C280, Computer Vision

Prof. Trevor Darrell
trevor@eecs.berkeley.edu

Lecture 18: Multiview and Photometric Stereo

Today

- Multiview stereo revisited
- Shape from large image collections
- Voxel Coloring
- Digital Forensics
- Photometric Stereo

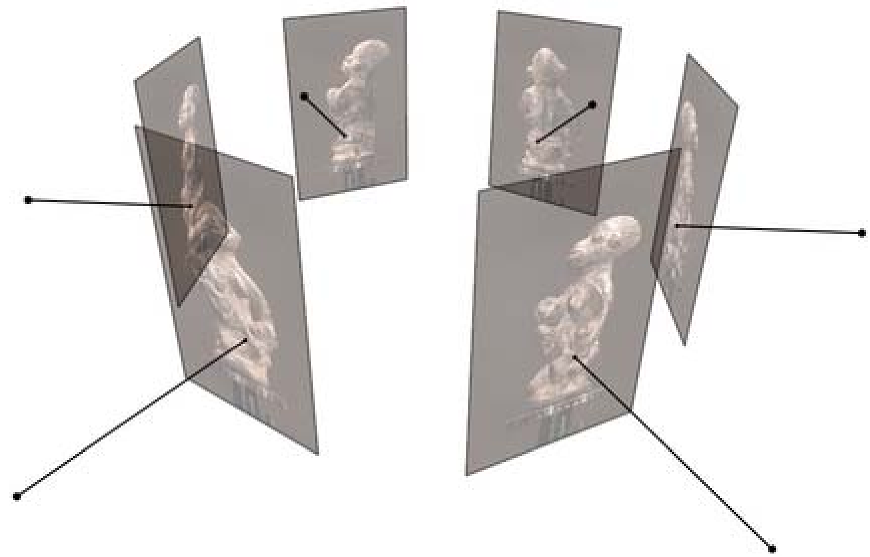
Multi-view Stereo

[Seitz]

Multi-view Stereo

Input: calibrated images from several viewpoints

Output: 3D object model

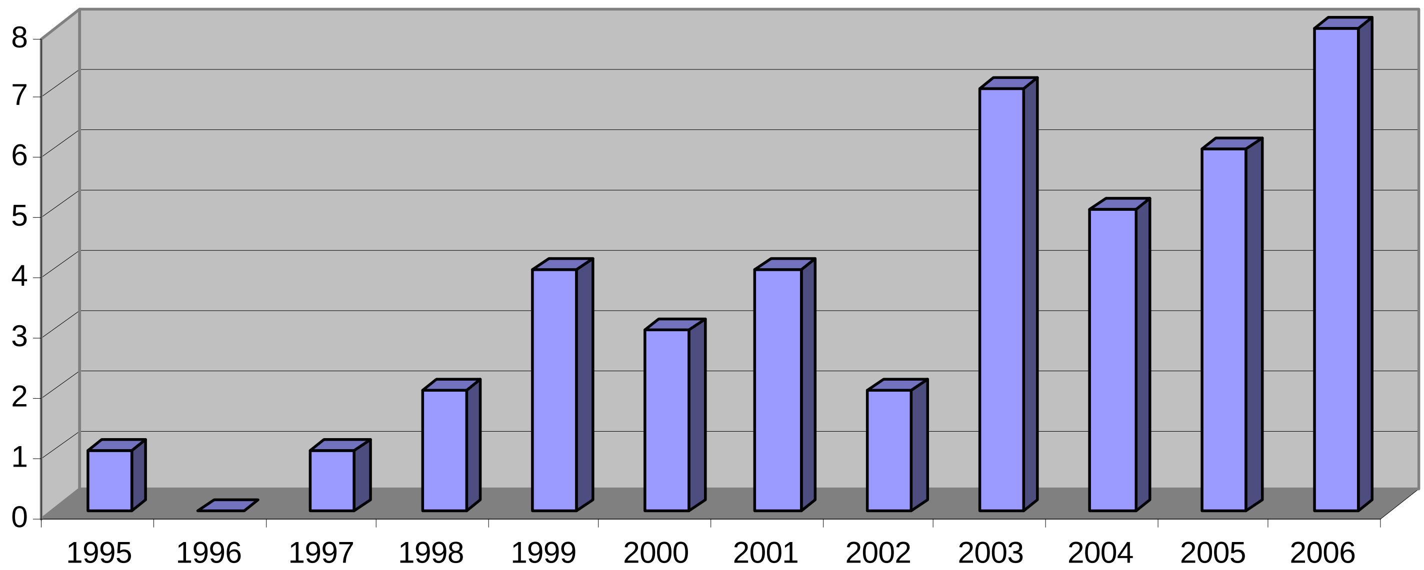


Figures by Carlos Hernandez

[Seitz]

History

number of papers in CVPR, ECCV, and ICCV, by year



Fua
Oriented
particles

Fromherz & Bichsel
shape from multiple cues,
ICYCS,
(precursor to Kutulakos &
Seitz, space carving)

Seitz & Dyer
Voxel coloring

Faugeras & Keriven
Level set stereo

Narayanan & Kanade
Virtualized Reality

Kutulakos & Seitz
Space carving

Kolmogorov & Zabih
multi-view graph cut

Hernandez

Pons

Vogiatzis

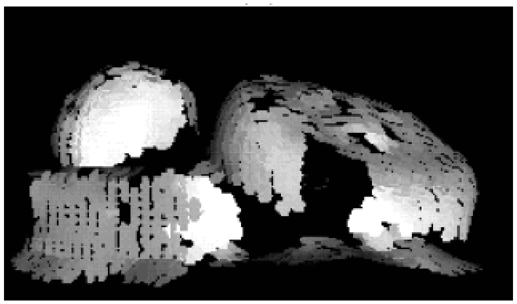
Tran

Hornung

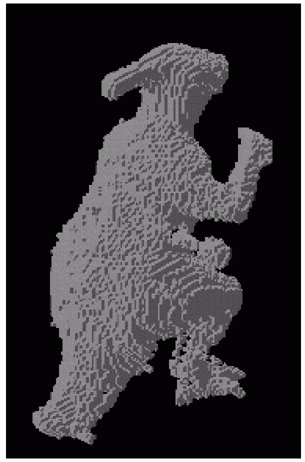
Goesele

Furukawa

[Seitz]



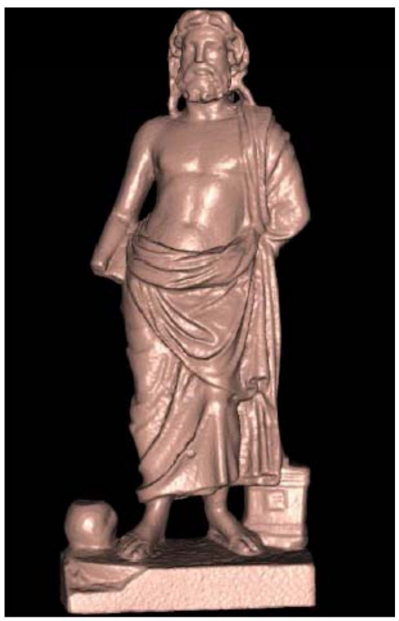
Fua
1995



Seitz, Dyer
1997

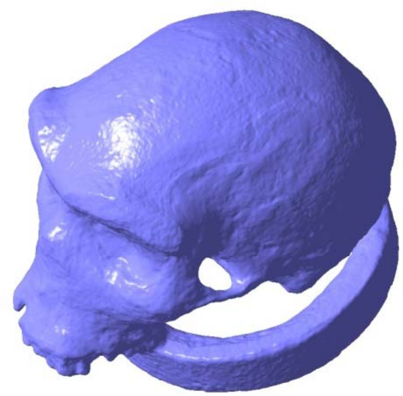
Narayanan, Rander, Kanade
1998

Faugeras, Keriven
1998

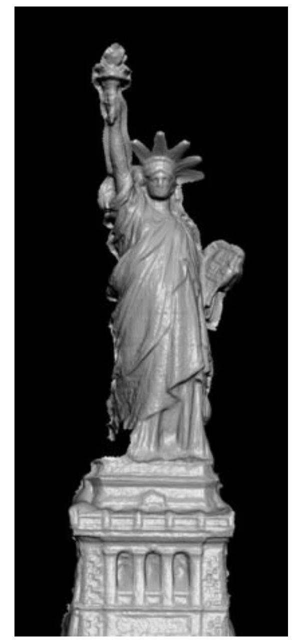


Hernandez, Schmitt
2004

Pons, Keriven, Faugeras
2005



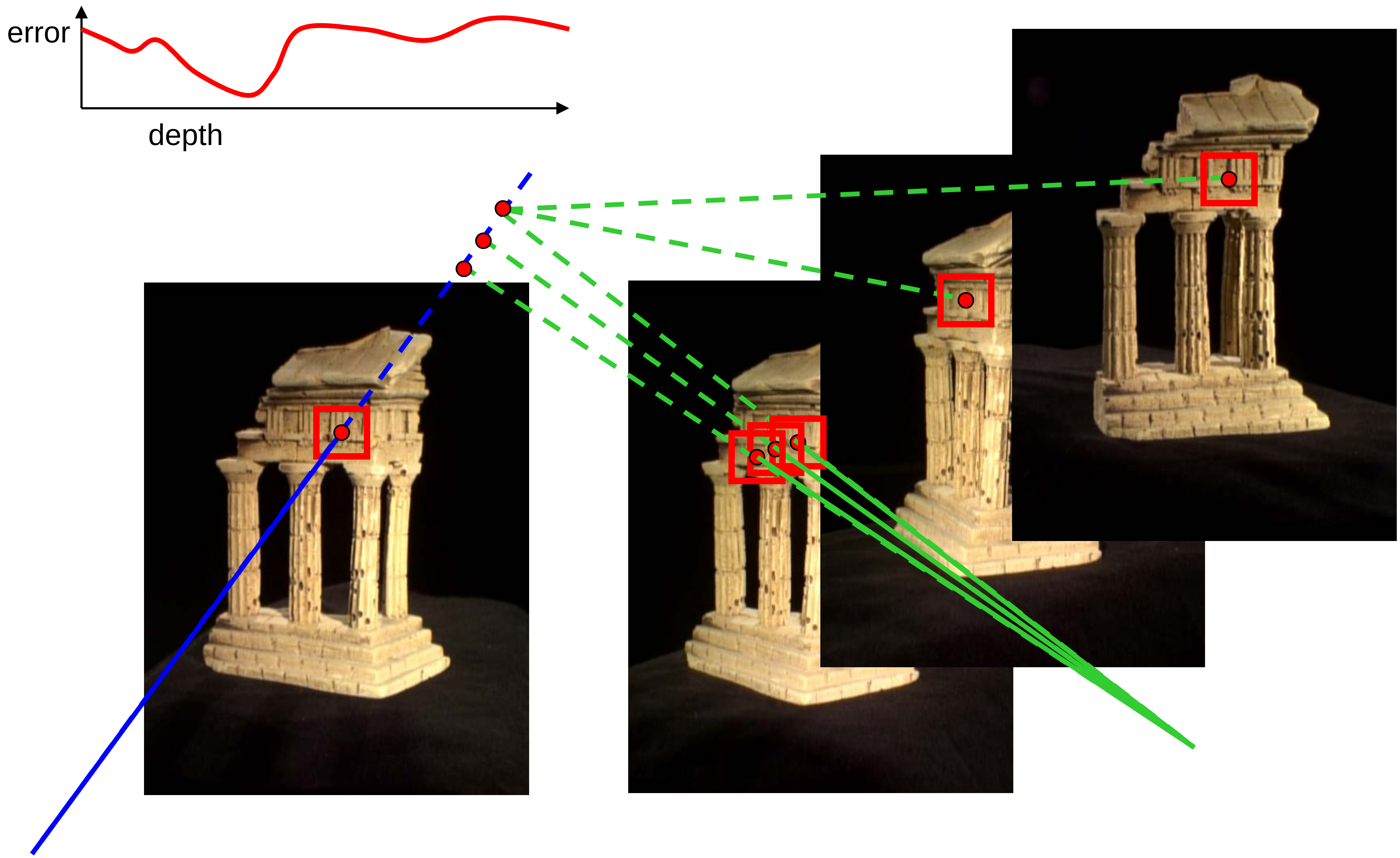
Furukawa, Ponce
2006



Goesele et al.
2007

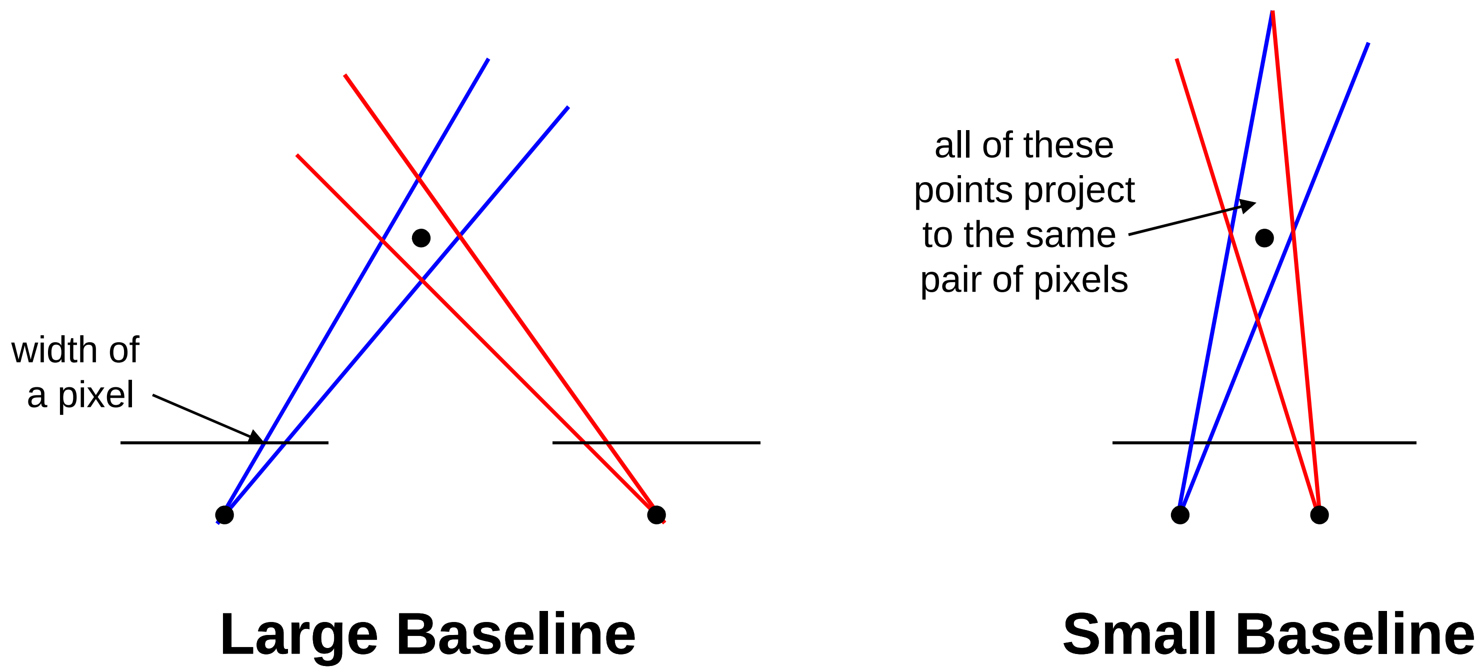
[Seitz]

Stereo: basic idea



[Seitz]

Choosing the stereo baseline



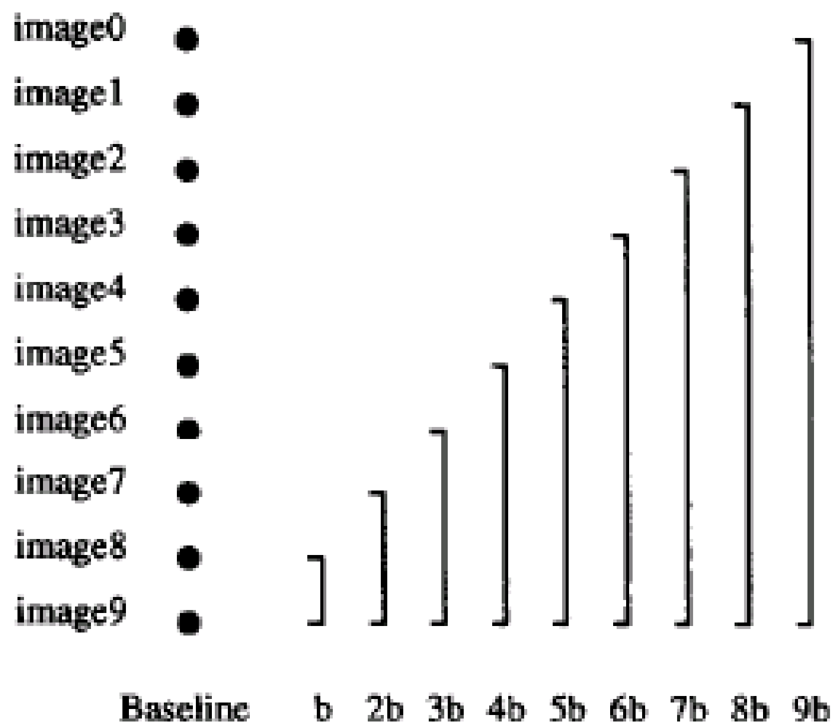
What's the optimal baseline?

