CS6670: Computer Vision

Noah Snavely

Lecture 23: Structure from motion and multi-view stereo

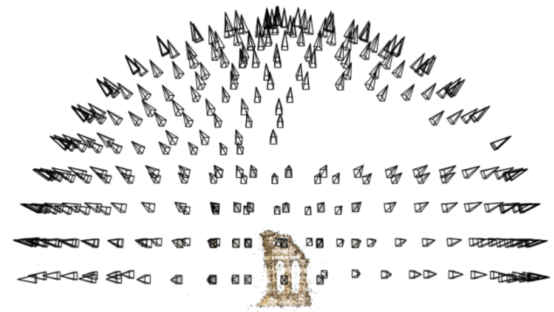
Readings

- Szeliski, Chapter 7.1 – 7.4, 11.6

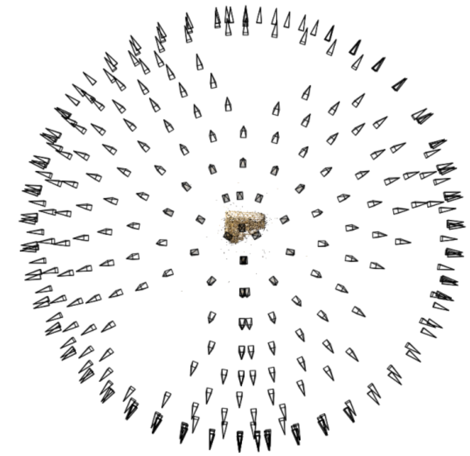
Announcements

- Project 2b due on Tuesday by 10:59pm
- Final project proposals feedback soon

Structure from motion



Reconstruction (side)



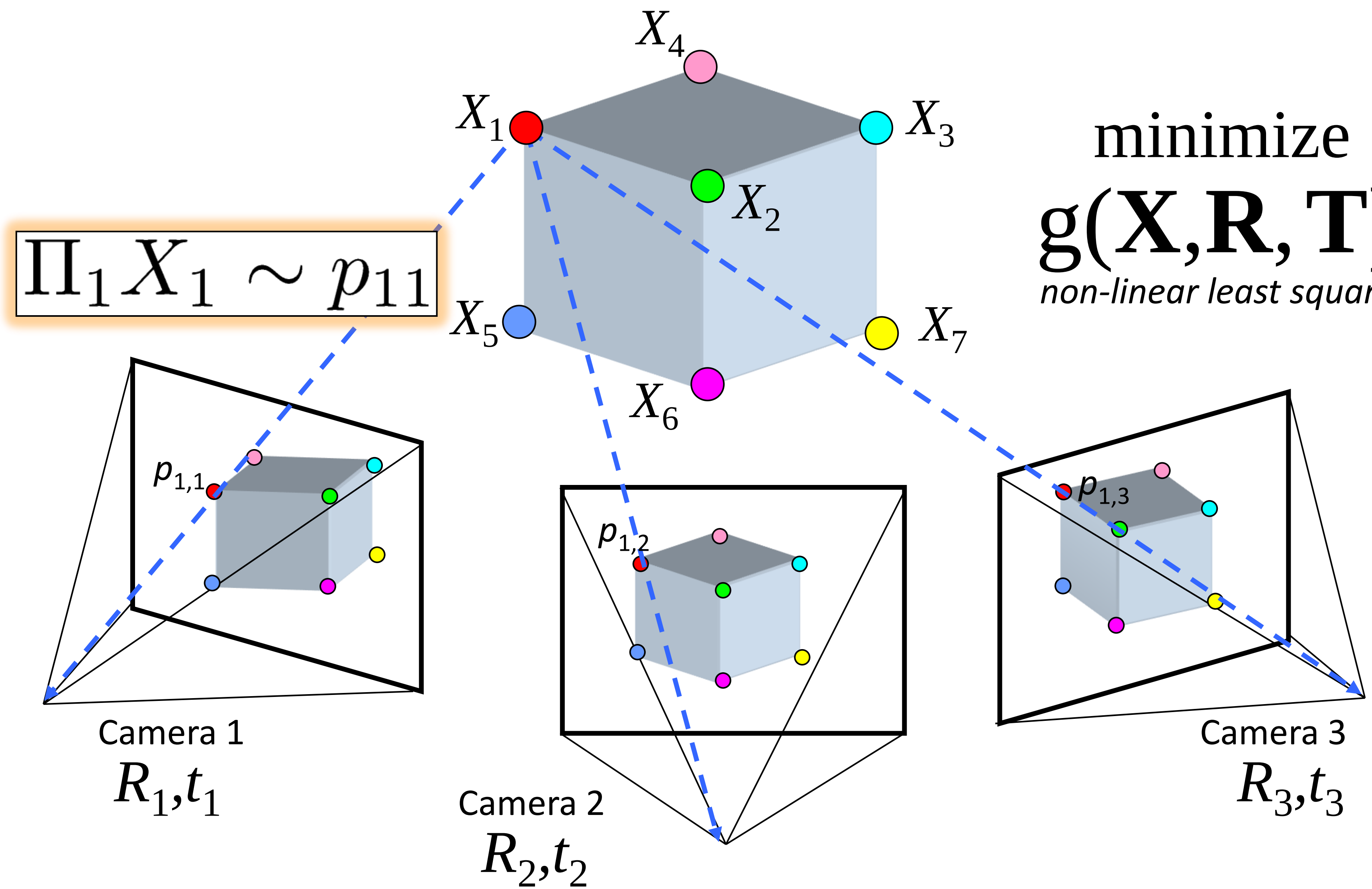
(top)

- Input: images with points in correspondence
  $p_{i,j} = (u_{i,j}, v_{i,j})$

- Output
  - structure: 3D location $\mathbf{x}_i$ for each point p_i
  - motion: camera parameters $\mathbf{R}_j$, $\mathbf{t}_j$ possibly $\mathbf{K}_j$

