

Structure from Motion

(or 3D from 2D)

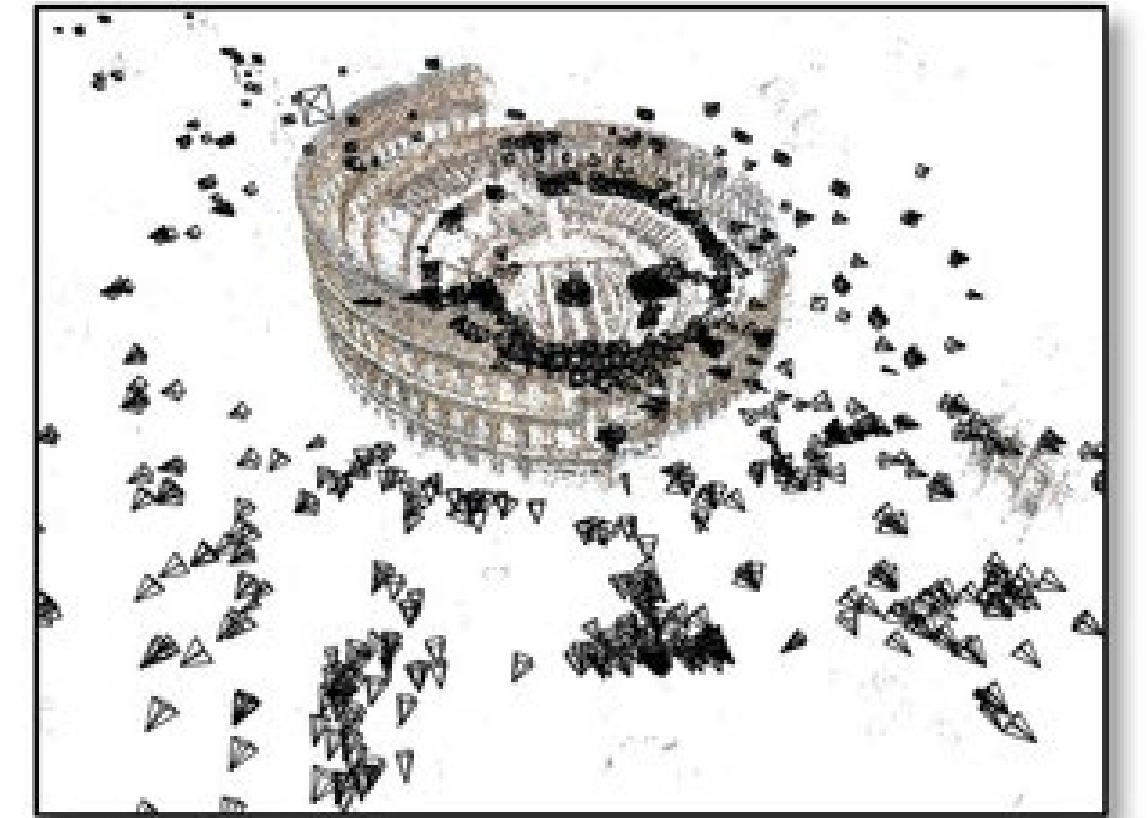


CS448V — Computational Video Manipulation

April 24th, 2019

Overview

- Structure from Motion
- Photo Tourism

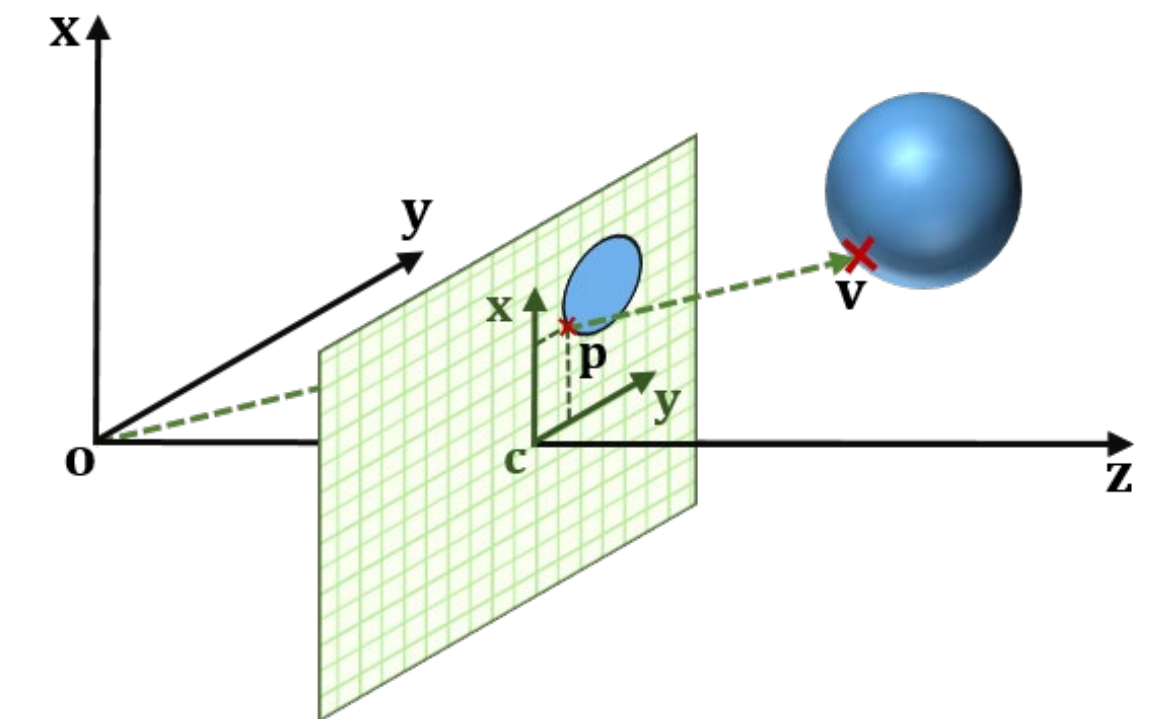


3D from 2D

- 3D? Really? Why 3D?
- Didn't I sign up for video manipulation?



- Yes, but our world is 3D



- Images are 'just' a 2D projection of our real world
- Many edits are easier to perform directly in 3D

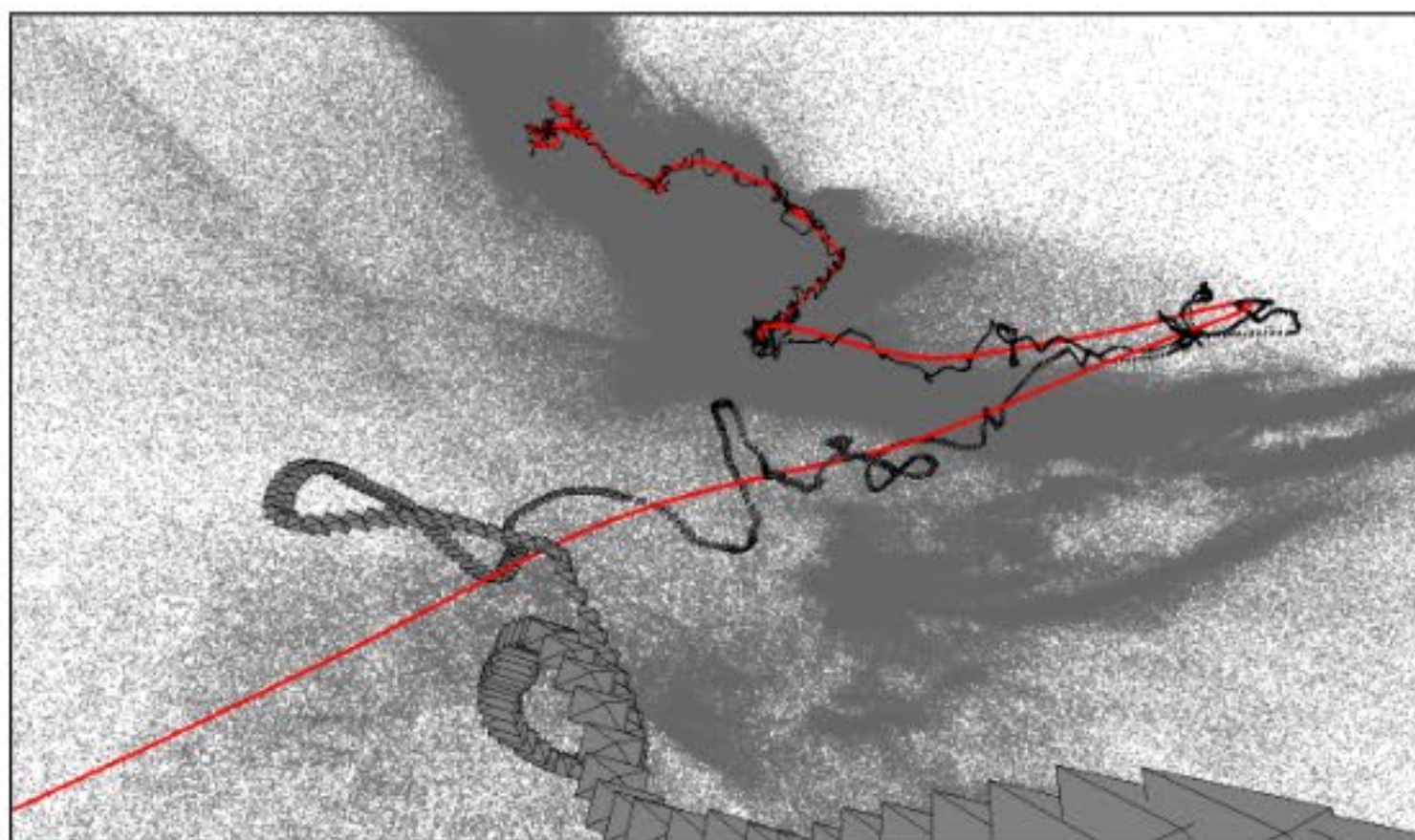
Video Stabilization in 3D

First-person Hyper-lapse Videos

Johannes Kopf
Microsoft Research

Michael F. Cohen
Microsoft Research

Richard Szeliski
Microsoft Research



(a) Scene reconstruction



(b) Proxy geometry



(c) Stitched & blended



Structure from Motion (SfM)

What does SfM recover?

1) Sparse 3D point cloud $\mathbf{v}_i \in \mathbb{R}^3$

2) Camera Extrinsic Parameters

- $\mathbf{t}_j \in \mathbb{R}^3$, position
- $\mathbf{R}_j \in \mathbf{SO}(3)$, orientation

3) Camera Intrinsic Parameters

- focal length f , skew s , optical center \mathbf{c} , lens distortion parameters, ...



Quiz: What is?

- **Structure from Motion**

- “3D pointcloud + cameras”

- **Bundle Adjustment**

- “a way to perform SfM”

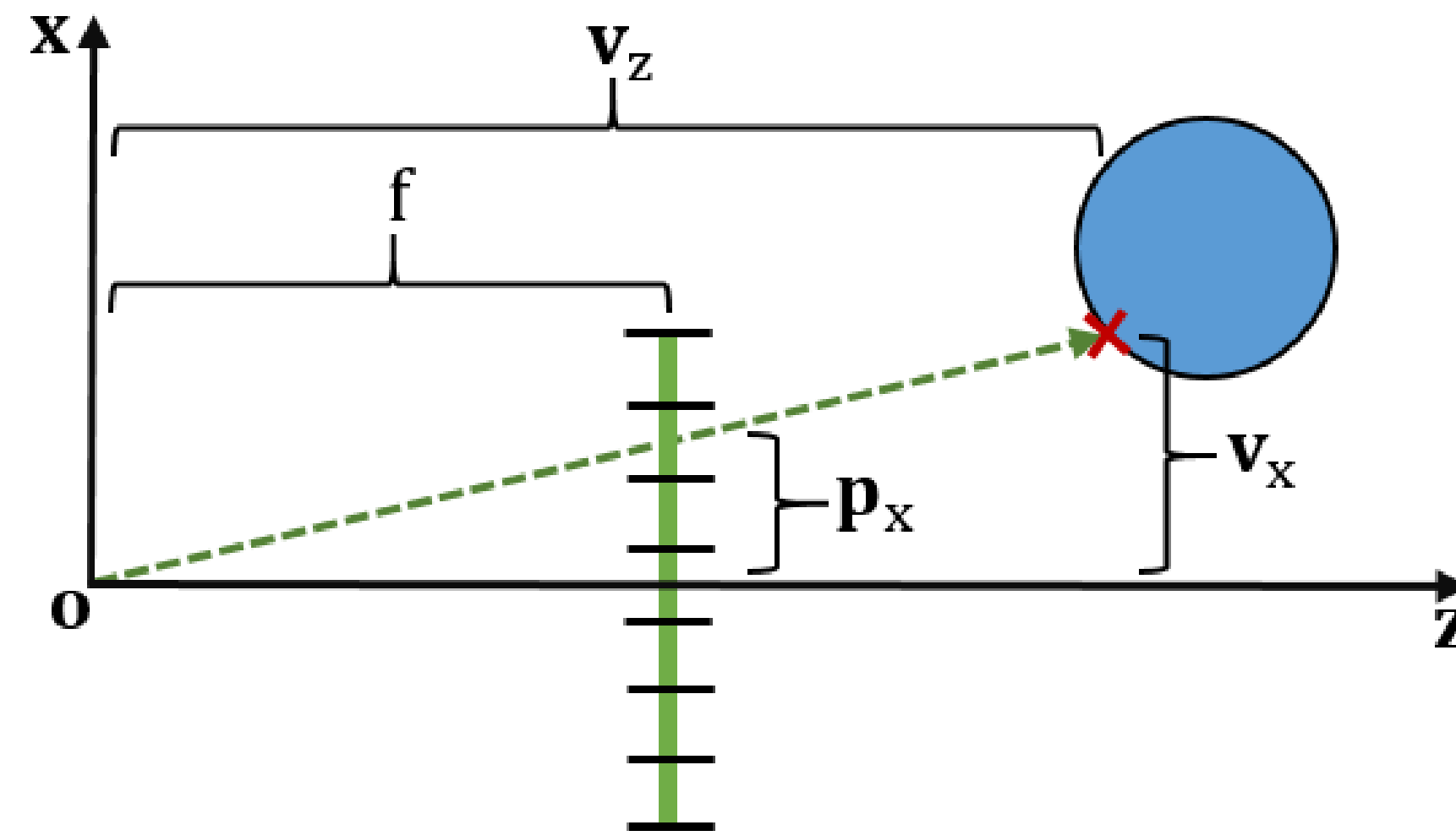
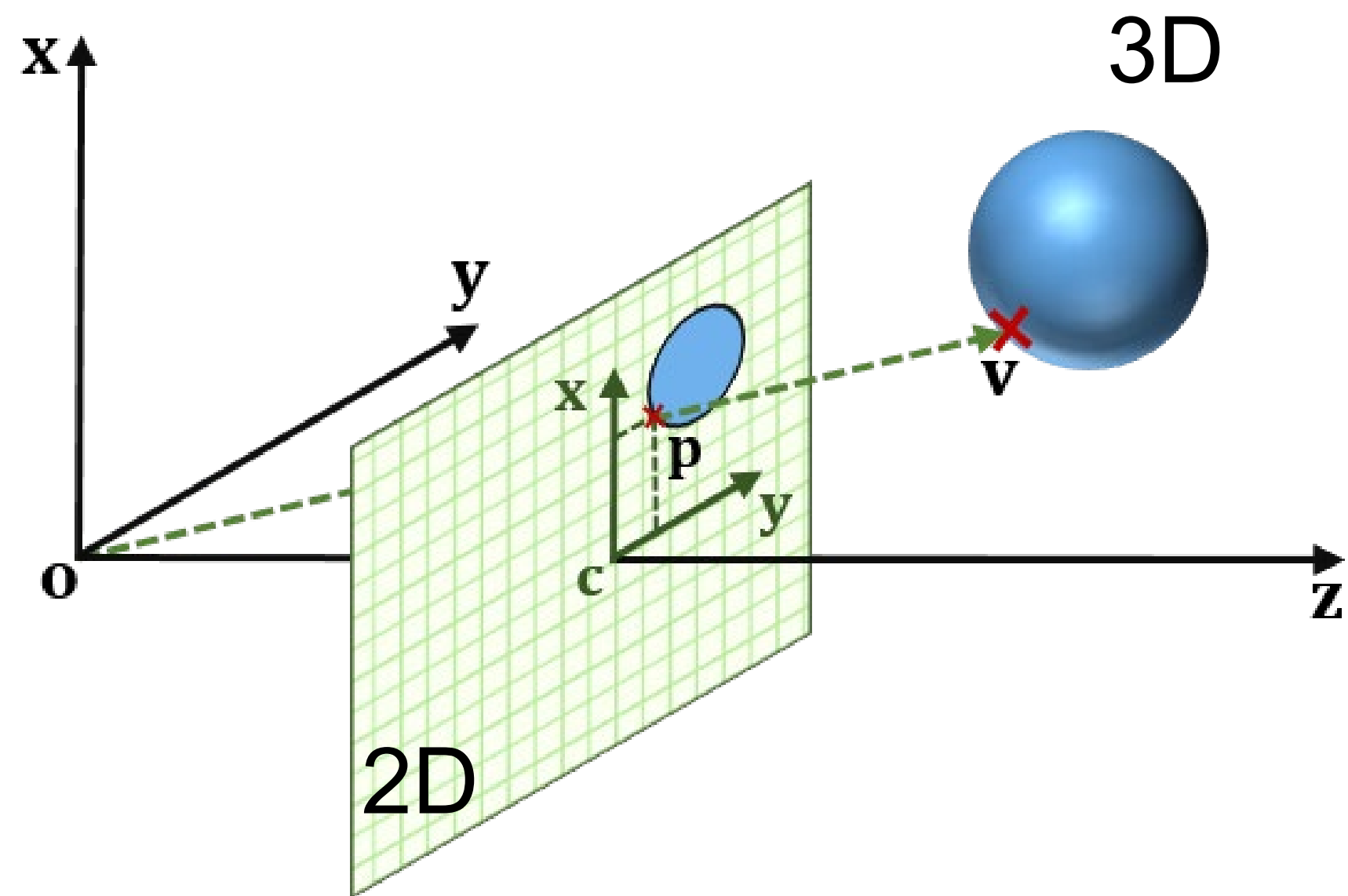
- **Multi-view Stereo**