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

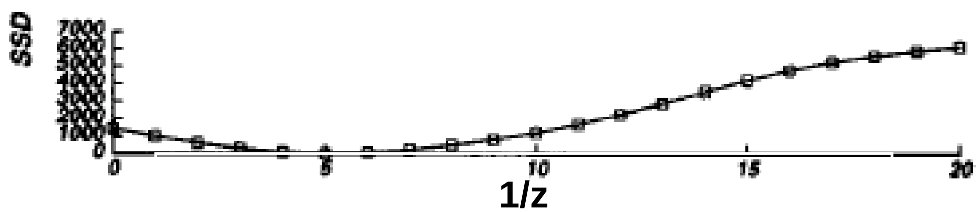
[Seitz]

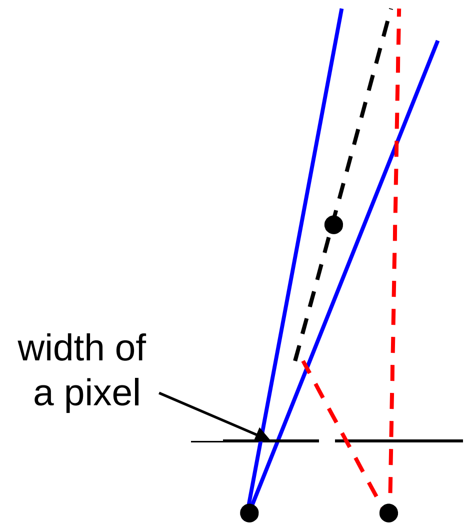
The Effect of Baseline on Depth Estimation

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

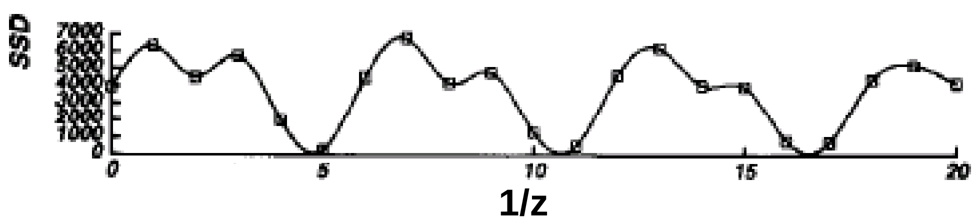


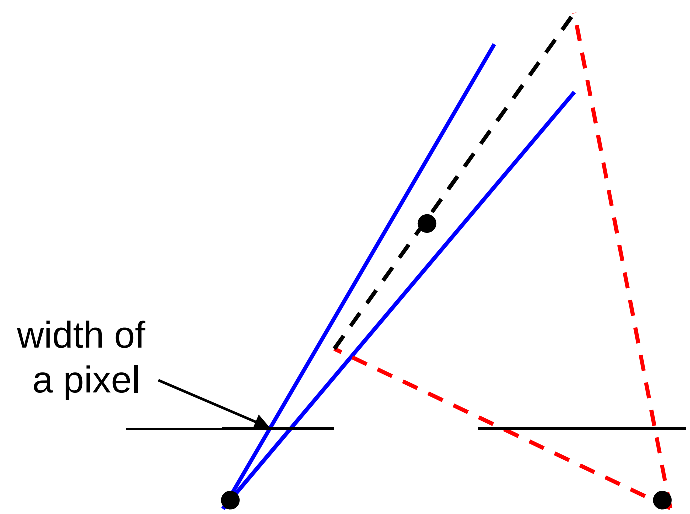
[Seitz]





pixel matching score





[Seitz]

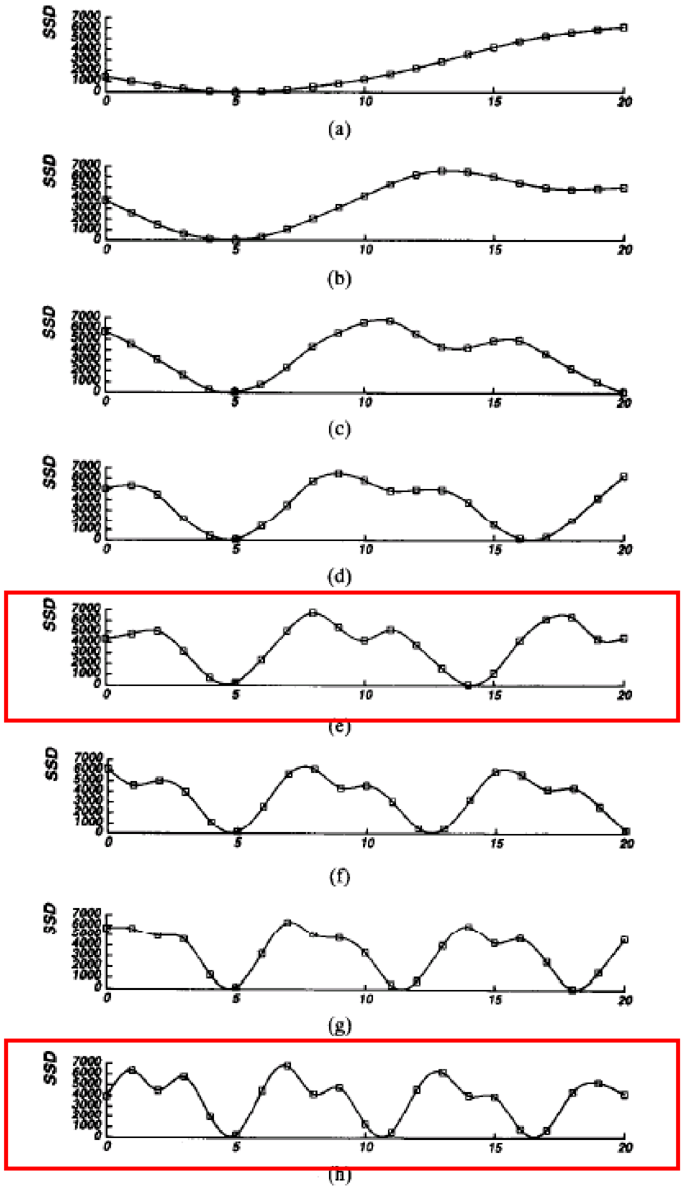


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

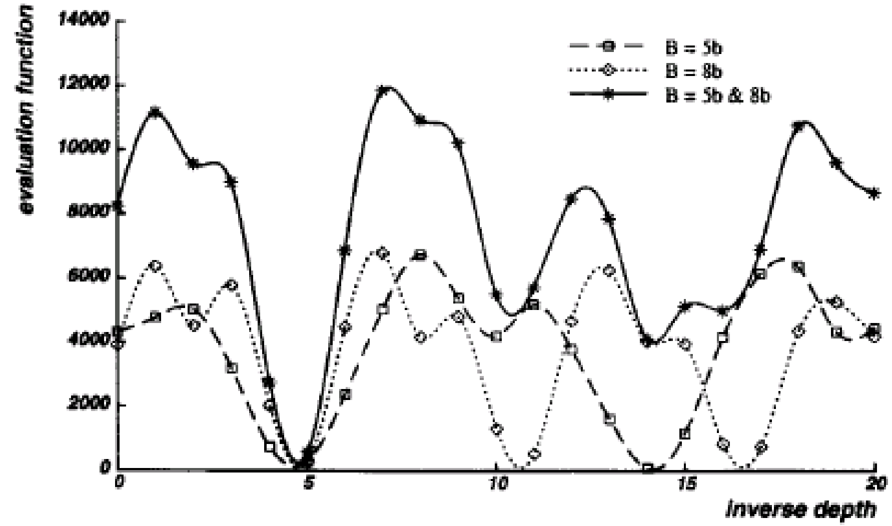


Fig. 6. Combining two stereo pairs with different baselines.

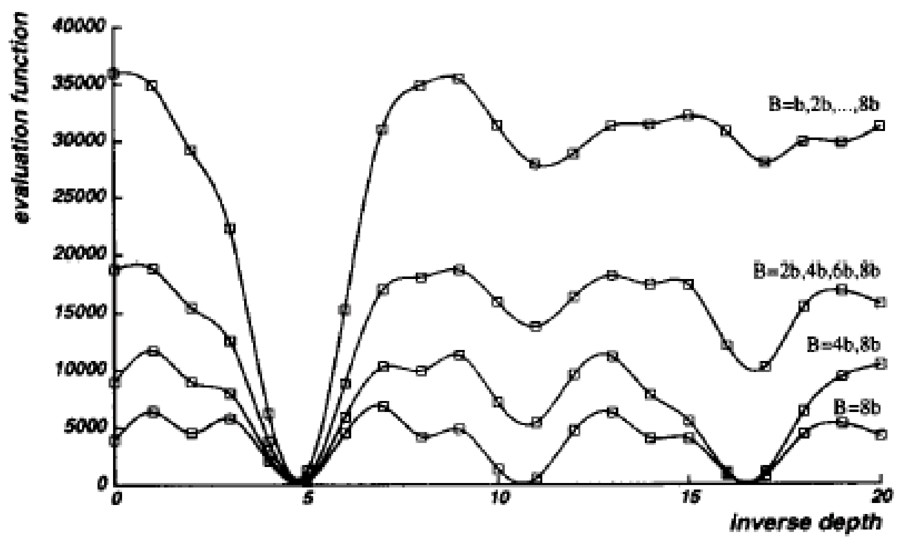


Fig. 7. Combining multiple baseline stereo pairs.

[Seitz]

Multibaseline Stereo

Basic Approach

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

Limitations

- Only gives a depth map (not an "object model")
- Won't work for widely distributed views:

[Seitz]

Problem: *visibility*

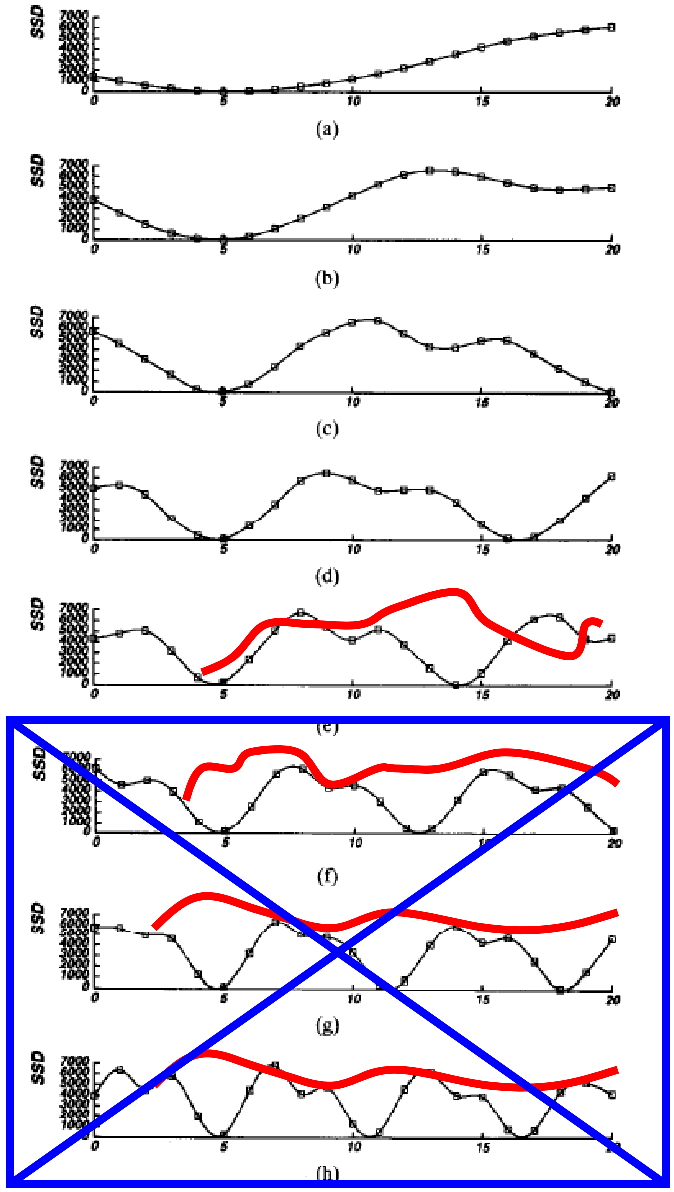


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

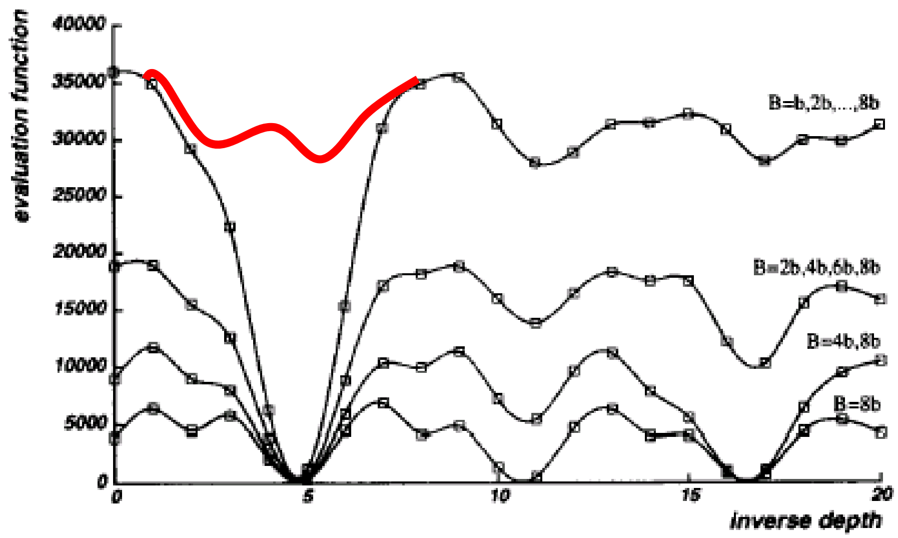


Fig. 7. Combining multiple baseline stereo pairs.

Some Solutions

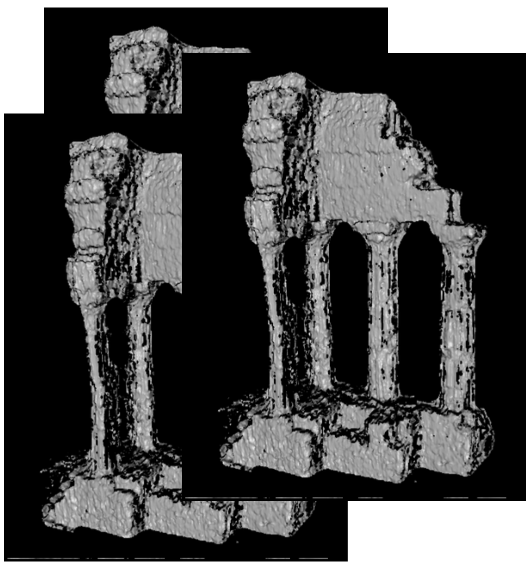
- Match only nearby photos [Narayanan 98]
- Use NCC instead of SSD,
  Ignore NCC values > threshold
  [Hernandez & Schmitt 03]

[Seitz]

Merging Depth Maps

vrip [Curless and Levoy 1996]

- compute weighted average of depth maps

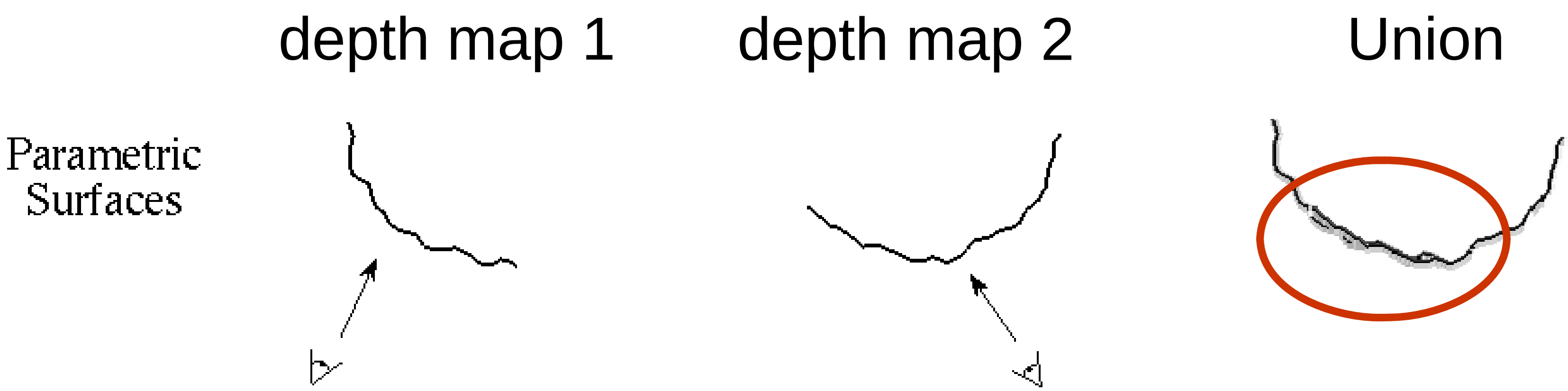


set of depth maps
(one per view)

merged surface
mesh

[Seitz]

Merging depth maps



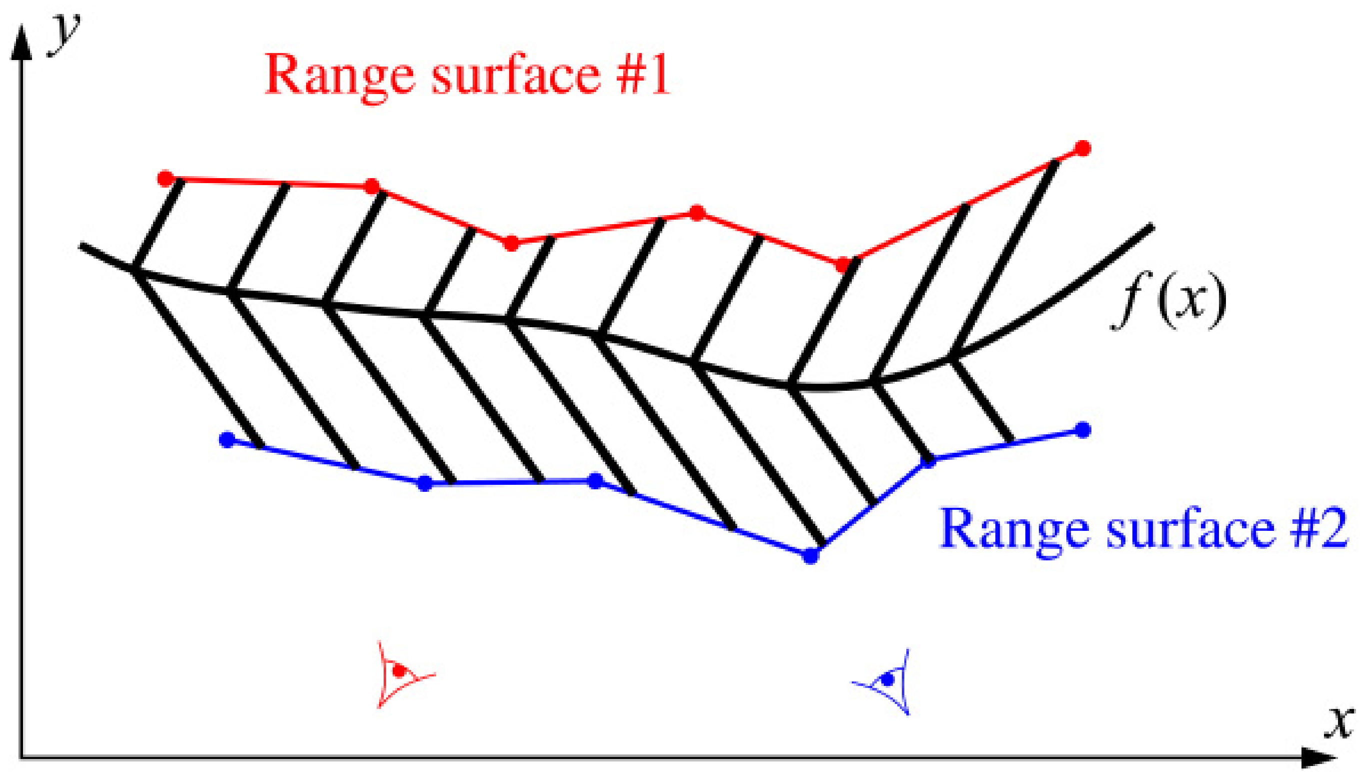
Naïve combination (union) produces artifacts