- Objective function: minimize *reprojection error*

Structure from motion



minimize

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

non-linear least squares

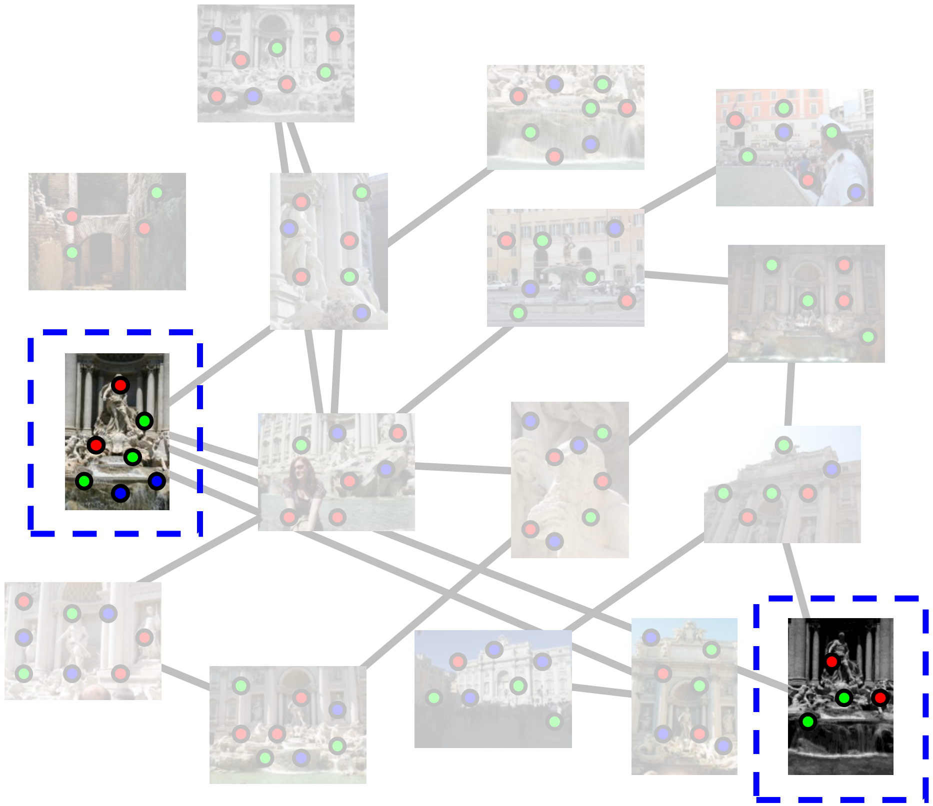
Structure from motion

- Minimize sum of squared reprojection errors:

$$g(\mathbf{X}, \mathbf{R}, \mathbf{T}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \underbrace{w_{ij}}_{\substack{\text{indicator variable:} \\ \text{is point } i \text{ visible in image } j\text{ ?}}} \cdot \left\| \underbrace{\mathbf{P}(\mathbf{x}_i, \mathbf{R}_j, \mathbf{t}_j)}_{\substack{\textit{predicted} \\ \text{image location}}} - \underbrace{\begin{bmatrix} u_{i,j} \\ v_{i,j} \end{bmatrix}}_{\substack{\textit{observed} \\ \text{image location}}} \right\|^2$$

- Minimizing this function is called *bundle adjustment*
  - Optimized using non-linear least squares, e.g. Levenberg-Marquardt

Incremental structure from motion



Incremental structure from motion

Incremental structure from motion

Photo Explorer

Demo


Why is this hard?

Libration

From Wikipedia, the free encyclopedia

Not to be confused with Liberation or Libation.

In astronomy **libration** (from the Latin verb *librare* "to balance, to sway", cf. *libra* "scales") refers to the various orbital conditions which make it possible to see more than 50% of the moon's surface over time, even though the front of the Moon is tidally locked to always face towards Earth. By extension, libration can also be used to describe the same phenomenon for other orbital bodies that are nominally locked to present the same face. As the orbital processes are repetitive, libration is manifested as a slow rocking back and forth (or up and down) of the face of the orbital body as viewed from the parent body, much like the rocking of a pair of scales about the point of balance.

Simulated views of the Moon over one month, demonstrating librations in latitude and longitude.

In the specific case of the Moon's librations, this motion permits a terrestrial observer to see slightly differing halves of the Moon's surface at different times. This means that a total of 59% of the Moon's surface can be observed from Earth.

There are three types of libration:

- *Libration in longitude* is a consequence of the Moon's orbit around Earth being somewhat eccentric, so that the Moon's rotation sometimes leads and sometimes lags its orbital position.
- *Libration in latitude* is a consequence of the Moon's axis of rotation being slightly inclined to the normal to the plane of its orbit around Earth. Its origin is analogous to the way in which the seasons arise from Earth's revolution about the Sun.
- *Diurnal libration* is a small daily oscillation due to the Earth's rotation, which carries an observer first to one side and then to the other side of the straight line joining Earth's center to the Moon's center, allowing the observer to look first around one side of the Moon and then around the other. This is because the observer is on the surface of the Earth, not at its centre.

