

Grounding Vision in the Real World

Noah Snavely

Google Inc., Cornell University

RSS RGB-D Workshop, 2014

The Age of Exapixel Image Data



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

What can we do with all this data?



- Use images to understand the world
 - Building the ultimate world map
 - Changes in cities over time
 - Traffic patterns
 - Surprising events
 - Forensics: what happened, when?
- **Challenge:** data is extremely unstructured


flickr® from YAHOO!

Home You Organize & Create Contacts Groups

Astrometry.net

Favorite Actions Share


Newer Older



NGC 2264 / Christmas Tree cluster / Cone nebula - astrometry.net

NGC 2264 / Christmas Tree cluster / Cone nebula

(reprocessed)



Hello, this is the blind astrometry solver. Your results are:
 (RA, Dec) center: (100.227235684, 9.93850271956) degrees
 (RA, Dec) center (H.M.S, D.M.S): (06:40:54.537, +9:56:18.610)
 Orientation: 146.00 deg E of N

Pixel scale: 3.17 arcsec/pixel

Parity: Reverse ("Left-handed")

Field size: 72.02 x 52.33 arcminutes

Your field contains:

The star 15Mon

NGC 2264 / Christmas Tree cluster / Cone nebula

[View in World Wide Telescope](#)

If you would like to have other images solved, please submit them to the [astrometry group](#).

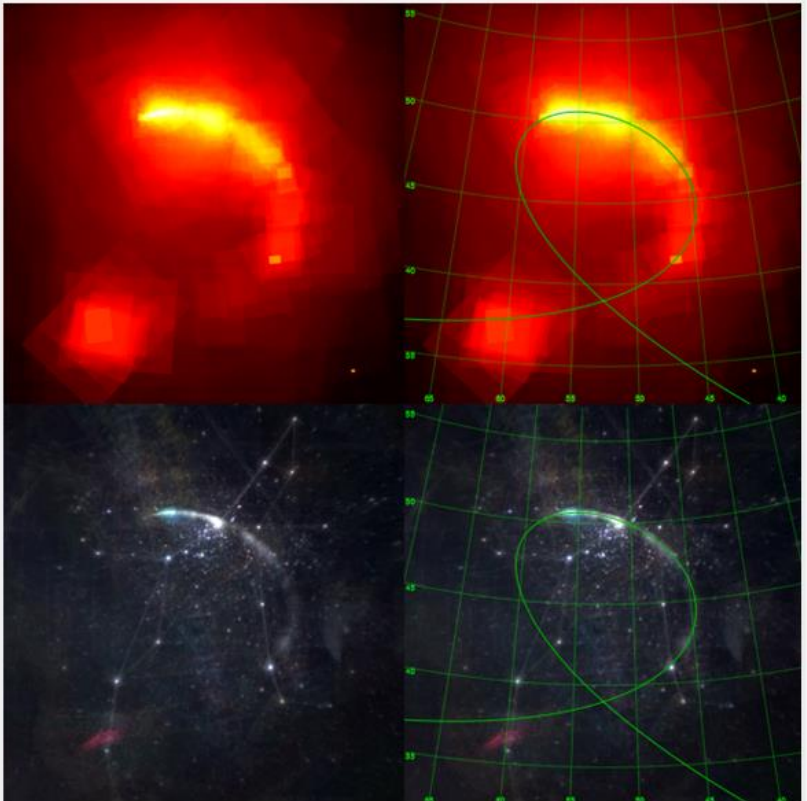
The Physics arXiv Blog

Astronomers Calculate Comet's Orbit Using Amateur Images From The Web

Amateur astrophotographs posted online represent a massive untapped resource. Now astronomers have worked out how to mine it

KFC 04/04/2011

5 COMMENTS



Lang and Hogg, Searching for comets on the World Wide Web: The orbit of 17P/Holmes from the behavior of photographers

Grounding photos in the real world

Place + Time

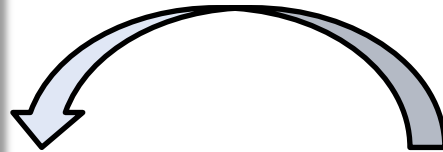
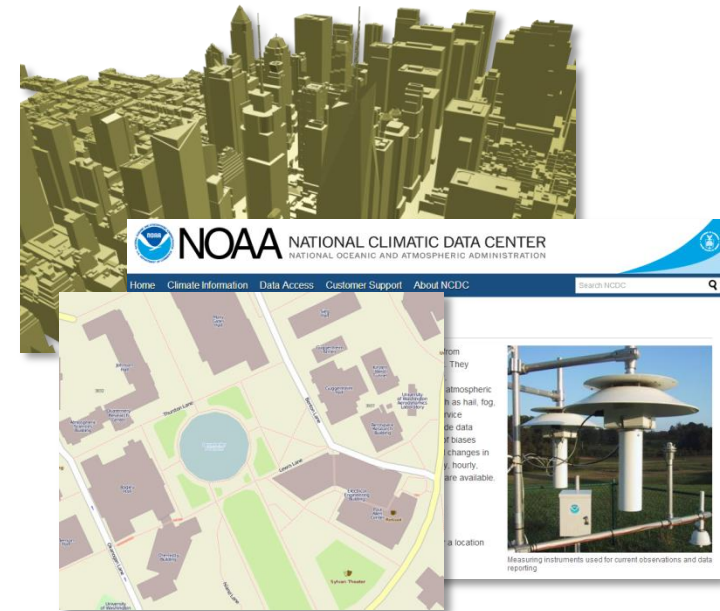


Image understanding



Prior world knowledge
(geographic / GIS data)



Photos

IM2GPS: estimating geographic information from a single image

James Hays and Alexei A. Efros
Carnegie Mellon University

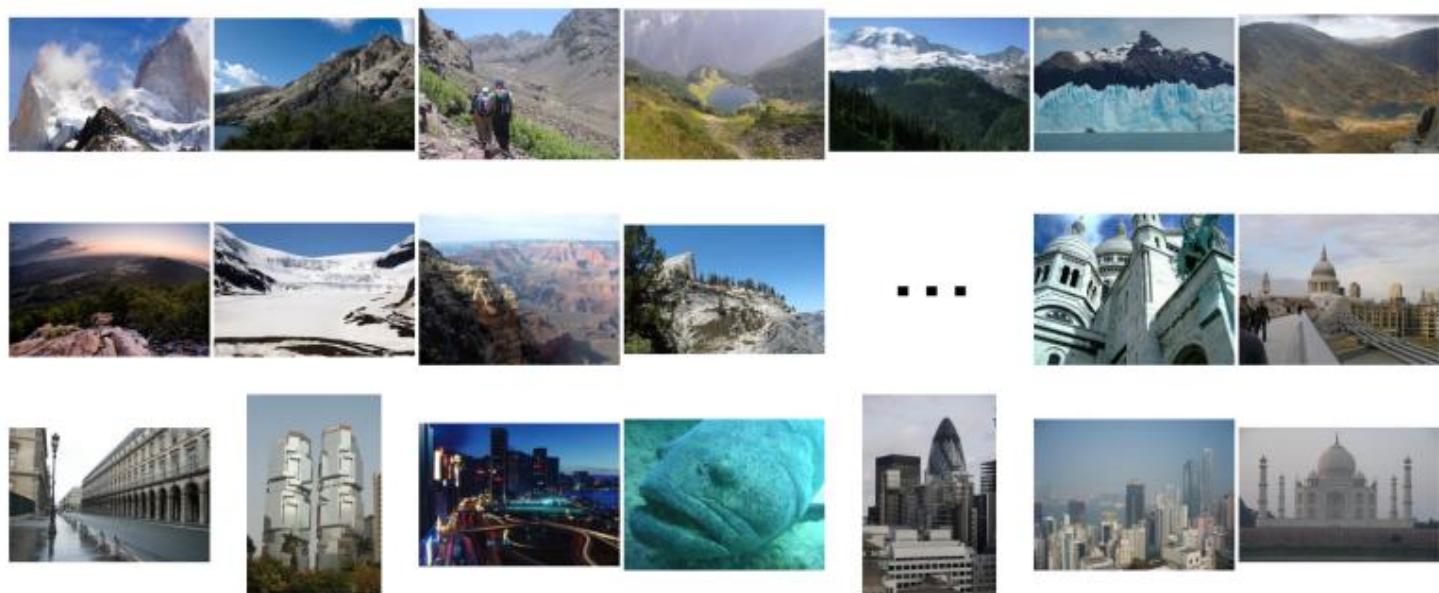


Figure 9. In scanline order, the test cases with the largest and smallest estimated elevation gradient.

