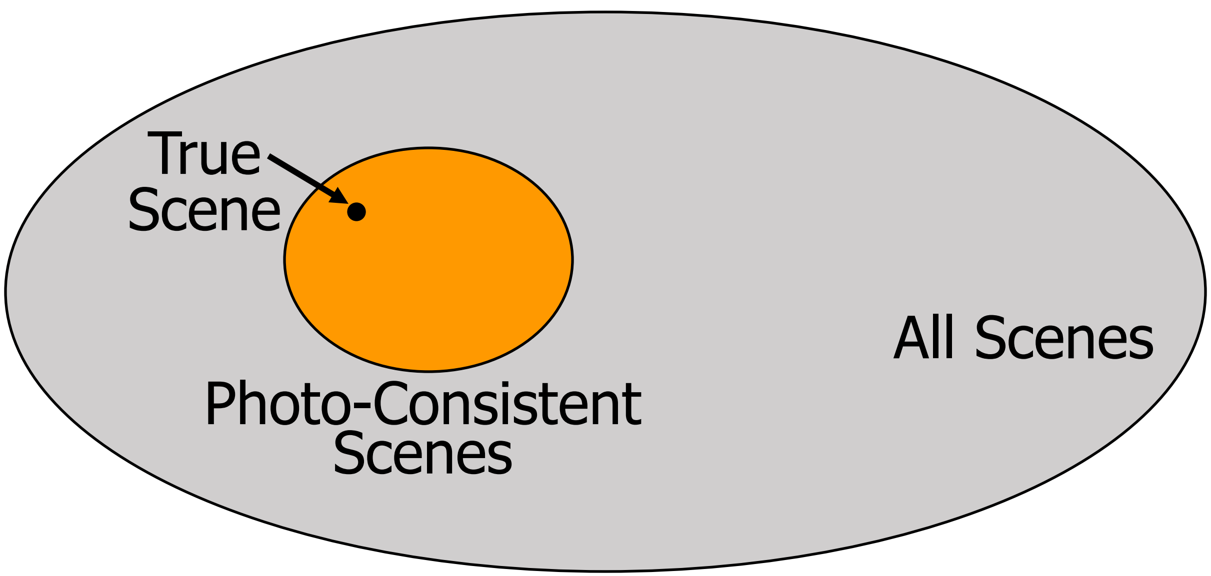
Volumetric stereo



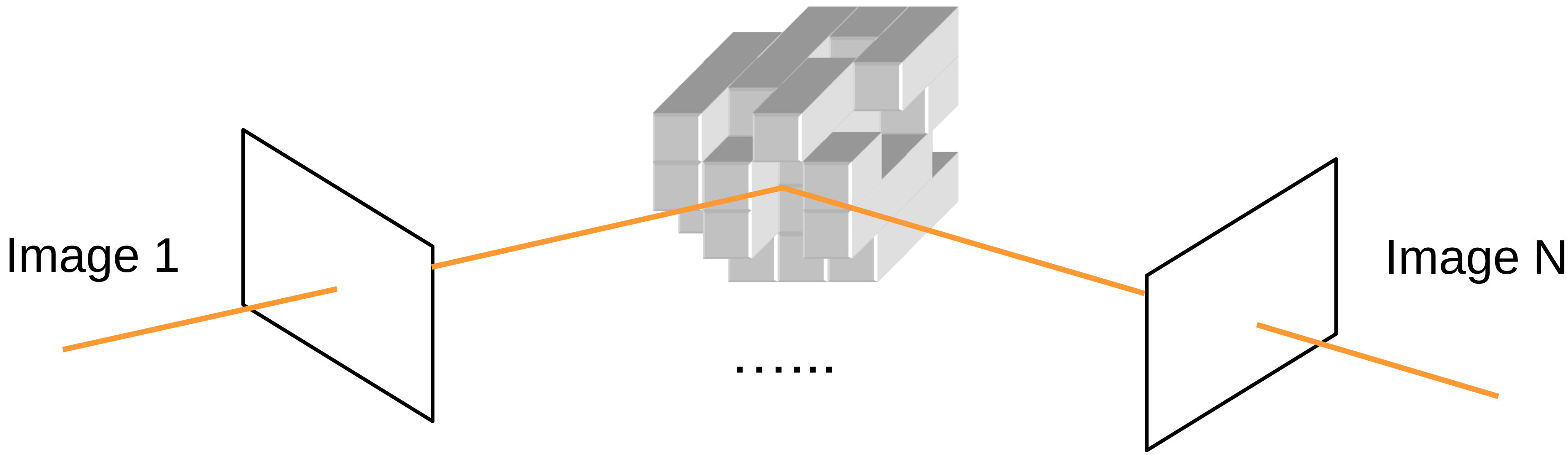
Goal: **Assign RGB values to voxels in V**
photo-consistent **with images**

Photo-consistency

- A *photo-consistent scene* is a scene that exactly reproduces your input images from the same camera viewpoints
- You can't use your input cameras and images to tell the difference between a photo-consistent scene and the true scene



Space Carving



Space Carving Algorithm

- Initialize to a volume V containing the true scene
- Choose a voxel on the outside of the volume
- Project to visible input images
- Carve if not photo-consistent
- Repeat until convergence

K. N. Kutulakos and S. M. Seitz, **A Theory of Shape by Space Carving**, *ICCV* 1999

Space Carving Results: African Violet

Input Image (1 of 45)