WIKIPEDIA
The Free Encyclopedia

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

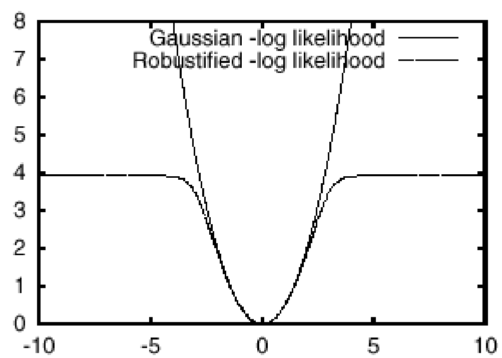
search
Go Search

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Extensions to SfM

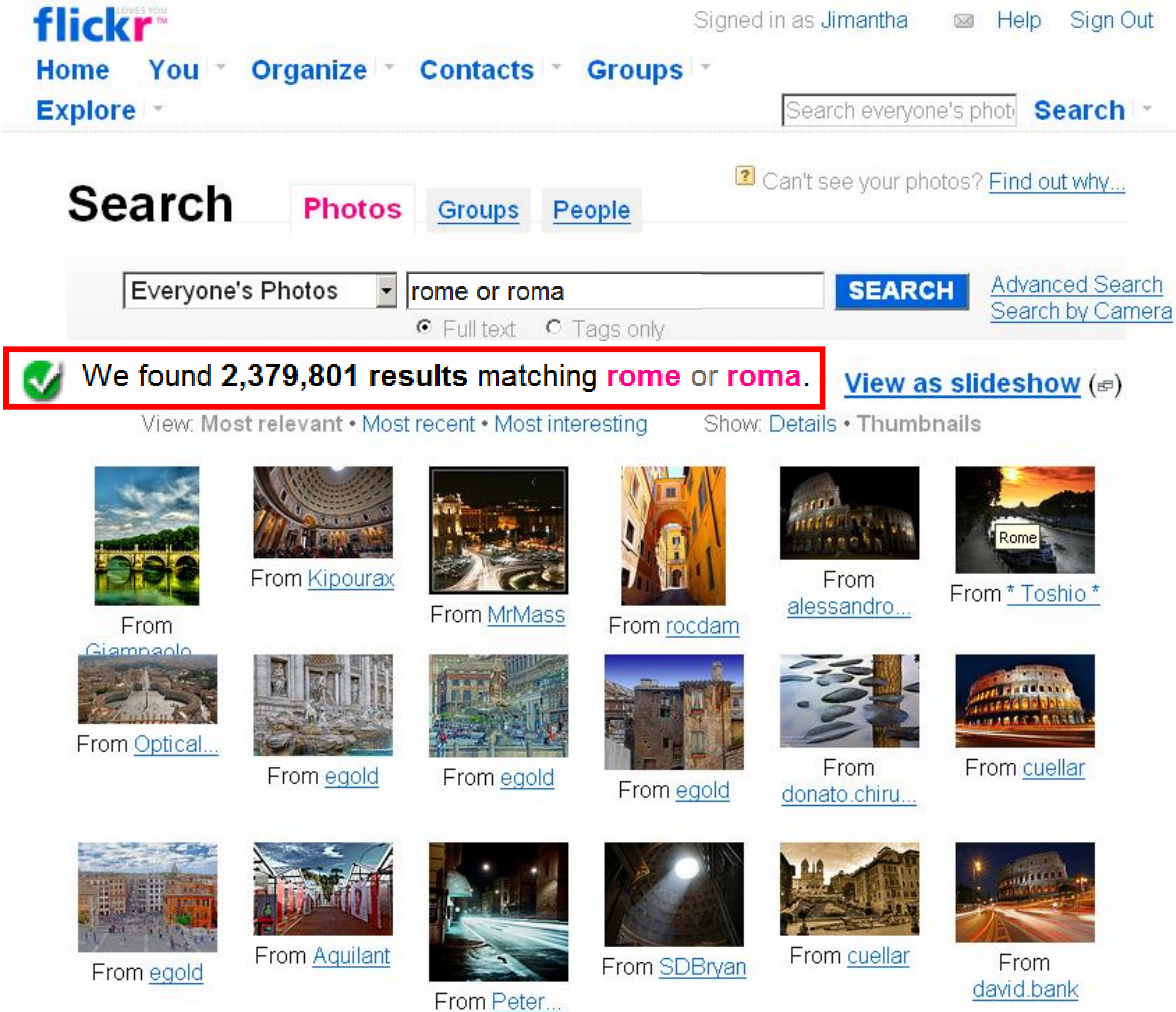
- Can also solve for intrinsic parameters (focal length, radial distortion, etc.)
- Can use a more robust function than squared error, to avoid fitting to outliers

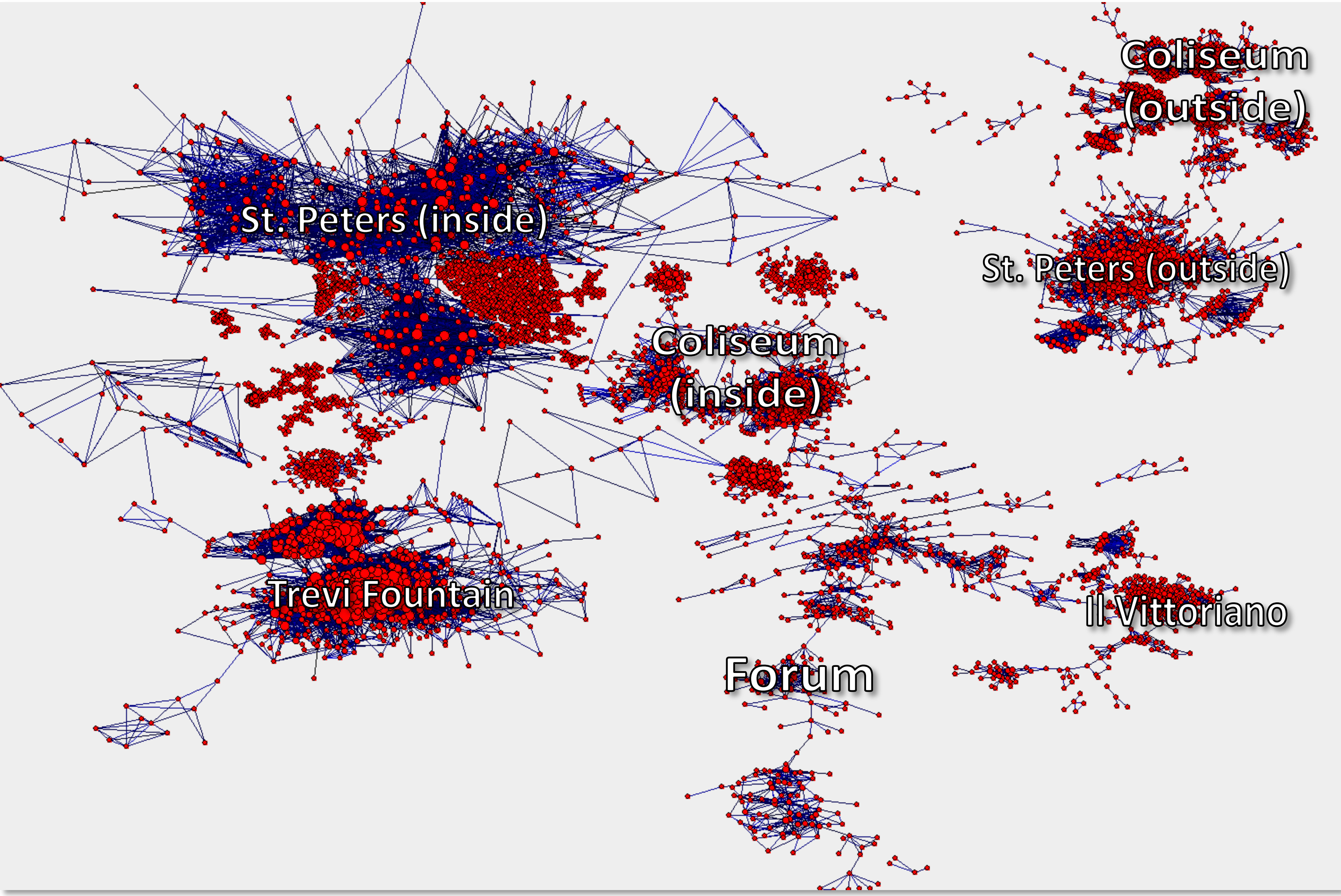


- For more information, see: Triggs, *et al*, “Bundle Adjustment – A Modern Synthesis”, *Vision Algorithms* 2000.

Questions?

