Photo Tourism:
Exploring Photo Collections in 3D

Noah Snavely

Steven M. Seitz

University of Washington

Richard Szeliski

Microsoft Research

Modified from authors' slides

15,464

37,383

76,389

15,464

37,383

76,389

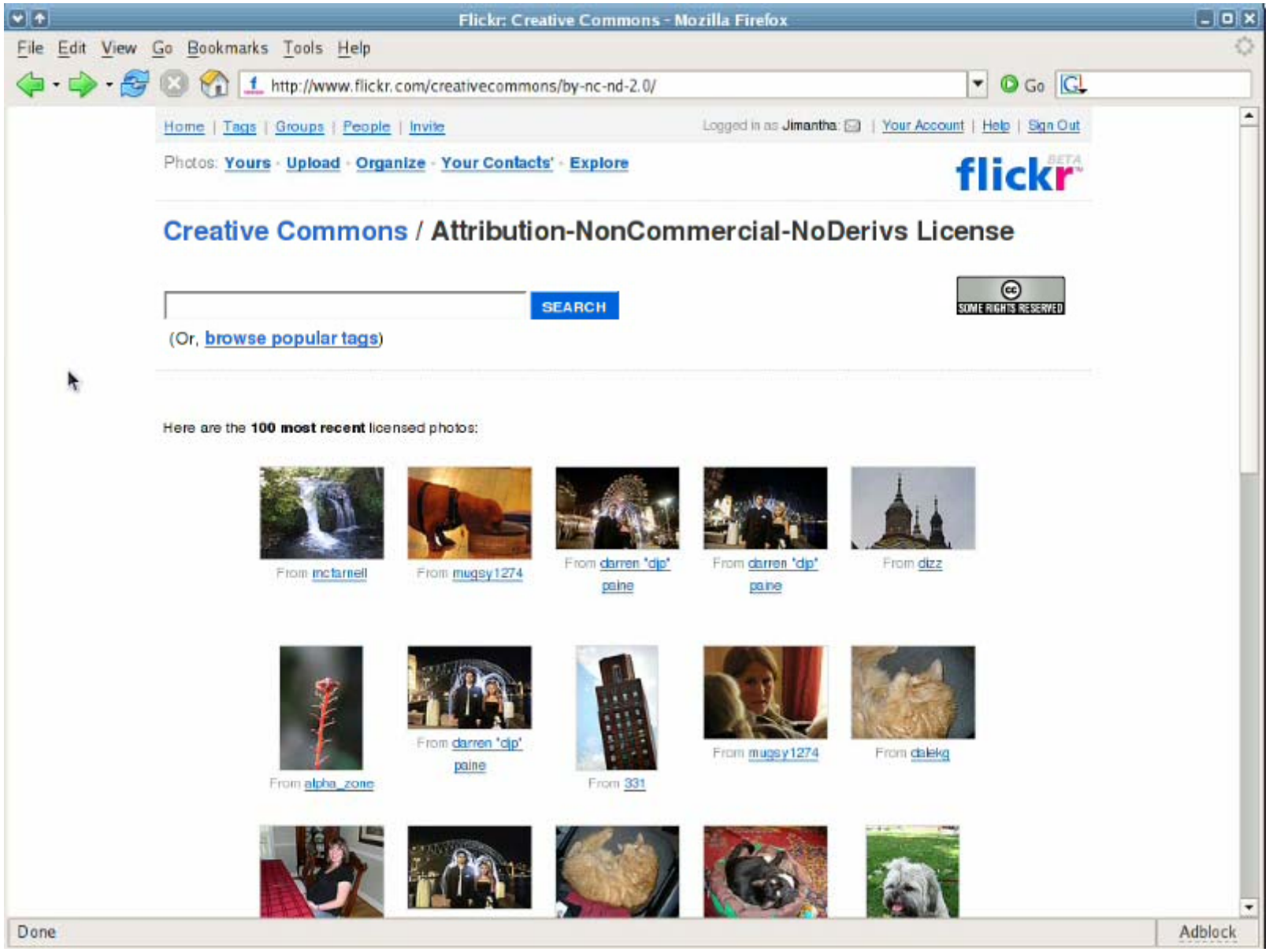
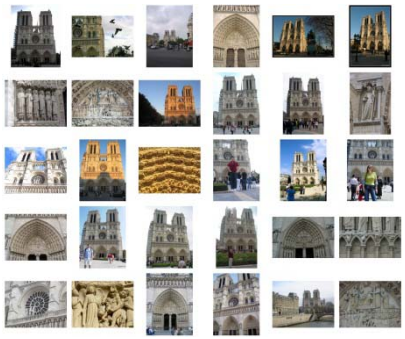


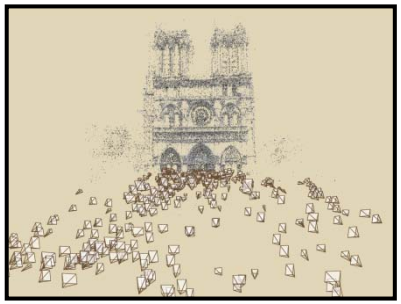
Photo Tourism

Photo Tourism overview



Input photographs

Scene reconstruction



Relative camera positions and orientations

Point cloud

Sparse correspondence

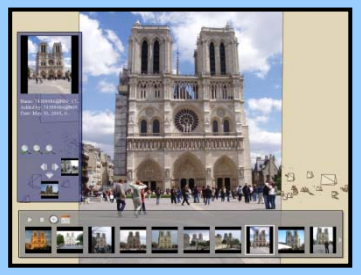


Photo Explorer

Related work

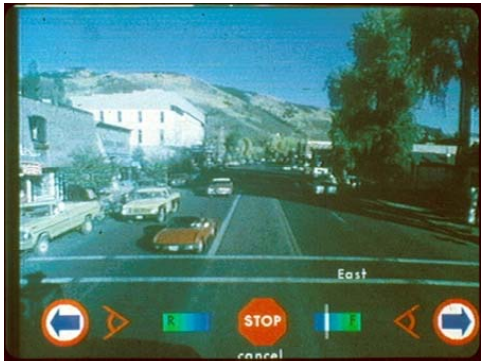
- Image-based modeling

Debevec, *et al.*
SIGGRAPH 1996

Schaffalitzky and Zisserman
ECCV 2002

Brown and Lowe
3DIM 2005

- Image-based rendering



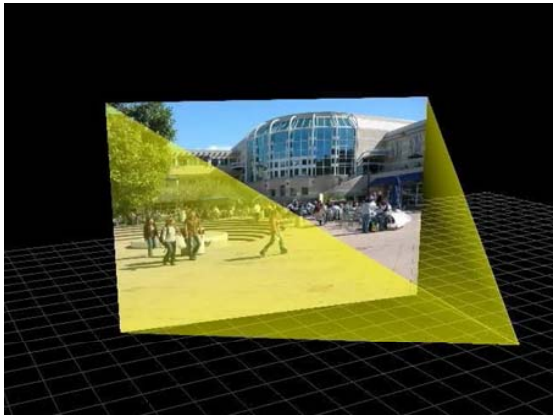
Aspen Movie Map
Lippman, *et al.*, 1978

Photorealistic IBR:
Levoy and Hanrahan, SIGGRAPH 1996
Gortler, *et al*, SIGGRAPH 1996
Seitz and Dyer, SIGGRAPH 1996
Aliaga, *et al*, SIGGRAPH 2001
and many others

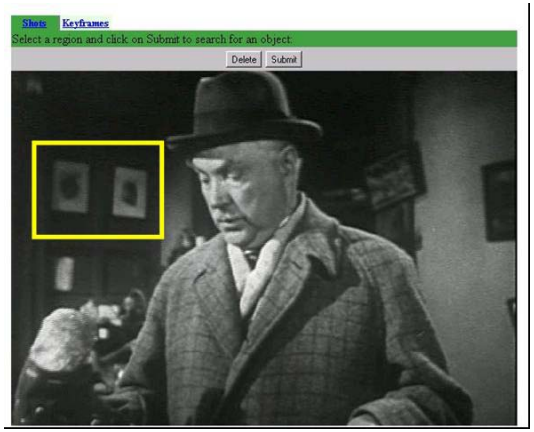
Related work

- Image browsing

Toyama, *et al*,
Int. Conf. Multimedia, 2003



McCurdy and Griswold
Mobisys 2003



Sivic and Zisserman
ICCV 2003

Photo Tourism overview

Input photographs

Scene reconstruction

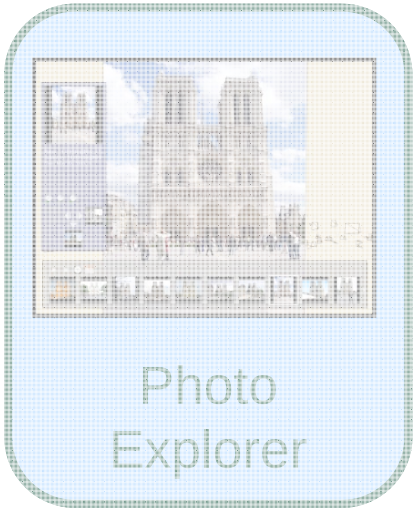
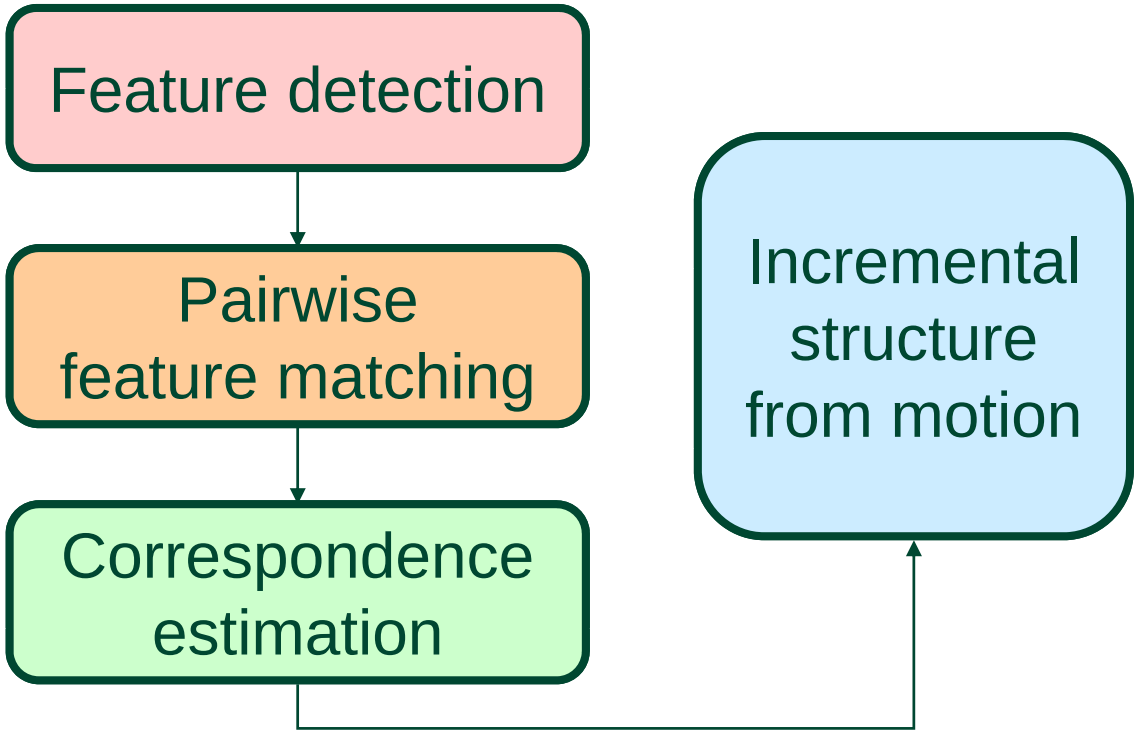


Photo Explorer

Scene reconstruction

- Automatically estimate
  - position, orientation, and focal length of cameras
  - 3D positions of feature points