Deep Photo: Model-Based Photograph Enhancement and Viewing

[Johannes Kopf](#)

University of Konstanz

[Boris Neubert](#)

University of Konstanz

[Billy Chen](#)

Microsoft

[Michael F. Cohen](#)

Microsoft Research

[Daniel Cohen-Or](#)

Tel Aviv University

[Oliver Deussen](#)

University of Konstanz

[Matt Uyttendaele](#)

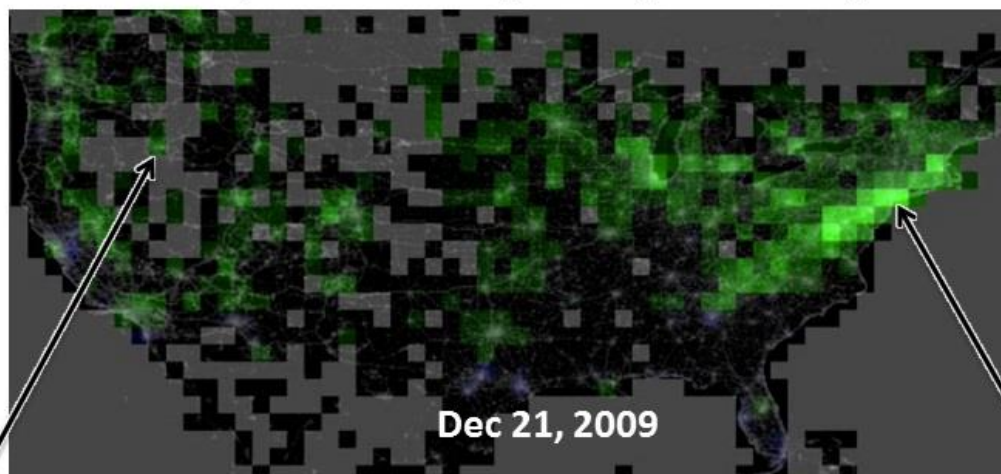
Microsoft Research

[Dani Lischinski](#)

The Hebrew University



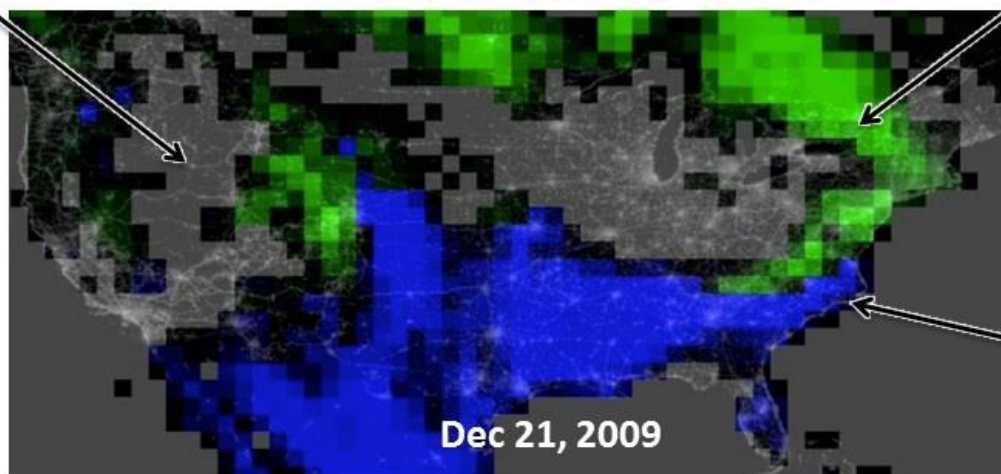
Map estimated by Flickr photo analysis



Missing
data
(black/gray)

Snow cover
(green)

Satellite map (1 degree geo bins)



No snow cover
(blue)

Snow cover estimation from Flickr photos

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?
- **Thesis:** Geometry is key to this problem

Sensor data alone is too noisy



By N. Pochanuk

Place

Location recognition



Image-based

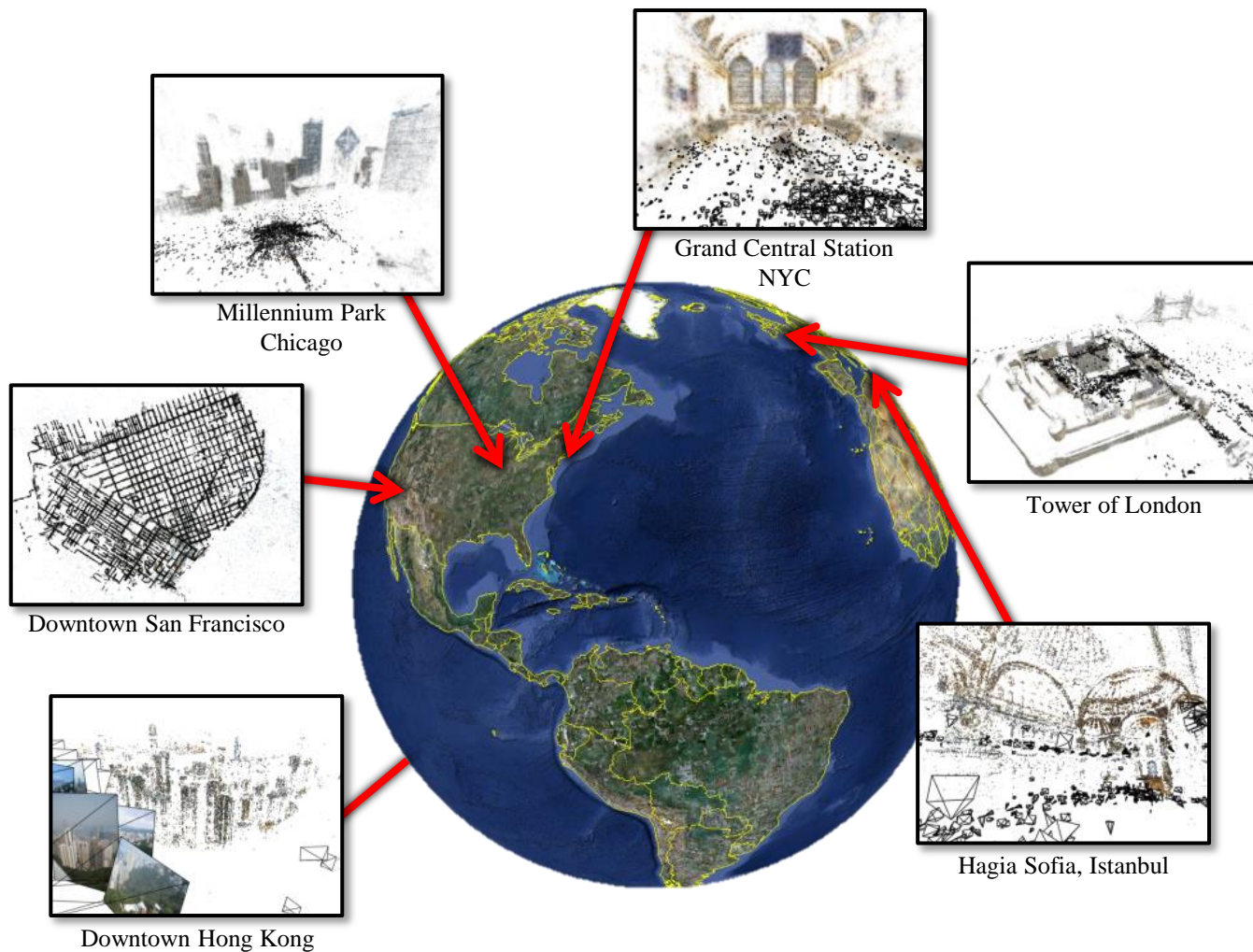
[Schindler, Brown, Szeliski '06]
[Hays & Efros '08]
[Kalogerakis *et al.* '09]
[Li, Crandall, Huttenlocher '09]
[Knopp, Sivic, Pajdla, '10]



