Geometry Beyond 3D

Noah Snavely
Google Inc., Cornell University

Bay Area Vision Meeting, 2014
Are we done with 3D modeling?

- Huge progress in the last 10 years

[Pollefeys et al. IJCV04]

[Snavely et al. SIGGRAPH06]

[Aerial models]

[Zhou & Koltun, SIGGRAPH14]
Are we done with 3D modeling?

[Agarwal et al. ICCV 2009]

[Klingner et al., ICCV 2013]
Are we done with 3D modeling?

• Not until we have a fully realistic, editable, semantically meaningful model of the entire world

• **Realistic** = correct geometry, materials, lighting; high-resolution; dynamic

• In other words, a model you can feed into your holodeck

See also the Visual Turing Test [Shan et al., 3DV 2013]
Times Square
What are the key challenges?

• Scale – we have made great progress here
• Robustness
• Time
• Materials
• Semantics / grounding

• My own biased view
Robustness
Are two things the same?

• How do we know what we are looking at is the same or different?
Structural similarities break SfM
Structural similarities break SfM
Other examples

Notre Dame Cathedral

St. Paul’s Cathedral
Tracks should contain one 3D point
Tracks can conflate distinct points
SfM Disambiguation

• Most methods reason about inconsistencies across many images
• Inconsistencies in
  – Loops of pairwise geometries
  – Visibility
  – Sequencing
  – Global geometry

[Zach et al., CVPR 2008], [Zach et al., CVPR 2010],
[Roberts et al., CVPR 2011], [Jiang et al., CVPR 2012]
SfM Disambiguation in the Large

• We wanted a solution that was
  – As simple as possible
  – Scalable to huge image collections

• Intuition: visibility of points is (often) transitive

[Wilson & Snavely, Network Principles for SfM. ICCV 2013]
Graph topology is a cue for ambiguities

Schematic of a scene with an ambiguous feature (in red)

Note that the two sides of the scene have different background (blue and green)

Graph topology is a cue for ambiguities

This structure can be seen in the visibility graph

[Wilson & Snavely, Network Principles for SfM. ICCV 2013]
Larger example

Bad tracks have more than one cluster of context. Measure this with the bipartite local clustering coefficient.
Bad tracks have more than one cluster of context. Measure this with the bipartite local clustering coefficient.
blcc is analogous to the local clustering coefficient

\[ lcc(\text{red node}) = \frac{\text{closed triangles}}{\text{possible triangles}} = \frac{3}{10} \]

\[ blcc(\text{red node}) = \frac{\text{closed 6-paths}}{\text{possible 6-paths}} \]
Filtering by \textit{blcc} removes bad tracks

ROC curve for classifying bad tracks

Algorithm:

1. Compute a covering subgraph
2. Compute blcc for each track
3. Remove tracks lower than a threshold
   Use lowest threshold that separates the graph into a user-predetermined number of components.
4. Reconstruct each component separately
5. Rigidly merge components if possible

Solid line: thresholding tracks on blcc.
Dotted line: same, but on a more uniform subgraph.
Disambiguation results

Sacre Coeur Basilica, Paris
Disambiguation results

Notre Dame Cathedral, Paris
Disambiguation results

Seville Cathedral
Outside the Louvre, Paris
Network Principles for SfM

+ Extremely fast method
+ Based on simple local reasoning
+ Very simple to implement

- Can sometimes oversegment models
- Theoretical guarantees?

See also [Heinly et al. ECCV 2014]
Feature matching as recognition

• Can’t we just solve this problem using appearance alone?
• Better features or image metrics?
Time
Places are dynamic
5pointz, Queens
How do we model these time-varying scenes?
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

Kevin Matzen and Noah Snavely, *Scene Chronology*, ECCV 2014 Best Paper Award Winner
Single 3D Model (from ~100,000 images)
Space-Time Point Clustering

Exploded View across Time
Re-time-stamping

Blue: original timestamp
Red: our predicted timestamp
Materials
OpenSurfaces
Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala
Cornell University

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Materials</th>
<th>Reflectances</th>
<th>Textures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good</strong></td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td><strong>All</strong></td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>Labeled Scenes</td>
<td>25,357</td>
<td>91,876</td>
<td></td>
</tr>
<tr>
<td>Whitesbalanced Photos</td>
<td>17,839</td>
<td>24,771</td>
<td></td>
</tr>
<tr>
<td>Segmentations</td>
<td>70,055</td>
<td>103,513</td>
<td></td>
</tr>
<tr>
<td>Planar Segmentations</td>
<td>36,482</td>
<td>70,005</td>
<td></td>
</tr>
<tr>
<td>Named Materials</td>
<td>56,625</td>
<td>68,761</td>
<td></td>
</tr>
<tr>
<td>Named Objects</td>
<td>31,697</td>
<td>42,203</td>
<td></td>
</tr>
<tr>
<td>Rectified Textures</td>
<td>16,829</td>
<td>22,219</td>
<td></td>
</tr>
</tbody>
</table>

Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala, SIGGRAPH 2013
http://opensurfaces.cs.cornell.edu/
Query

Results: wood floors in kitchens, sorted by diffuse color

Query

Results: fabric sofas in living rooms, sorted by diffuse color similarity
Intrinsic Images in the Wild

Sean Bell, Kavita Bala, Noah Snavely
Cornell University

Semantics / Grounding
Every image tells a story...
Grounding vision in the world

OpenStreetMap

3D city models

Weather data

Bus schedules
<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>3. Zip Codes Map Social Services geographic, location, map, cartography, zip, code, ... Polygon representing the boundary of the zip codes in the city.</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>5. Restaurant Inspection Results Health restaurant inspection results, ... NYC restaurant inspection results</td>
</tr>
<tr>
<td>6. Basic Description of Colleges and Universities Education doit gis, geographic, location, map, cartography, ... Location of colleges and universities with basic descriptive information.</td>
</tr>
<tr>
<td>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</td>
</tr>
<tr>
<td>8. Mapped In NY Companies Business jobs, tech, jobs and economic mobility Raw data which powers the Mapped In NY site at <a href="http://www.mappedinny.com/">http://www.mappedinny.com/</a></td>
</tr>
<tr>
<td>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</td>
</tr>
</tbody>
</table>

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: NYC3DCars

[Kevin Matzen and Noah Snavely, ICCV 2013]
NYCOpenData Roadbeds
Vision grounded in the real world

Input photo

Overlayed GIS data (roads / sidewalks / medians)

Overlayed Google Earth models
Annotated 3D Vehicles
Video
3D Detection
Appearance score

Ground coverage score

Elevation score

3D orientation score
Results

Precision / Recall

- $S_{C \circ O \circ E \circ V}$: 0.487
- $S_{V}$: 0.455
- $S_{E \circ V}$: 0.485
- $S_{O \circ V}$: 0.465
- $S_{C \circ V}$: 0.457

Orientation similarity / Recall

- $S_{C \circ O \circ E \circ V}$: 0.483
- $S_{V}$: 0.426
- $S_{E \circ V}$: 0.446
- $S_{O \circ V}$: 0.461
- $S_{C \circ V}$: 0.427
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

• Many interesting challenges in modeling the world

• Contributions from every area (cf. much wonderful recent work):
  – Scene understanding, object detection, material recognition, illumination modeling, ...
  – Learning?
Acknowledgements

Students

Sean Bell  Song Cao  Daniel Hauagge  Kevin Matzen

Paul Upchurch  Chun-Po Wang  Scott Wehrwein  Kyle Wilson

Collaborators

Kavita Bala  Dan Huttenlocher  Yunpeng Li  Dave Crandall

• 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/~snavely/