Reconstruction

Reconstruction

Reconstruction

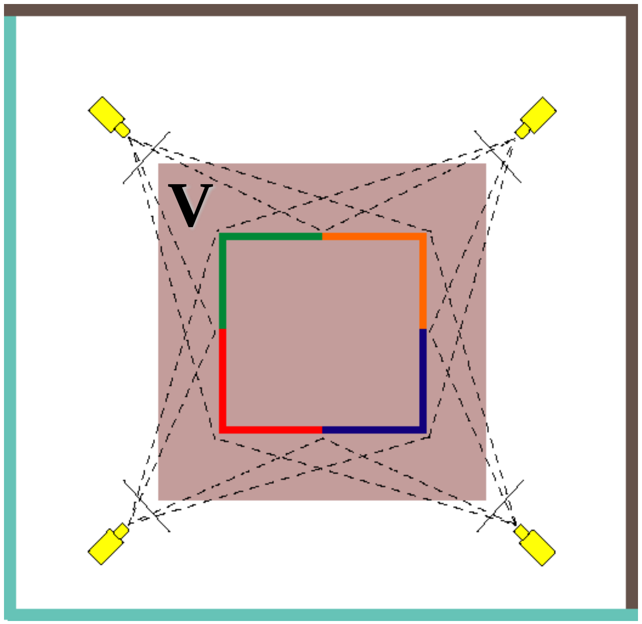
Source: S. Seitz

Space Carving Results: Hand

Input Image
(1 of 100)

Views of Reconstruction

Which shape do you get?



True Scene

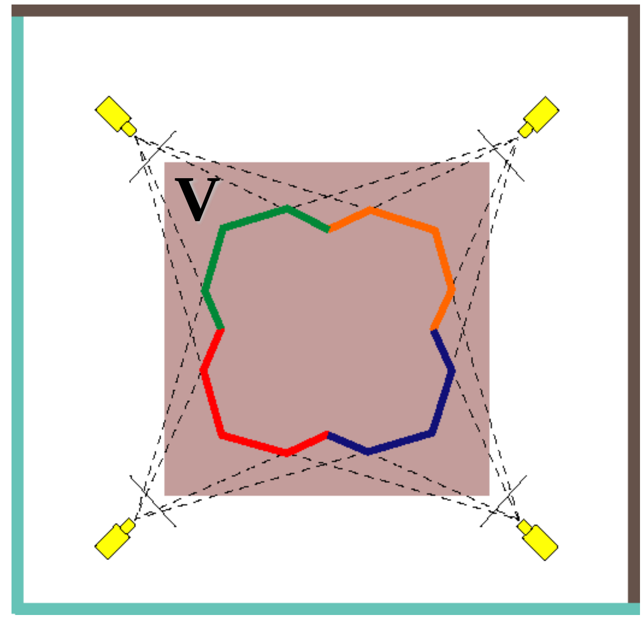


Photo Hull

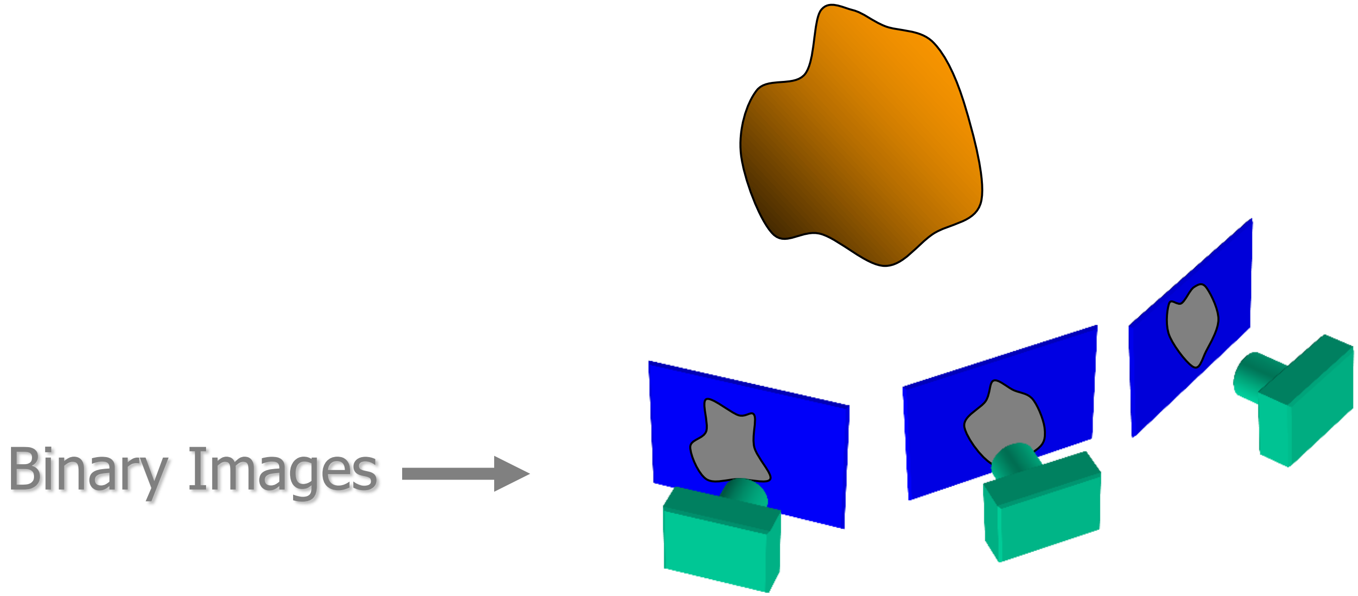
The Photo Hull *is the UNION of all photo-consistent scenes in V*

