3D Exploitation of Large Urban Photo Archives*

High Performance Embedded Computing Workshop

Professor Noah Snavely
Computer Science Department
Cornell University

Peter Cho & Ross Anderson
Active Optical Systems Group
MIT Lincoln Laboratory

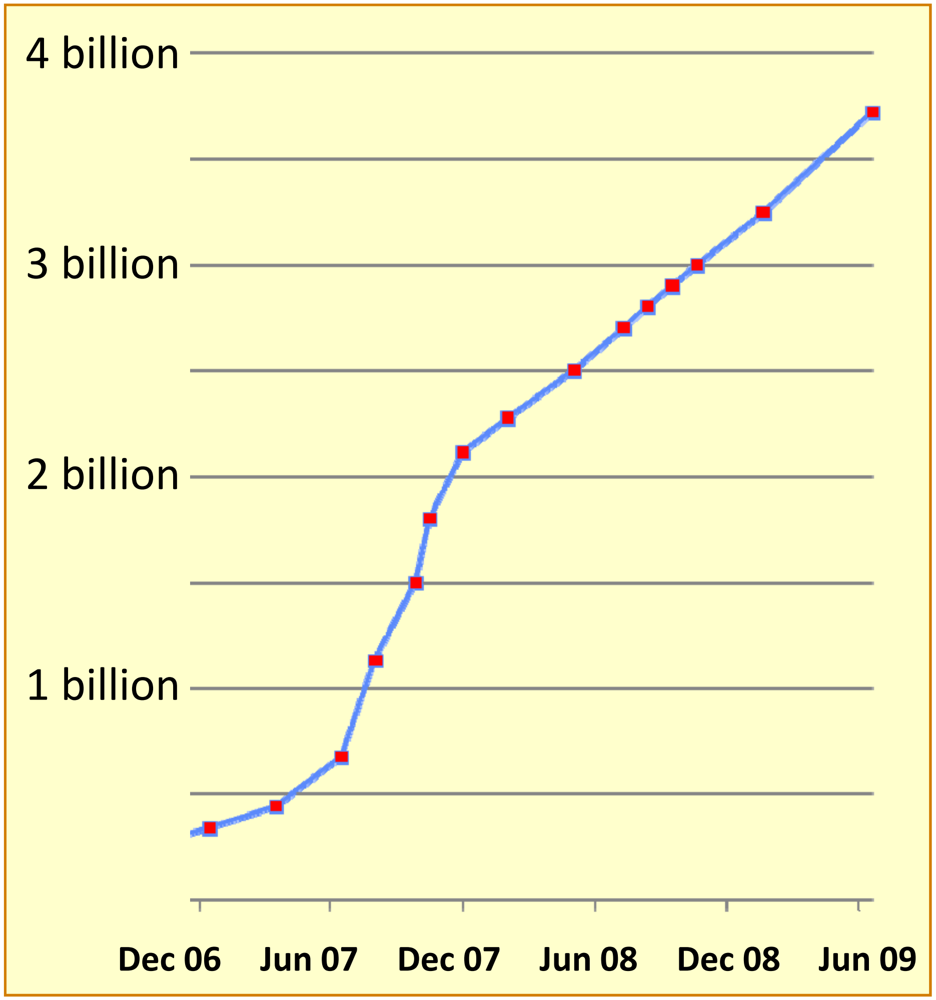
September 2009

*This work was sponsored by the Department of the Air Force under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

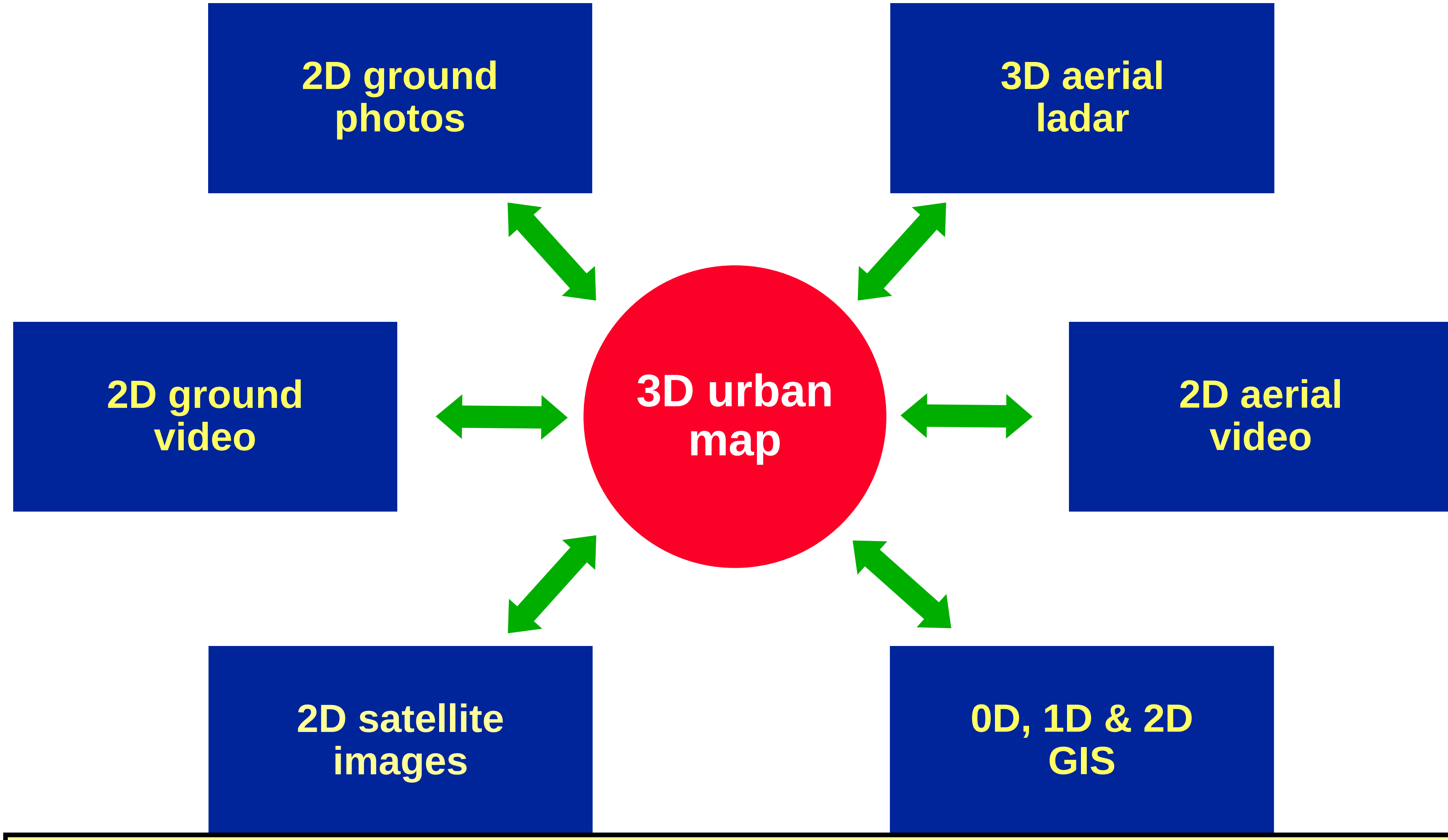
Problem Statement: Urban Digital Imagery Explosion

- **Quantity & quality of digital urban imagery are rapidly increasing over time**
- **Vast numbers of photos shot by inexpensive digital cameras can be accessed via web**
  - **Recent Google search found 237 million images matching New York City, 5.3 million for Baghdad & 2.4 million for Kabul**
  - **But no connection exists between retrieved thumbnails besides their having been shot in same metropolitan area**
- **Organizing principle is needed for navigating & exploiting large urban imagery archives**

Total number of flickr website photos vs year



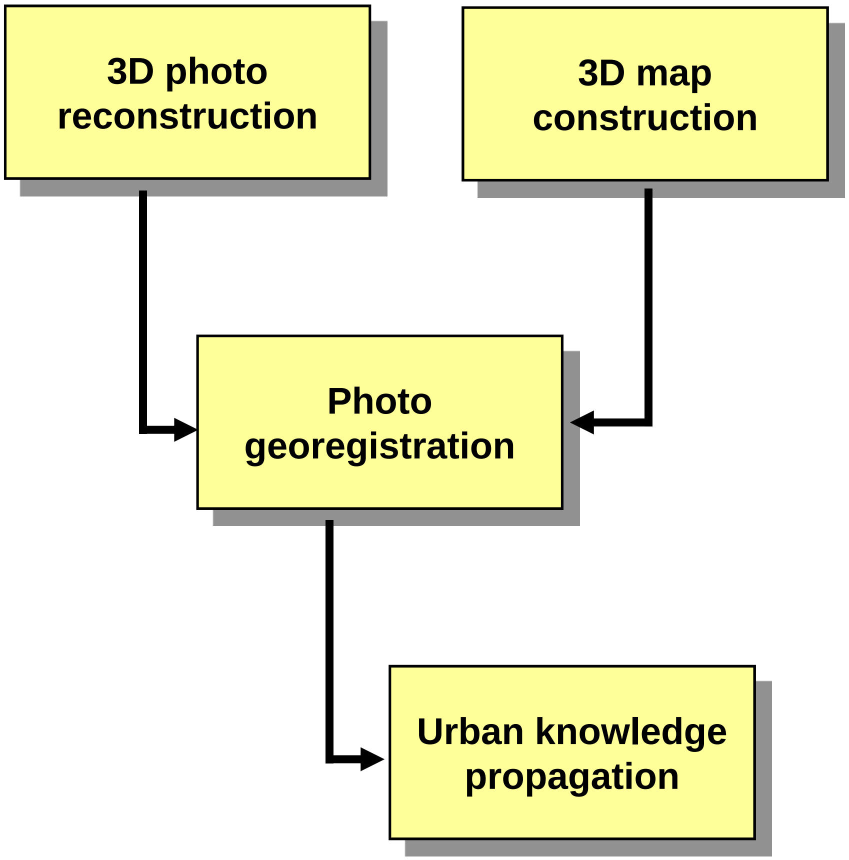
Multi-Source Image Organization



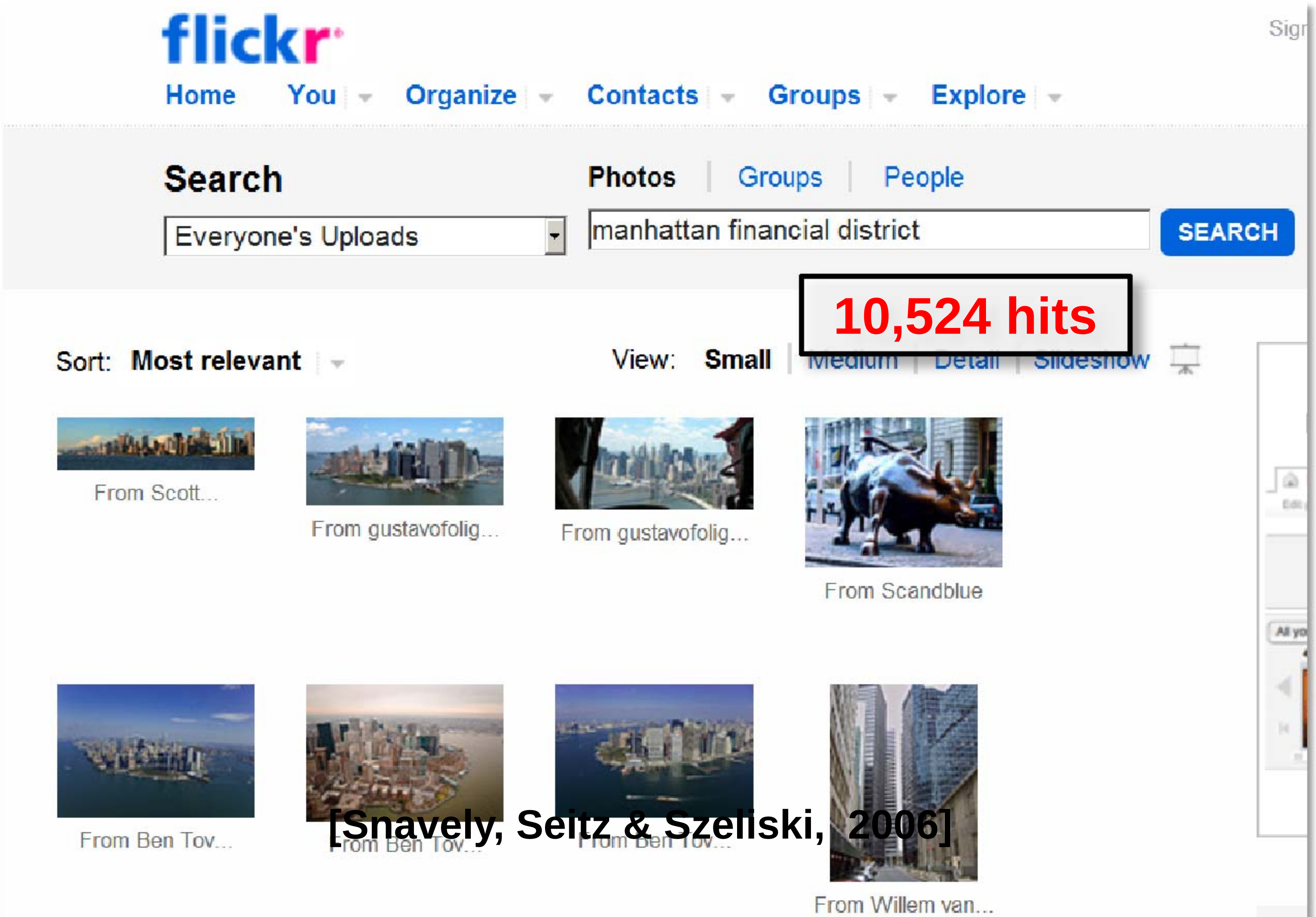
3D urban map provides geometrical framework for organizing imagery collected at different times, places, perspectives & resolutions

Outline

- 3D photo reconstruction
  - SIFT feature detection & matching
  - Structure from motion
  - LLGrid parallelization
- 3D map construction
- Photo georegistration
- Urban knowledge propagation
- Ongoing & future work