Geometry-based

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

A Database of 3D Geometry



Search

[Photos](#)

[Groups](#)

[People](#)

colosseum rome

[SEARCH](#)

[Advanced Search](#)
[Search by Camera](#)

☒ Full text ☐ Tags only



From [Jeremy...](#)

From [Jeremy...](#)

From [Jeremy...](#)



From [Jeremy...](#)



From [Jeremy...](#)



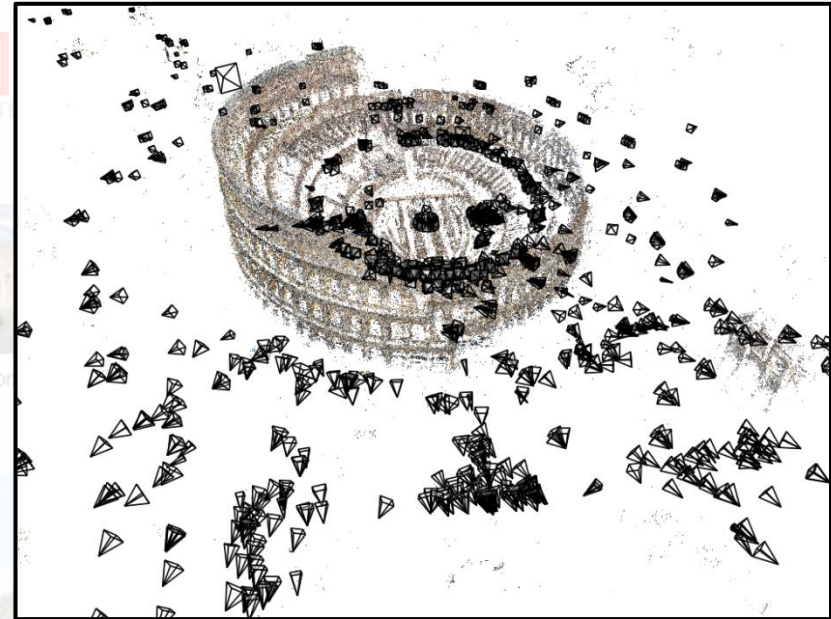
From [Jeremy...](#)



From [Jeremy...](#)

um and rome.

Show: Details • T



[www.florenceart.it](#)

[Colosseum Tickets](#)

Save on Colosseum Tickets. Most 10-15% Less than other Sites.

[www.TicketsPlus.com](#)

[Hotels near the Colosseum](#)

Save up to 75% on Italy hotels. Pay at check-in. No booking fees.

[www.priceline-europe.com](#)

A collage of 25 Polaroid-style photographs of Dubrovnik, Croatia. The photos are arranged in a dense, overlapping layout. Key features include: the Dubrovnik City Walls and towers; the red-tiled roofs of the old town; the Rector's Palace; the Dubrovnik Cathedral; the Dubrovnik Harbor with boats; and various street scenes and views of the city from different angles. The photos are set against a light gray background.



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



[Crandall, Backstrom, Huttenlocher, and Kleinberg. WWWo9]

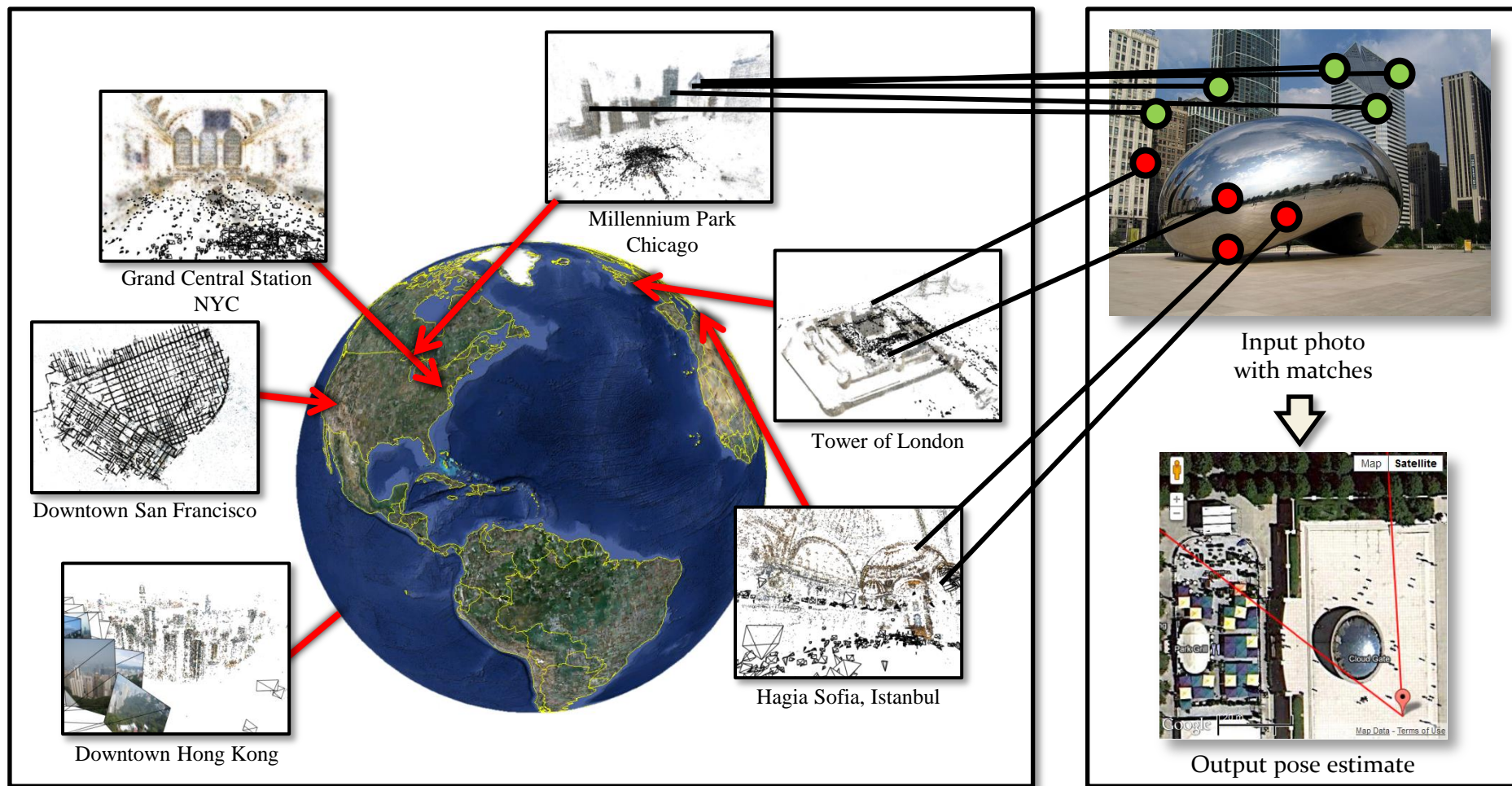
NAVTEQ SF Street View Dataset

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



Model of San Francisco

World-wide Pose Estimation



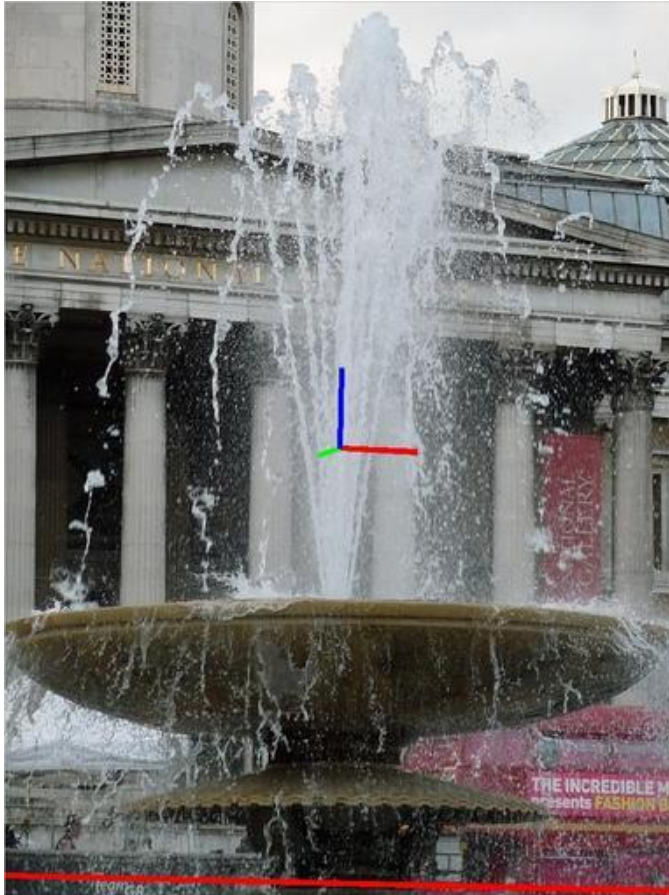
Matching becomes challenging as # of points grows very large