Better solution: find "average" surface

- Surface that minimizes sum (of squared) distances to the depth maps

[Seitz]

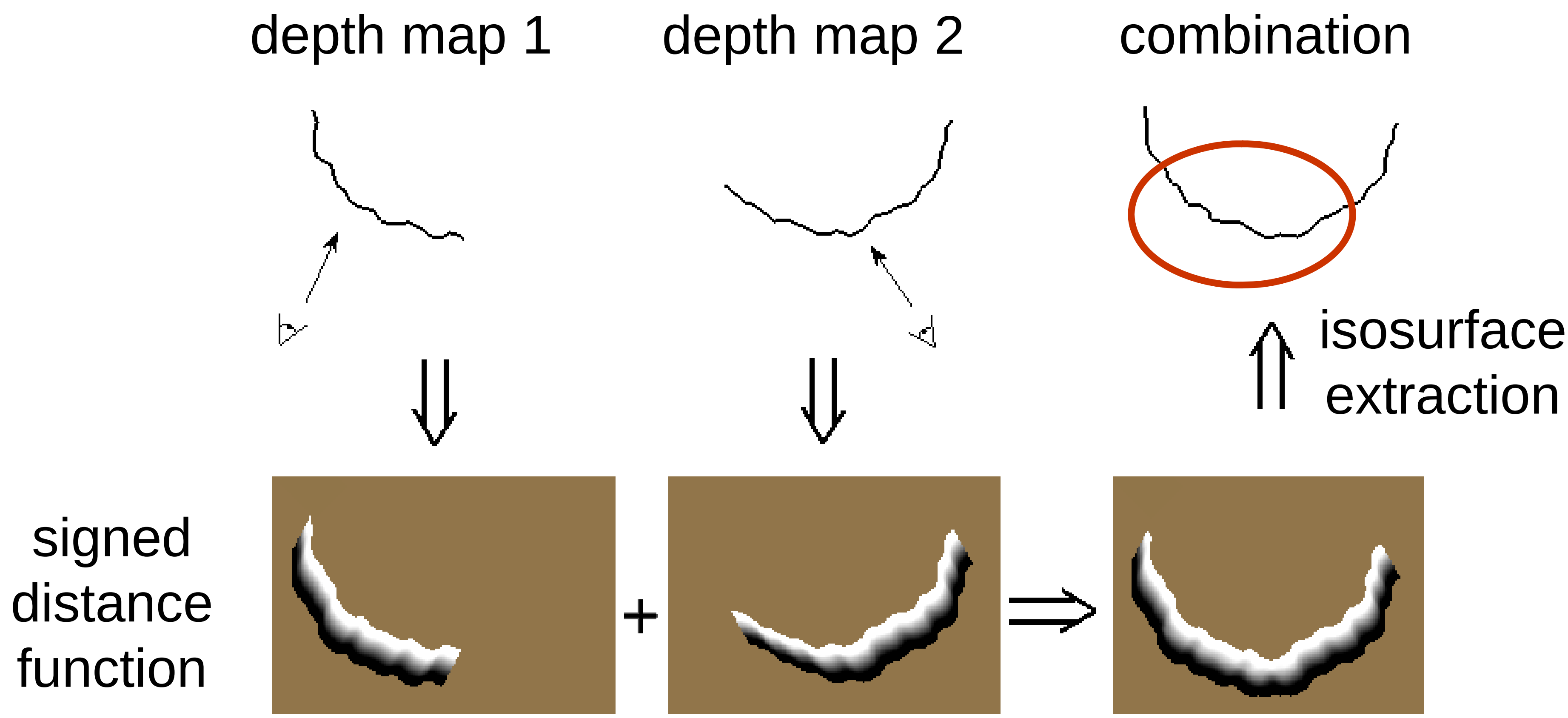
Least squares solution



$$E(f) = \sum_{i=1}^{N} \int d_i^2(x, f)\,dx$$

[Seitz]

VRIP [Curless & Levoy 1996]



[Seitz]

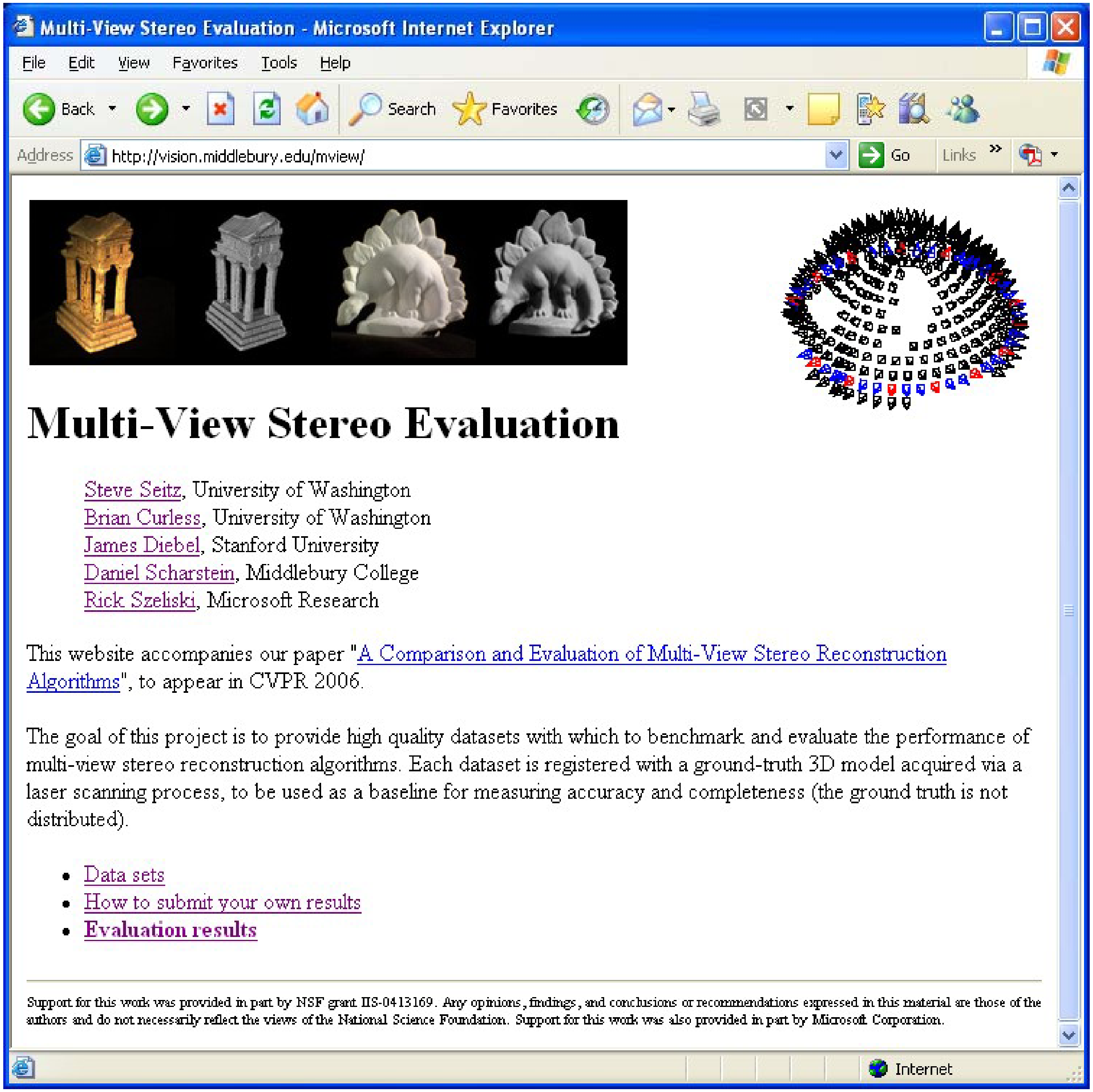
Merging Depth Maps: Temple Model

input image

317 images
(hemisphere)

ground truth model

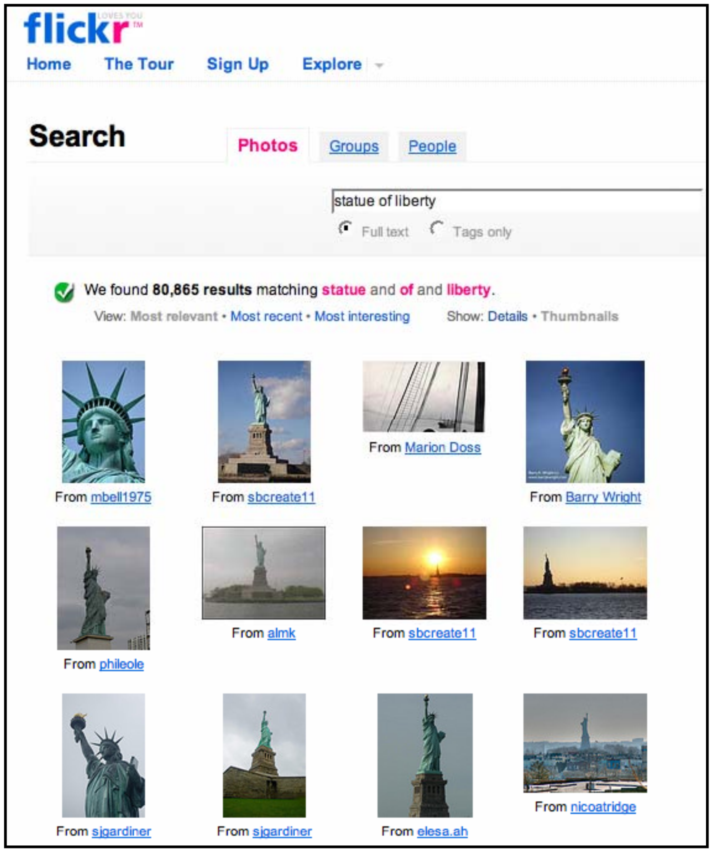
Goesele, Curless, Seitz, 2006


Multi-View Stereo Evaluation - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://vision.middlebury.edu/mview/
Go Links
Multi-View Stereo Evaluation
Steve Seitz, University of Washington
Brian Curless, University of Washington
James Diebel, Stanford University
Daniel Scharstein, Middlebury College
Rick Szeliski, Microsoft Research
This website accompanies our paper "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms", to appear in CVPR 2006.
The goal of this project is to provide high quality datasets with which to benchmark and evaluate the performance of multi-view stereo reconstruction algorithms. Each dataset is registered with a ground-truth 3D model acquired via a laser scanning process, to be used as a baseline for measuring accuracy and completeness (the ground truth is not distributed).
• Data sets
• How to submit your own results
• Evaluation results
Support for this work was provided in part by NSF grant IIS-0413169. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Support for this work was also provided in part by Microsoft Corporation.
Internet

[Seitz]

Multi-view stereo from Internet Collections

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



[Seitz]

Challenges

- appearance variation

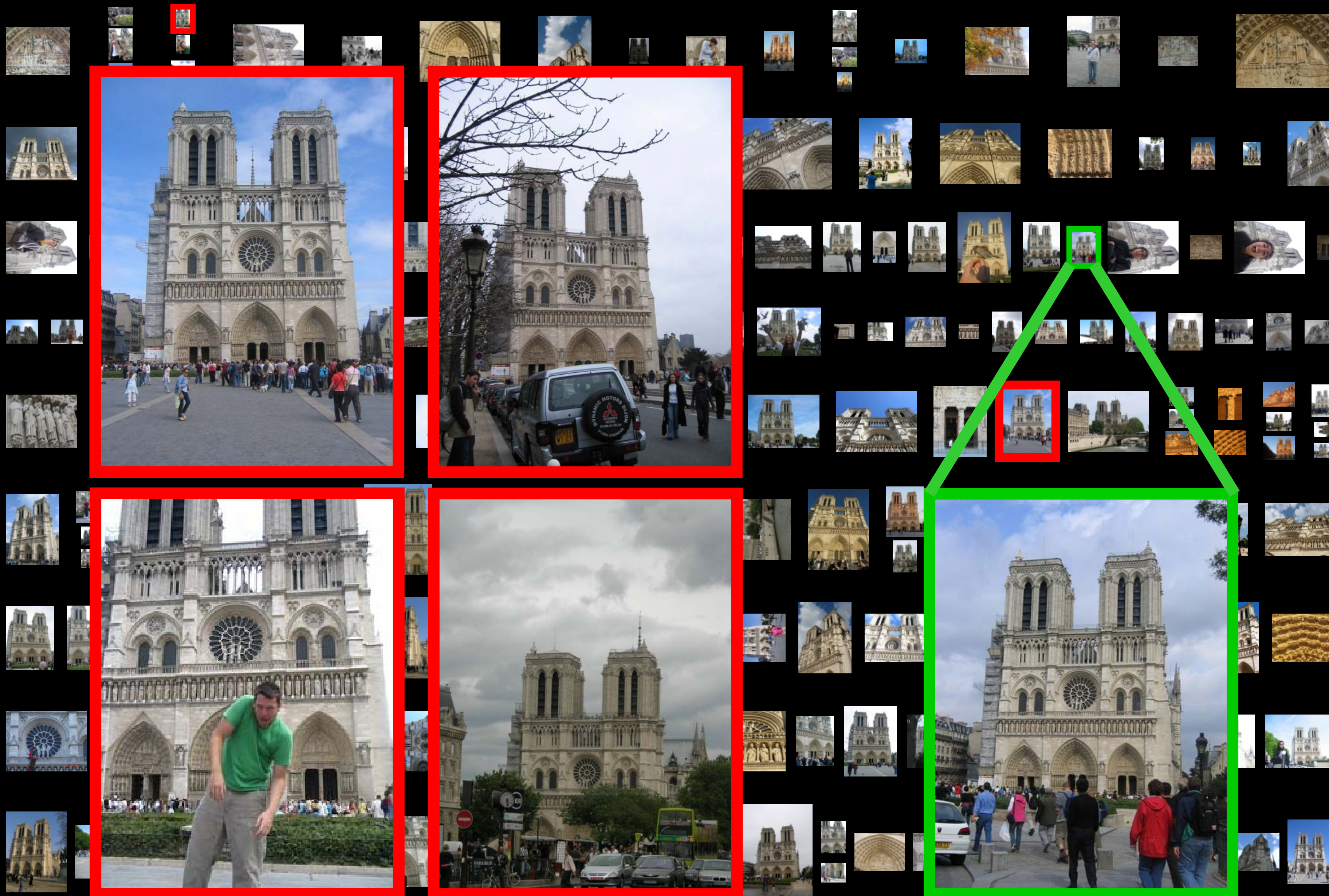
- resolution

- massive collections

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

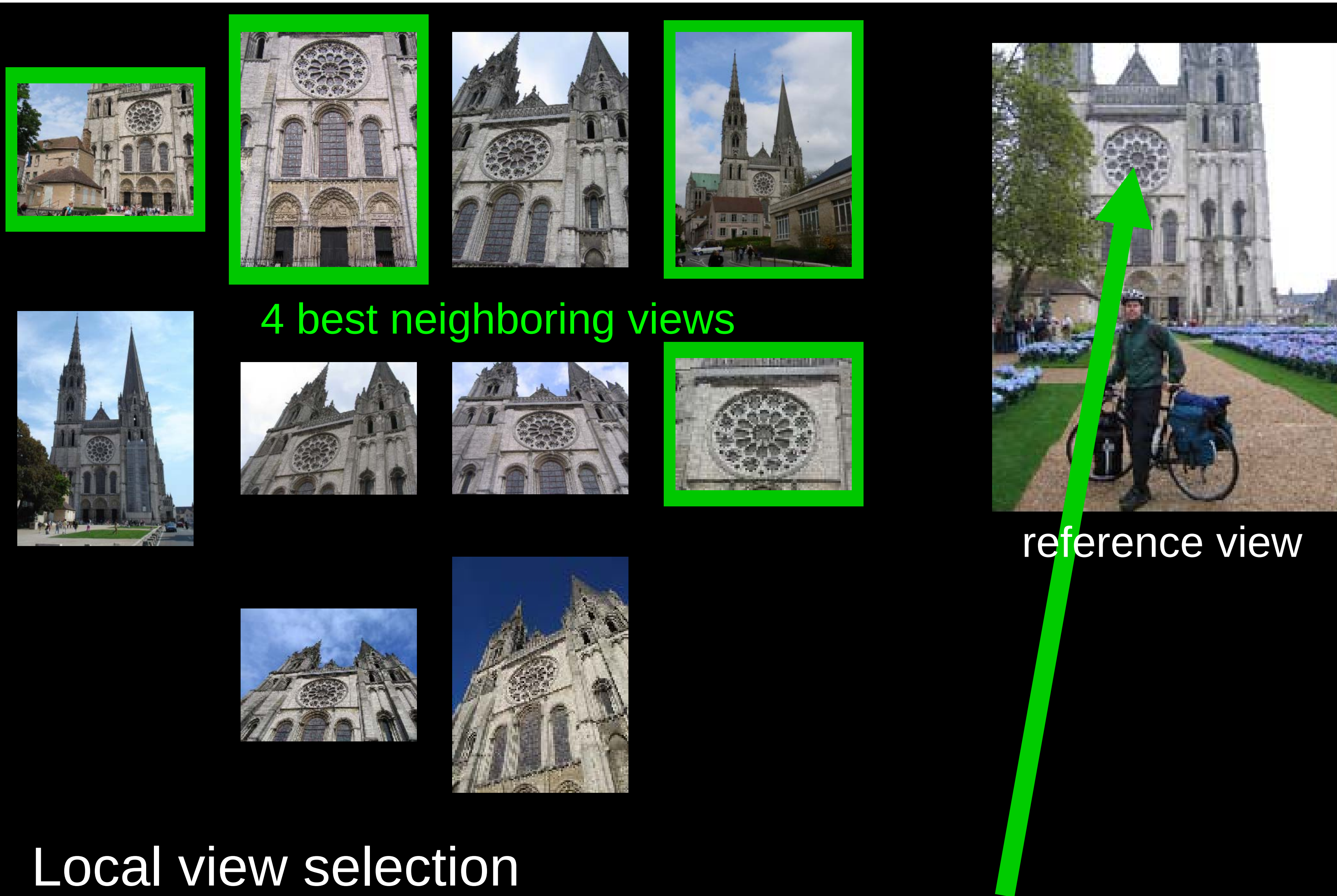
[Seitz]