Feature detection

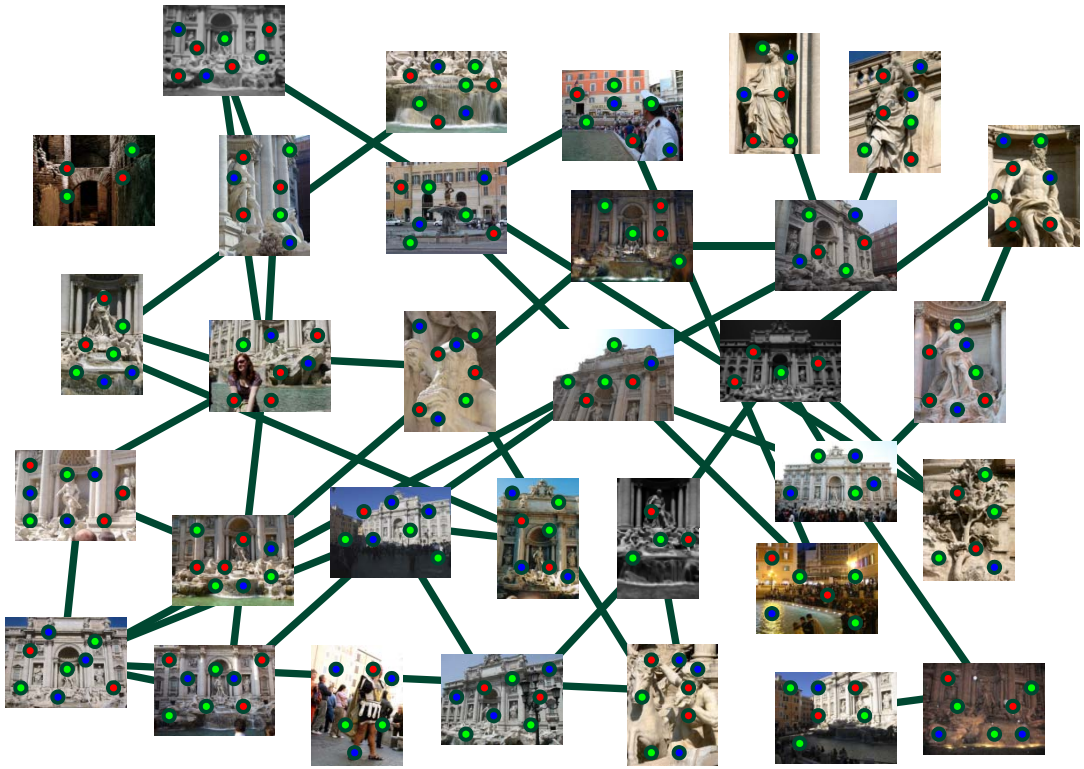
- Detect features using SIFT [Lowe, IJCV 2004]

Feature detection

- Detect features using SIFT [Lowe, IJCV 2004]

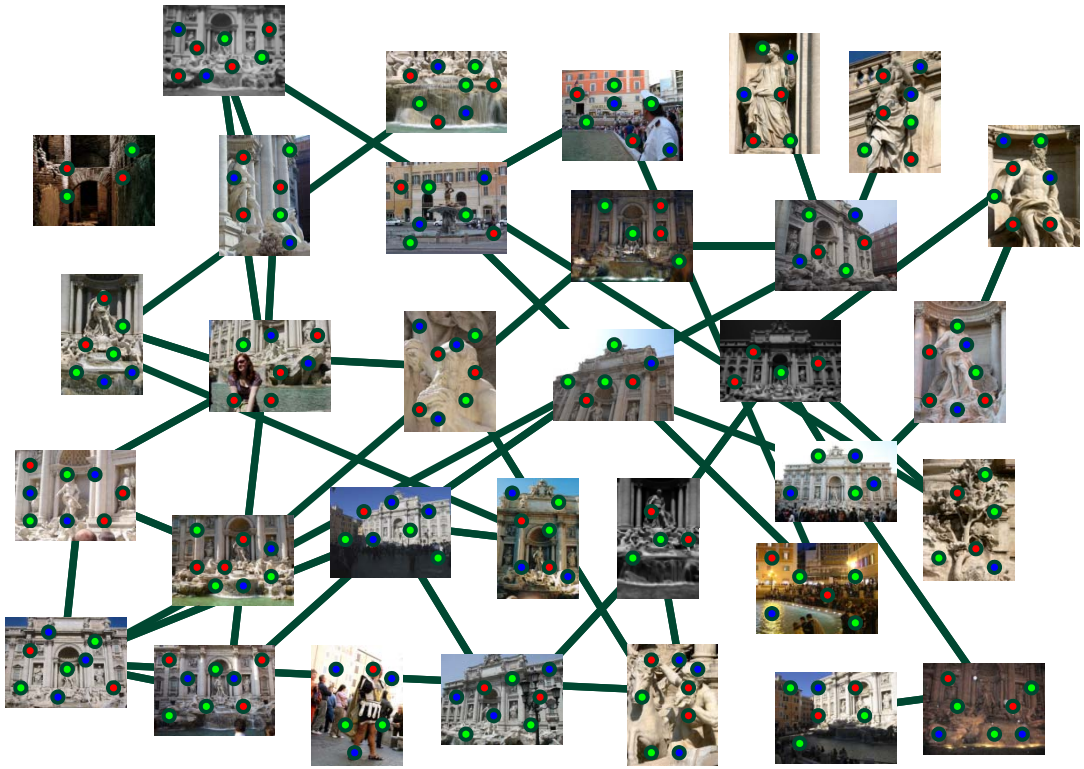
Pairwise feature matching

- Match features between each pair of images



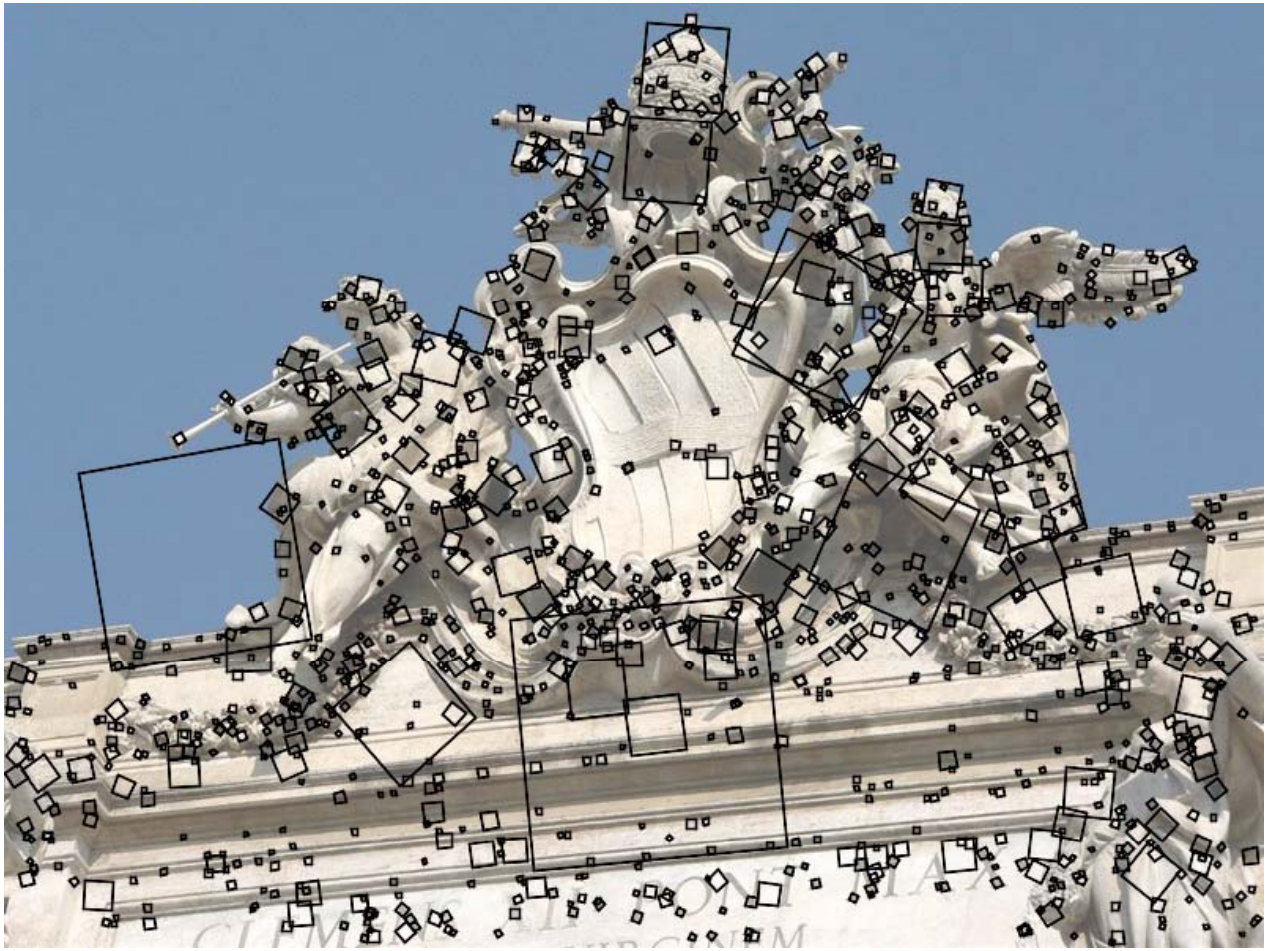
Pairwise feature matching

- Refine matching using RANSAC [Fischler & Bolles 1987] to estimate fundamental matrices between pairs



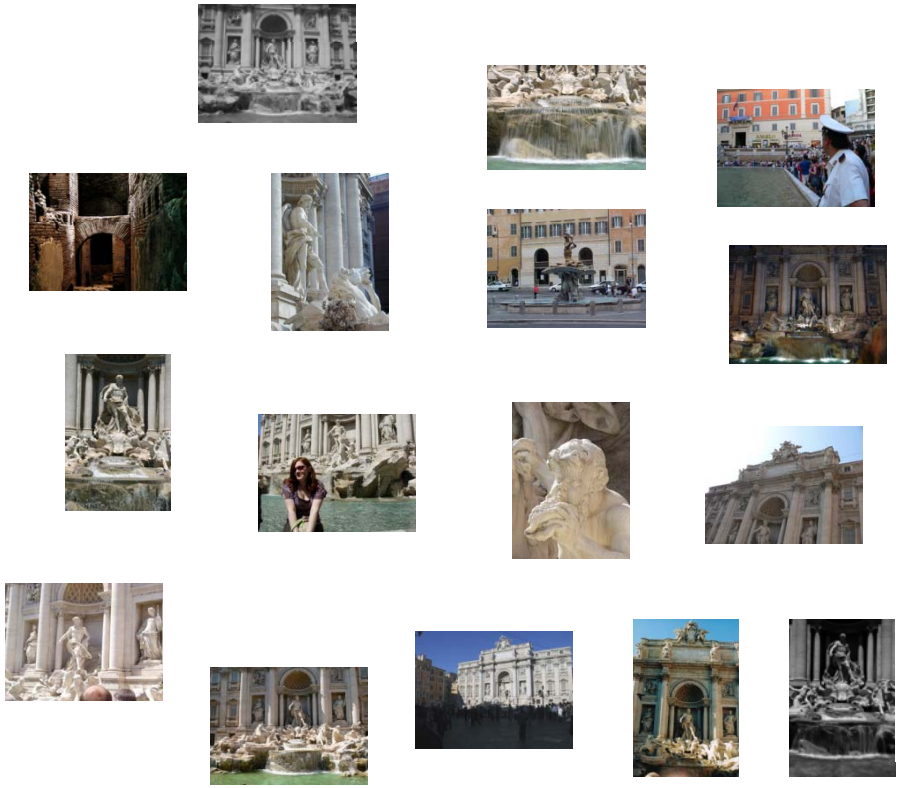
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



Feature detection

Detect features using SIFT [Lowe, IJCV 2004]