Can we reconstruct entire cities?


flickr
Signed in as Jimantha
Help
Sign Out
Home
You
Organize
Contacts
Groups
Explore
Search everyone's phot
Search
Can't see your photos? Find out why...
Search
Photos
Groups
People
Everyone's Photos
rome or roma
SEARCH
Advanced Search
Search by Camera
Full text
Tags only
We found 2,379,801 results matching rome or roma.
View as slideshow
View: Most relevant • Most recent • Most interesting
Show: Details • Thumbnails
From Giampaolo
From Kipourax
From MrMass
From rocdam
From alessandro...
Rome
From * Toshio *
From Optical...
From egold
From egold
From egold
From donato.chiru...
From cuellar
From egold
From Aquilant
From Peter...
From SDBryan
From cuellar
From david.bank


Coliseum (outside)
St. Peters (inside)
St. Peters (outside)
Coliseum (inside)
Trevi Fountain
Il Vittoriano
Forum

Gigantic matching problem

- 1,000,000 images → 500,000,000,000 pairs
  - Matching all of these on a 1,000-node cluster would take more than a year, even if we match 10,000 every second
  - And involves TBs of data

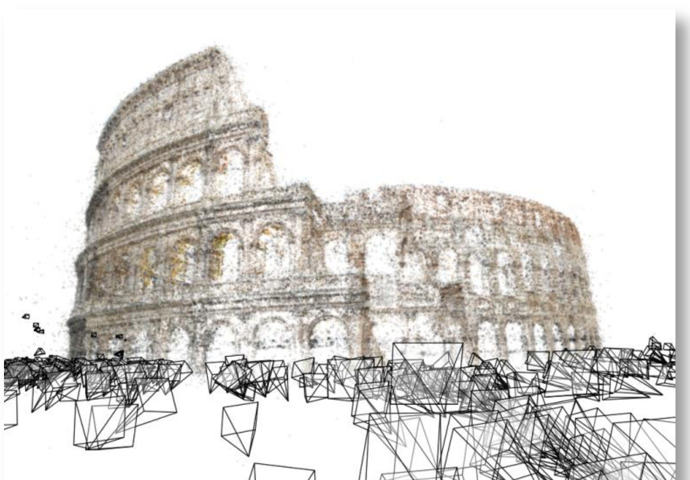
- The vast majority (>99%) of image pairs do *not* match
- There are better ways of finding matching images (more on this later)

Gigantic SfM problem

- Largest problem size we’ve seen:
  - 15,000 cameras
  - 4 million 3D points
  - more than 12 million parameters
  - more than 25 million equations
- Huge optimization problem
- Requires *sparse least squares* techniques

Building Rome in a Day

St. Peter's Basilica



Colosseum

Trevi Fountain

Rome, Italy. Reconstructed 150,000 in 21 hours on 496 machines

Dubrovnik, Croatia. 4,619 images (out of an initial 57,845).
Total reconstruction time: 23 hours
Number of cores: 352

Dubrovnik

Dubrovnik, Croatia. 4,619 images (out of an initial 57,845).
Total reconstruction time: 23 hours
Number of cores: 352

San Marco Square

San Marco Square and environs, Venice. 14,079 photos, out of an initial 250,000.
Total reconstruction time: 3 days. Number of cores: 496.

