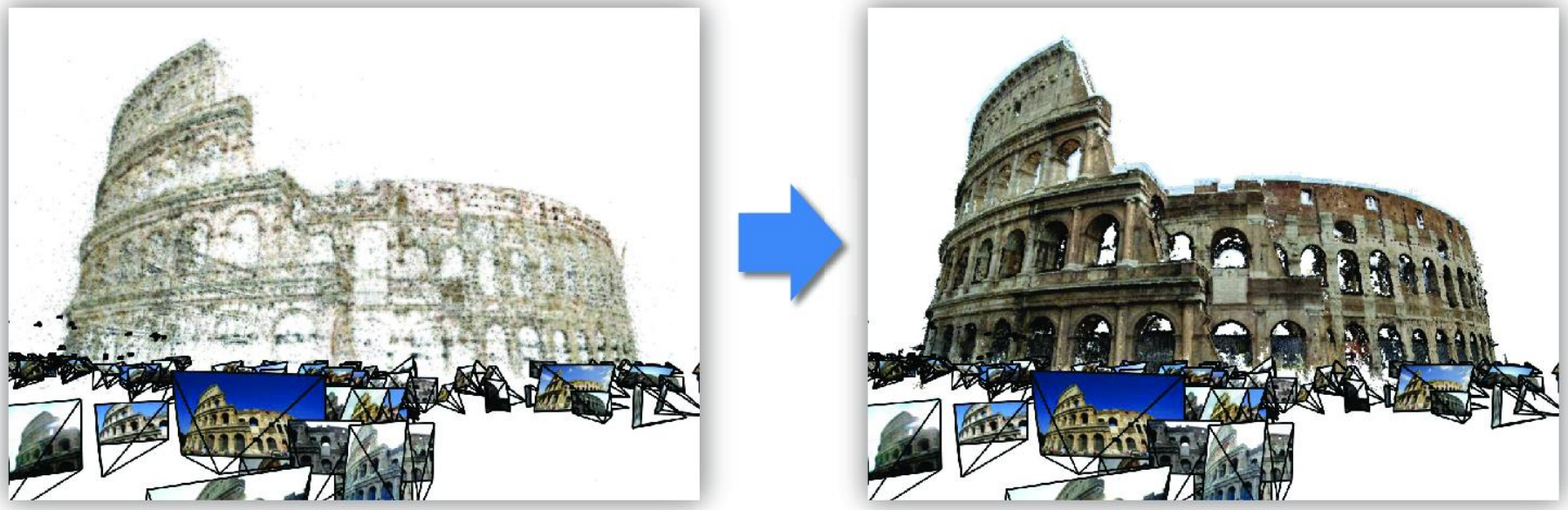
CS5670: Computer Vision

Noah Snavely

Multi-view stereo



Announcements

- Project 3 (“Autostitch”) due Monday 4/17 by 11:59pm

Recommended Reading

Szeliski Chapter 11.6

Multi-View Stereo: A Tutorial

Furukawa and Hernandez, 2015

http://www.cse.wustl.edu/~furukawa/papers/fnt_mvs.pdf

Multi-view Stereo

Stereo

Multi-view stereo

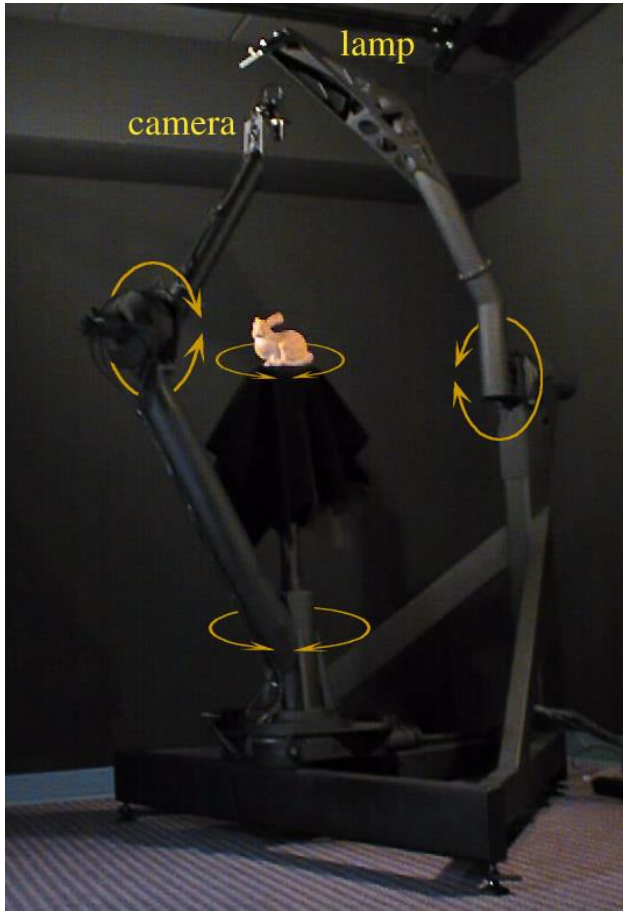
Multi-view Stereo

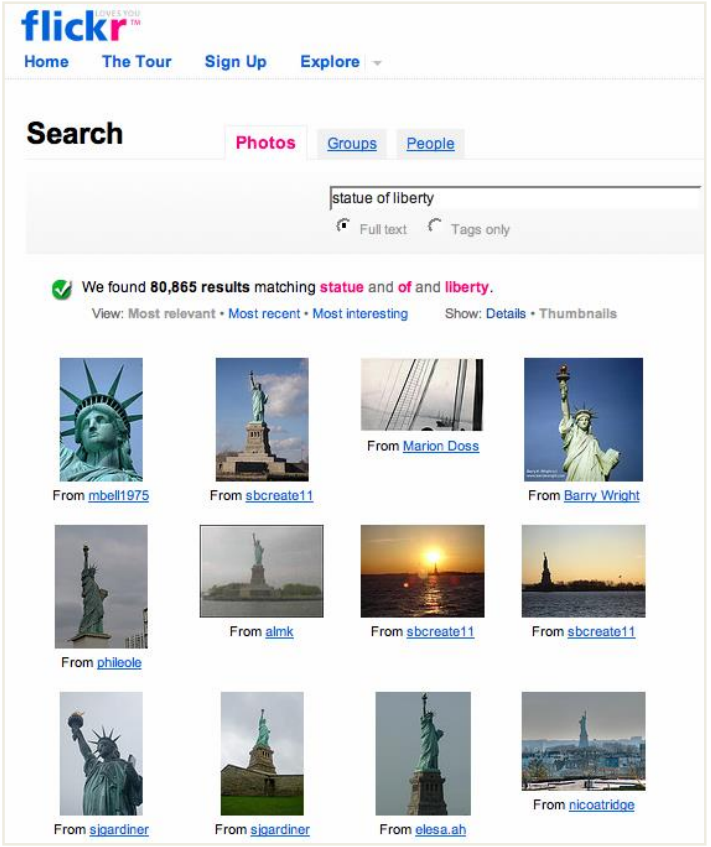
Point Grey's Bumblebee XB3

Point Grey's ProFusion 25

CMU's 3D Room

Multi-view Stereo

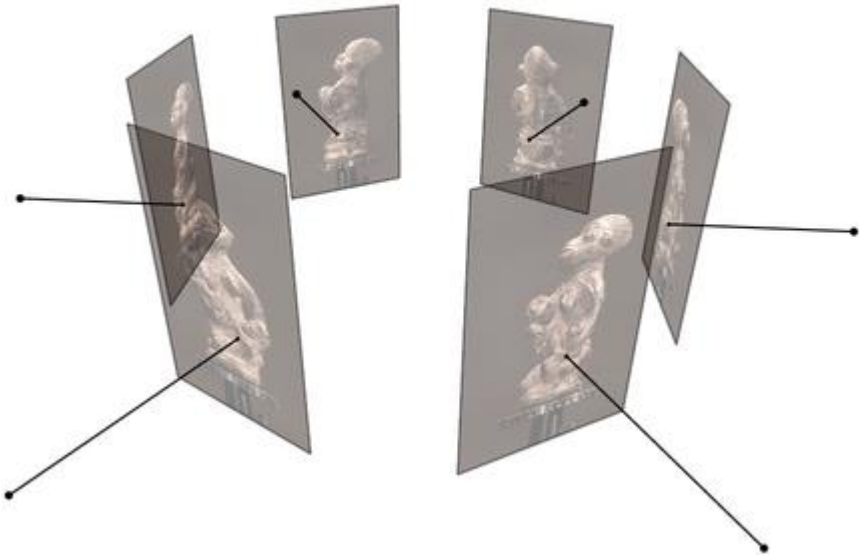




Multi-view Stereo

Input: calibrated images from several viewpoints

Output: 3D object model



Figures by Carlos Hernandez

Faugeras, Keriven
1998

Hernandez, Schmitt
2004

Pons, Keriven, Faugeras
2005

Furukawa, Ponce
2006

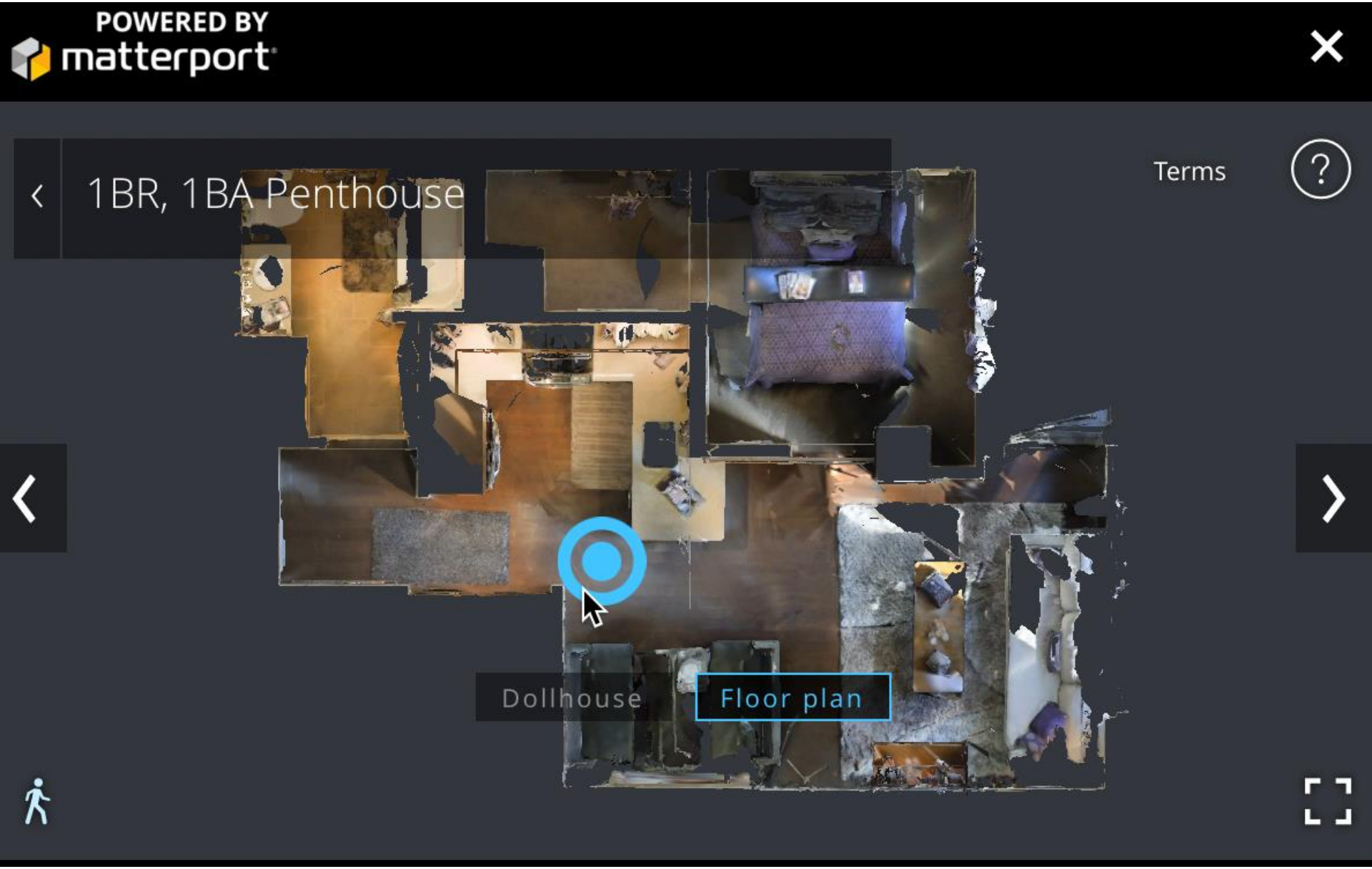
Goesele et al.
2007

Applications


Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Google earth

でんじろんさん

by じろう でん 266,


POWERED BY
matterport
1BR, 1BA Penthouse
Terms
Dollhouse
Floor plan

Whistle in the Form of Female Figure *600 AD - 900 AD*

Details Los Angeles County Museum of Art

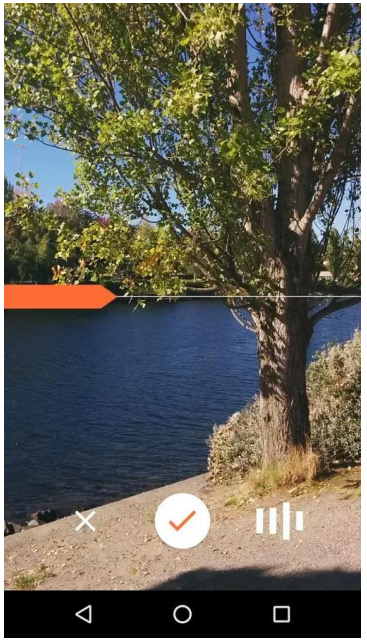
Los Angeles County Museum of Art | Sculpture | Mexico

Share | Compare | Saved 0 | Discover

Google

Google

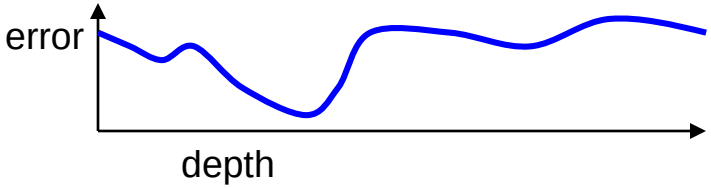
.JUMP




c|net


http://bit.ly/surround360

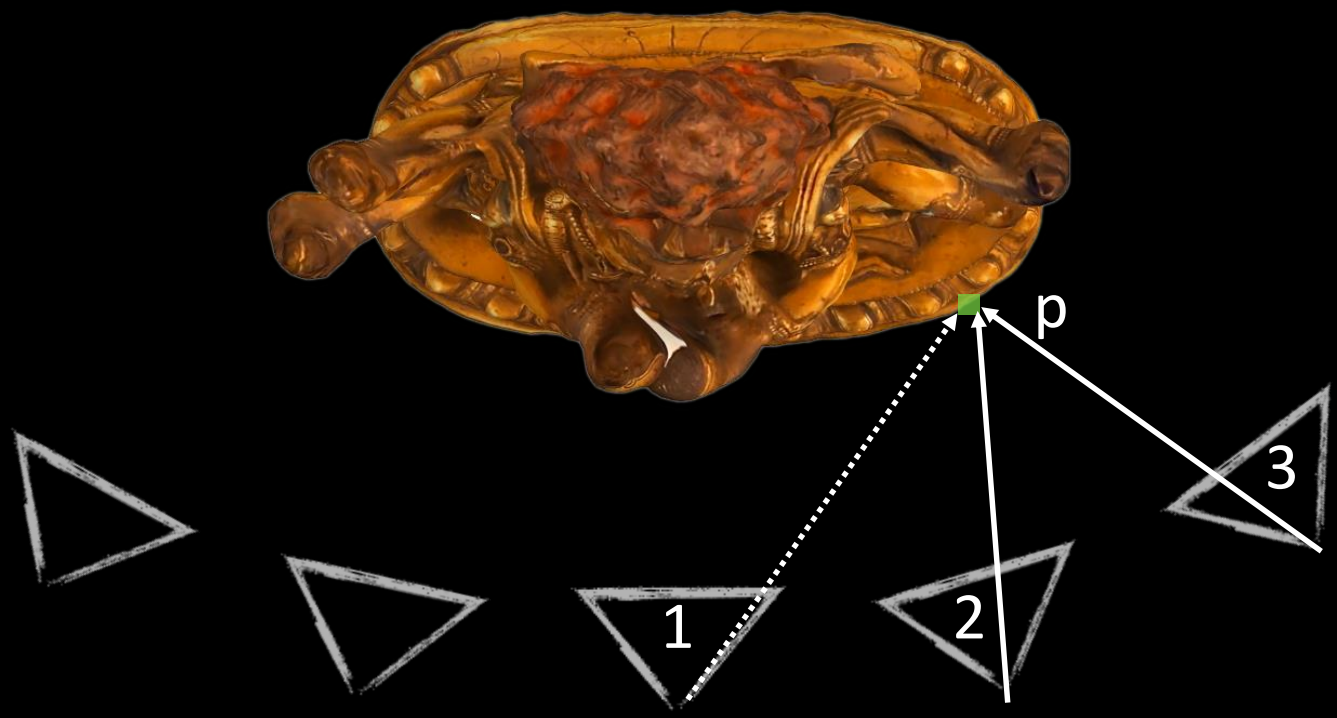
Stereo: another view



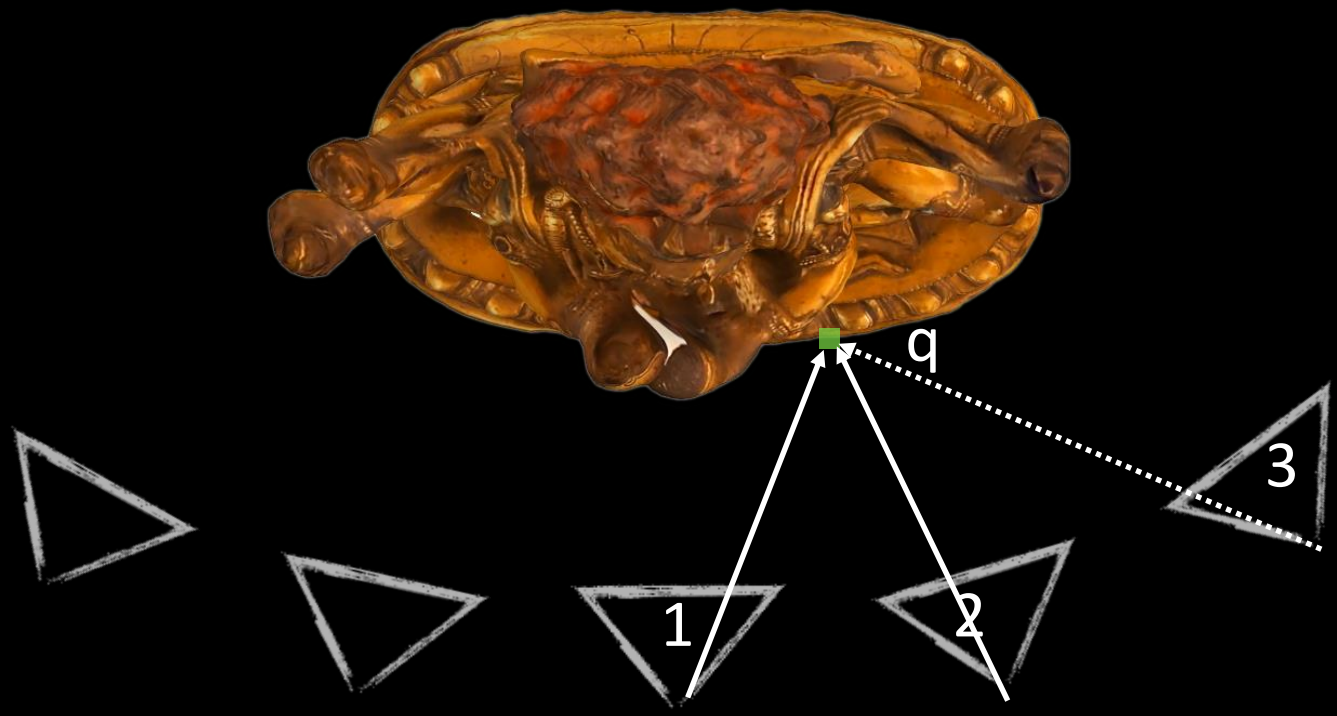


Why MVS?