Large Image Collections



206 *Flickr* images taken by 92 photographers

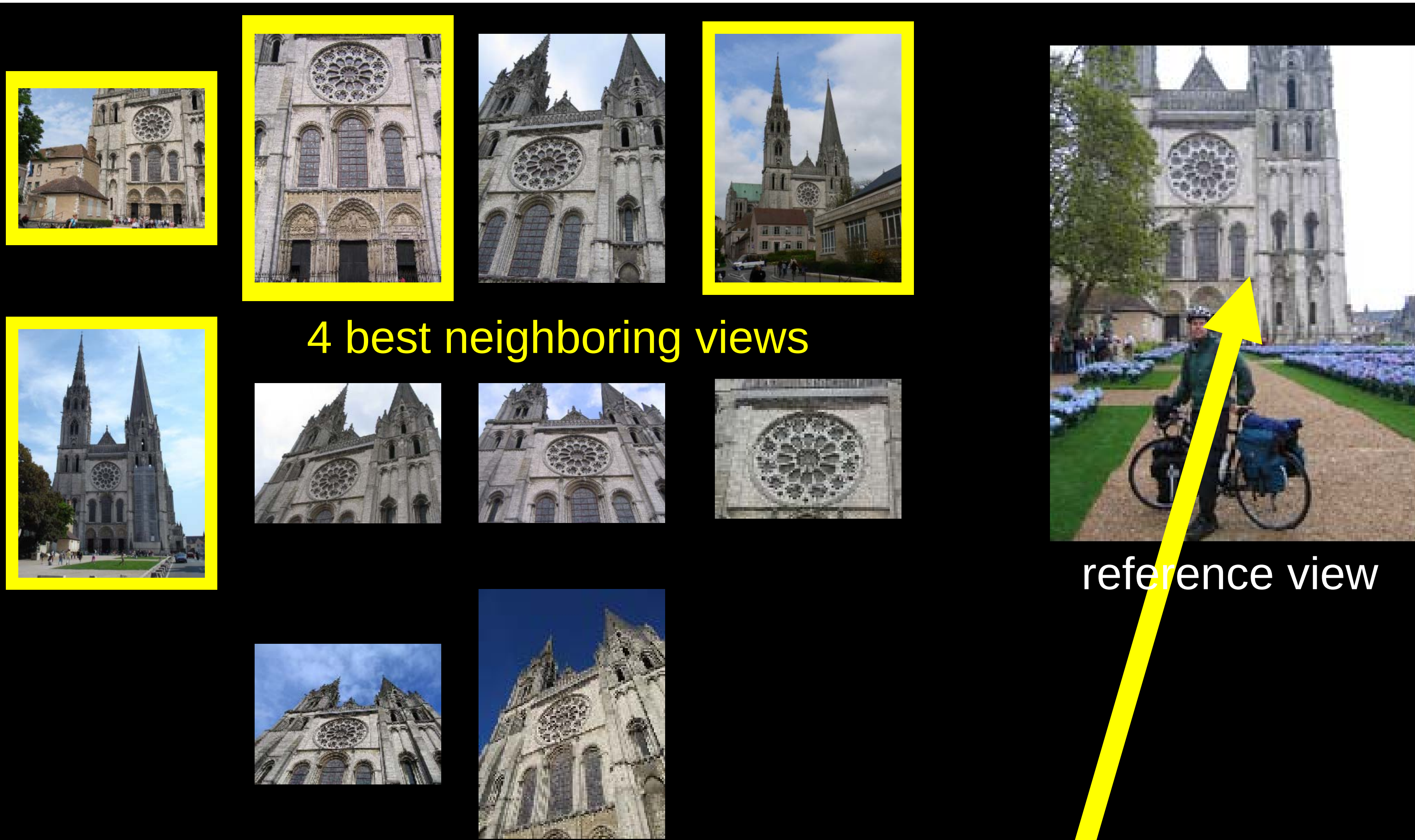
[Seitz]



Local view selection

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

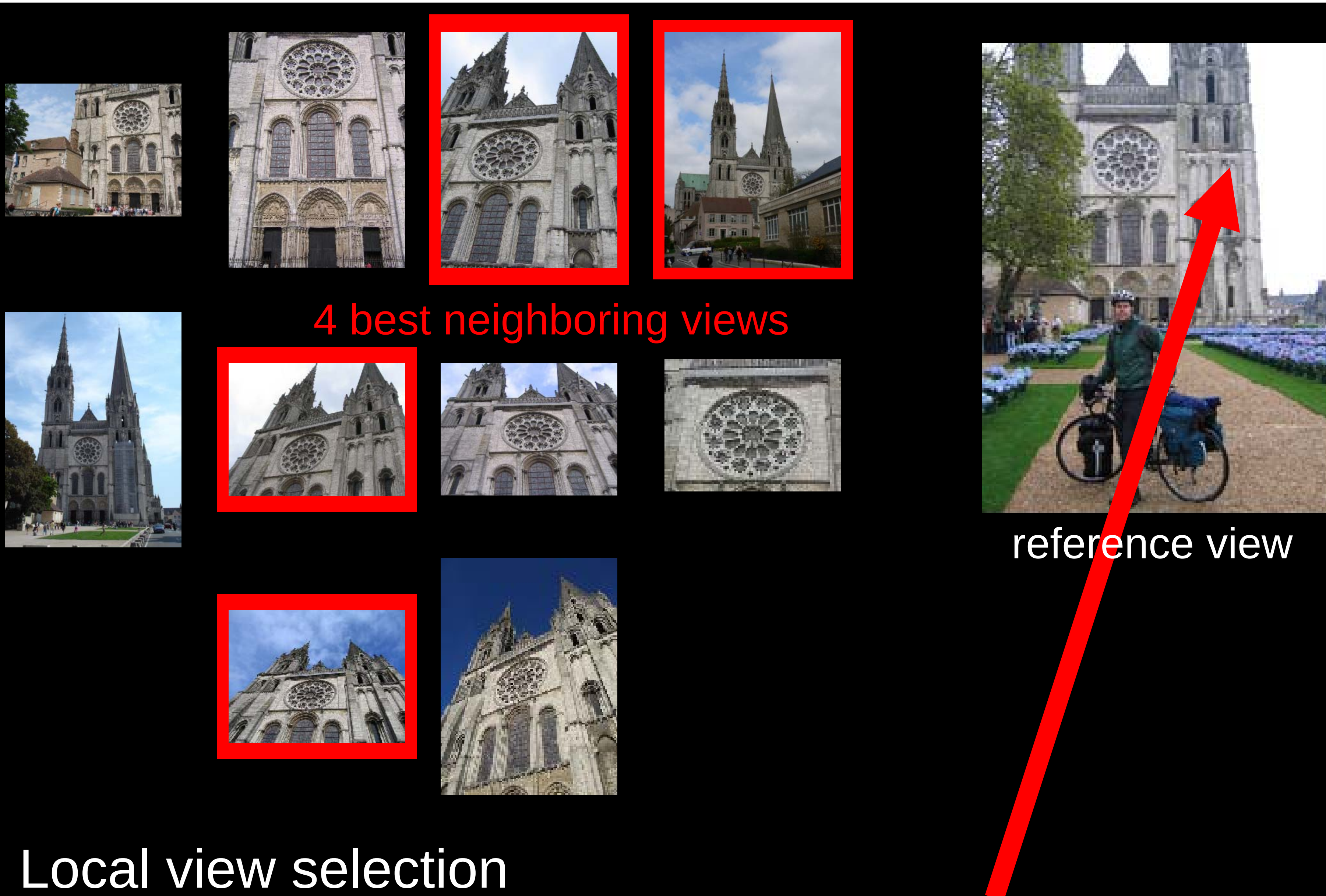
[Seitz]



Local view selection

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

[Seitz]



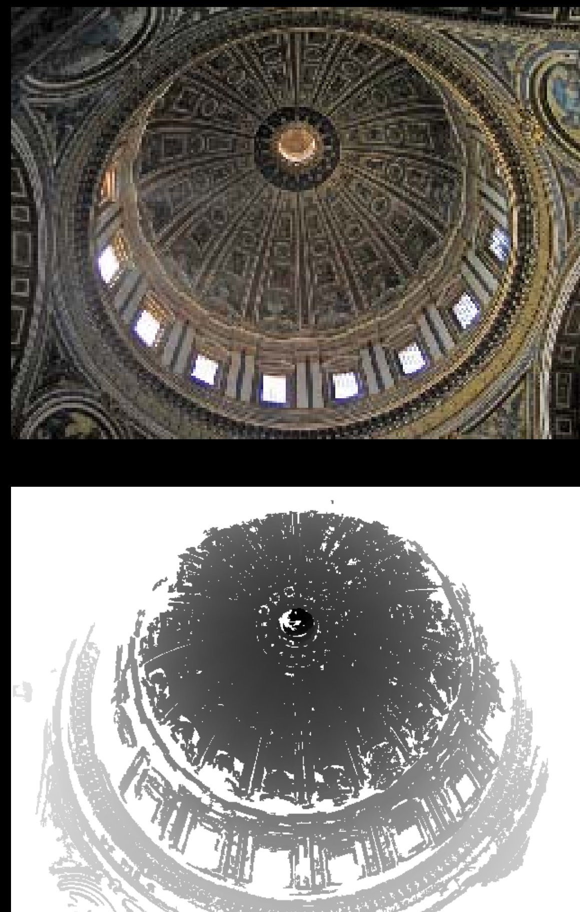
Local view selection

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

[Seitz]

Results

Mt. Rushmore
160 images
60 photographers



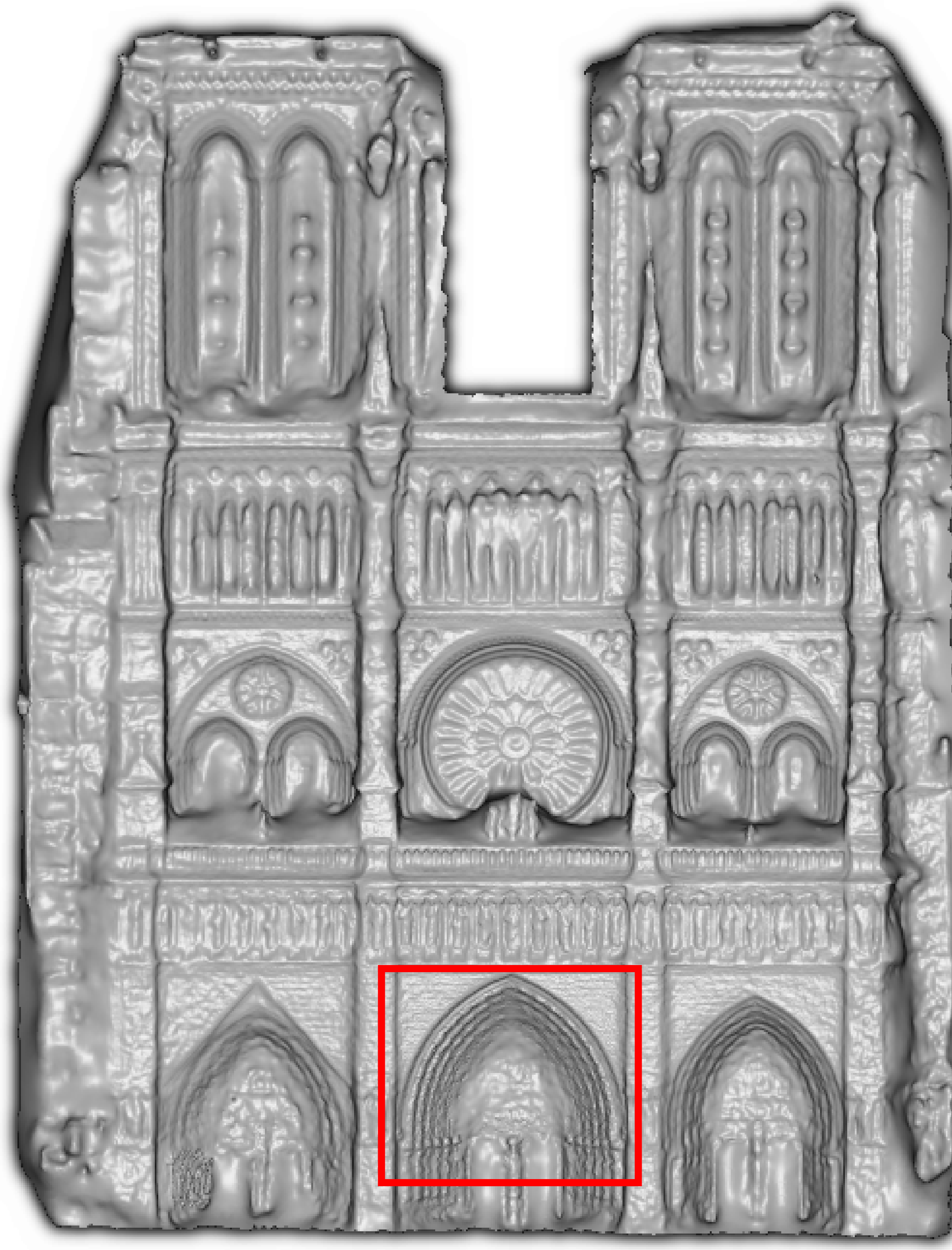
St. Peter
151 images
50 photographers

Trevi Fountain
106 images
51 photographers

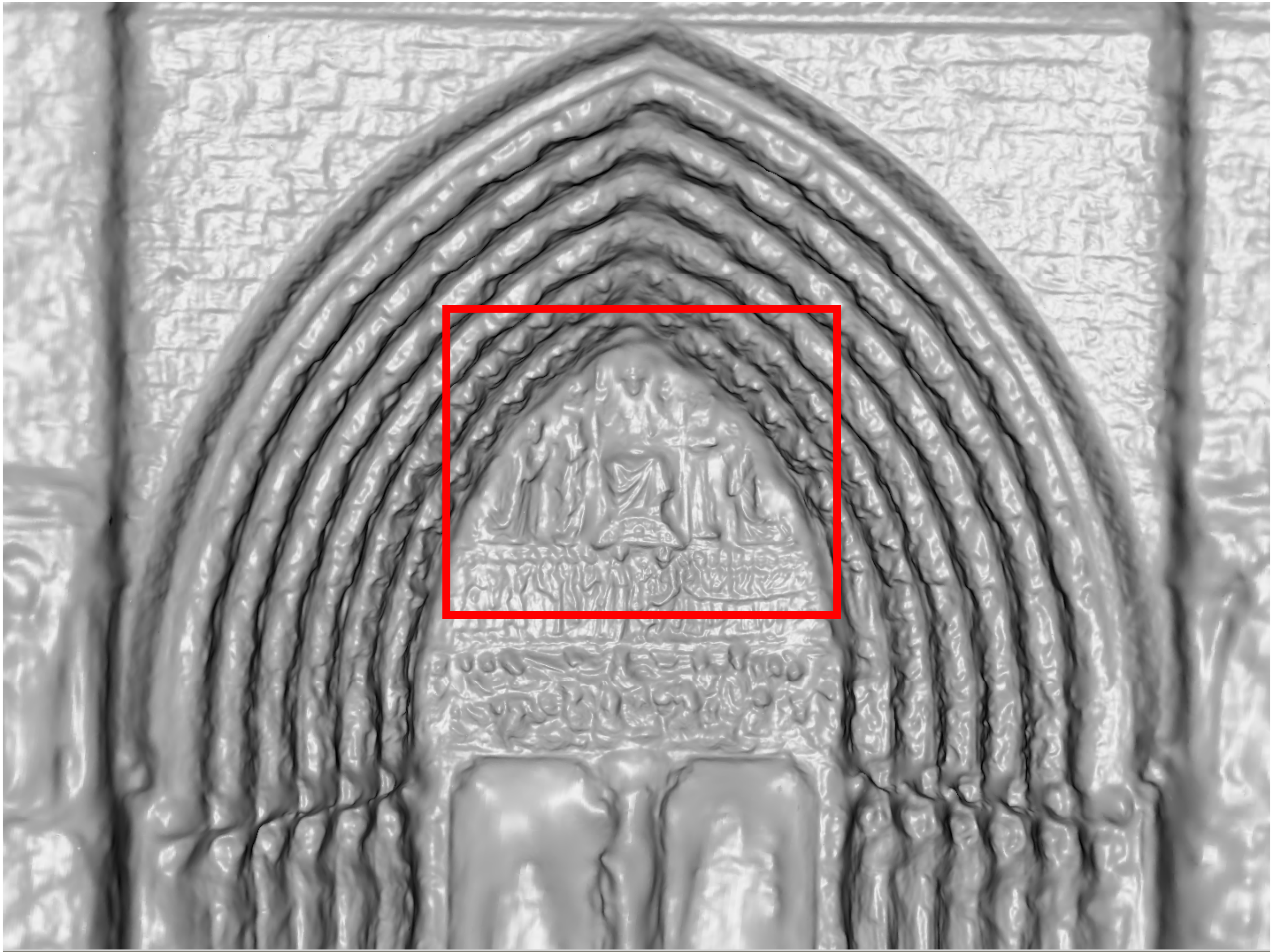
[Seitz]