Multi-view stereo

Stereo

Multi-view stereo

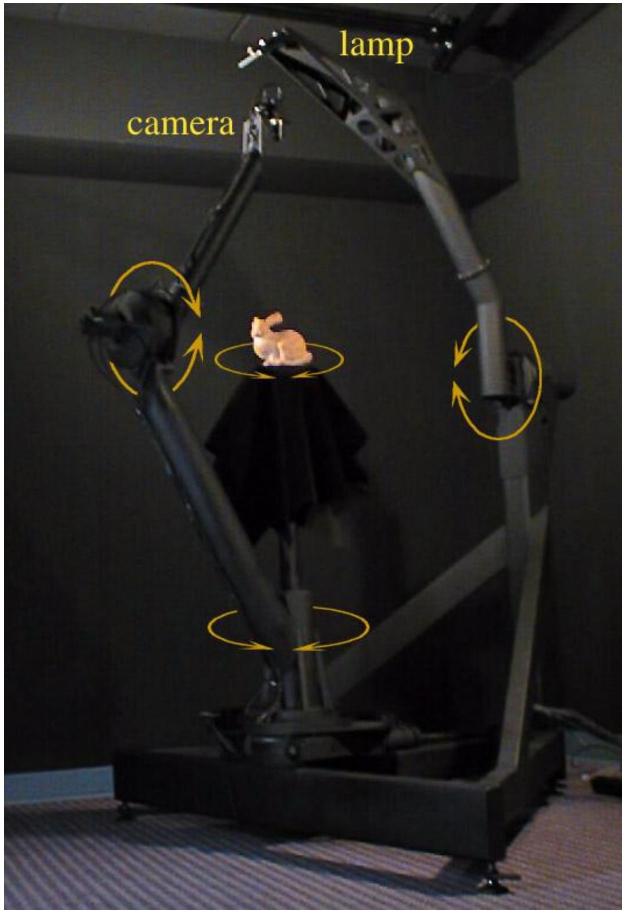
Multi-view Stereo

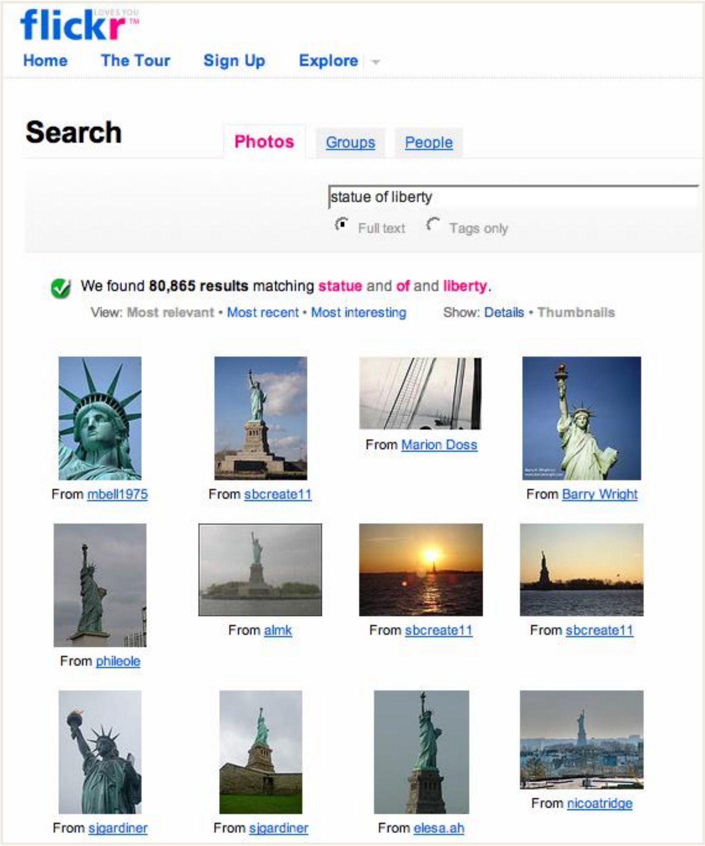
Point Grey's Bumblebee XB3

Point Grey's ProFusion 25

CMU's 3D Room

Multi-view Stereo

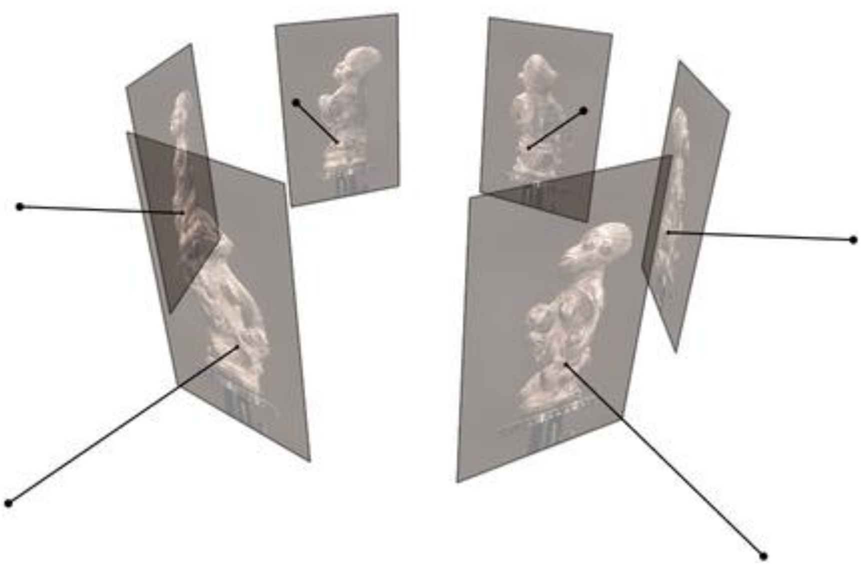




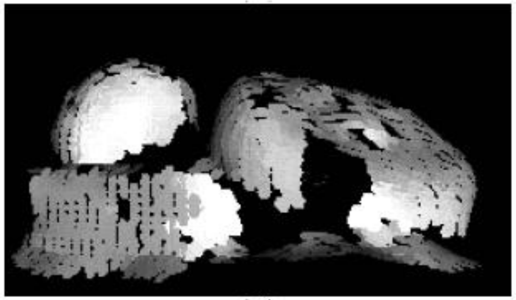
Multi-view Stereo

Input: calibrated images from several viewpoints

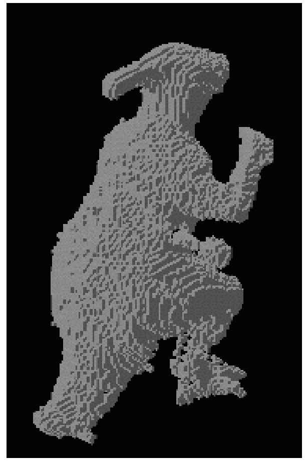
Output: 3D object model



Figures by Carlos Hernandez