- Different points on the object's surface will be more clearly visible in some subset of cameras
  - Could have high-res closeups of some regions
  - Some surfaces are foreshortened from certain views



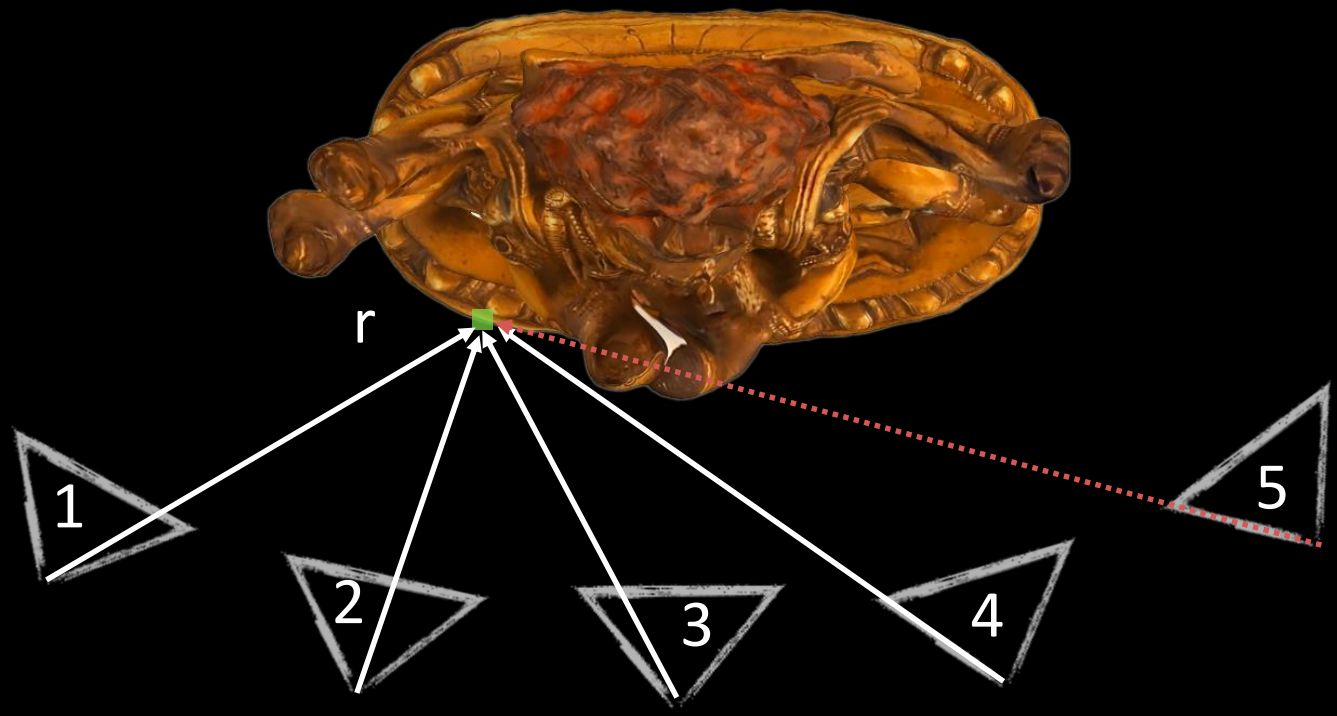
Cameras 2 and 3 can more clearly see point p.



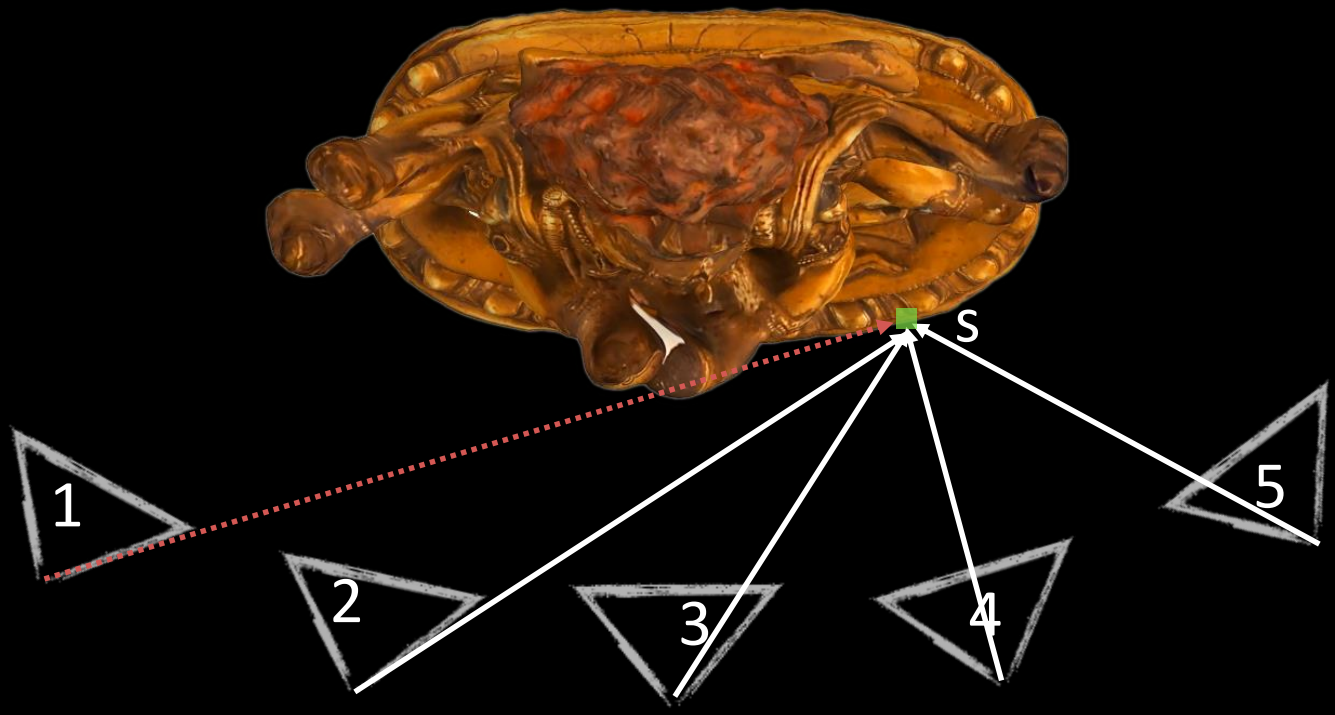
Cameras 1 and 2 can more clearly see point q.

Why MVS?

- Different points on the object's surface will be more clearly visible in some subset of cameras
  - Could have high res closeups of some regions
  - Some surfaces are foreshortened from certain views
- Some points may be occluded entirely in certain views



Camera 5 can’t see point r.



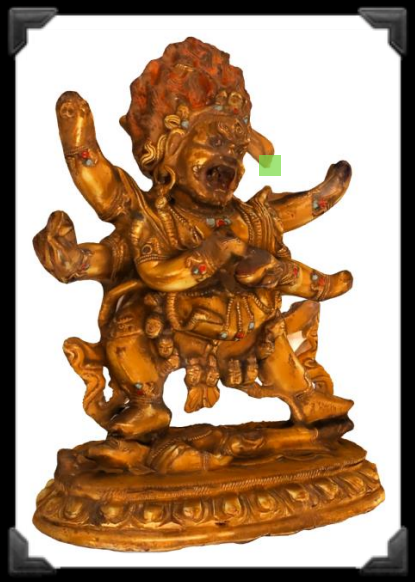
Camera 1 can’t see point s.

Why MVS?

- Different points on the object's surface will be more clearly visible in some subset of cameras
  - Could have high res closeups of some regions
  - Some surfaces are foreshortened from certain views
- Some points may be occluded entirely in certain views
- More measurements per point can reduce error