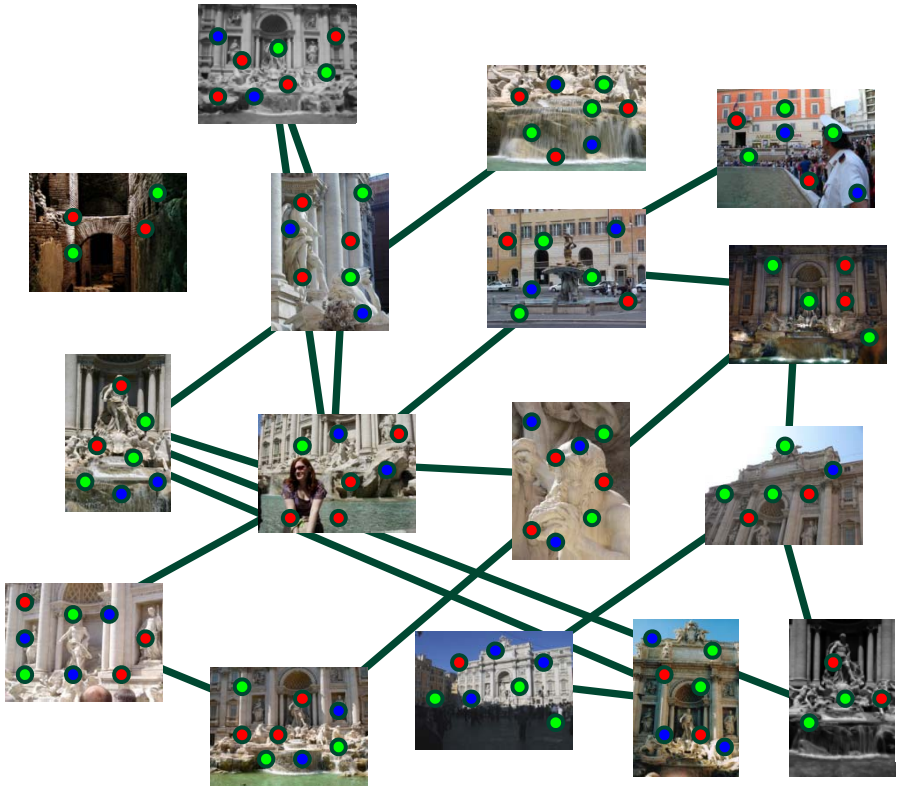


Feature detection

Detect features using SIFT [Lowe, IJCV 2004]

Feature matching

Match features between each pair of images



Correspondence estimation

- Link up pairwise matches to form connected components of matches across several images

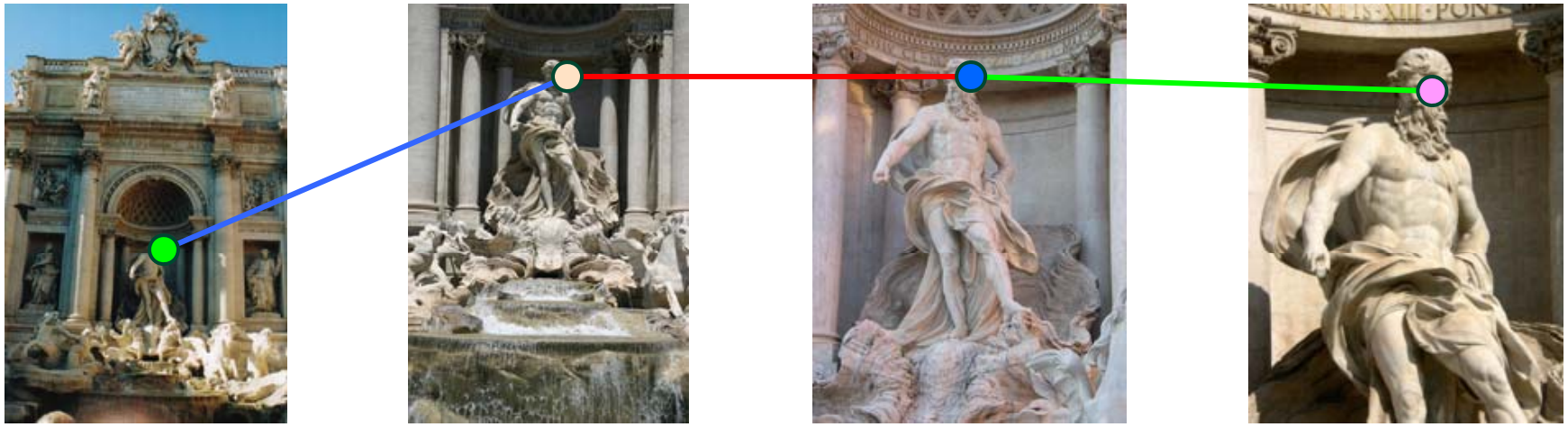
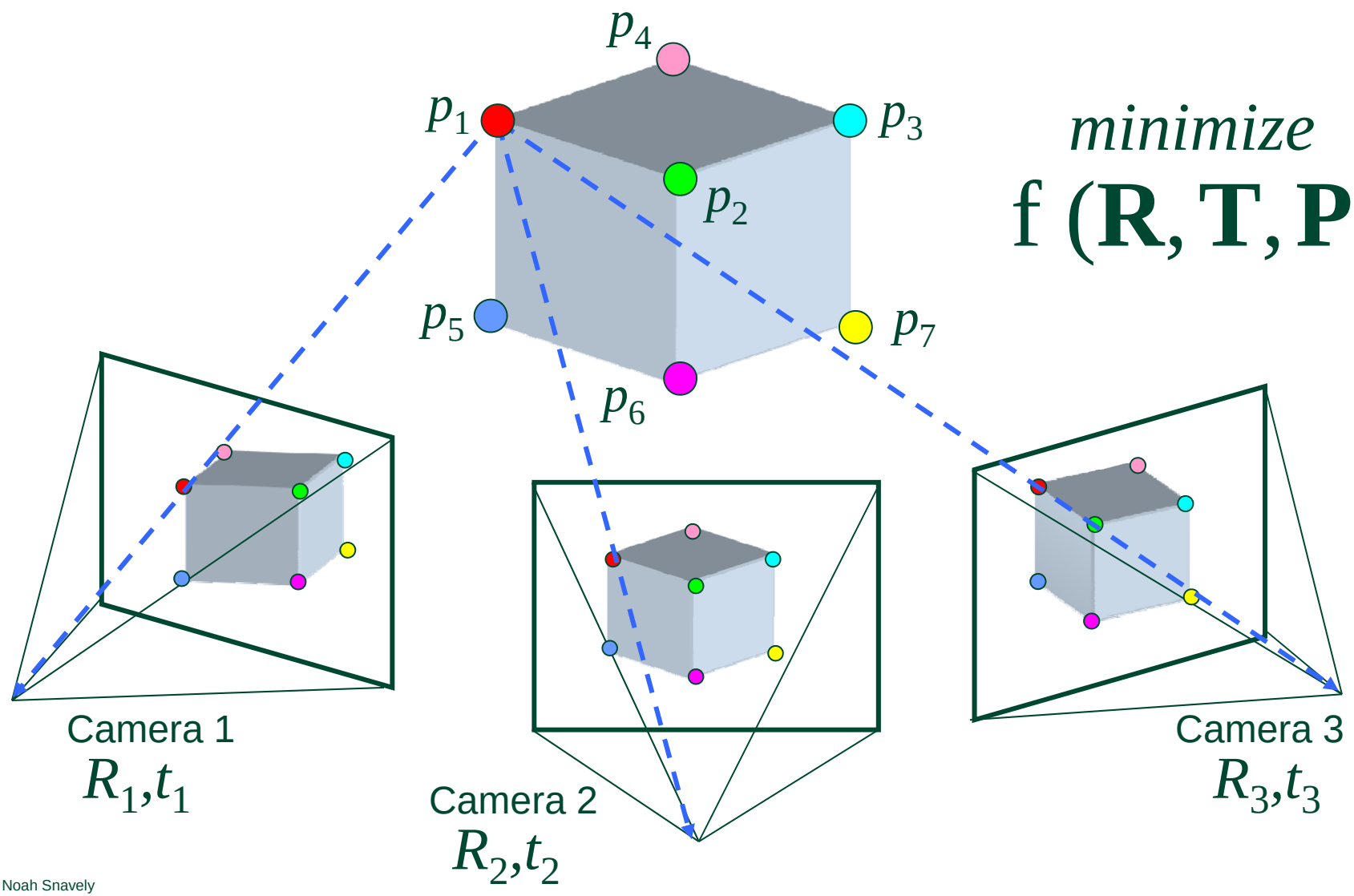


Image 1 Image 2 Image 3 Image 4

Structure from motion

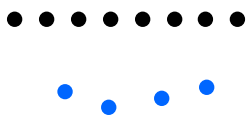


minimize

$$f(\mathbf{R},\mathbf{T},\mathbf{P})$$

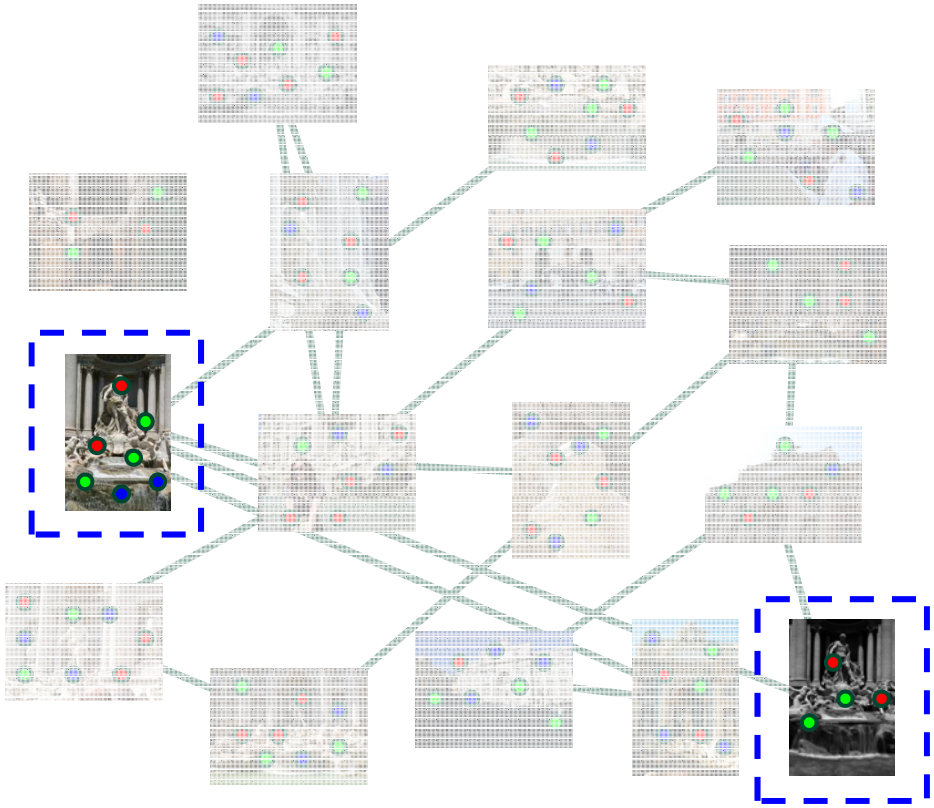
Incremental structure from motion

Incremental structure from motion

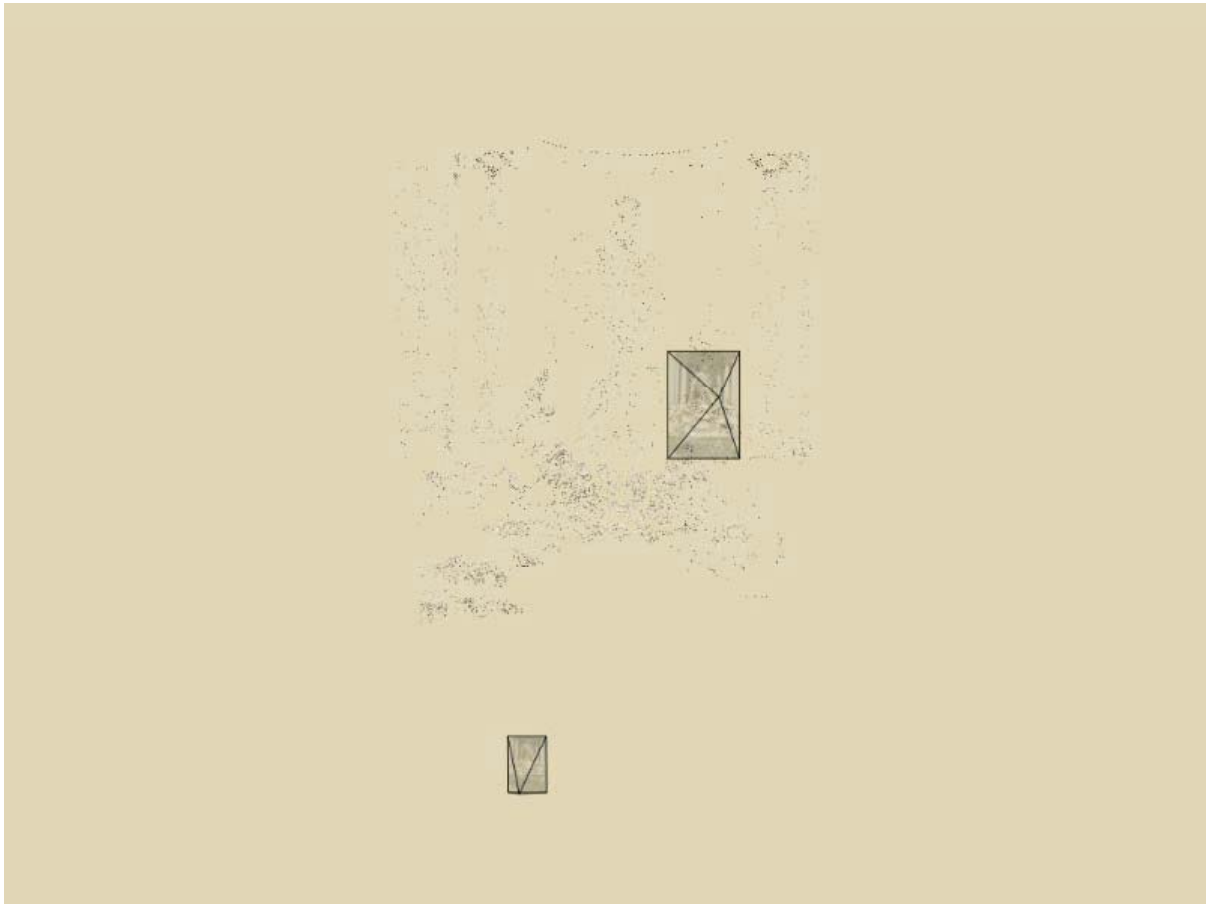


Incremental structure from motion

Incremental structure from motion



Incremental structure from motion



Incremental structure from motion

Reconstruction performance

- For photo sets from the Internet, 20% to 75% of the photos were registered
- Most unregistered photos belonged to different connected components