- It is a photo-consistent scene reconstruction
- Tightest possible bound on the true scene

Source: S. Seitz

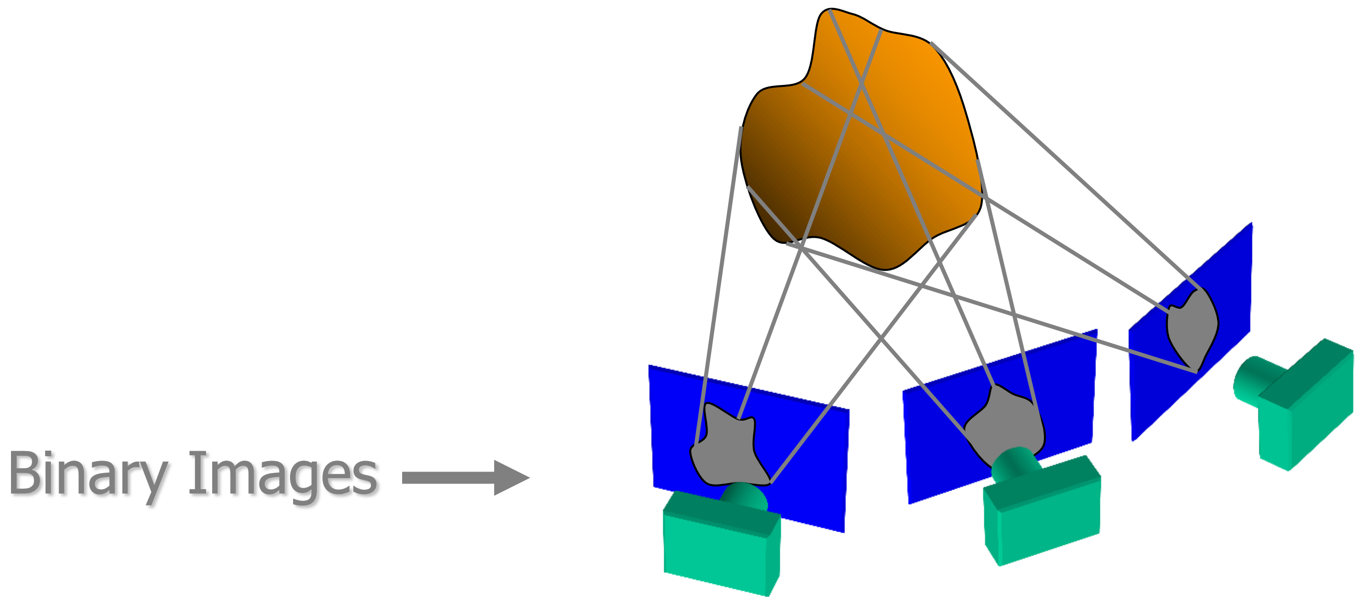
Reconstruction from Silhouettes

- The case of binary images: a voxel is photo-consistent if it lies inside the object's silhouette in all views



Reconstruction from Silhouettes

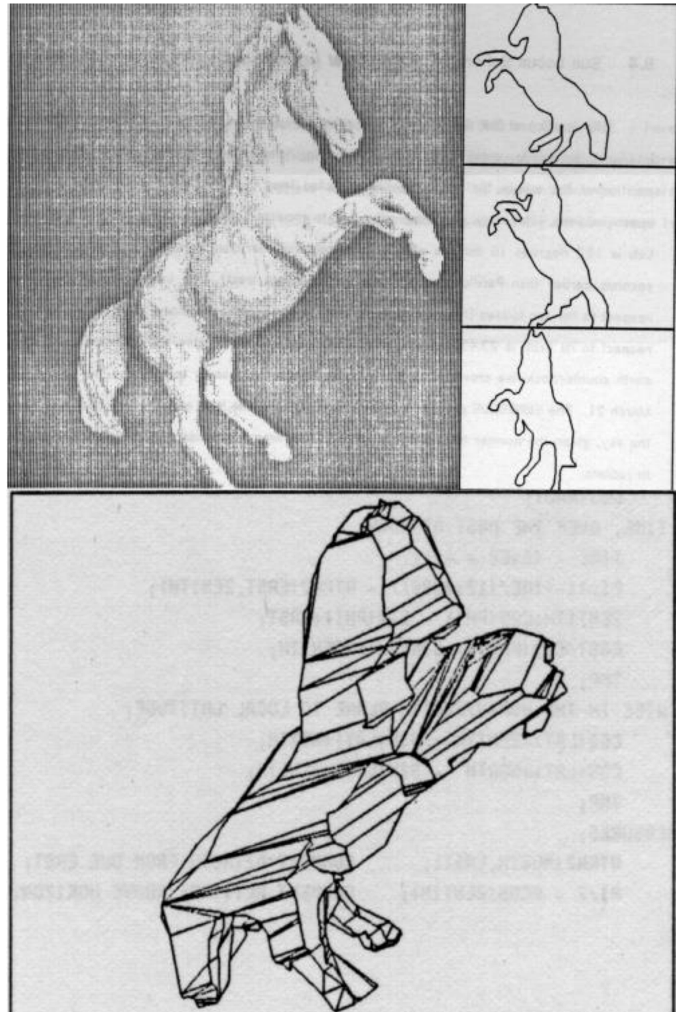
- The case of binary images: a voxel is photo-consistent if it lies inside the object’s silhouette in all views