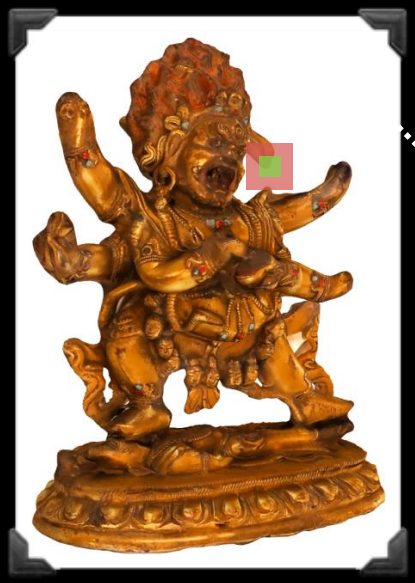
1

2



1

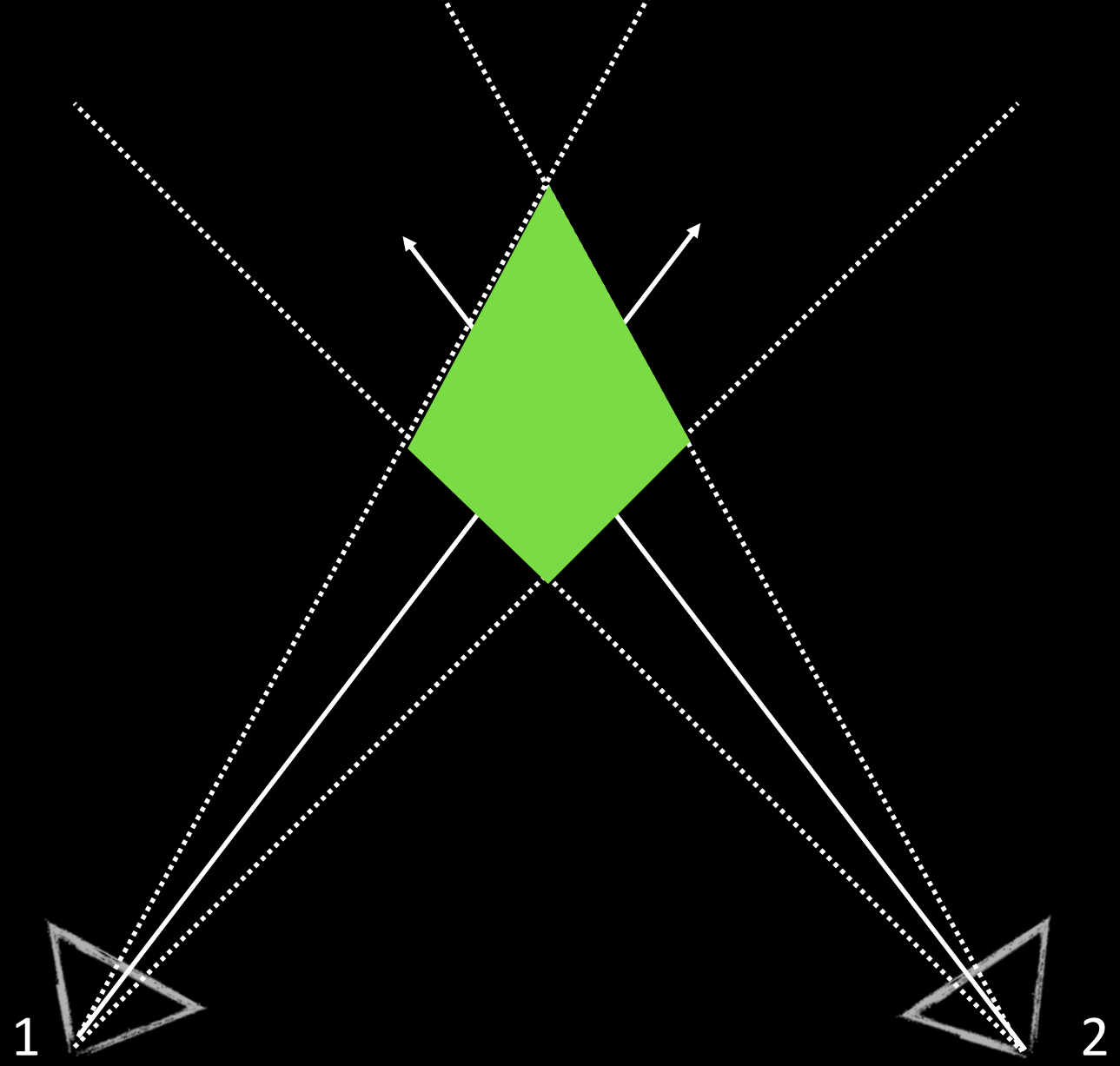
2



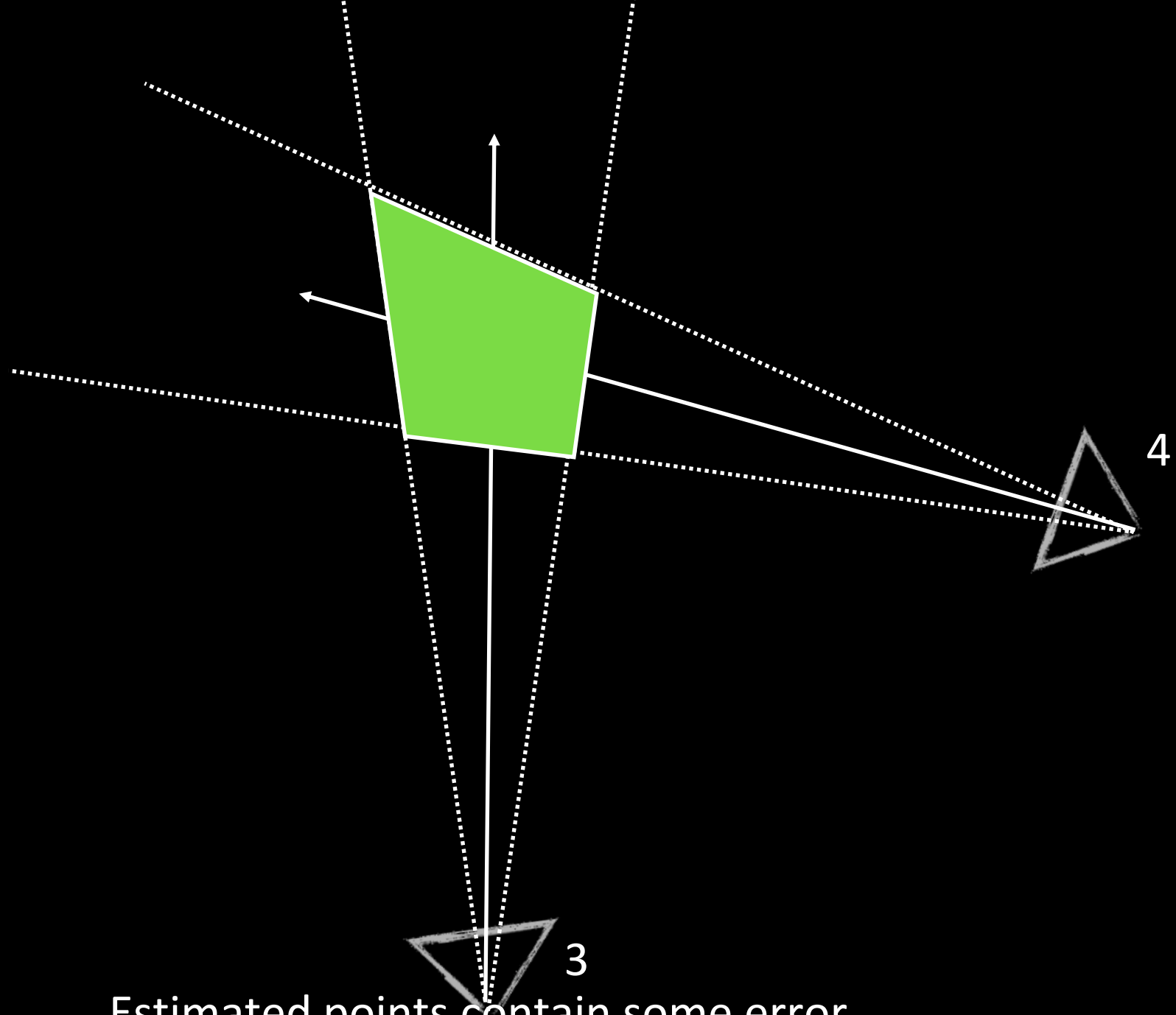
1

2

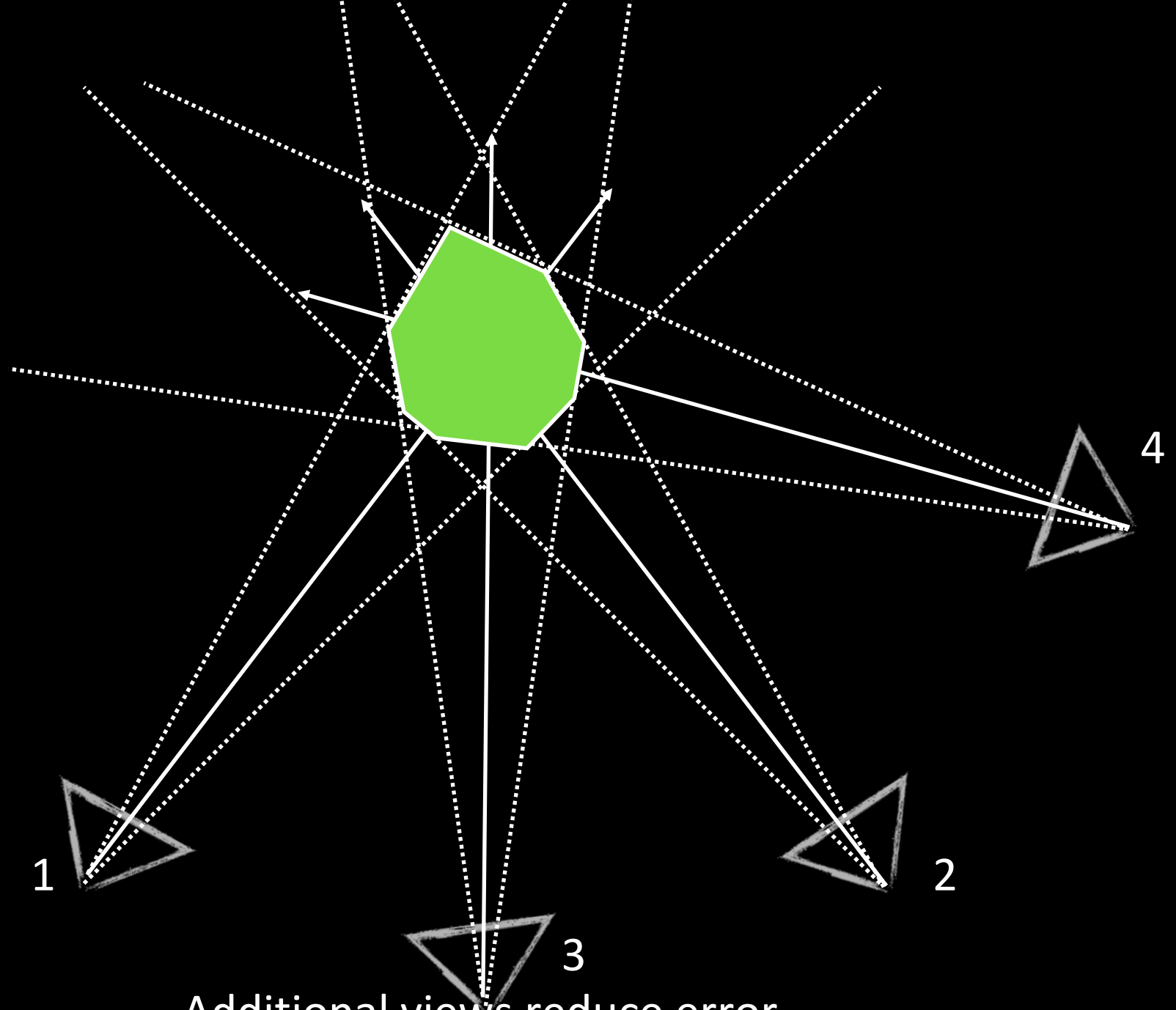
Estimated points contain some error.



Estimated points contain some error.

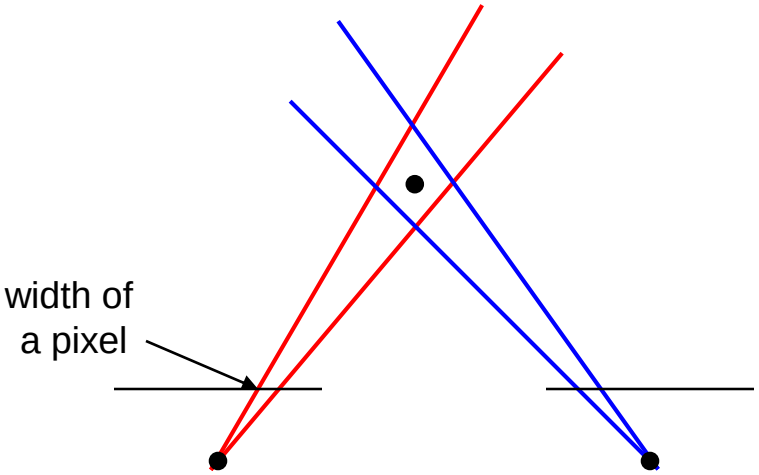


Estimated points contain some error.

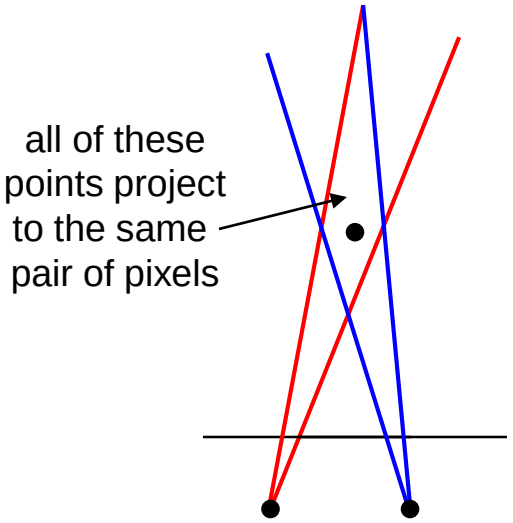


Additional views reduce error.

Choosing the stereo baseline



Large Baseline



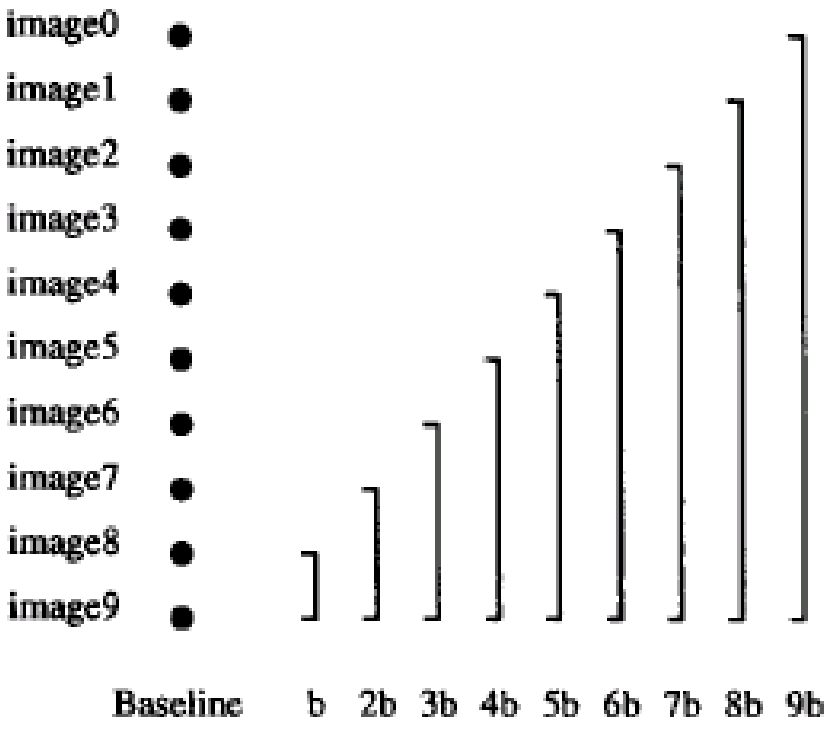
Small Baseline

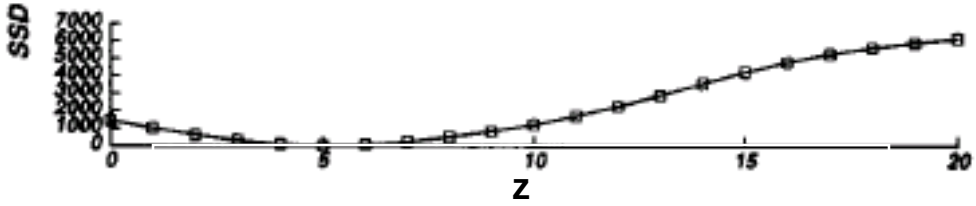
What's the optimal baseline?

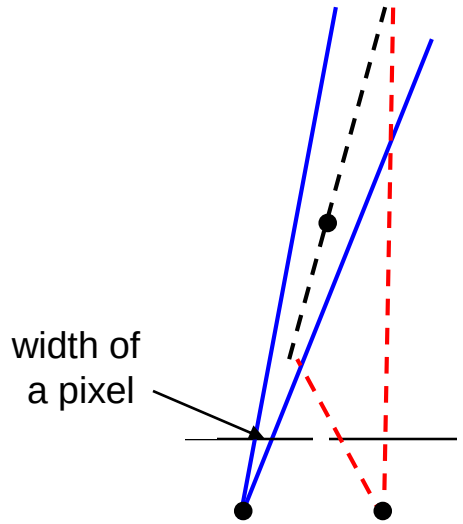
- Too small: large depth error
- Too large: difficult search problem

The Effect of Baseline on Depth Estimation

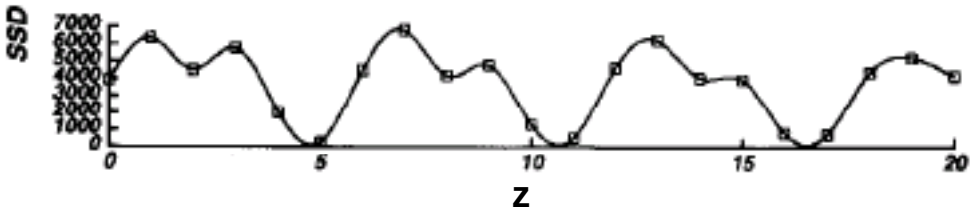
Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.

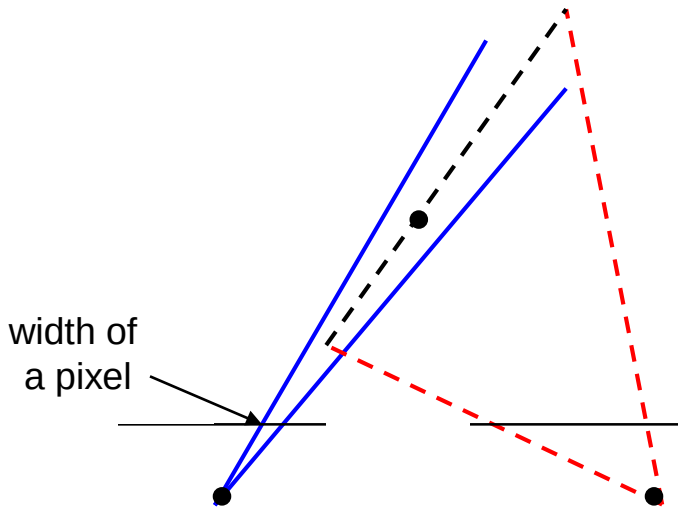






pixel matching score





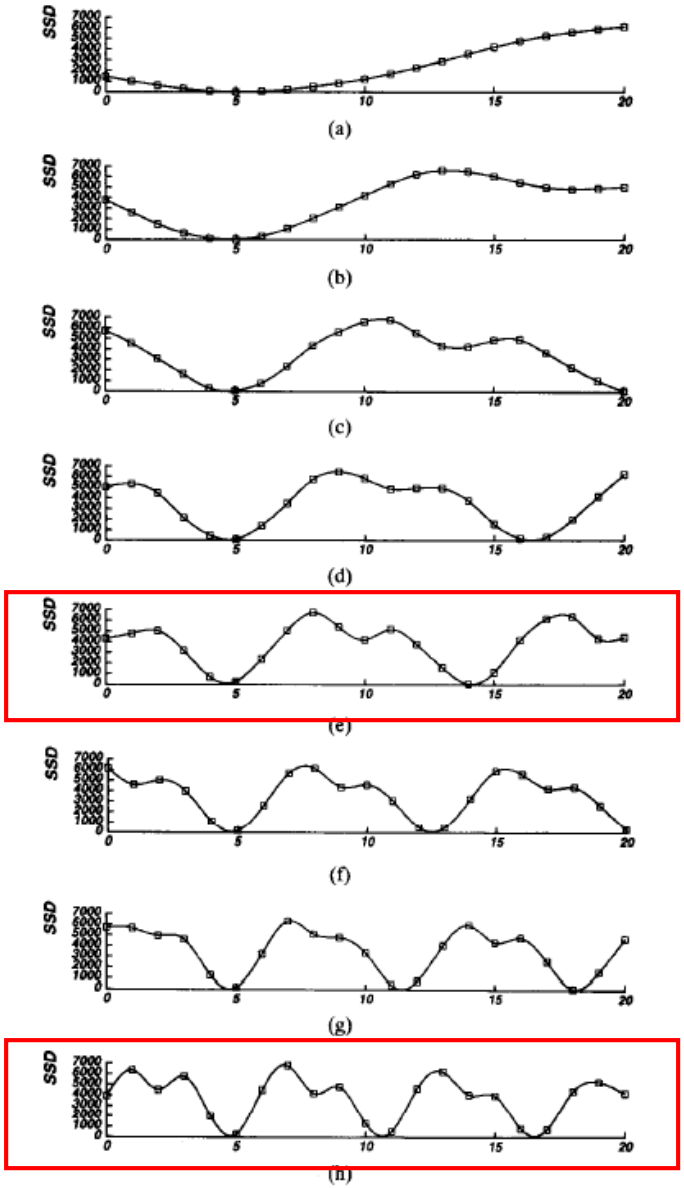


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

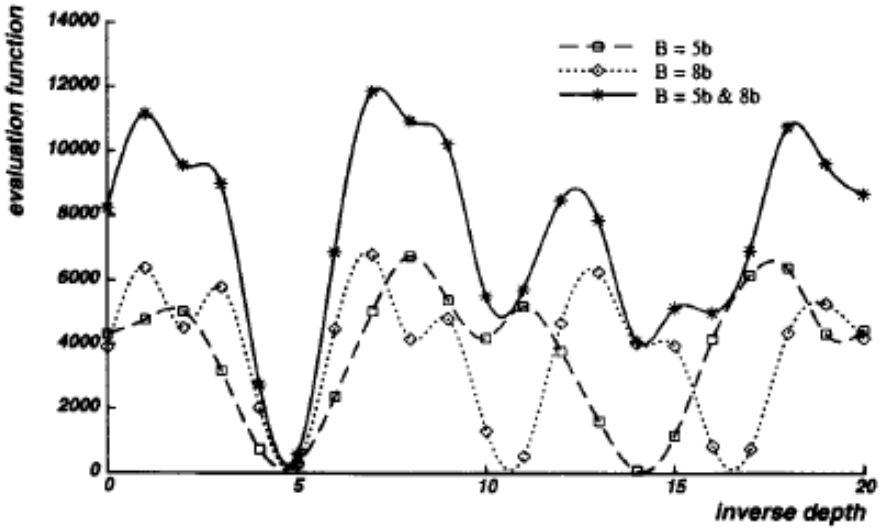


Fig. 6. Combining two stereo pairs with different baselines.

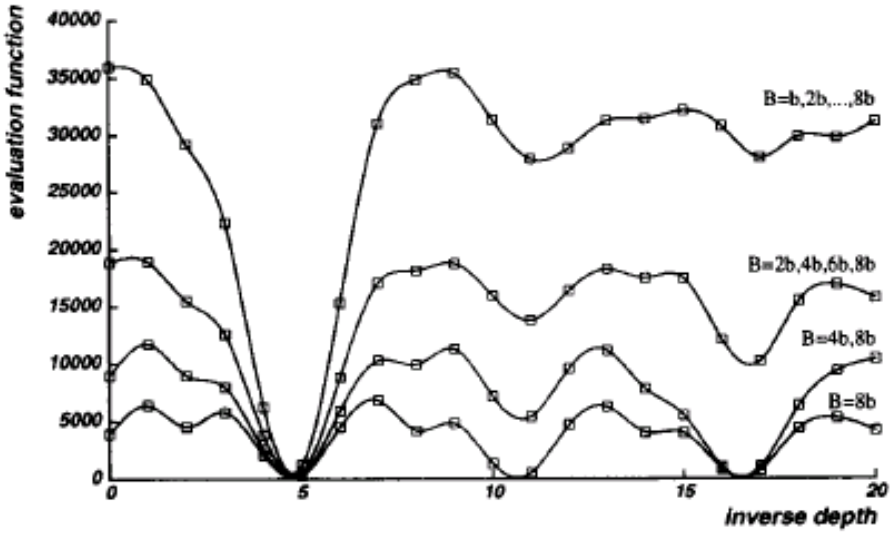


Fig. 7. Combining multiple baseline stereo pairs.