- “dense reconstruction step after SfM”

- **Photogrammetry**

- “dense reconstruction step after SfM”

Intrinsics of a Perspective Camera



“rule of equal triangles”

$$1) \frac{p_x}{f} = \frac{v_x}{v_z}$$

“math”

$$2) \begin{aligned} p_x &= \frac{f \cdot v_x}{v_z} \\ p_y &= \frac{f \cdot v_y}{v_z} \end{aligned}$$

“vectorize/homogeneous coordinates”

$$3) \quad \mathbf{p} = \mathbf{dehom}(\hat{\mathbf{p}})$$

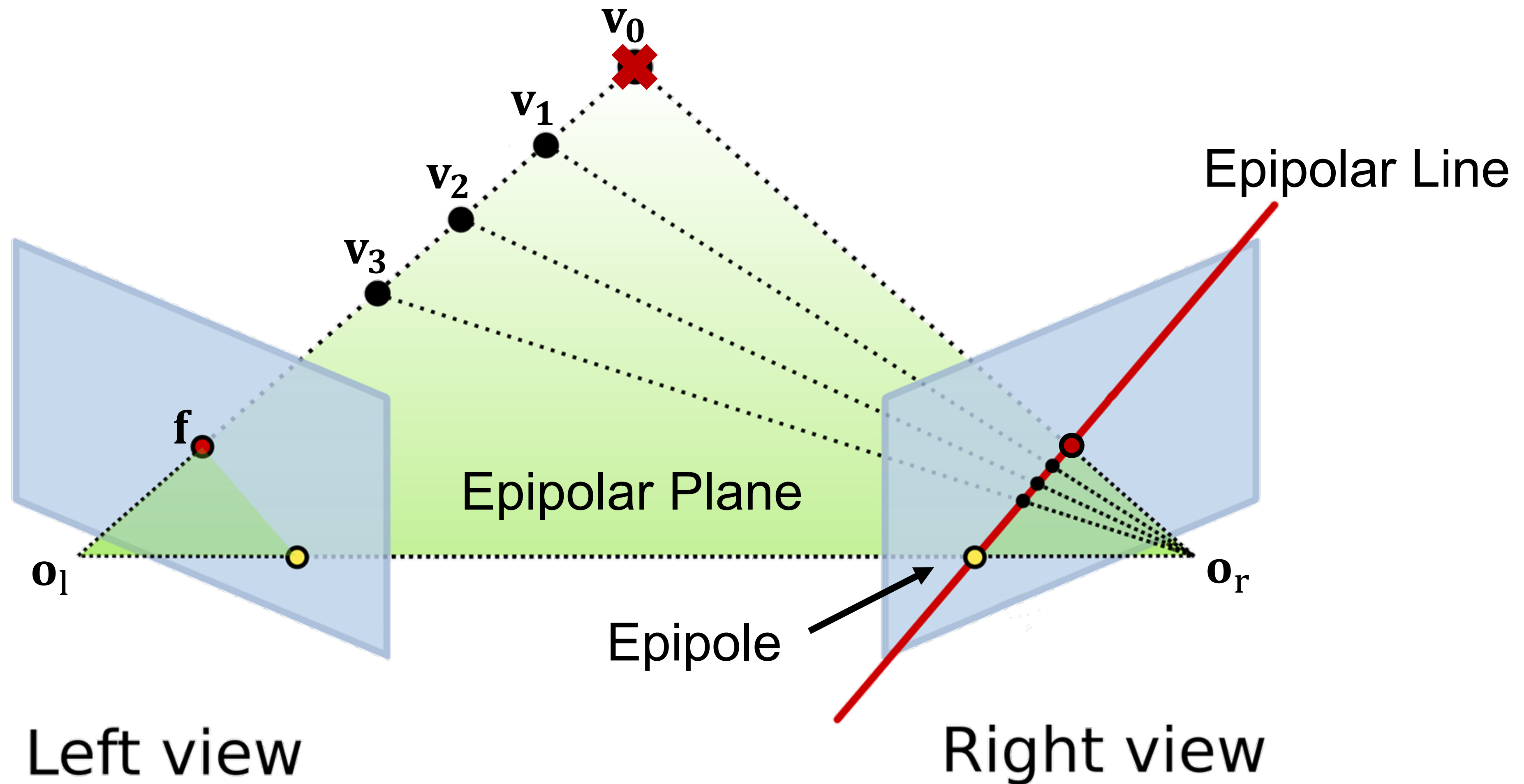
$$\begin{bmatrix} \hat{p}_x \\ \hat{p}_y \\ \hat{p}_z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$$

$$\hat{\mathbf{p}} = \mathbf{Kv}$$

“more general”

$$\mathbf{K} = \begin{bmatrix} f & s & \mathbf{c}_x \\ 0 & f & \mathbf{c}_y \\ 0 & 0 & 1 \end{bmatrix}$$

Epipolar Geometry & Triangulation



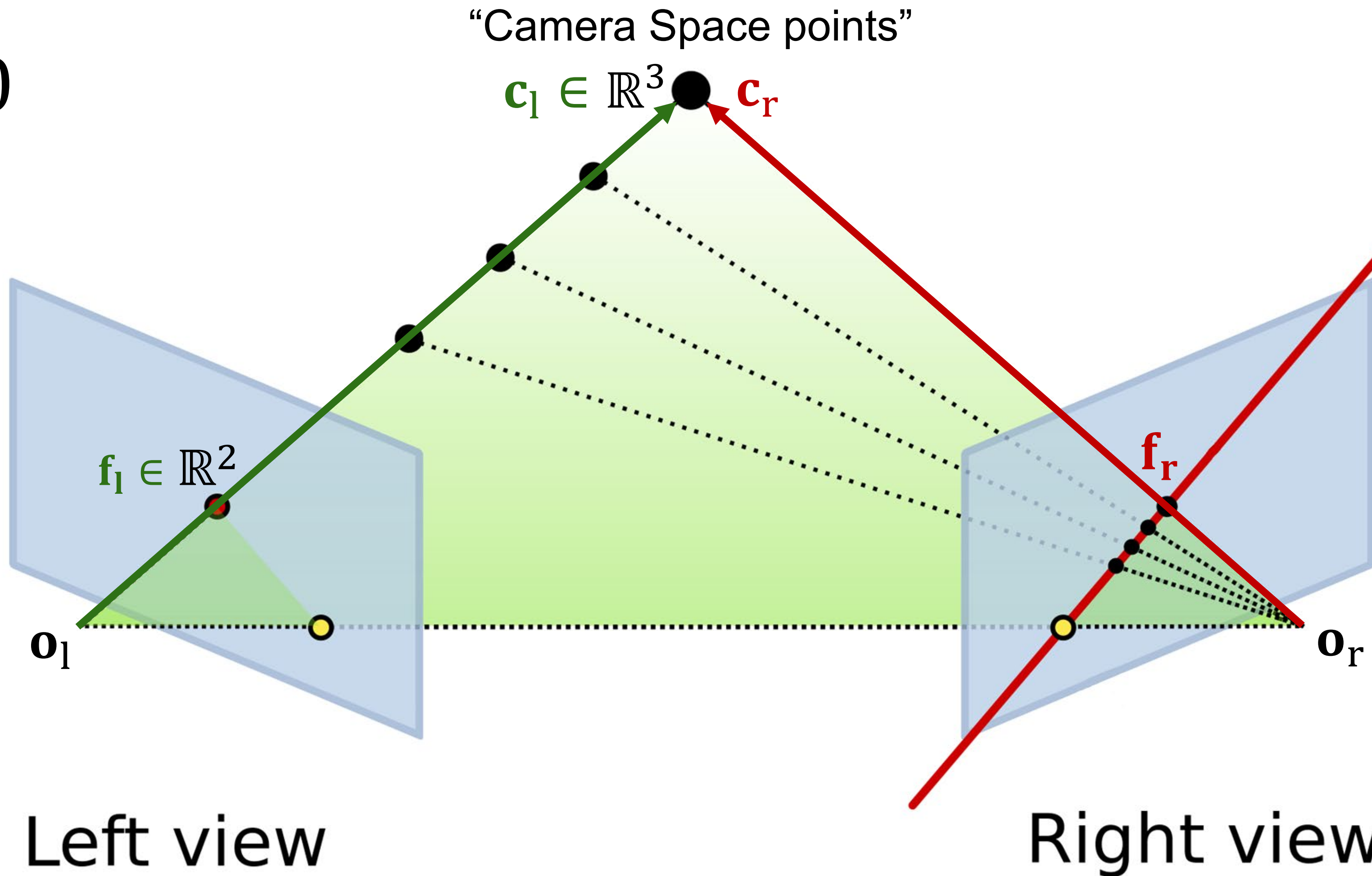
Essential/Fundamental Matrix

$$\mathbf{c}_r^T \mathbf{E} \mathbf{c}_l = 0$$

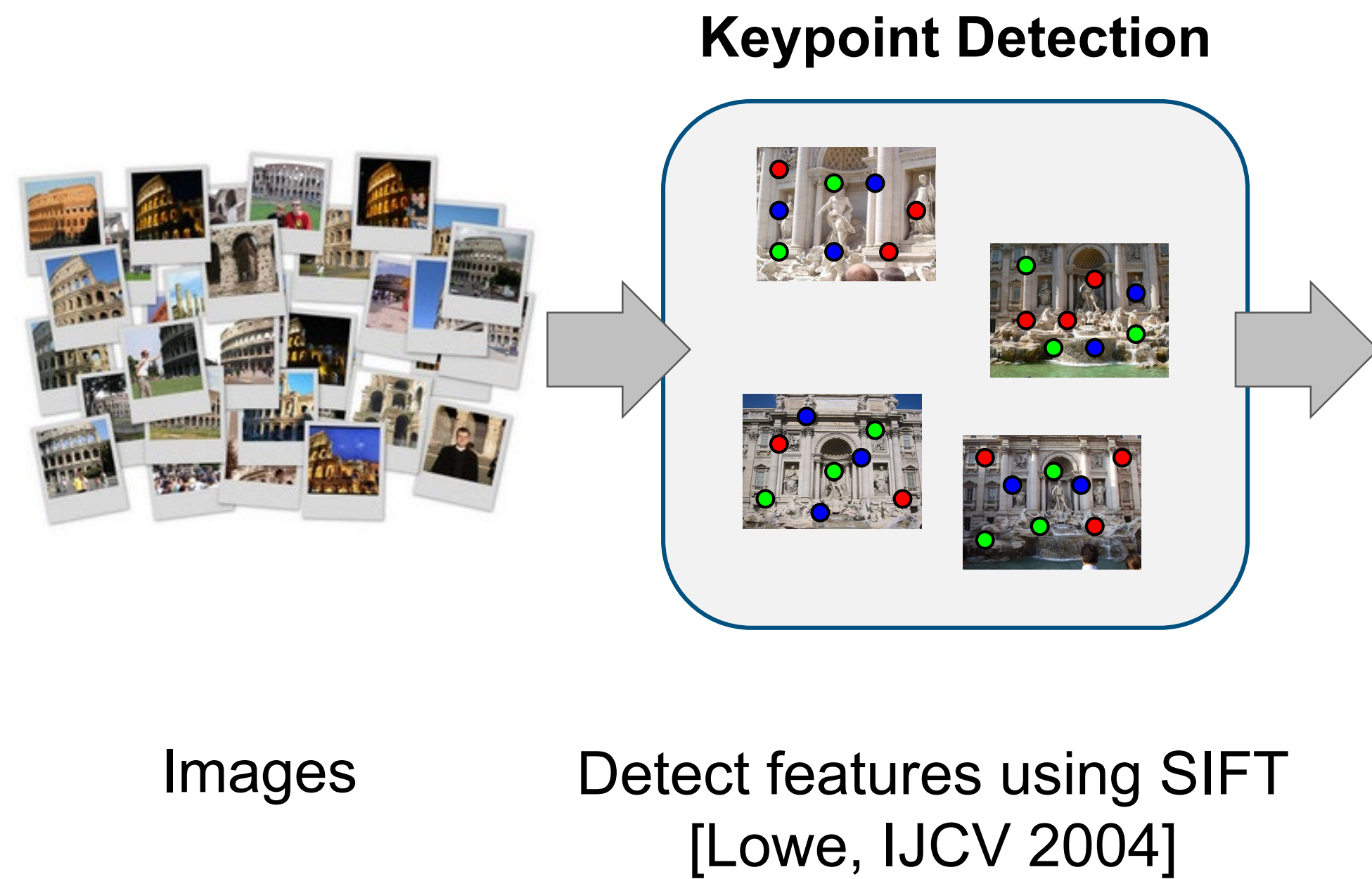
↑
Essential Matrix
"3D to 3D"

$$\bar{\mathbf{f}}_l^T \mathbf{F} \bar{\mathbf{f}}_r = 0$$

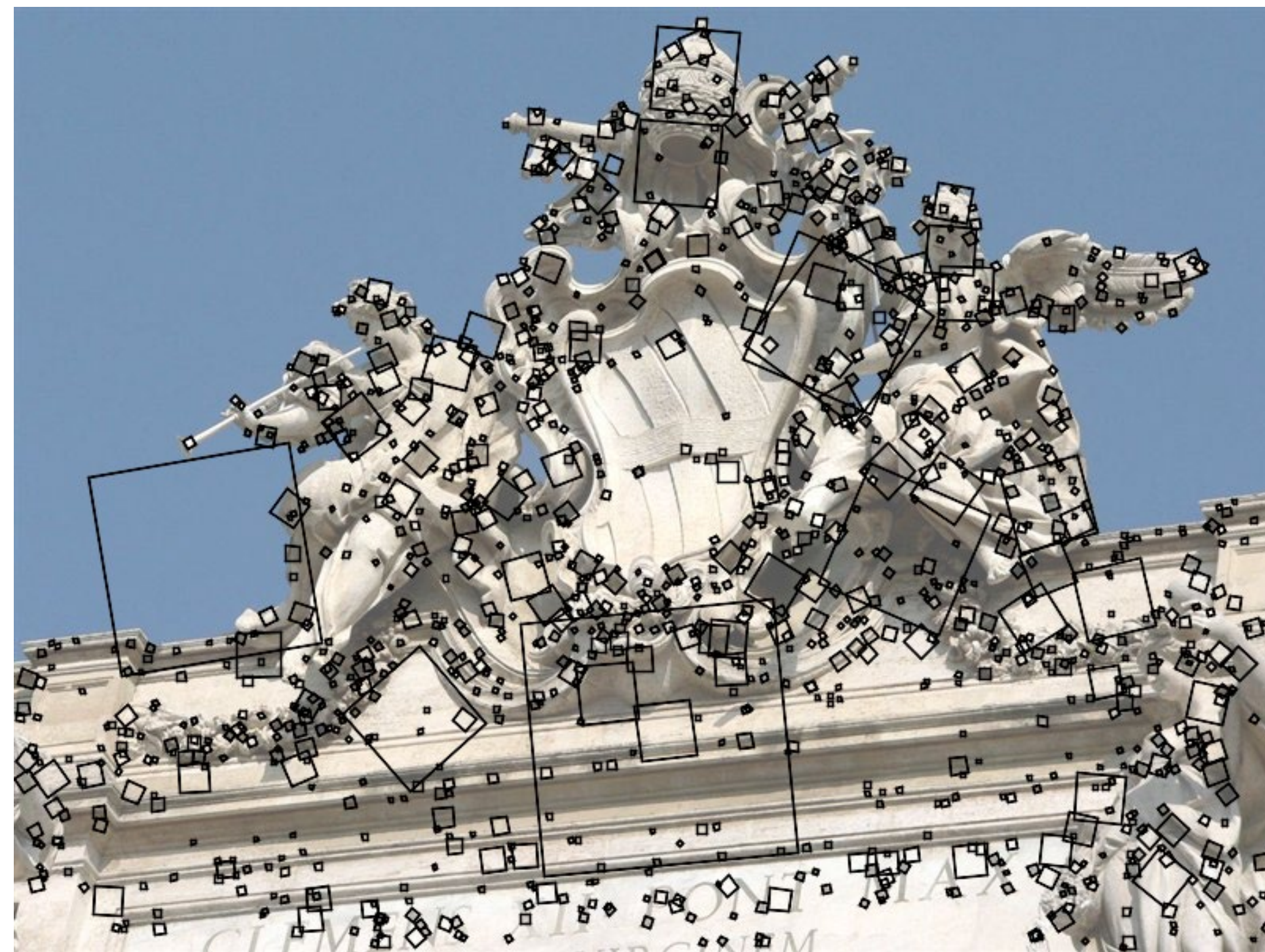
↑
Fundamental Matrix
"2D to 2D"
"homogeneous"