Finding the silhouette-consistent shape (*visual hull*):

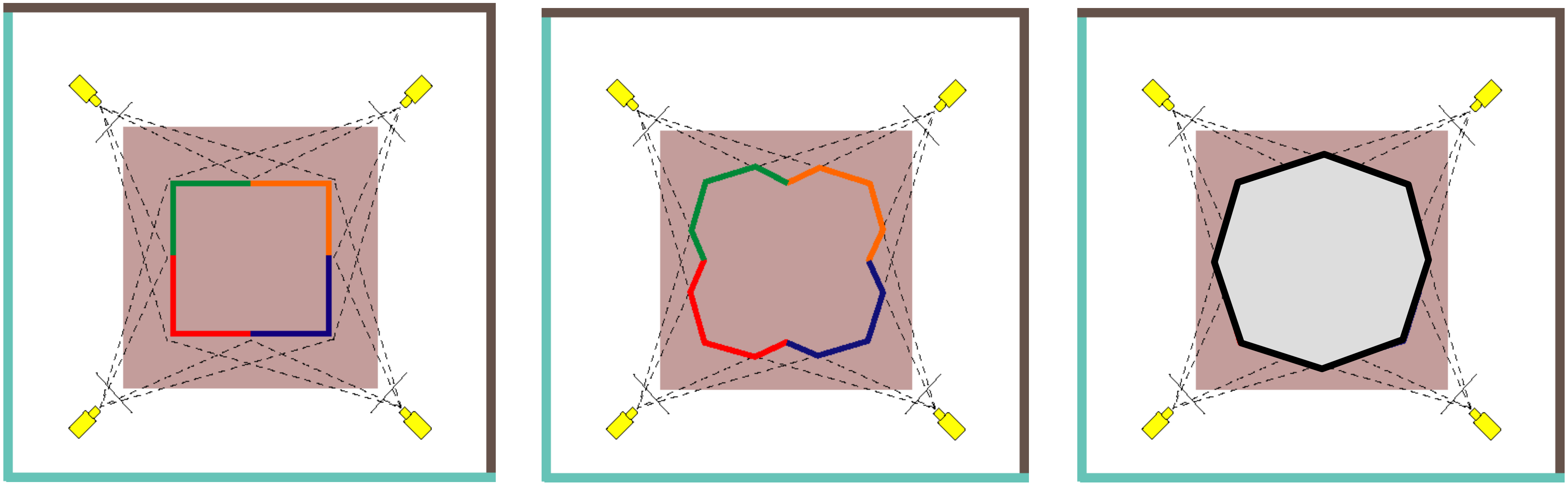
- *Backproject* each silhouette
- Intersect backprojected volumes

Volume intersection



B. Baumgart, *Geometric Modeling for Computer Vision*, Stanford Artificial Intelligence Laboratory, Memo no. AIM-249, Stanford University, October 1974.

Photo-consistency vs. silhouette-consistency



True Scene **Photo Hull** **Visual Hull**

Carved visual hulls

- The visual hull is a good starting point for optimizing photo-consistency
  - Easy to compute
  - Tight outer boundary of the object
  - Parts of the visual hull (rims) already lie on the surface and are already photo-consistent

Yasutaka Furukawa and Jean Ponce, **Carved Visual Hulls for Image-Based Modeling**, ECCV 2006.

Carved visual hulls

1. Compute visual hull
2. Use dynamic programming to find rims (photo-consistent parts of visual hull)
3. Carve the visual hull to optimize photo-consistency keeping the rims fixed

Yasutaka Furukawa and Jean Ponce, **Carved Visual Hulls for Image-Based Modeling**, ECCV 2006.

From feature matching to dense stereo

1. Extract features
2. Get a sparse set of initial matches
3. Iteratively expand matches to nearby locations
4. Use visibility constraints to filter out false matches
5. Perform surface reconstruction

Yasutaka Furukawa and Jean Ponce, **Accurate, Dense, and Robust Multi-View Stereopsis**, CVPR 2007.