- Running time: < 1 hour for 80 photos

  > 1 week for 2600 photo

Photo Tourism overview

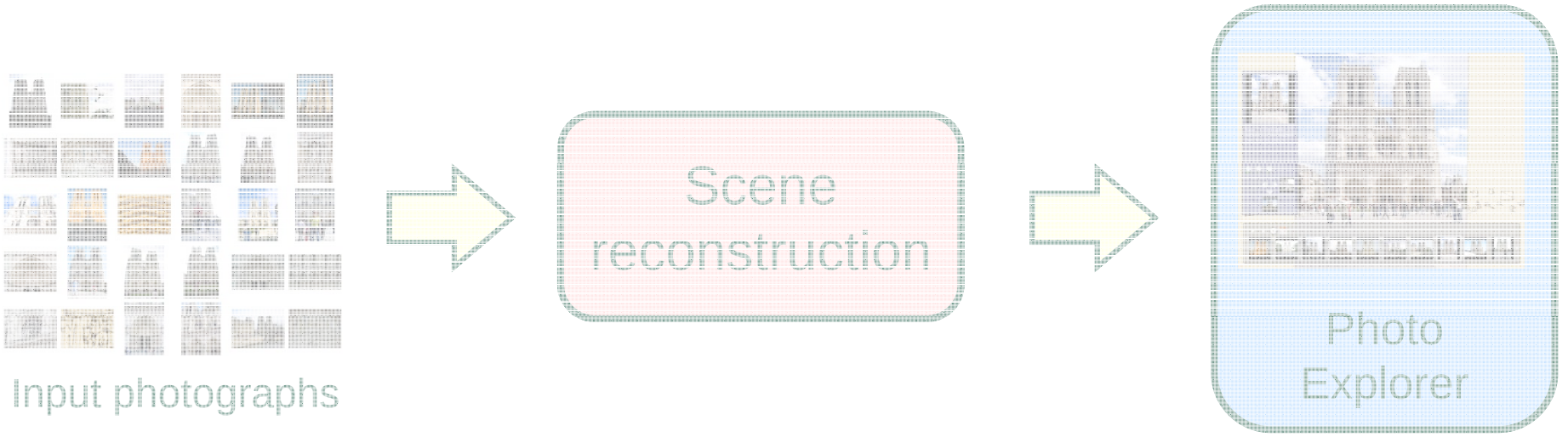


Photo Explorer

Demo

Photo Tourism overview

Input photographs

Scene reconstruction

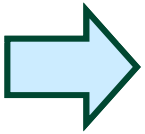
Photo Explorer

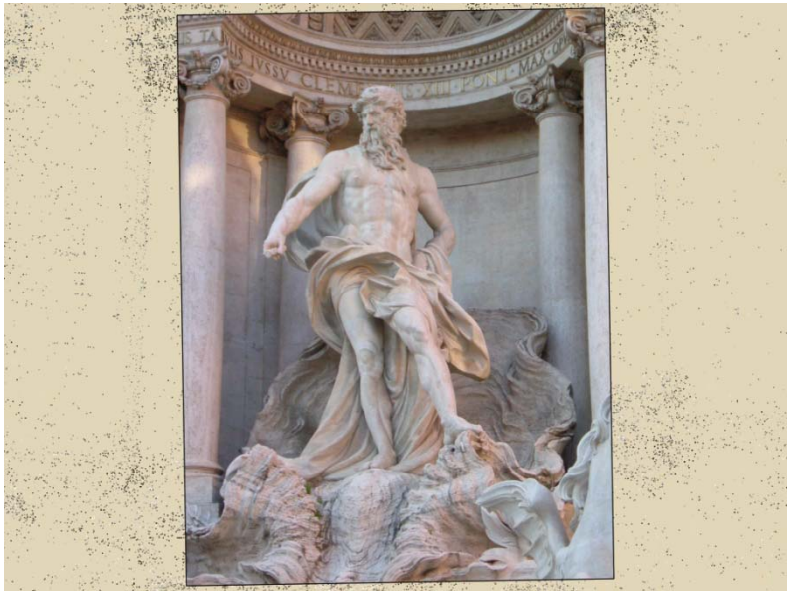
- Navigation
- Rendering
- Annotations

Navigation controls

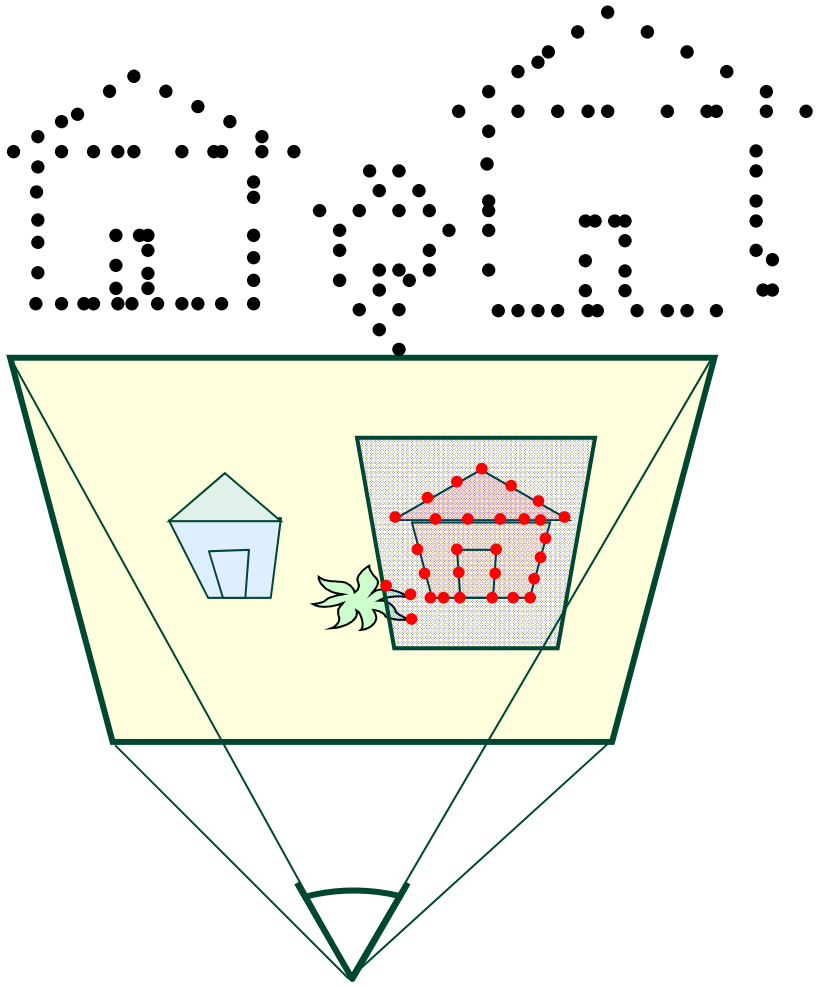
- Free-flight navigation
- Object-based browsing
- Relation-based browsing
- Overhead map