Pipeline: Structure from Motion

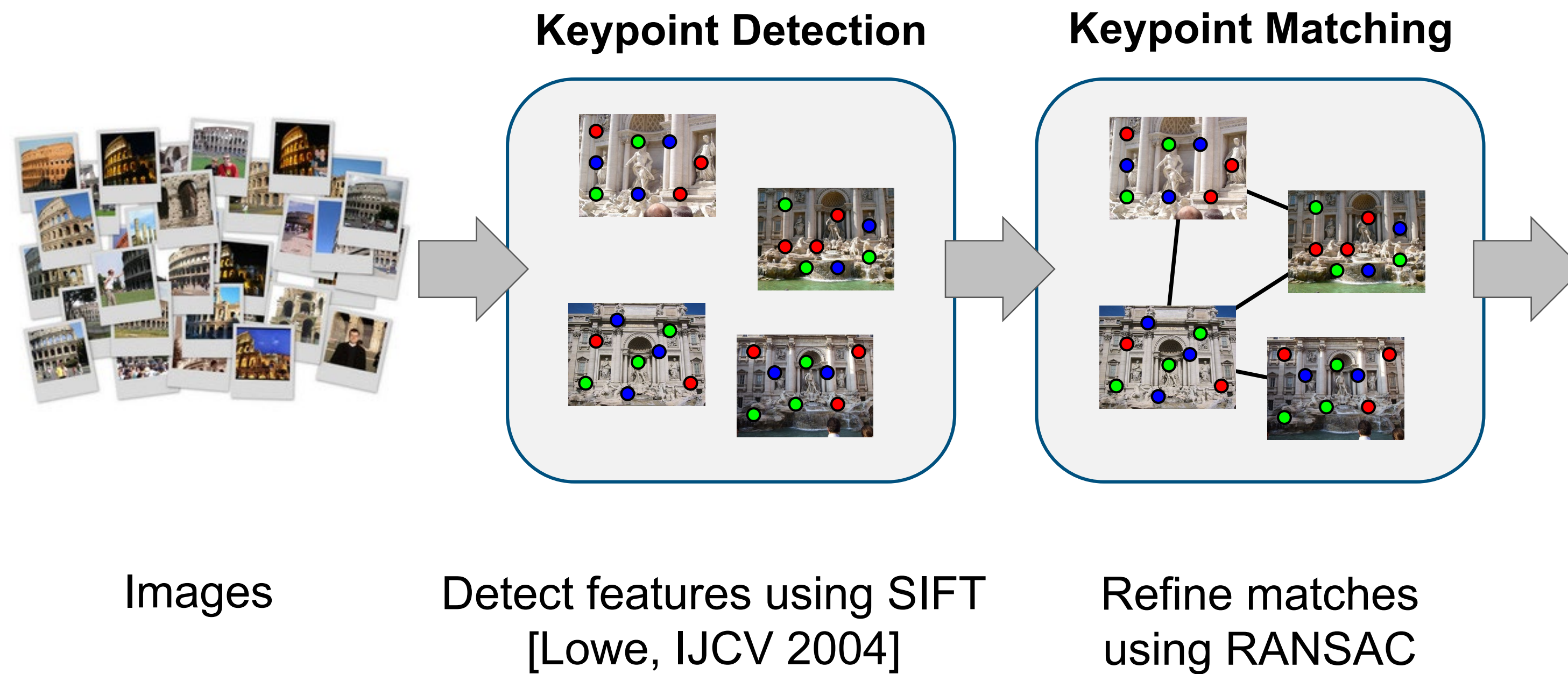


Feature detection

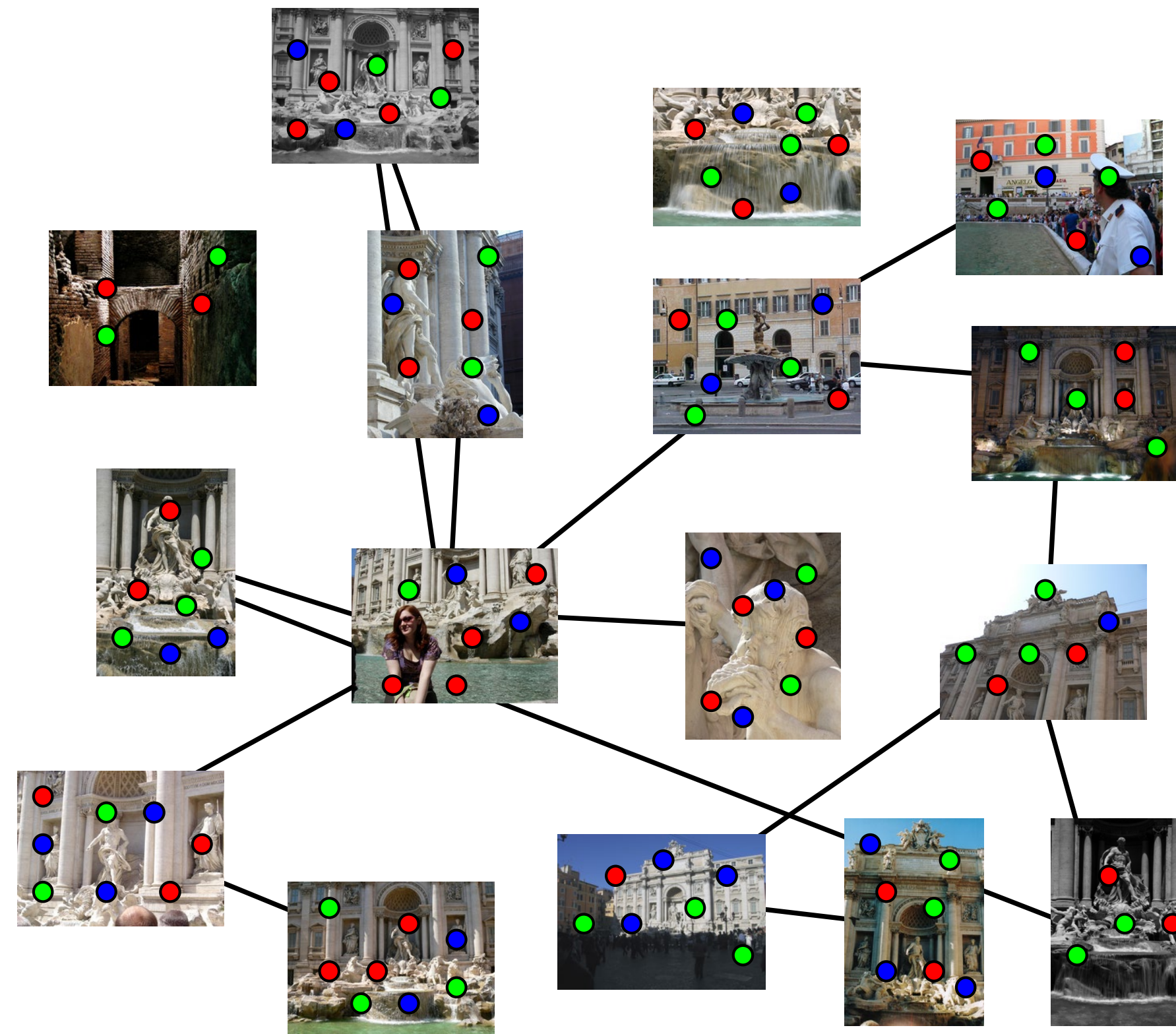


Detect features using SIFT [Lowe, IJCV 2004]

Pipeline: Structure from Motion



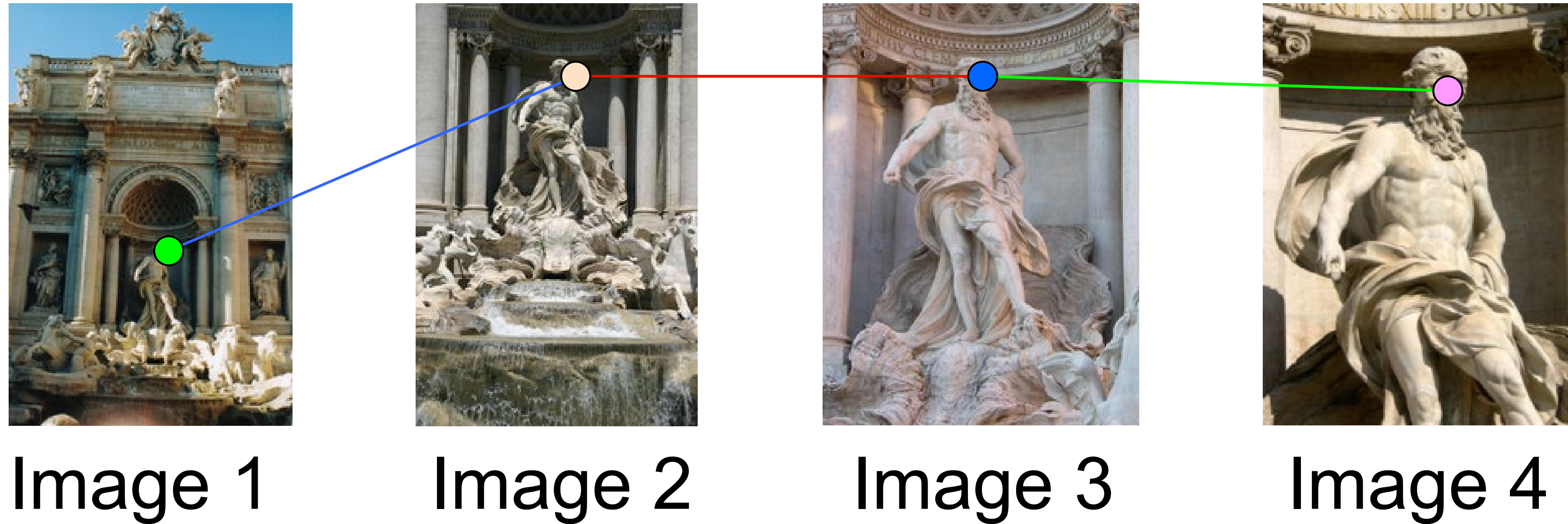
Match Features between Images



Refine matches using RANSAC [Fischler & Bolles 1987]

to estimate fundamental matrices between pairs

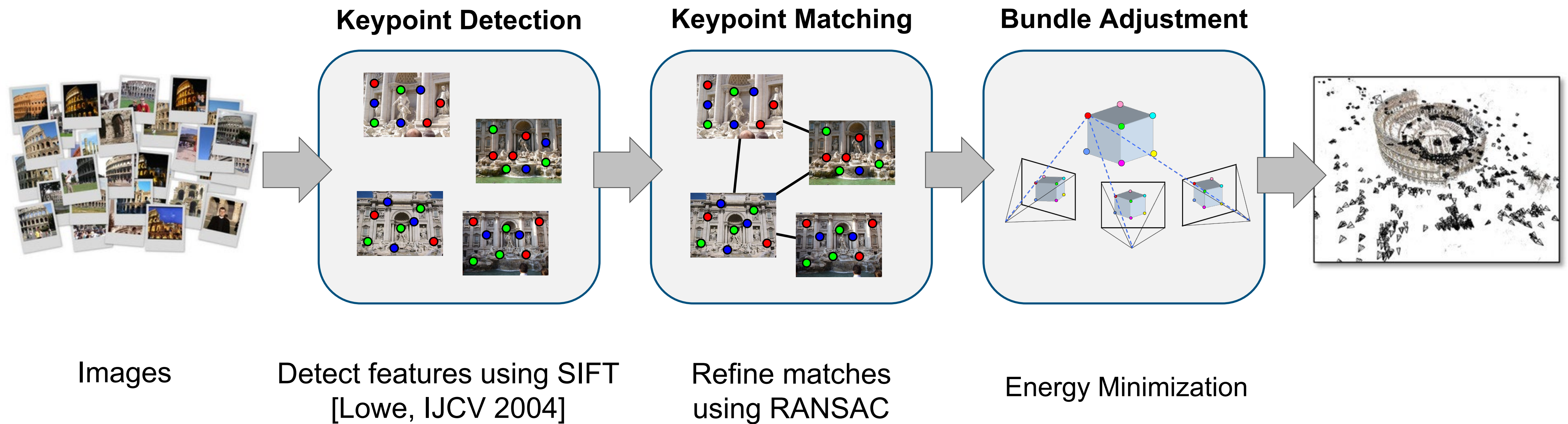
Correspondence estimation



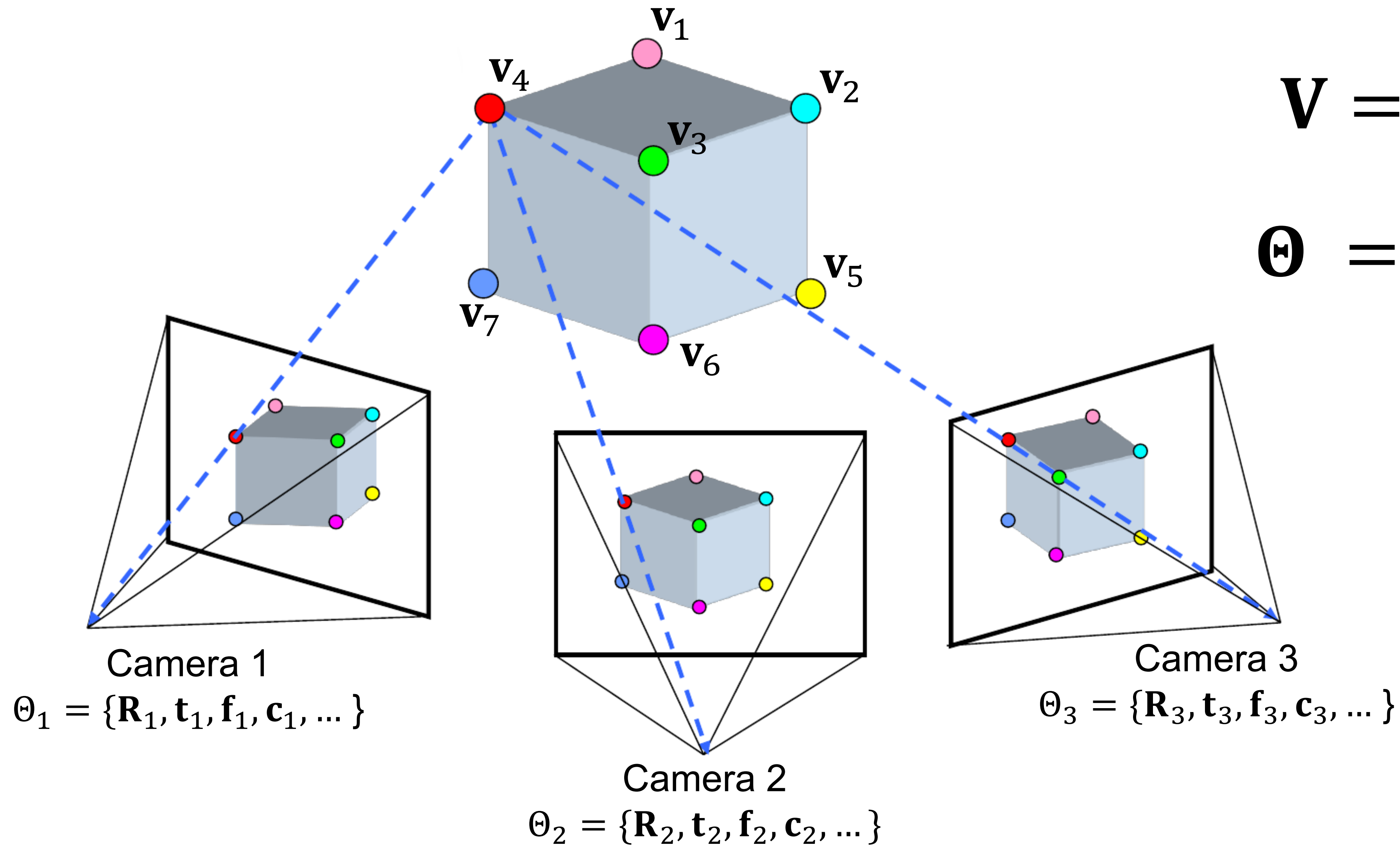
Link up pairwise matches to form connected components

➤ Single 3D Point

Pipeline: Structure from Motion



Multi-view Geometry



$$\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_7\}$$

$$\Theta = \{\Theta_1, \dots, \Theta_3\}$$

Bundle Adjustment

- Energy Minimization based Approach for SfM

$$\Theta^*, V^* = \underset{\Theta, V}{\operatorname{argmin}} E(\Theta, V)$$

$$V = \{v_1, \dots, v_M\}$$

M points

$$\Theta = \{\Theta_1, \dots, \Theta_N\}$$

N cameras

Re-projection Error

$$E(\Theta, V) = \sum_{i=1}^N \sum_{j=1}^M w_{i,j} \|\mathbf{p}_{i,j} - \Pi_i(\mathbf{v}_j)\|_2^2$$

Square?

$\Pi_i(\mathbf{v}_j)$ projection of point j in camera i

$$w_{i,j} = \begin{cases} 1, & \text{camera } i \text{ observes point } j \\ 0, & \text{otherwise} \end{cases}$$