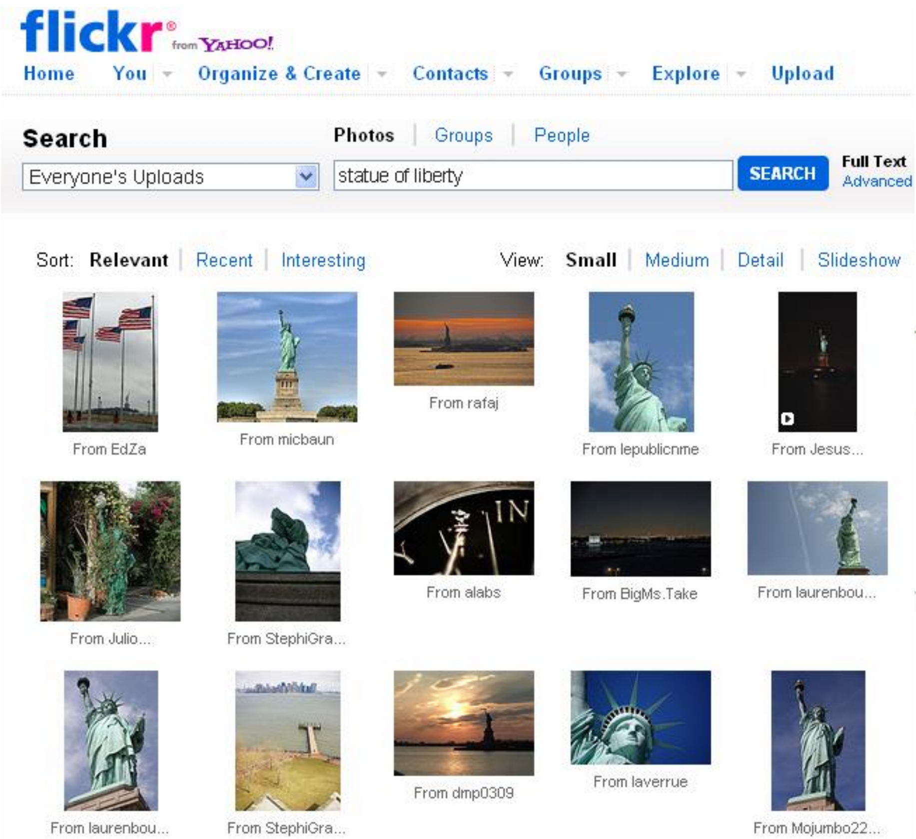
From feature matching to dense stereo

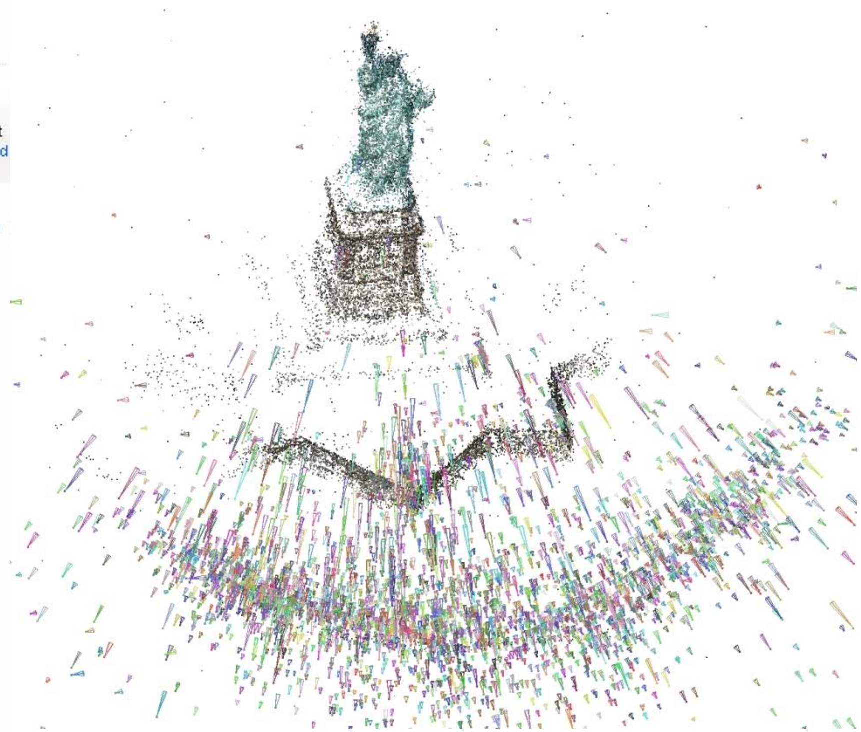
http://www.cs.washington.edu/homes/furukawa/gallery/

Yasutaka Furukawa and Jean Ponce, **Accurate, Dense, and Robust Multi-View Stereopsis**, CVPR 2007.