Very large search problem

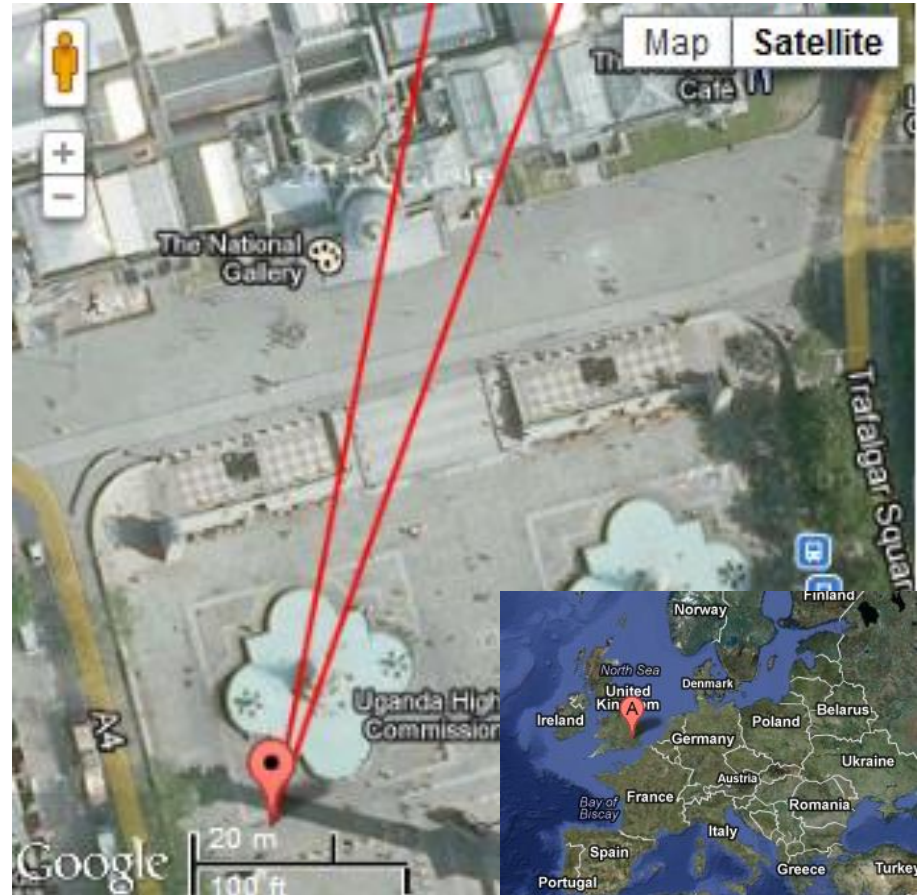
- Largest model we've created:
 - About 500M 3D points
 - ... from several million images
- Each 3D point has 1 or more SIFT descriptors
 - We index these using standard kd-trees
- Finding good matches at this scale is challenging
 - We have some new tricks to estimate location quickly and robustly

Example results

Input Photo

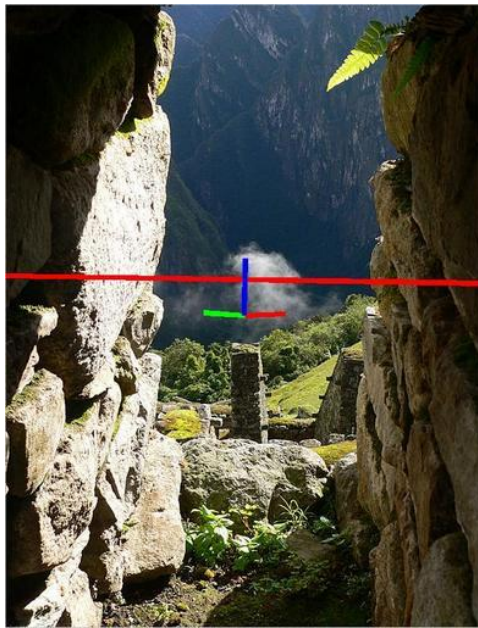


Estimated Camera Pose

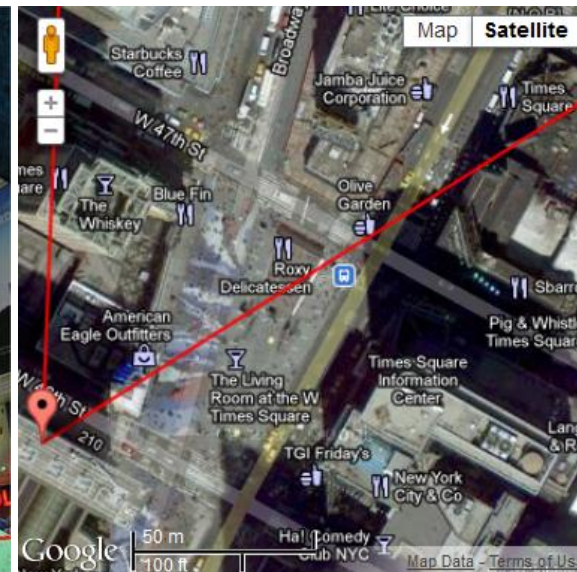


latitude: 51.5079 deg
longitude: -0.1283 deg
altitude: 0.718 m

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



Machu Picchu, Peru



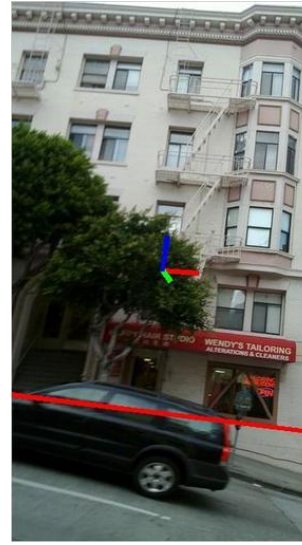
Times Square



Corner of Beach and Jones (San Francisco)



Sutter St.

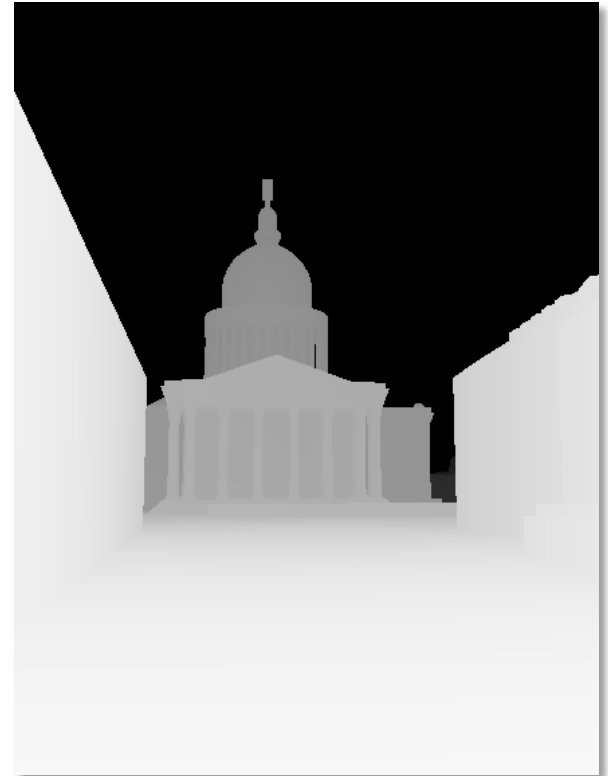
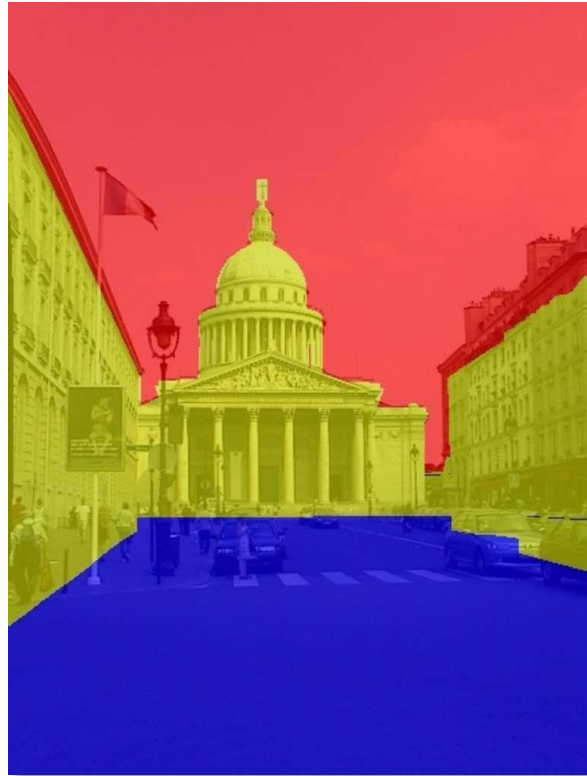


