The Distributed Camera

Noah Snavely
Cornell University

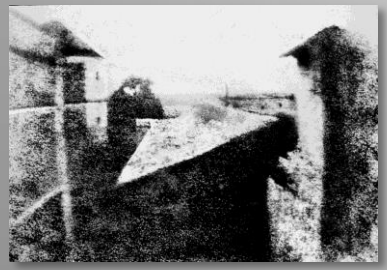
Microsoft Faculty Summit
June 16, 2013

The Age of Exapixel Image Data

- Over a *trillion* photos available online
- Millions uploaded every hour
- Interconnected
- The Internet is becoming a **living visual record** of our world

Photos over time

1820's
Early photography



1900
Kodak Brownie

1990's
Digital cameras

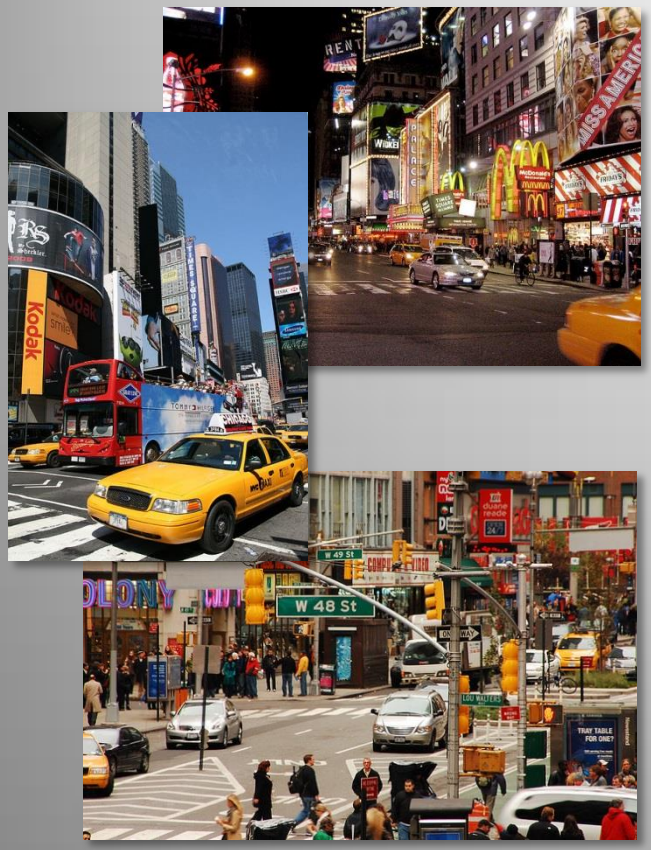
2000's
Photo sharing

350 million photos per day (Facebook)

100 hours of video per minute (YouTube)

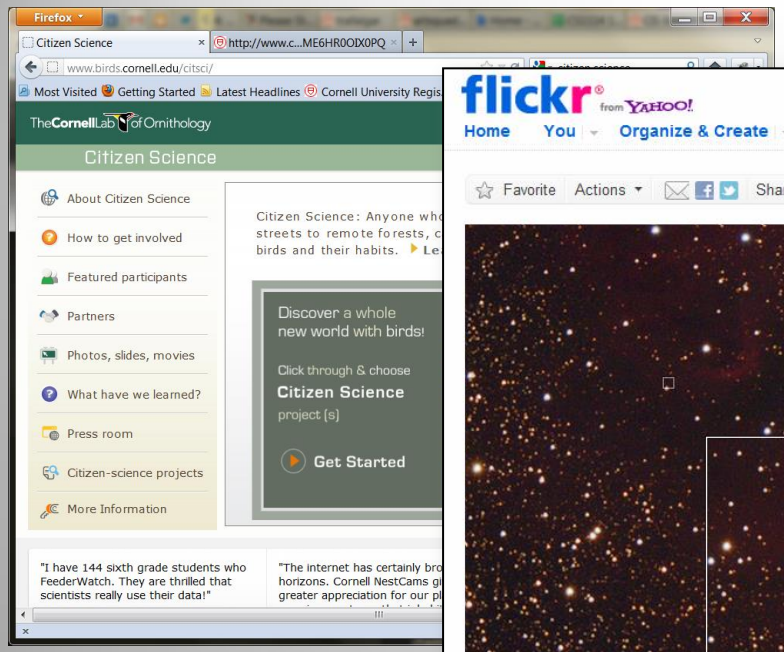
"Every 2 minutes today we capture as many photos as the whole of humanity took in the 1800s." [1000memories]

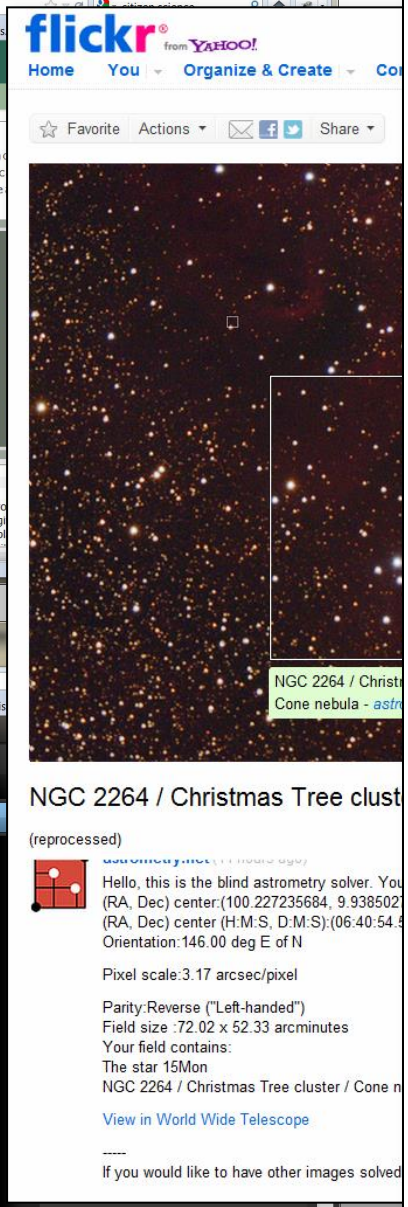
What can we do with all this data?

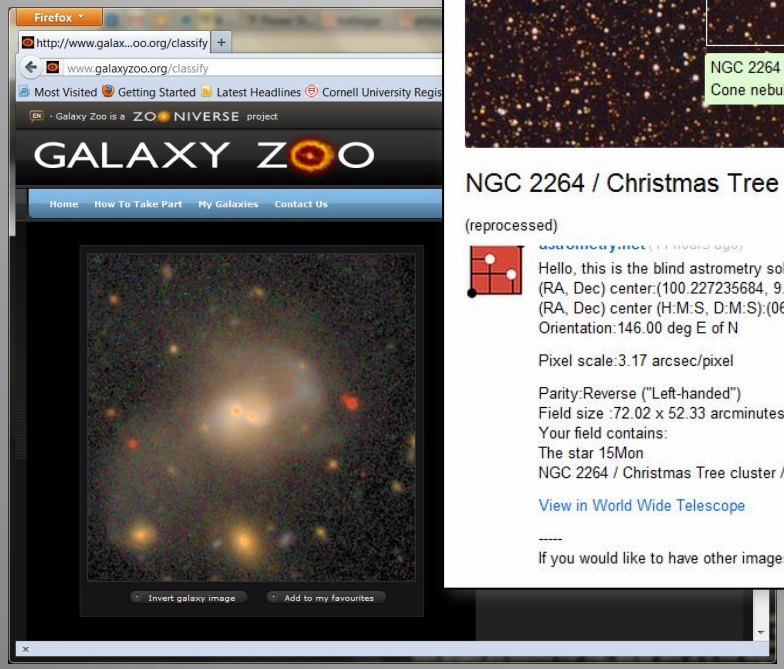


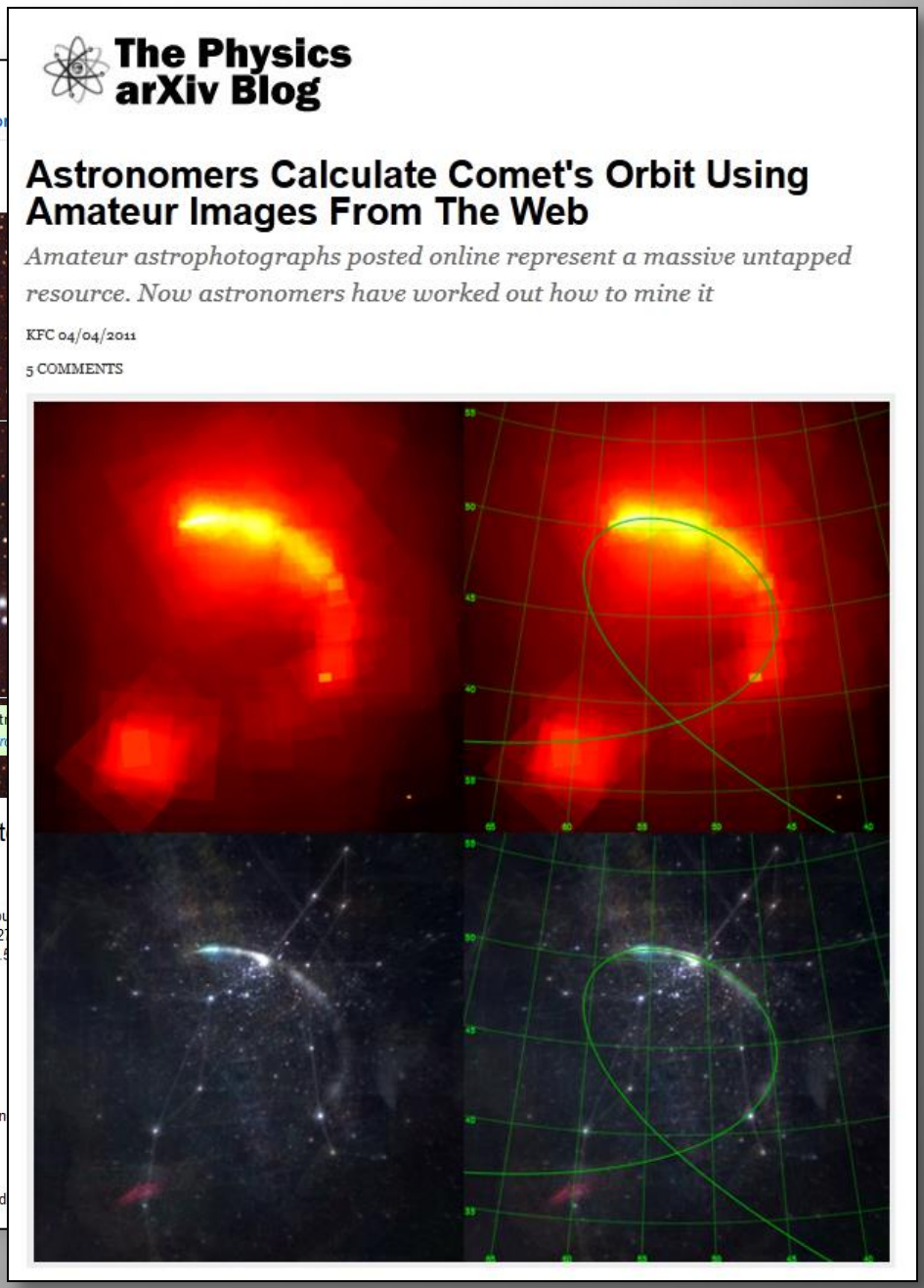
- Use images to understand the world
  - Changes in cities and environments over time
  - High-level behaviors, e.g. traffic patterns, pedestrian moments
  - Surprising events
  - Forensics – what happened, when?