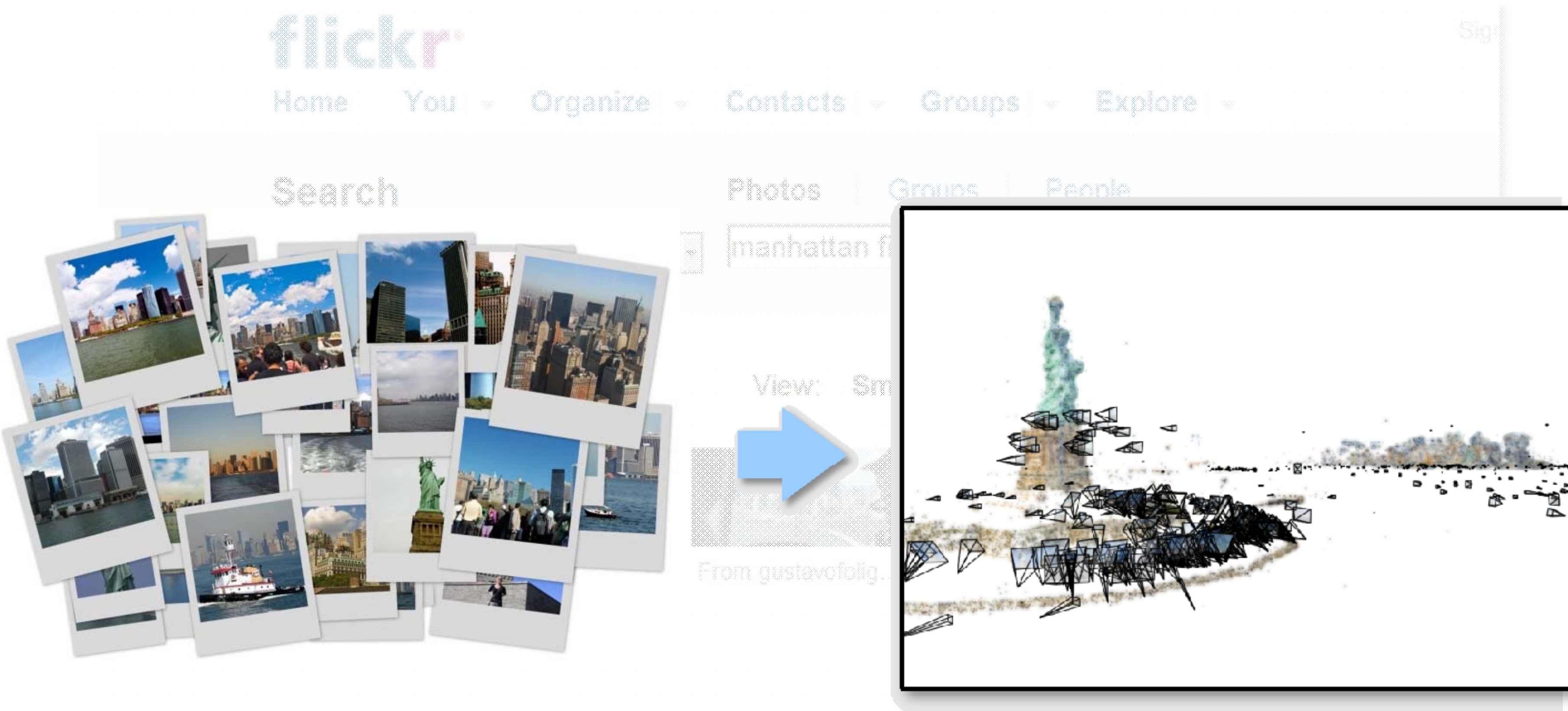
Structured Output from Unstructured Input



10,524 hits

[Snavely, Seitz & Szeliski, 2006]

Structured Output from Unstructured Input



[Snavely, Seitz & Szeliski, 2006]

SIFT Feature Detection & Matching

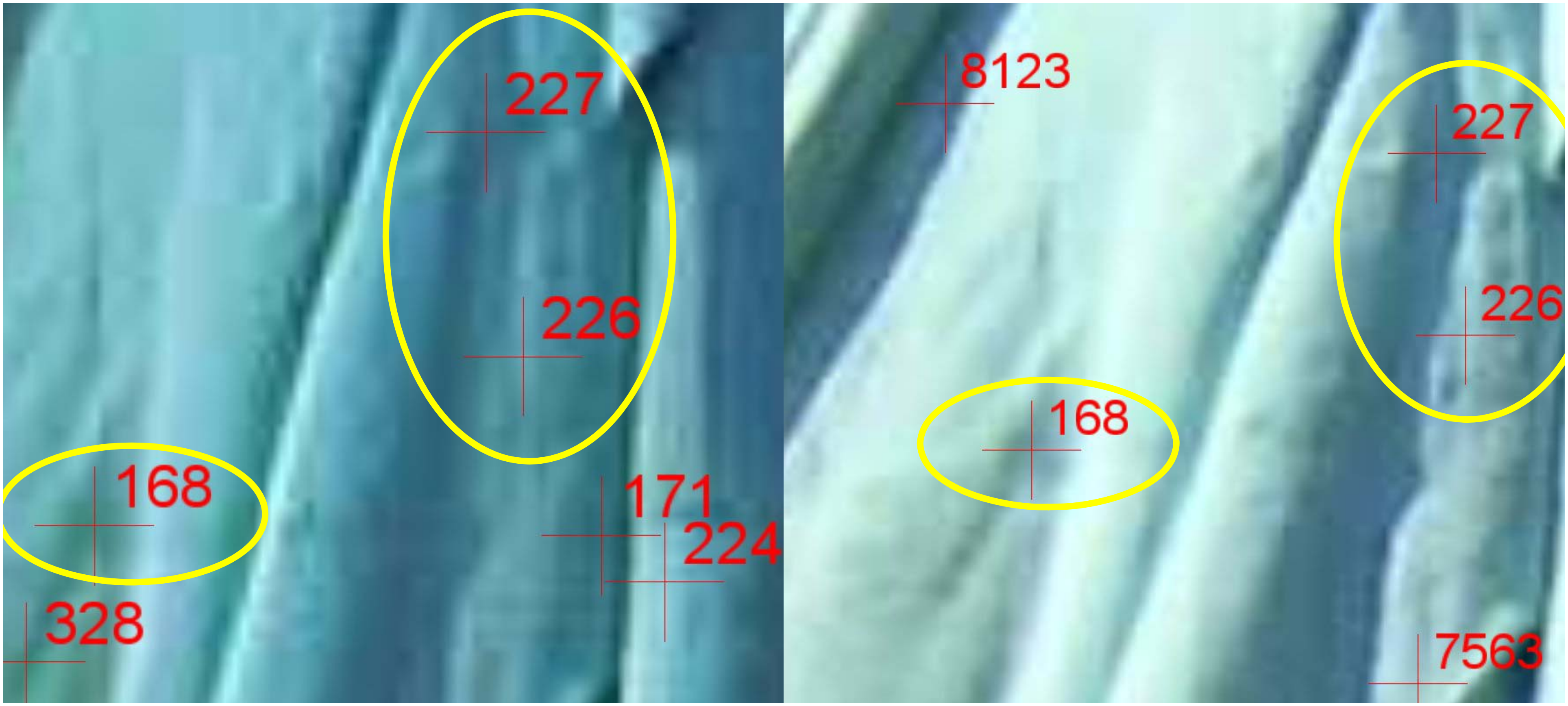
- **Extract Scale Invariant Feature Transform (SIFT) features [Lowe, 2004]**

SIFT features extracted from 2 photos with partially overlapping views

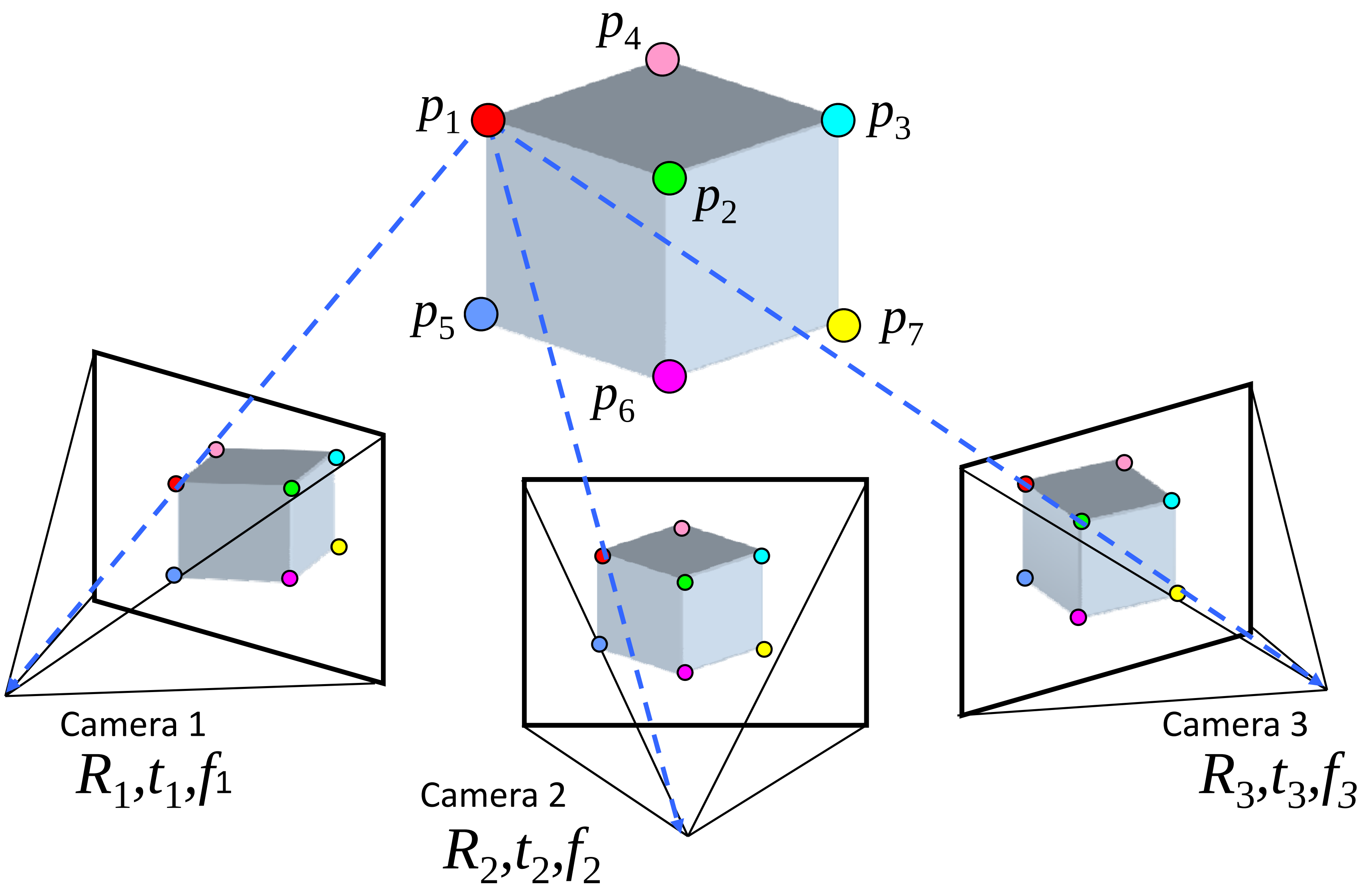
SIFT Feature Detection & Matching

- **Extract Scale Invariant Feature Transform (SIFT) features [Lowe, 2004]**
- **Identify feature tiepoint matches**
  - **Use Approximate Nearest Neighbor algorithm to search 128-dim vector space**
- **Employ Random Sample Consensus to minimize false tiepoint pairings**

SIFT features extracted from 2 photos with partially overlapping views



Structure from Motion



Statue & Skyline Reconstruction

- **Relative positions & poses for 1012 cameras automatically determined**
  - **Camera covariance matrices also provide reconstruction uncertainty estimates**
- **Target 3D geometry contained in sparse point cloud output**

Play Video

Naïve Computational Complexity

- **SIFT feature detection**
  - O(n) where n = number of photos
  - Embarrassingly parallel (each image can be independently processed)

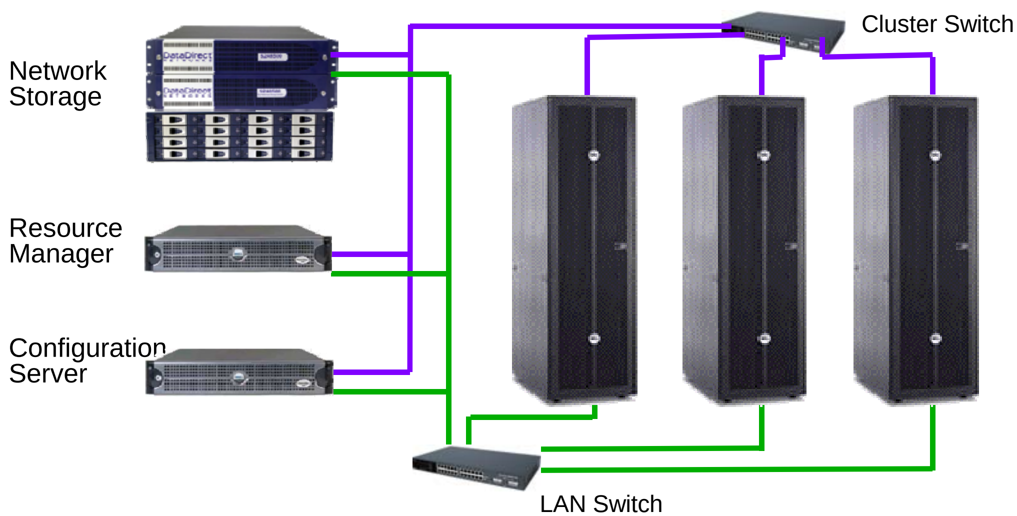
- **SIFT feature matching**
  - O(n^2) when matching must be performed for each pair of uncooperatively collected photos
  - Embarrassingly parallel (each match can be independently processed)