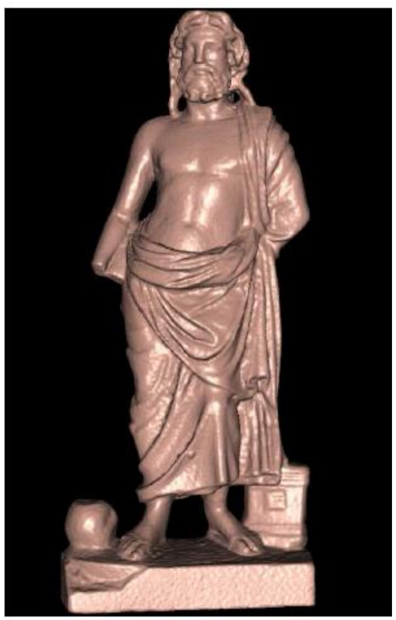
Fua
1995



Seitz, Dyer
1997

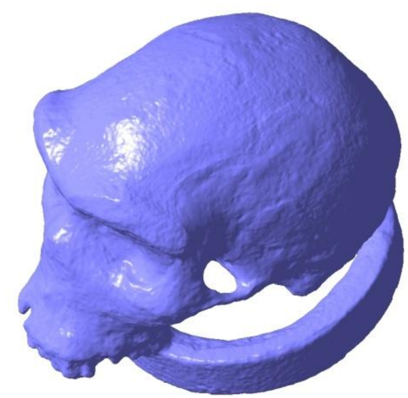
Narayanan, Rander, Kanade
1998

Faugeras, Keriven
1998



Hernandez, Schmitt
2004

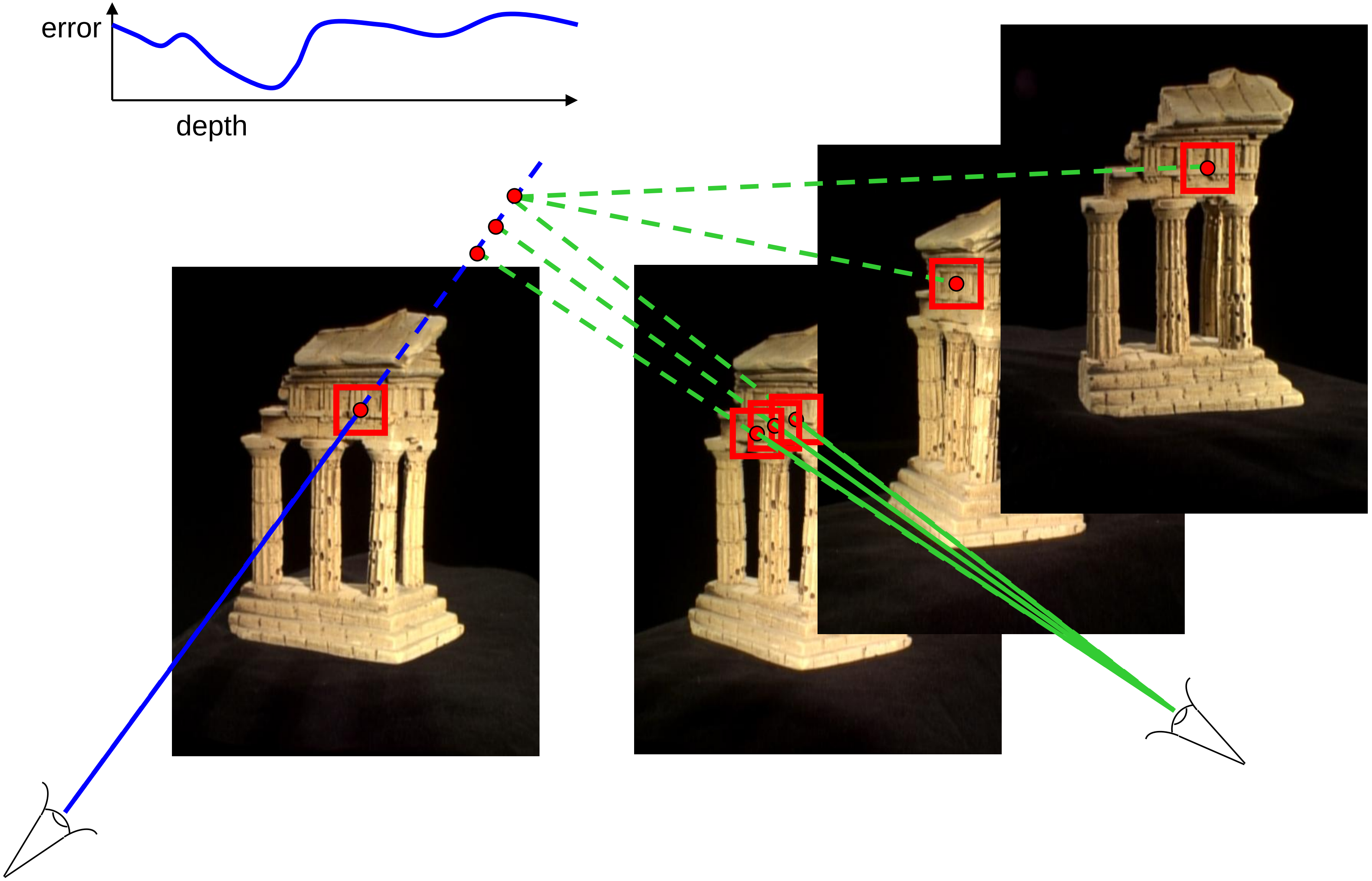
Pons, Keriven, Faugeras
2005



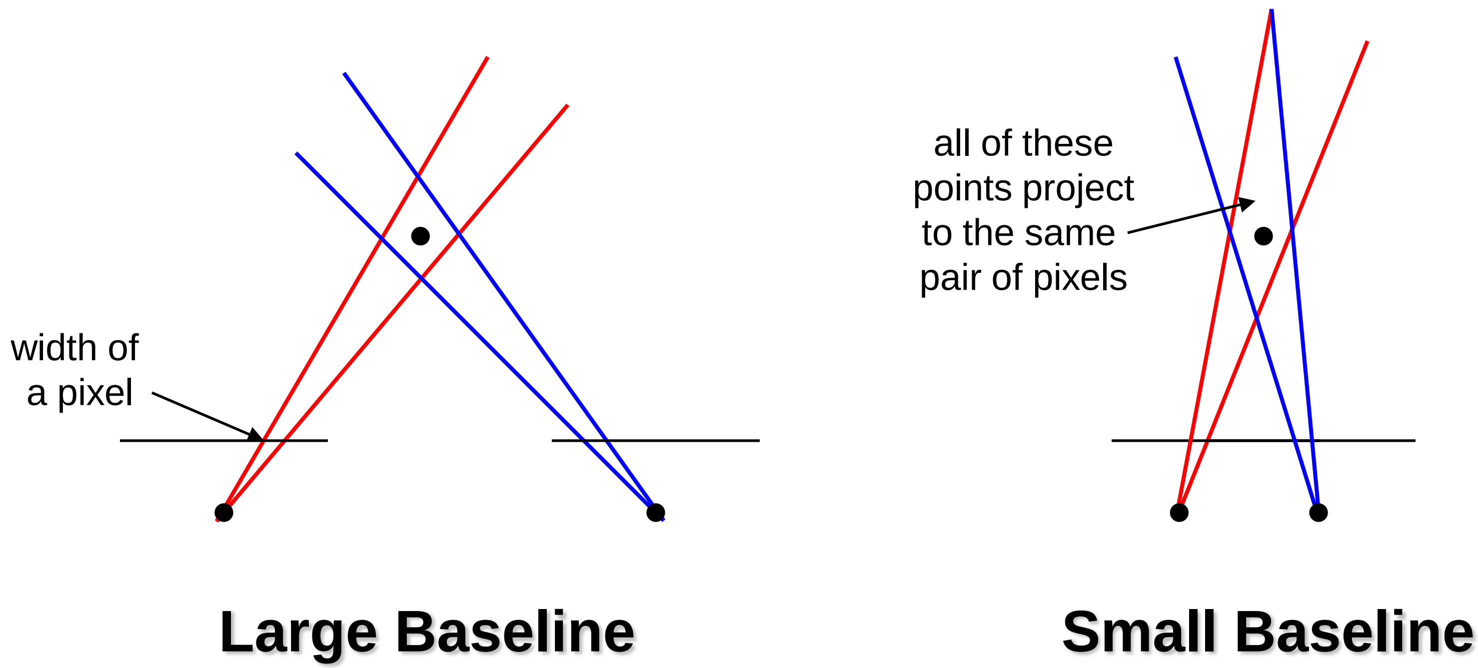
Furukawa, Ponce
2006

Goesele et al.
2007

Stereo: basic idea



Choosing the stereo baseline

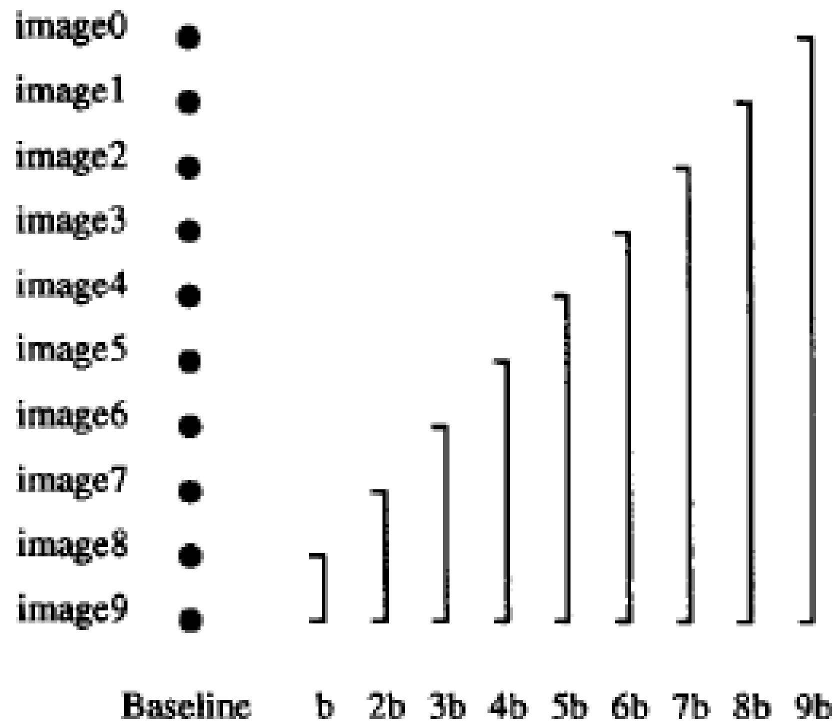


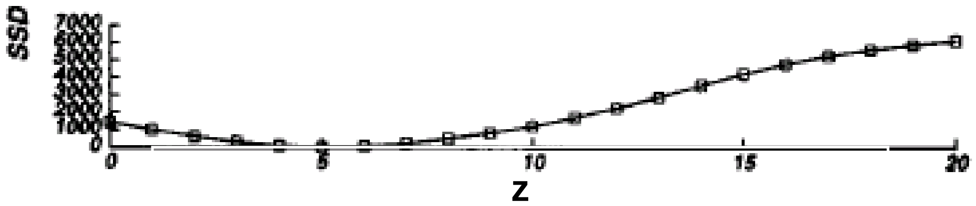
What's the optimal baseline?