Pine St.

“Pixel-accurate” alignment



3D world model rendered from
estimated viewpoint



Creates a “stage” for the action to take place in

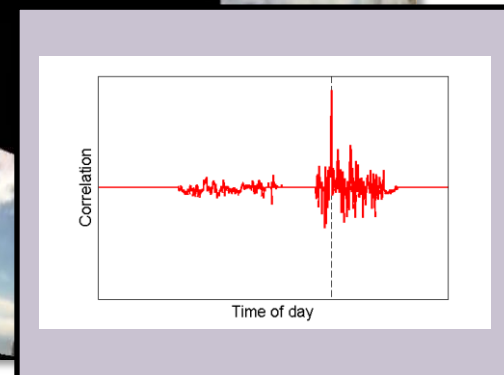
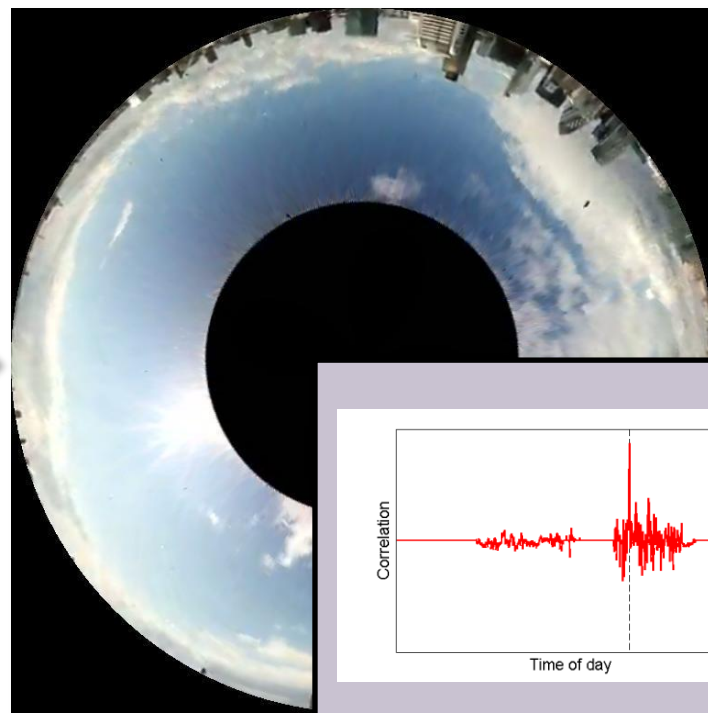
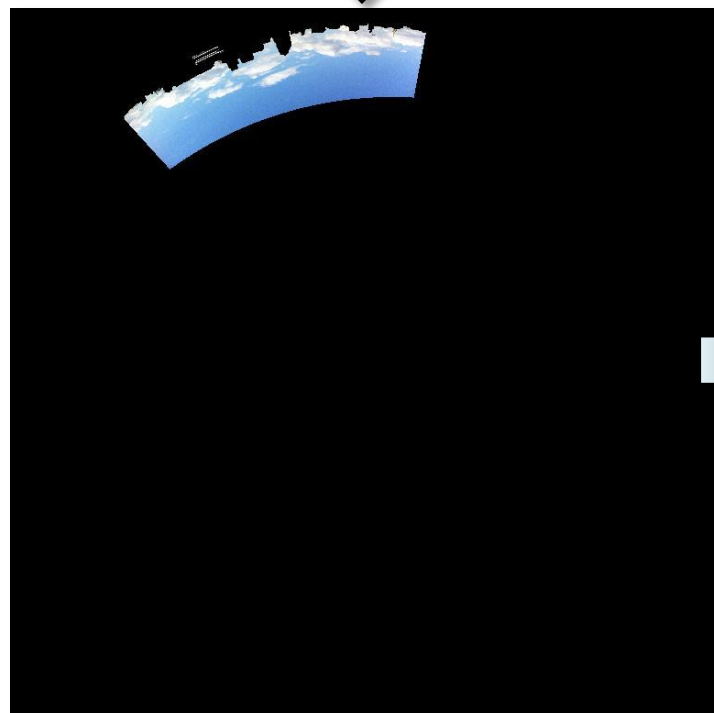
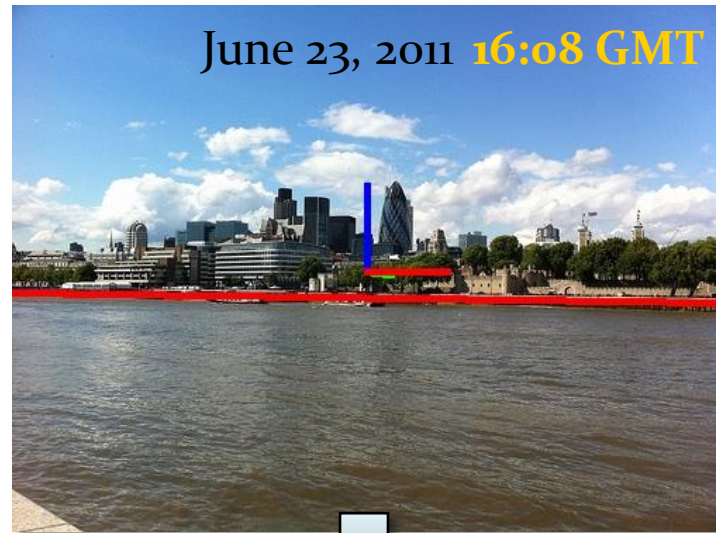
Time



The Monument, London



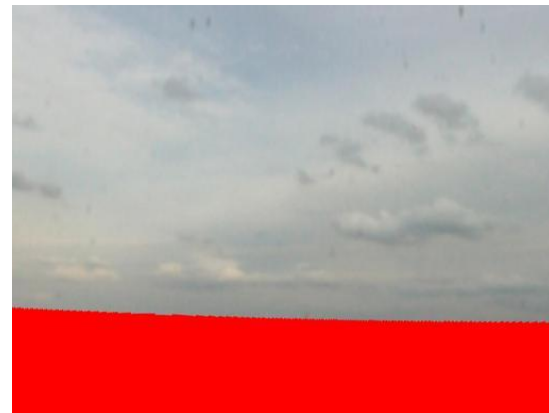
[credit: Chris Meighs-Andrews]



Input photo



Best matching webcam
frame (cropped)



[joint work with
Chun-Po Wang]

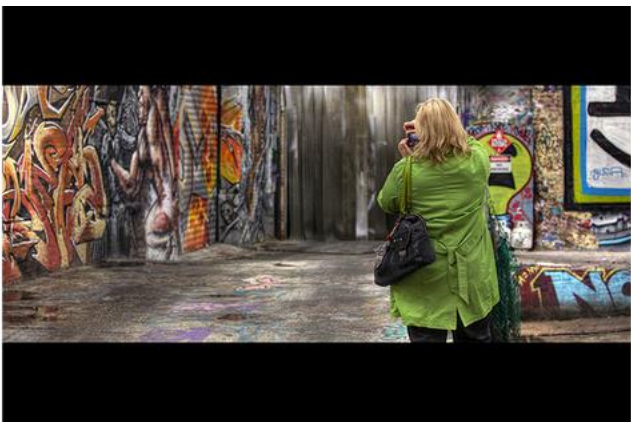


For images that contain clouds and within $\sim 0.75\text{km}$ of the webcam, we can identify correct time stamps (± 5 minutes) $\sim 40\%$ of the time

Scenes are dynamic



Times Square



5pointz, Brooklyn

Graffiti Archaeology

GRAFFITI ARCHAEOLOGY

[about](#) [credits](#) [links](#) [news](#) [faq](#) [submit](#) [donate](#)

Locations & Walls.