- **Structure from motion**
  - Optimization problem for camera & target point parameters
  - Thousands of cameras & millions of points generate millions of parameters to determine via nonlinear objective function
  - Solving nonlinear least squares optimization is at worst O(n^4) (but often faster)
  - More difficult to parallelize (but investigating parallel least squares solvers)

LLGrid Processing

- **3D photo reconstruction has been parallelized on Lincoln Lab's Grid [Travinin-Bliss et al, 2006]**

Compute Nodes	**150**
Processors	**130**
Peak Throughput (TFLOPs)	**3.84**
Total RAM (GB)	**1,200**
Central Storage (TB)	**4.3**
Total Local Disk Space (TB)	**90.0**

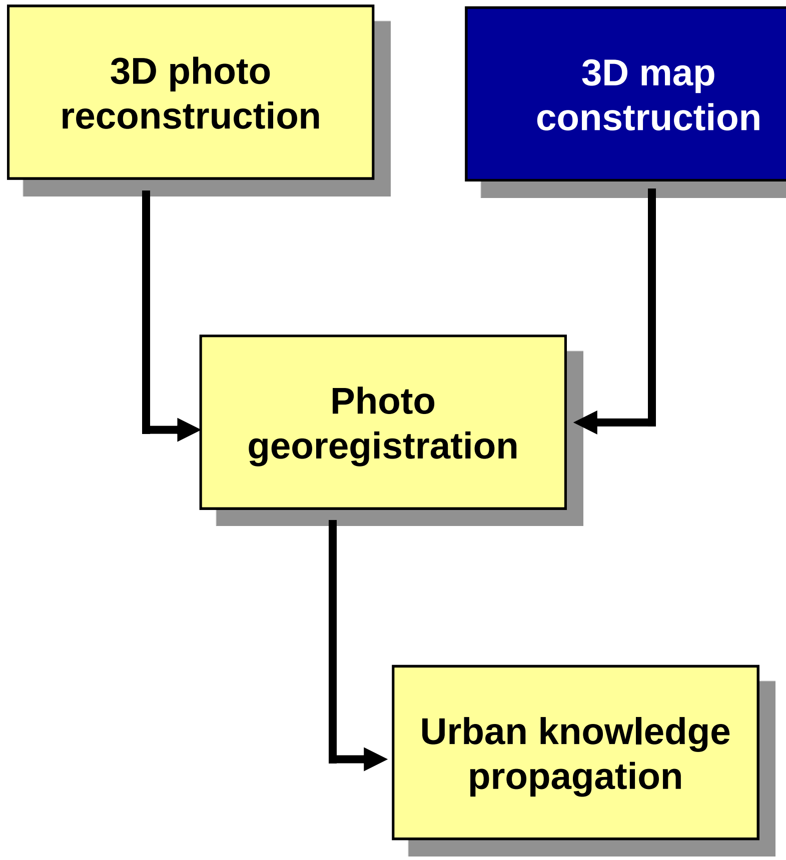


- **Timing results for 1012 NYC internet photo collection processed using 128 LLGrid compute nodes**

SIFT feature detection	**5 minutes**
SIFT feature matching	**2 hours**
Structure from motion	**2 hours**
Total run time	**4 hours**

Outline

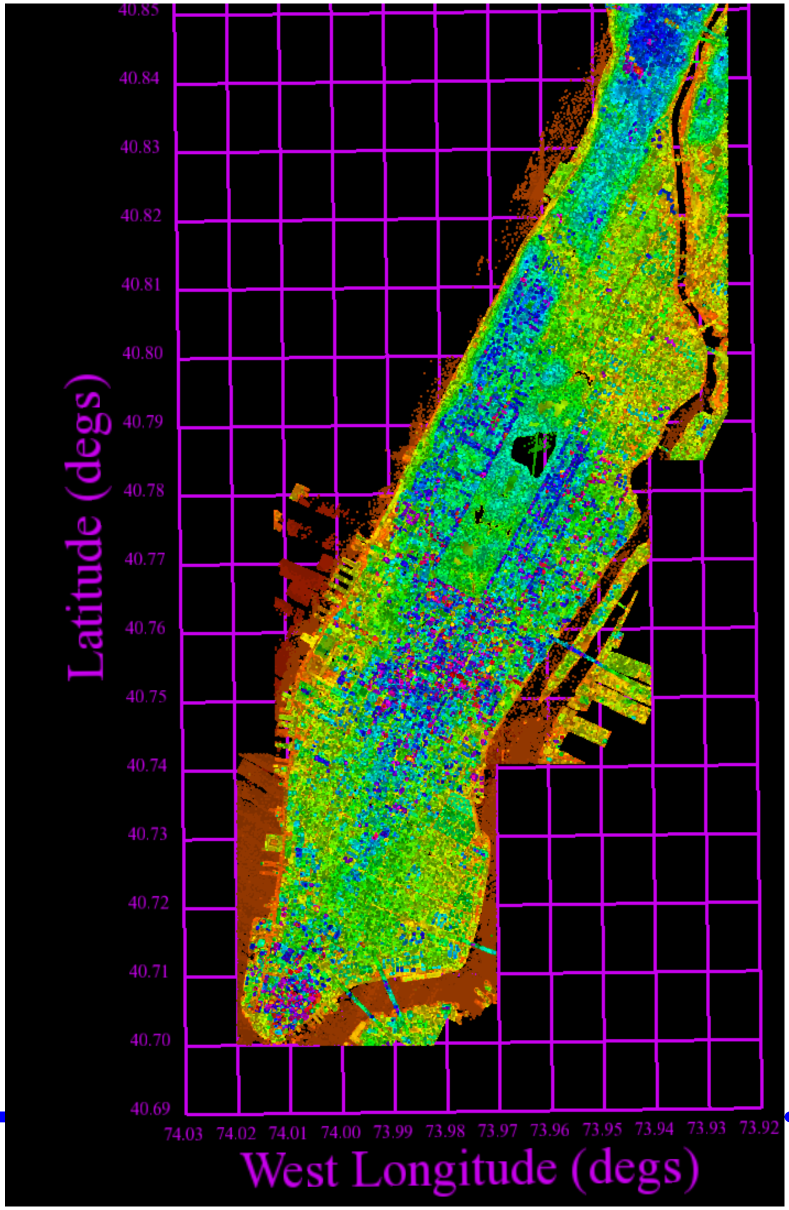
- 3D photo reconstruction
- **3D map construction**
  - **Ladar, EO & GIS data fusion**
- Photo georegistration
- Urban knowledge propagation
- Ongoing & future work



Urban Laser Radar Underlay

- **High-resolution ladars provide 3D geometry backdrops for urban regions**
  - **Wide area ladar maps of entire cities are commercially available**
- **As a representative example, work with Rapid Terrain Visualization ladar imagery collected over New York City in Oct 2001**
  - **1 meter ground sampling distance**
  - **Polynomial warp used to lock point cloud onto absolute longitude/latitude grid**
  - **Maximum local georegistration error ≈ 2 meter**

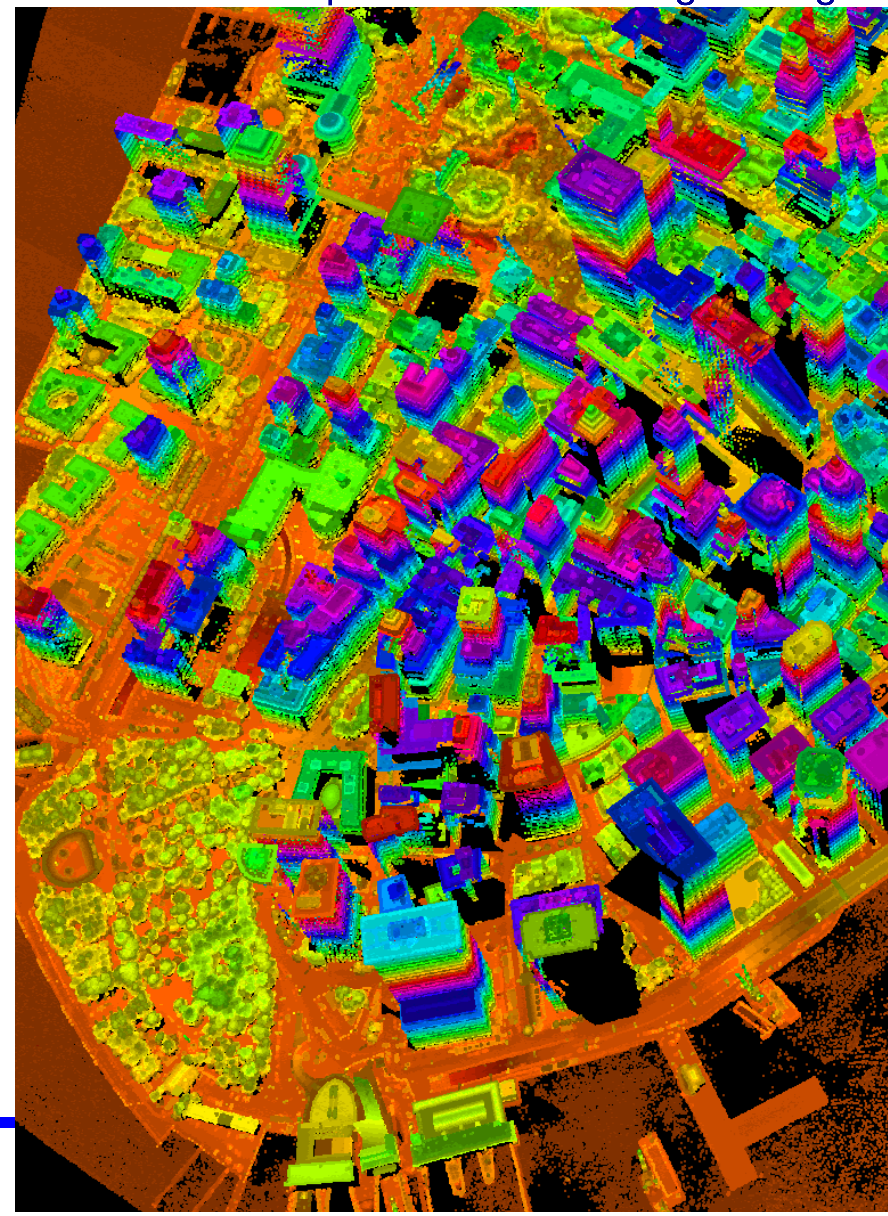
NYC ladar map colored according to height



Urban Laser Radar Underlay

- **High-resolution ladars provide 3D geometry backdrops for urban regions**
  - **Wide area ladar maps of entire cities are commercially available**
- **As a representative example, work with Rapid Terrain Visualization ladar imagery collected over New York City in Oct 2001**
  - **1 meter ground sampling distance**
  - **Polynomial warp used to lock point cloud onto absolute longitude/latitude grid**
  - **Maximum local georegistration error ≈ 2 meter**

NYC ladar map colored according to height

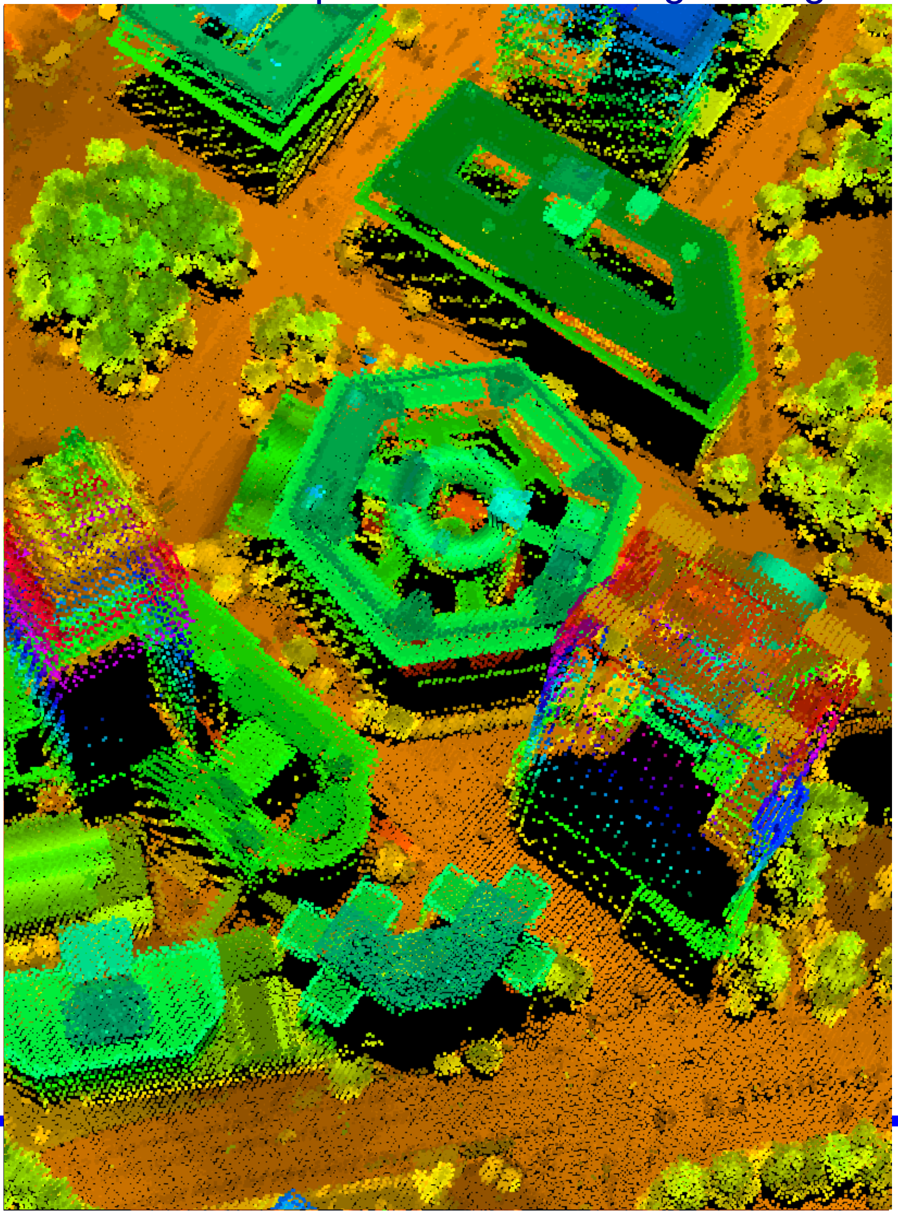


Urban Laser Radar Underlay

- **High-resolution ladars provide 3D geometry backdrops for urban regions**
  - **Wide area ladar maps of entire cities are commercially available**