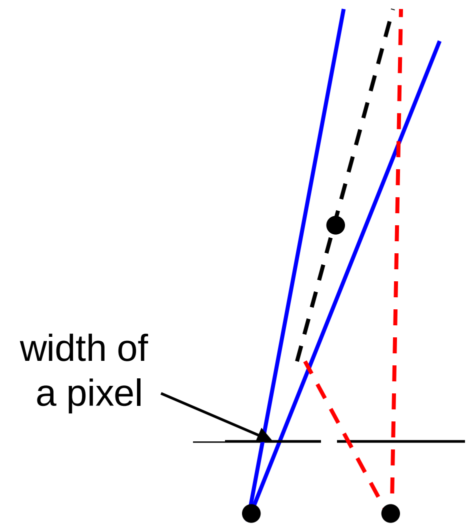
- Too small: large depth error
- Too large: difficult search problem

The Effect of Baseline on Depth Estimation

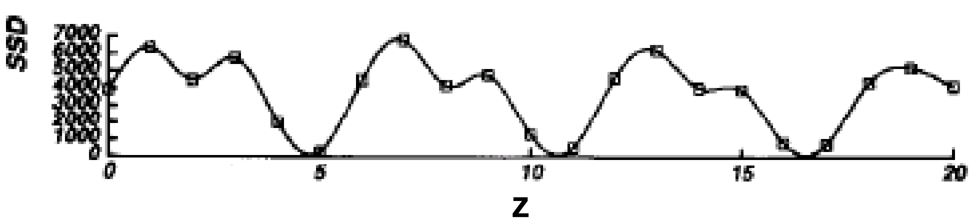
Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.

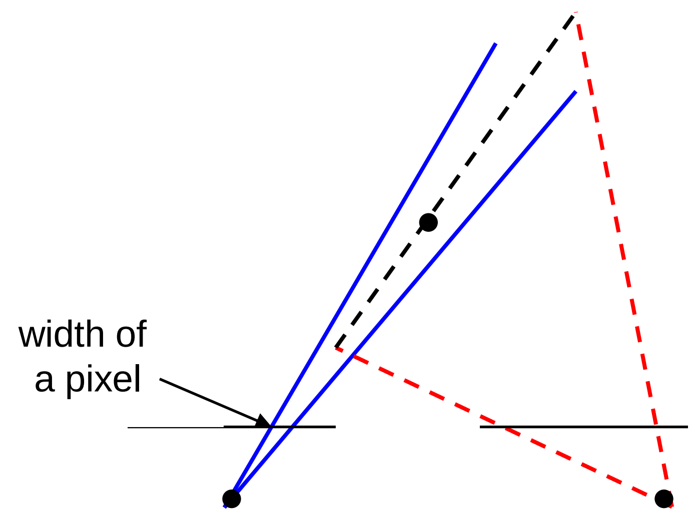






pixel matching score





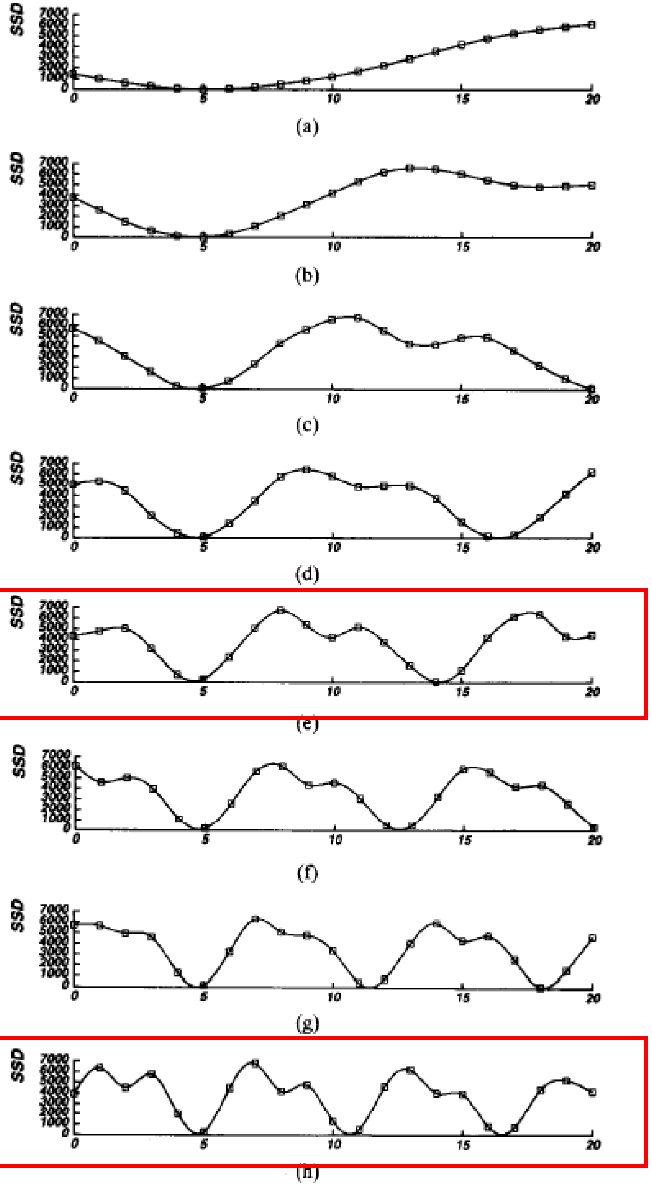


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

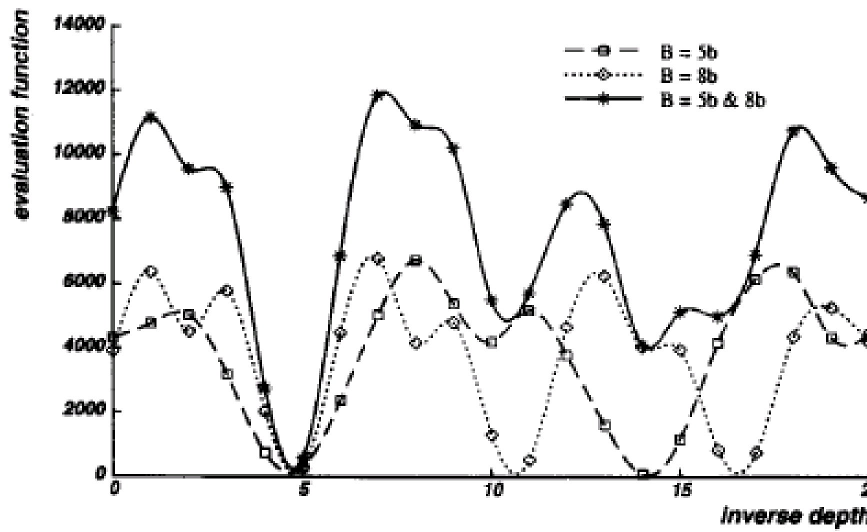


Fig. 6. Combining two stereo pairs with different baselines.