Multibaseline Stereo

Basic Approach

- Choose a reference view
- Use your favorite stereo algorithm BUT
  - replace two-view SSD with SSSD over all baselines

Limitations

Problem: *visibility*

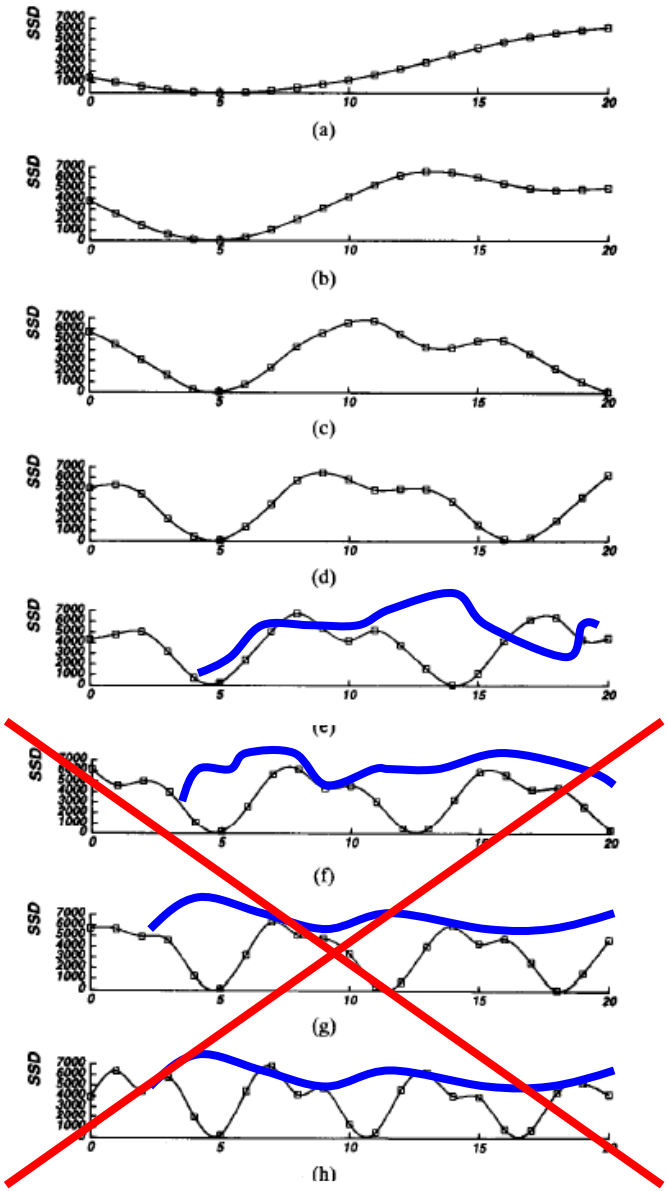


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

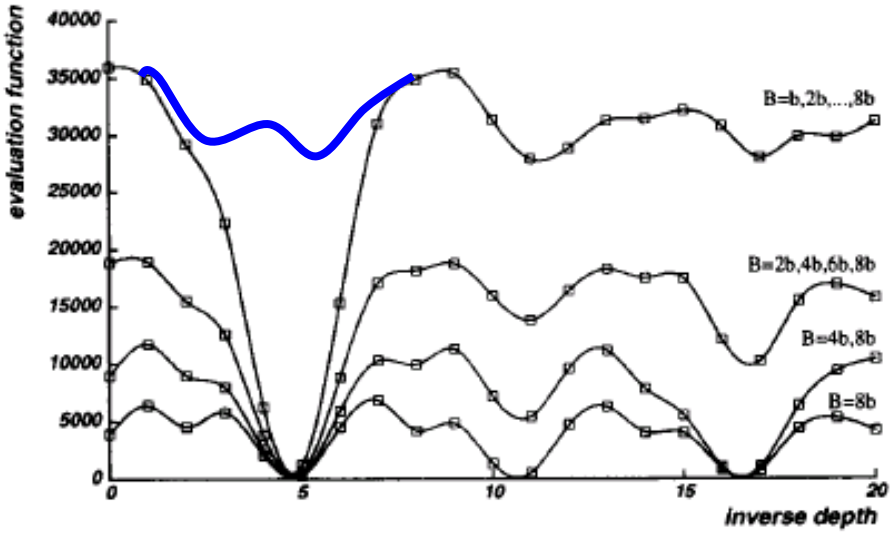


Fig. 7. Combining multiple baseline stereo pairs.

Some Solutions

- Match only nearby photos [Narayanan 98]
- Use NCC instead of SSD,
  Ignore NCC values > threshold
  [Hernandez & Schmitt 03]

Popular matching scores

- SSD (Sum Squared Distance)

$$\sum_{x,y} |W_1(x,y) - W_2(x,y)|^2$$

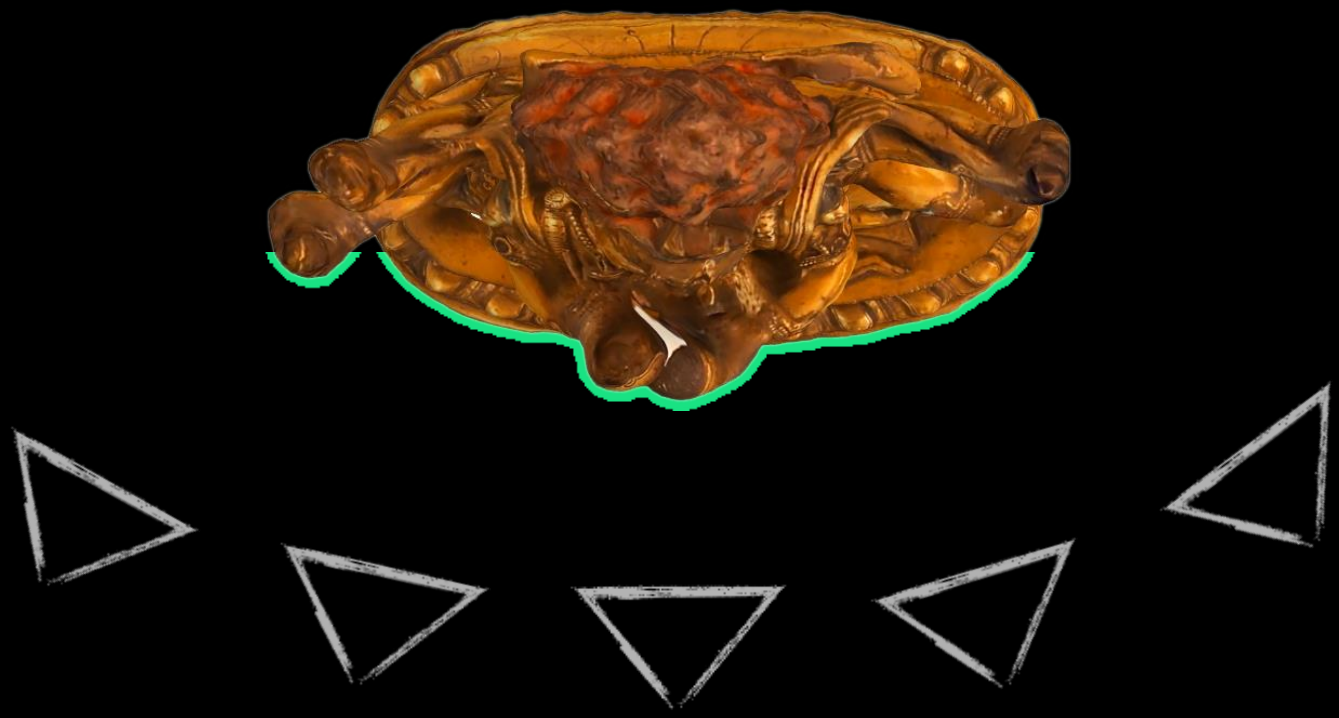
- NCC (Normalized Cross Correlation)

$$\frac{\sum_{x,y}(W_1(x,y) - \overline{W_1})(W_2(x,y) - \overline{W_2})}{\sigma_{W_1}\sigma_{W_2}}$$

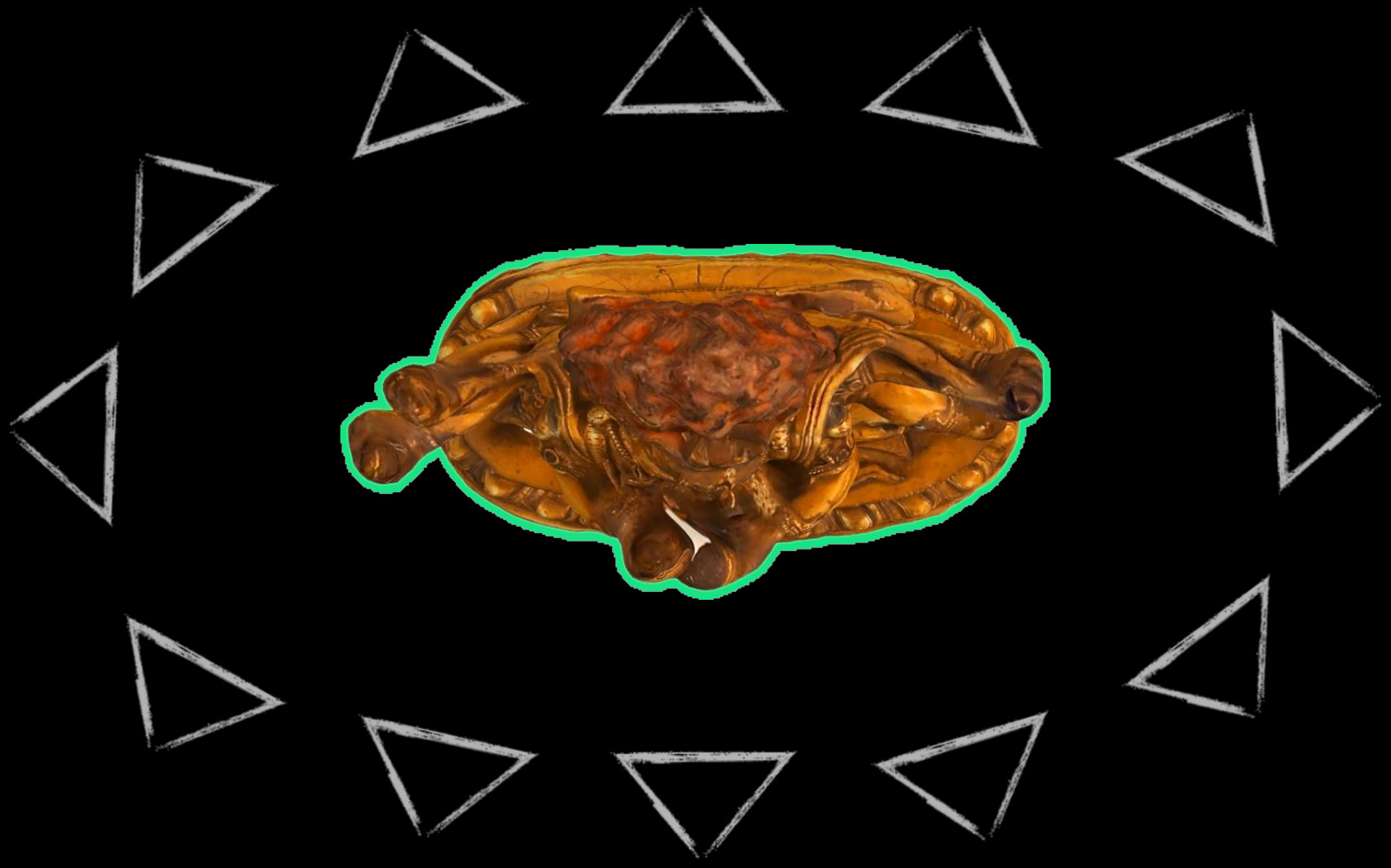
  - where $\overline{W_i} = \frac{1}{n}\sum_{x,y} W_i \qquad \sigma_{W_i} = \sqrt{\frac{1}{n}\sum_{x,y}(W_i - \overline{W_i})^2}$

  - what advantages might NCC have?