Object-based browsing

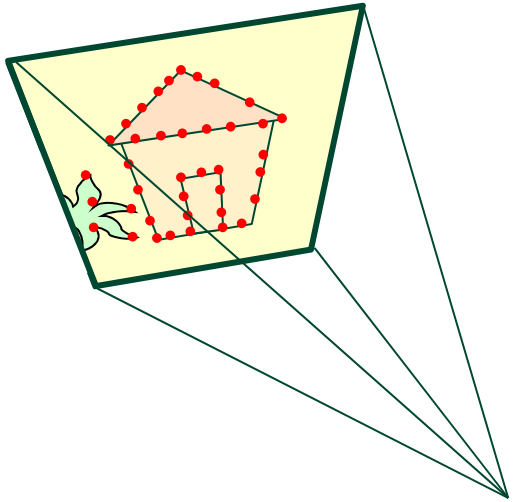




Object-based browsing



- Visibility
- Resolution
- Head-on view



Relation-based browsing

Find all similar images

Find all details

Find all zoom outs

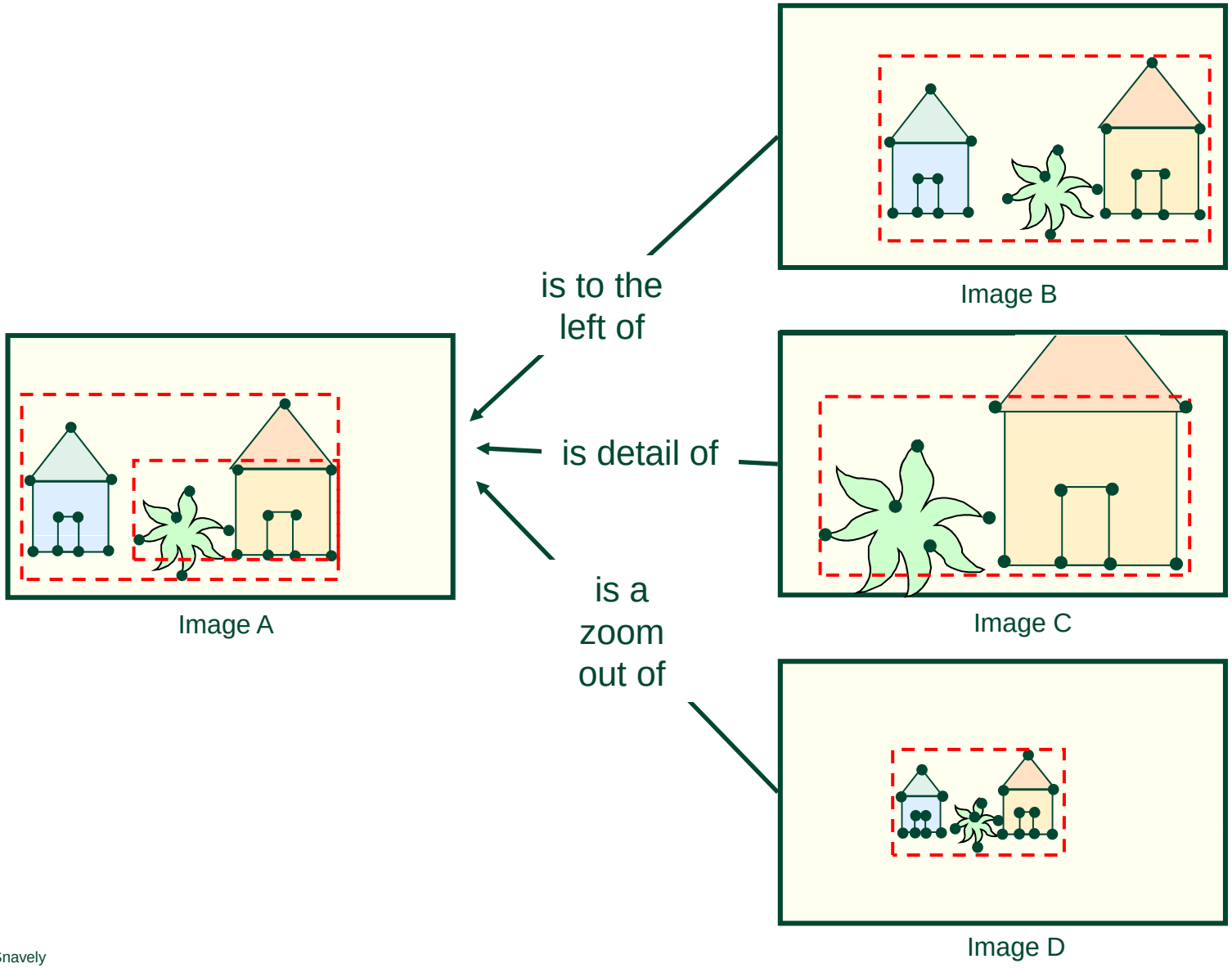
Zoom in

Move left

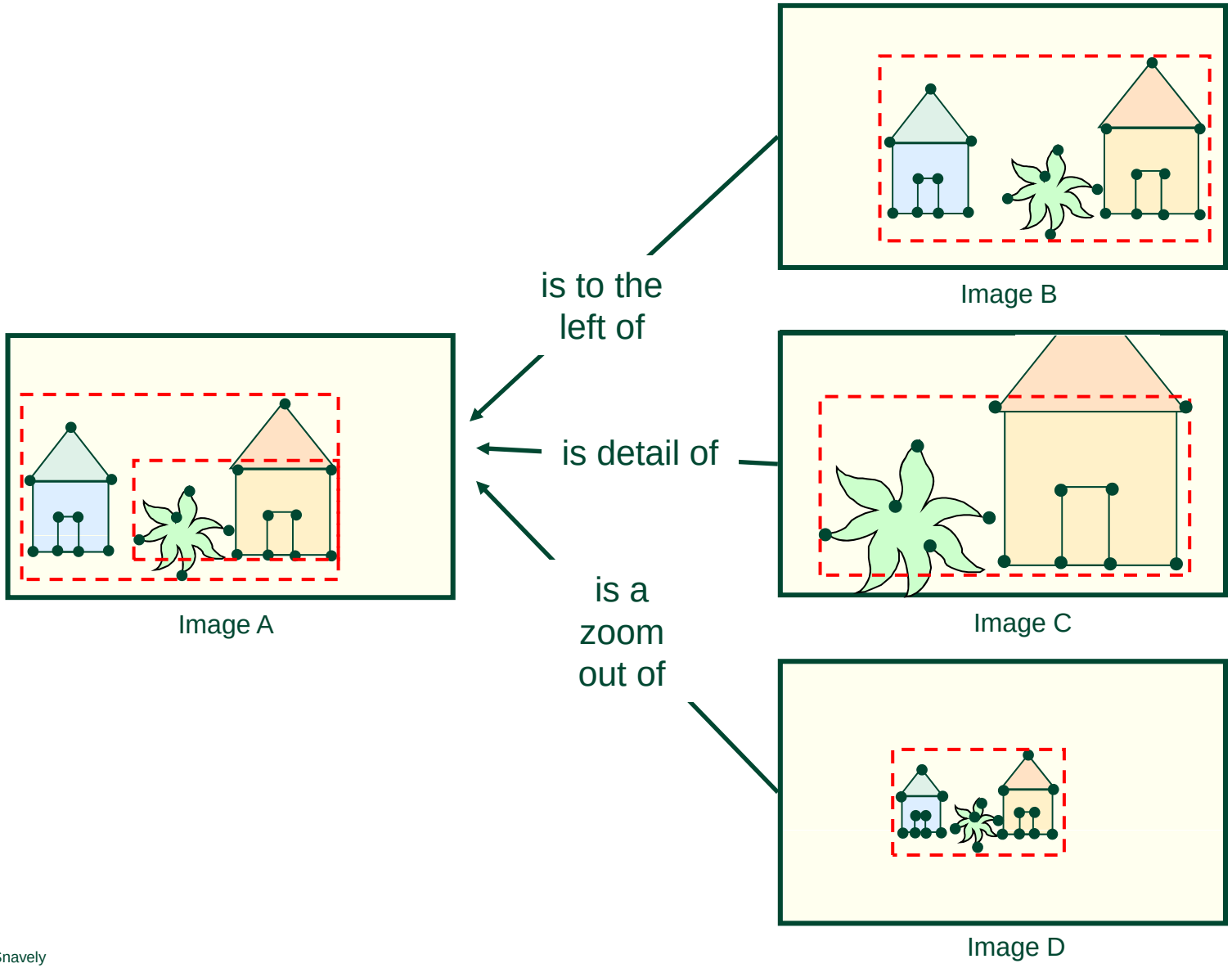
Move right

Zoom out

Relation-based browsing



Relation-based browsing



Relation-based browsing

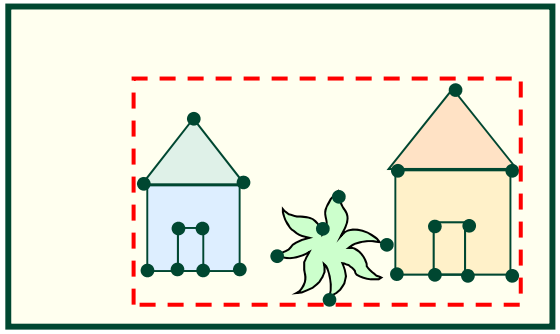


Image A

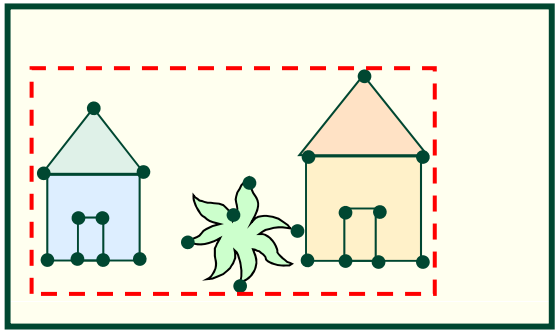