▼ bluxome:

eastA
eastB
eastC

► undercroft:

► cavern:

► harveys:

► olympia:

► 22nd:

► 19th:

► candle:

► florida:

► ghost:

► cove:

► belmont:

► 25th:

► williams:

Current Layer: 3 october 2004. Photos by Cassidy Curtis



4D Cities



[Frank Dellaert, Grant Schindler, et al.]

Scene Chronology

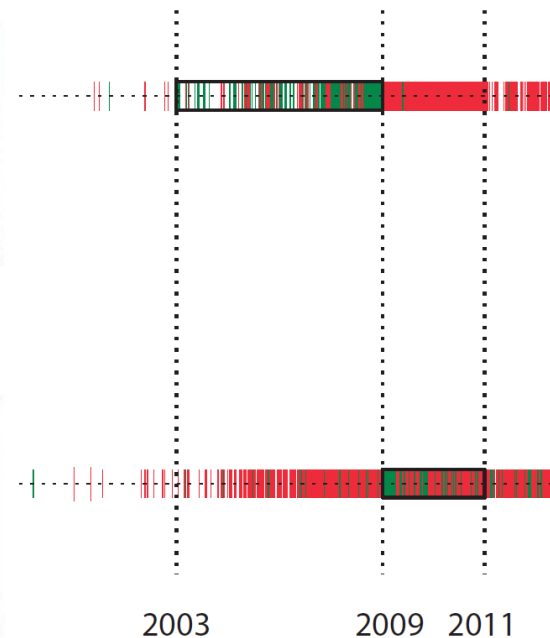
Step 1: Download photos from Flickr

Step 2: Reconstruct a single 3D model with all times mixed up together

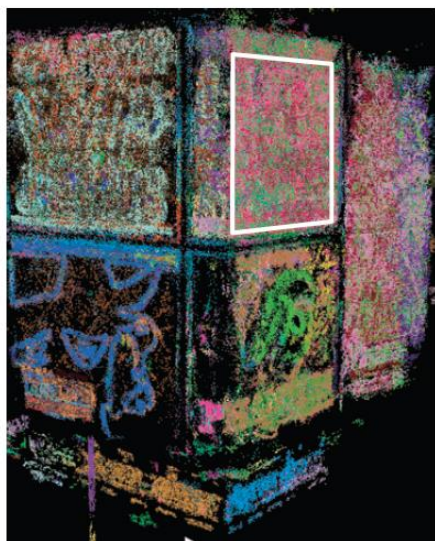
Step 3: Recover the chronology of the scene



Single 3D Model (from ~100,000 images)



Per-Point Time Observations



Plane-Time point clustering



Reconstructed structure across time

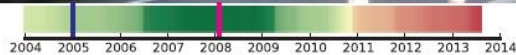
Video

Re-time-stamping

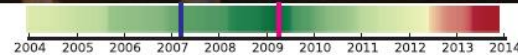
March 2008 - Late 2007



January 2005 - Early/Mid 2007



April 2007 - 2008/2009



Blue: original timestamp

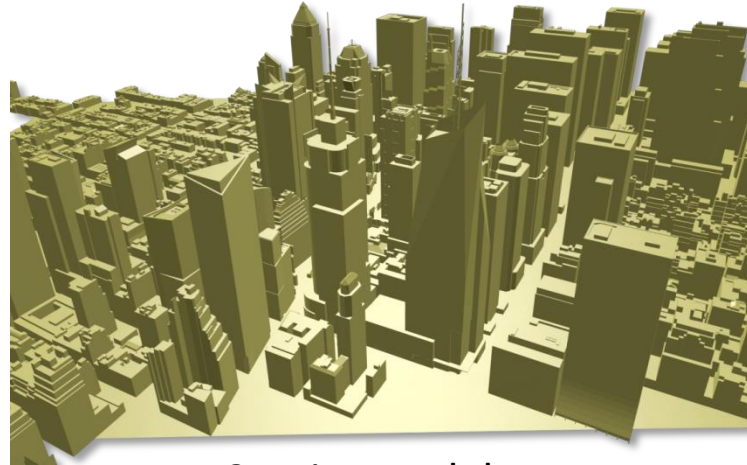
Red: our predicted timestamp

Geography

Grounding vision in the world



OpenStreetMap



3D city models

NOAA NATIONAL CLIMATIC DATA CENTER
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Home Climate Information Data Access Customer Support About NCDC Search NCDC

Quick Links

- Land-Based Station
 - Datasets
 - Find a Station
 - Station Metadata
 - Climate Data Online
 - Data Publications
- Satellite
- Radar
- Model
- Weather Balloon
- Marine / Ocean
- Paleoclimate
- Severe Weather

HOME > DATA ACCESS

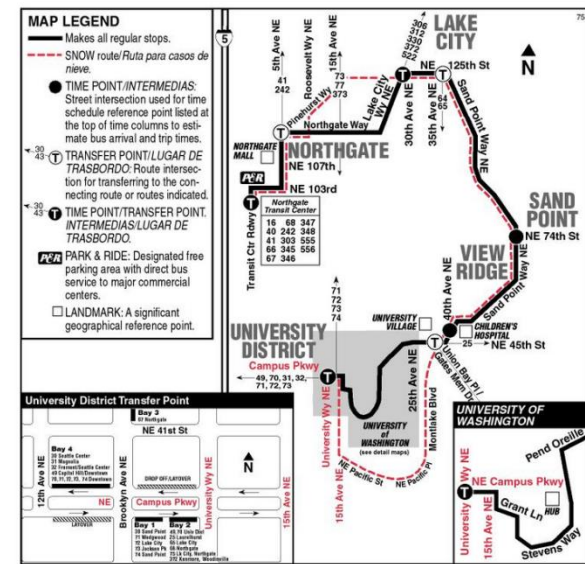
Land-Based Station Data

Land-based (*in situ*) observations are collected from instruments sited at locations on every continent. They include temperature, dew point, relative humidity, precipitation, wind speed and direction, visibility, atmospheric pressure, and types of weather occurrences such as hail, fog, and thunder. NCDC provides a broad level of service associated with *in situ* observations. These include data collection, quality control, archive, and removal of biases associated with factors such as urbanization and changes in instrumentation through time. Data on sub-hourly, hourly, daily, monthly, annual, and multiyear timescales are available.

- Datasets and Products
 - Access NCDC's land-based datasets directly.
- Find a Station
 - Locate a station by using either a map tool or a location and data search tool.
- Station Metadata

Measuring instruments used for current observations and data reporting

Weather data



Bus schedules

Search & Browse Datasets and Views

Name	
1.	311 Service Requests from 2010 to Present Social Services 311, 311 service requests, 2010, 2011, 2012, ... All 311 Service Requests from 2010 to present. This information is automatically updated daily.
2.	Electric Consumption by ZIP Code - 2010 Environment electricity, energy, environment, planning, power, ... 2010 electricity consumption in kWh and GJ, by ZIP code, building type, and utility company.
3.	Zip Codes Map Social Services geographic, location, map, cartography, zip, code, ... Polygon representing the boundary of the zip codes in the city.
4.	MTA Data Transportation traffic, vehicles, route, schedules, clean web Information pertaining to MTA (Metropolitan Transportation Authority of the State of New York) subways, buses, commuter rail,
5.	Restaurant Inspection Results Health restaurant inspection results, ... NYC restaurant inspection results
6.	Basic Description of Colleges and Universities Education doitt gis, geographic, location, map, cartography, ... Location of colleges and universities with basic descriptive information.
7.	SAT (College Board) 2010 School Level Results Education lifelong learning New York City school level College Board SAT results for the graduating seniors of 2010. Records contain 2010 College-bound
8.	Mapped In NY Companies Business jobs, tech, jobs and economic mobility Raw data which powers the Mapped In NY site at http://www.mappedinny.com/
9.	Filming Locations (Scenes from the City) Business film, movie, scene, scenes from the city List of filming locations mentioned in the book Scenes from the City
10.	2012 NYC Noise Complaints - Heat Map Other 311, 311 service requests, 2010, 2011, ... NYC Noise complaints for 2012

<https://nycopendata.socrata.com> (<https://data.sfgov.org/>, <https://data.seattle.gov/>, ...)