- **Challenge**: data is extremely unstructured

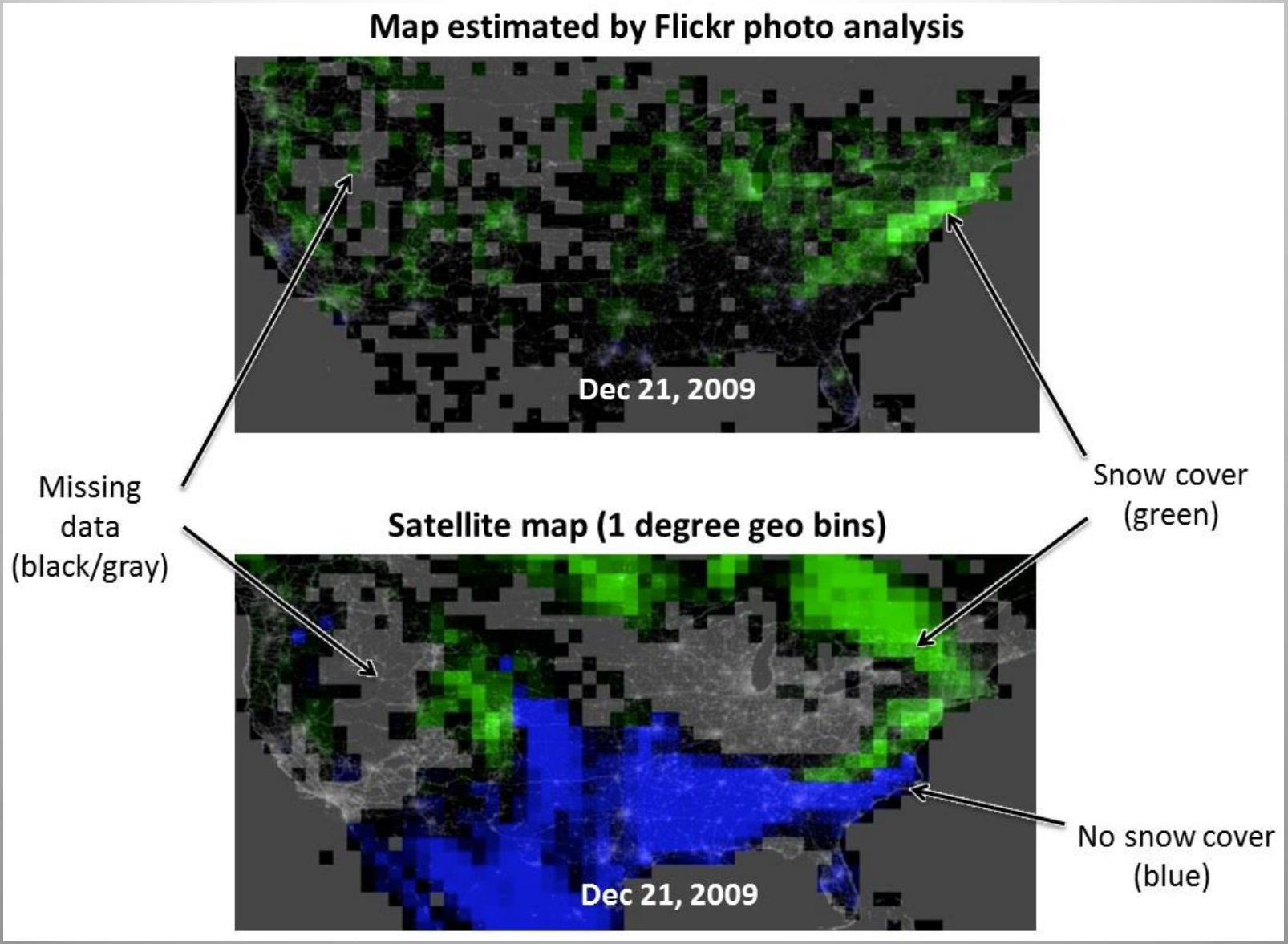








http://www.technologyreview.com/blog/arxiv/26603/

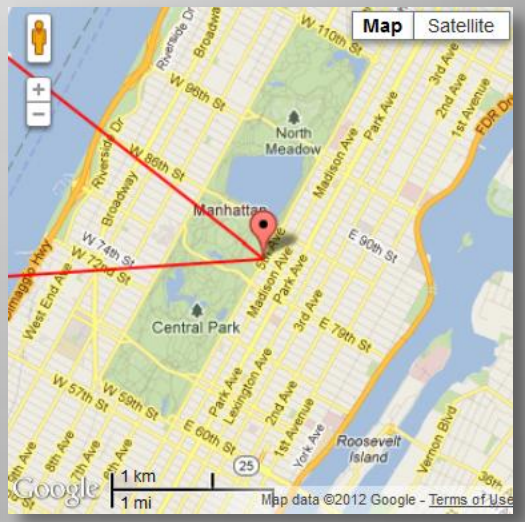


Snow cover from Flickr photos

[Zheng, Korayem, Crandall, LeBuhn, WWW 2012]

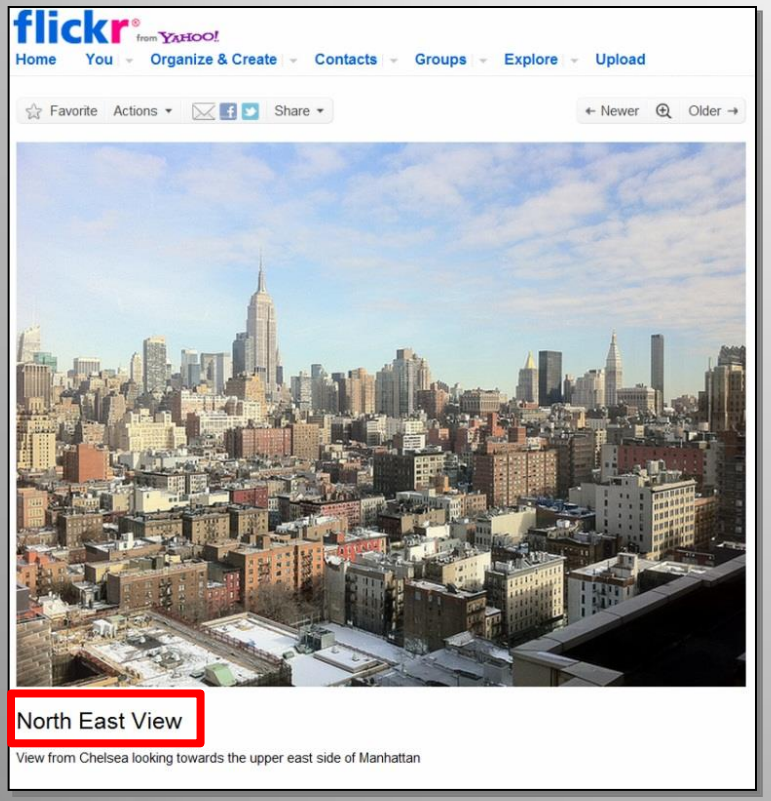
Calibrating the distributed camera

- For any photo on the web



- Where was it taken? In what direction?
- What time was it taken?
- What is visible in the image? Where?
- Our work: vision tools to provide basic calibration data

What about sensor data?



GPS Latitude	40 deg 44' 30.00" N
GPS Longitude	74 deg 0' 13.20" W
GPS Altitude Ref	Above Sea Level
GPS Altitude	10.3827476 m
GPS Time Stamp	17:53:26.86
GPS Img Direction Ref	True North
GPS Img Direction	171.3049327