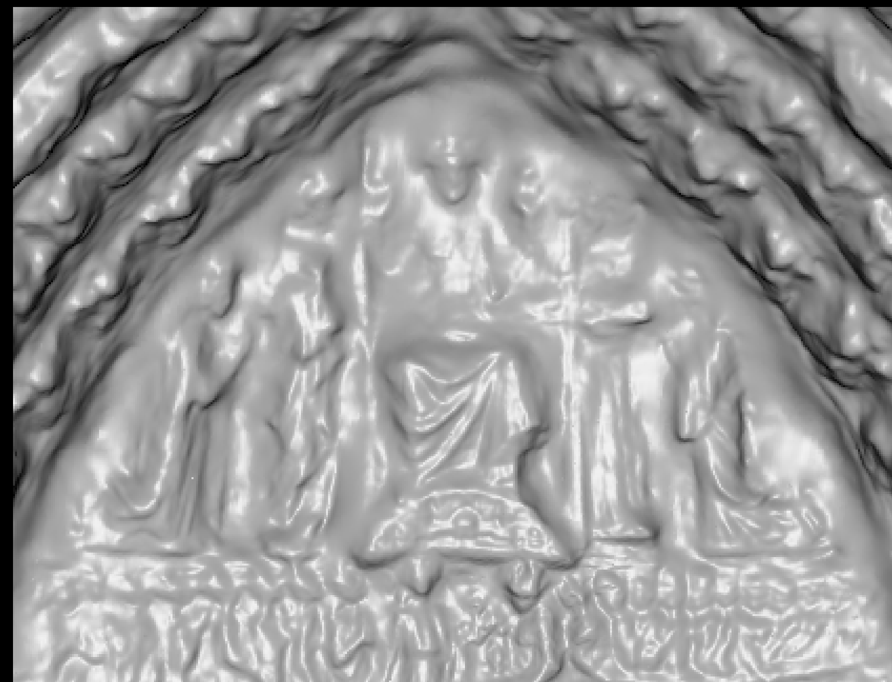
Notre Dame de Paris

653 images
313 photographers



[Seitz]



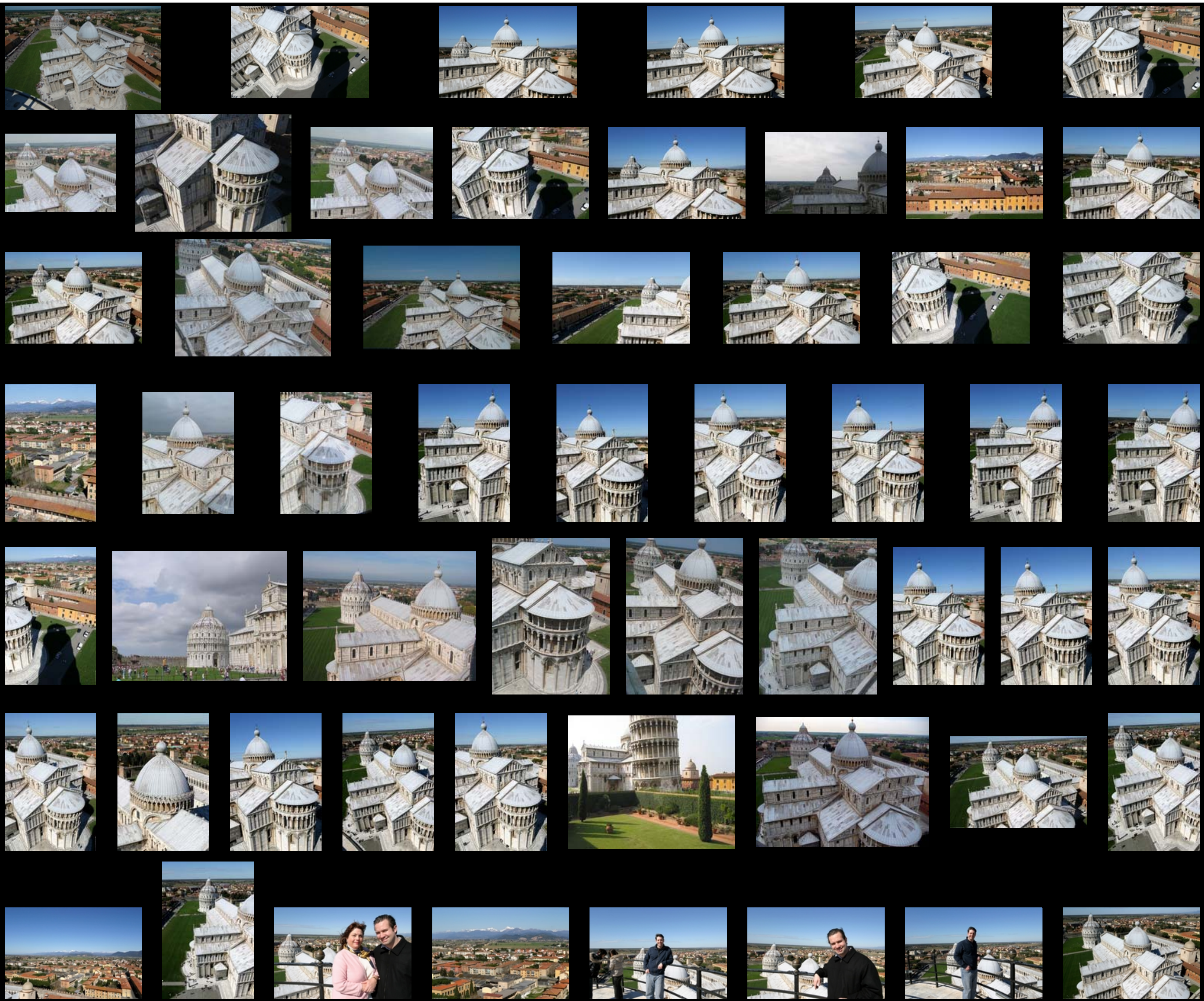


[Seitz]

129 *Flickr* images taken by 98 photographers [Seitz]

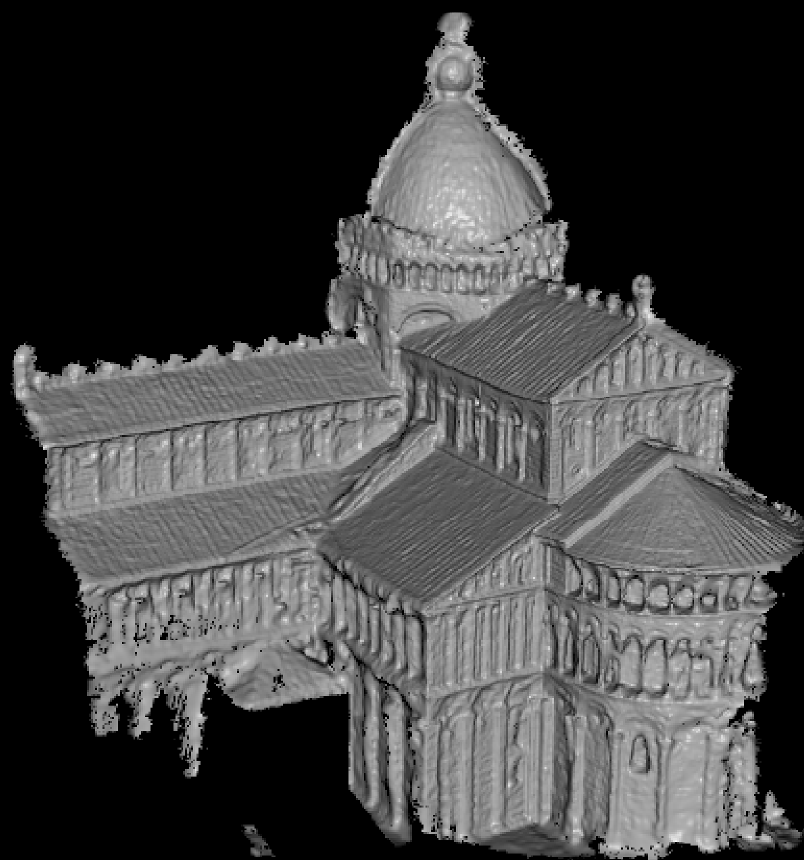
merged model of Venus de Milo

[Seitz]



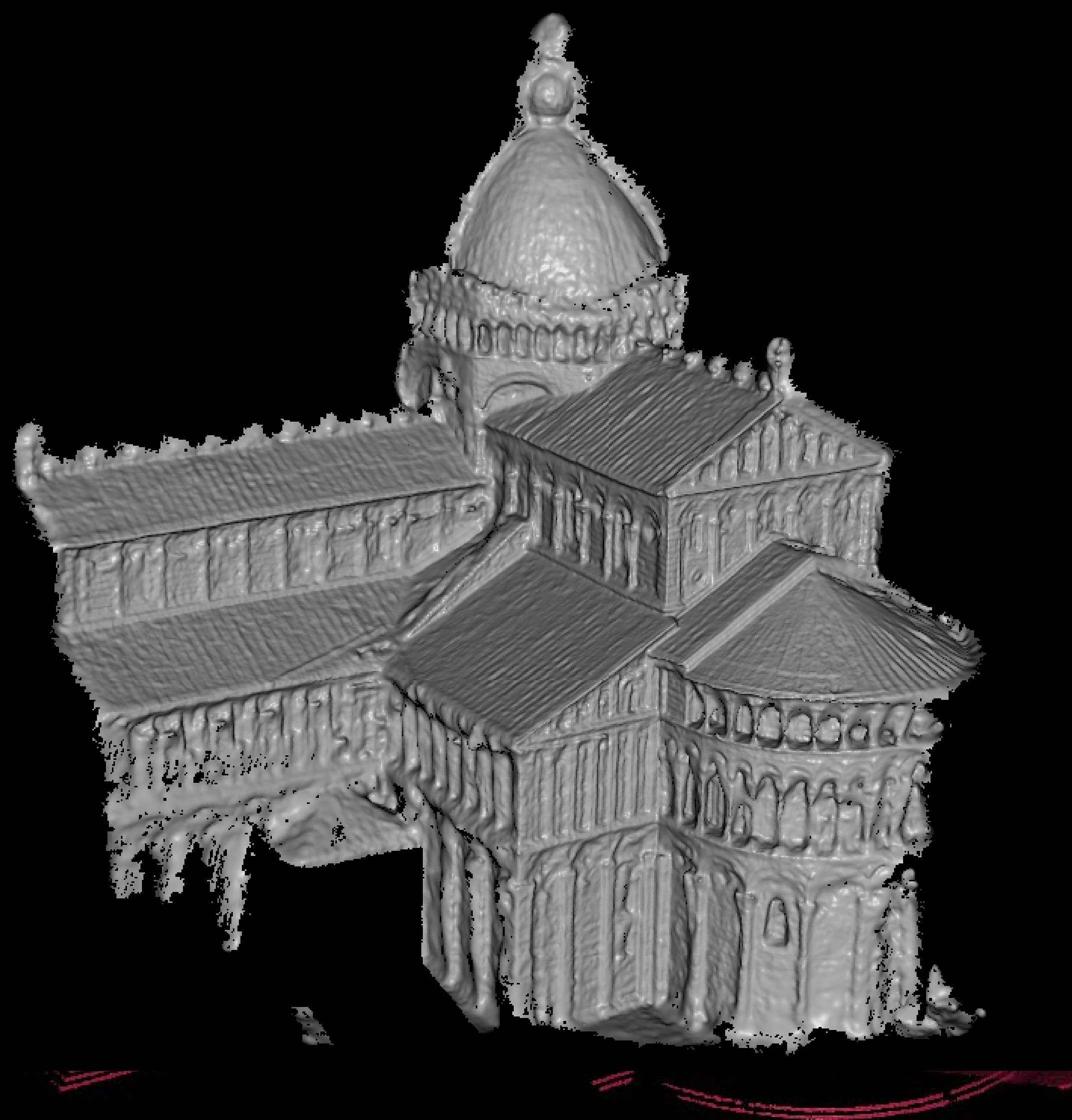
56 *Flickr* images taken by 8 photographers

[Seitz]



merged model of Pisa Cathedral

[Seitz]



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

Problem: *visibility*

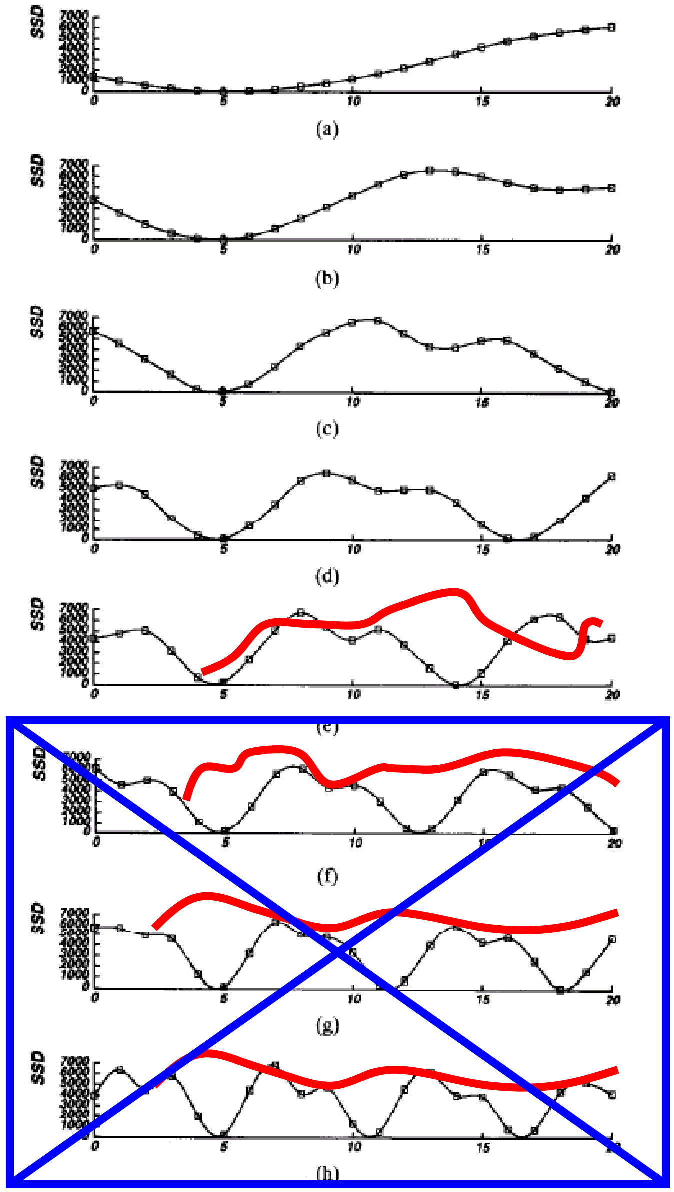


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

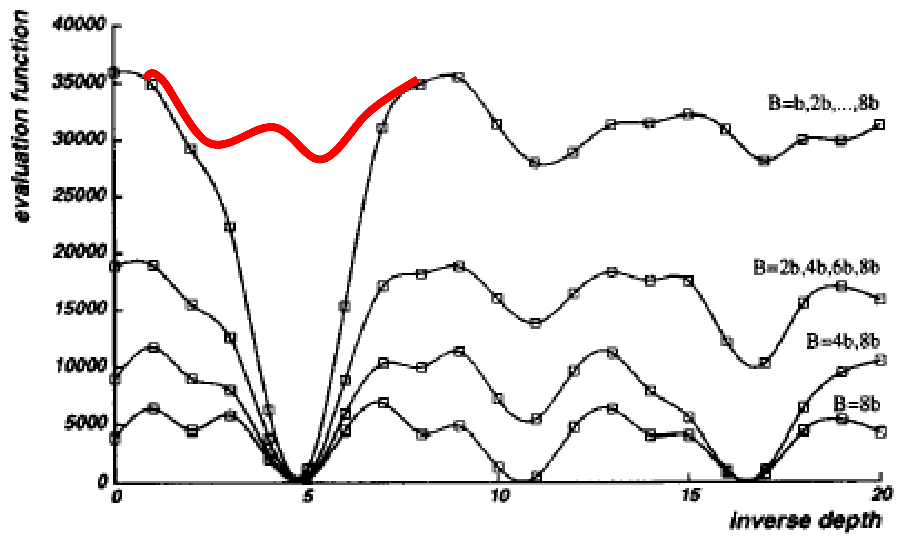


Fig. 7. Combining multiple baseline stereo pairs.

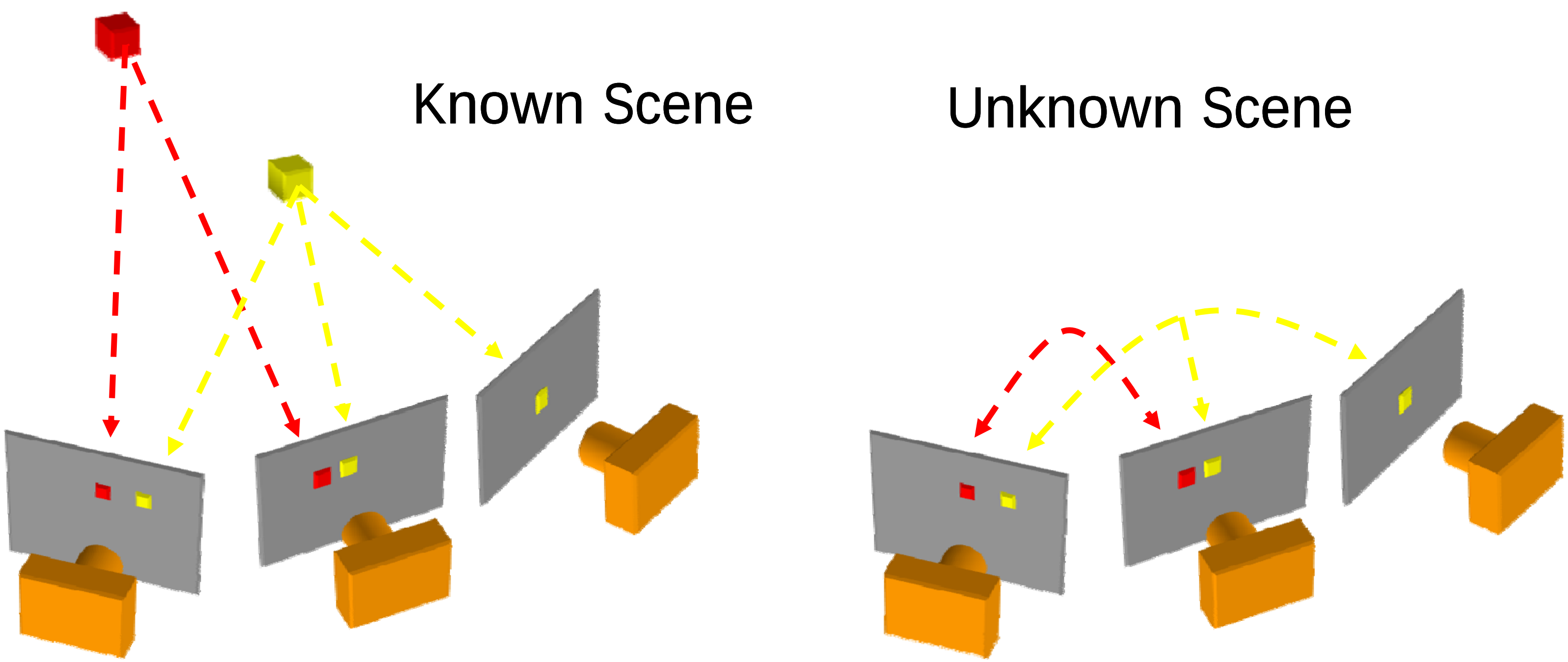
Some Solutions

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

[Seitz]

The visibility problem

Which points are visible in which images?