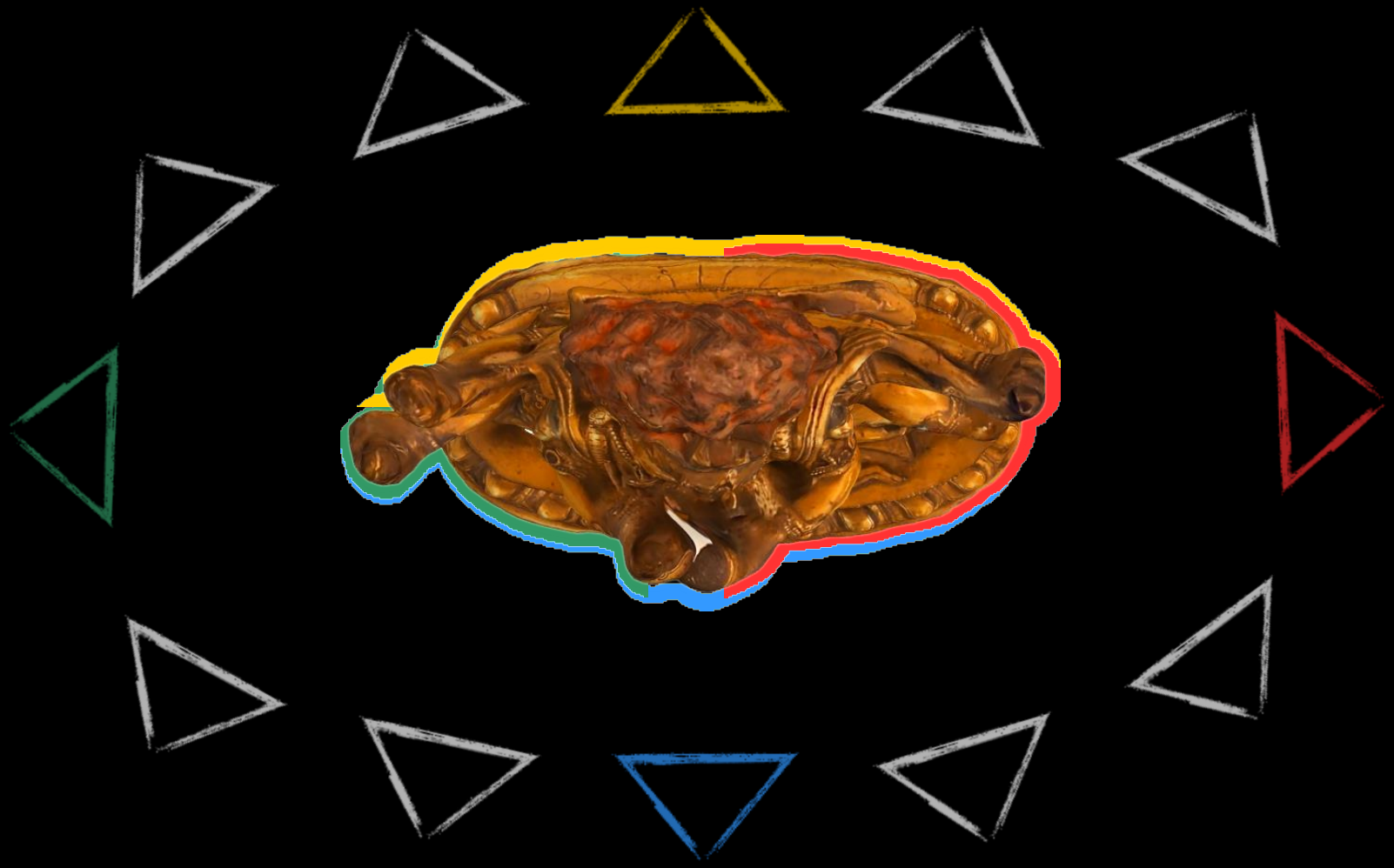
Single depth map often isn't enough



Really want full coverage

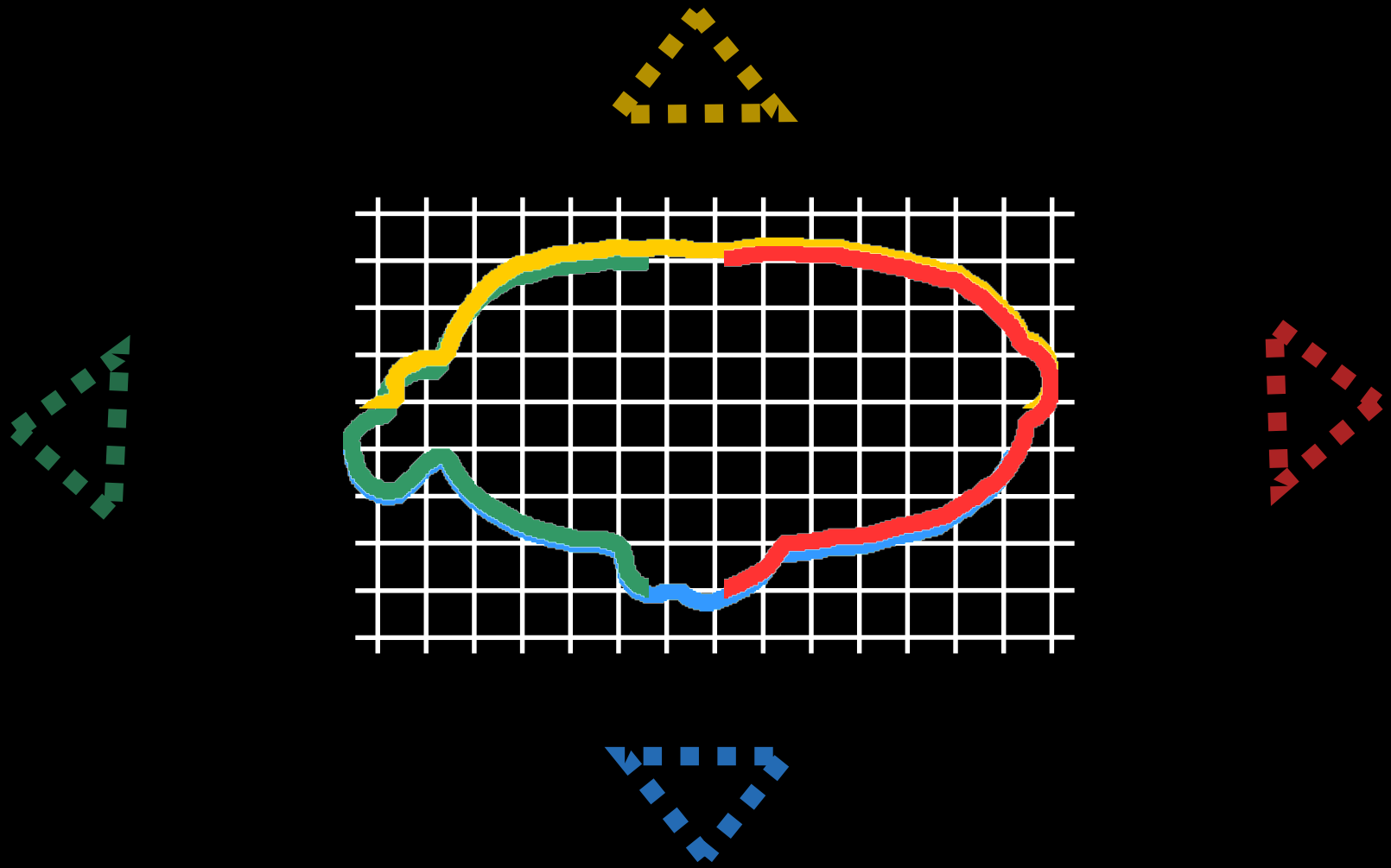


Idea: Combine many depth maps



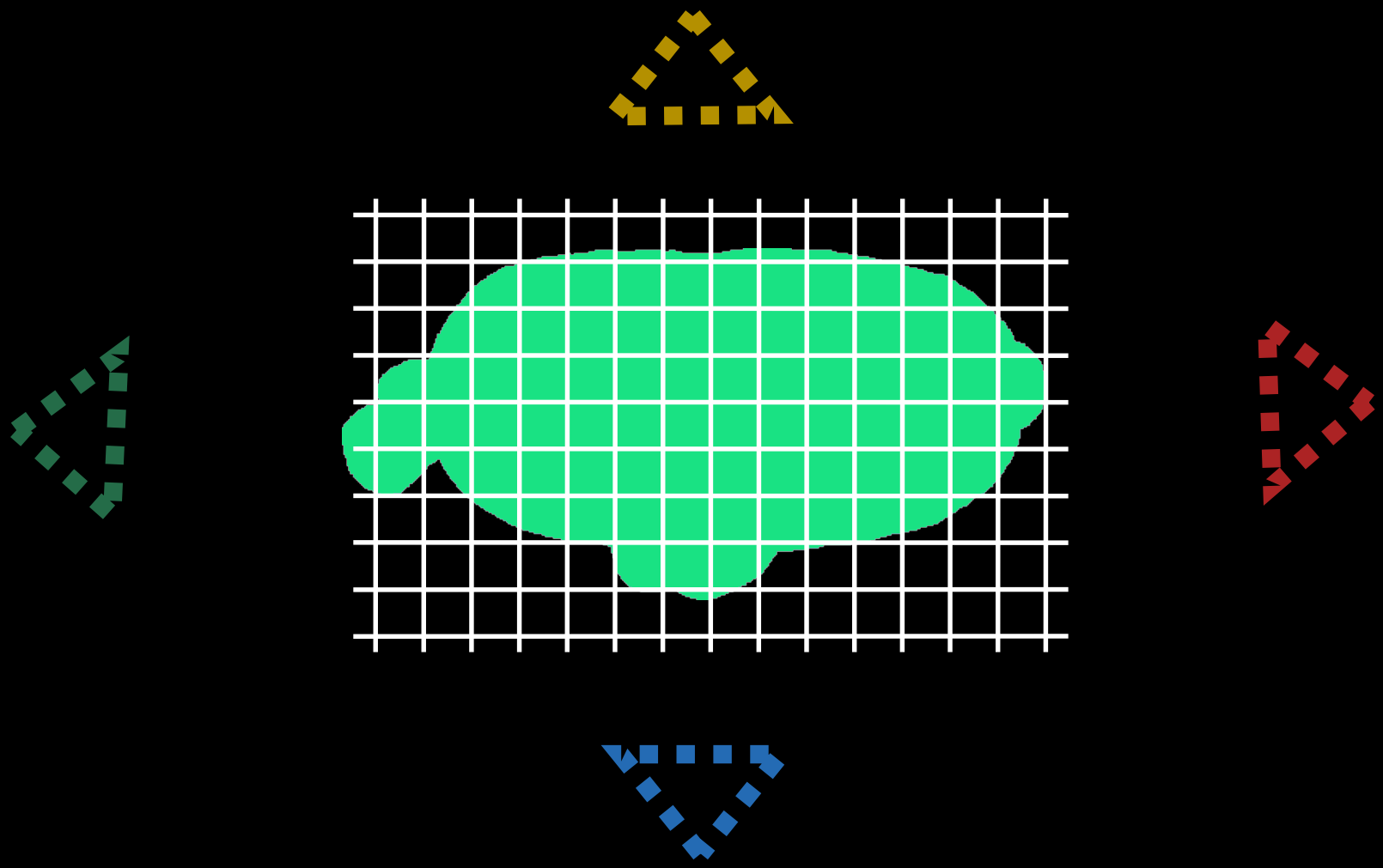
Many depth maps, each with error. How can we fuse these?

Volumetric fusion

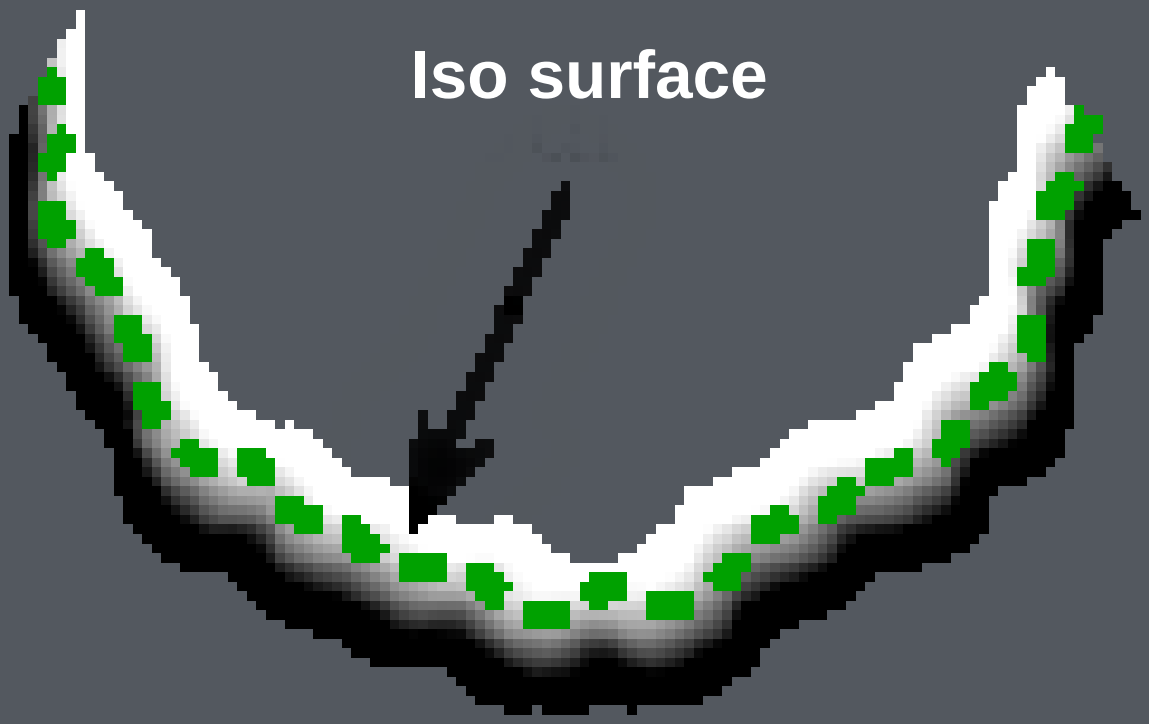


A common world-space coordinate system.

Volumetric fusion



A common world-space coordinate system.


Iso surface

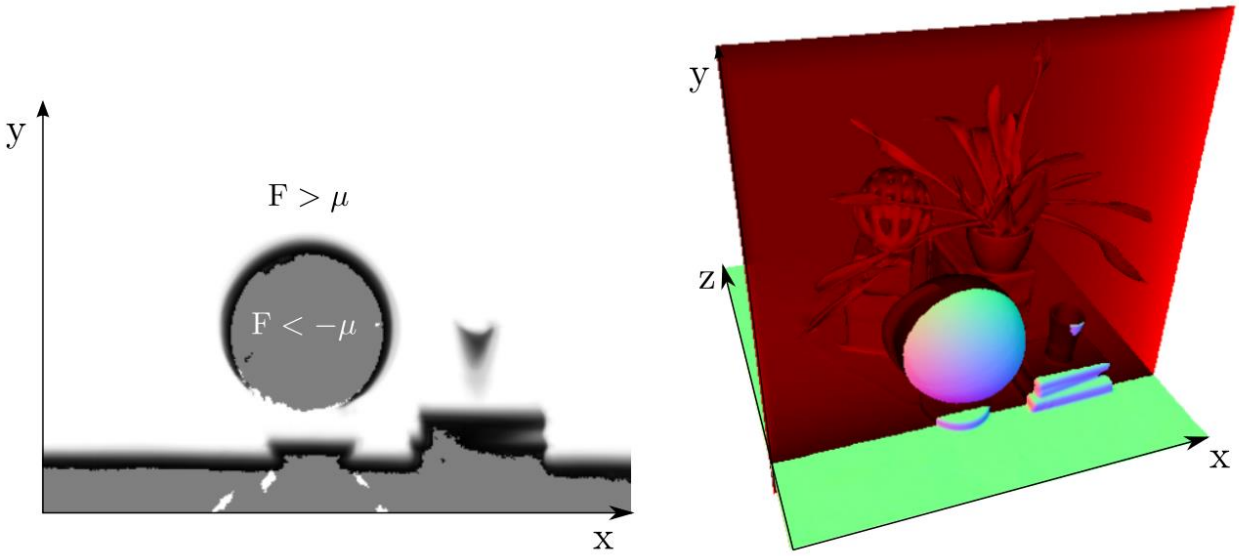
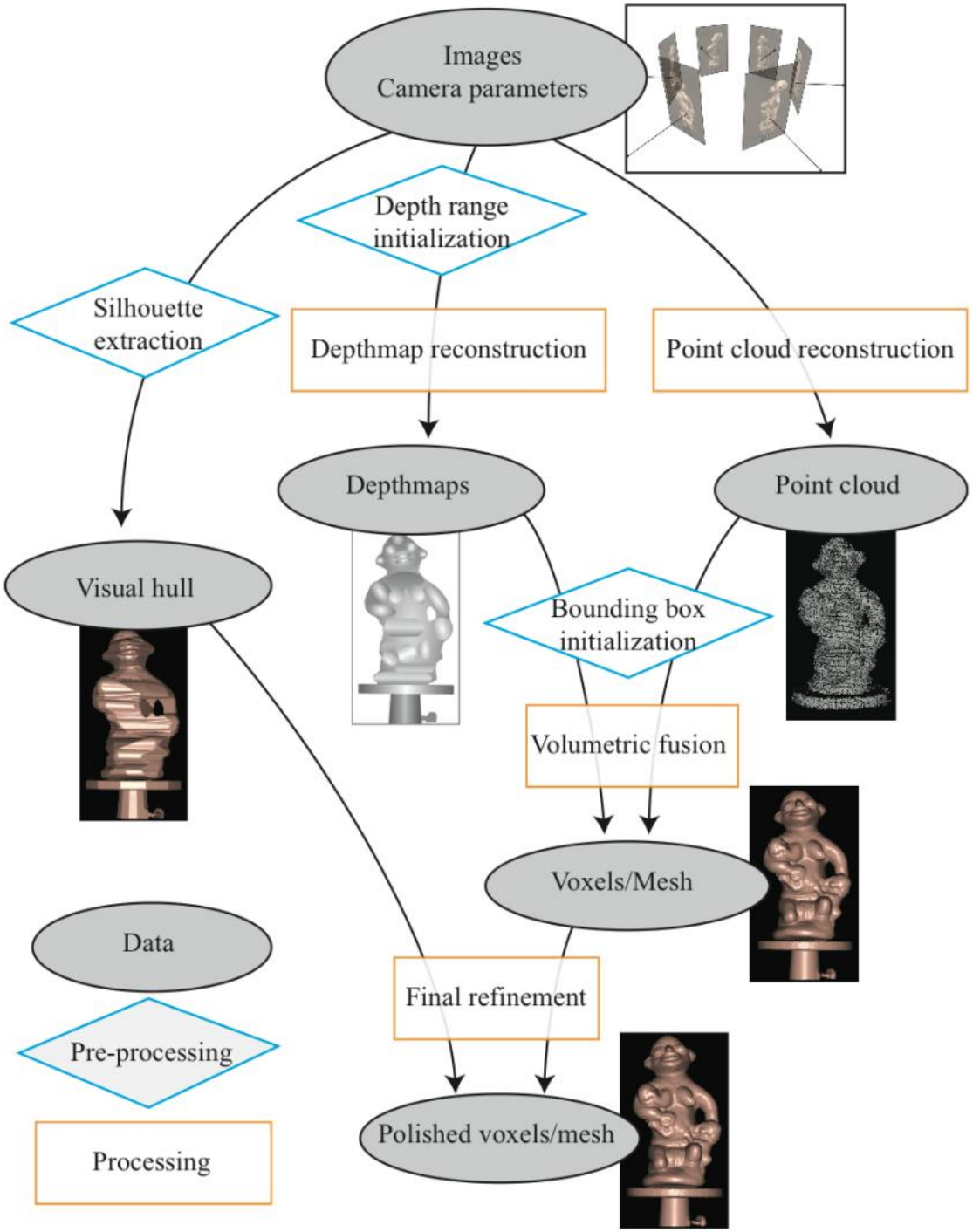
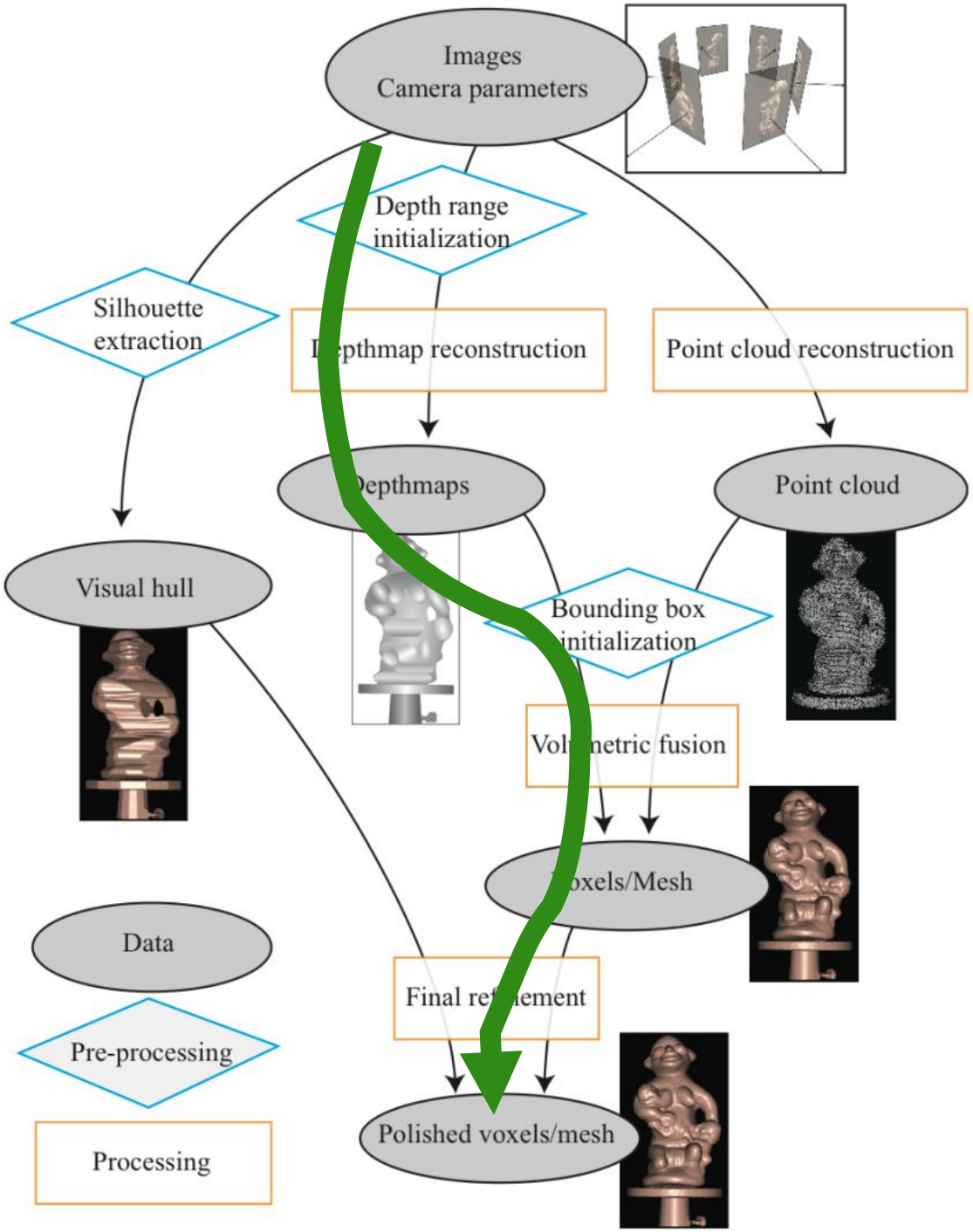
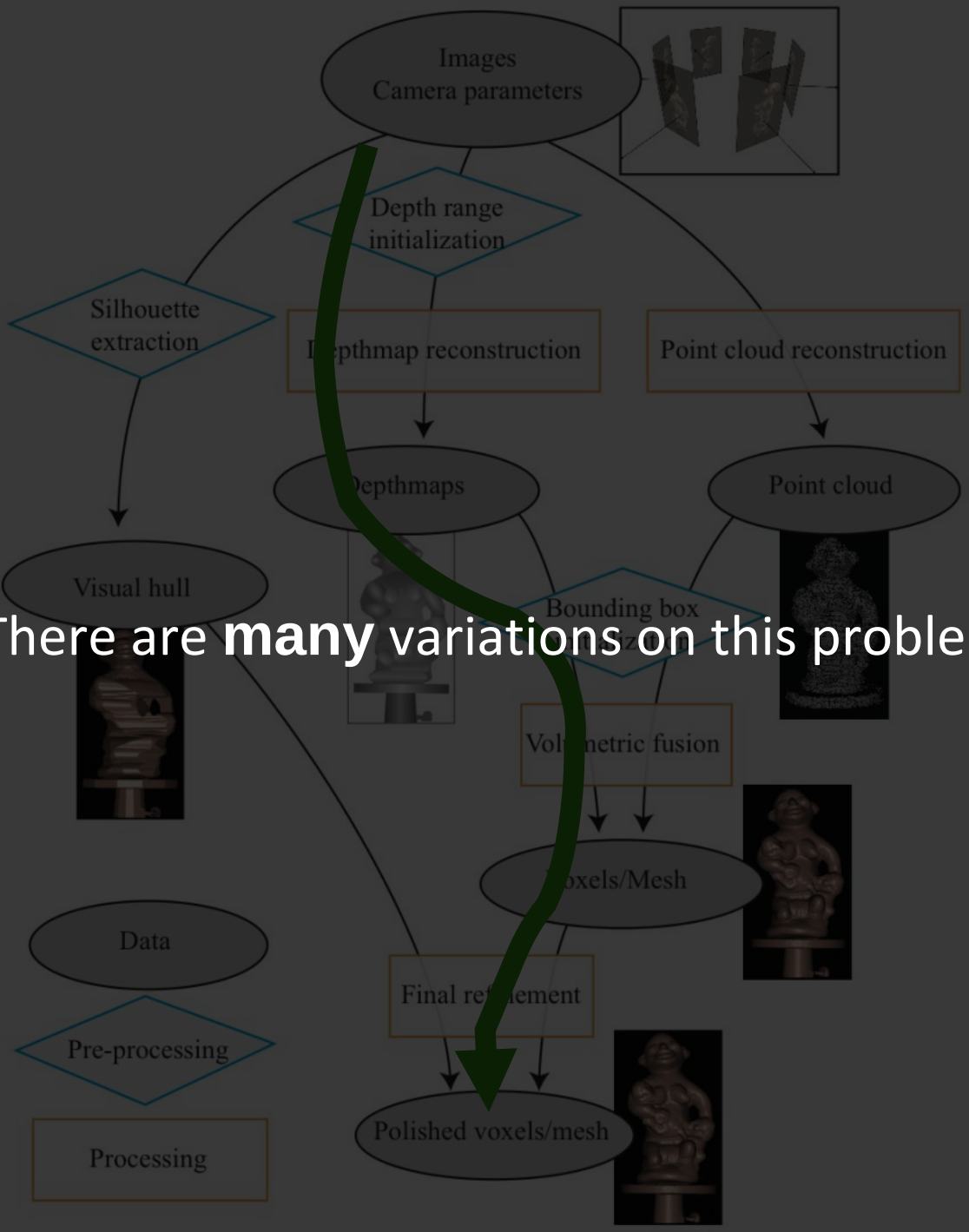


Figure 4: A slice through the truncated signed distance volume showing the truncated function $F > \mu$ (white), the smooth distance field around the surface interface $F = 0$ and voxels that have not yet had a valid measurement(grey) as detailed in eqn. 9.