- Provides a weak signal, but we want *pixel-accurate* localization

Location recognition

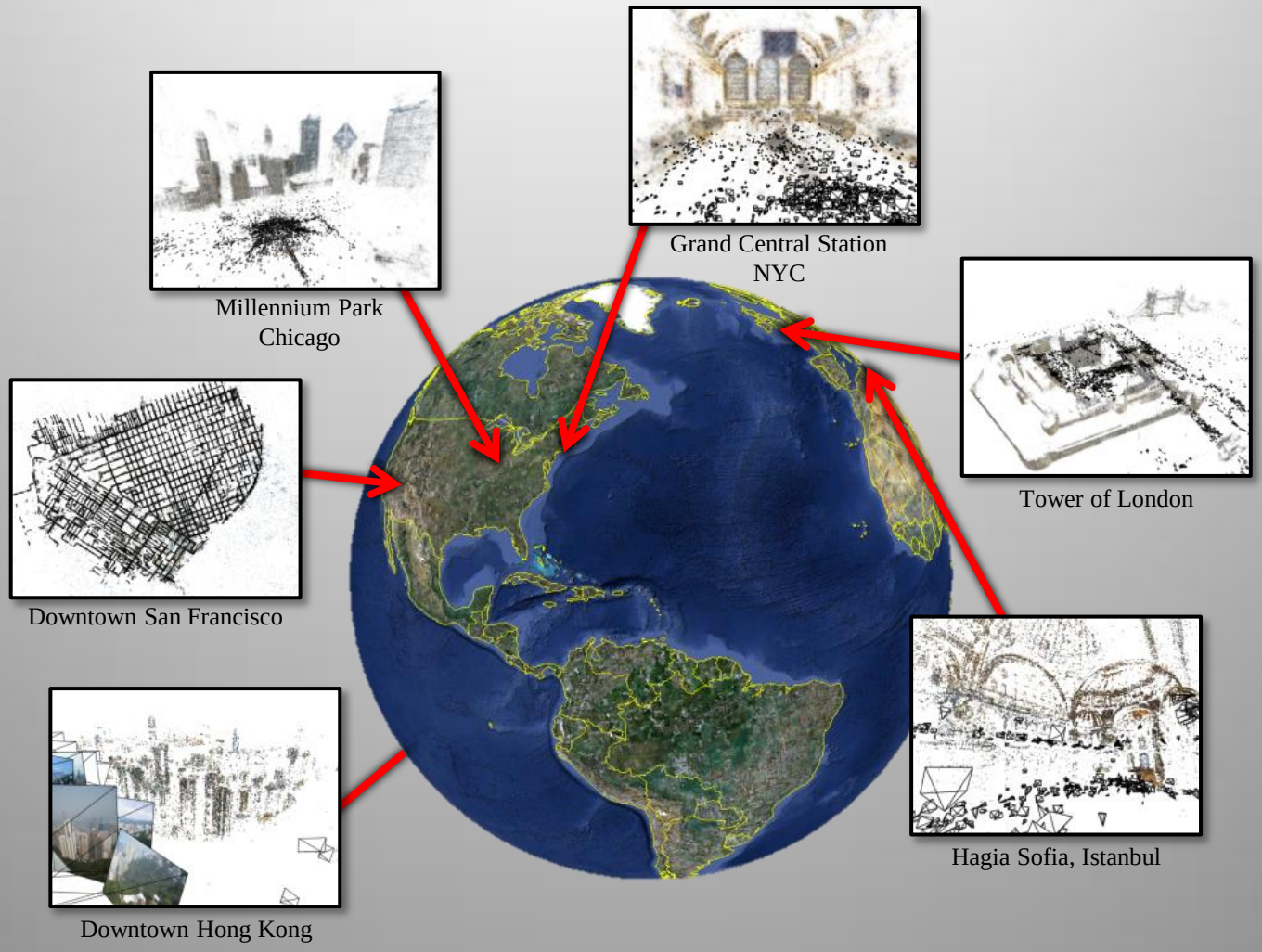
Image-based

[Schindler, Brown, Szeliski '06]
[Hays & Efros '08]
[Kalogerakis *et al.* '09]
[Li, Crandall, Huttenlocher '09]
[Knopp, Sivic, Pajdla, '10]

Geometry-based

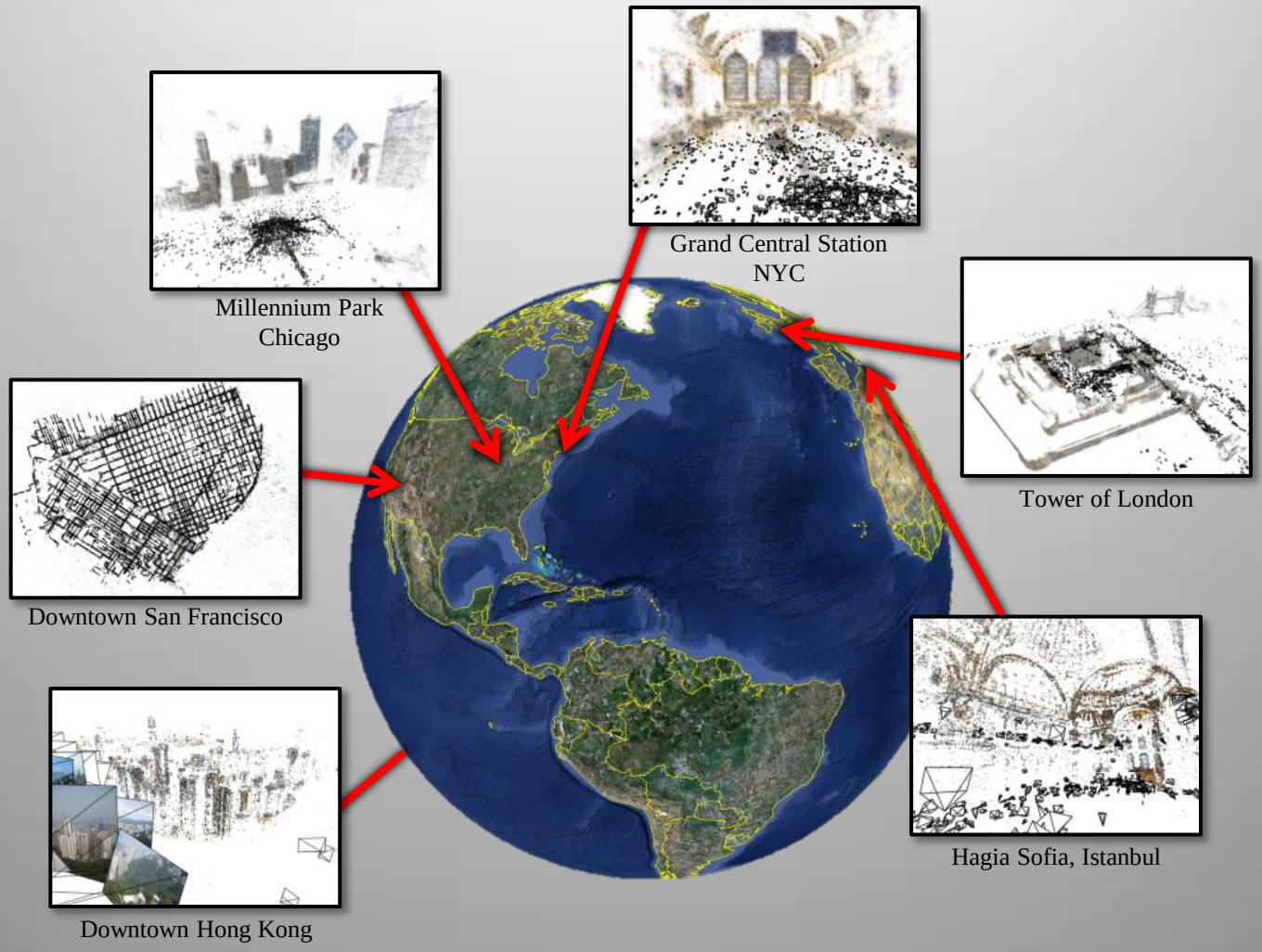
[Li, *et al.* '10]
[Sattler & Leibe '11]
[Lim *et al.*, '11]
[Li, Snavely, Huttenlocher, Fua '12]

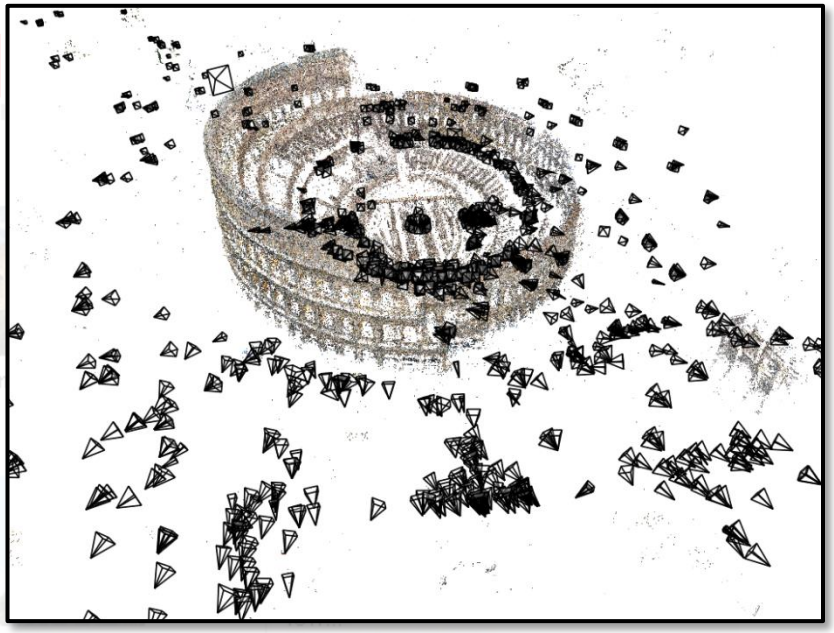
A Database of 3D Geometry



Demo

A Database of 3D Geometry





[Snavely, Seitz, Szeliski, 2006]

Dubrovnik, Croatia

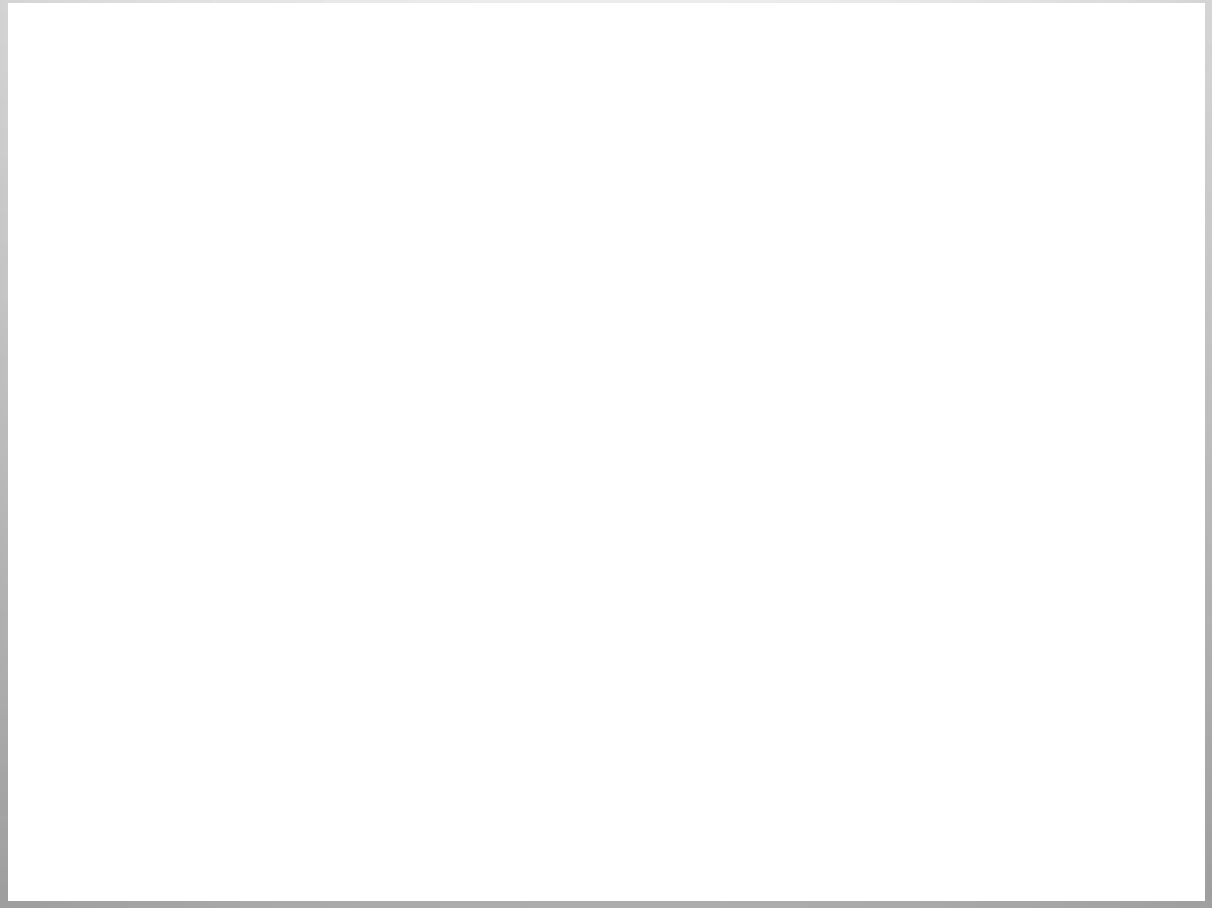
[*Building Rome in a Day*, Agarwal, Snavely, Simon, Seitz, Szeliski, ICCV 2009]