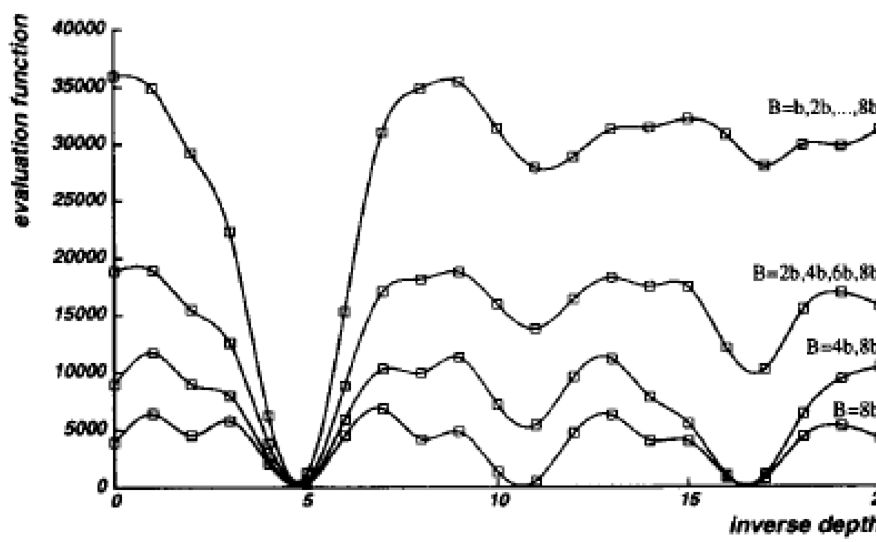


Fig. 7. Combining multiple baseline stereo pairs.

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*

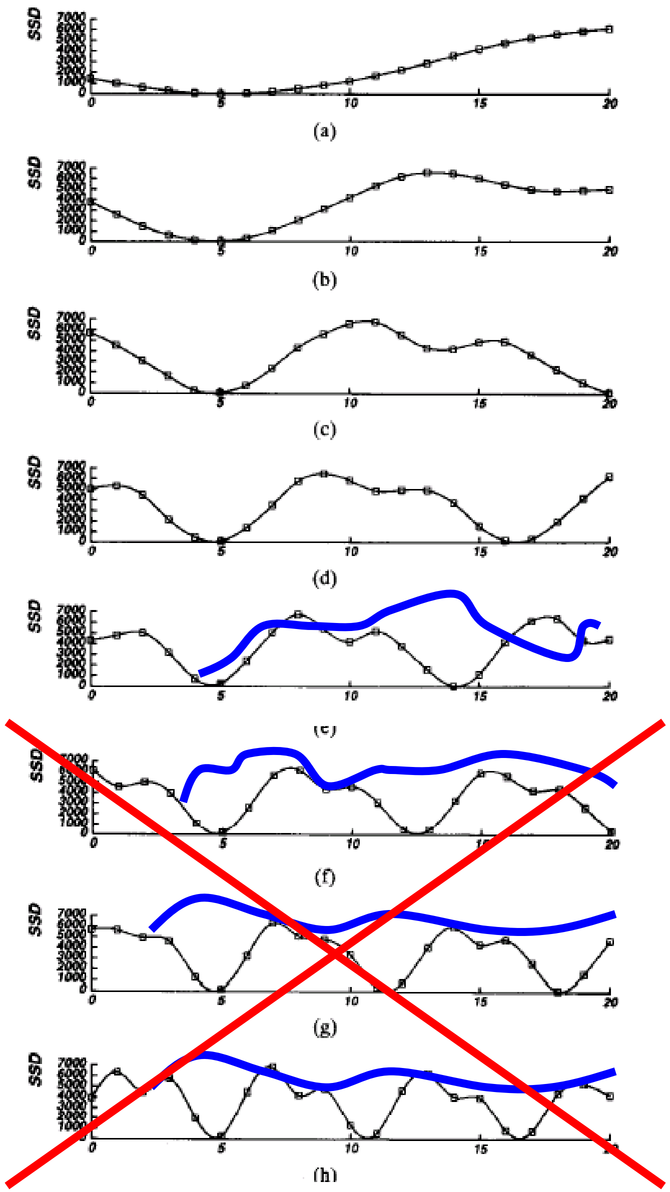


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

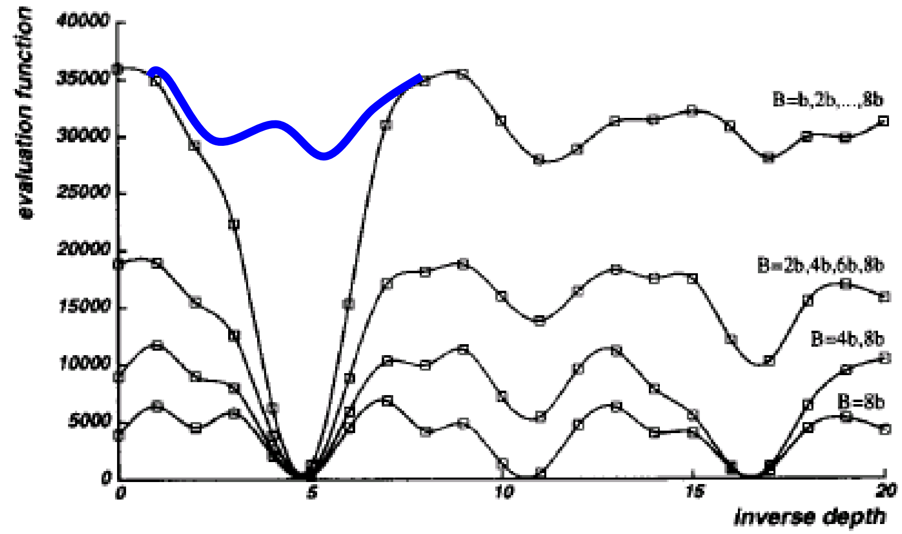


Fig. 7. Combining multiple baseline stereo pairs.

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?

Questions?