Newcombe, et. al. ISMAR 2011.


Images
Camera parameters
Depth range initialization
Silhouette extraction
Depthmap reconstruction
Point cloud reconstruction
Depthmaps
Point cloud
Visual hull
Bounding box initialization
Volumetric fusion
Voxels/Mesh
Final refinement
Polished voxels/mesh
Data
Pre-processing
Processing

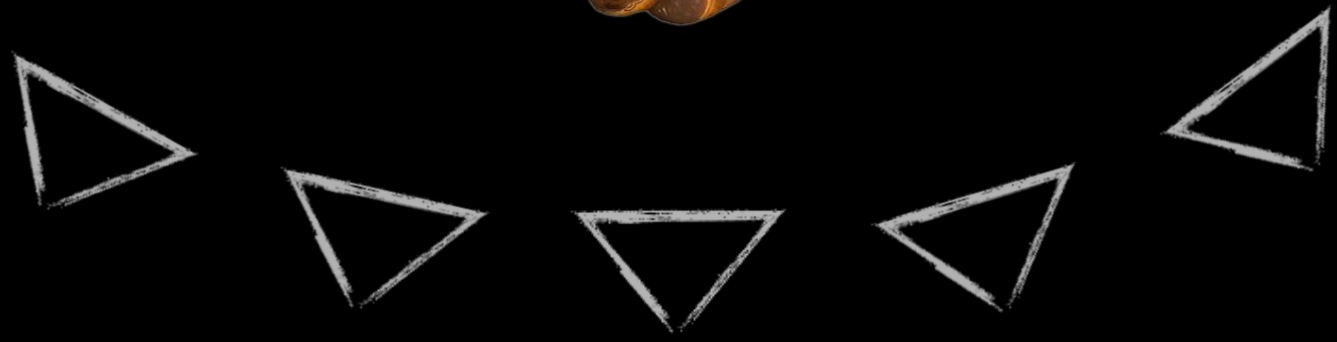

Images
Camera parameters
Depth range initialization
Silhouette extraction
Depthmap reconstruction
Point cloud reconstruction
Depthmaps
Point cloud
Visual hull
Bounding box initialization
Volumetric fusion
Voxels/Mesh
Final refinement
Polished voxels/mesh
Data
Pre-processing
Processing



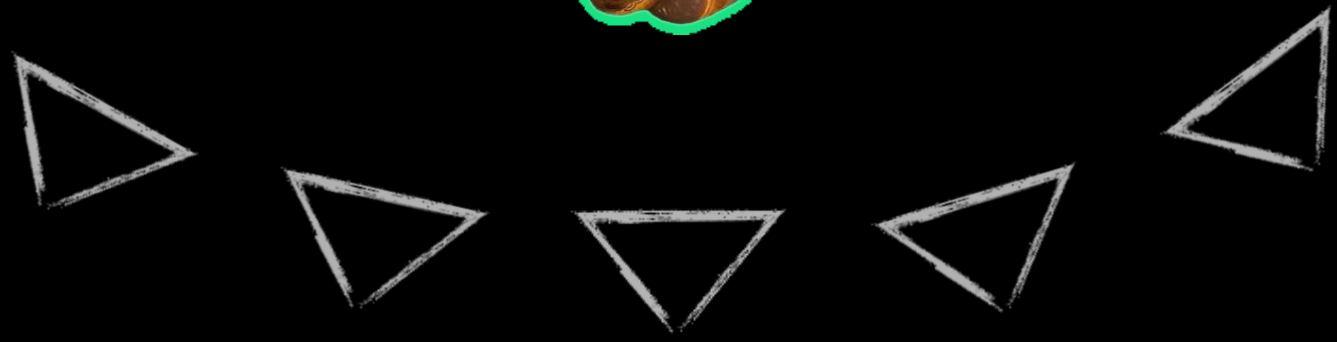
There are **many** variations on this problem.

Questions?

Are depth maps enough?



Are depth maps enough?



Depends on your application…

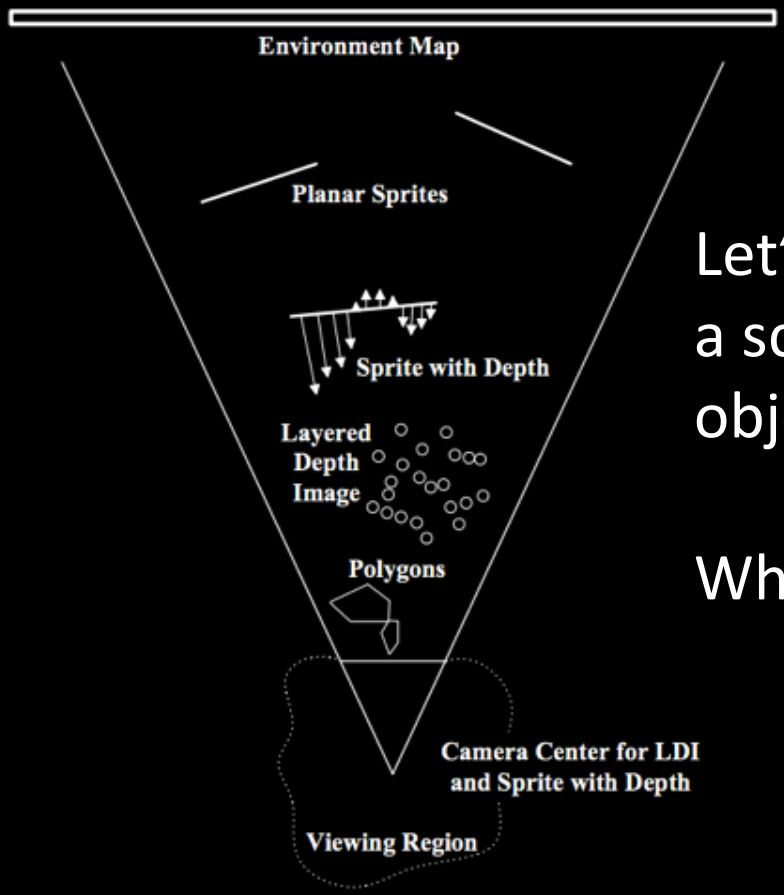


Figure 1 Different image based primitives can serve well depending on distance from the camera

Let's assume you want to render a scene containing your real-world object.

What should you capture?

Shade, et. al. SIGGRAPH 98.

Depends on your application…

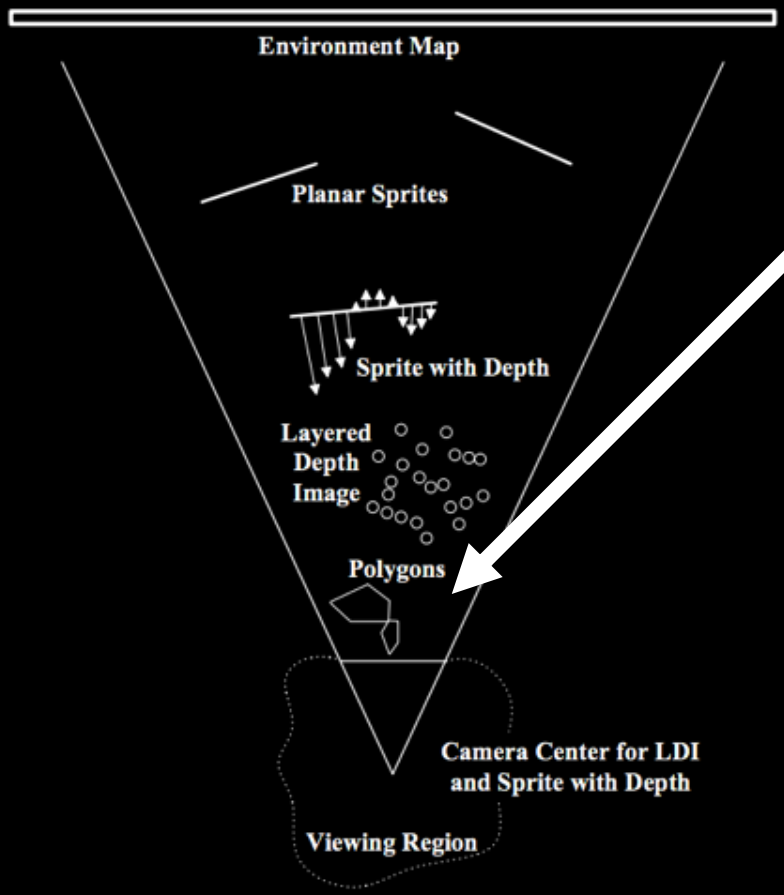


Figure 1 Different image based primitives can serve well depending on distance from the camera

The object is right in front of you and you want to look all around it.

Shade, et. al. SIGGRAPH 98.

Depends on your application…

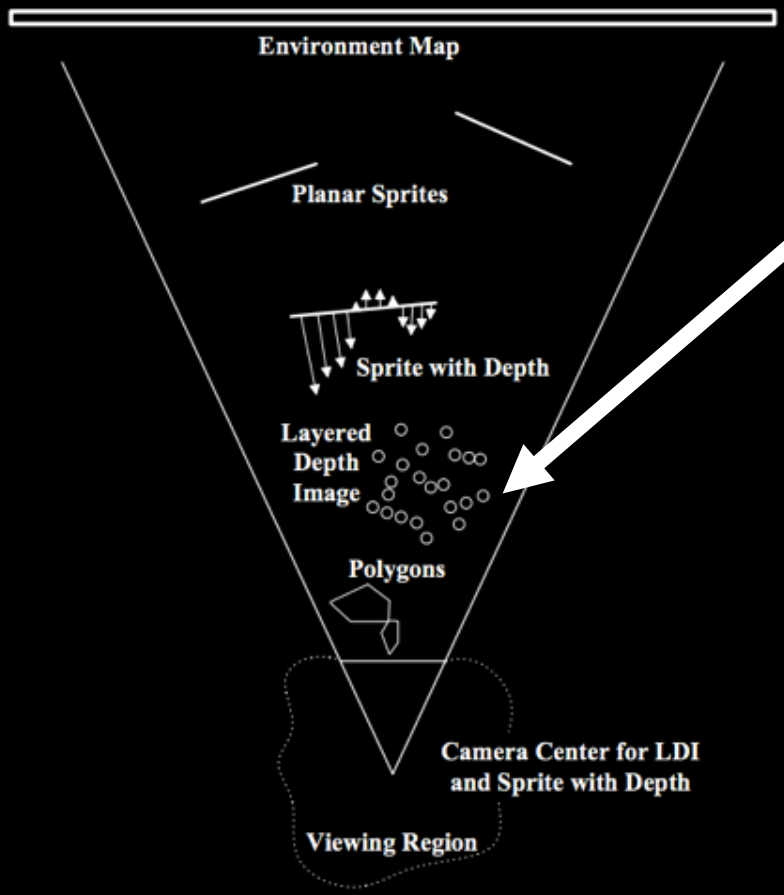


Figure 1 Different image based primitives can serve well depending on distance from the camera

Obvious dis-occlusions if a single depth map is used.

Shade, et. al. SIGGRAPH 98.


Dis-occlusion

Depends on your application…

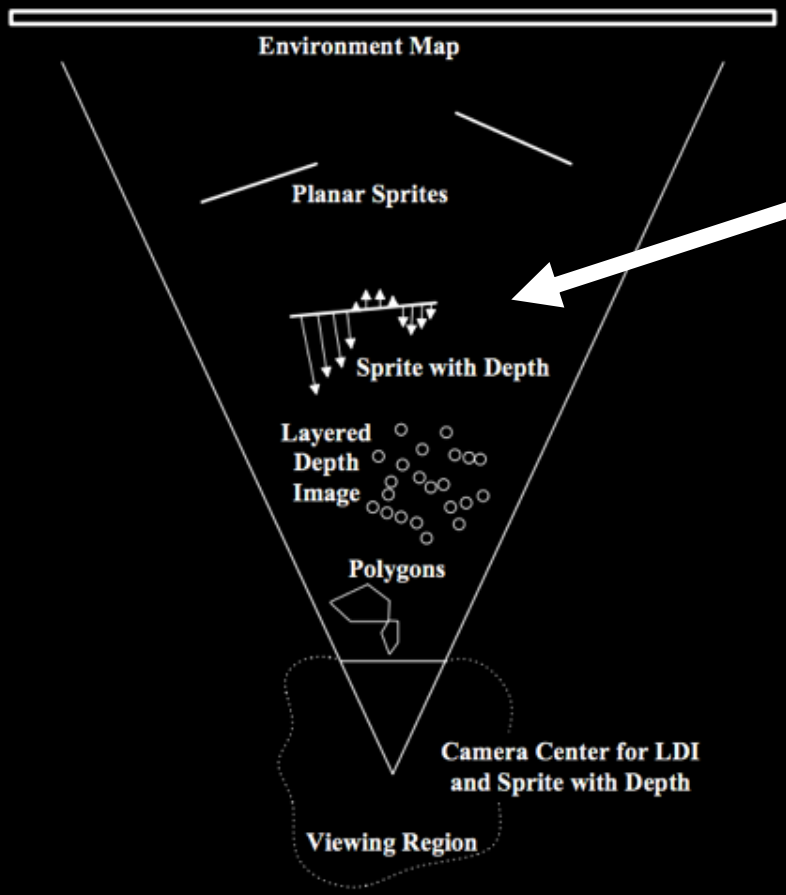


Figure 1 Different image based primitives can serve well depending on distance from the camera

Close enough that you can tell it shouldn't be flat, but **dis-occlusions** are minimal.

Shade, et. al. SIGGRAPH 98.

Depends on your application…

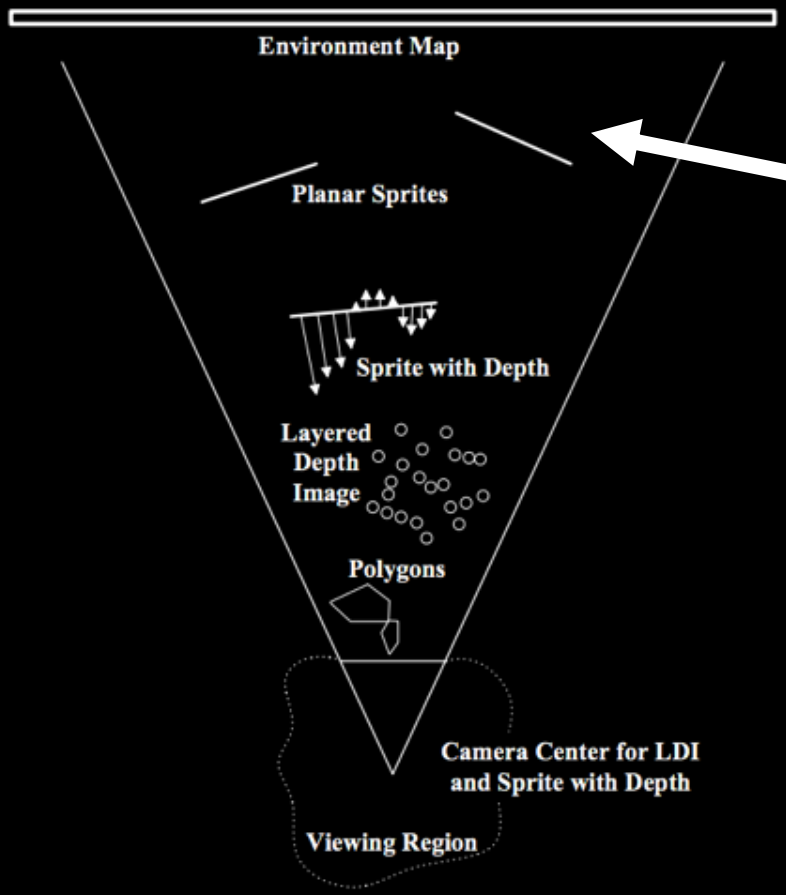


Figure 1 Different image based primitives can serve well depending on distance from the camera

Far enough away that a plane can approximate it.

Shade, et. al. SIGGRAPH 98.