Image B

Relation-based browsing

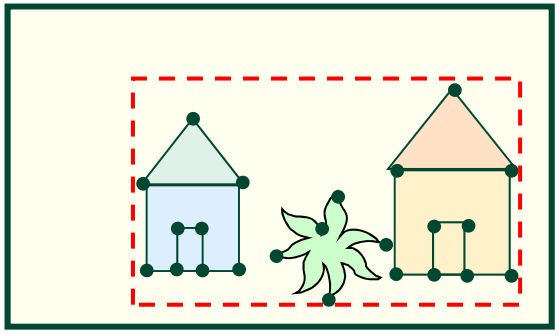


Image A

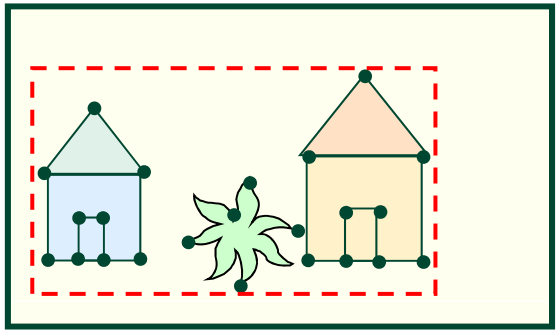


Image B

Relation-based browsing

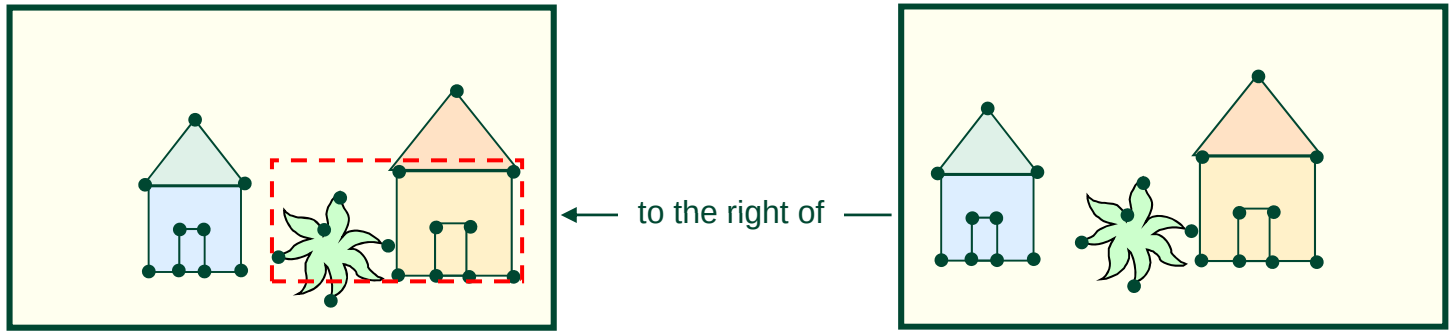


Image A

Image B

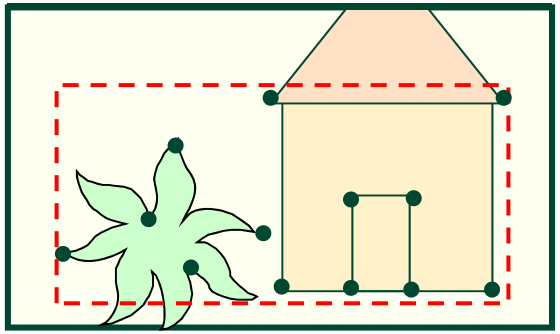


Image C

Relation-based browsing

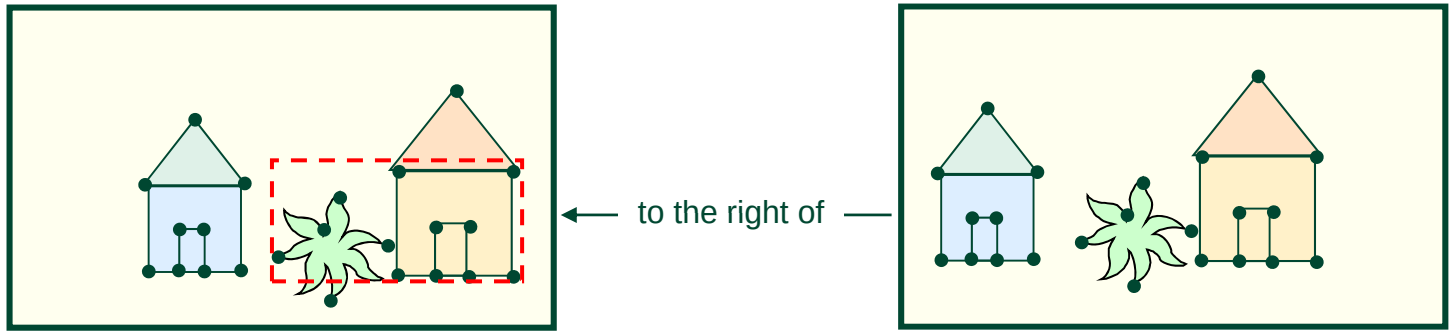


Image A

Image B

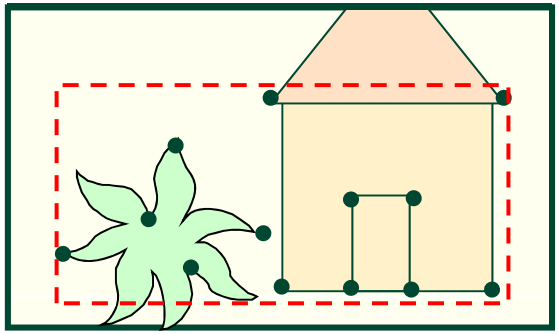
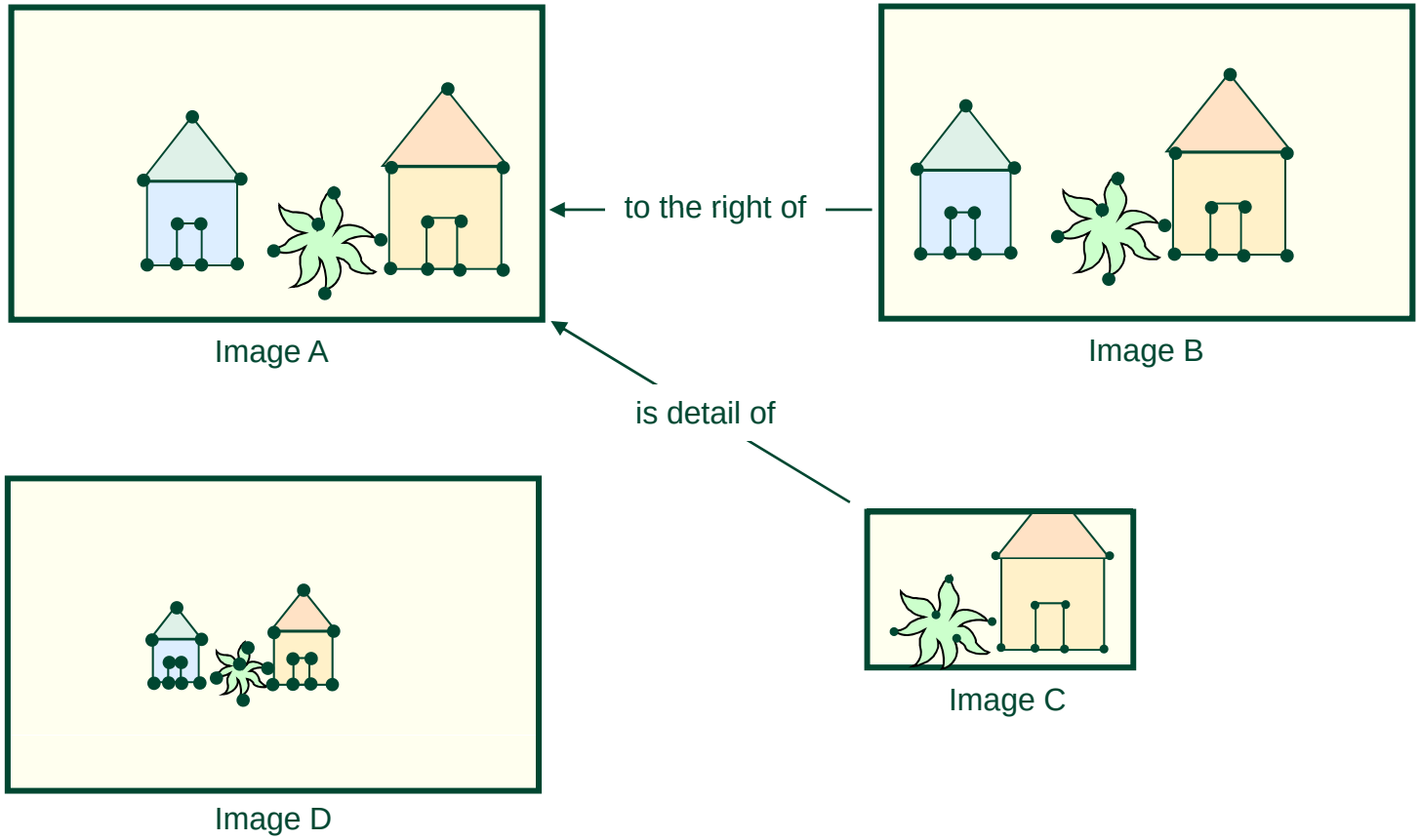
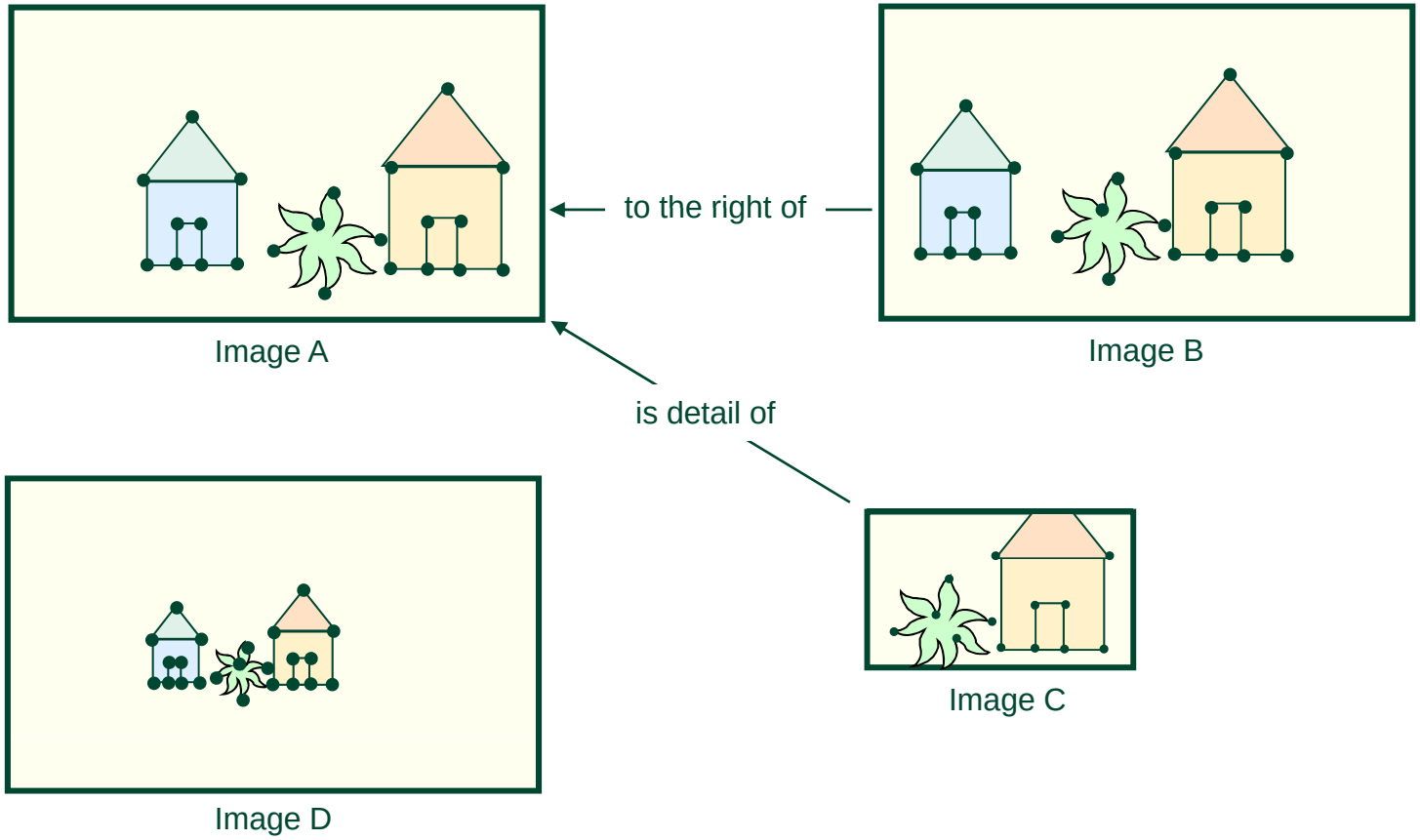