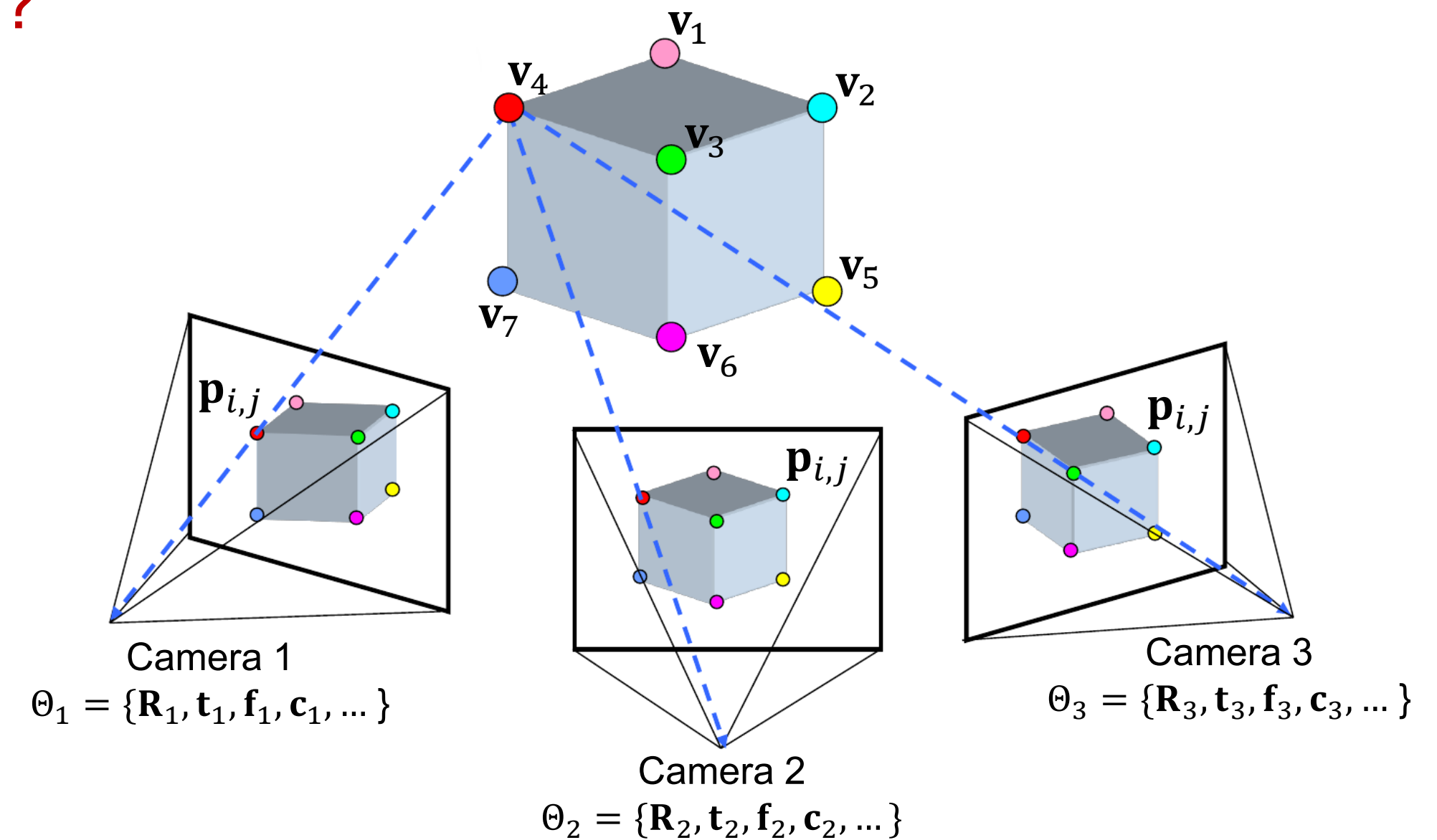


Photo Tourism: Exploring Photo Collections in 3D

Noah Snavely
University of Washington

Steven M. Seitz
University of Washington

Richard Szeliski
Microsoft Research

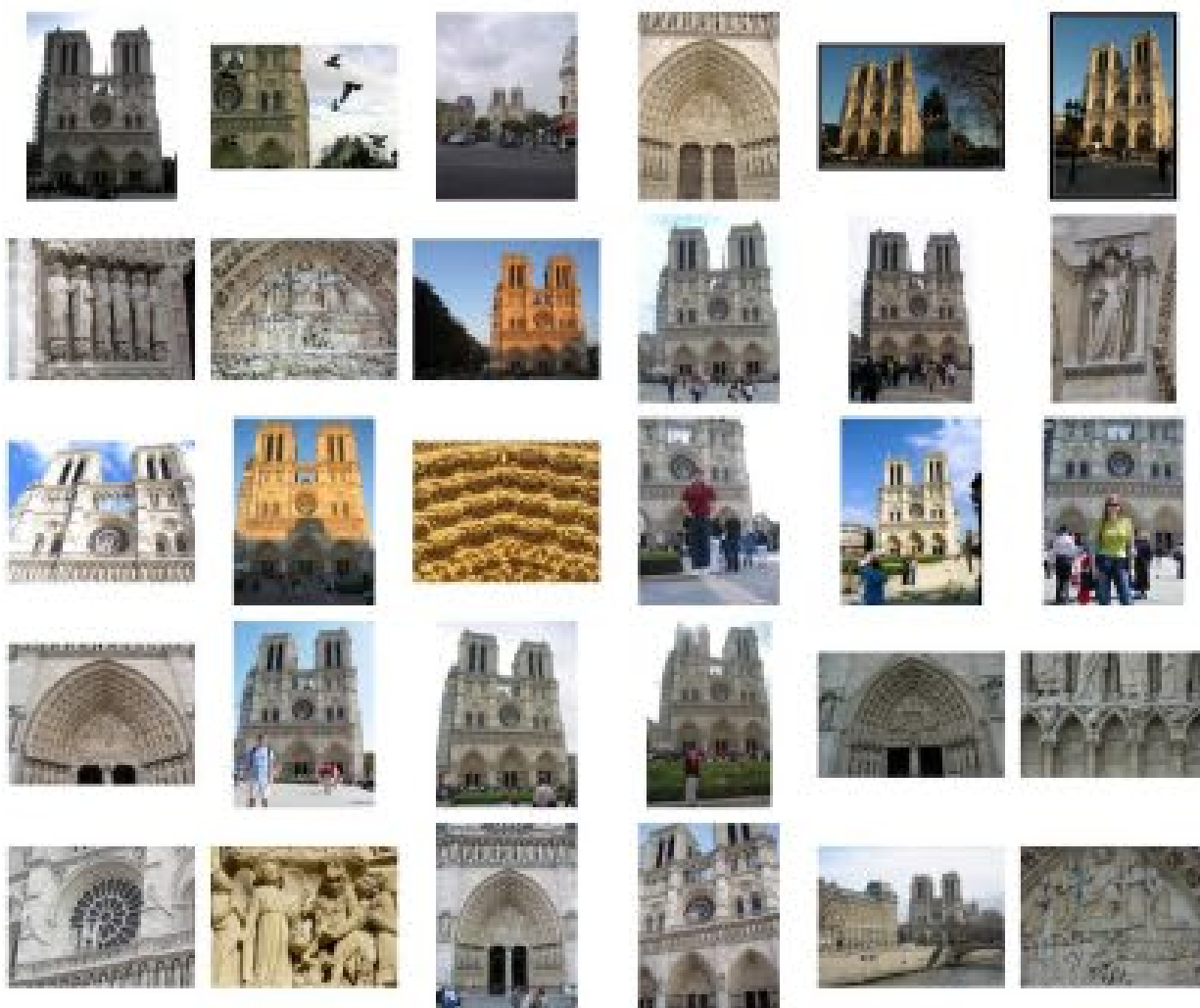
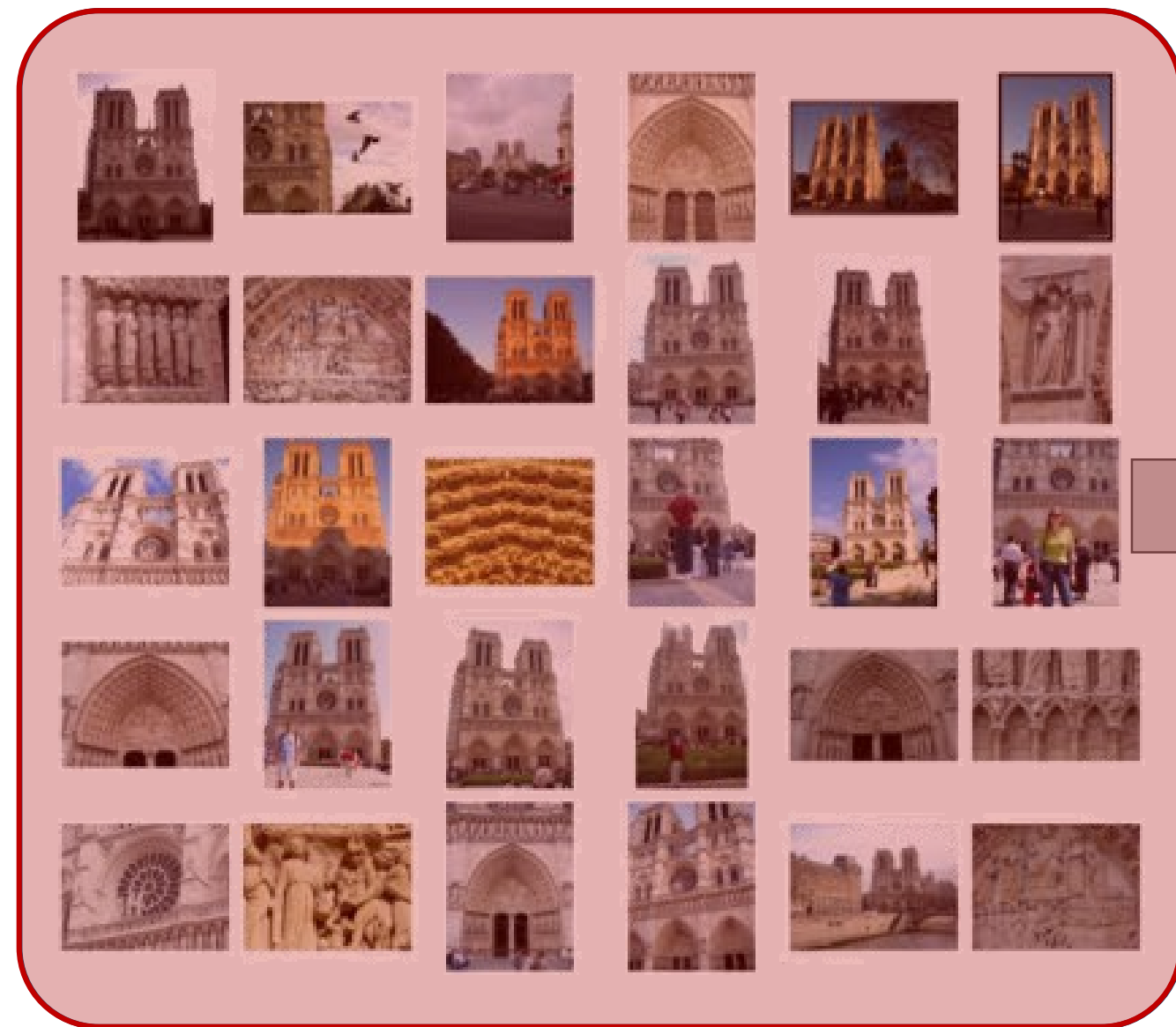
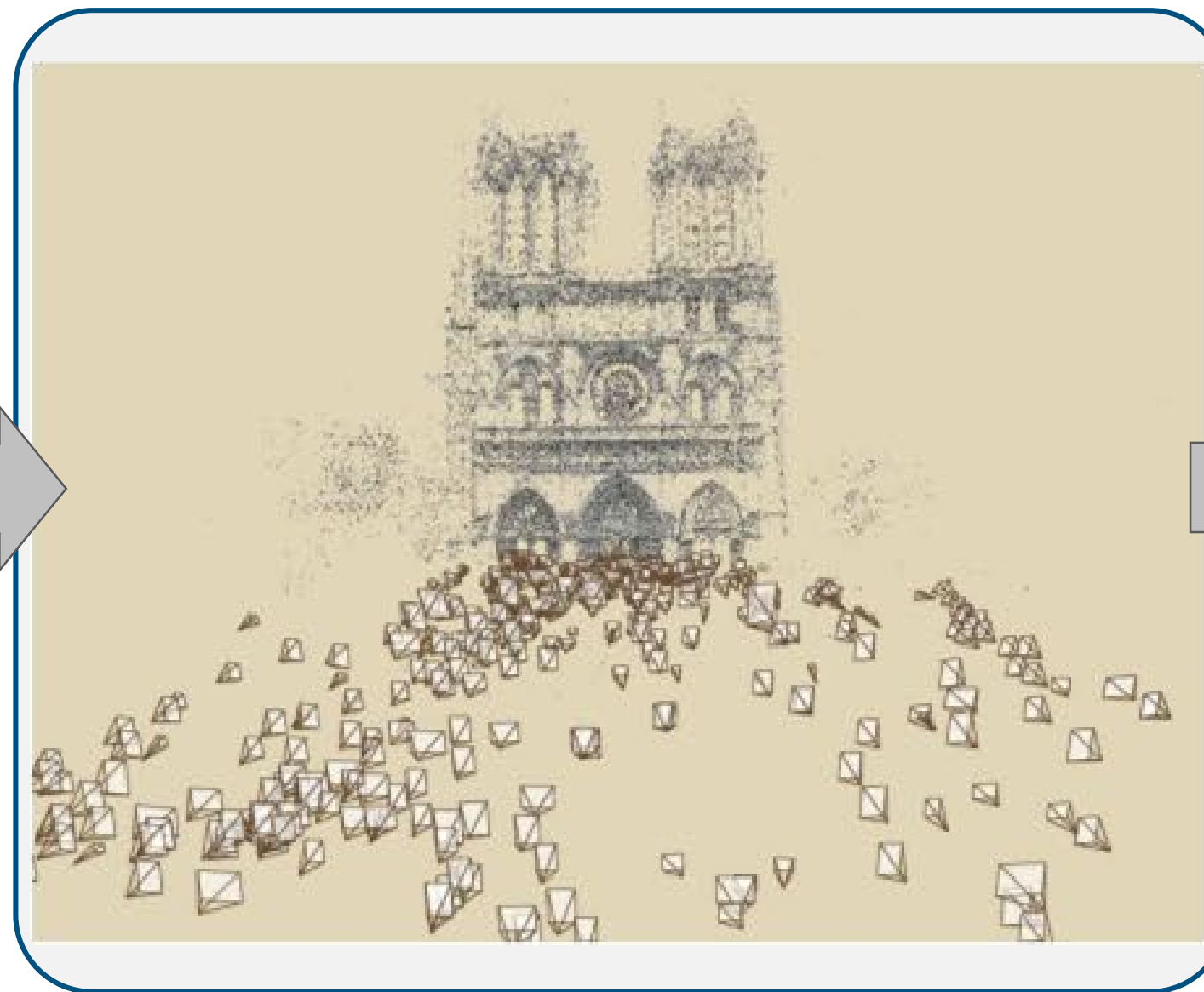


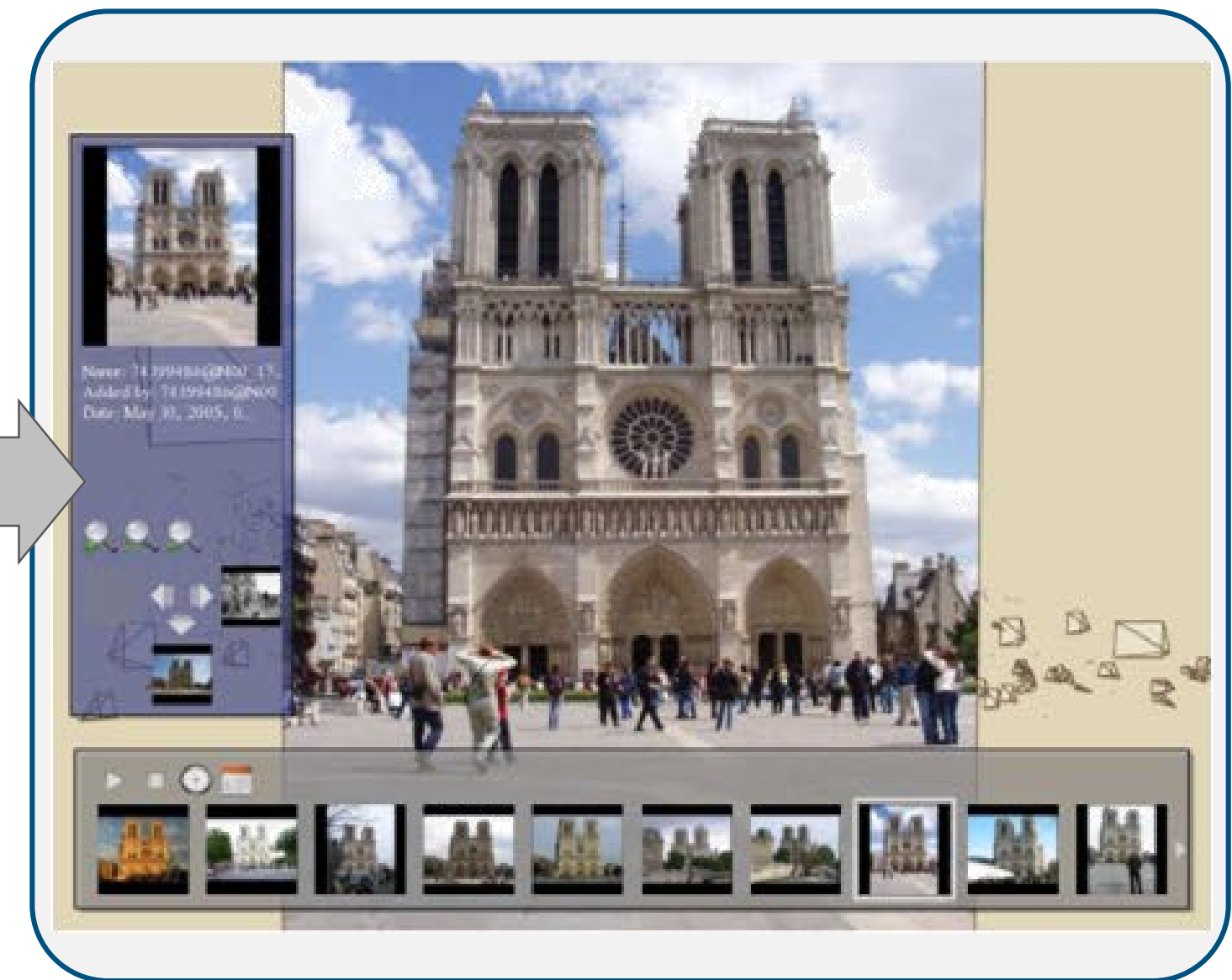
Photo Tourism



1) Images from the Web



2) Structure from Motion



3) Photo Explorer



15,464



37,383



76,389

Flickr: Creative Commons - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.flickr.com/creativecommons/by-nc-nd-2.0/

Home | Tags | Groups | People | Invite

Logged in as Jimantha: | Your Account | Help | Sign Out


Photos: Yours · Upload · Organize · Your Contacts · Explore

flickr BETA

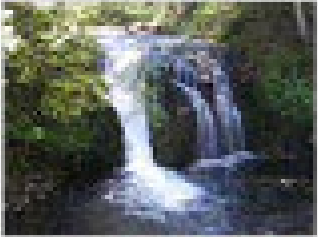
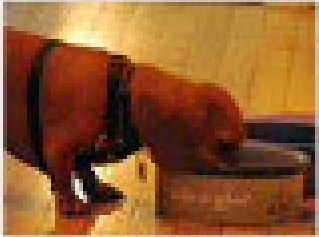
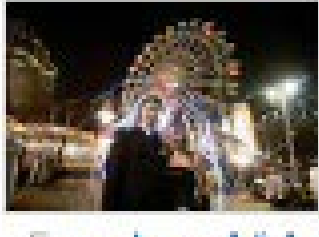
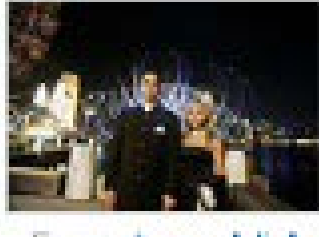
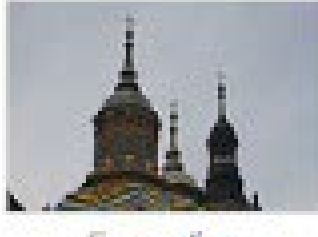
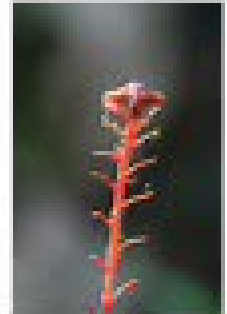
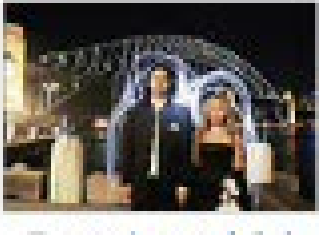
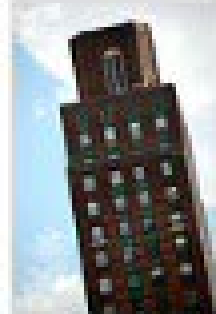
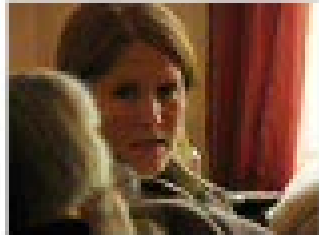
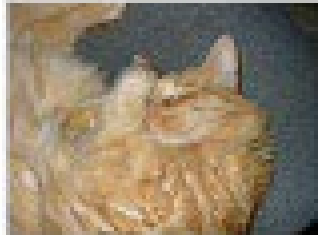
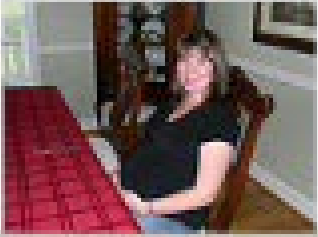
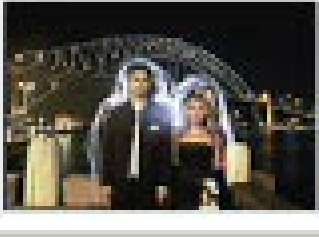