Depends on your application…

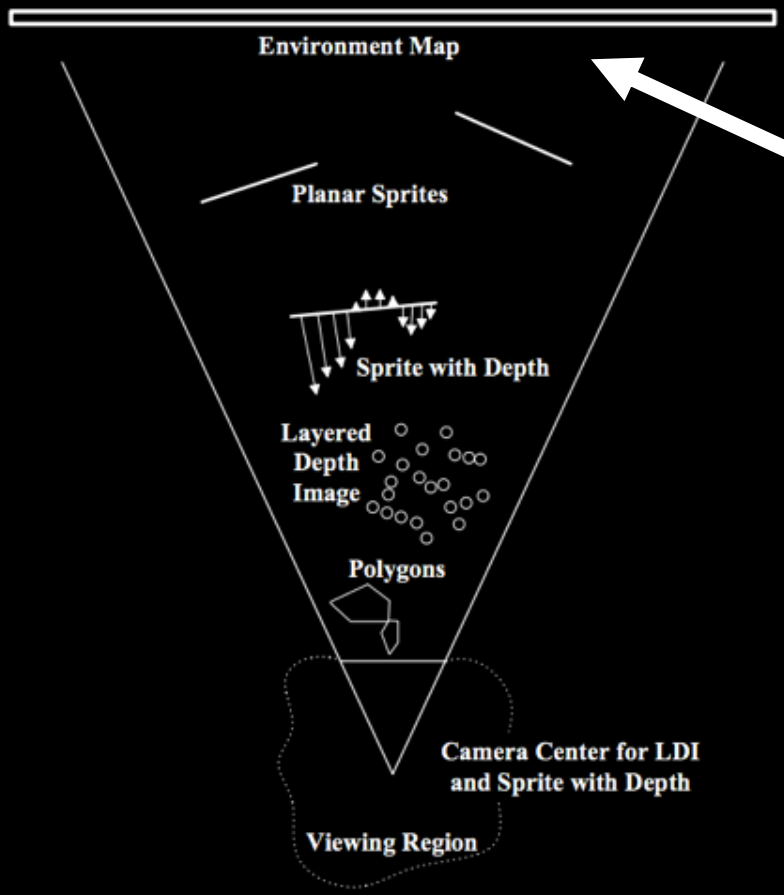


Figure 1 Different image based primitives can serve well depending on distance from the camera

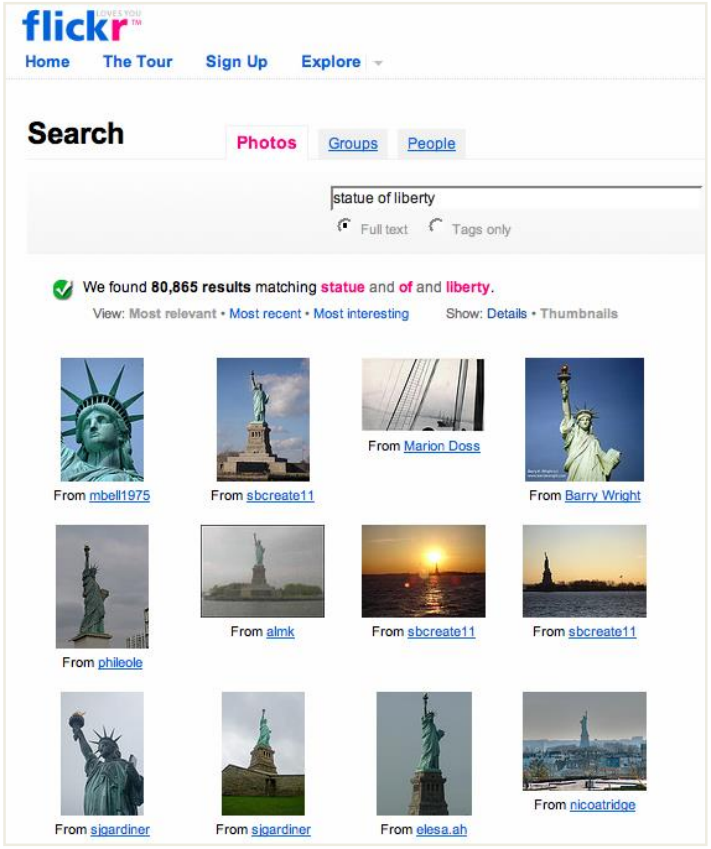
So far away it's effectively at infinity.

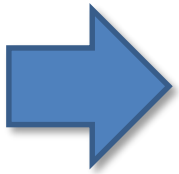
Shade, et. al. SIGGRAPH 98.

Questions?

Multi-view stereo from Internet Collections

[Goesele, Snavely, Curless, Hoppe, Seitz, ICCV 2007]





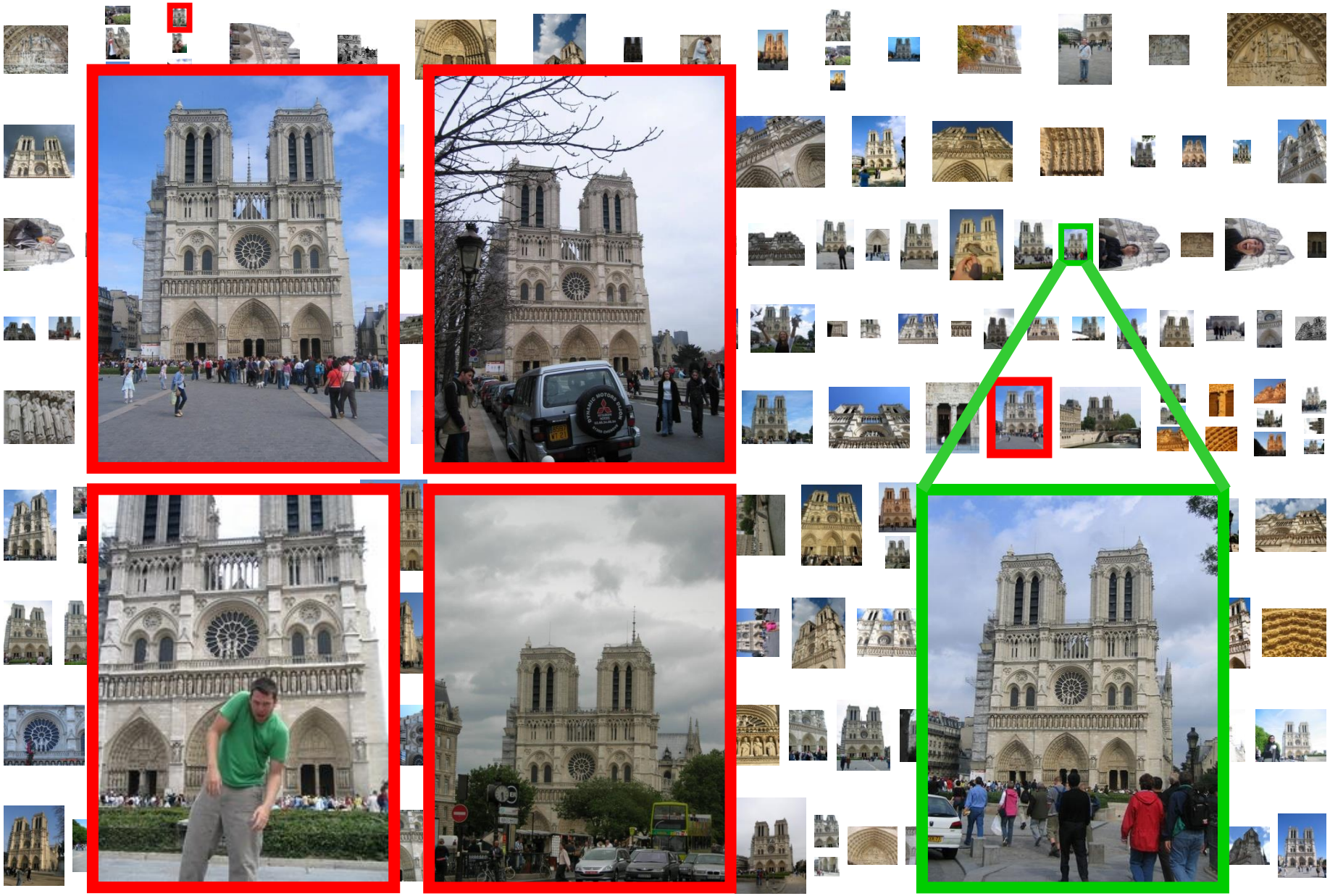
Challenges

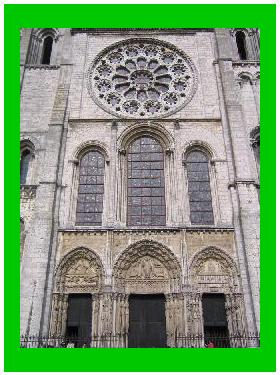
- appearance variation

- resolution

- massive collections

82,754 results for photos matching **notre** and **dame** and **paris**.