- **As a representative example, work with Rapid Terrain Visualization ladar imagery collected over New York City in Oct 2001**
  - **1 meter ground sampling distance**
  - **Polynomial warp used to lock point cloud onto absolute longitude/latitude grid**
  - **Maximum local georegistration error ≈ 2 meter**

NYC ladar map colored according to height



Urban Electro-Optical Layer

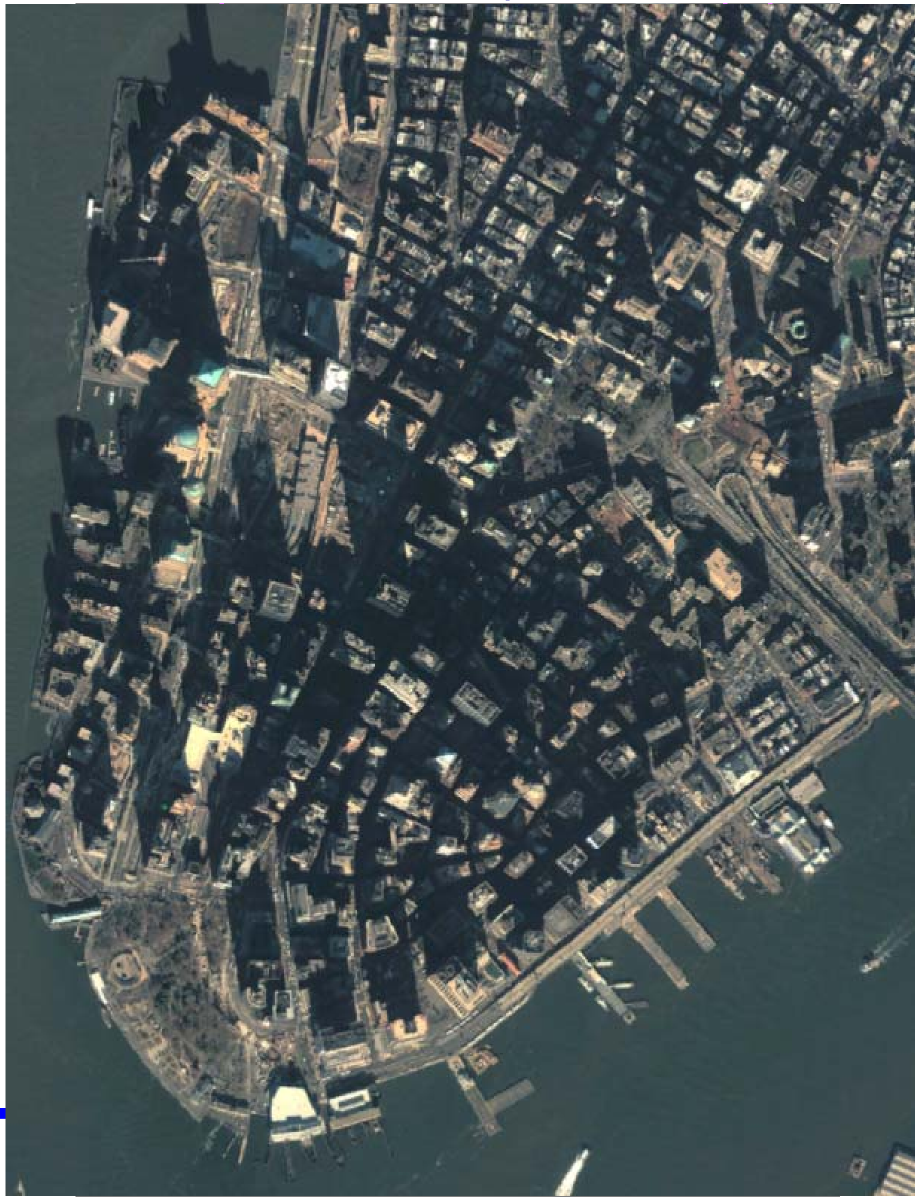
- **Fuse ladar voxels with EO pixels to correlate urban geometry & intensity information**
  - **Building lean is constant [variable] in overhead satellite [aerial] imagery**
  - **Parallax compensation is straightforward [tedious] when photos from satellites [aircraft] are fused with ladar data**
- **Work with NYC satellite image collected in Nov 2006**
  - **0.8 meter ground sampling distance**
  - **Mid-morning collection time avoided cloud obscuration but yielded significant shadowing in extreme urban canyons**

Quickbird satellite photo of NYC

Urban Electro-Optical Layer

- **Fuse ladar voxels with EO pixels to correlate urban geometry & intensity information**
  - **Building lean is constant [variable] in overhead satellite [aerial] imagery**
  - **Parallax compensation is straightforward [tedious] when photos from satellites [aircraft] are fused with ladar data**
- **Work with NYC satellite image collected in Nov 2006**
  - **0.8 meter ground sampling distance**
  - **Mid-morning collection time avoided cloud obscuration but yielded significant shadowing in extreme urban canyons**

Quickbird satellite photo of NYC

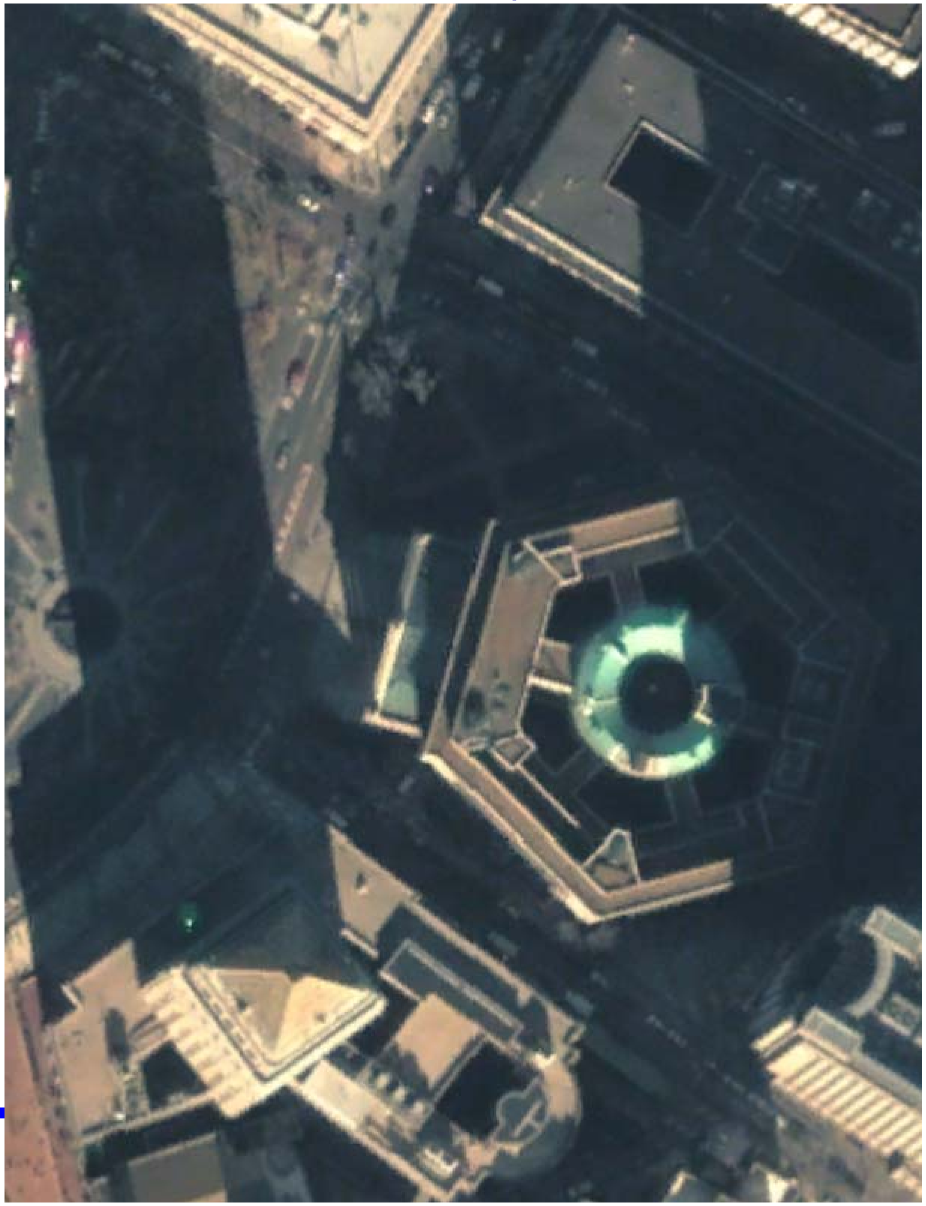


Urban Electro-Optical Layer

- **Fuse ladar voxels with EO pixels to correlate urban geometry & intensity information**
  - **Building lean is constant [variable] in overhead satellite [aerial] imagery**
  - **Parallax compensation is straightforward [tedious] when photos from satellites [aircraft] are fused with ladar data**

- **Work with NYC satellite image collected in Nov 2006**
  - **0.8 meter ground sampling distance**
  - **Mid-morning collection time avoided cloud obscuration but yielded significant shadowing in extreme urban canyons**

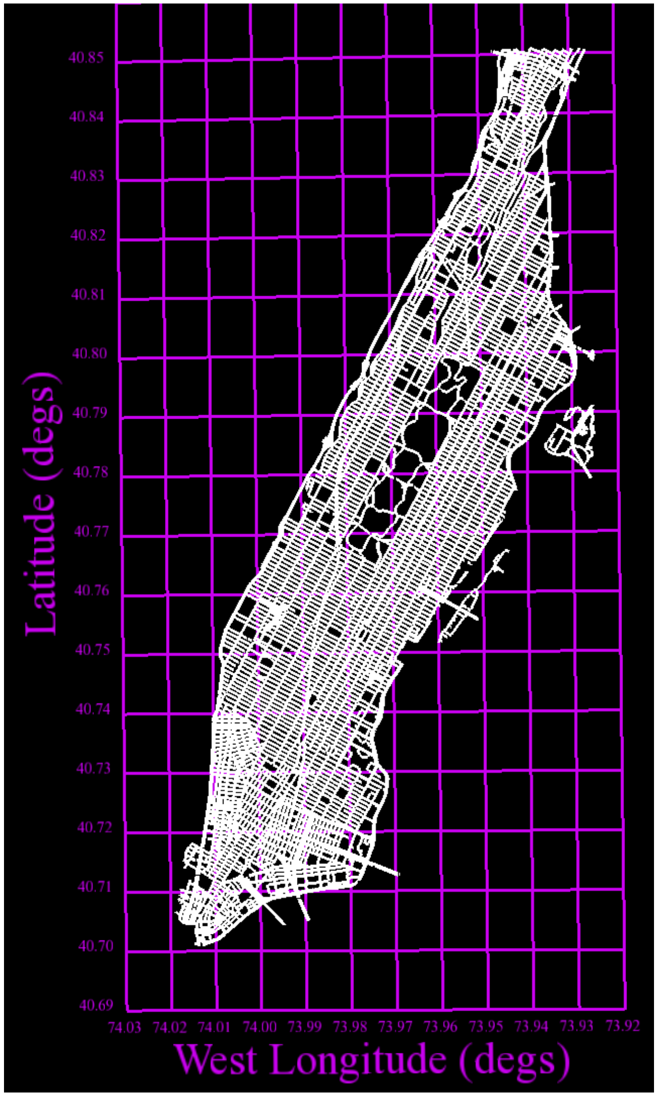
Quickbird satellite photo of NYC



Urban Geographic Information System Layers

- **Fuse ladar voxels & EO pixels with GIS data to combine urban geometry, intensity & network information**
  - **e.g. 2D political districts, 1D transportation routes, 0D landmarks**
  - **Use ladar data to supply altitude information not generally included in GIS databases**
- **GIS layers for New York City have become more tightly controlled after Sept 11, 2001**
  - **But some roadway, subway & local landmark shape files are still accessible on the web**

Road network & points of interest in NYC

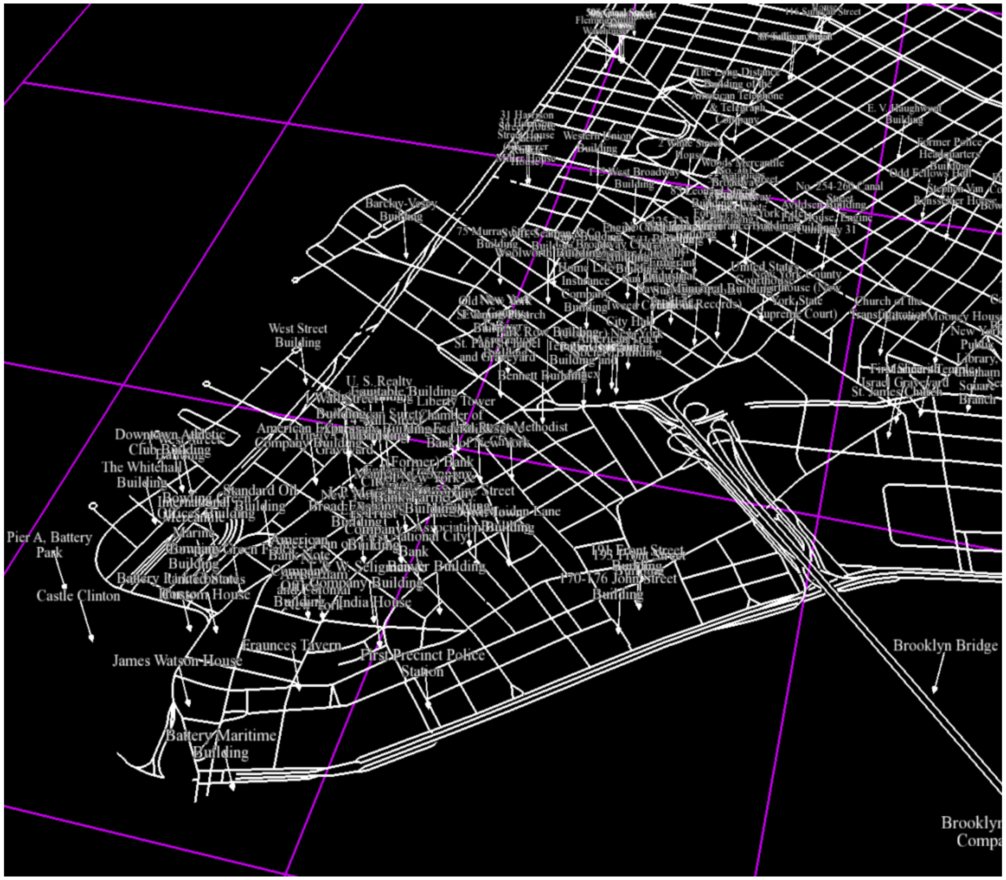


Urban Geographic Information System Layers

- **Fuse ladar voxels & EO pixels with GIS data to combine urban geometry, intensity & network information**
  - **e.g. 2D political districts, 1D transportation routes, 0D landmarks**
  - **Use ladar data to supply altitude information not generally included in GIS databases**