[Crandall, Backstrom, Huttenlocher, and Kleinberg. WWW09]

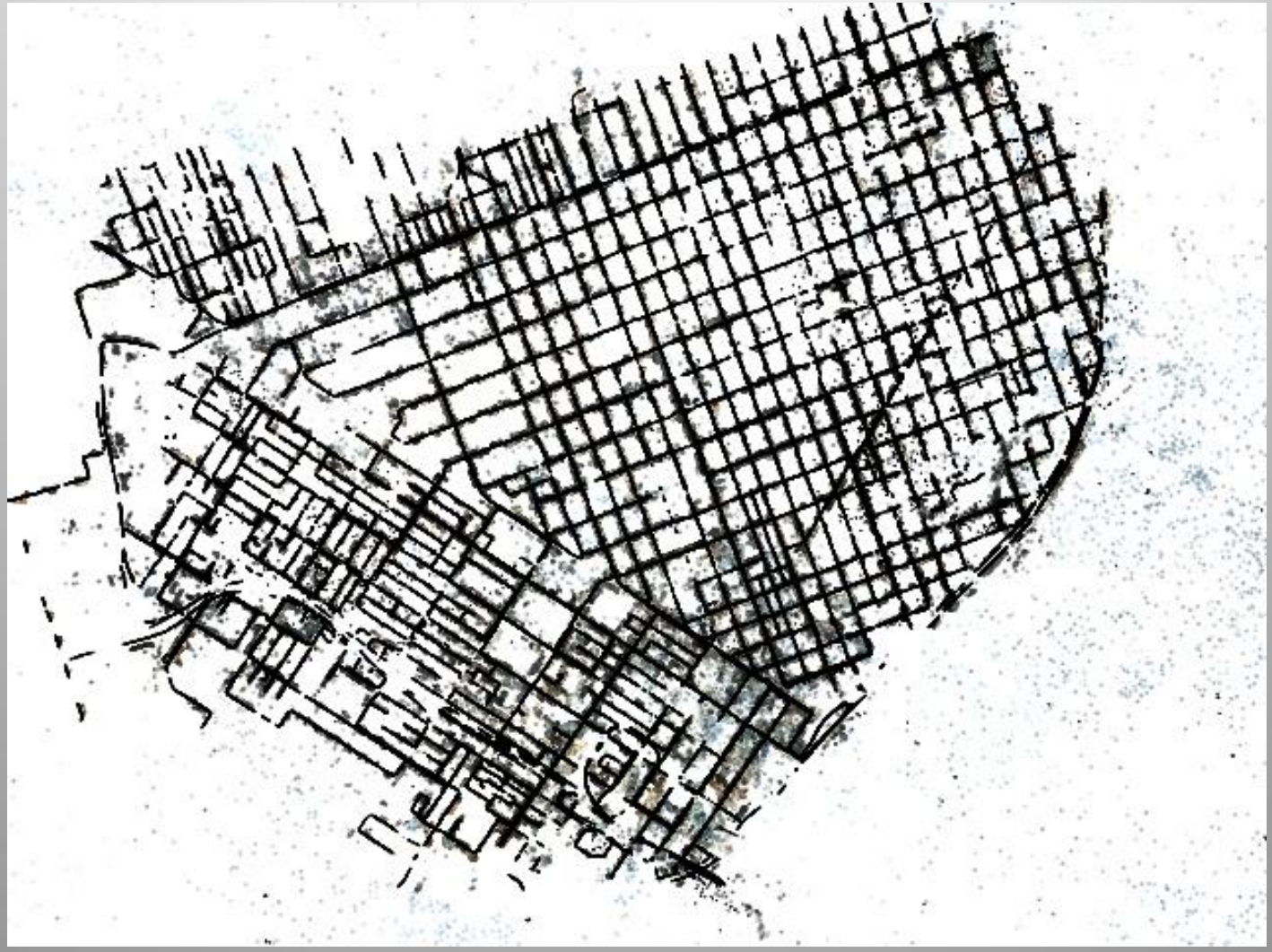
Status was last updated: 03/21/2012 08:55:50.
844 out of 10000 landmarks are processed (8.44%)

Landmark ID[1]	Tags (click to hide)	Cleaned[2]	# components	Component 1	Component 2	Compone 3
0000	paris eiffeltower eiffel toureiffel tower paris france					
0001	london trafalgarsquare trafalgar square london england uk		12	3951.0 (93.21%, 6380/6845)	1457.0 (87.62%, 92/105)	3613.0 (100.00% 26/26)
0002	london londoneye eye thames london england uk		17	4771.0 (81.68%, 6102/7471)	4771.7 (80.69%, 163/202)	4771.6 (85.71%, 48/56)
0003	london bigben westminster england london uk		5	57.0 (86.67%, 4654/5370)	57.1 (88.33%, 106/120)	57.2 (89.66%, 26/29)
0004	paris cathedral notre dame notredame church france paris		49	1008.0 (7981/--)	1008.0.0 (7974/--)	1008.1 (2101/--)
0005	paris pyramid louvre museum paris france		68	508.1 (73.04%, 4273/5850)	508.6 (90.83%, 307/338)	508.12 (99.21%, 126/127)
0006	london tatemodern tate modern bridge thames london uk england		18	1076.0 (89.01%, 2851/3203)	1076.3 (73.54%, 164/223)	1076.4 (61.27%, 106/173)
0007	rome colosseum colosseo rome italy italia roma		8	33.0 (100.00%, 49/49)	120.0 (100.00%, 38/38)	67.0 (100.00% 28/28)

NAVTEQ SF Street View Dataset

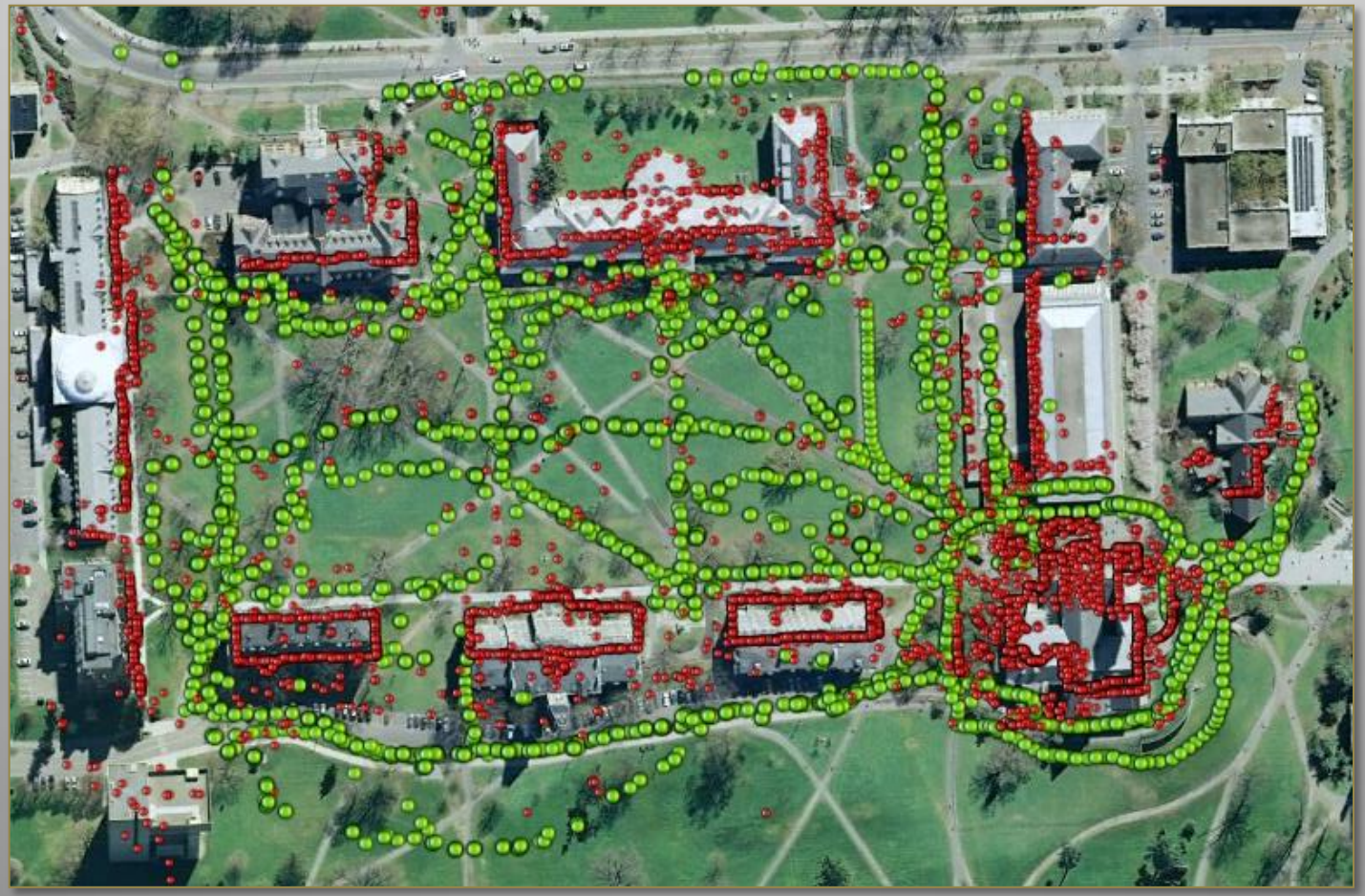


Chen et al. City-scale landmark identification on mobile devices.[CVPR 2011]