Stereo from community photo collections

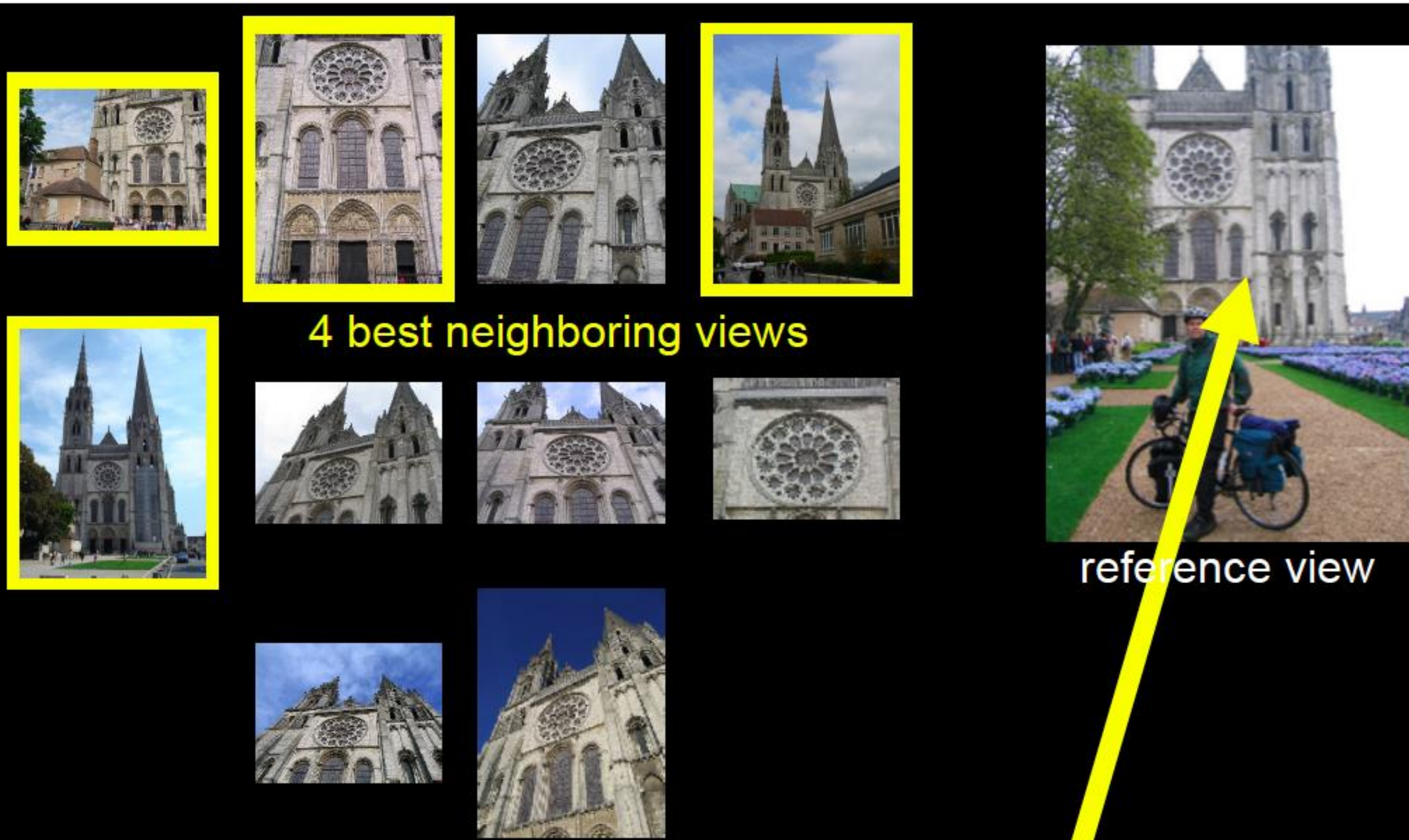
- Need *structure from motion* to recover unknown camera parameters
- Need *view selection* to find good groups of images on which to run dense stereo