- **GIS layers for New York City have become more tightly controlled after Sept 11, 2001**
  - **But some roadway, subway & local landmark shape files are still accessible on the web**

Road network & points of interest in NYC

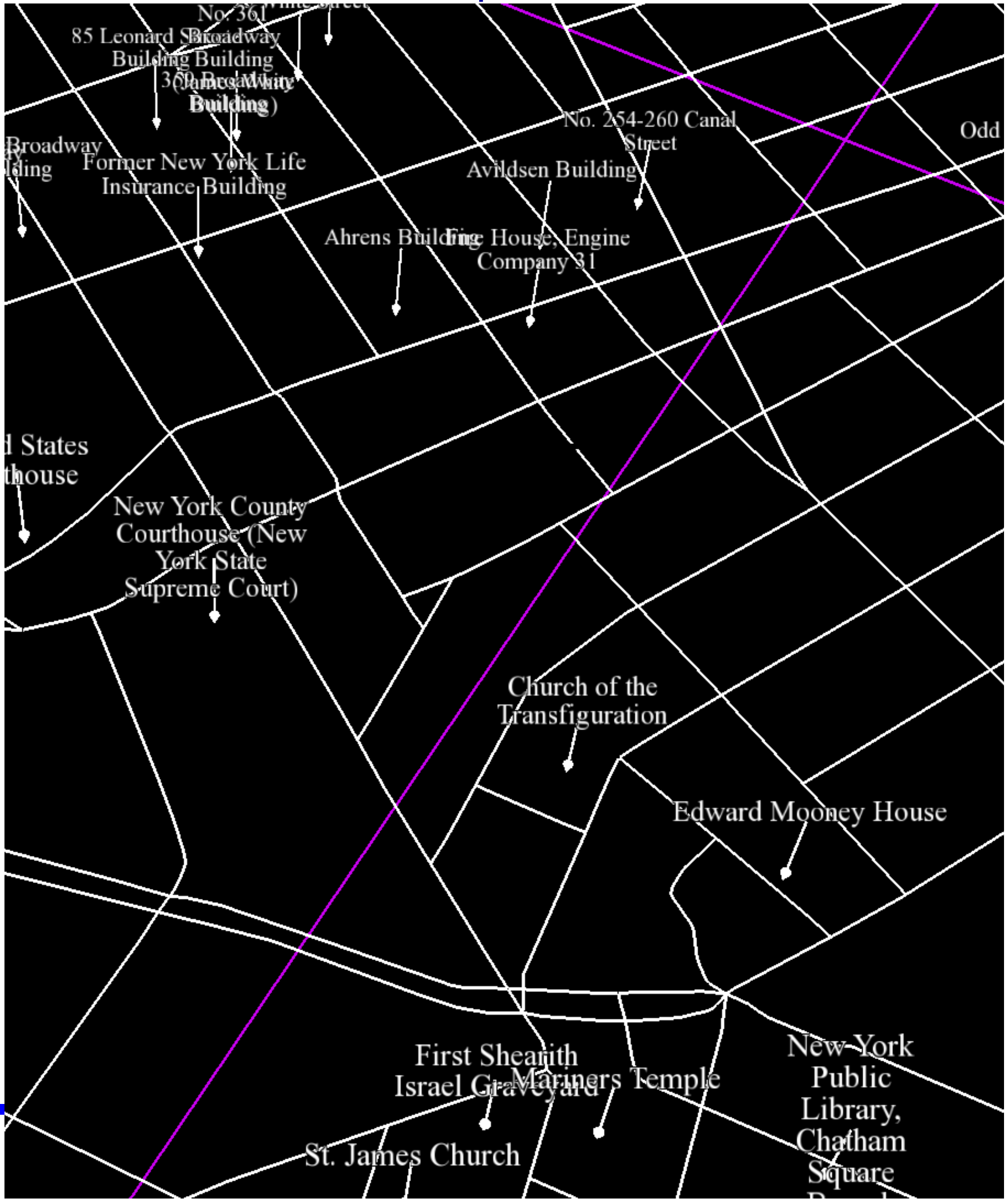


Urban Geographic Information System Layers

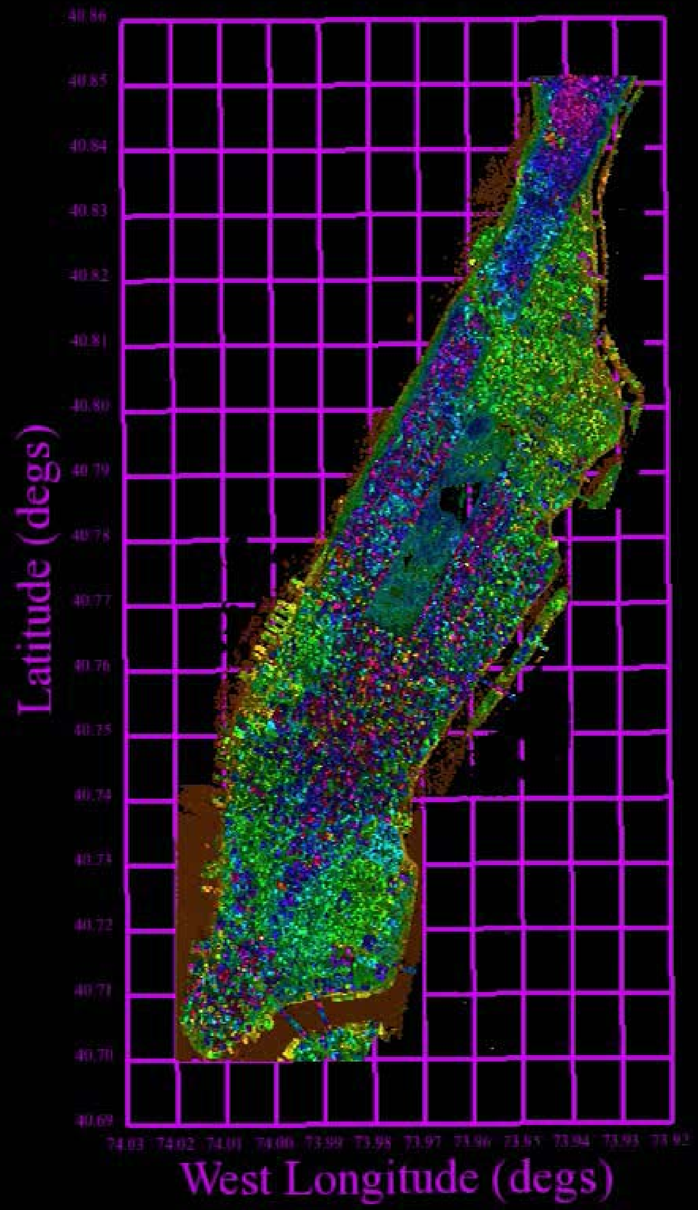
- **Fuse ladar voxels & EO pixels with GIS data to combine urban geometry, intensity & network information**
  - **e.g. 2D political districts, 1D transportation routes, 0D landmarks**
  - **Use ladar data to supply altitude information not generally included in GIS databases**

- **GIS layers for New York City have become more tightly controlled after Sept 11, 2001**
  - **But some roadway, subway & local landmark shape files are still accessible on the web**

Road network & points of interest in NYC



Fused 3D NYC Map [Cho, 2007]



Play Video

Outline

- 3D photo reconstruction
- 3D map construction
- **Photo georegistration**
  - **Dual map & graph displays**
- Urban knowledge propagation
- Ongoing & future work

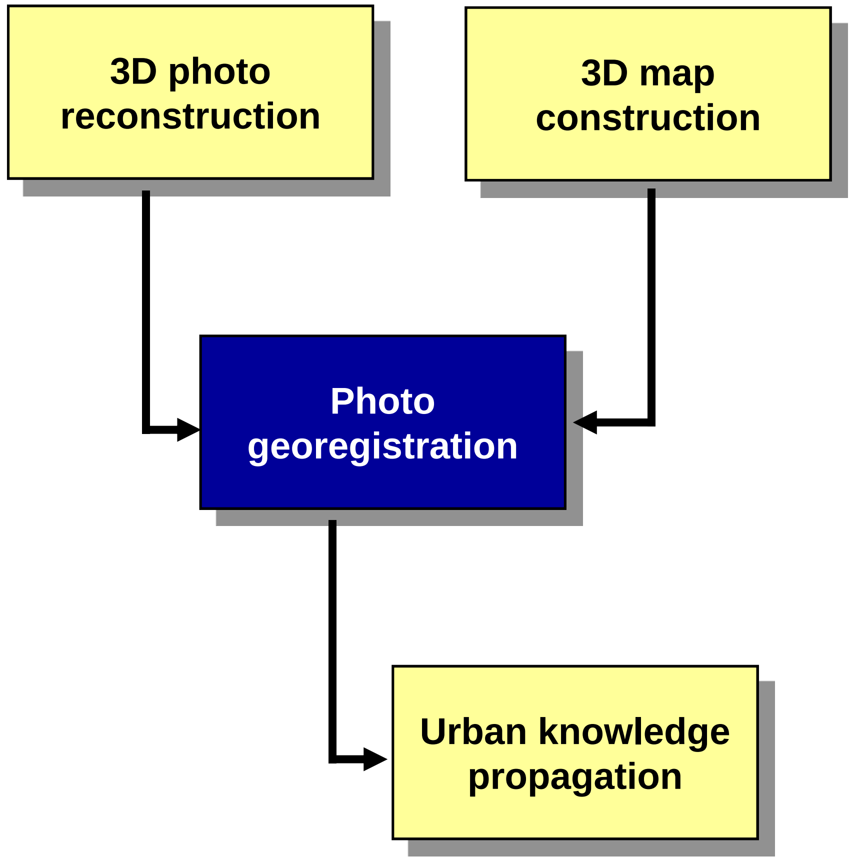
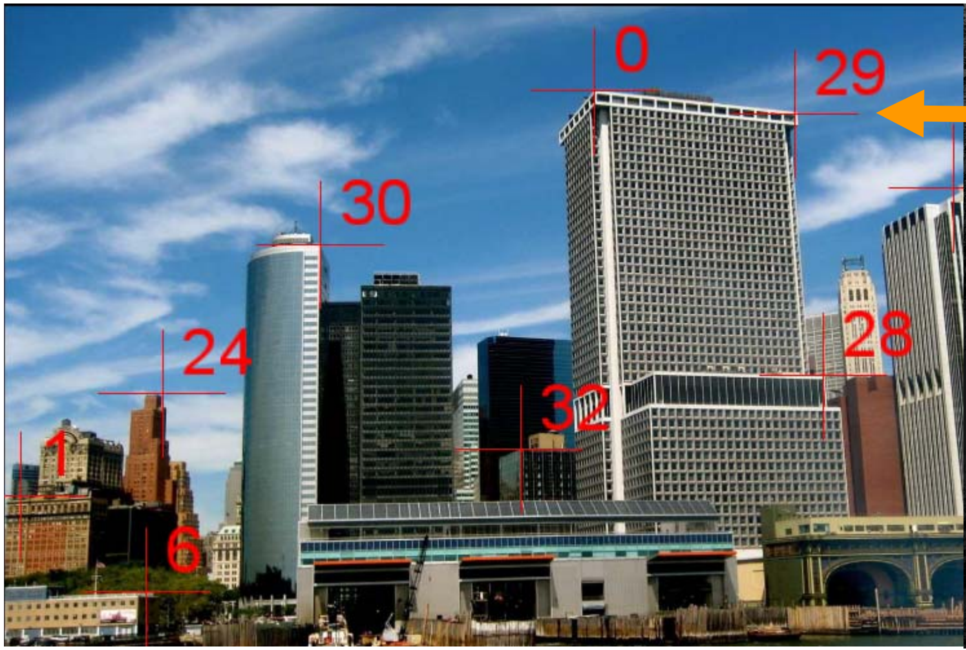