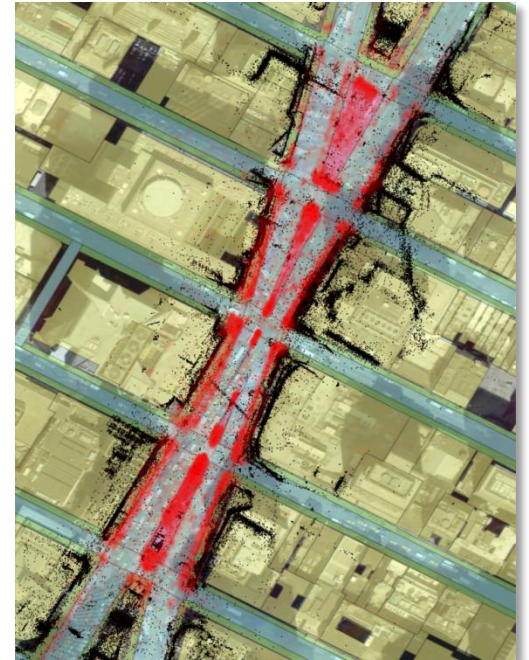
Grounding vision in the world



- 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

First steps: NYC_{3D}Cars



[Kevin Matzen and Noah Snavely, ICCV 2013]

NYCOpenData Roadbeds



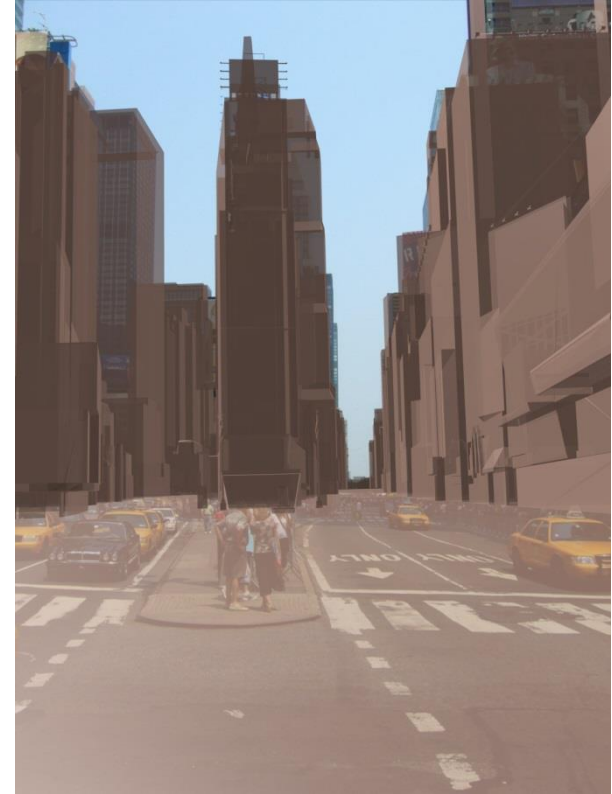
Vision grounded in the real world



Input photo



Overlaid GIS data
(roads / sidewalks / medians)



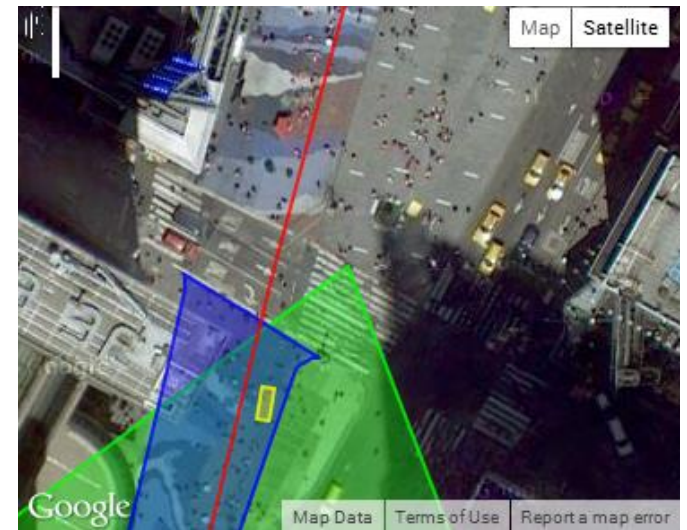
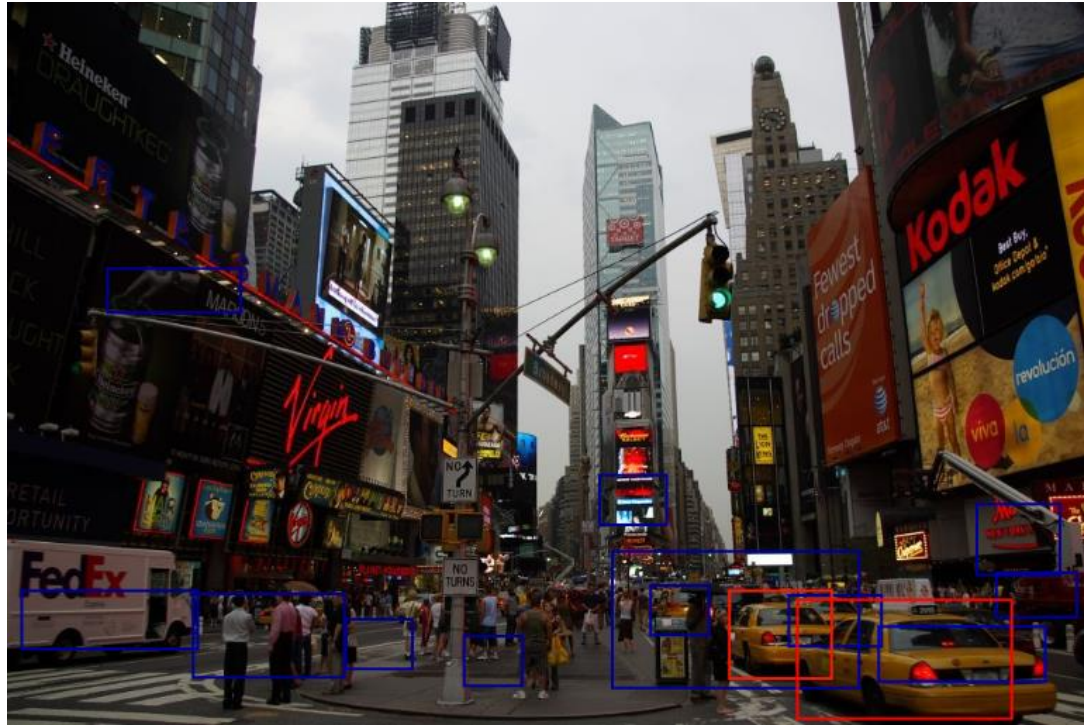
Overlaid Google Earth models