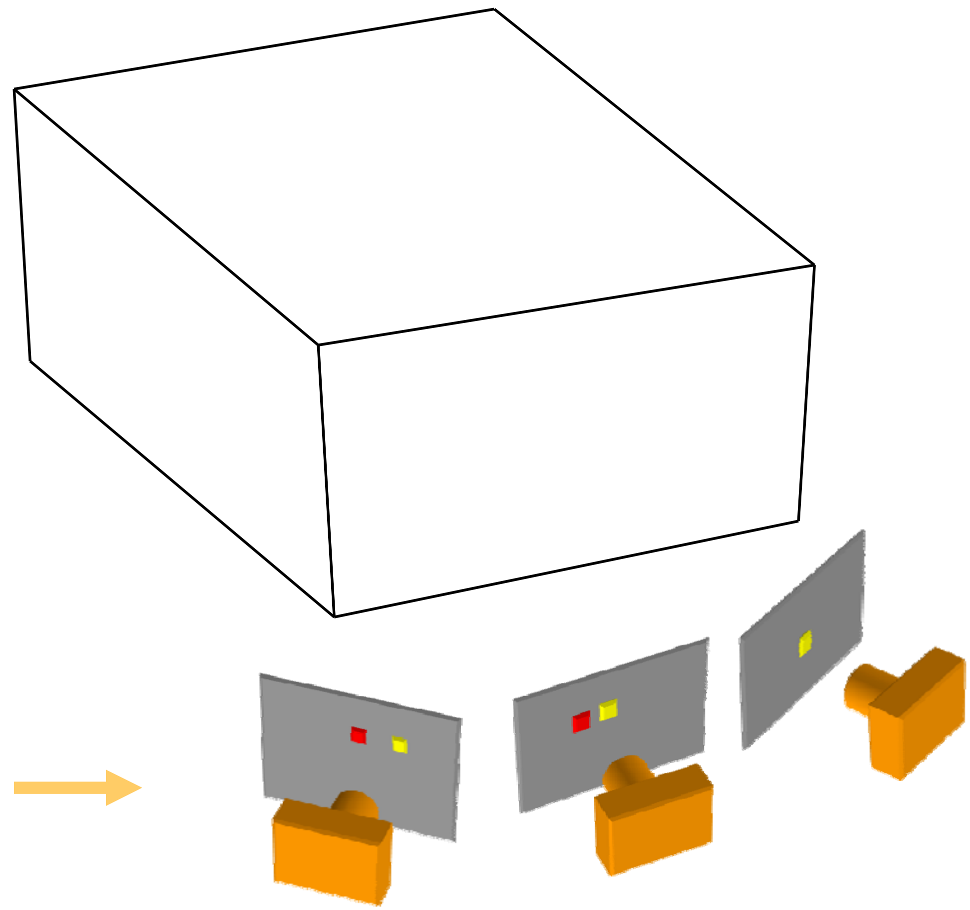


Forward Visibility *Inverse Visibility*

[Seitz]

Volumetric stereo

Scene Volume
V



Input Images
(Calibrated)

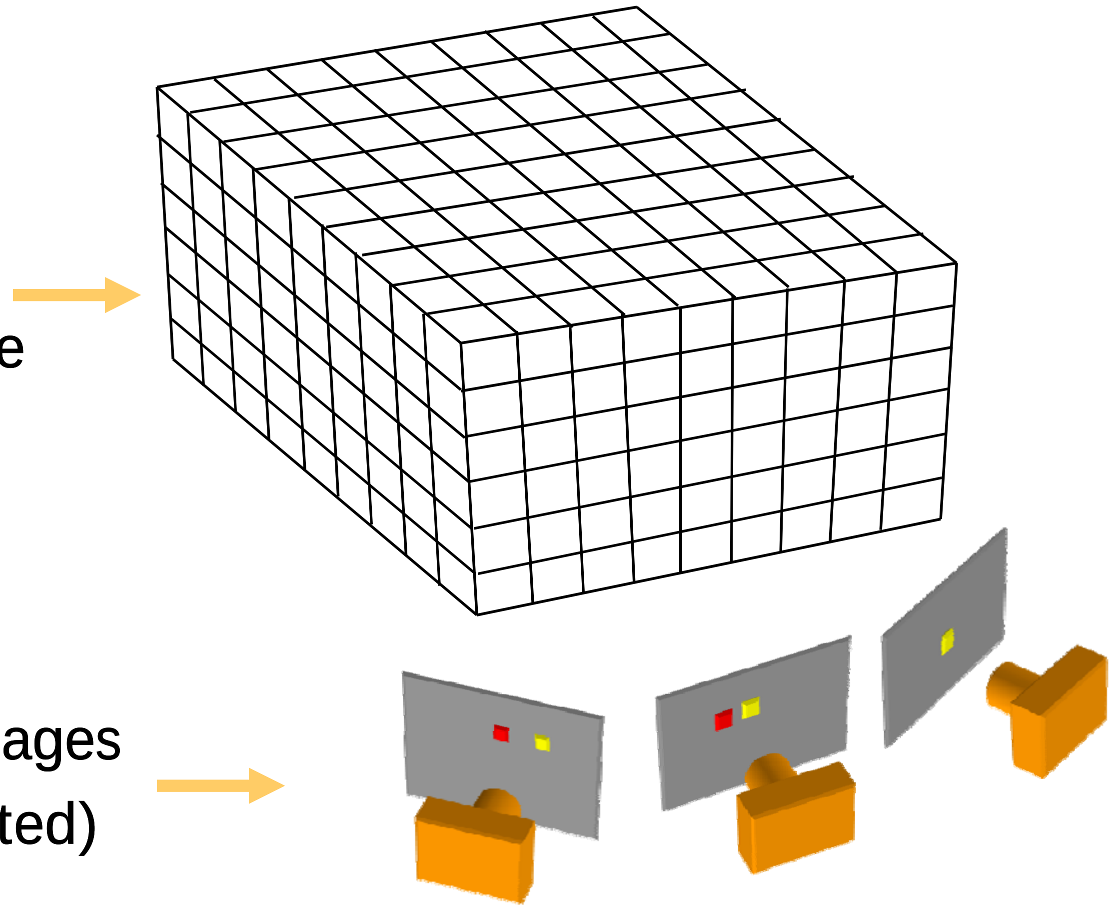
Goal: **Determine occupancy, “color” of points in V**

[Seitz]

Discrete formulation: Voxel Coloring

Discretized Scene Volume

Input Images (Calibrated)



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

[Seitz]

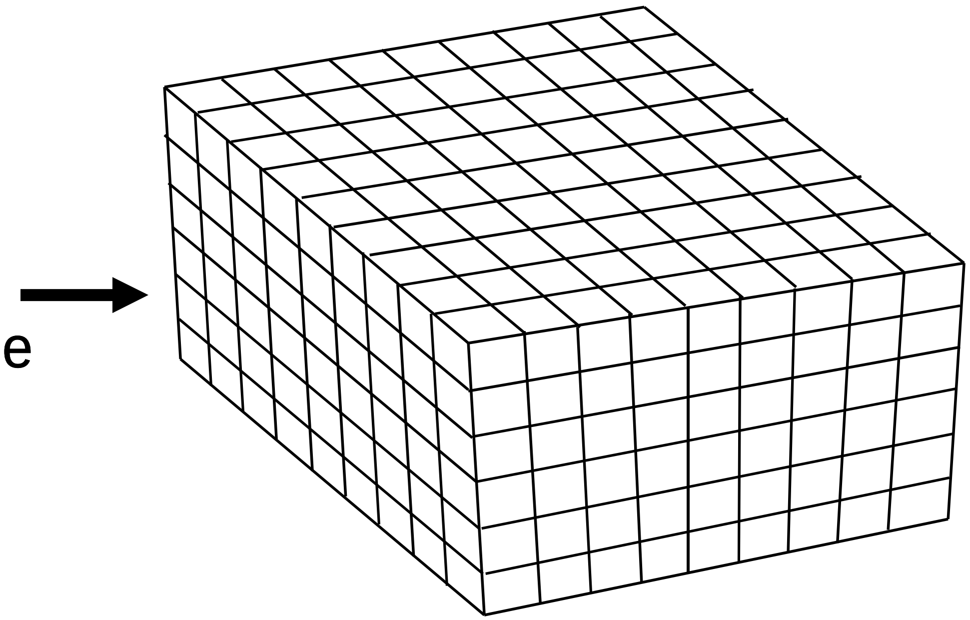
Complexity and computability

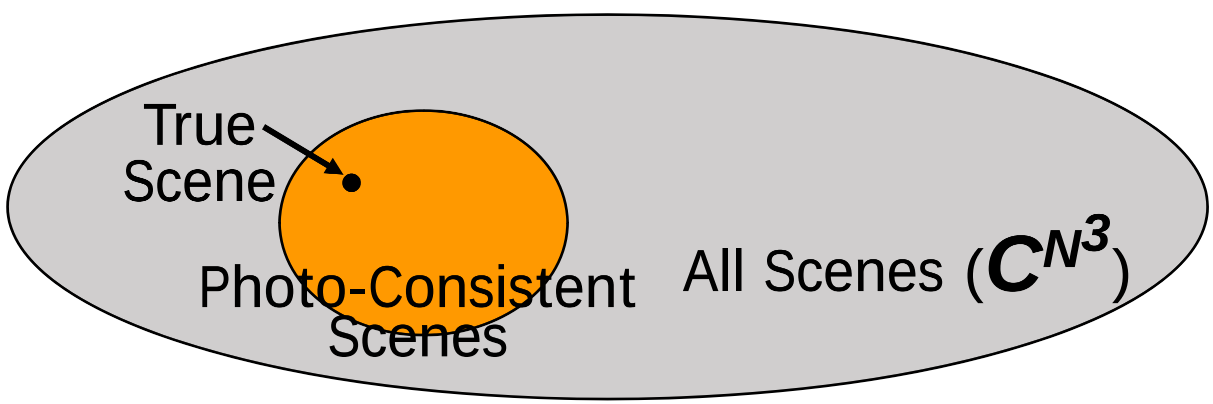
Discretized

Scene Volume

N^3 voxels

C colors





[Seitz]

Issues

Theoretical Questions

- Identify class of *all* photo-consistent scenes

Practical Questions

- How do we compute photo-consistent models?

[Seitz]

Voxel coloring solutions

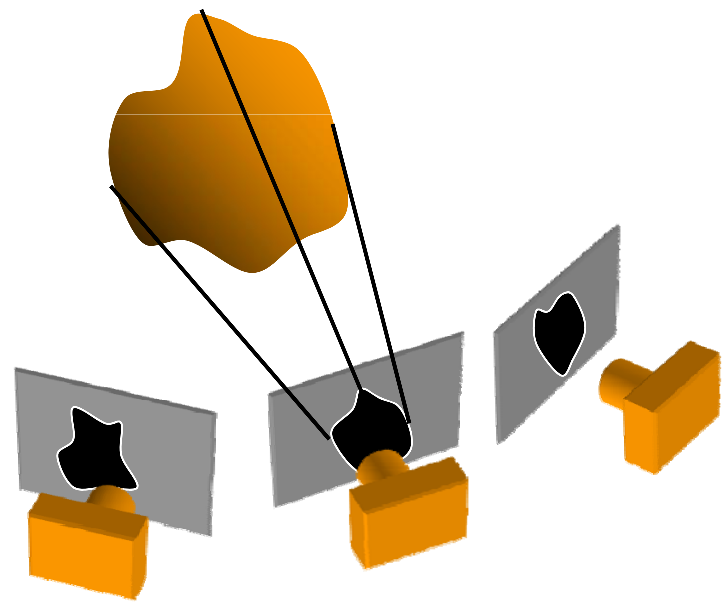
1. C=2 (shape from silhouettes)
   - Volume intersection [Baumgart 1974]
     > For more info: *Rapid octree construction from image sequences.* R. Szeliski, CVGIP: Image Understanding, 58(1):23-32, July 1993. (this paper is apparently not available online) or

     > W. Matusik, C. Buehler, R. Raskar, L. McMillan, and S. J. Gortler, *Image-Based Visual Hulls*, SIGGRAPH 2000 (pdf 1.6 MB)
2. C unconstrained, viewpoint constraints
   - Voxel coloring algorithm [Seitz & Dyer 97]
3. General Case
   - Space carving [Kutulakos & Seitz 98]

[Seitz]

Reconstruction from Silhouettes (C = 2)

Binary Images →

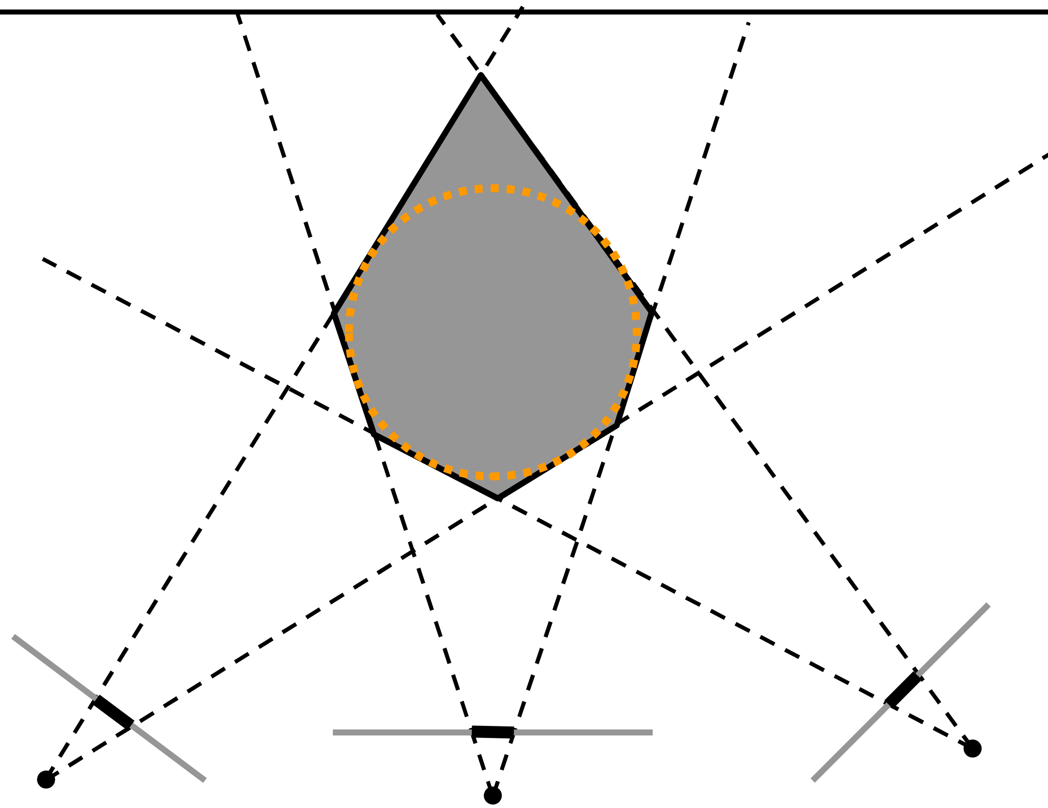


Approach:

- *Backproject* each silhouette
- Intersect backprojected volumes

[Seitz]

Volume intersection

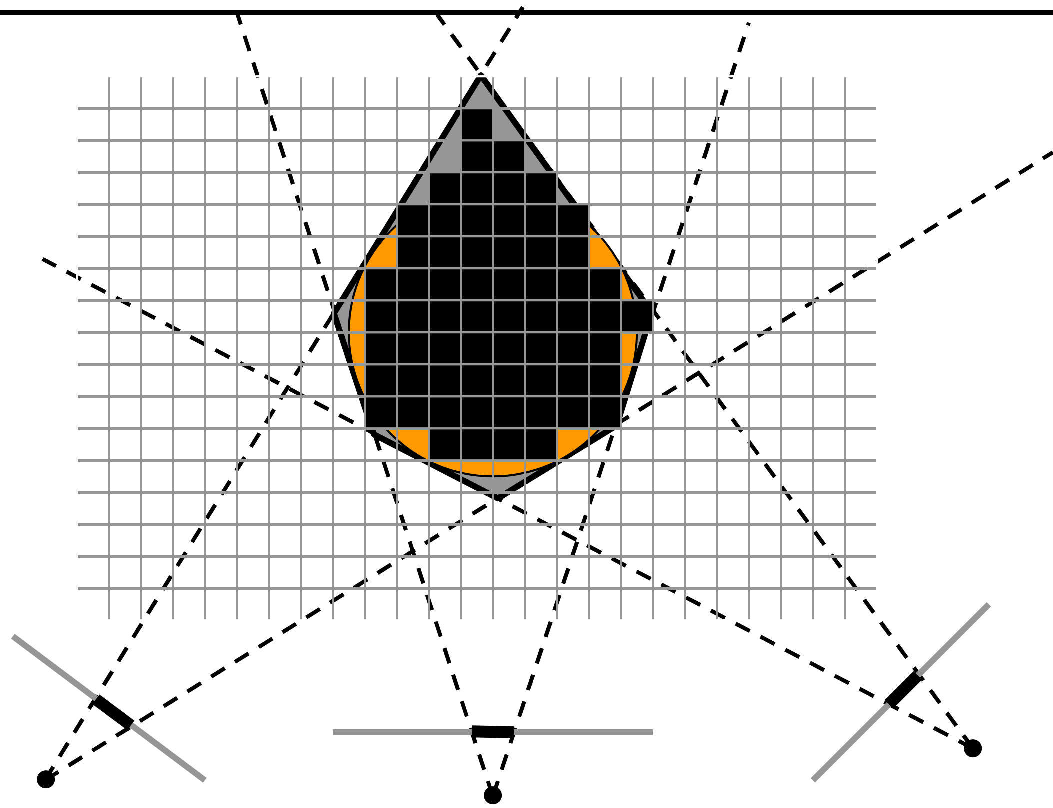


Reconstruction Contains the True Scene

- But is generally not the same
- In the limit (all views) get *visual hull*
  - > Complement of all lines that don't intersect S

[Seitz]

Voxel algorithm for volume intersection



Color voxel black if on silhouette in every image

- O(?), for M images, N^3 voxels O(MN^3)
- Don't have to search 2^{N^3} possible scenes!