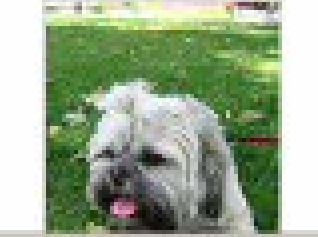
Creative Commons / Attribution-NonCommercial-NoDerivs License

SEARCH

(Or, [browse popular tags](#))

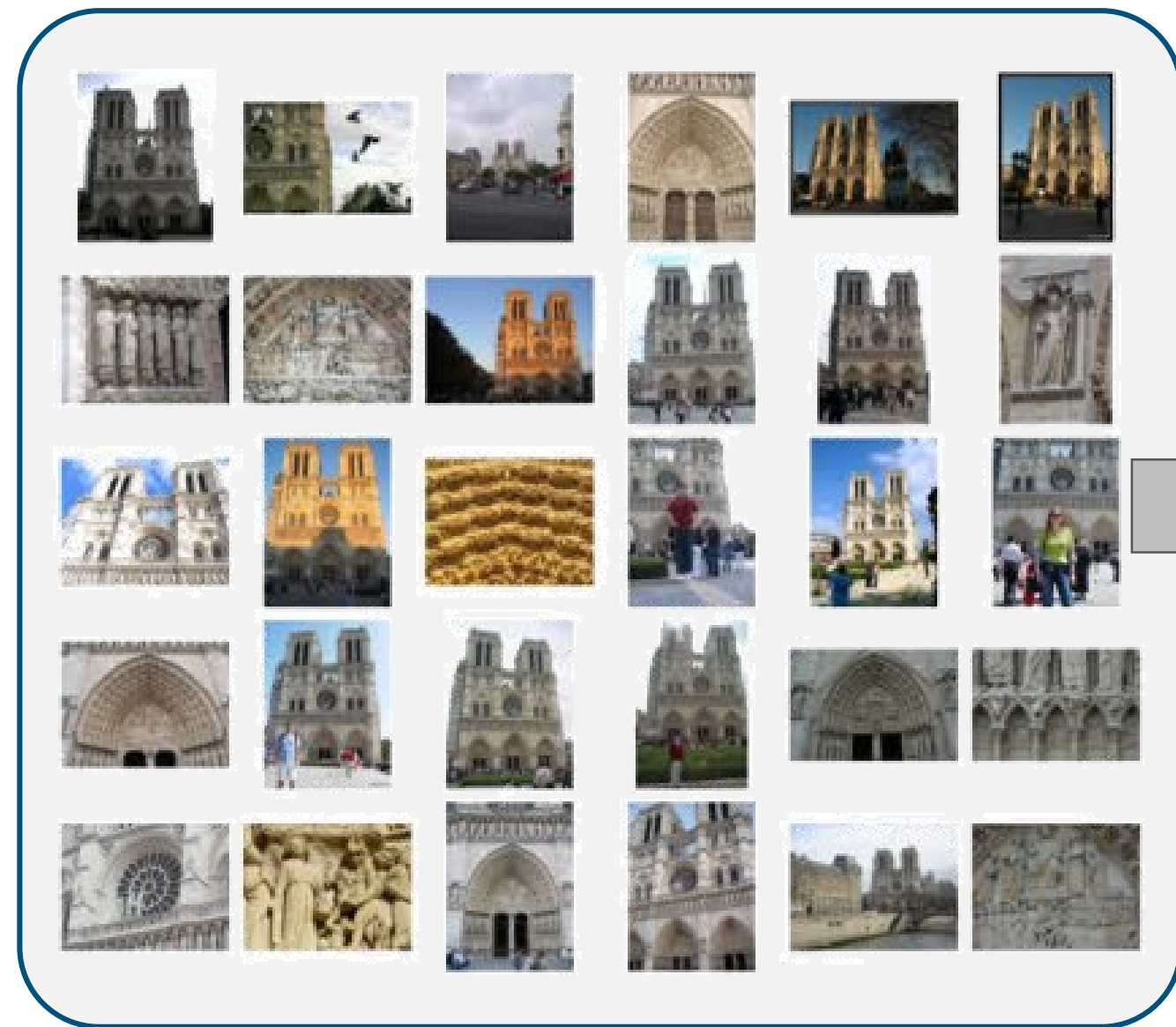
 SOME RIGHTS RESERVED

Here are the **100 most recent** licensed photos:

				
From inclarnell	From mugsy1274	From darren 'djp' paine	From darren 'djp' paine	From dizz
				
From alpha_zone	From darren 'djp' paine	From 331	From mugsy1274	From dalekg
				

Done Adblock

Photo Tourism



1) Images from the Web



2) Structure from Motion



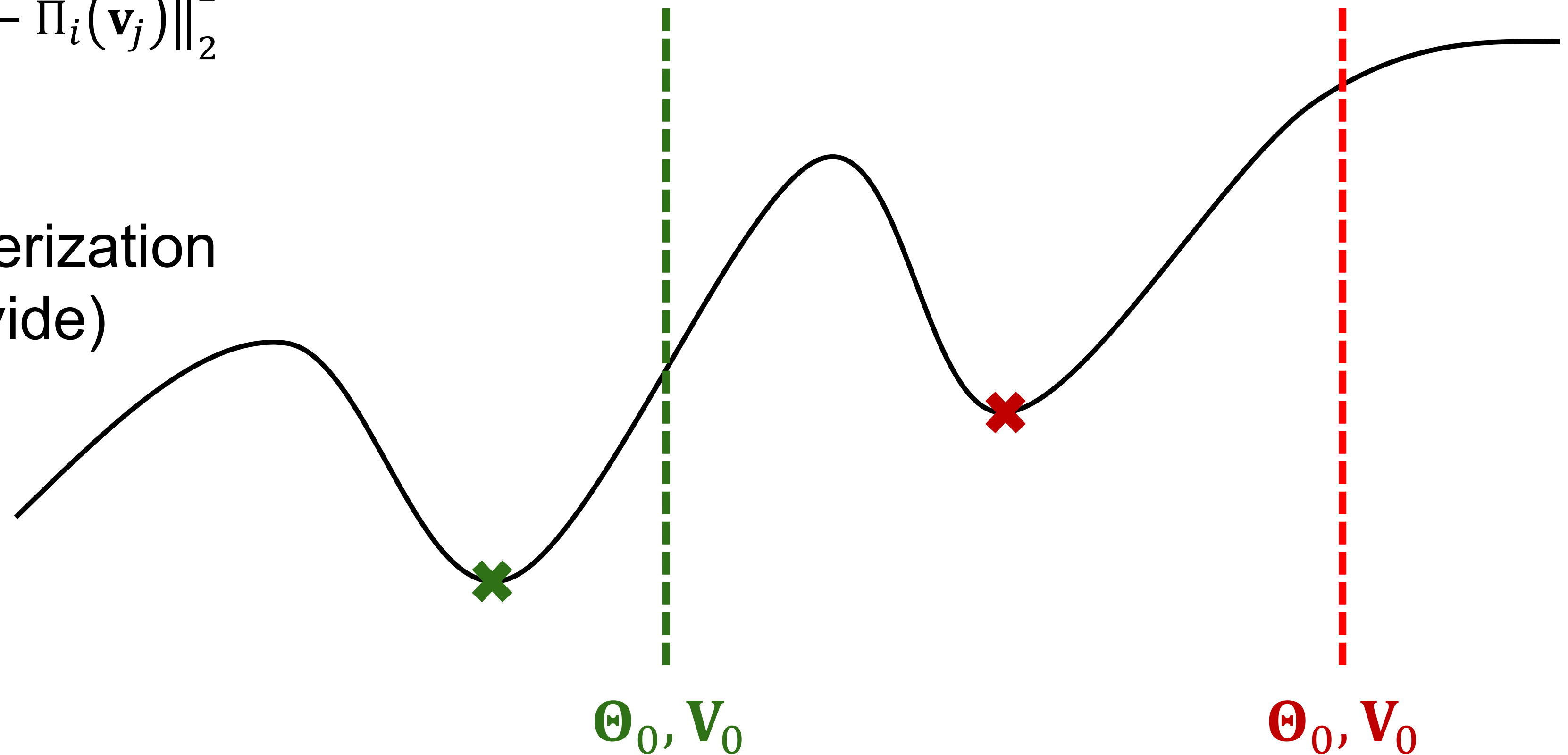
3) Photo Explorer

SfM / Non-Convex Optimization

$$E(\Theta, P) = \sum_{i=1}^N \sum_{j=1}^M w_{i,j} \|\mathbf{p}_{i,j} - \Pi_i(\mathbf{v}_j)\|_2^2$$

Why non-linear?

- Rotation parameterization
- Projection ('z'-divide)



We need a good initialization!

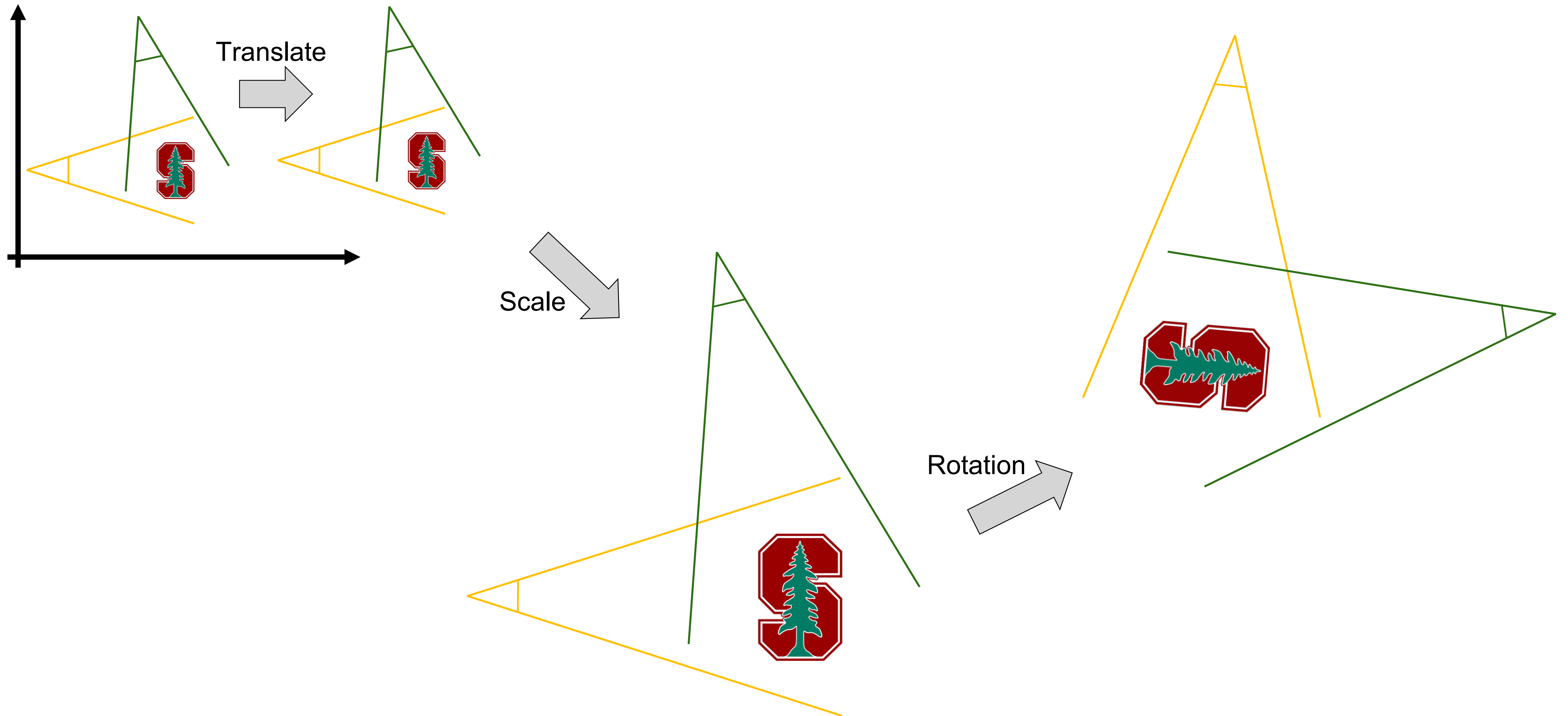
Incremental SfM

- Select two “suitable” images/cameras
- Find camera parameters using the 8-Point algorithm / RANSAC
- **Iterate:**
 - Add one camera using the Direct Linear Transform (DLT)
 - Bundle Adjustment (Levenberg Marquardt)
 - Prune “bad tracks” after each iteration