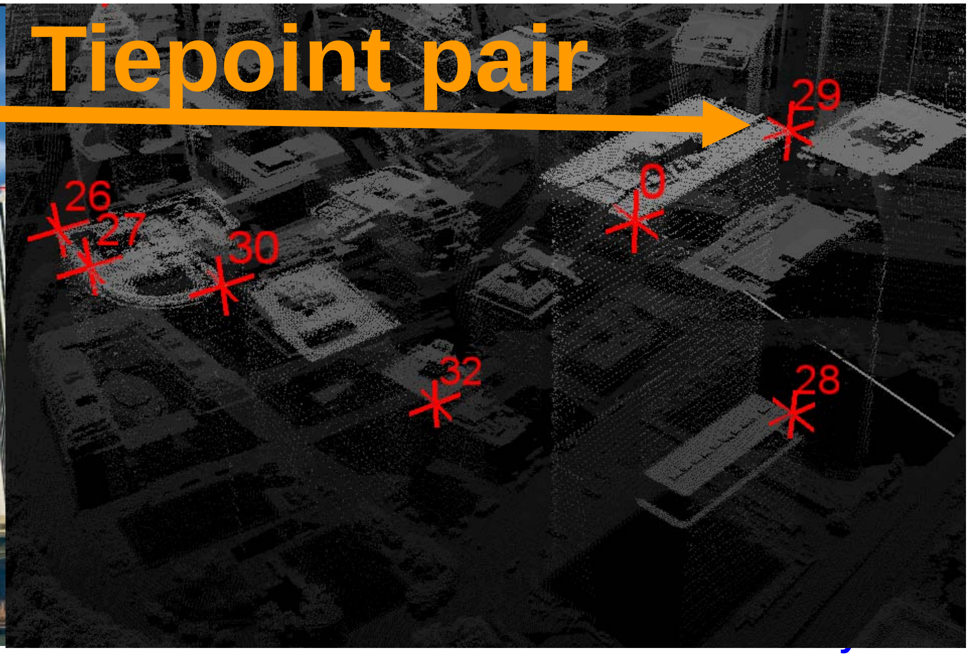
Photo & Ladar Data Matching

- **Select 10 from 1012 reconstructed photos with low covariance traces**
- **Manually establish 33 tiepoint pairings between the 10 photos & ladar map**
- **Compute single global translation, rotation & scaling needed to register reconstructed photos with 3D NYC data**

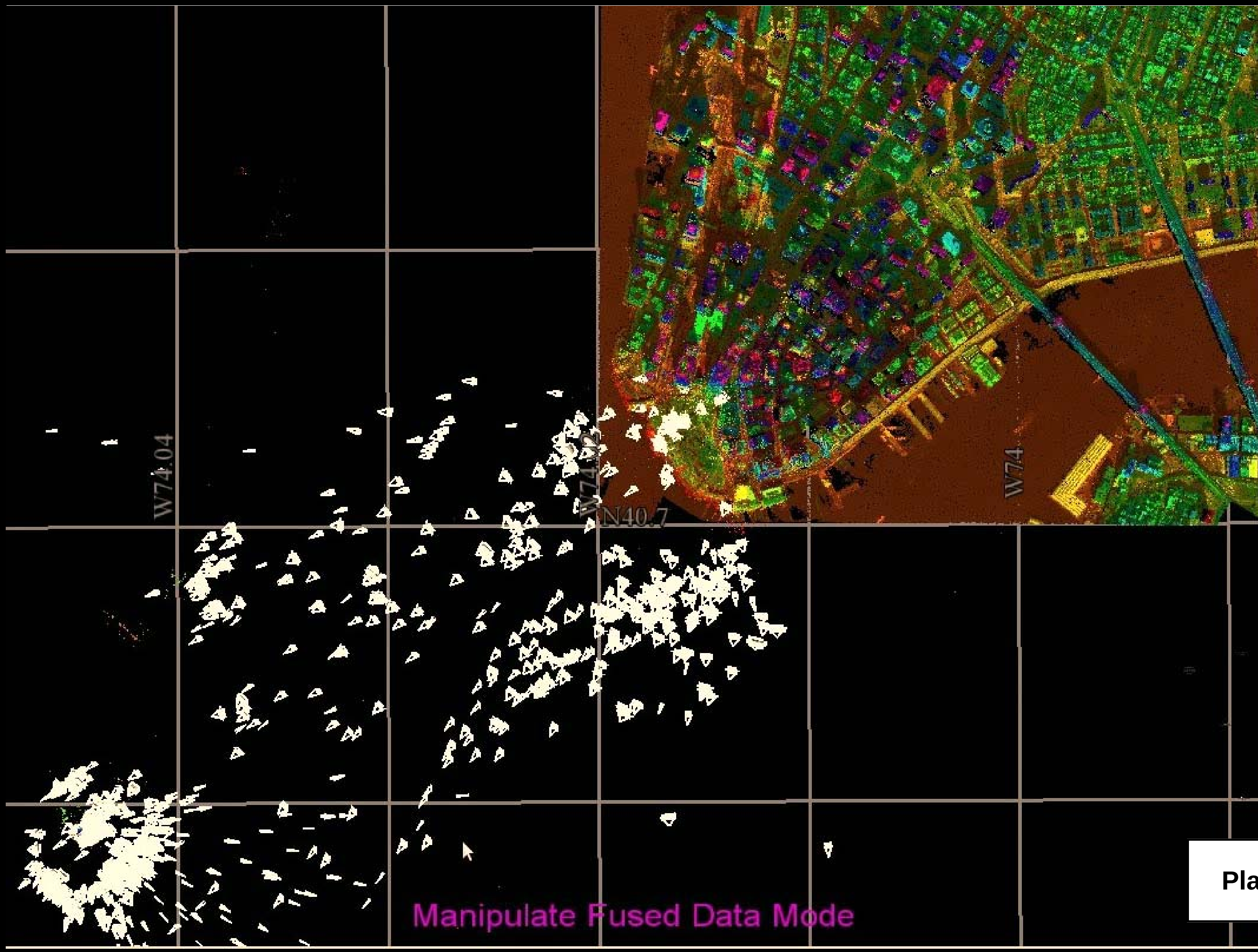
Lower Manhattan skyline photo



Grey-scale colored NYC ladar map



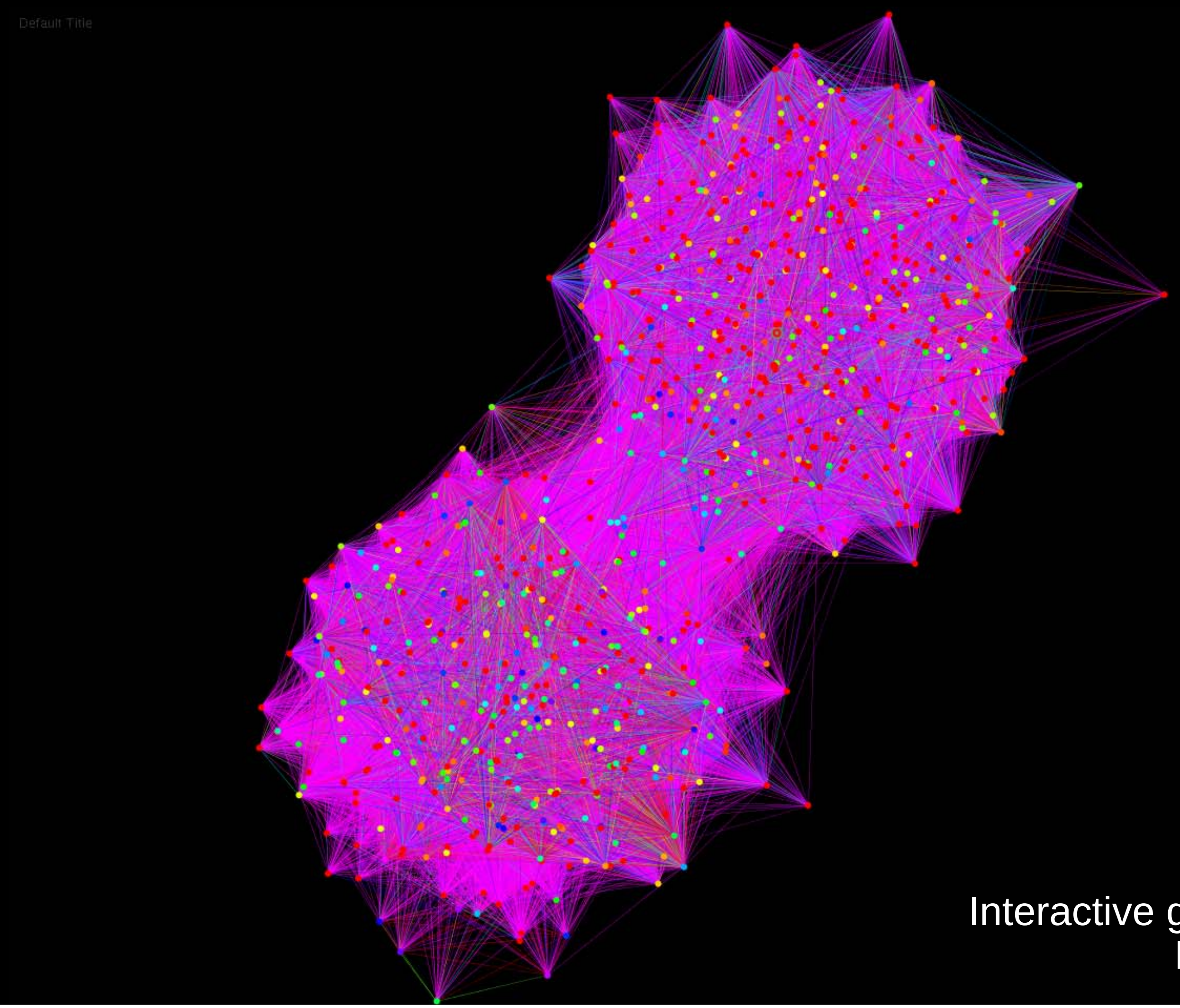
1012 Photos Georegistered with 3D Map



Play Video

Graph Display

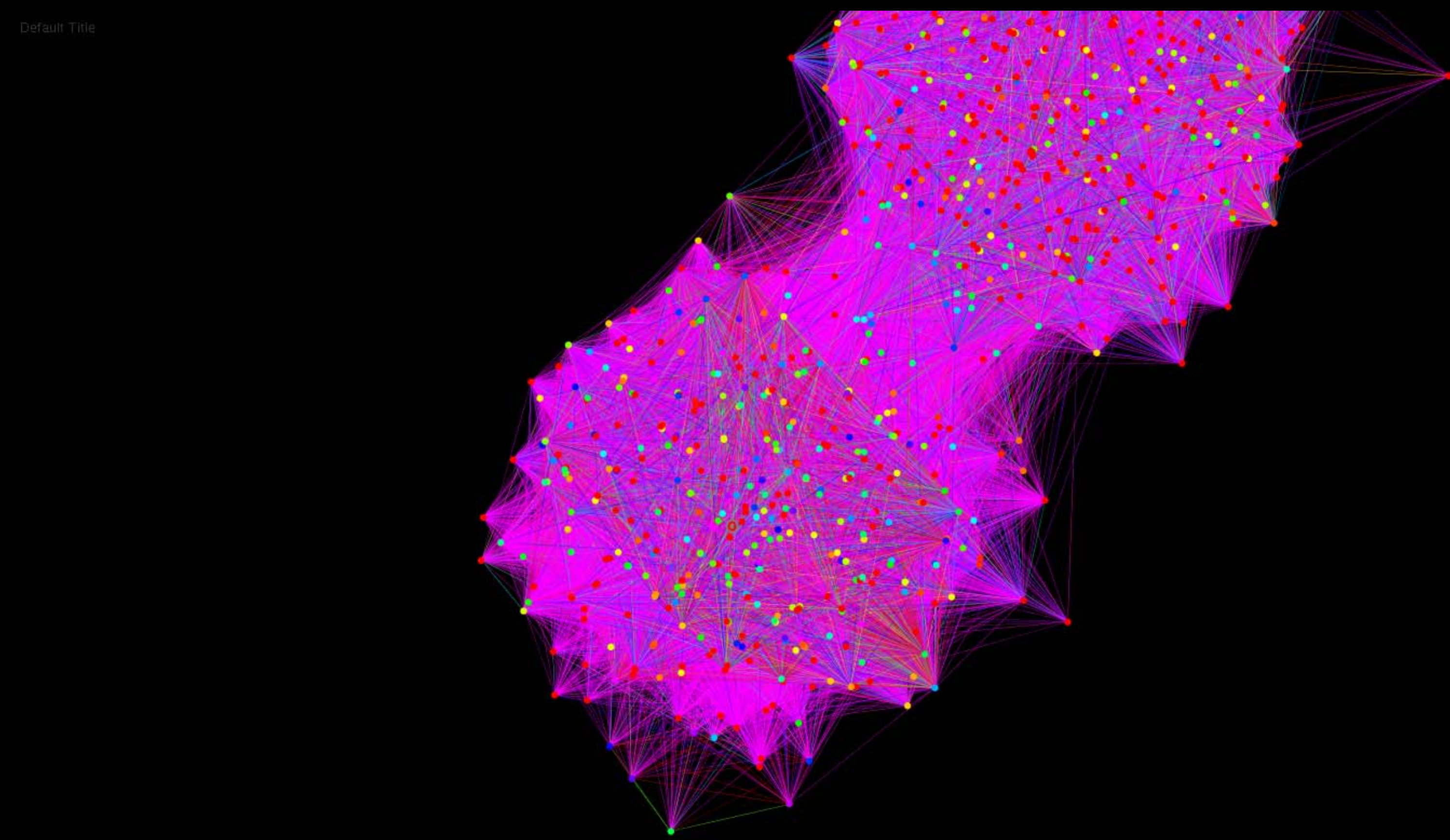
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