Model of San Francisco

Automatic georeferencing

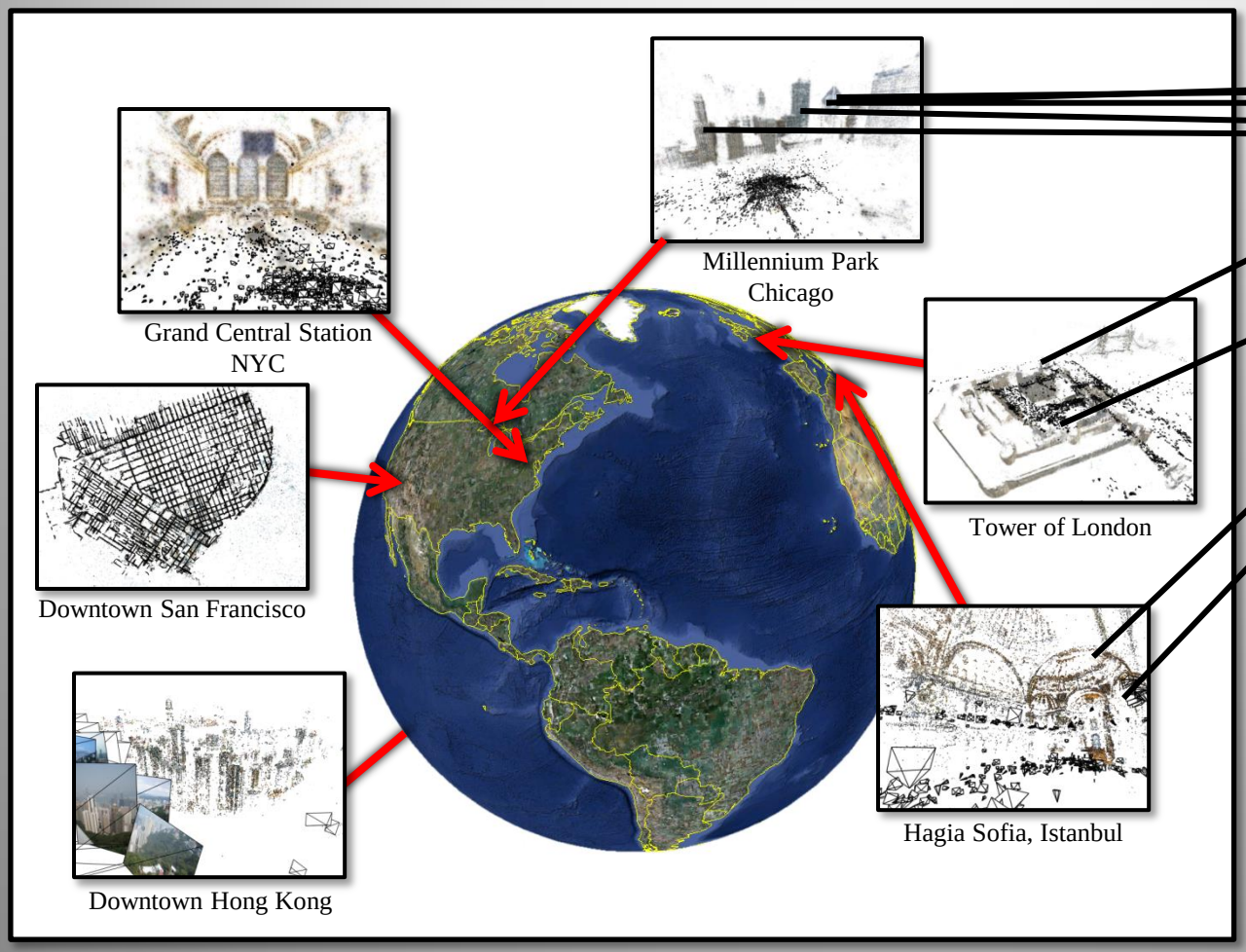


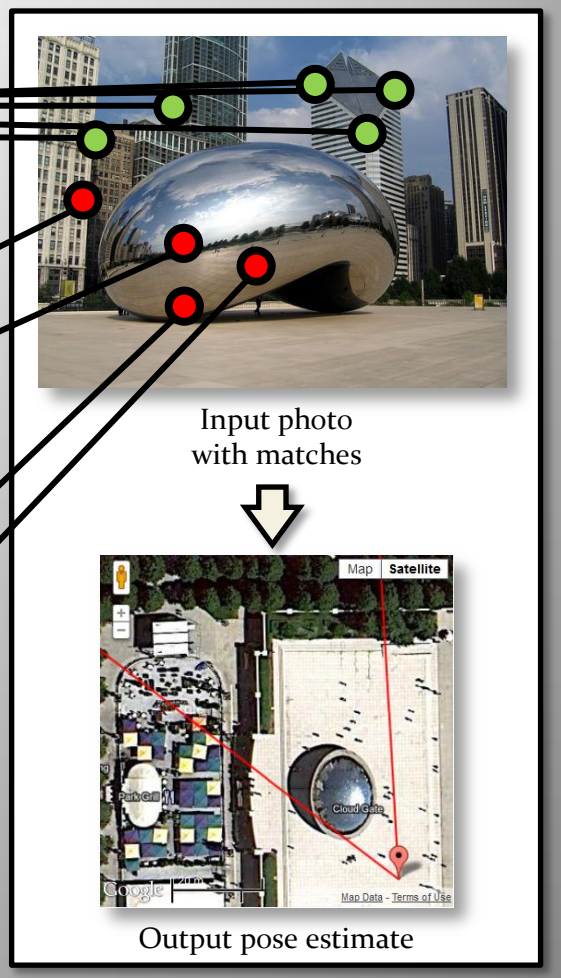
Where was this photo taken?

Utrecht, Netherlands

Havana, Cuba

World-wide Pose Estimation





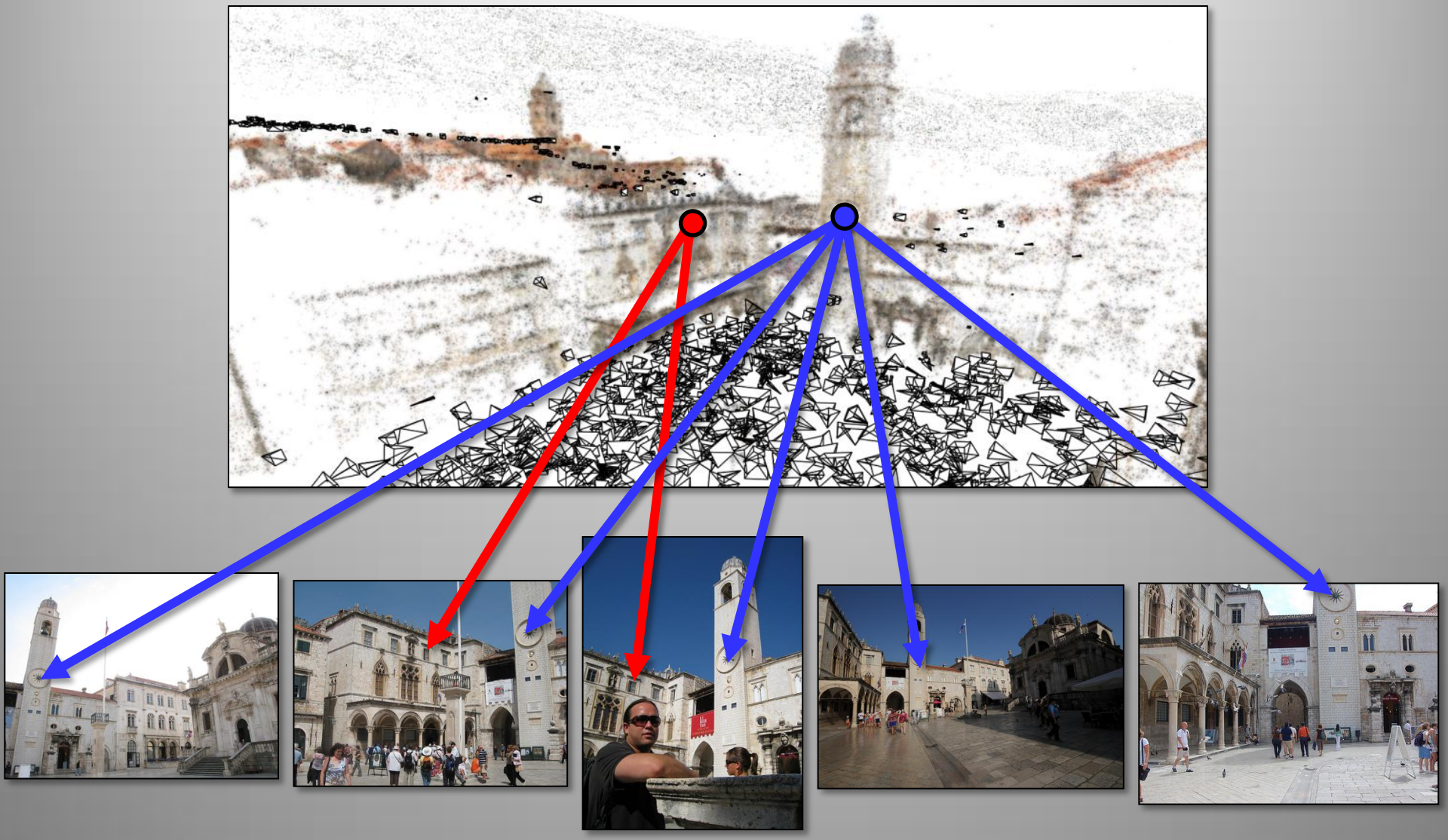
Matching becomes challenging as # of points grows very large

[Li, Snavely, Huttenlocher, Fua. ECCV 2012]

Very large search problem

- Largest model we've created:
  - About 500M 3D points
  - … from several million images
- Each 3D point has 1 or more SIFT descriptors
  - We index these using standard kd-trees
- Finding good matches at this scale is challenging
  - We have to come up with new tricks