Incremental Structure from Motion



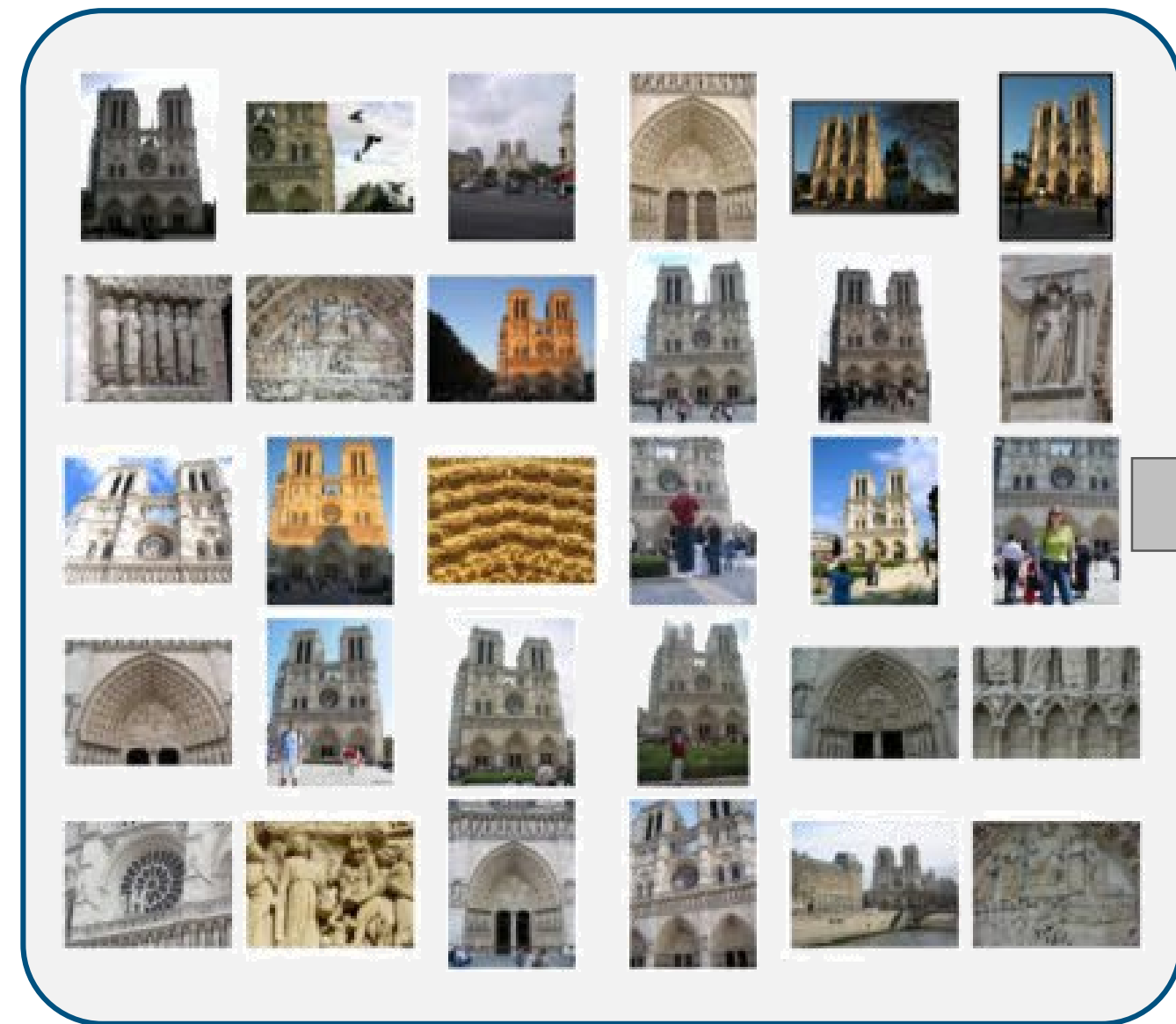
Scale, Rotation, Position Ambiguity



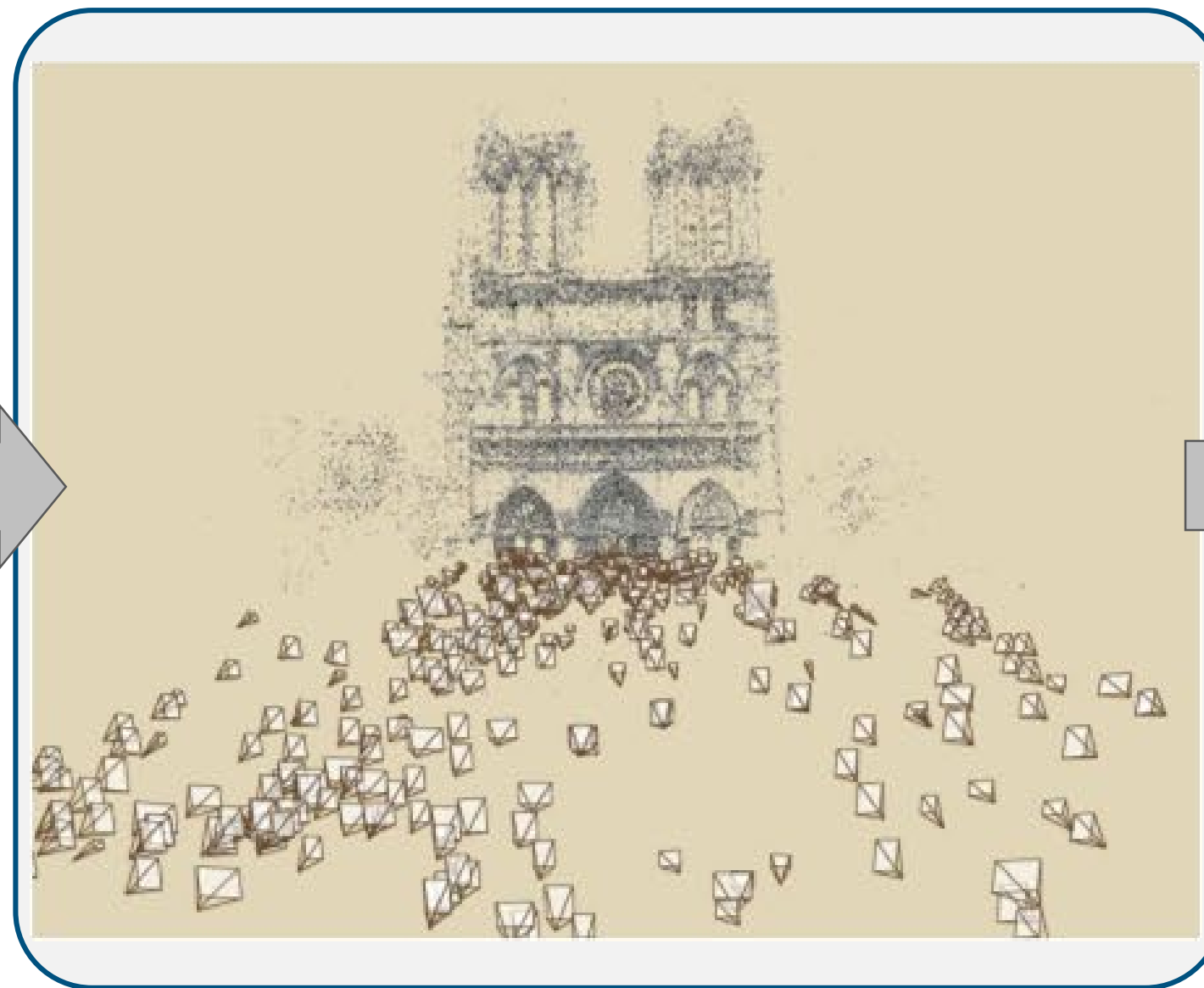
Geo-Registration



Photo Tourism



1) Images from the Web



2) Structure from Motion



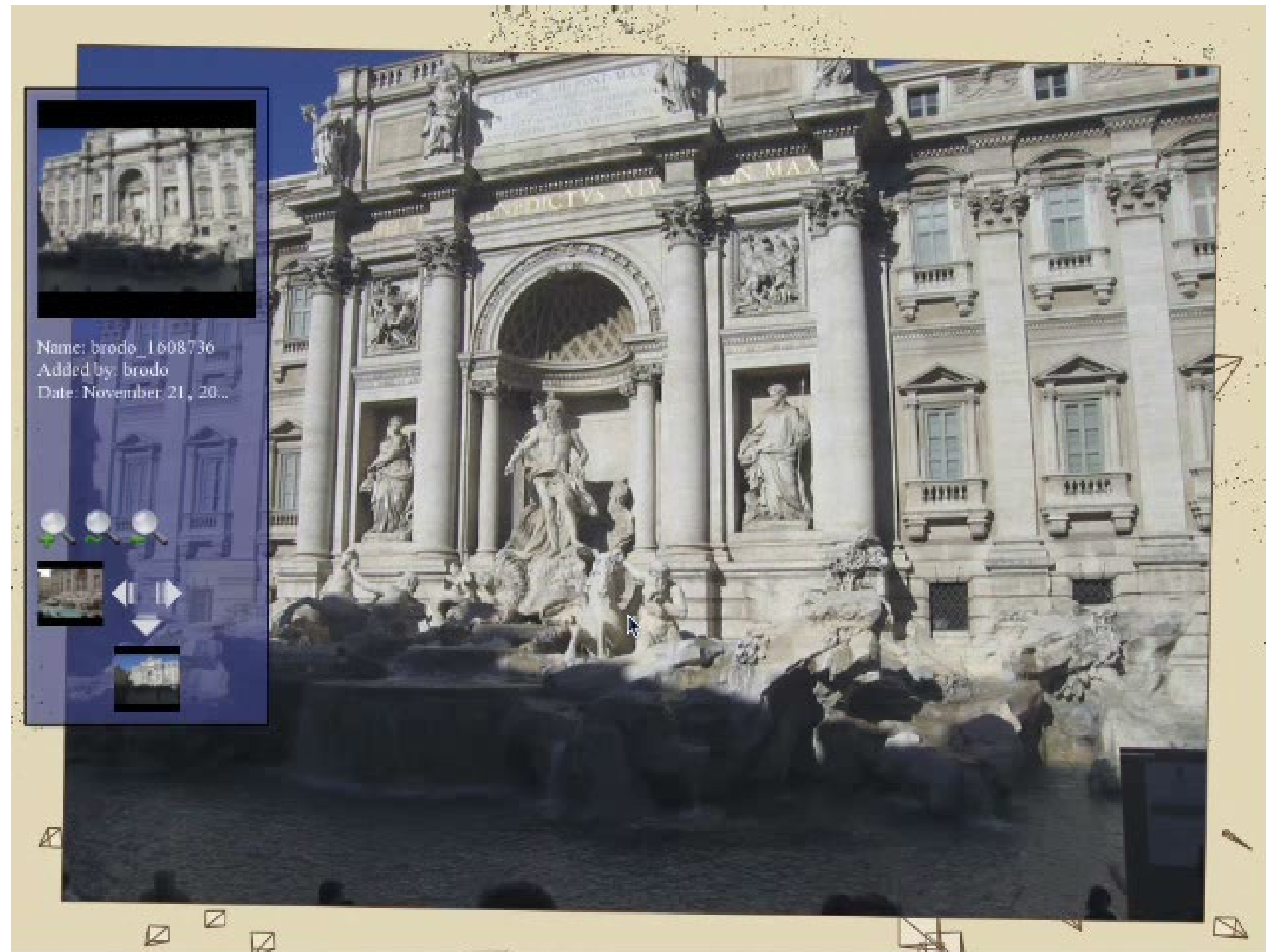
3) Photo Explorer

- Browsing
- Rendering
- Annotation

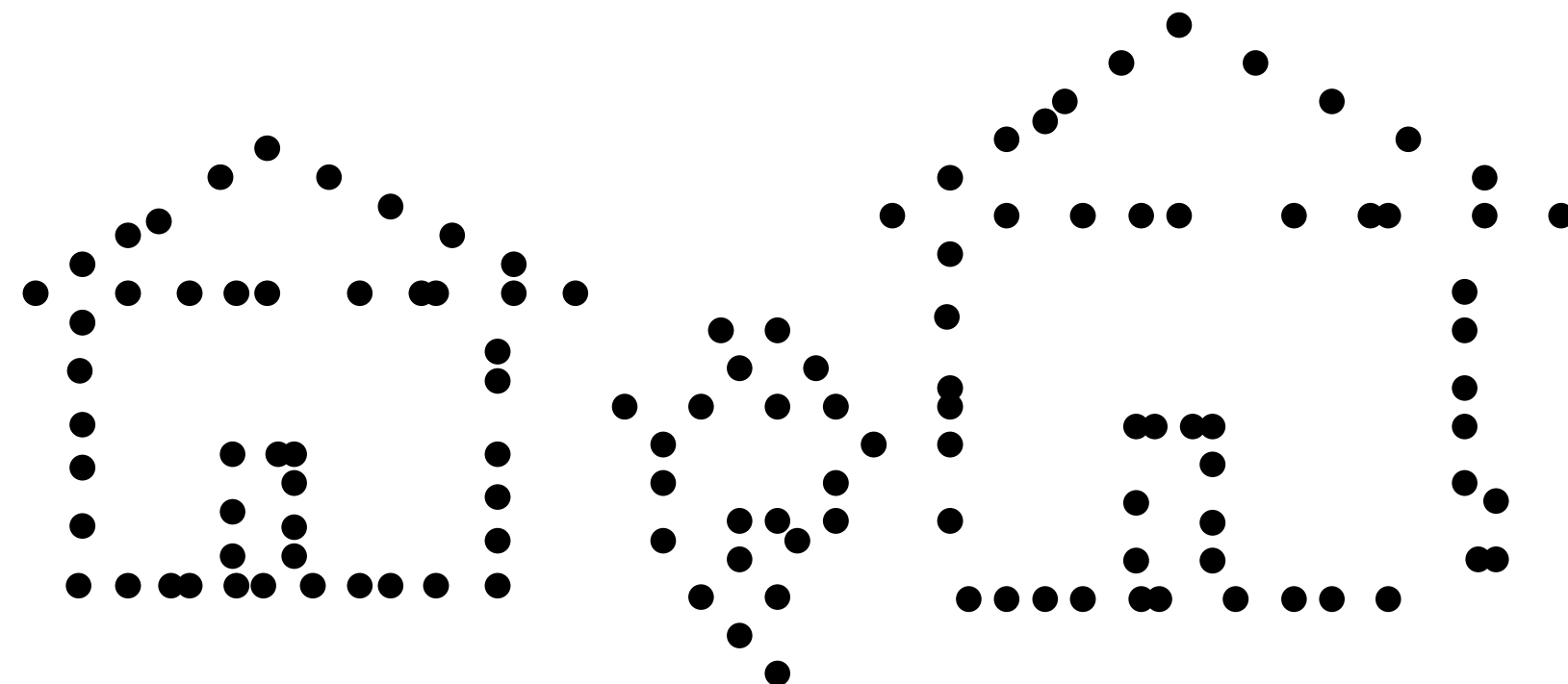
3D Browsing



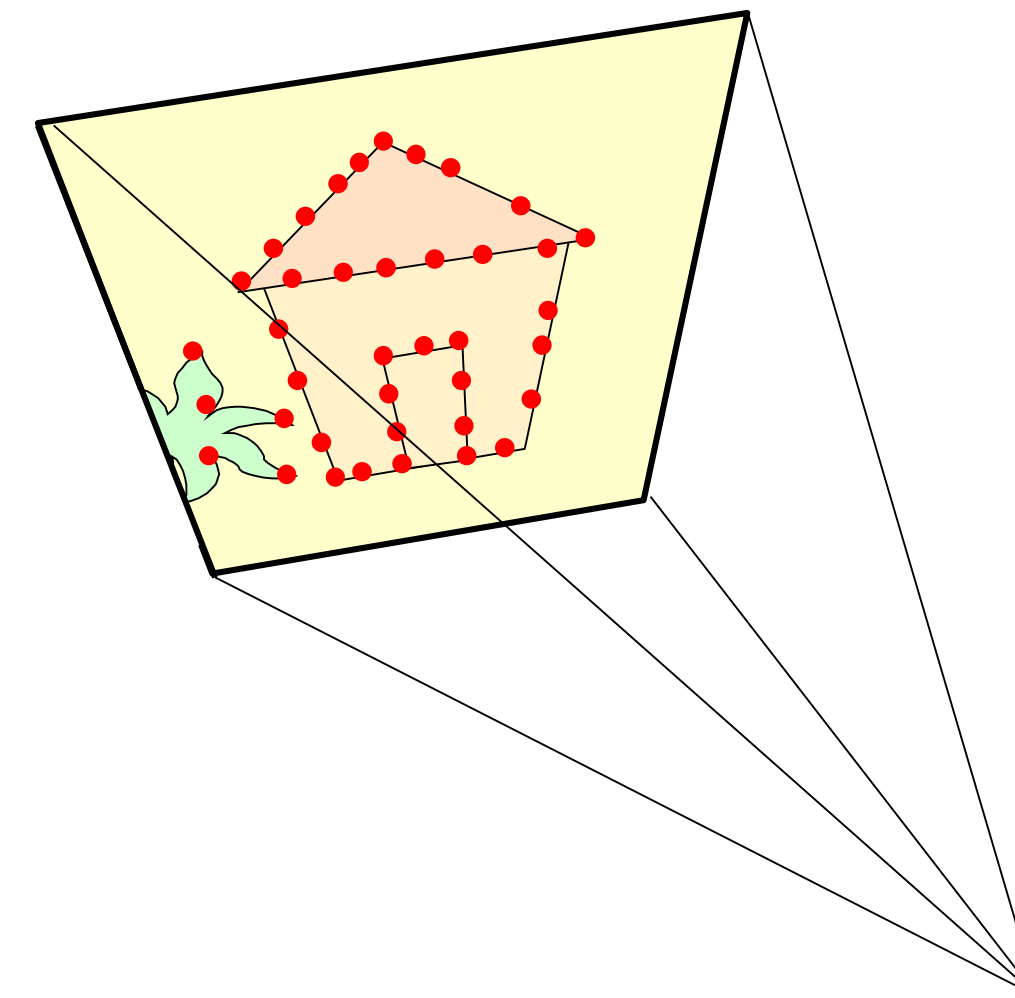
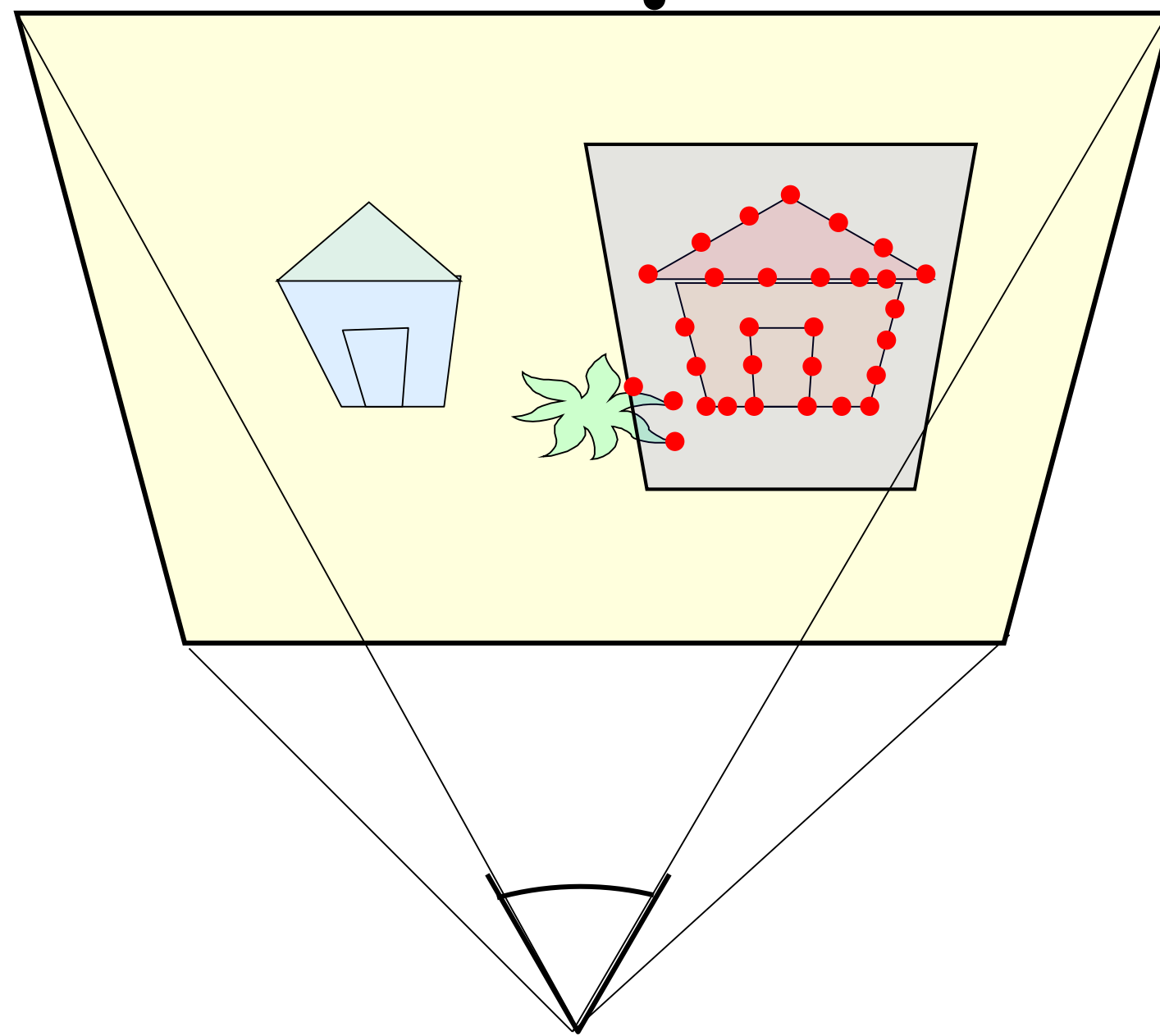
Object-based browsing