Image C

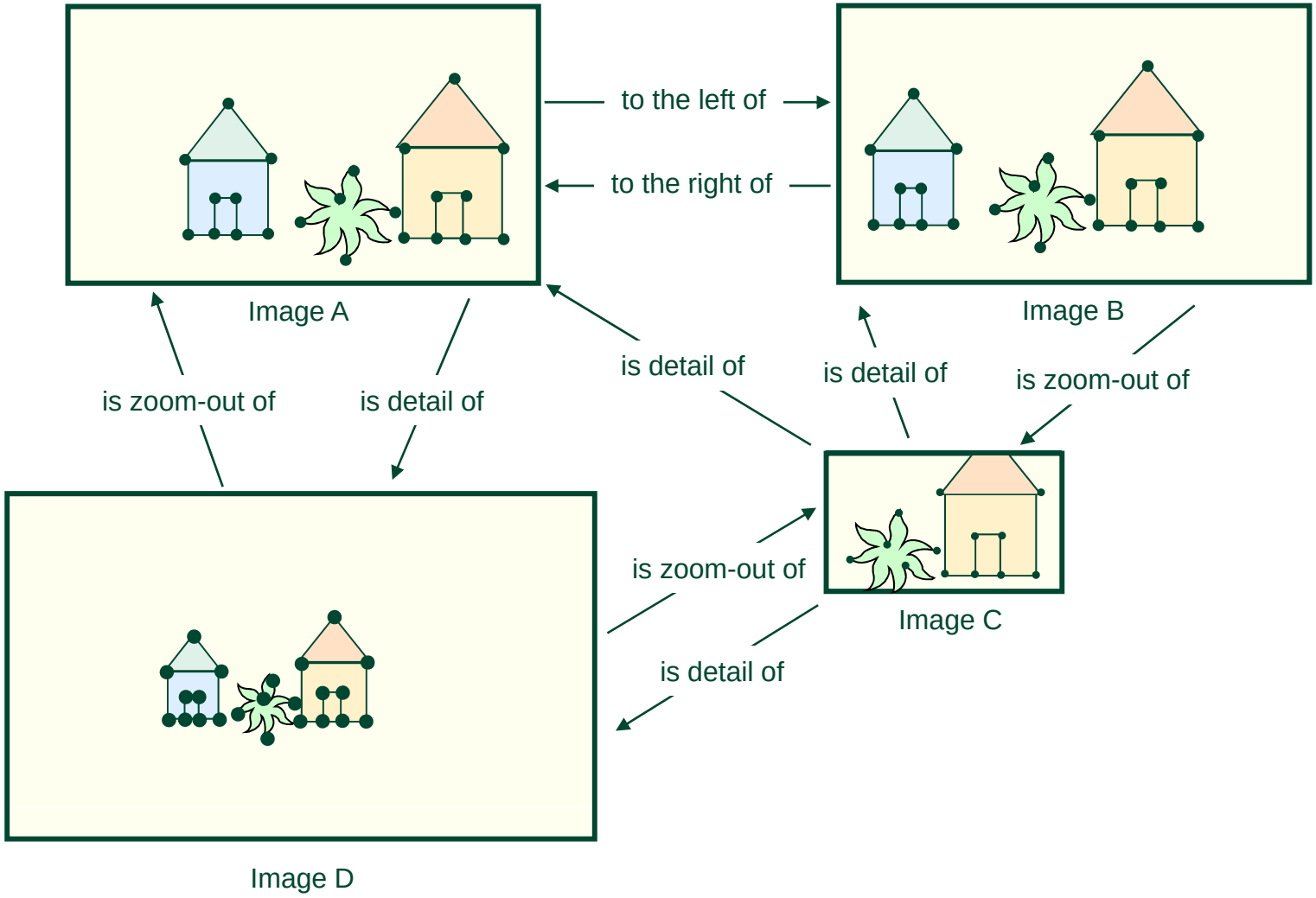
Relation-based browsing



Relation-based browsing



Relation-based browsing



Prague Old Town Square

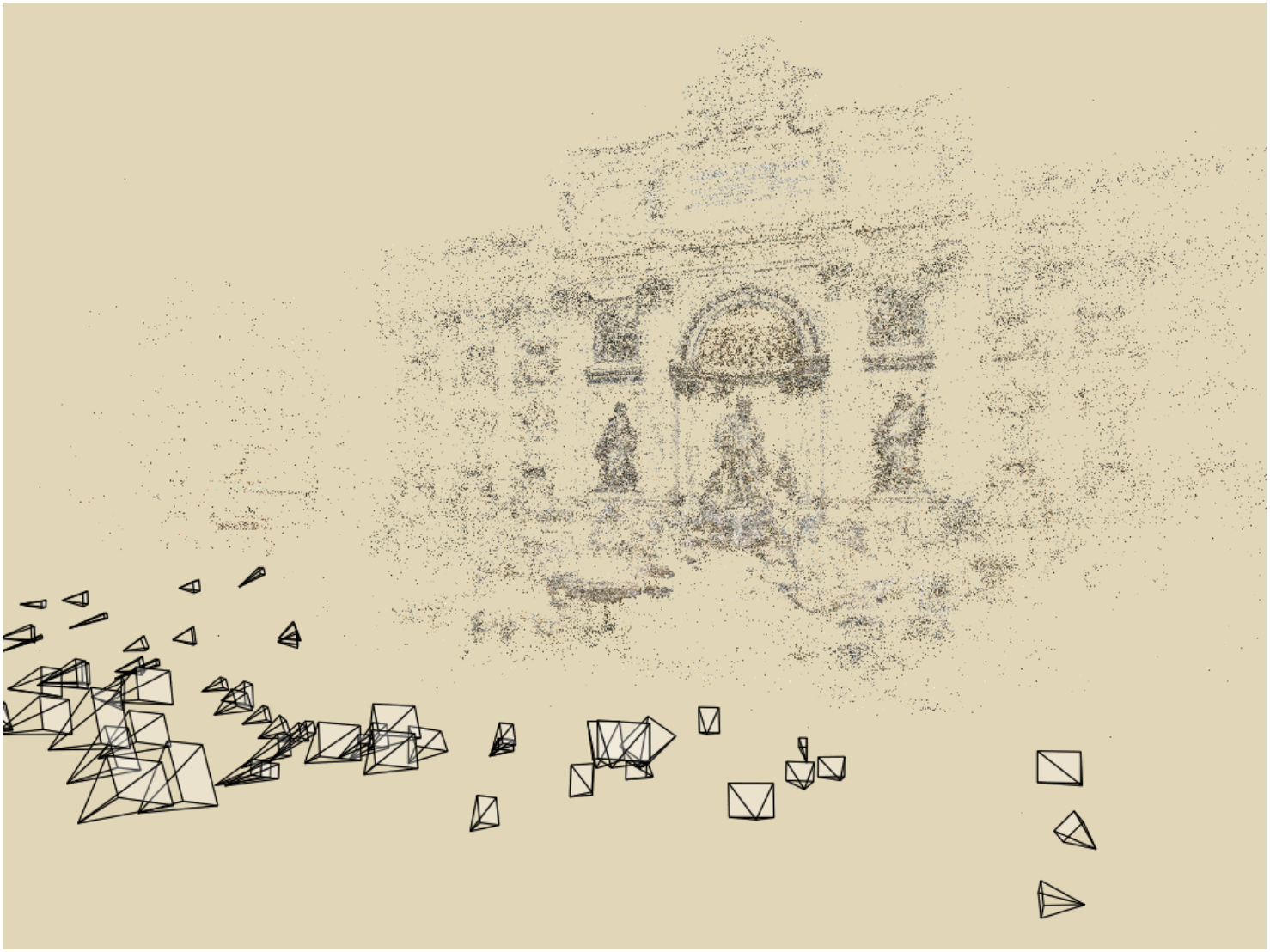
Photo Tourism overview

Input photographs

Scene reconstruction

- Navigation
- Rendering
- Annotations

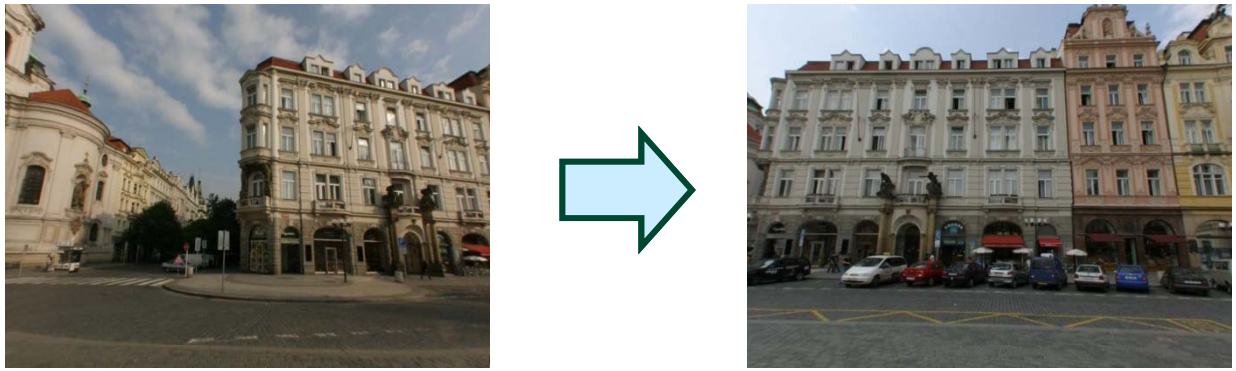
Rendering



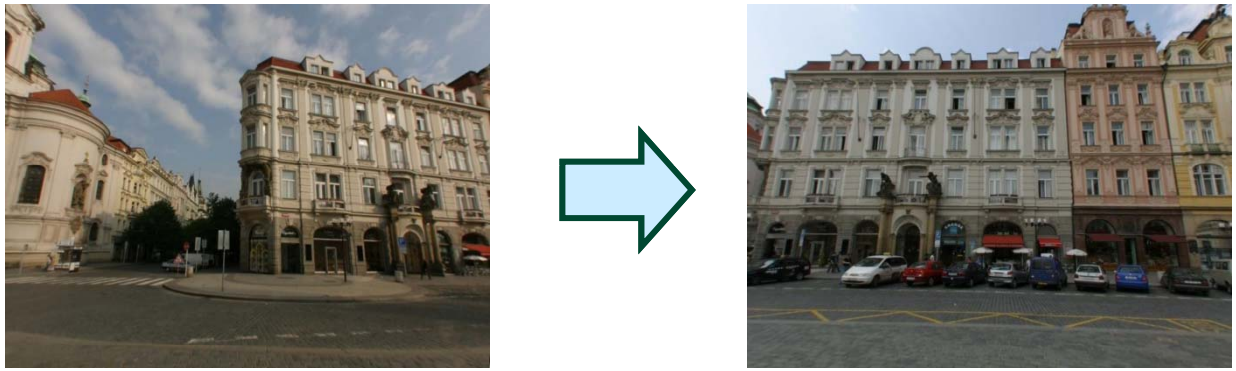
Rendering

Rendering

Rendering transitions



Rendering transitions



Rendering transitions

Rendering transitions

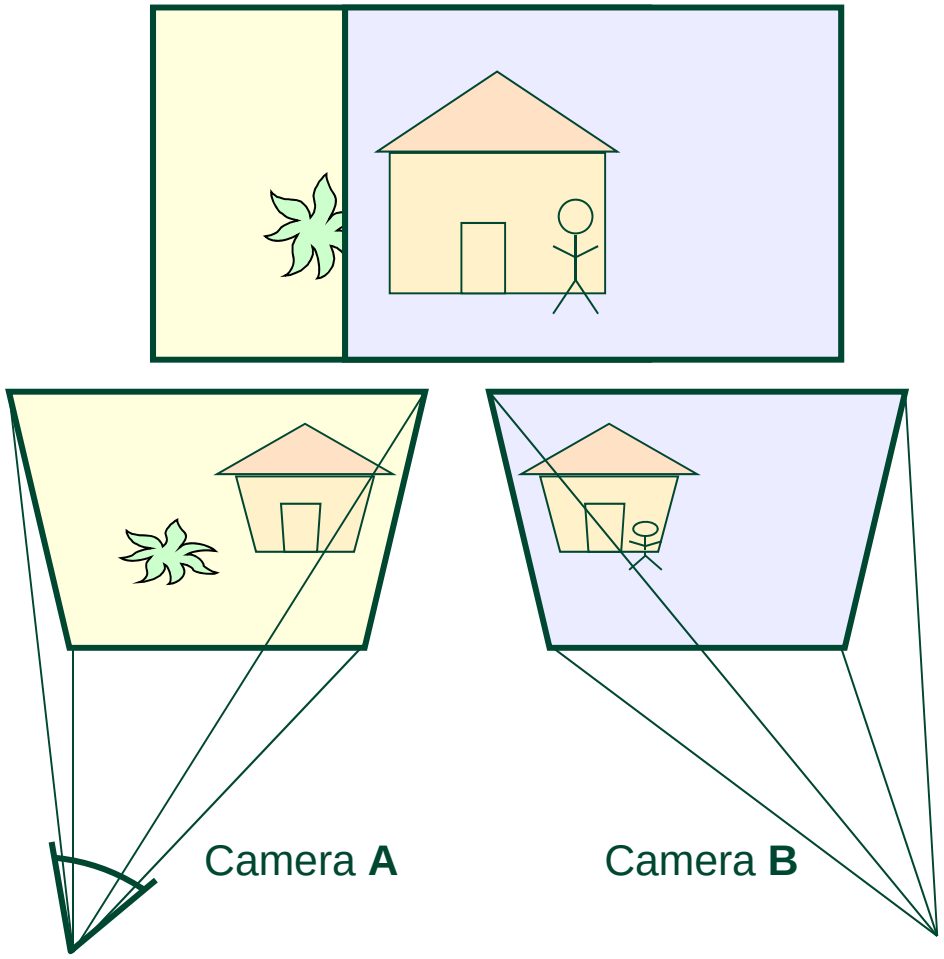


Photo Tourism overview

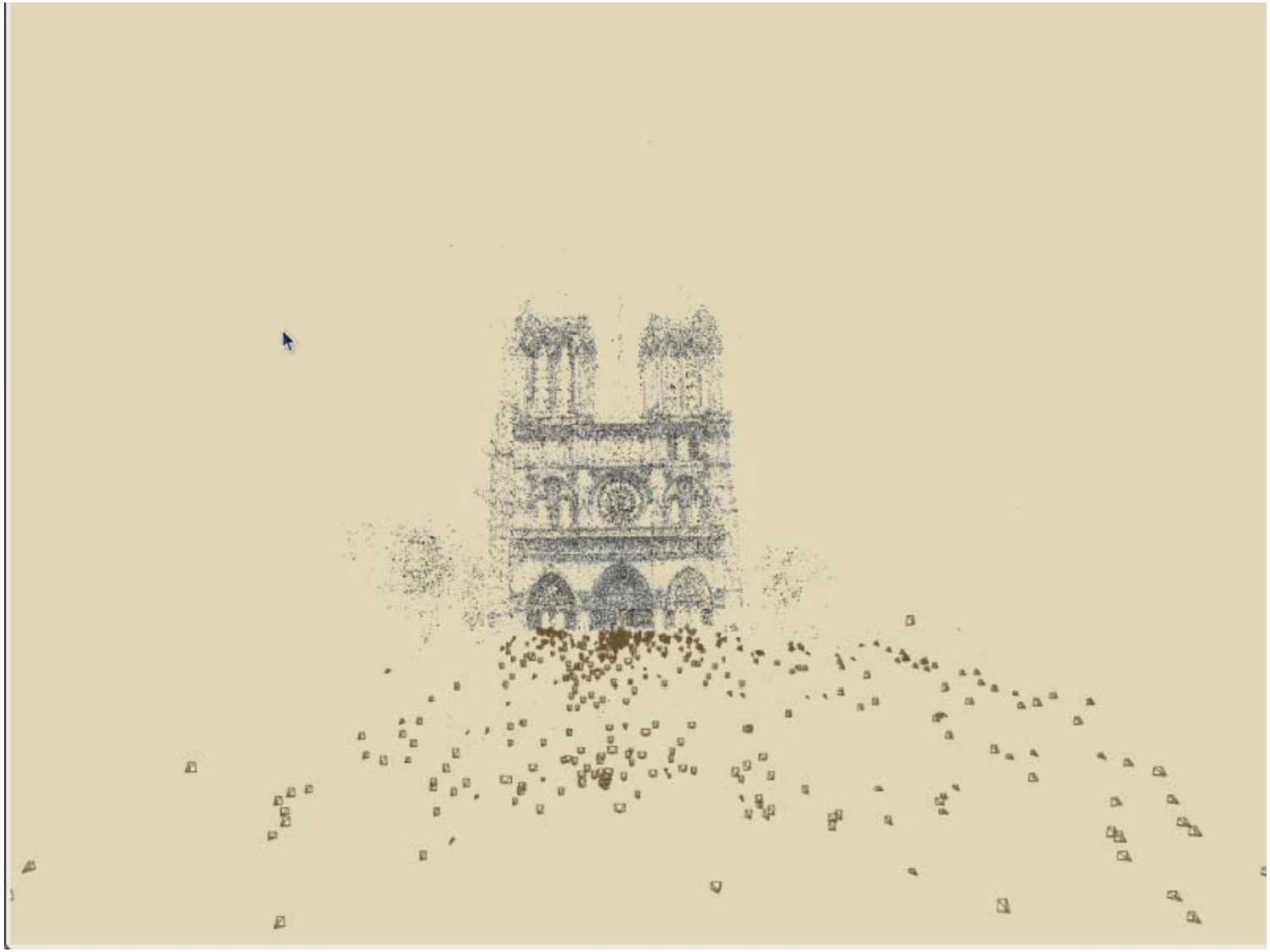
Input photographs

Scene reconstruction

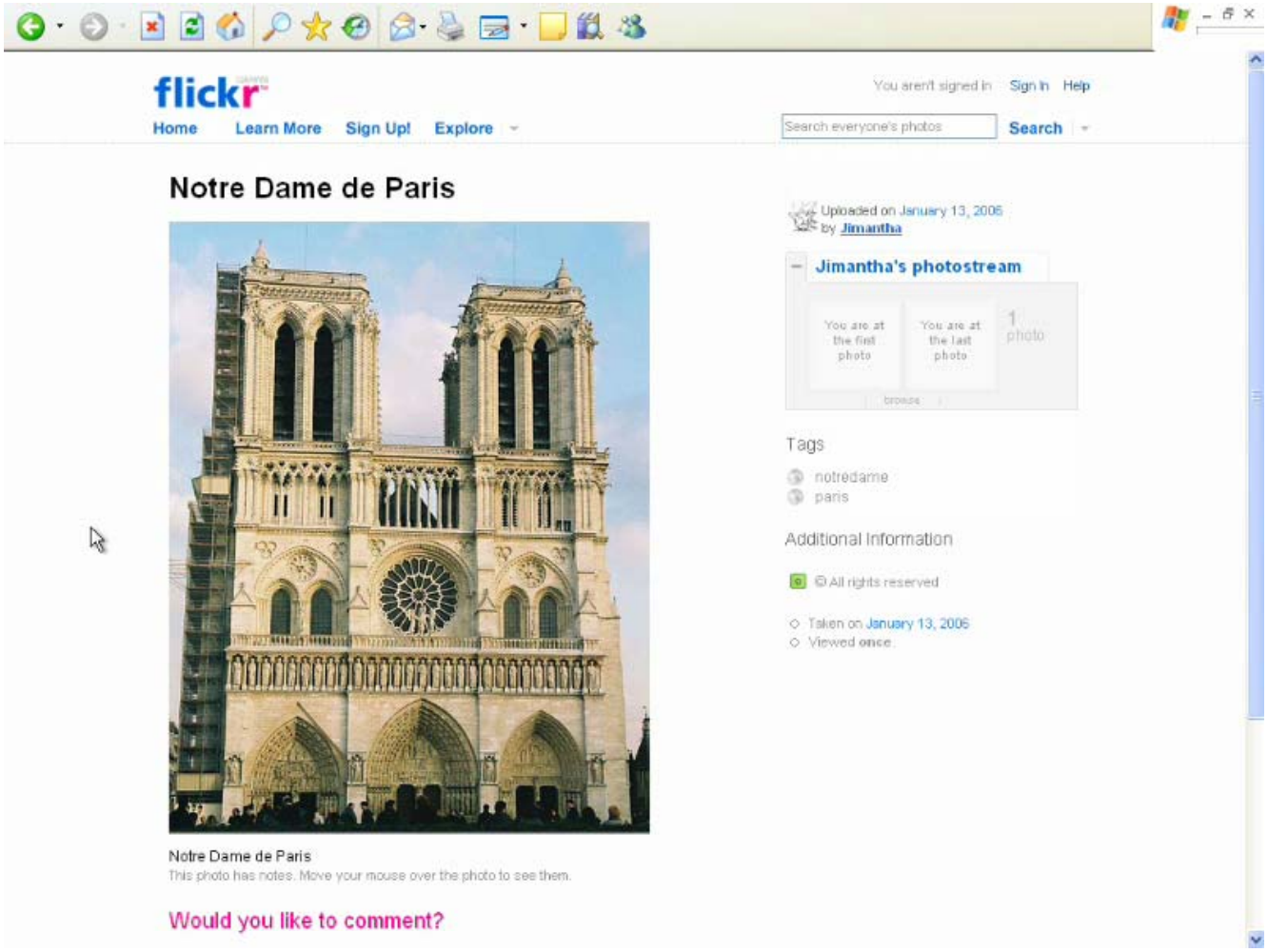
Photo Explorer

- Navigation
- Rendering
- Annotations

Annotations



Annotations



Annotations

Yosemite

Contributions

- Automated system for registering photo collections in 3D for interactive exploration
- Structure from motion algorithm demonstrated on hundreds of photos from the Internet
- Photo exploration system combining new image-based rendering and photo navigation techniques

Limitations / Future work

- Not all photos can be reliably matched

- Structure from motion scalability

  → More ns

- Plane-base

Limitations / Future work

Limitations / Future work

- Not all photos can be reliably matched
  - → Better feature detection / matching
  - → Integrating GPS & other localization info.
- Structure from motion scalability
  - → More efficient (sparse) algorithms
- Plane-based transitions lack parallax
  - → Richer transitions
- Photo explorer scalability…

Future work

- Photo explorer scalability
  - Design client-server architecture for streaming images and geometry at required resolution