Graph Display

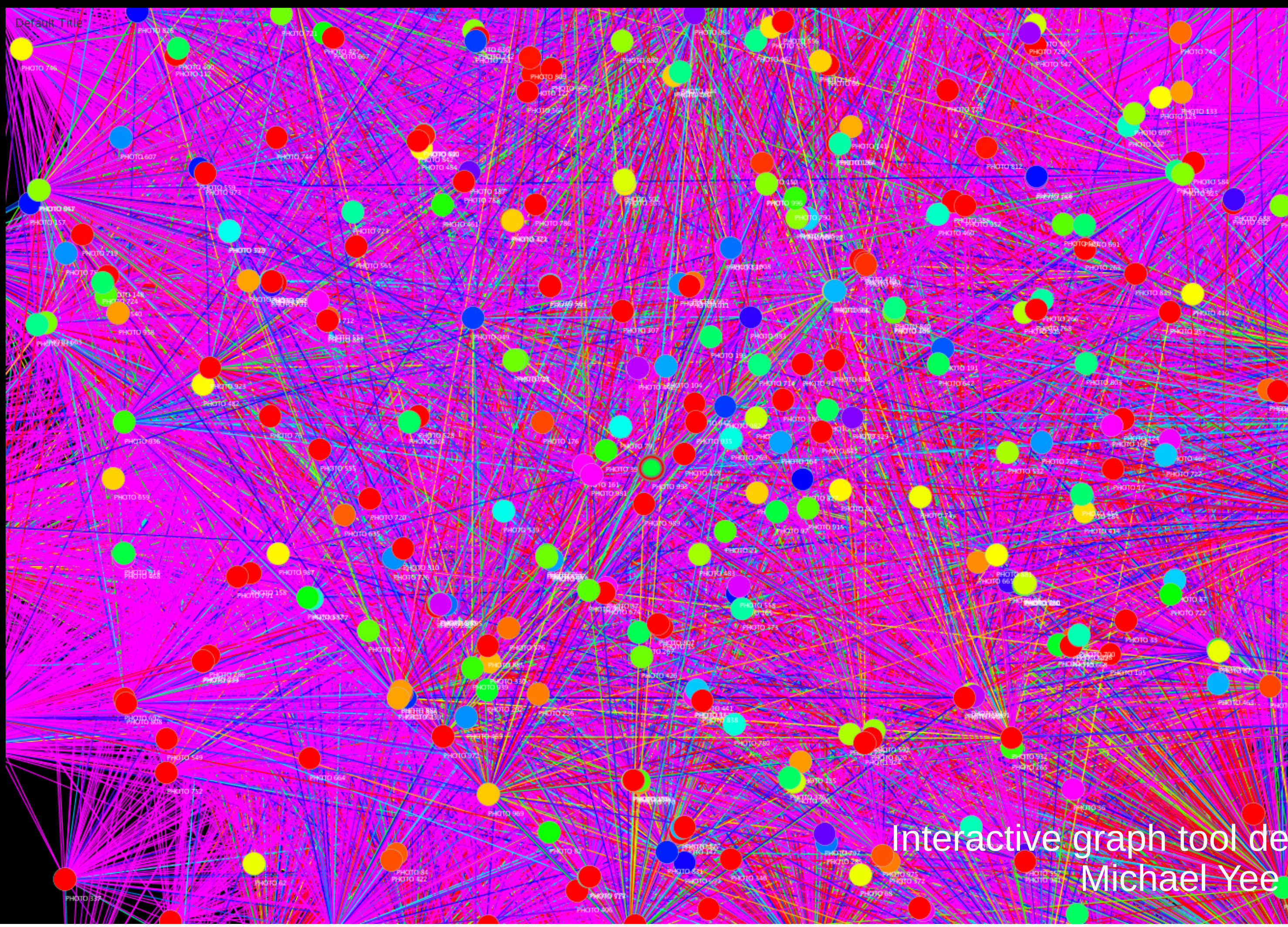
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by
Michael Yee

Graph Display

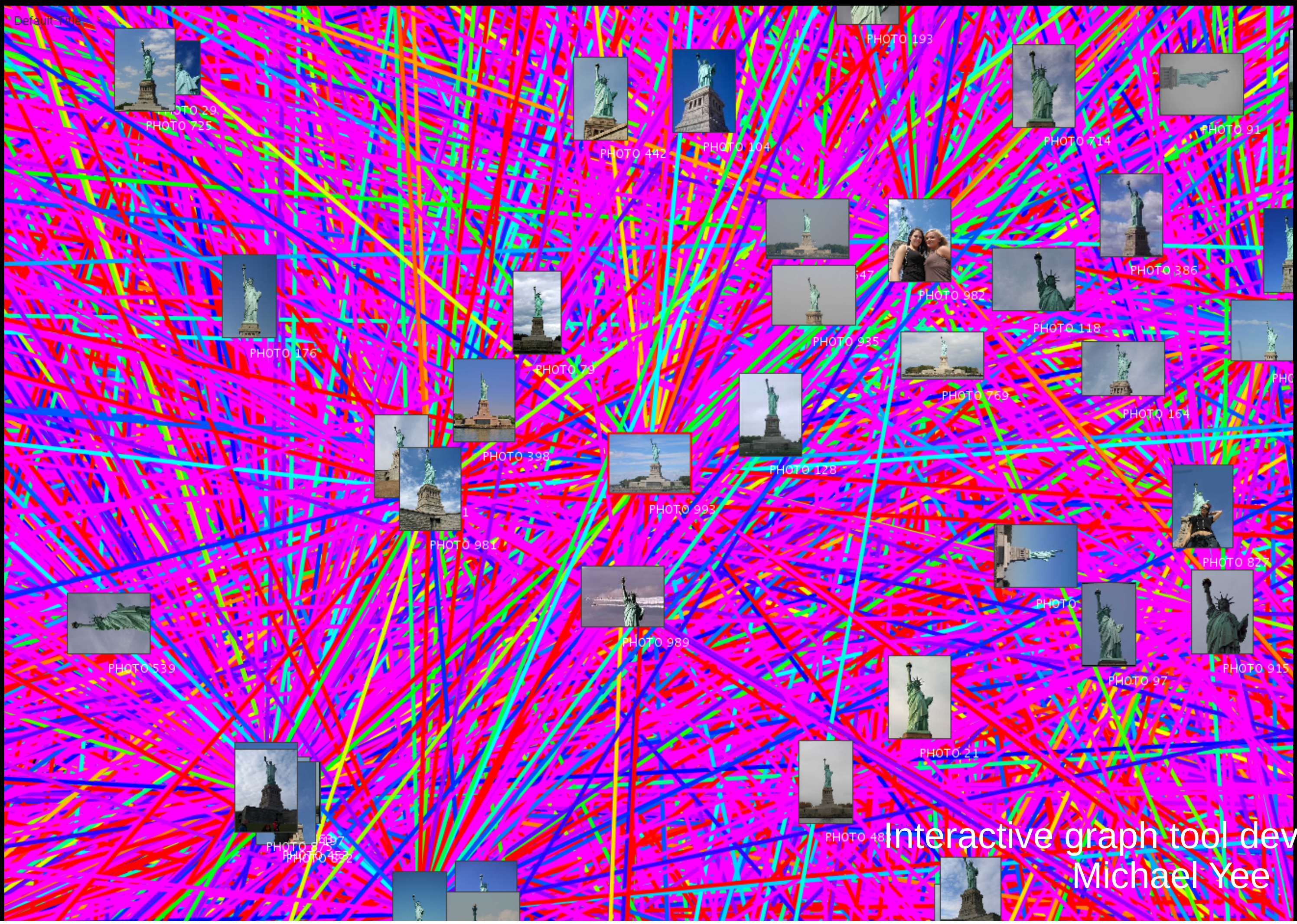
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

Graph Display

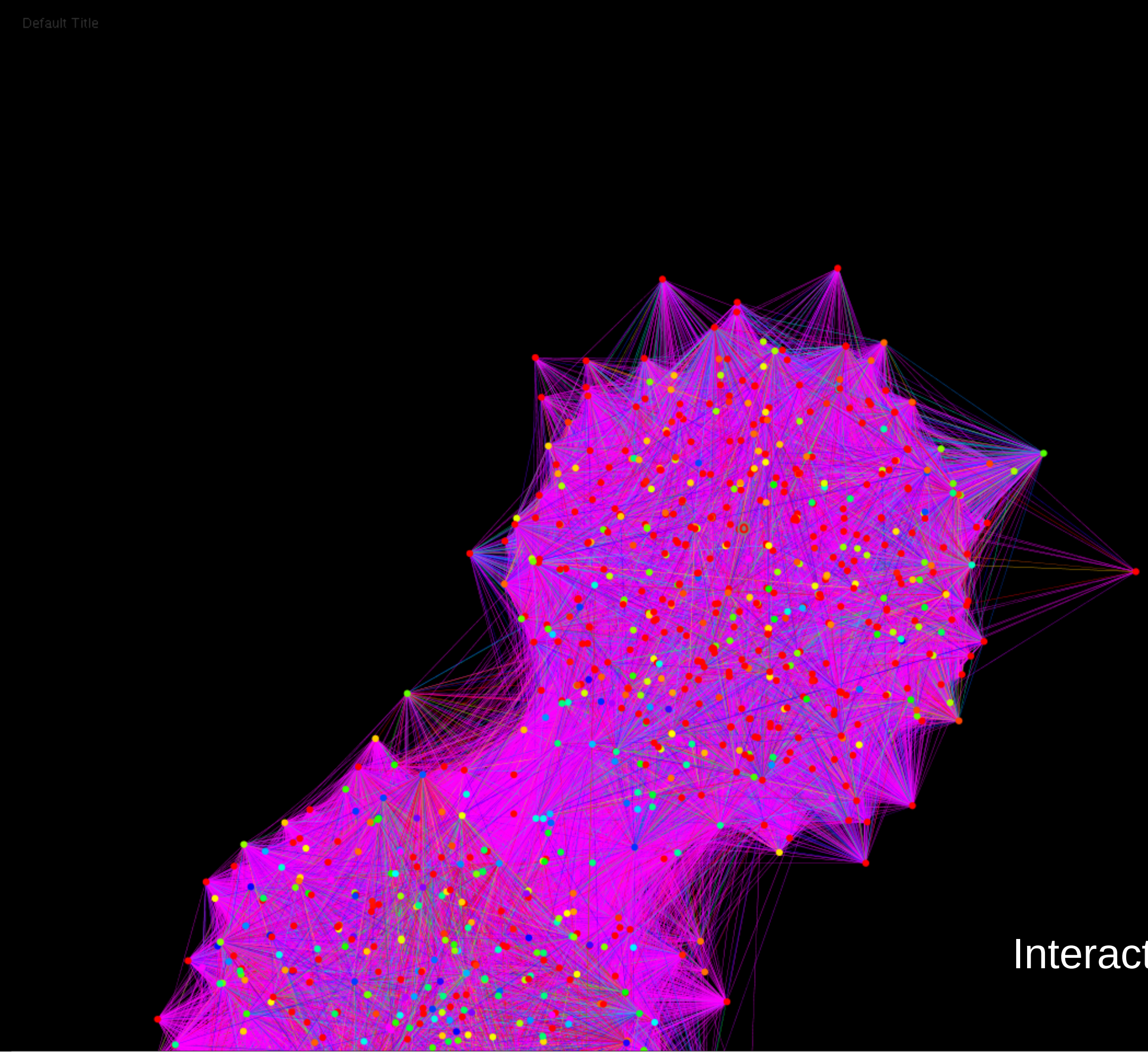
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

Graph Display

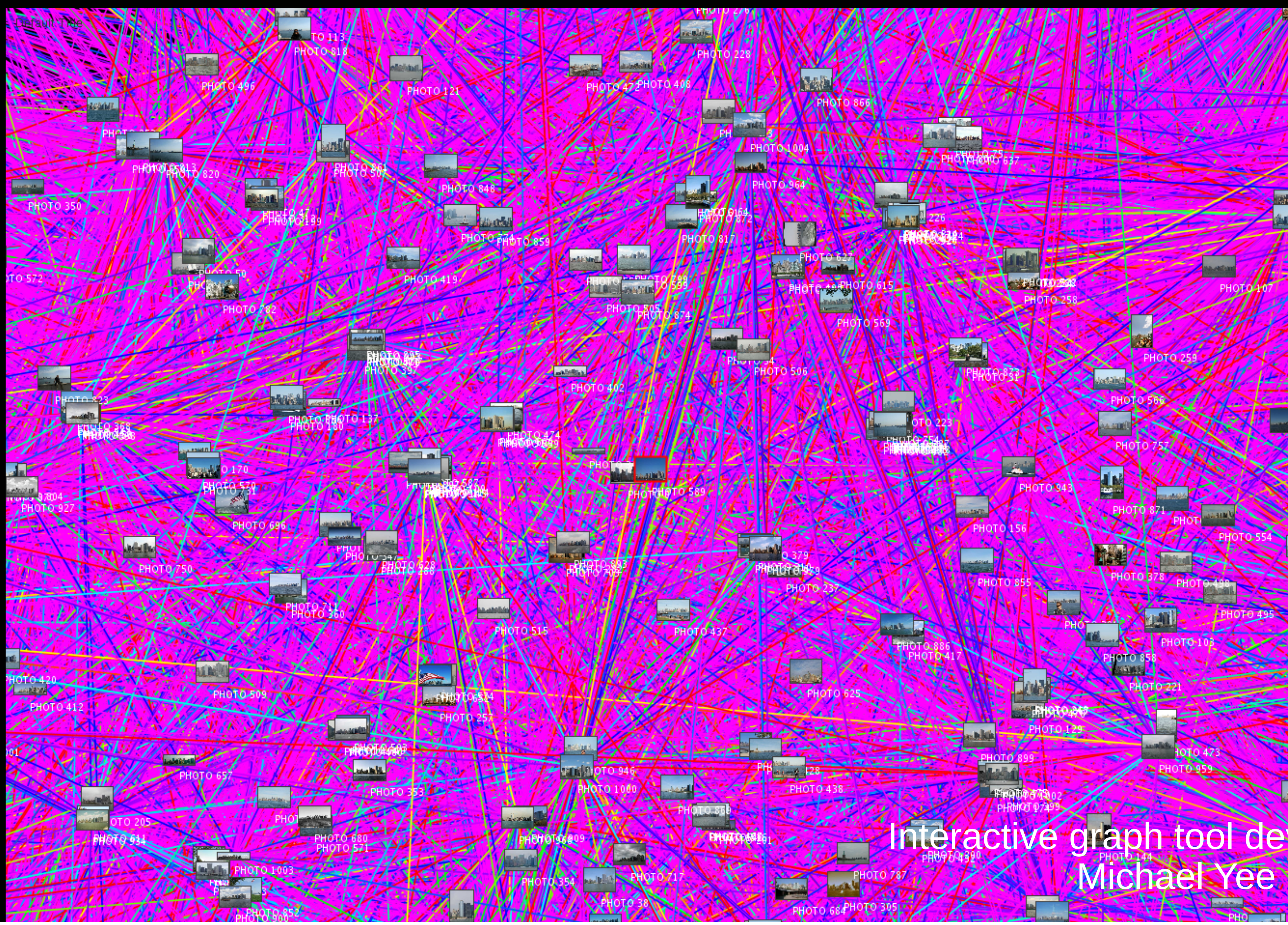
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