[Seitz]

Properties of Volume Intersection

Pros

- Easy to implement, fast
- Accelerated via octrees [Szeliski 1993] or interval techniques [Matusik 2000]

Cons

- No concavities
- Reconstruction is not photo-consistent
- Requires identification of silhouettes

[Seitz]

Voxel Coloring Solutions

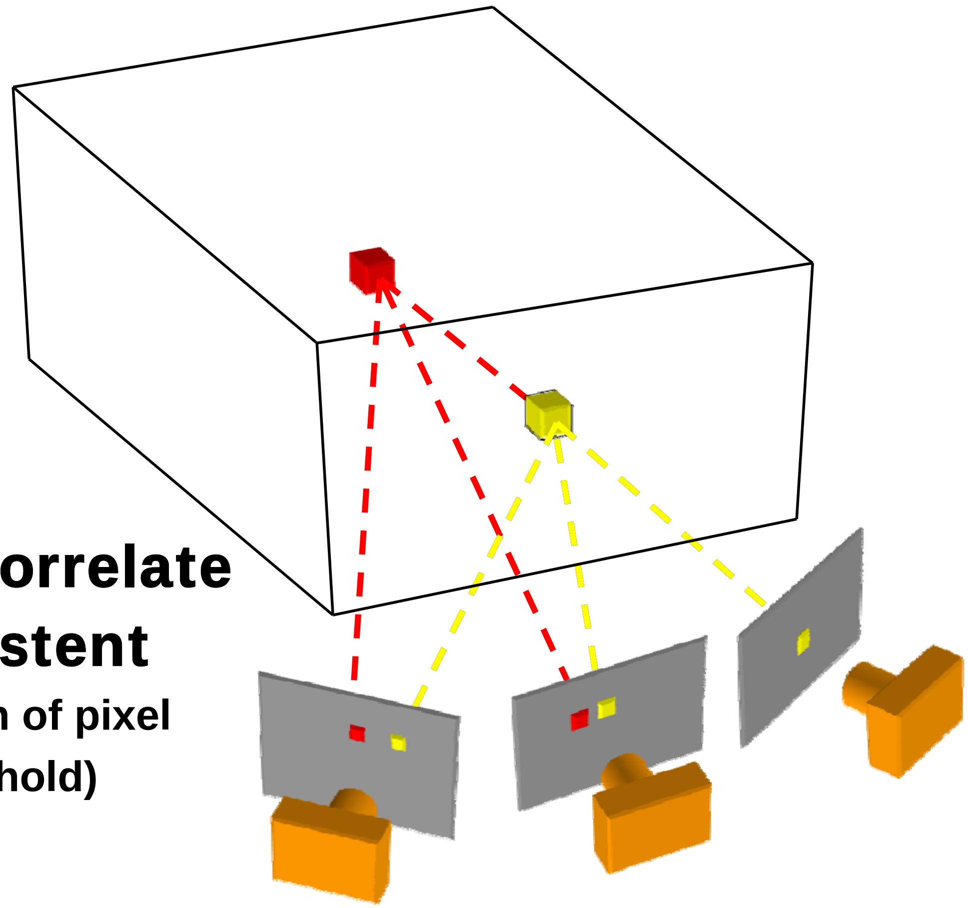
1. C=2 (silhouettes)
   - Volume intersection [Baumgart 1974]

2. C unconstrained, viewpoint constraints
   - Voxel coloring algorithm [Seitz & Dyer 97]
     - For more info: http://www.cs.washington.edu/homes/seitz/papers/ijcv99.pdf

3. General Case
   - Space carving [Kutulakos & Seitz 98]

[Seitz]

Voxel Coloring Approach

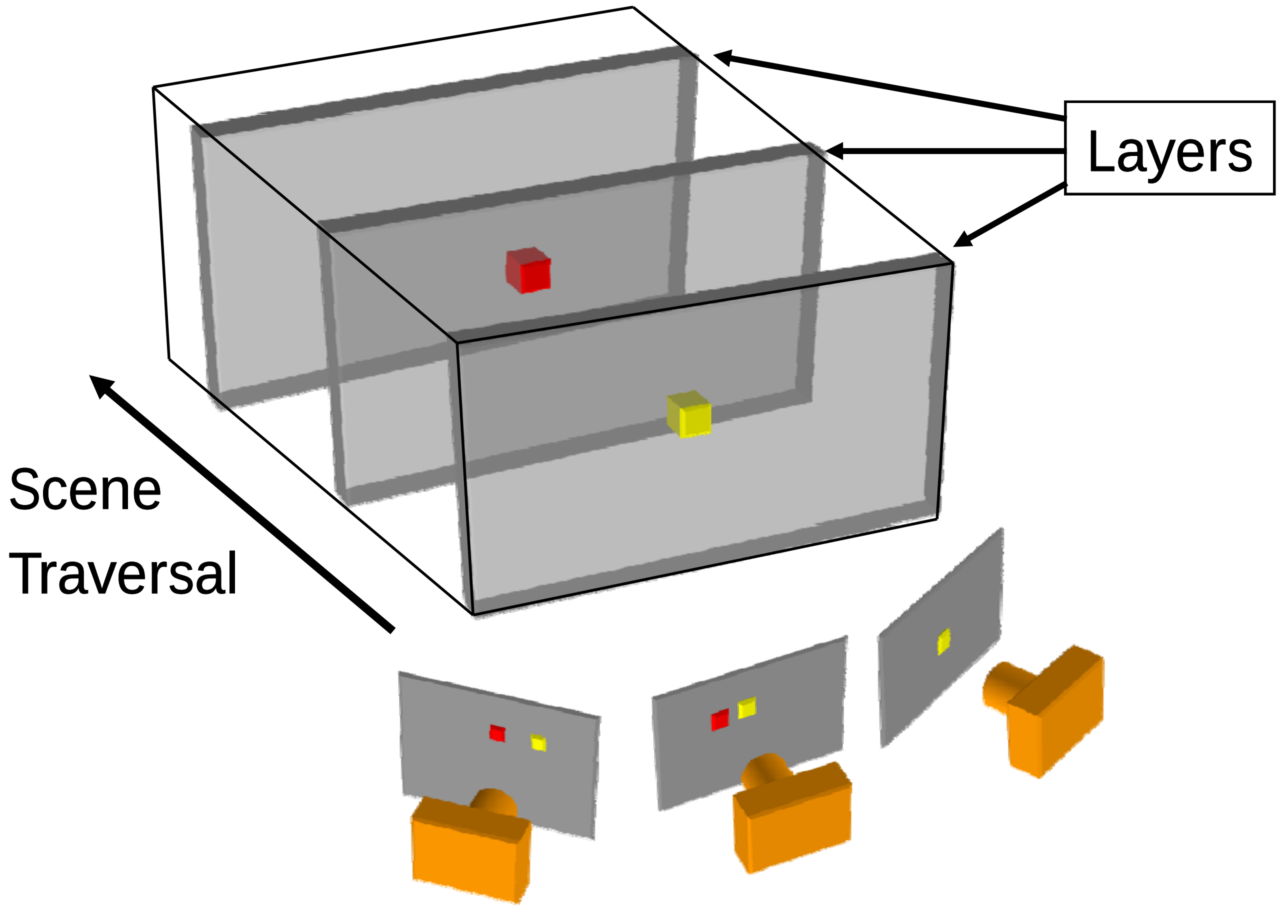


1. **Choose voxel**
2. **Project and correlate**
3. **Color if consistent**
   (standard deviation of pixel colors below threshold)

Visibility Problem: in which images is each voxel visible?

[Seitz]

Depth Ordering: visit occluders first!

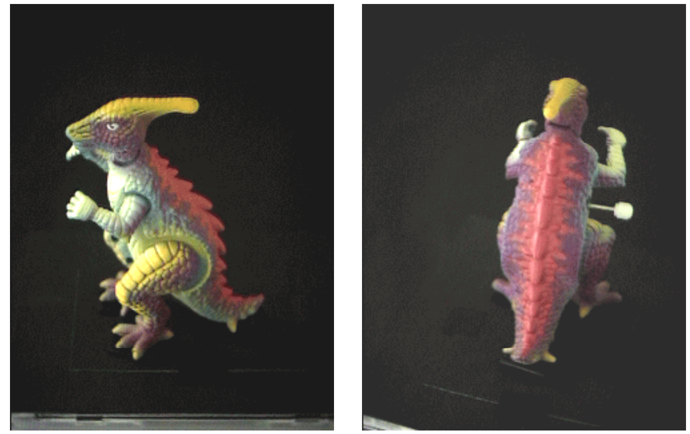


Condition: **depth order is the *same for all input views***

[Seitz]

Calibrated Image Acquisition

Calibrated Turntable



Selected Dinosaur Images

Selected Flower Images

[Seitz]

Voxel Coloring Results (Video)

Dinosaur Reconstruction
72 K voxels colored
7.6 M voxels tested
7 min. to compute
on a 250MHz SGI

Flower Reconstruction
70 K voxels colored
7.6 M voxels tested
7 min. to compute
on a 250MHz SGI

[Seitz]

Improvements

Unconstrained camera viewpoints

- Space carving [Kutulakos & Seitz 98]

Evolving a surface

- Level sets [Faugeras & Keriven 98]
- More recent work by Pons et al.

Global optimization

- Graph cut approaches
  - > [Kolmogoriv & Zabih, ECCV 2002]
  - > [Vogiatzis et al., PAMI 2007]

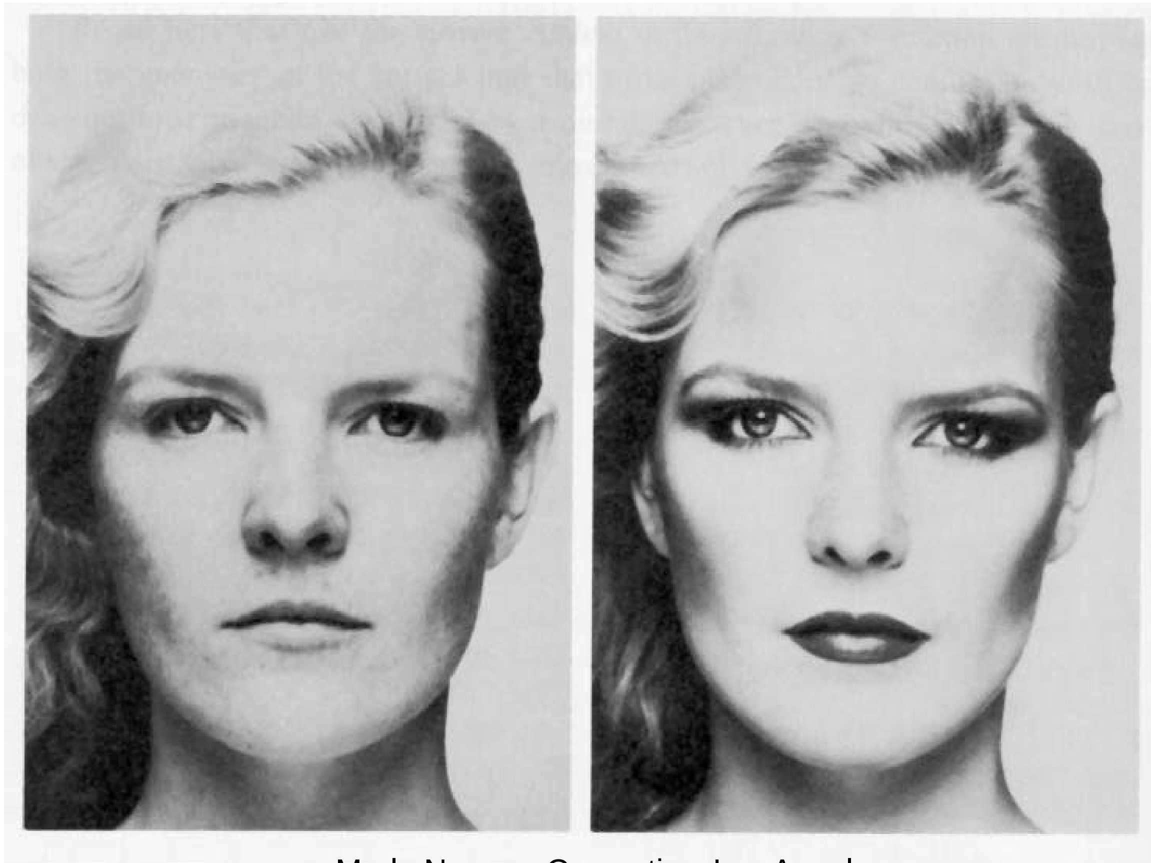
Modeling shiny (and other reflective) surfaces

- e.g., Zickler et al., *Helmholtz Stereopsis*

See today's reading for an overview of the state of the art

[Seitz]

Photometric Stereo



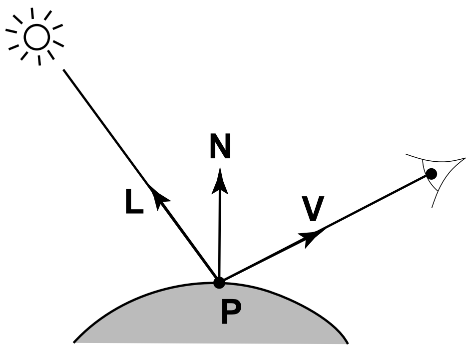
Merle Norman Cosmetics, Los Angeles

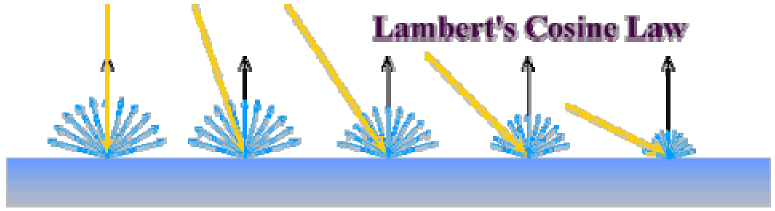
Readings

- R. Woodham, *Photometric Method for Determining Surface Orientation from Multiple Images*. Optical Engineering 19(1)139-144 (1980). (PDF)

[Seitz]

Diffuse reflection





$$R_e = k_d \mathbf{N} \cdot \mathbf{L} R_i$$

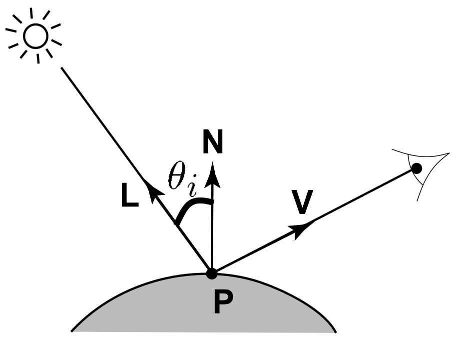
image intensity of **P** → $$I = k_d \mathbf{N} \cdot \mathbf{L}$$

Simplifying assumptions

- $I = R_e$: camera response function f is the identity function:
  - can always achieve this in practice by solving for f and applying f^{-1} to each pixel in the image
- $R_i = 1$: light source intensity is 1
  - can achieve this by dividing each pixel in the image by R_i

[Seitz]

Shape from shading



Suppose $k_d = 1$

$$
\begin{aligned}
I &= k_d \mathbf{N} \cdot \mathbf{L} \\
&= \mathbf{N} \cdot \mathbf{L} \\
&= cos\ \theta_i
\end{aligned}
$$

You can directly measure angle between normal and light source

- Not quite enough information to compute surface shape
- But can be if you add some additional info, for example
  - assume a few of the normals are known (e.g., along silhouette)
  - constraints on neighboring normals—“integrability”
  - smoothness
- Hard to get it to work well in practice
  - plus, how many real objects have constant albedo?