  - Scale to *all* of the world's photos (and videos…)
  - Photosynth project at Microsoft Live Labs (live demo)

Acknowledgements

- National Science Foundation
- Achievement Rewards for College Scientists (ARCS)
- The many people who allowed use of their photos
- UW GRAIL Lab
- MSR Interactive Visual Media Lab
- Kevin Chiu and Andy Hou for writing the Java applet

Conclusion

Indexing everyone's photos provides a new way to share and experience our world

To find out more:

- http://phototour.cs.washington.edu
- http://research.microsoft.com/IVM/PhotoTourism
- http://labs.live.com/photosynth
- Exhibition booth #2619

Saint Basil's Cathedral

Trafalgar Square

Rockefeller Center

Mount Rushmore

Statistics

Dataset	# input	# registered
Trevi Fountain	466	360
Yosemite	325	1,893
Notre Dame	597	2,635
Prague	197	235
Great Wall	82	120
Trafalgar Square	278	1,893

Reconstruction running time

- **Great Wall**: 82 / 120 photos registered

  Running time: ~ 3 hours

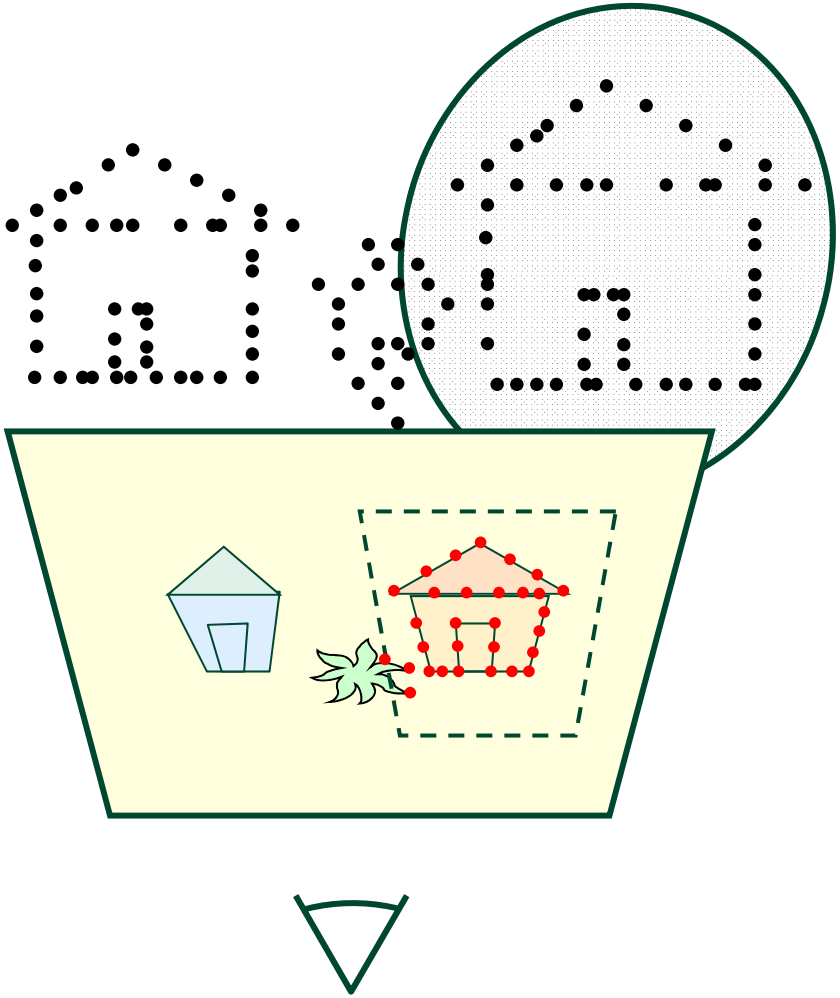
- **Notre Dame**: 597 / 2,635 photos registered

  Running time: ~ 2 weeks

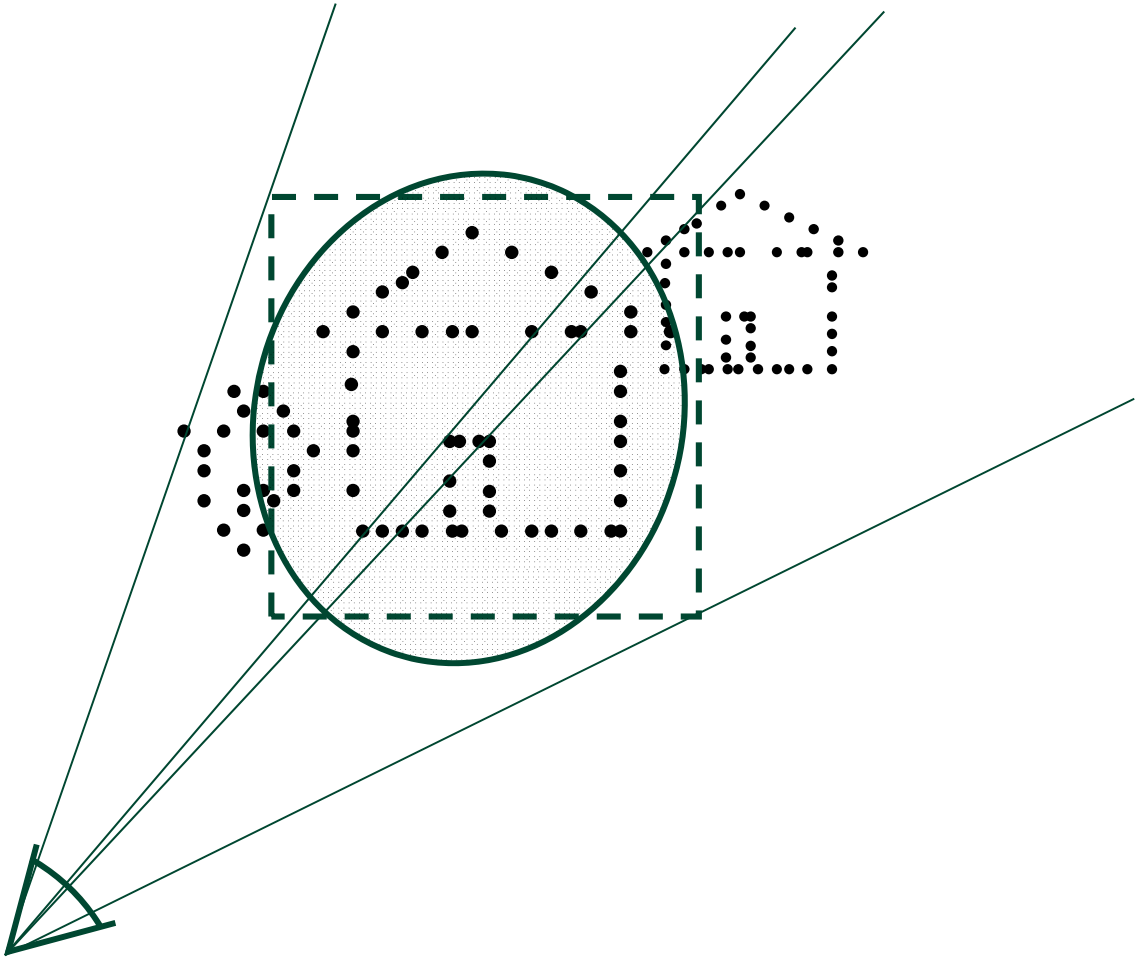
Future work

- Incorporate other metadata (e.g., time, photographer) and media (e.g., panoramas, video)
- Enhanced morphing
- Scale up structure from motion algorithm

Visibility



Visibility



Advantages of 3D over 2D

- 3D geometry has multi-image consistency
- Can annotate point cloud directly
- Can import annotations from georeferenced sources (e.g., landmark databases)
- Can use depth as cue for rejecting outliers in selection

Post-processing the reconstruction

- Compute gravity direction
- Center point cloud at the origin
- Scale model to unit variance