Not all 3D points are created equal…



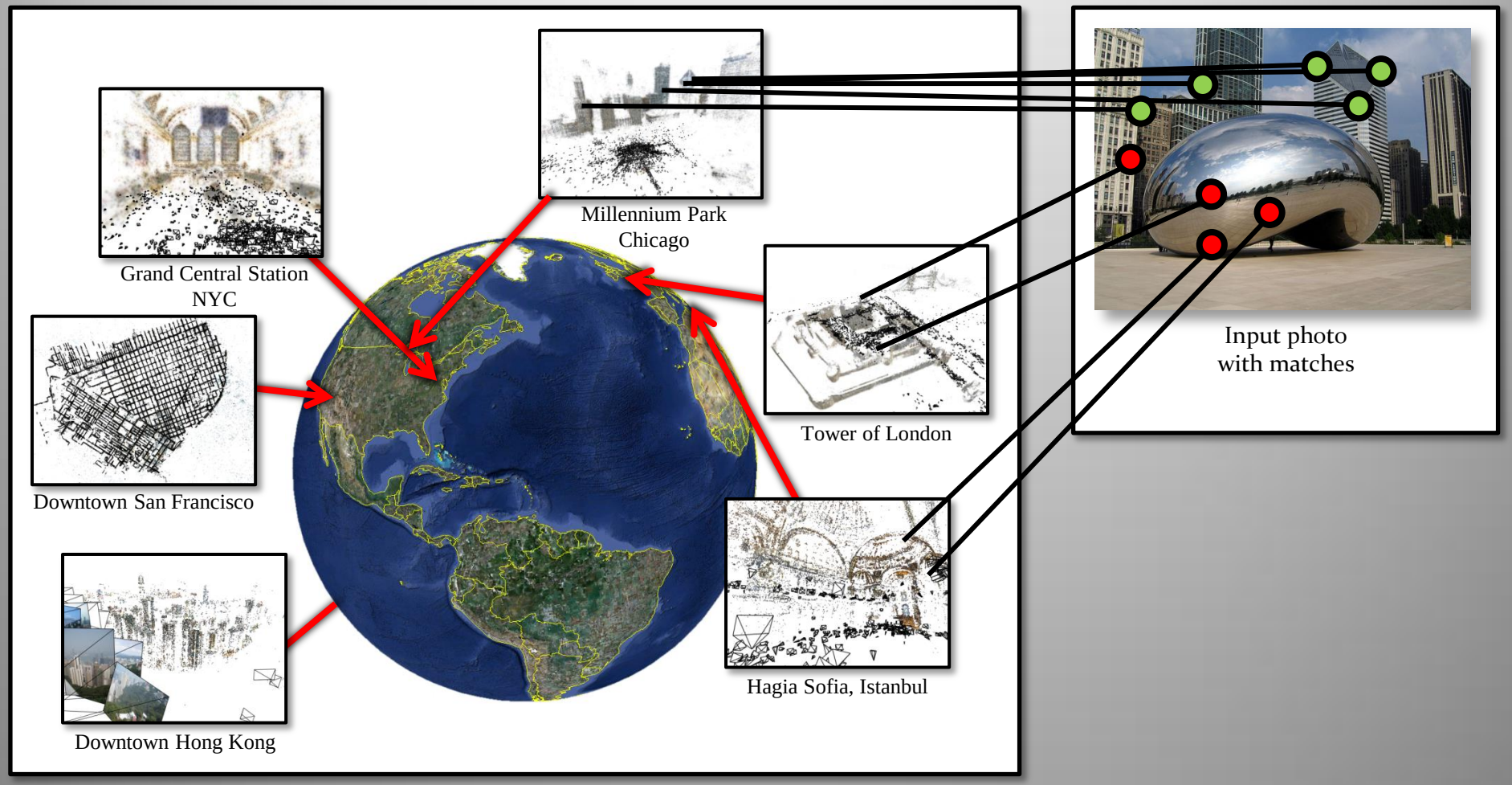
[Li, Snavely, Huttenlocher, ECCV 2010]

Point Co-occurrence

Examples of empirically frequently co-occurring triplets of points

- We can use these rich statistics over point frequency and co-occurrence to make hypothesis testing much more efficient

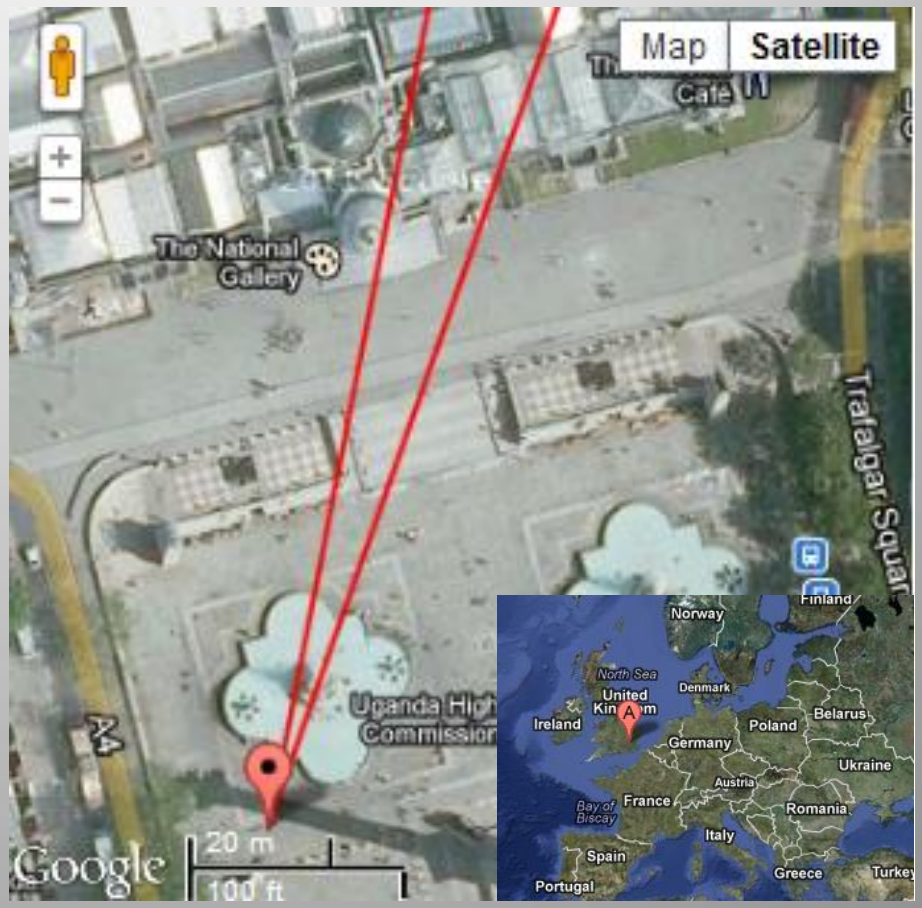
Sampling based on co-occurrence



Example result

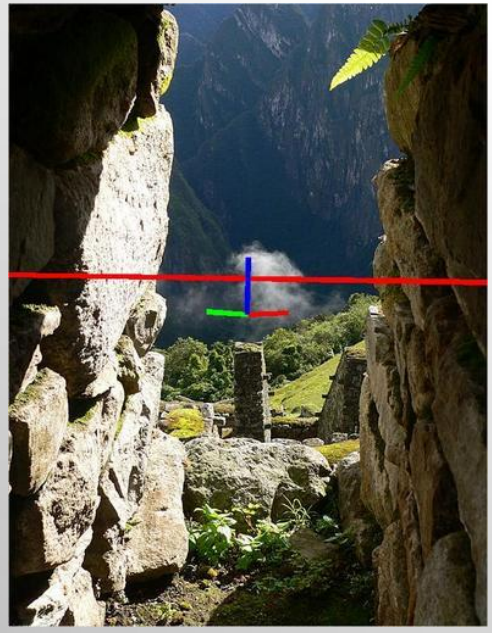
Input Photo

Estimated Camera Pose

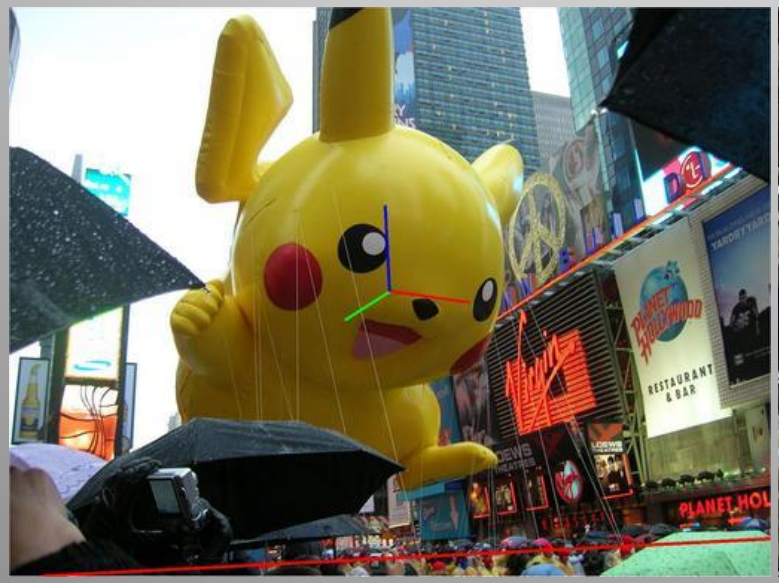


latitude: 51.5079 deg
longitude: -0.1283 deg
altitude: 0.718 m

zenith: 82.2991 deg
azimuth: -8.7291 deg
roll: -0.0391 deg
focalLength: 1610.01 px



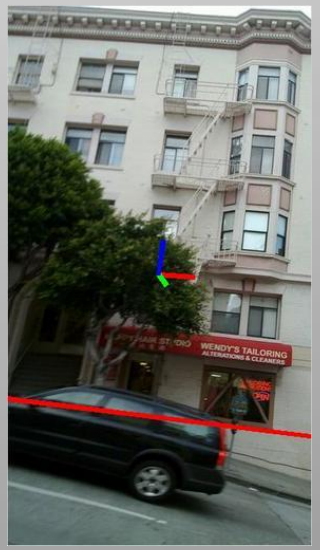
Machu Picchu, Peru



Times Square

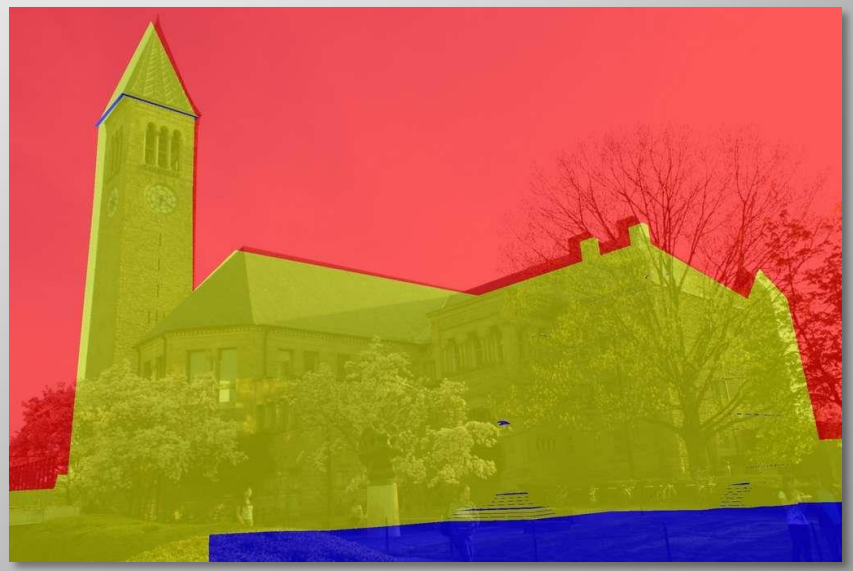
Corner of Beach and Jones (San Francisco)

Sutter St.