4 best neighboring views






reference view

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





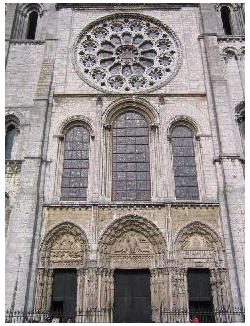
4 best neighboring views





reference view

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



4 best neighboring views



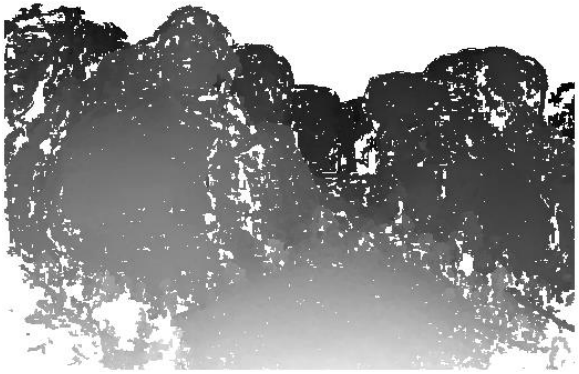


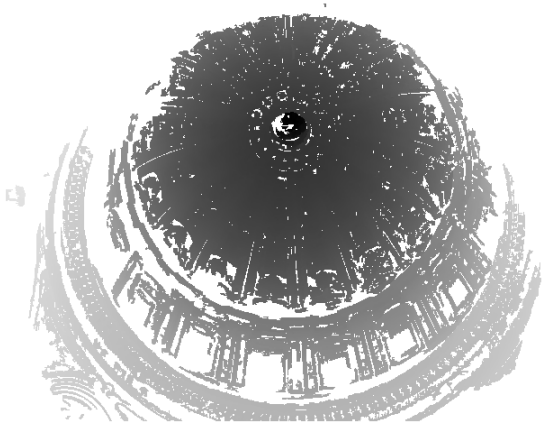

reference view

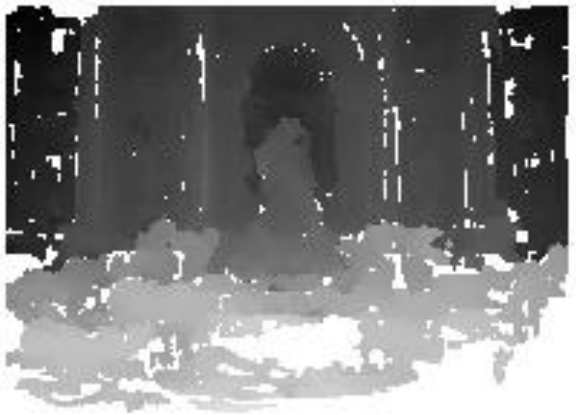
Local view selection

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

Results

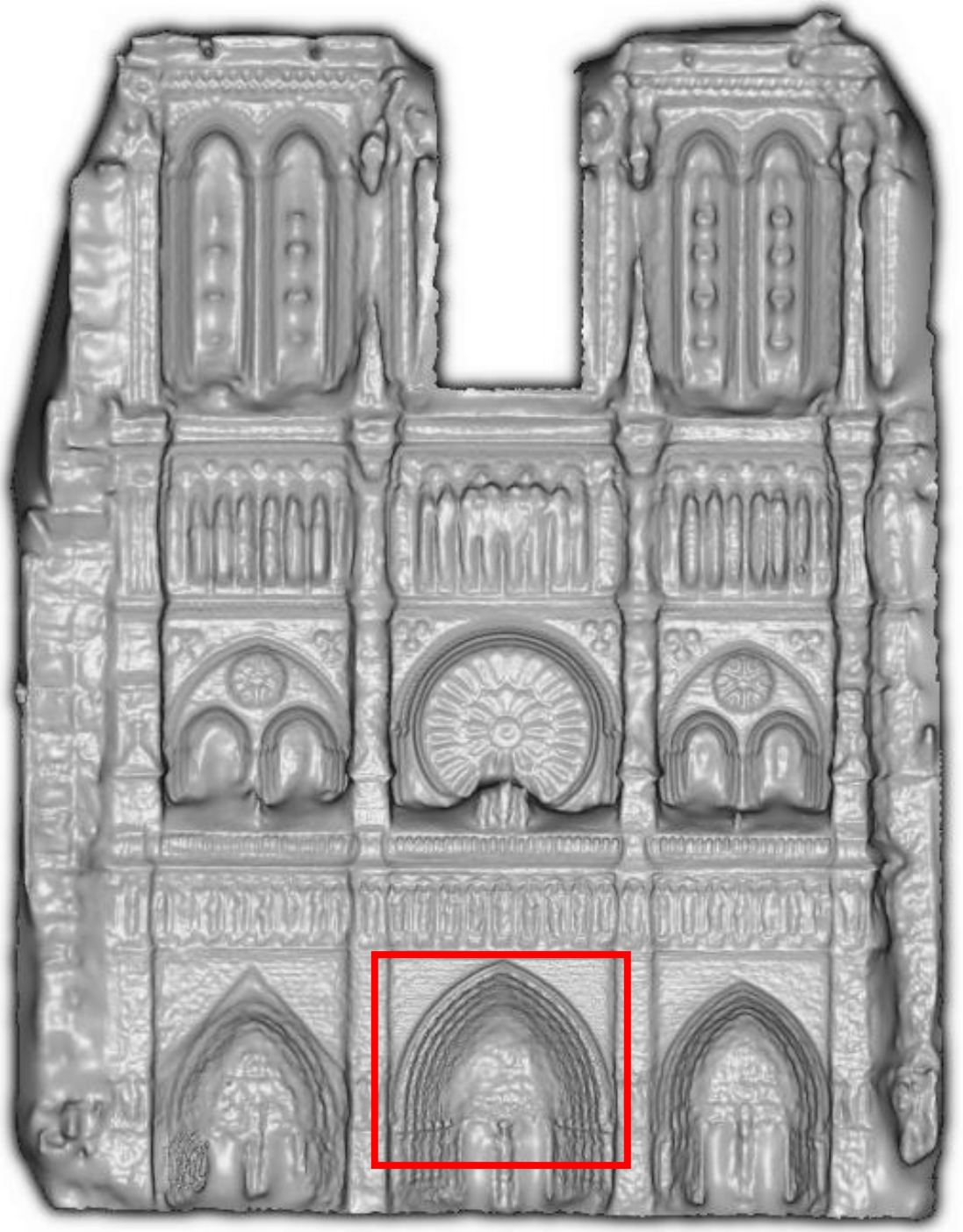


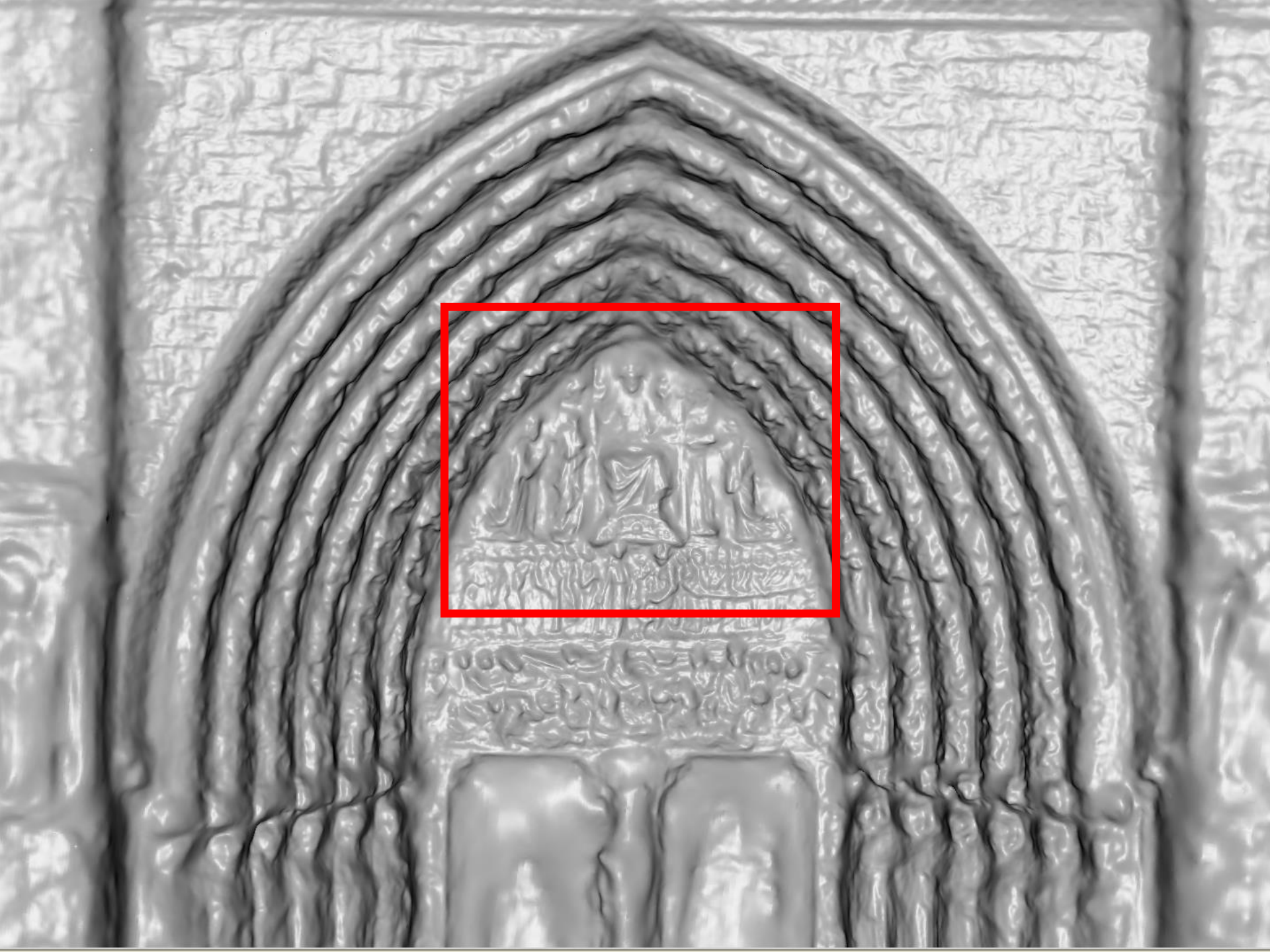


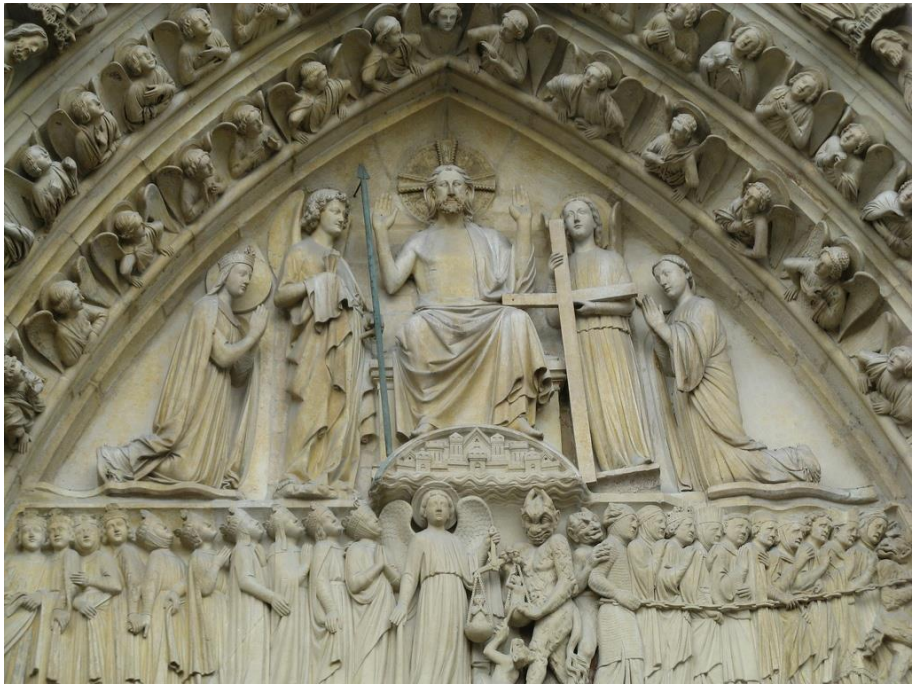


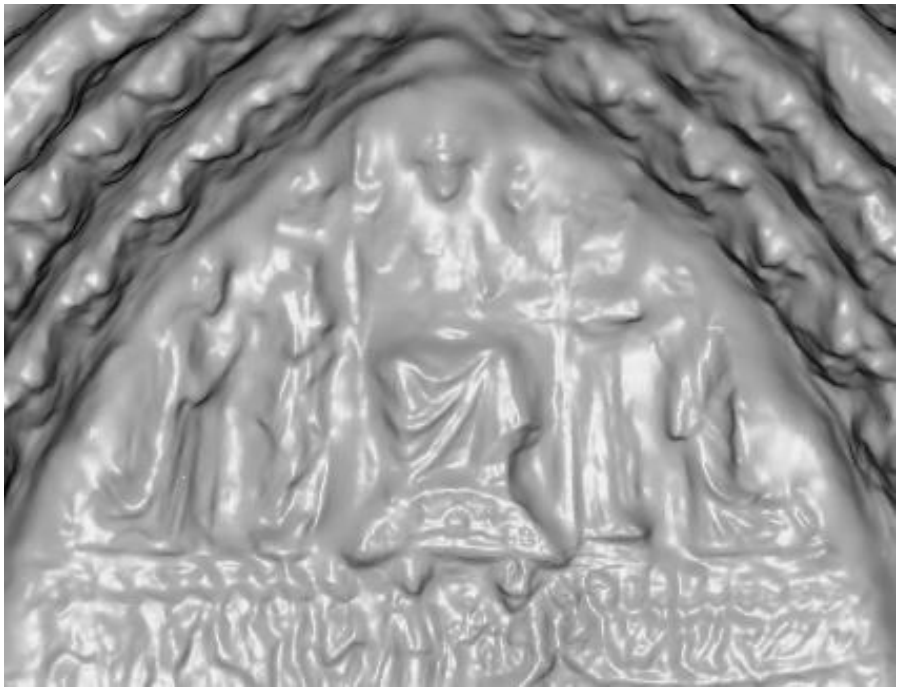
Notre Dame de Paris

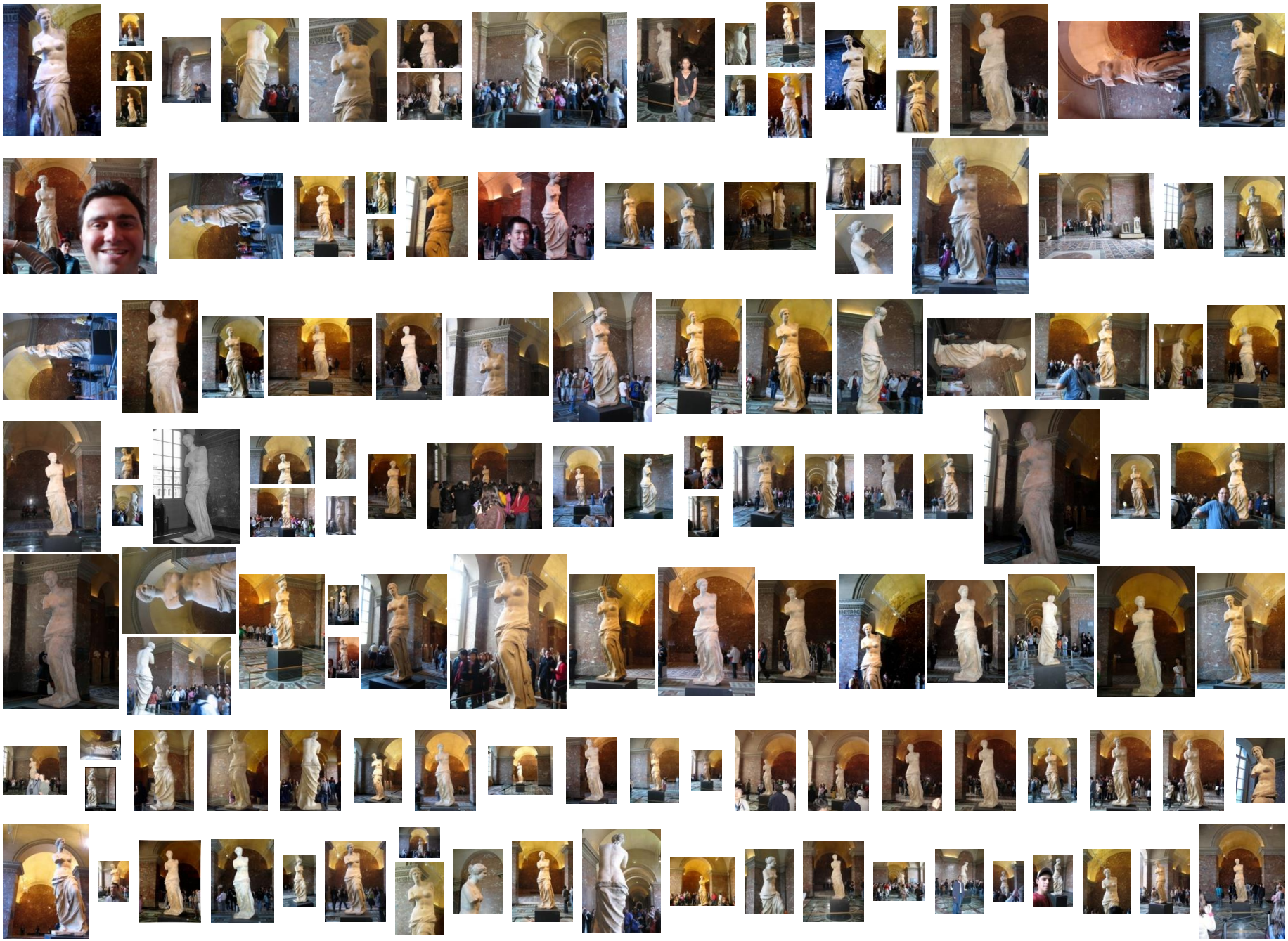
653 images
313 photographers



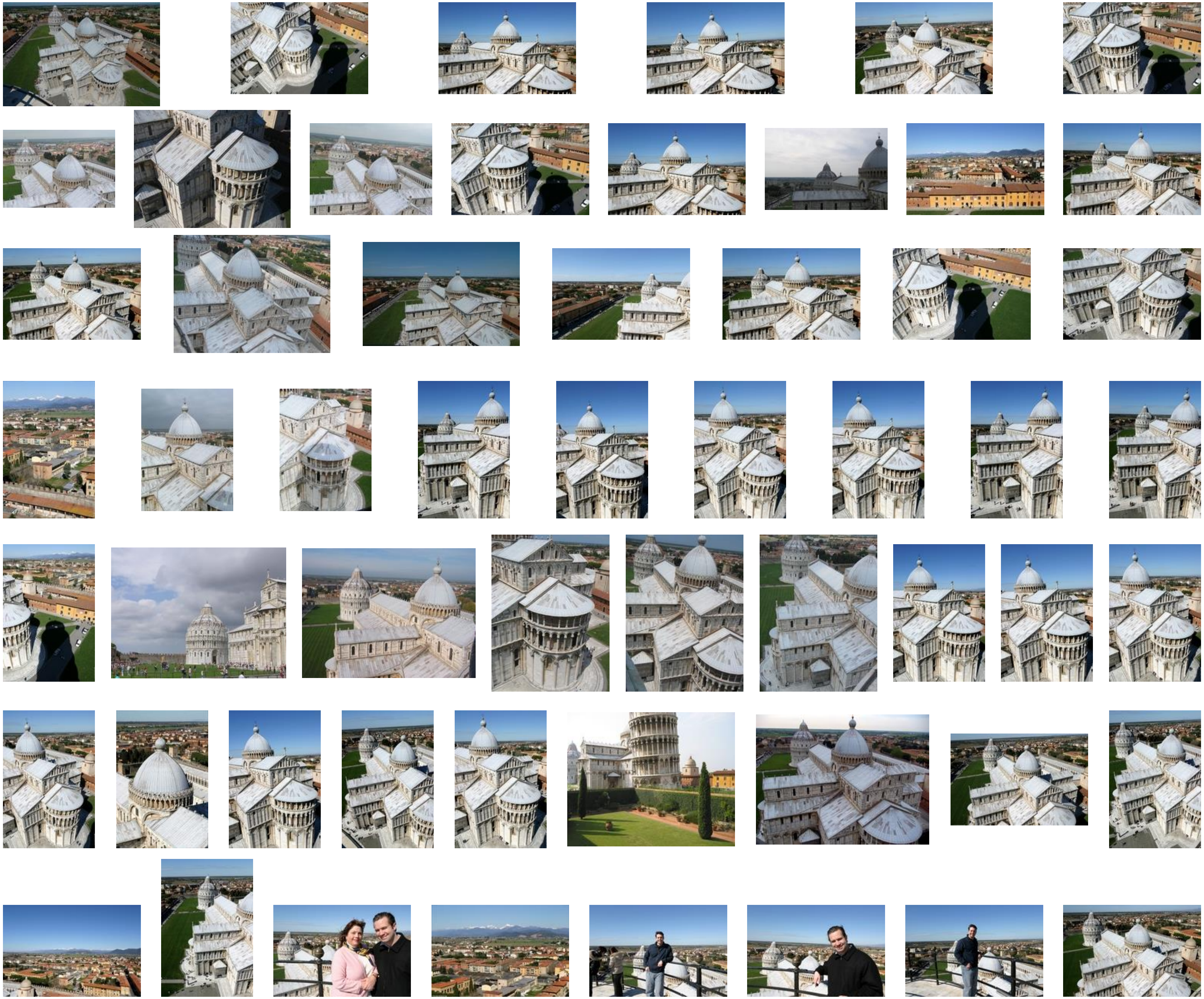


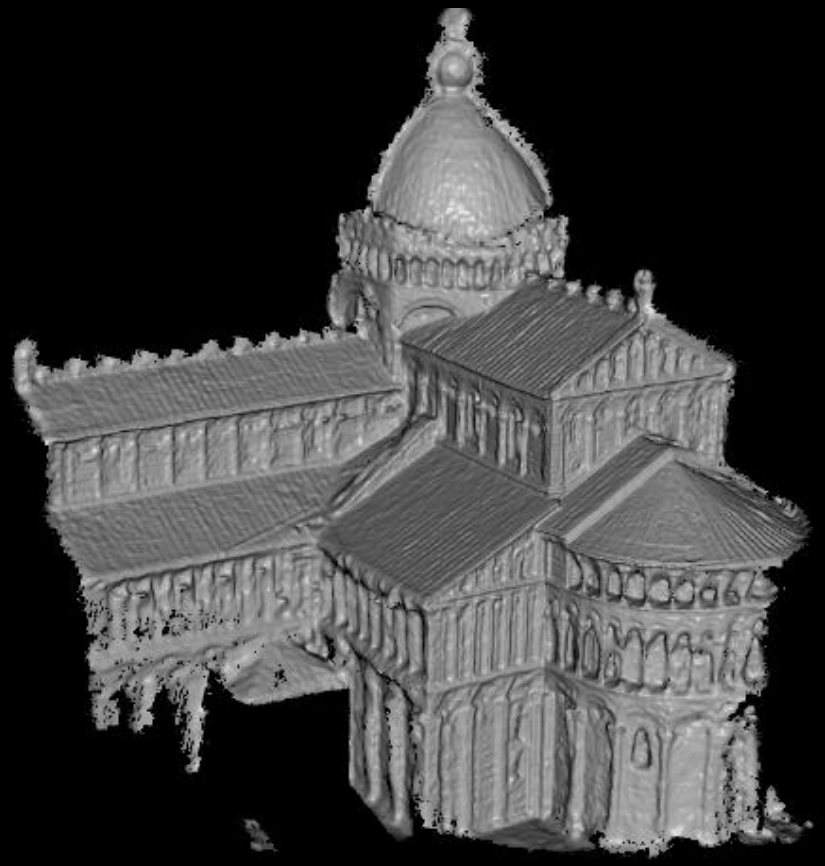




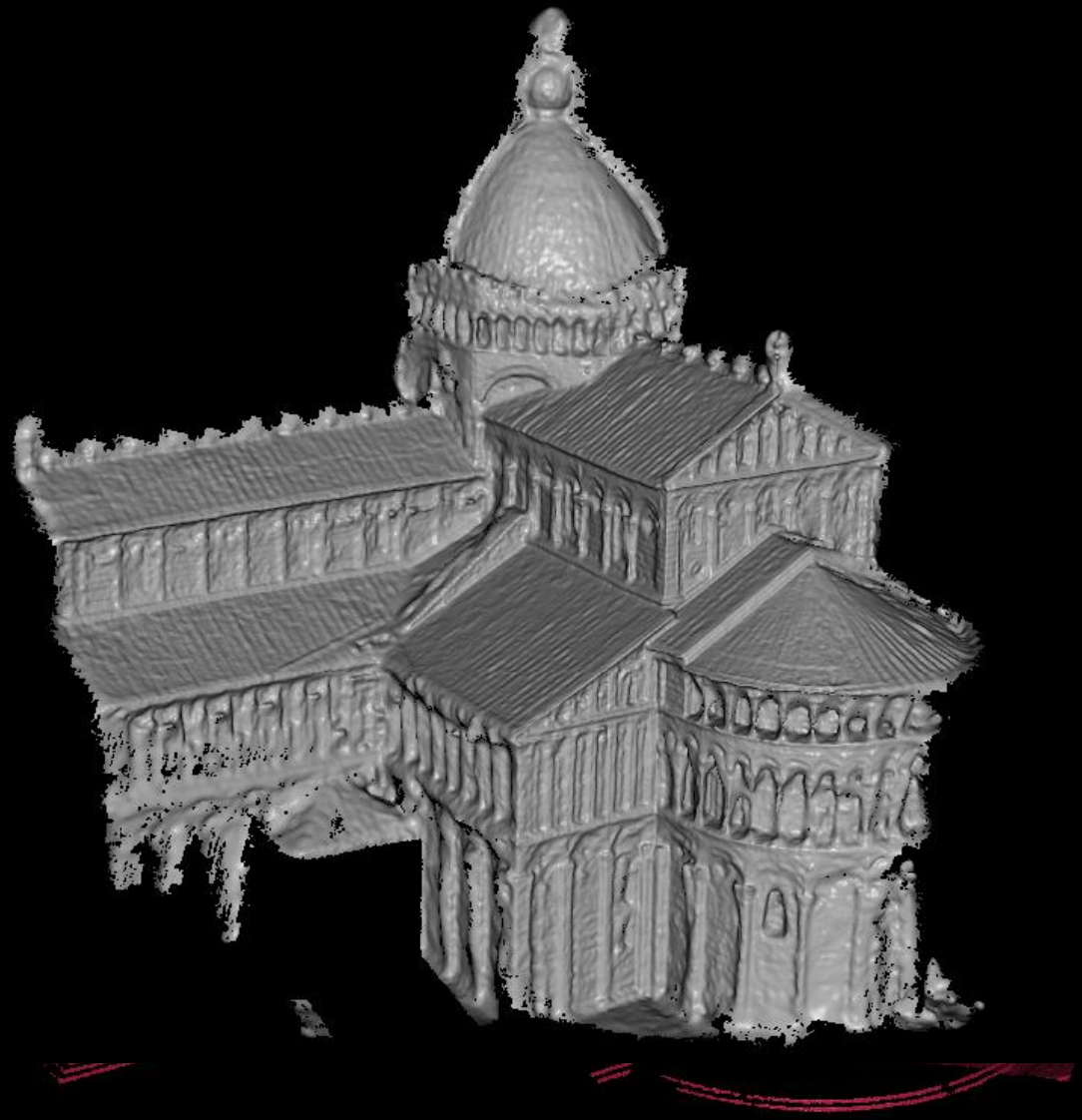


merged model of Venus de Milo





merged model of Pisa Cathedral



Accuracy compared to laser scanned model:
90% of points within *0.25%* of ground truth