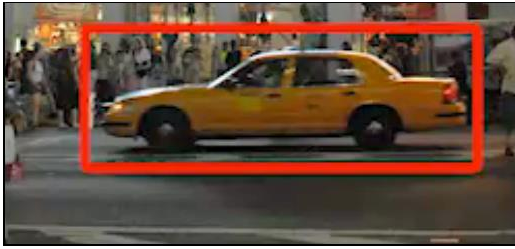
Annotated 3D Vehicles



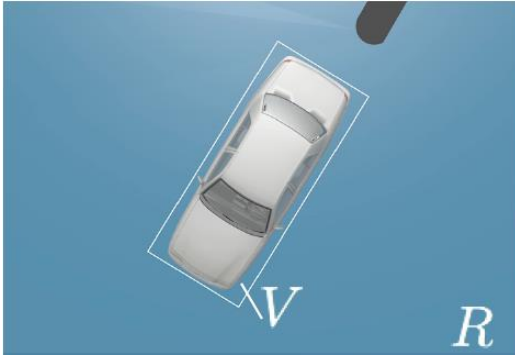
Video

3D Detection

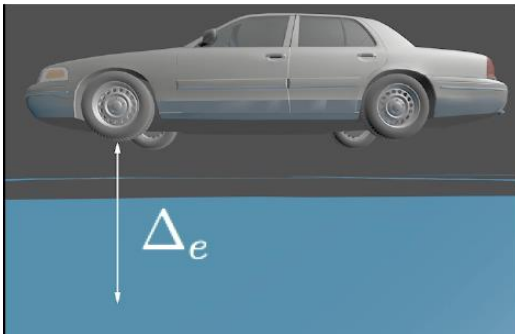




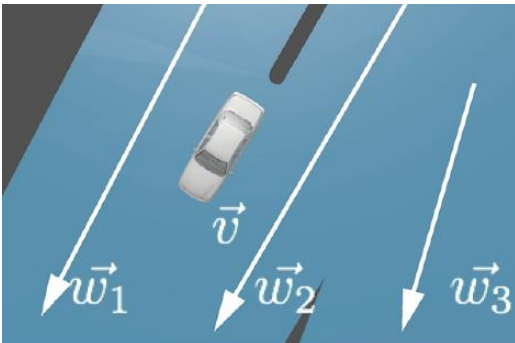
Appearance score



Ground coverage score

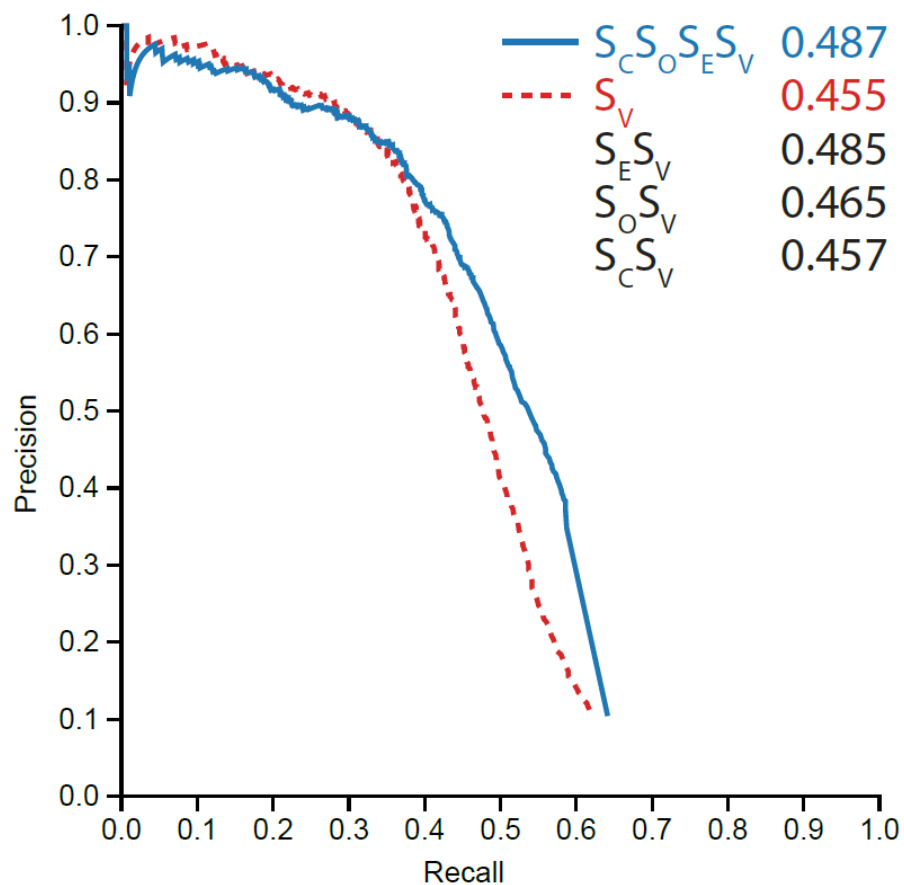


Elevation score

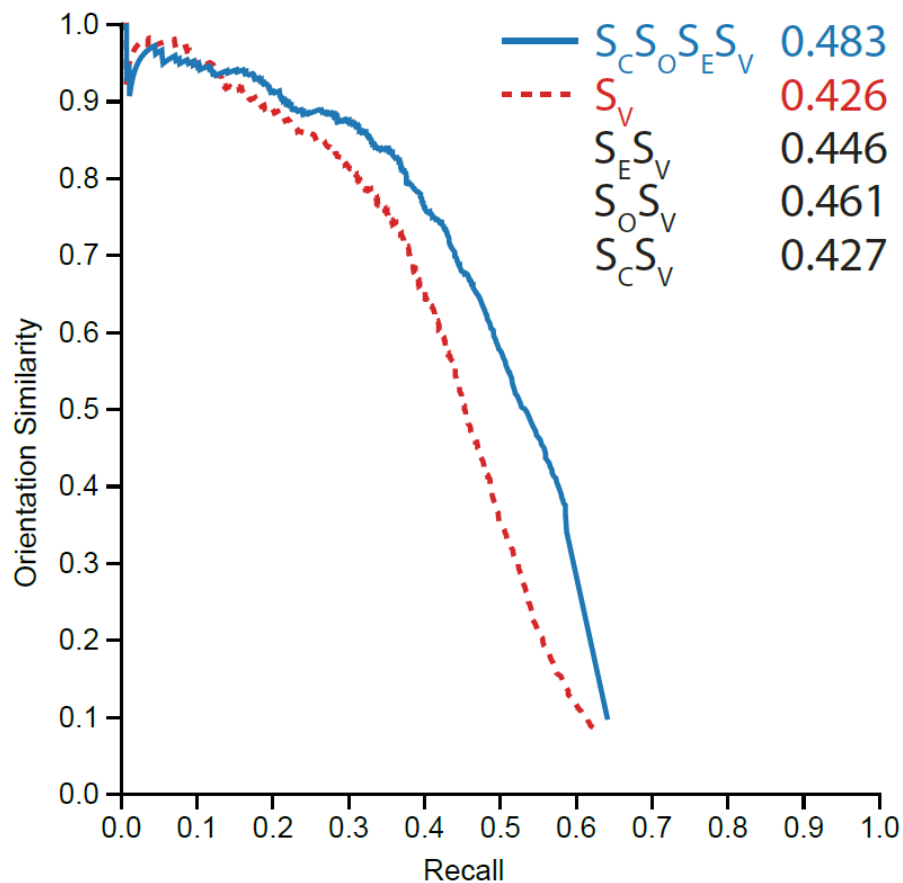


3D orientation score

Results



Precision / Recall



Orientation similarity / Recall



A vehicle detection database for vision tasks set in the real world.



3D Reconstructions

Each photograph in NYC3DCars has been geo-registered to the Earth, providing full camera intrinsics and extrinsics in an Earth-Centered, Earth-Fixed coordinate system enabling seamless integration with existing geospatial data.



Geographic Data

Companion databases such as those provided by [OpenStreetMap](#) and [NYC OpenData](#) have been integrated for easy access to geographic features such as road, sidewalk, and median polygons as well as road network connectivity.



Vehicle Annotations

Human annotators have provided detailed descriptions for vehicles contained in the database. Annotations include a full 6 degree of freedom vehicle pose, vehicle type, 2D vehicle bounding box, and approximate photo time of day.

<http://nyc3d.cs.cornell.edu/>

Summary

- Massive image collections are a rich source of information about our world
- In vision we should be leveraging all available data: images, depth, **geographic context**
- 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
- National Science Foundation
- Intel Center for Science and Technology – Visual Computing
- Amazon AWS for Education

Thank you!

More information at

<http://www.cs.cornell.edu/~snave/>