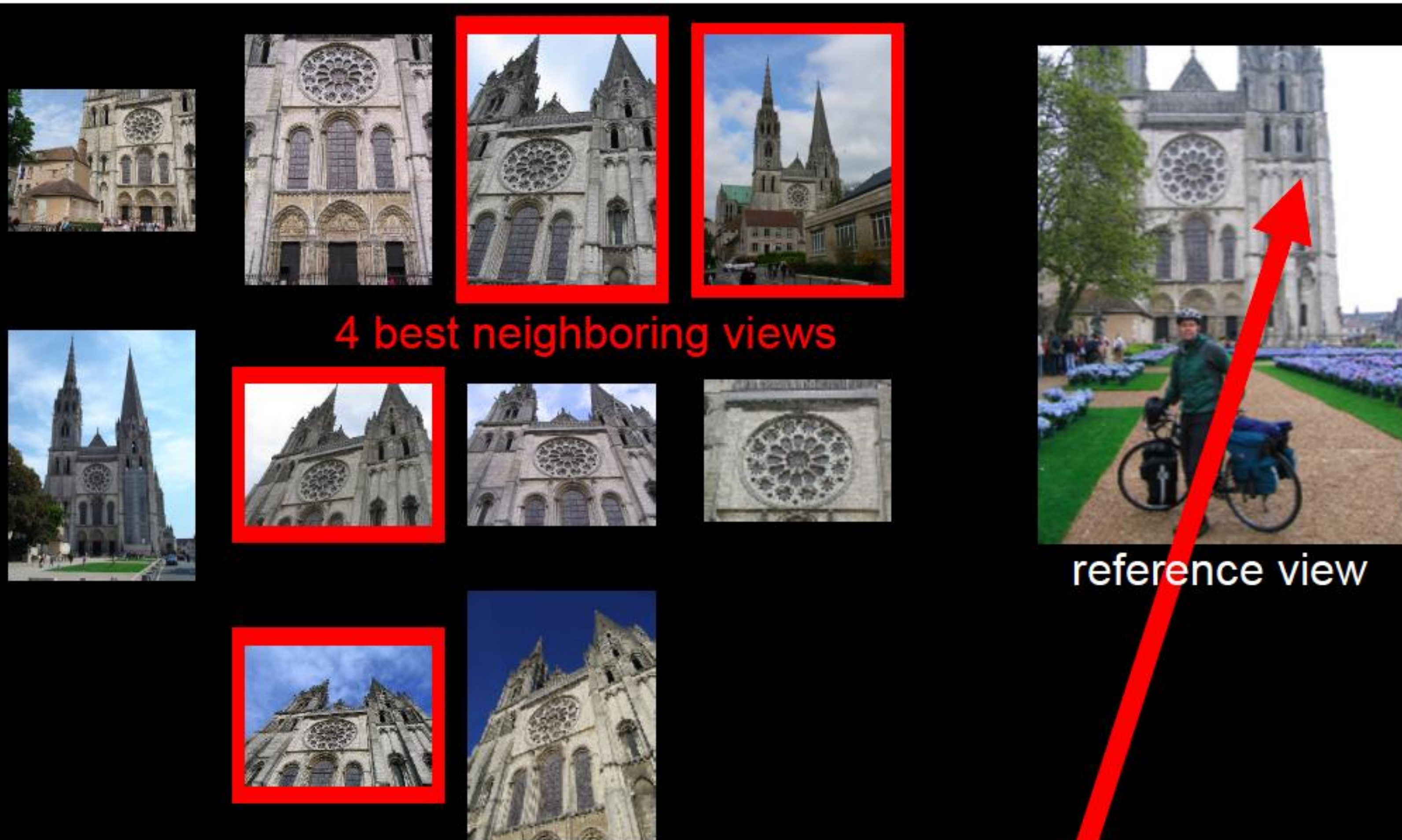
Local view selection

- Automatically select neighboring views for each point in the image
- Desiderata: good matches AND good baselines



Local view selection

- Automatically select neighboring views for each point in the image
- Desiderata: good matches AND good baselines



Local view selection

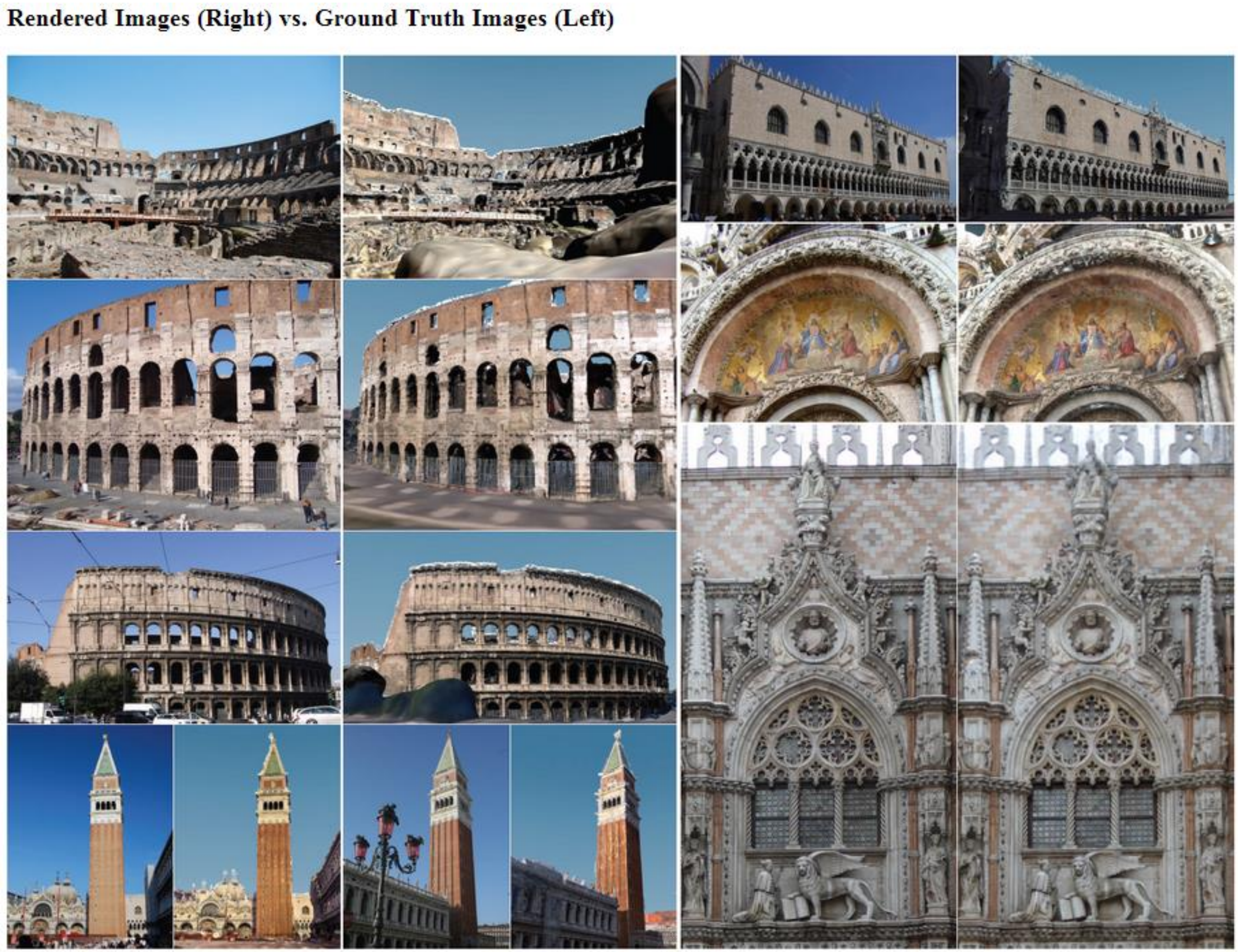
- Automatically select neighboring views for each point in the image
- Desiderata: good matches AND good baselines

Towards Internet-Scale Multi-View Stereo

[YouTube video](), [high-quality video]()

Yasutaka Furukawa, Brian Curless, Steven M. Seitz and Richard Szeliski, [Towards Internet-scale Multi-view Stereo](),CVPR 2010.