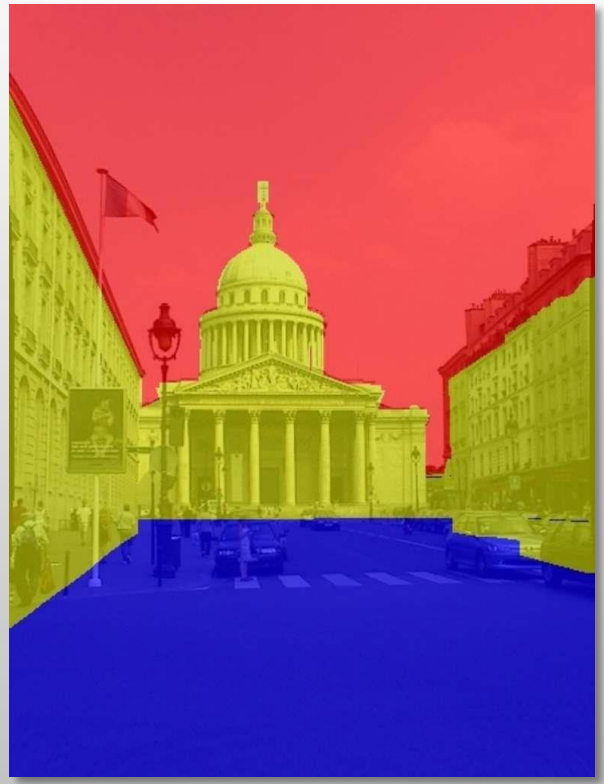


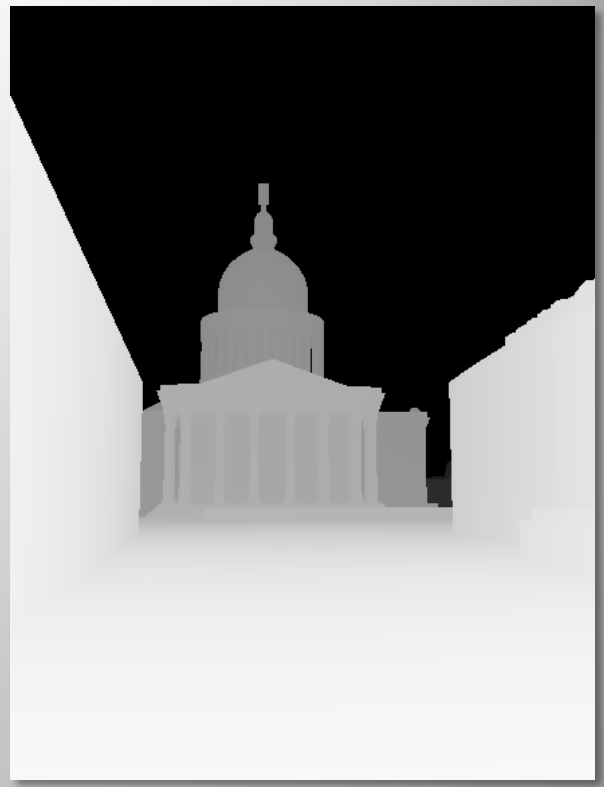
Pine St.

“Pixel-accurate” alignment



3D world model rendered from estimated viewpoint





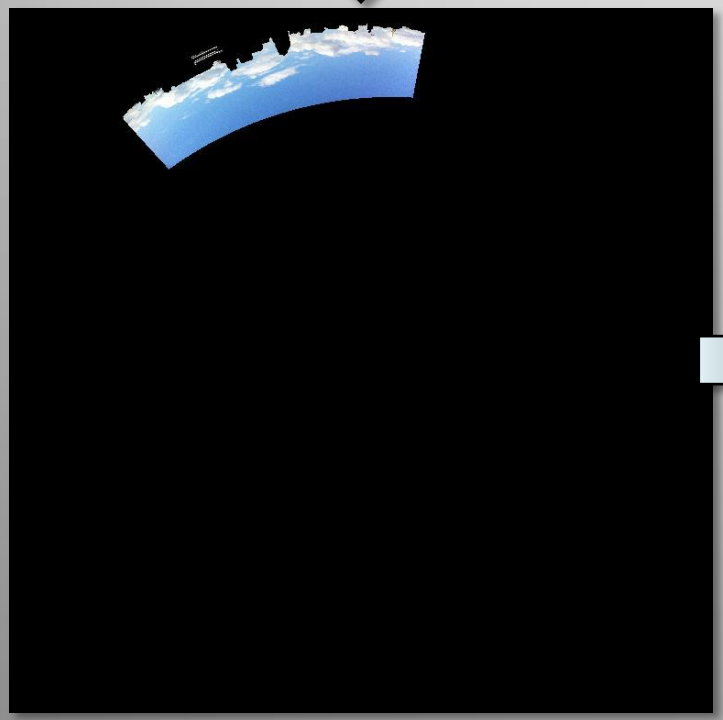
See also “Deep Photo,” Kopf *et al.* SIGGRAPH Asia 2008

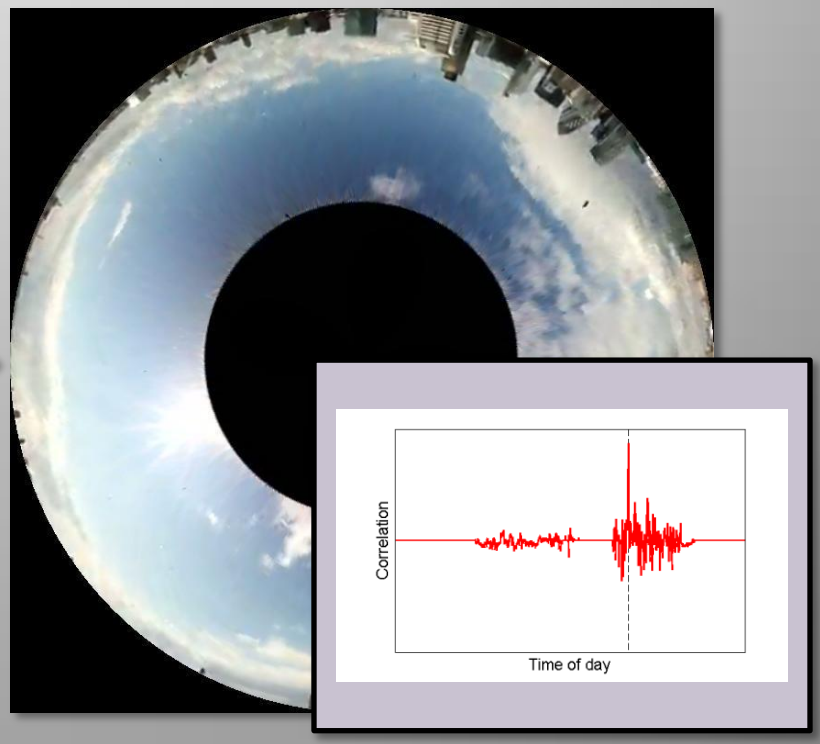
What about time?

The Monument, London

[credit: Chris Meighs-Andrews]