[Seitz]

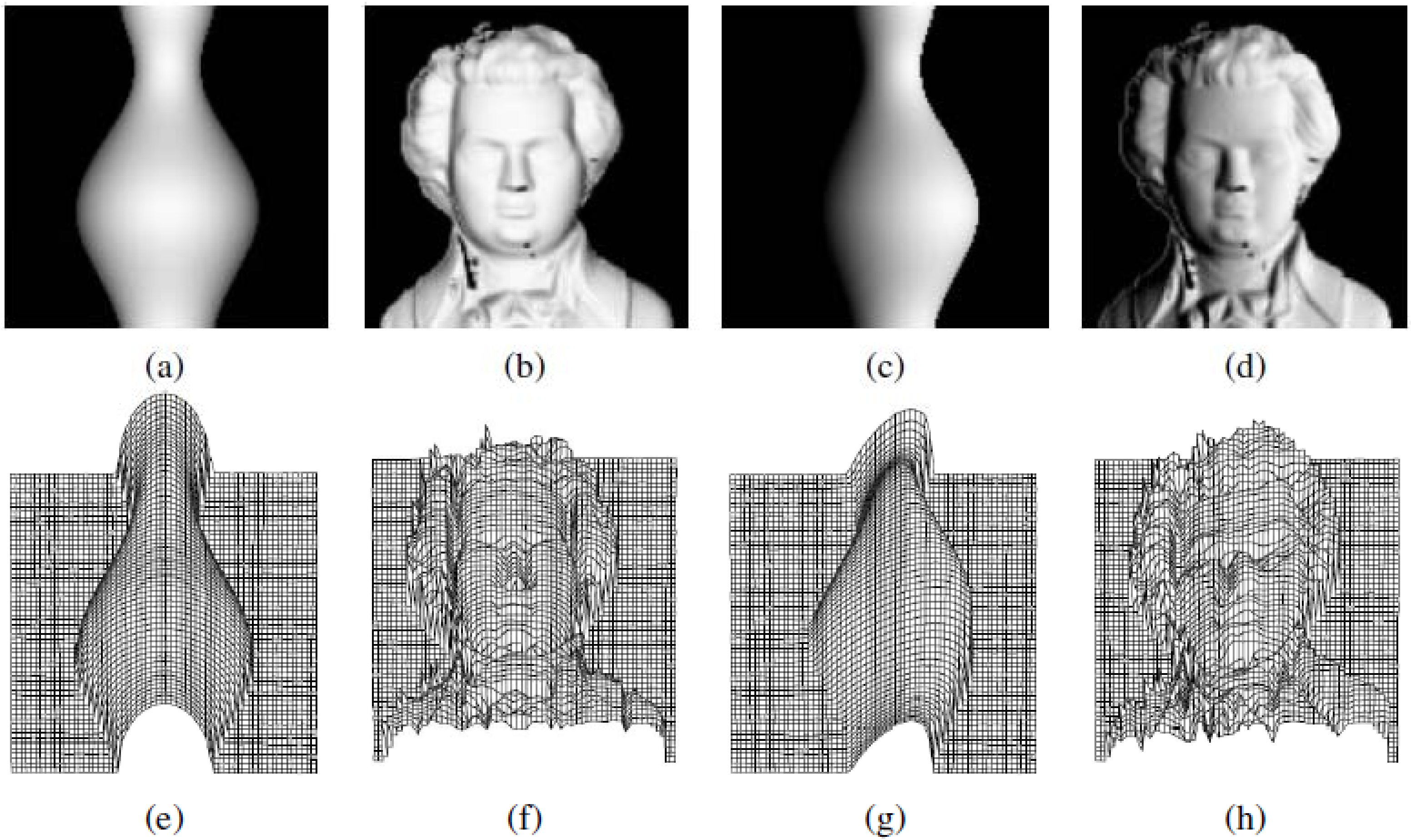


Figure 12.2: *Synthetic shape from shading example (Zhang* et al. *1999a): (a–d) shaded images, with light from in front* $(0, 0, 1)$ *and from the front right* $(1, 0, 1)$*; (e–f) corresponding shape from shading reconstructions using the technique of Tsai and Shah (1994).*

[Seitz]

Shape from Shading

Most shape from shading algorithms assume that the surface under consideration is of a uniform albedo and reflectance, and that the light source directions are either known or can be calibrated by the use of a reference object. Under the assumptions of distant light sources and observer, the variation in intensity (*irradiance equation*) become purely a function of the local surface orientation,

$$I(x, y) = R(p(x, y), q(x, y)), \tag{12.1}$$

where $(p, q) = (z_x, z_y)$ are the depth map derivatives and $R(p, q)$ is called the *reflectance map*. For example, a diffuse (Lambertian) surface has a reflectance map that is the (non-negative) dot product (2.87) between the surface normal $\hat{n} = (p, q, 1)/\sqrt{1 + p^2 + q^2}$ and the light source direction $v = (v_x, v_y, v_z)$,

$$R(p, q) = \max\left(0, \rho\frac{pv_x + qv_y + v_z}{\sqrt{1 + p^2 + q^2}}\right), \tag{12.2}$$

where ρ is the surface reflectance factor (albedo).

[Seitz]

Shape from Shading

In principle, (12.1–12.2) can be used to estimate (p, q) using non-linear least squares or some other method. Unfortunately, unless additional constraints are imposed, there are more unknowns per pixel (p, q) than there are measurements (I). One commonly used constraint is the smoothness constraint,

$$\mathcal{E}_s = \int p_x^2 + p_y^2 + q_x^2 + q_y^2 \, dx \, dy = \int \|\nabla p\|^2 + \|\nabla q\|^2 \, dx \, dy, \tag{12.3}$$

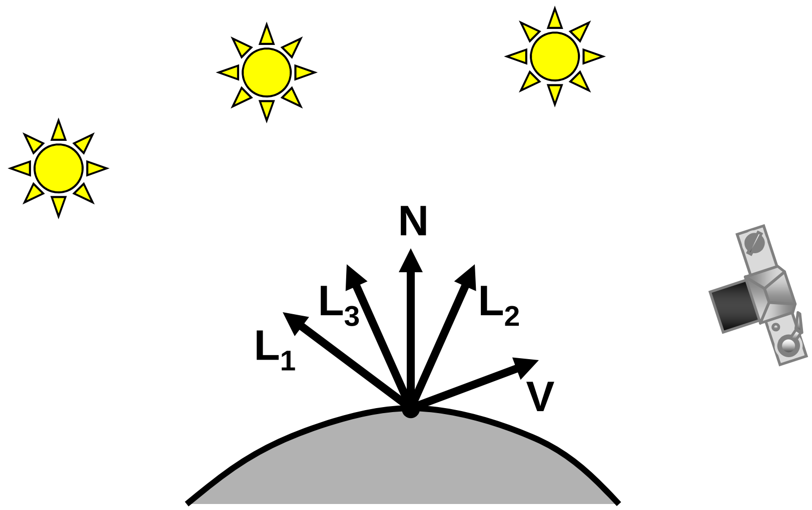
which we already saw in §3.6.1 (3.93). The other is the *integrability constraint*,

$$\mathcal{E}_i = \int (p_y - q_x)^2 \, dx \, dy, \tag{12.4}$$

which arises naturally, since for a valid depth map $z(x, y)$ with $(p, q) = (z_x, z_y)$, we have $p_y = z_{xy} = z_{yx} = q_x$.

[Seitz]

Photometric stereo



$$I_1 = k_d \mathbf{N} \cdot \mathbf{L_1}$$
$$I_2 = k_d \mathbf{N} \cdot \mathbf{L_2}$$
$$I_3 = k_d \mathbf{N} \cdot \mathbf{L_3}$$

Can write this as a matrix equation:

$$\left[\begin{array}{ccc} I_1 & I_2 & I_3 \end{array}\right] = k_d \mathbf{N}^T \left[\begin{array}{ccc} \mathbf{L_1} & \mathbf{L_2} & \mathbf{L_3} \end{array}\right]$$

[Seitz]

Solving the equations

$$\underbrace{\begin{bmatrix} I_1 & I_2 & I_3 \end{bmatrix}}_{\substack{\mathbf{I} \\ 1 \times 3}} = \underbrace{k_d \mathbf{N}^T}_{\substack{\mathbf{G} \\ 1 \times 3}} \underbrace{\begin{bmatrix} \mathbf{L_1} & \mathbf{L_2} & \mathbf{L_3} \end{bmatrix}}_{\substack{\mathcal{L} \\ 3 \times 3}}$$

$$\mathbf{G} = \mathbf{I}\mathbf{L}^{-1}$$

$$k_d = \|\mathbf{G}\|$$

$$\mathbf{N} = \frac{1}{k_d}\mathbf{G}$$

[Seitz]

More than three lights

Get better results by using more lights

$$\begin{bmatrix} I_1 & \dots & I_n \end{bmatrix} = k_d \mathbf{N}^T \begin{bmatrix} \mathbf{L_1} & \dots & \mathbf{L_n} \end{bmatrix}$$

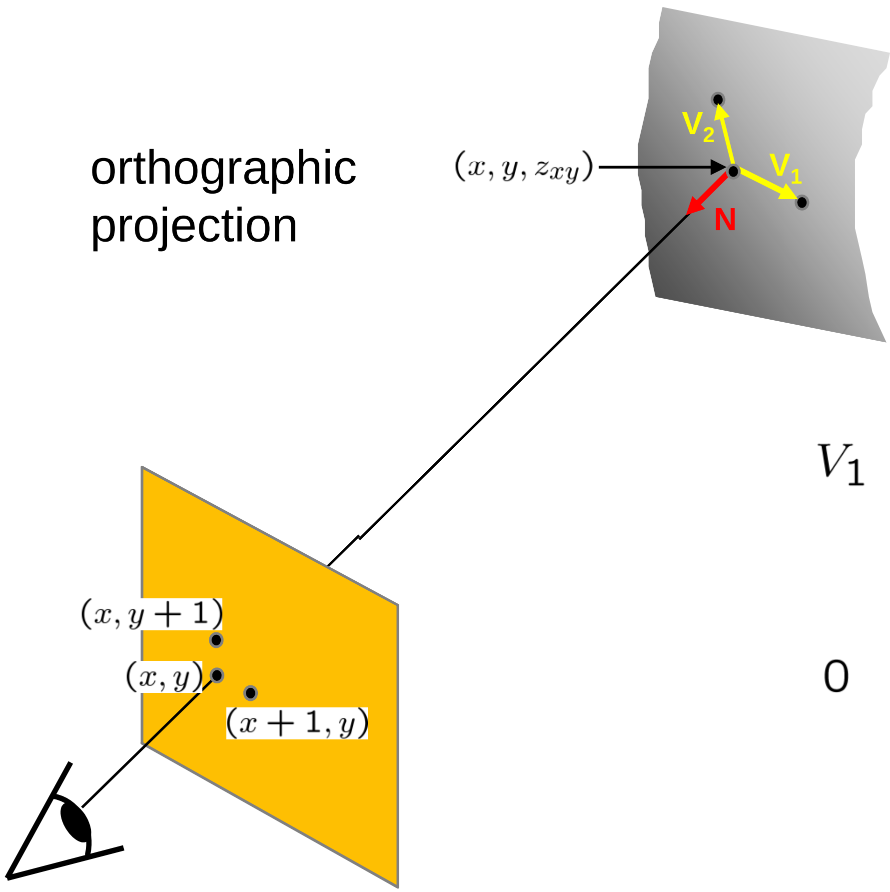
Least squares solution:

$$\begin{aligned} \mathbf{I} &= \mathbf{GL} \\ \mathbf{IL^T} &= \mathbf{GLL^T} \\ \mathbf{G} &= (\mathbf{IL^T})(\mathbf{LL^T})^{-1} \end{aligned}$$

Solve for N, k_d as before

[Seitz]

Depth from normals



$$
\begin{aligned}
V_1 &= (x+1, y, z_{x+1,y}) - (x, y, z_{xy}) \\
&= (1, 0, z_{x+1,y} - z_{xy})
\end{aligned}
$$

$$
\begin{aligned}
0 &= N \cdot V_1 \\
&= (n_x, n_y, n_z) \cdot (1, 0, z_{x+1,y} - z_{xy}) \\
&= n_x + n_z(z_{x+1,y} - z_{xy})
\end{aligned}
$$

Get a similar equation for $\mathbf{V_2}$

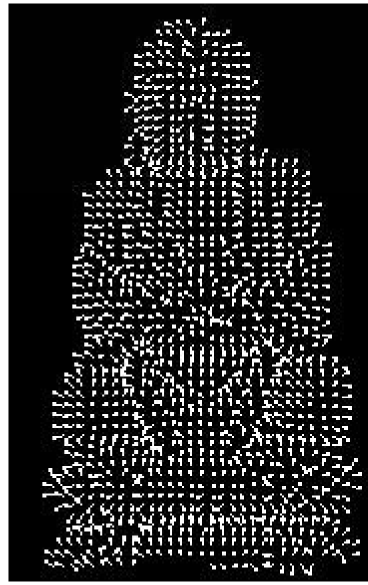
- Each normal gives us two linear constraints on z
- compute z values by solving a matrix equation

[Seitz]

Results…

Input
(1 of 12)

Normals



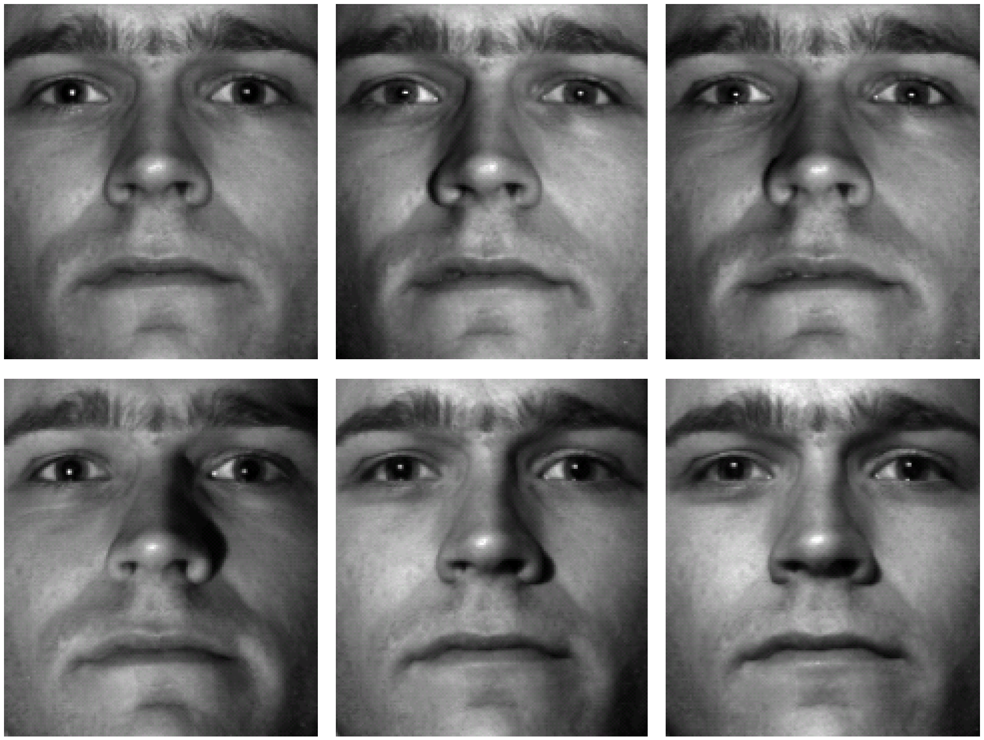
Normals

Shaded
rendering

Textured
rendering

[Seitz]

Results…



from Athos Georghiades
http://cvc.yale.edu/people/Athos.html

[Seitz]

Limitations

Big problems

- doesn't work for shiny things, semi-translucent things
- shadows, inter-reflections

Smaller problems

- camera and lights have to be distant
- calibration requirements
  - measure light source directions, intensities
  - camera response function

Newer work addresses some of these issues

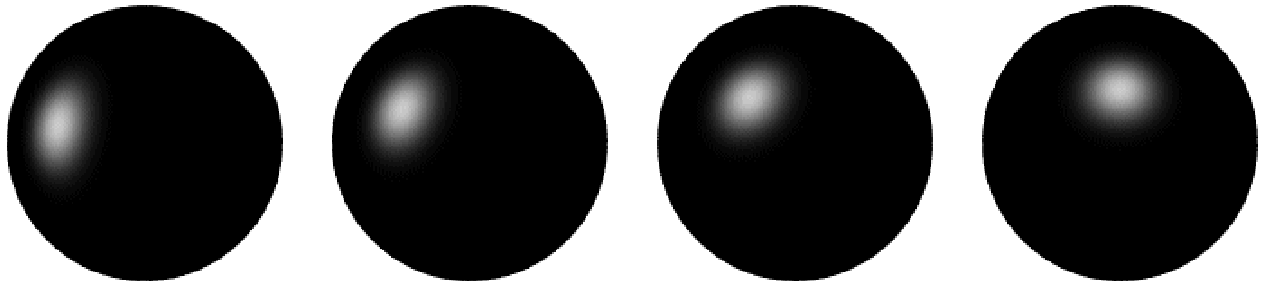
Some pointers for further reading:

- Zickler, Belhumeur, and Kriegman, *"Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction."* IJCV, Vol. 49 No. 2/3, pp 215-227.
- Hertzmann & Seitz, "*Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs*." IEEE Trans. PAMI 2005

[Seitz]

Computing light source directions

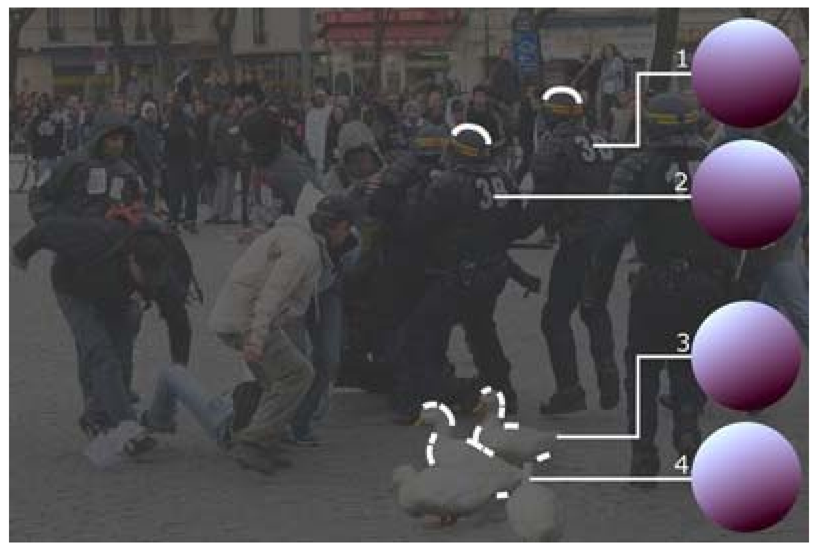
Trick: place a chrome sphere in the scene



- the location of the highlight tells you where the light source is

[Seitz]

An Aside: Digital Forensics (H. Farid)



[H. Farid, Dartmouth]

This photograph of the host and judges for the popular television show American Idol was scheduled for publication when it caught the attention of a photo-editor who was concerned that the image was doctored.

We found traces of tampering when we examined each eye. First, the inconsistencies in the shape of the specular highlight on the eyes suggest that each person was originally photographed under different lighting conditions. Second, from the location of these highlights, we can mathematically determine the 3-D location of the lights under which each person was originally photographed. We have determined that the positions of the lights in this image are inconsistent and conclude that this image is a composite of at least three images.

[H. Farid, Dartmouth]