The Visual Turing Test for Scene Reconstruction



Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz, "The Visual Turing Test for Scene Reconstruction," 3DV 2013.

Fast stereo for Internet photo collections

- Start with a cluster of registered views
- Obtain a depth map for every view using plane sweeping stereo with normalized cross-correlation

Frahm et al., [“Building Rome on a Cloudless Day,”]() ECCV 2010.

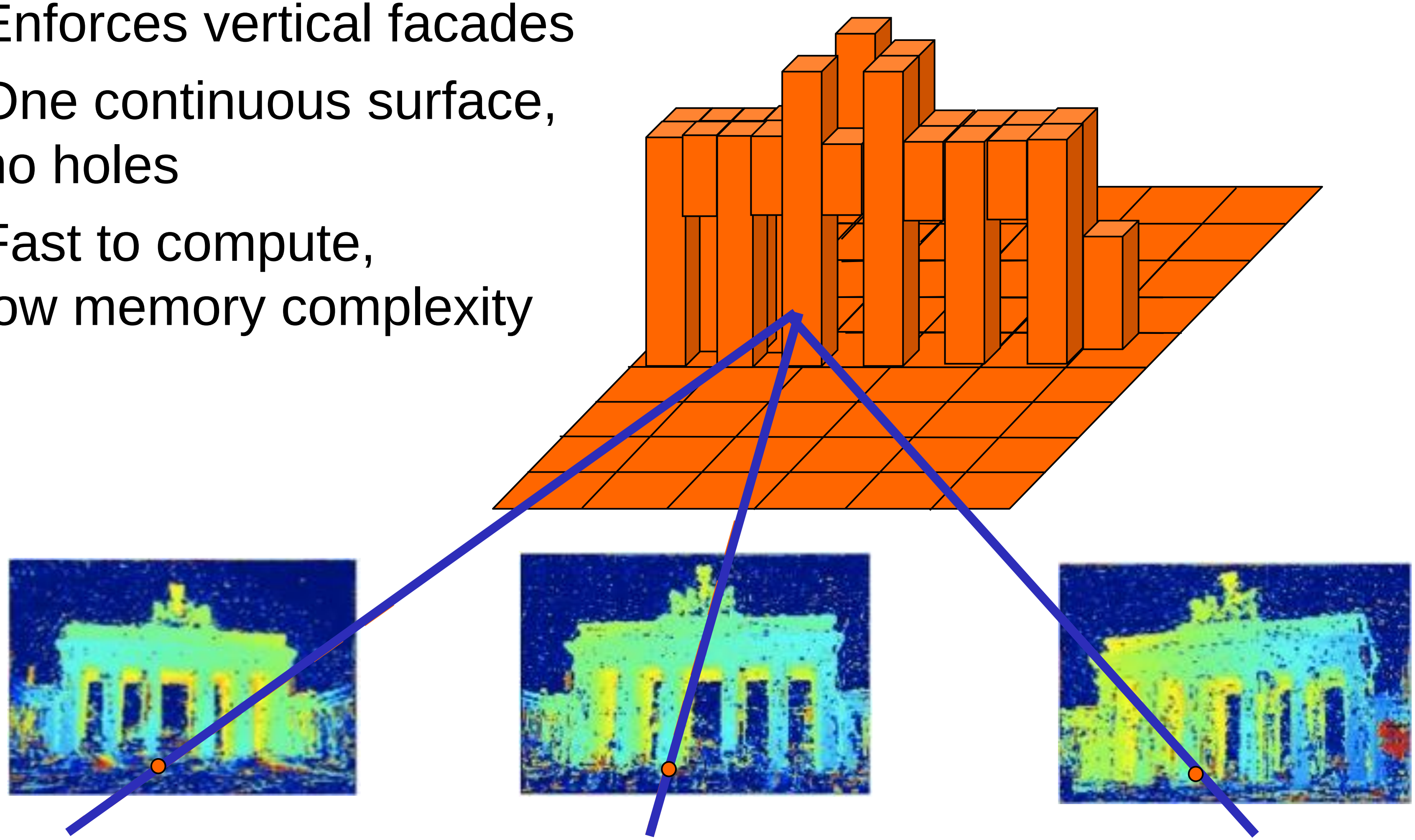
Plane sweeping stereo

- Need to register individual depth maps into a single 3D model
- Problem: depth maps are very noisy

Frahm et al., [“Building Rome on a Cloudless Day,”] ECCV 2010.

Robust stereo fusion using a heightmap

- Enforces vertical facades
- One continuous surface, no holes
- Fast to compute, low memory complexity



David Gallup, Marc Pollefeys, Jan-Michael Frahm, “3D Reconstruction using an n-Layer Heightmap”, DAGM 2010

Results

[YouTube Video](#)

Frahm et al., [“Building Rome on a Cloudless Day,”](#) ECCV 2010.

Slide Credits

Rob Fergus – NYU

Darell Trevor – UC Berkeley

Fei Fei Li - Stanford

Svetlana Lazebnik – UIUC

David A. Forsyth - UIUC

Next class

- RANSAC
  - Reading -
    - Forsyth & Ponce 10.1-10.4
    - Szeliski 4.3

Questions