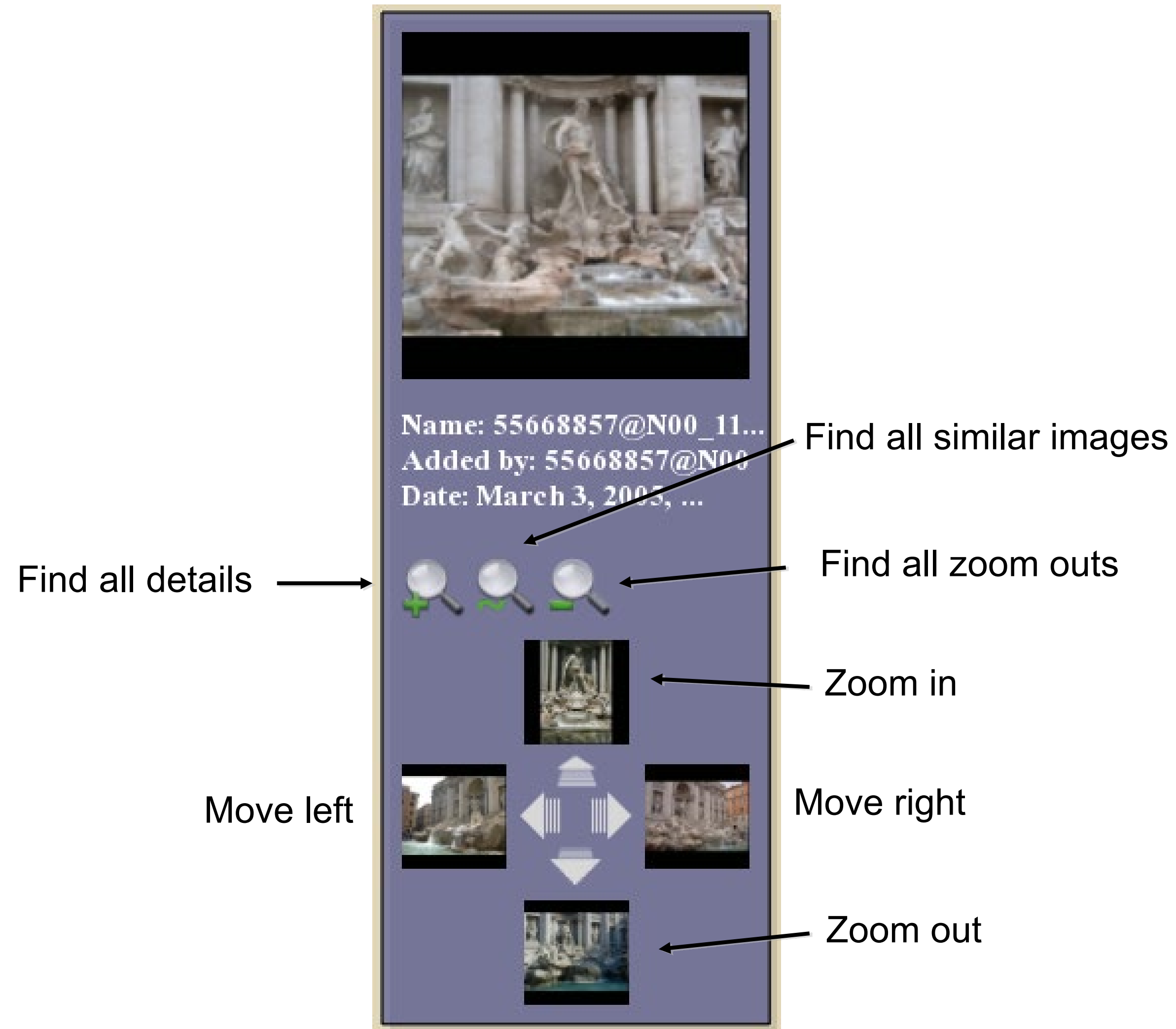
Object-based browsing



- Visibility
- Resolution
- Head-on view



Relation-based browsing



Relation-based browsing



Relation-based browsing

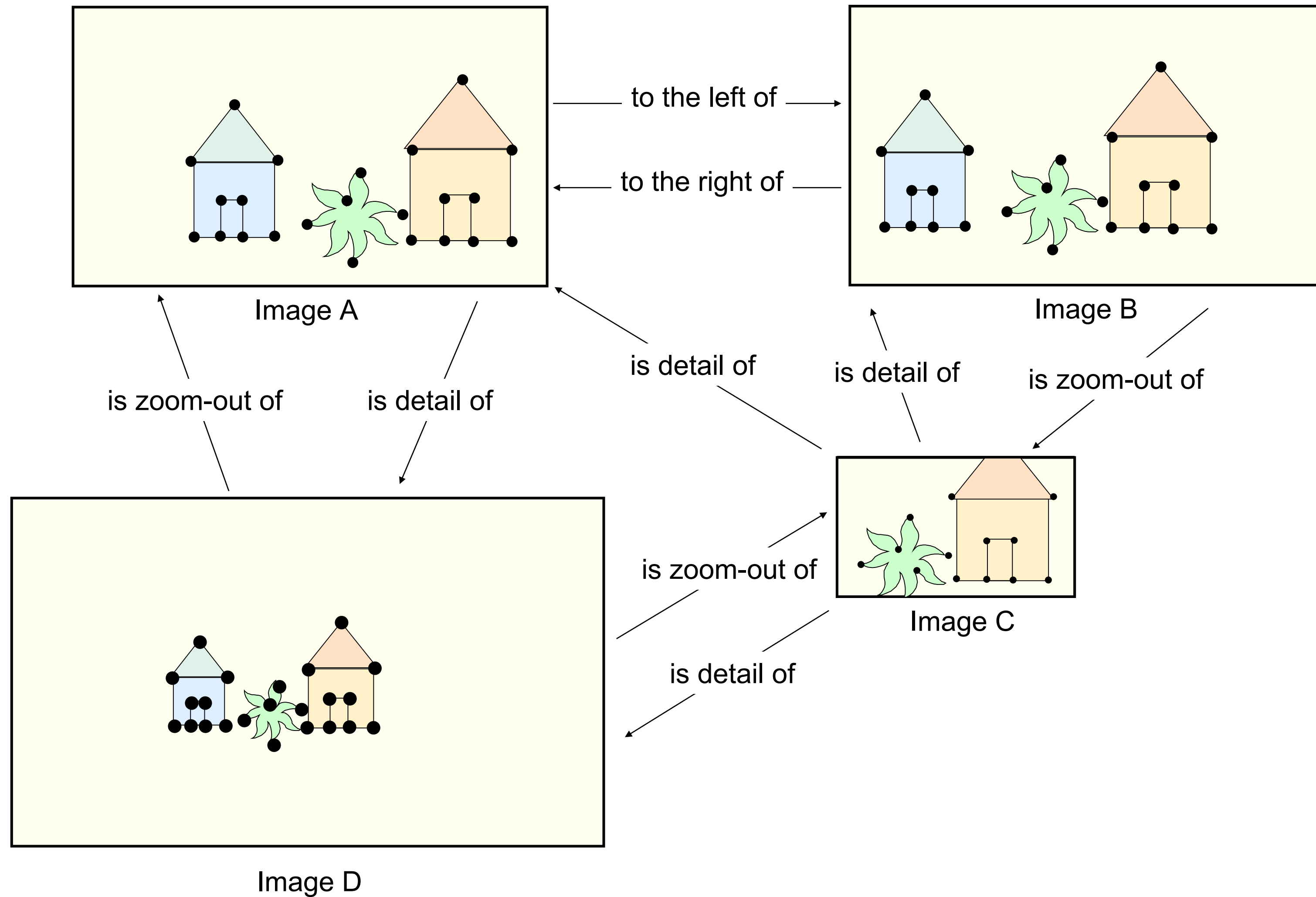
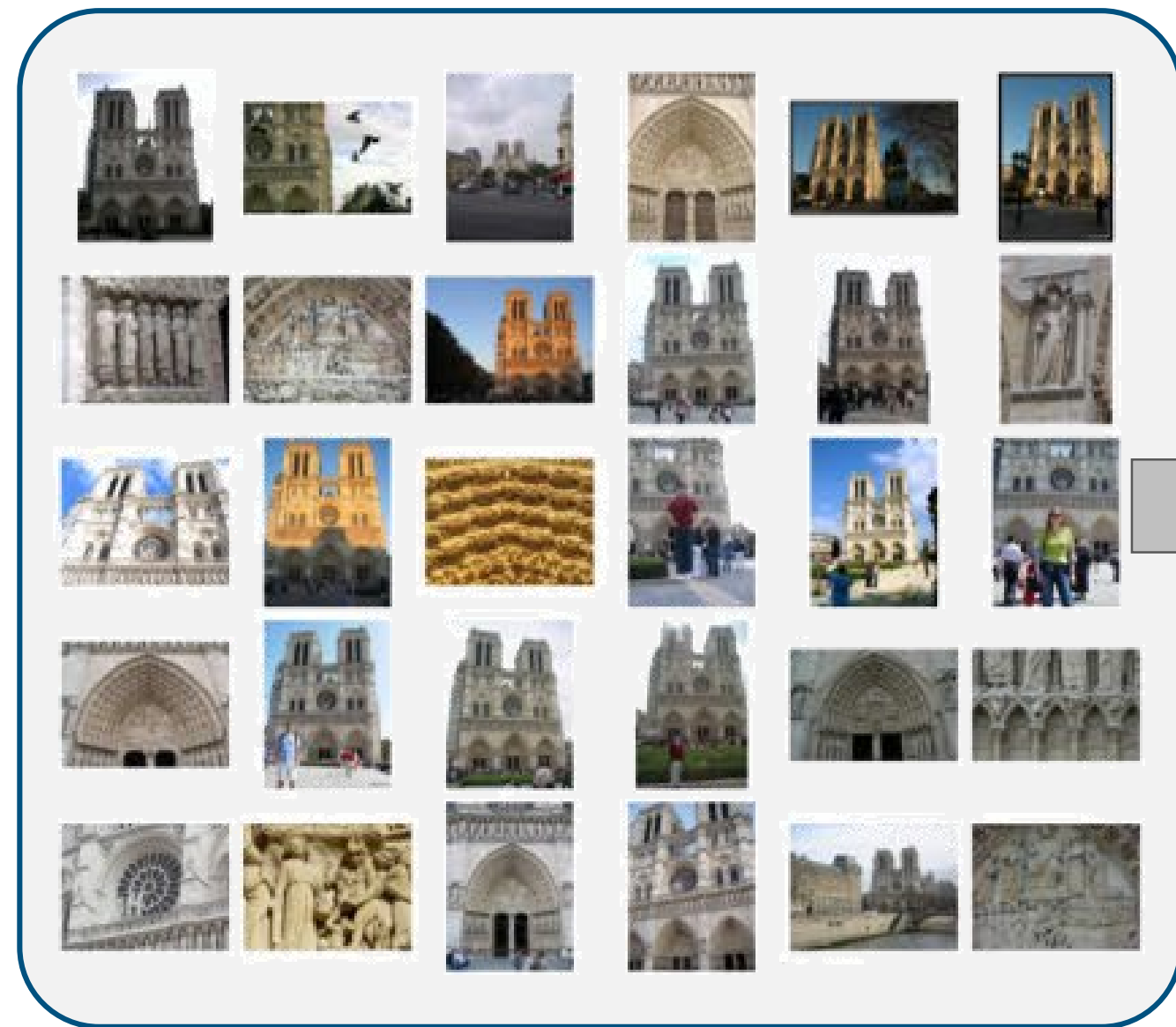
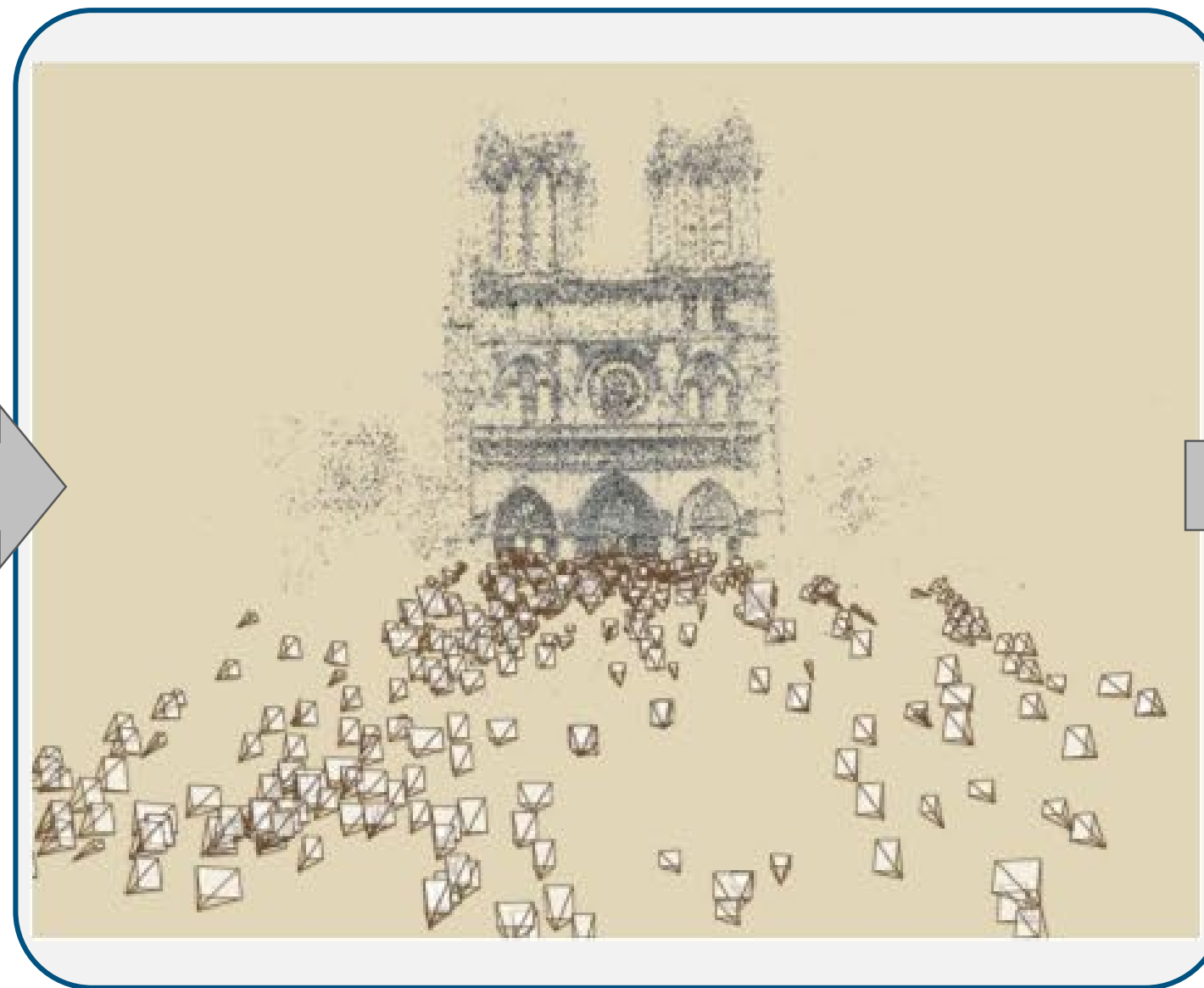