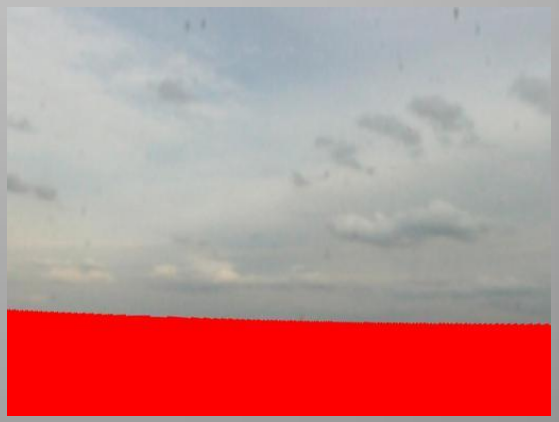

June 23, 2011 16:08 GMT
Map
Satellite



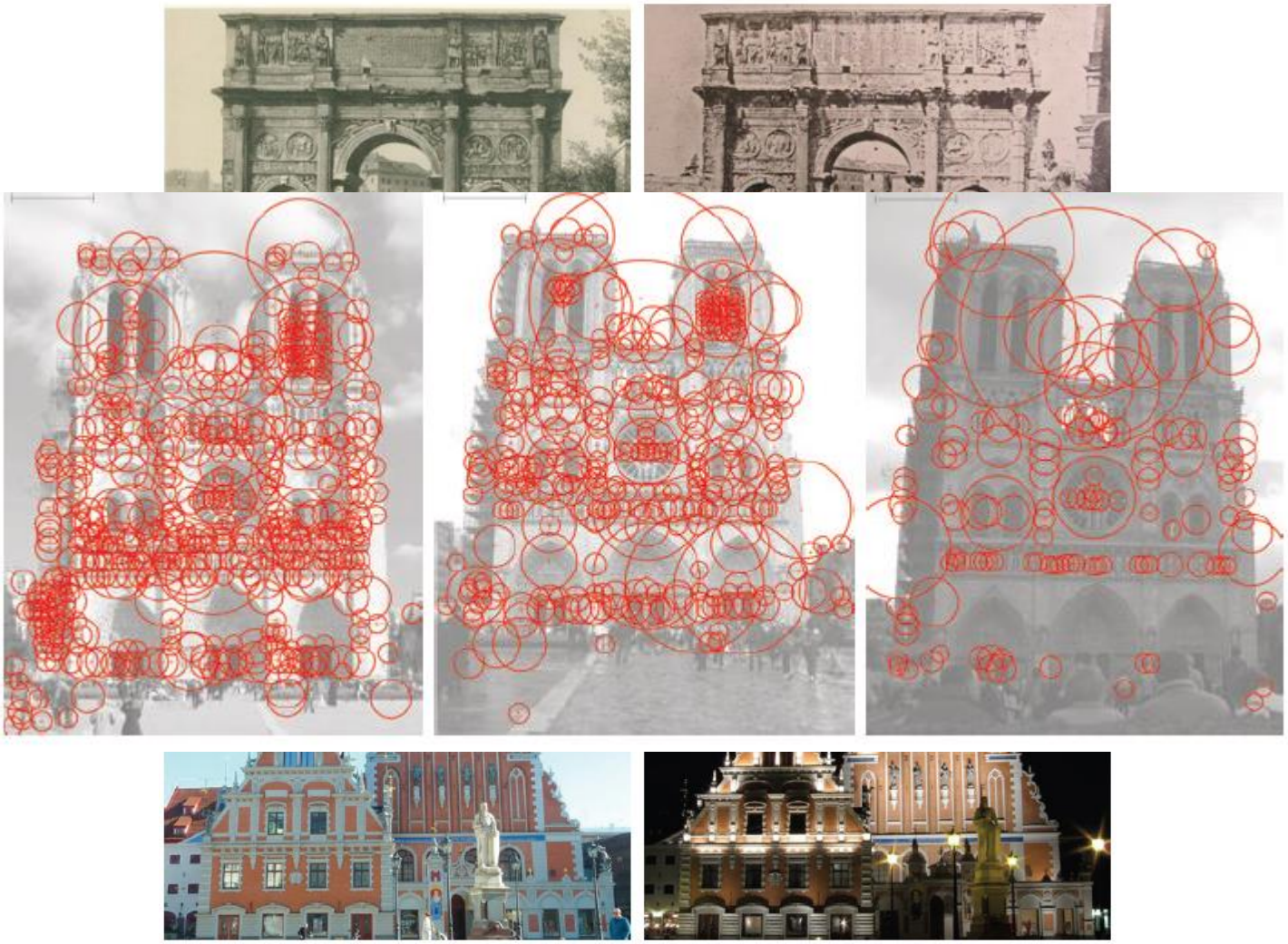

Correlation
Time of day

Input photo

Best matching webcam frame



Matching features across time



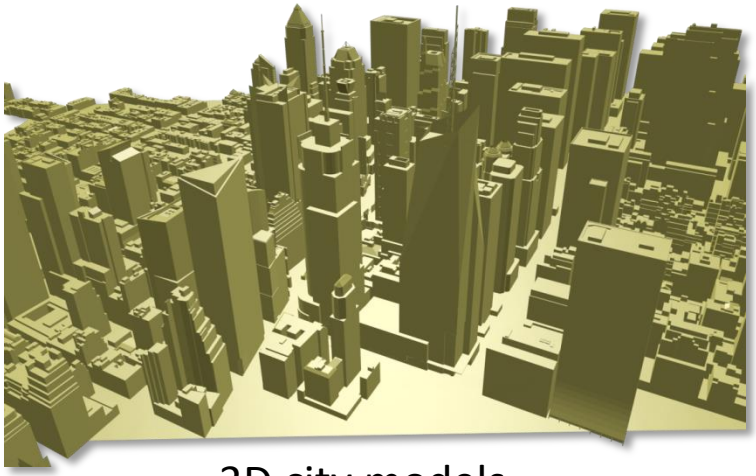
[Hauagge and Snavely, CVPR 2012]

Next steps

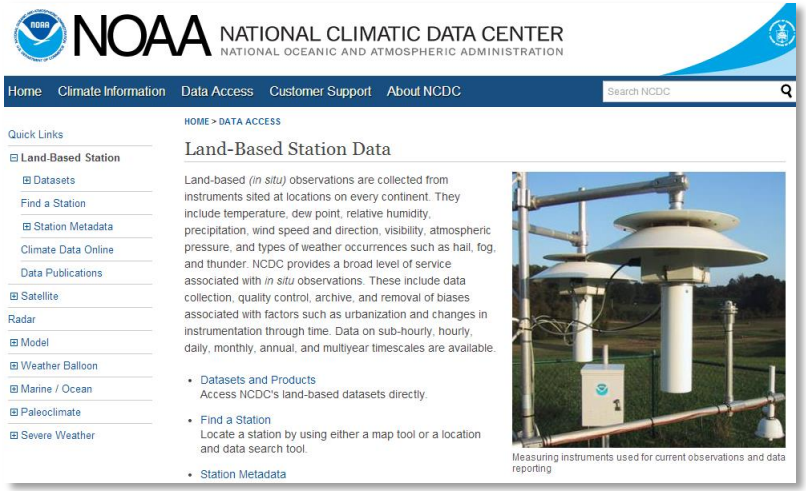
Scene appearance

Using geographic data

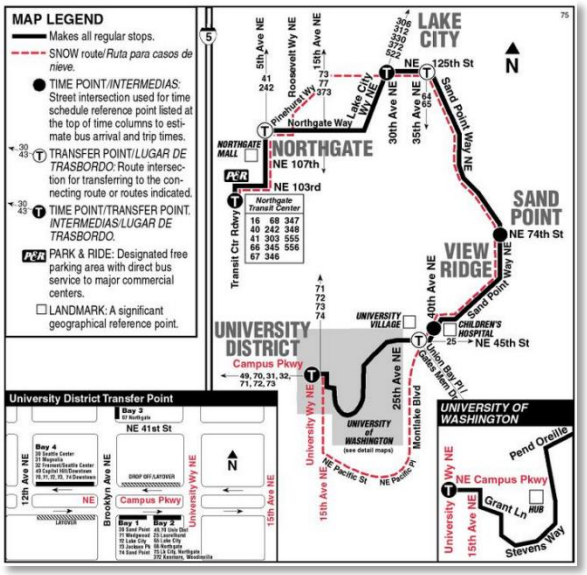
OpenStreetMap



3D city models



Weather data



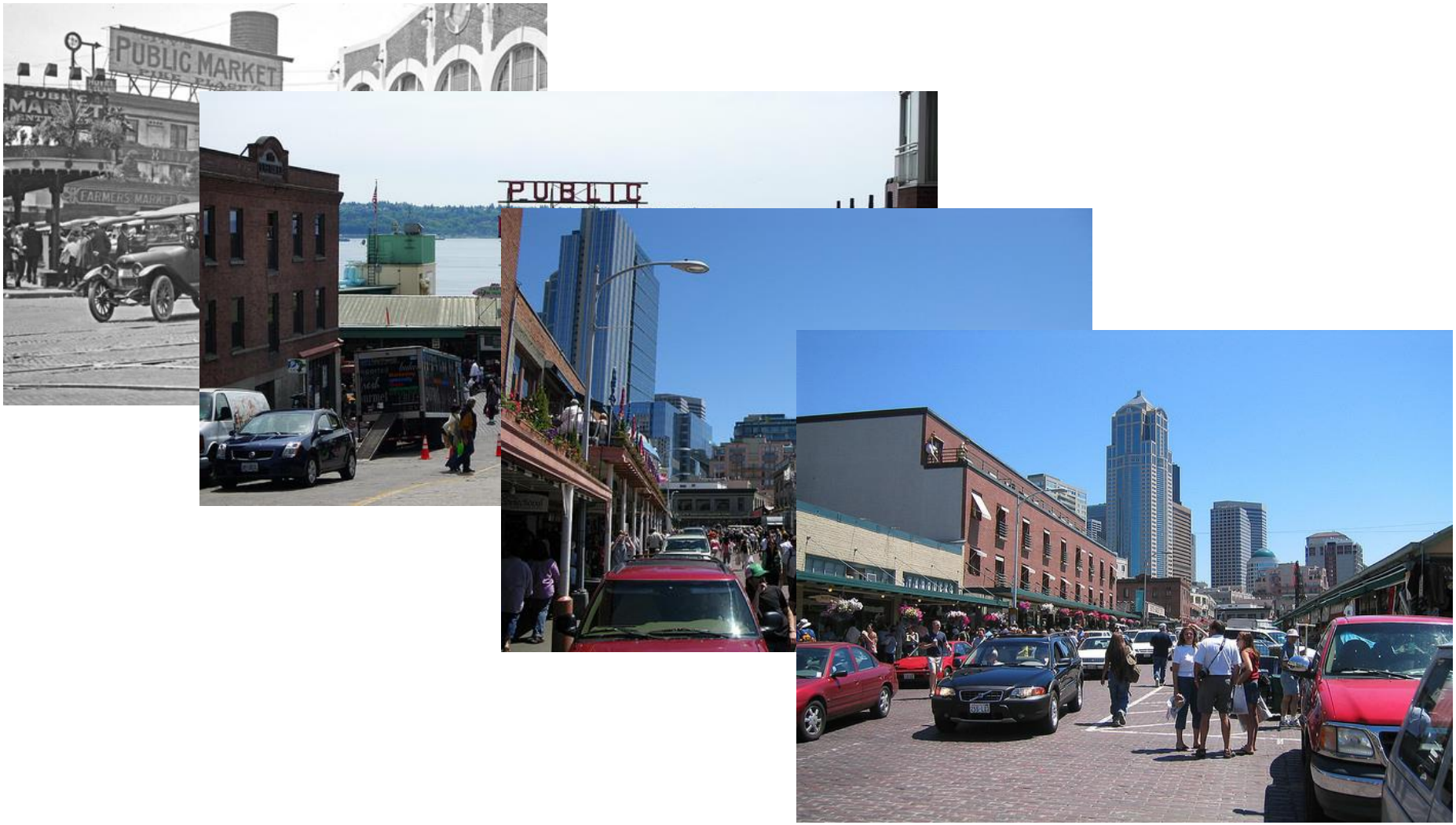
Bus schedules

Relating geographic data to vision

- Which direction is north?
- What is the shape of the buildings?
- What was the weather like?
- Where are streets?
- What is the #51 bus schedule in Rome?

Goal: Integrate images into this ecosystem of geographic data

Understanding scenes over time



OpenSurfaces

Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala
Cornell University

Statistics

	Good	All
Labeled Scenes	25,357	91,876
Whitebalanced Photos	17,839	24,771
Segmentations	70,005	103,513
Planar Segmentations	36,482	70,005
Named Materials	56,625	68,761
Named Objects	31,697	42,203
Rectified Textures	16,829	22,219

Materials

Reflectances

Textures

http://opensurfaces.cs.cornell.edu/, SIGGRAPH 2013

Demo

Summary

- Massive image collections can help reveal information about our world
- We're taking steps toward organizing this massive data source
- Lots of interesting challenges

Acknowledgements

- Sean Bell
- Daniel Cabrini Hauagge
- Kevin Matzen
- Andrew Owens
- Chun-Po Wang
- Kyle Wilson

- Yunpeng Li
- Dan Huttenlocher
- David Crandall
- Kavita Bala

Thank you!

More information at
http://www.cs.cornell.edu/~snavely/