Photo Tourism



1) Images from the Web



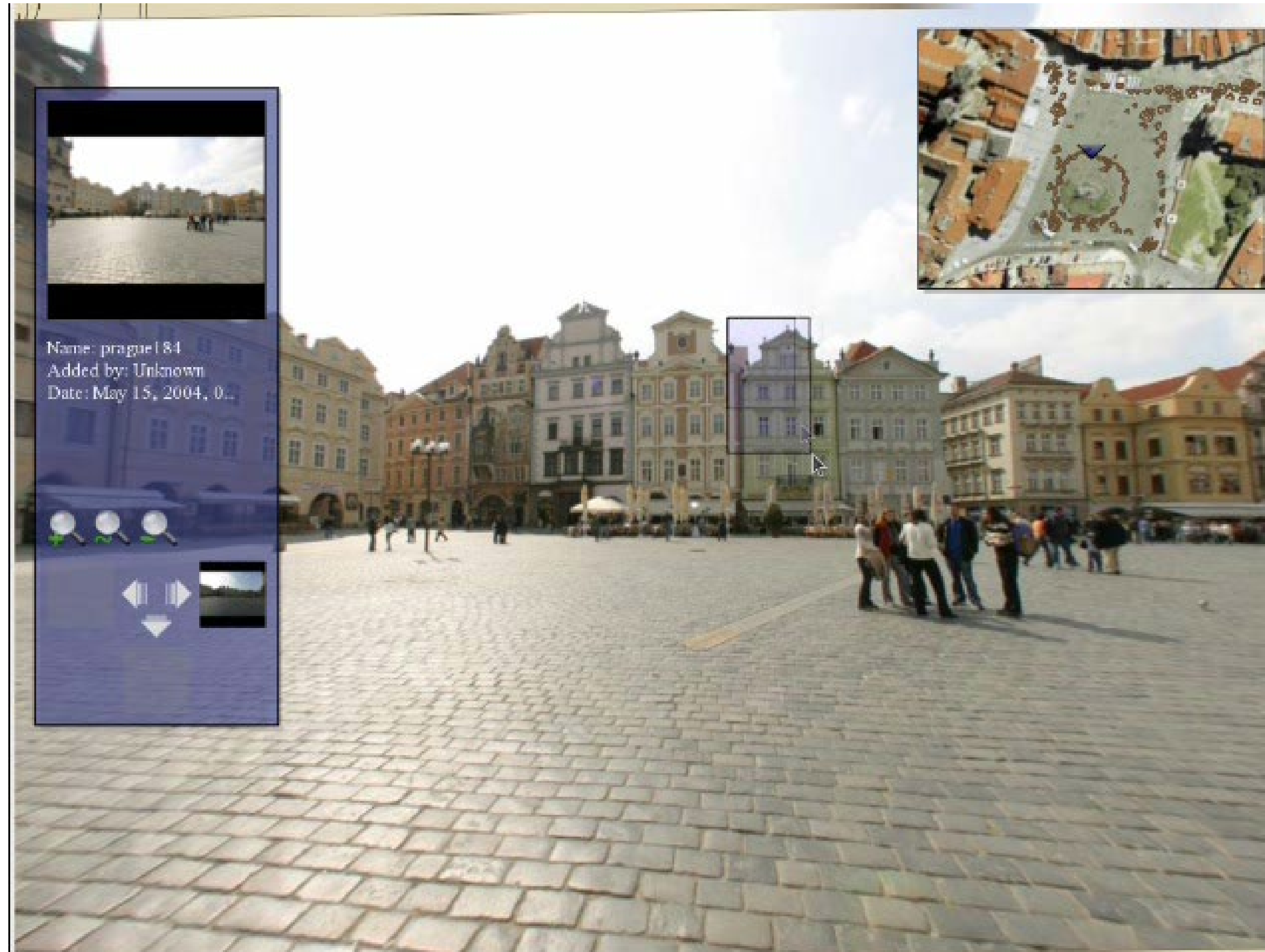
2) Structure from Motion



3) Photo Explorer

- Browsing
- Rendering
- Annotation

Rendering Transitions



Rendering Transitions

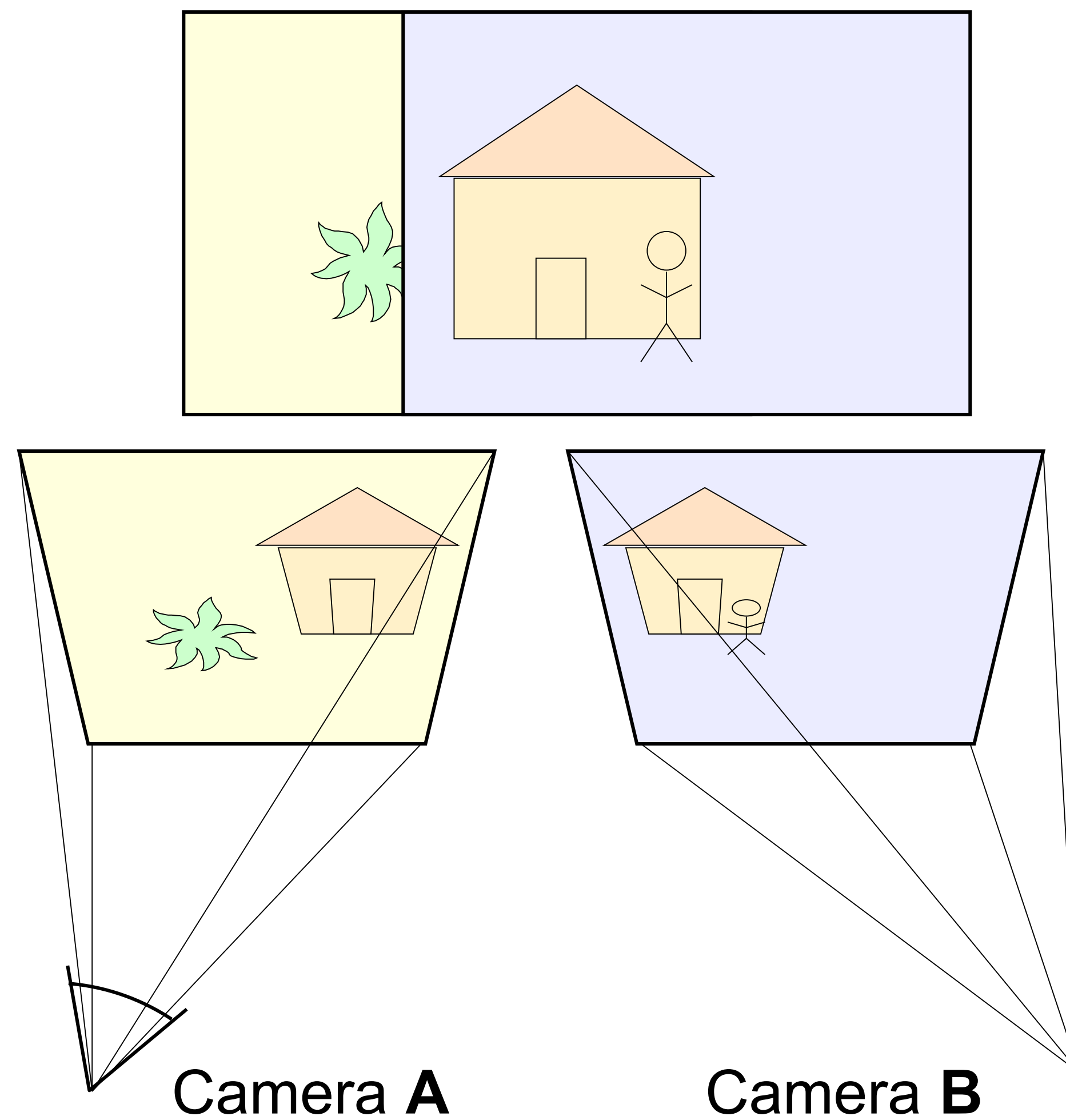
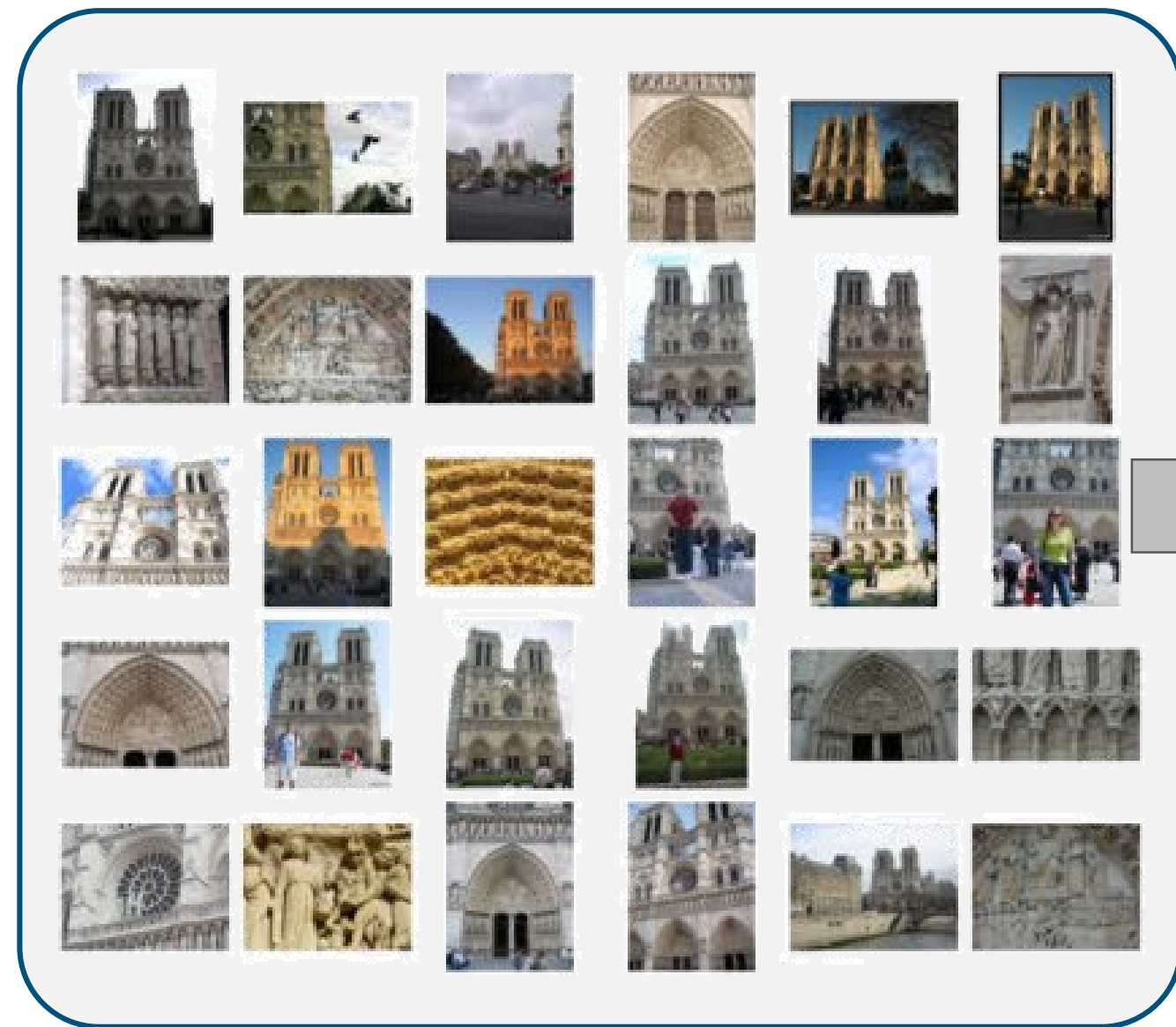
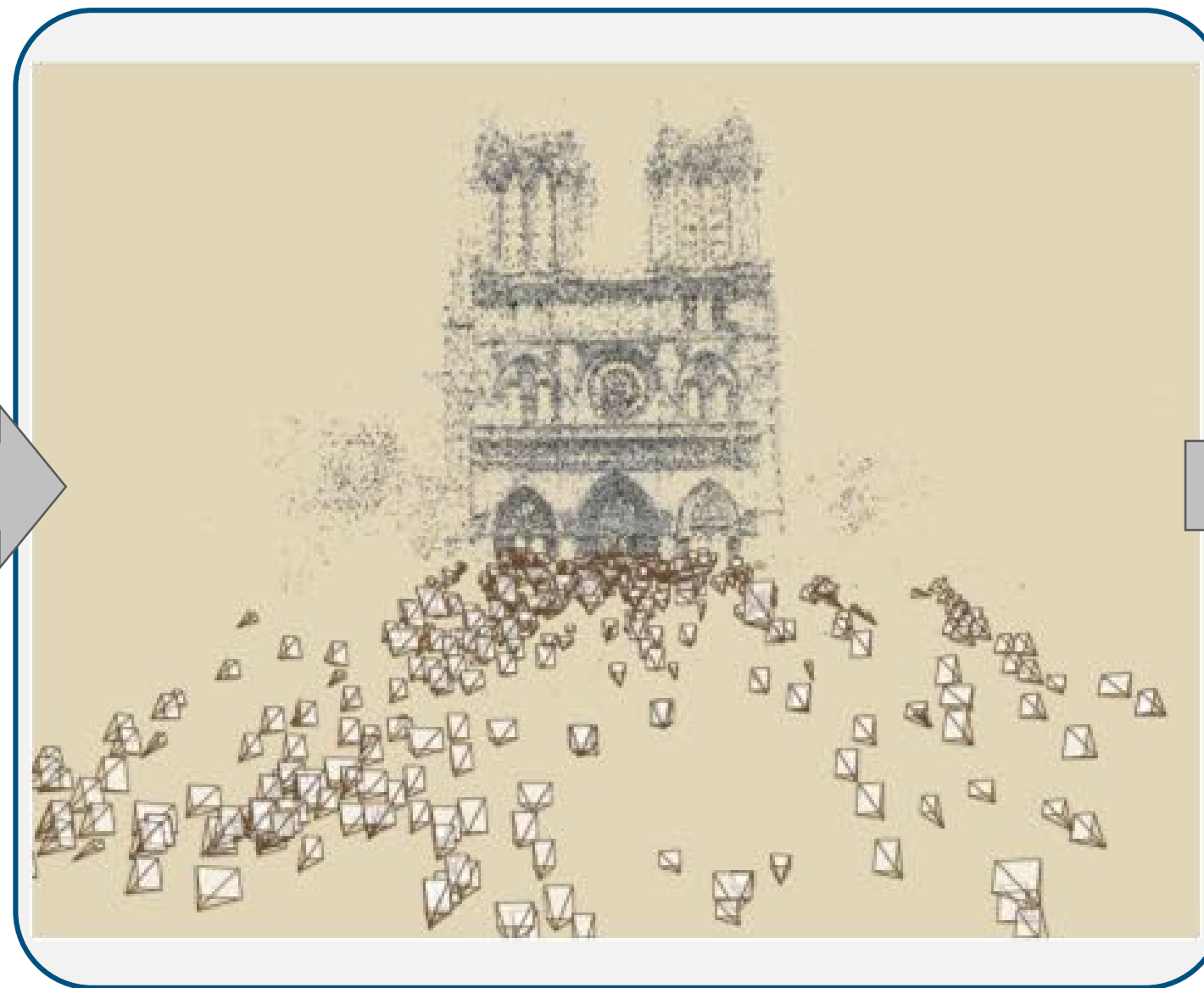


Photo Tourism



1) Images from the Web



2) Structure from Motion



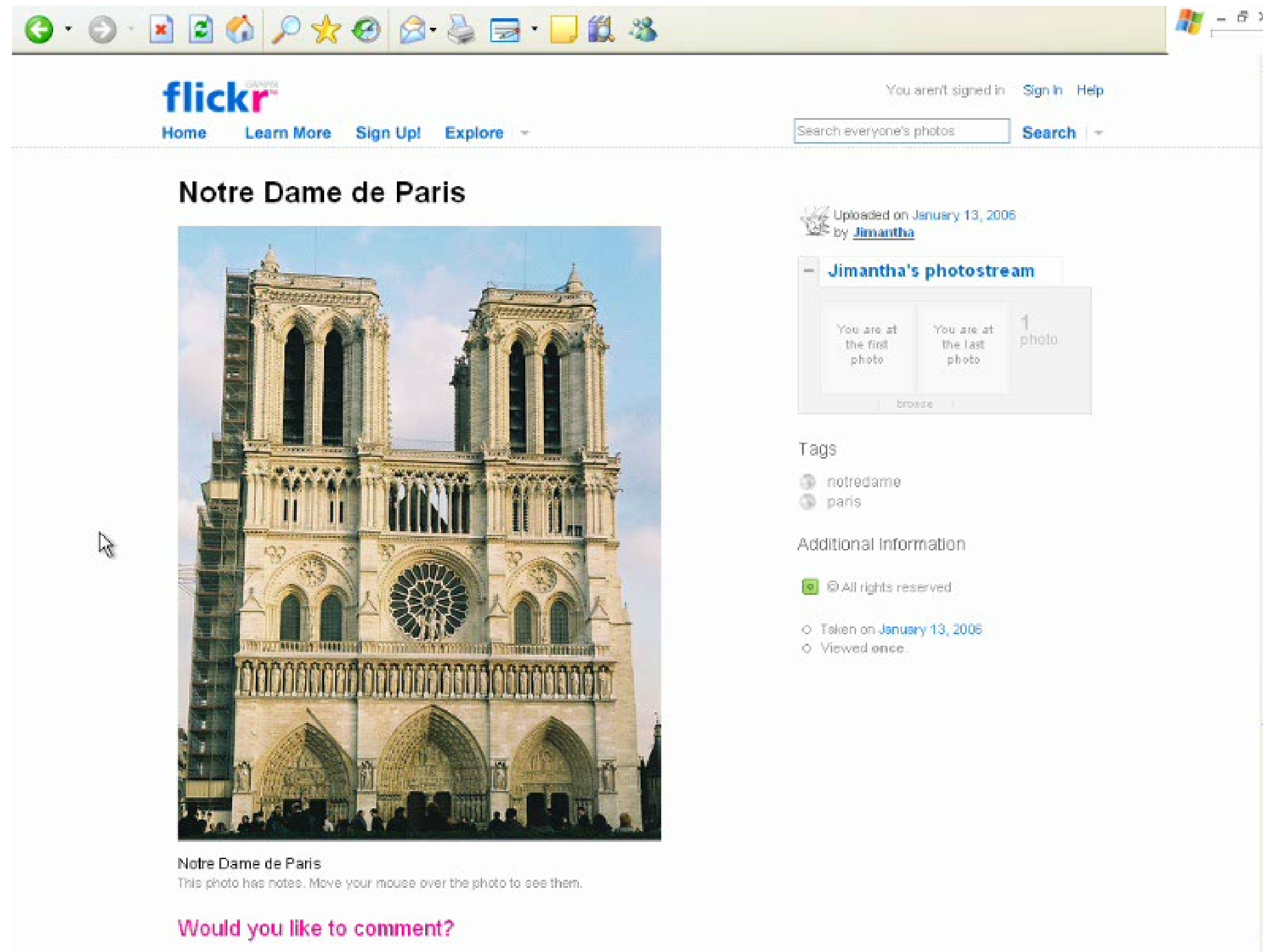
3) Photo Explorer

- Browsing
- Rendering
- Annotation

Annotations



Annotations



The screenshot shows a web browser window displaying a Flickr page for a photo of Notre Dame de Paris. The browser's address bar is empty, and the toolbar contains various icons for navigation and utility. The Flickr page header includes the logo, navigation links (Home, Learn More, Sign Up!, Explore), and a search bar. The main content area features the title "Notre Dame de Paris" above a large photograph of the cathedral's facade. To the right of the photo, there is a metadata section indicating the photo was uploaded on January 13, 2006, by user "Jimantha". Below this is a "photostream" widget showing the user is at the first and last photo of a single-photo stream. Further down, there are tags for "notredame" and "paris", and an "Additional information" section with a copyright notice and statistics: "Taken on January 13, 2006" and "Viewed once". At the bottom of the page, a pink text prompt asks "Would you like to comment?".

Notre Dame de Paris

Uploaded on January 13, 2006
by Jimantha

Jimantha's photostream

You are at the first photo | You are at the last photo | 1 photo

Tags

- notredame
- paris

Additional information

© All rights reserved

Taken on January 13, 2006
Viewed once

Would you like to comment?

Reproduced with permission of Yahoo! Inc. © 2005 by Yahoo! Inc.
YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc.

Limitations

1. Some photos can not be registered

- Reason: blur, noise, no overlap, dynamic objects, ...

2. Runtime Speed (up to several weeks)

- Fix: Hierarchical approaches, Parallelization, ...

Building Rome in a Day



The old city of Dubrovnik, 4,619 images, 3,485,717 points

Small Detour: Bundle Fusion

Live Capture



Limitations

1. Some photos can not be registered

- Reason: blur, noise, no overlap, dynamic objects, ...

2. Runtime Speed (up to several weeks)

- Fix: Hierarchical approaches, Parallelization, ...

3. Transitions do not look “real”

- Better Image-based or Neural Rendering?

Enhancing Transitions

Neural Rerendering in the Wild

Moustafa Meshry¹, Dan B Goldman², Sameh Khamis², Hugues Hoppe², Rohit Pandey²,
Noah Snavely², Ricardo Martin-Brualla²

¹University of Maryland, ²Google Inc.

Summary

- 3D is important
- Perspective Cameras
- Structure from Motion
- Photo Tourism