Graph Display

- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

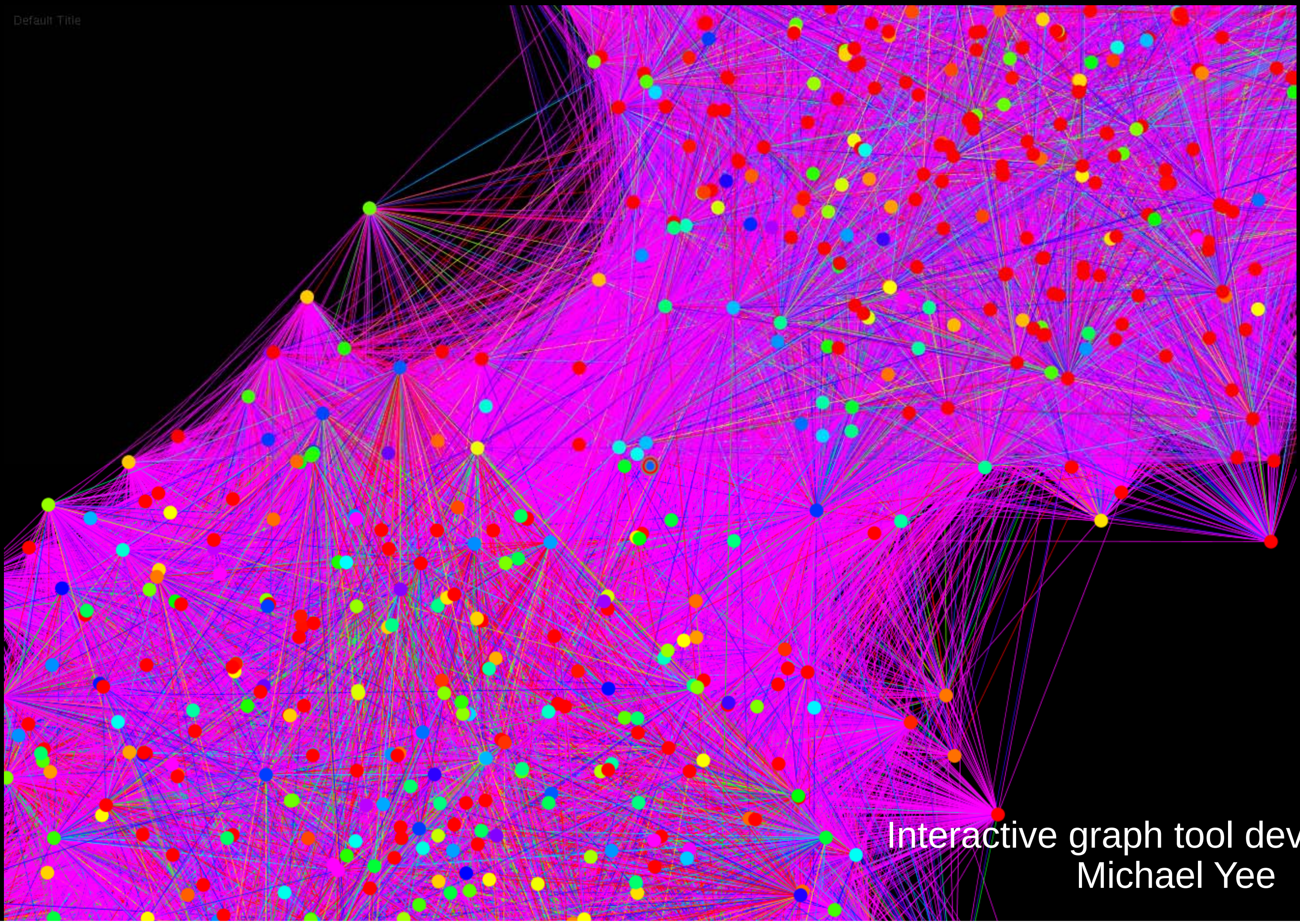
Graph Display

- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**

Interactive graph tool developed by Michael Yee

Graph Display

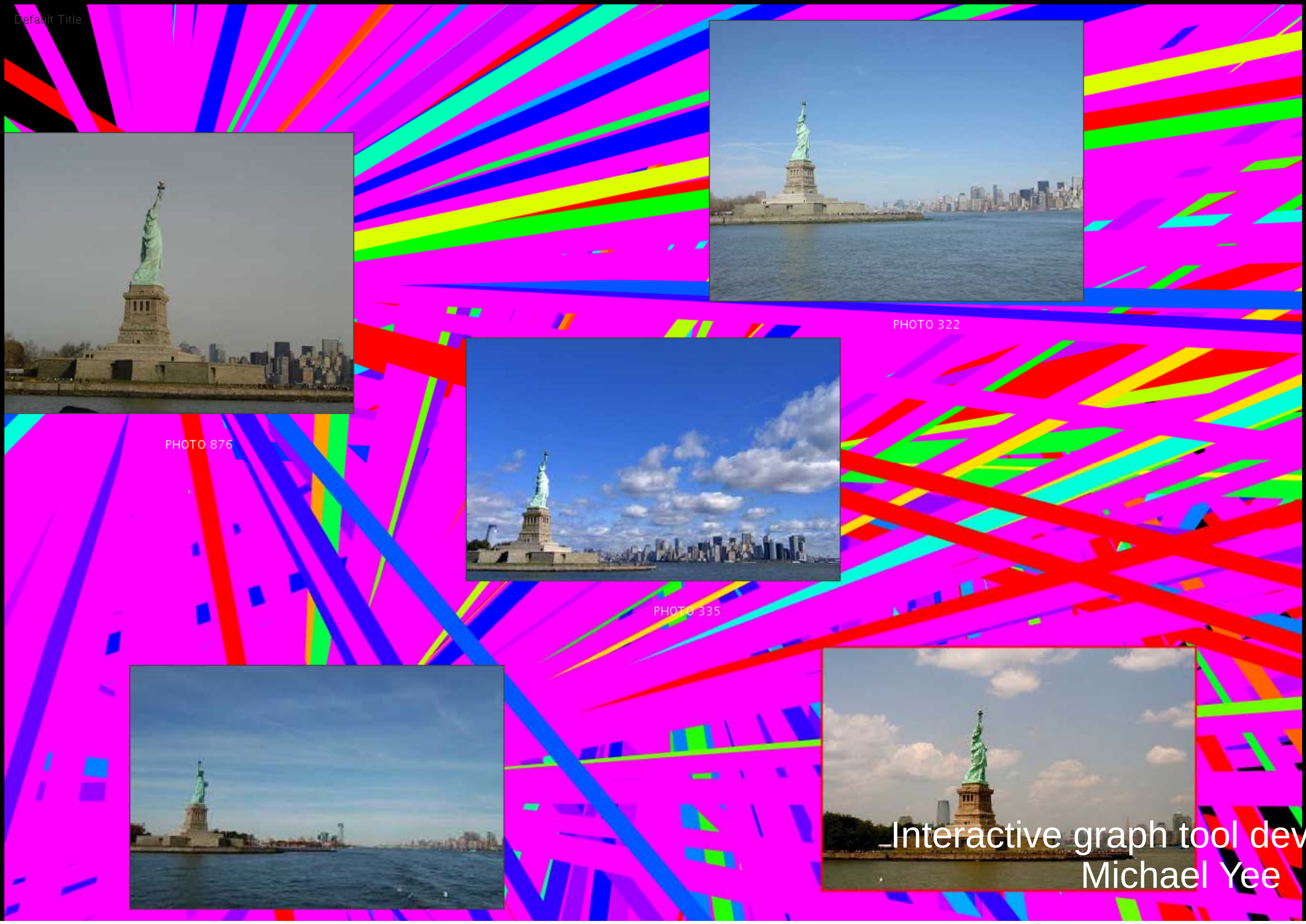
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

Graph Display

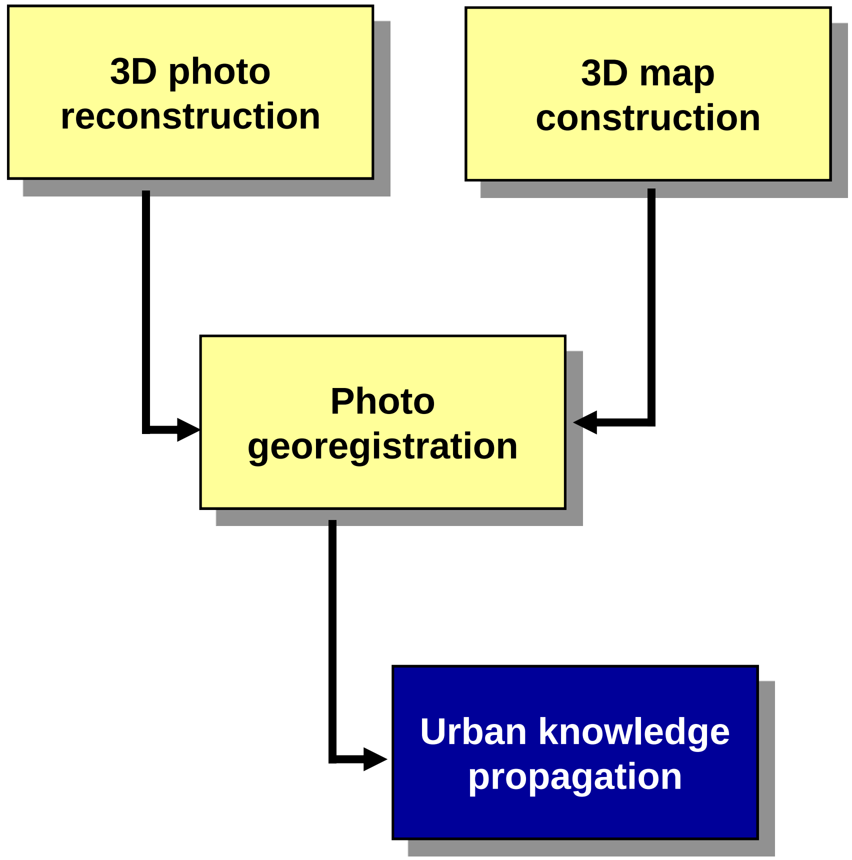
- **Capture topological relationships among reconstructed photos in a graph**
  - **Nodes colored according to covariance trace**
  - **Edges colored according to SIFT feature overlap**



Interactive graph tool developed by Michael Yee

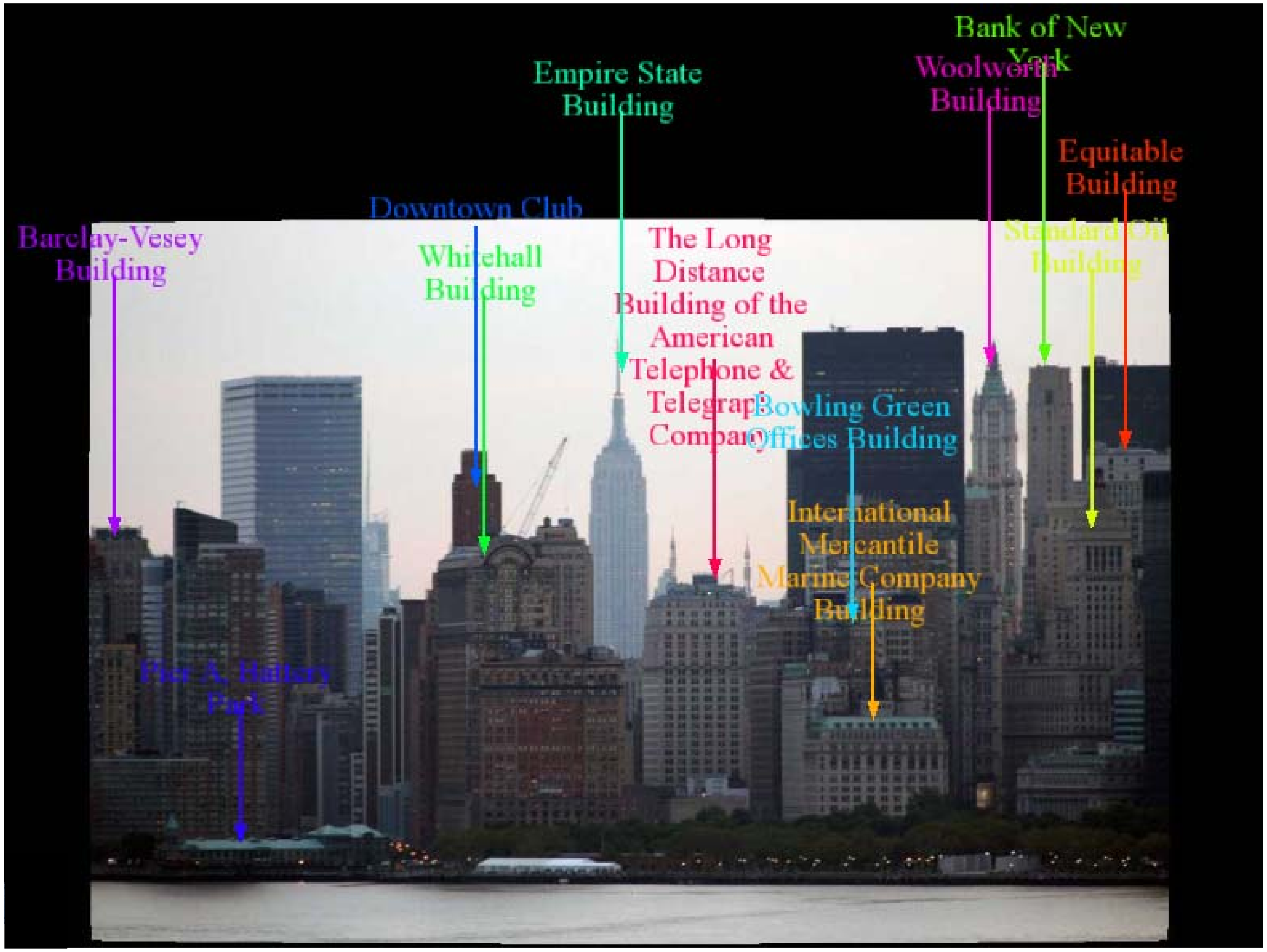
Outline

- 3D photo reconstruction
- 3D map construction
- Photo georegistration
- **Urban knowledge propagation**
  - **Automatic feature annotation**
  - **Image based querying**
- Ongoing & future work



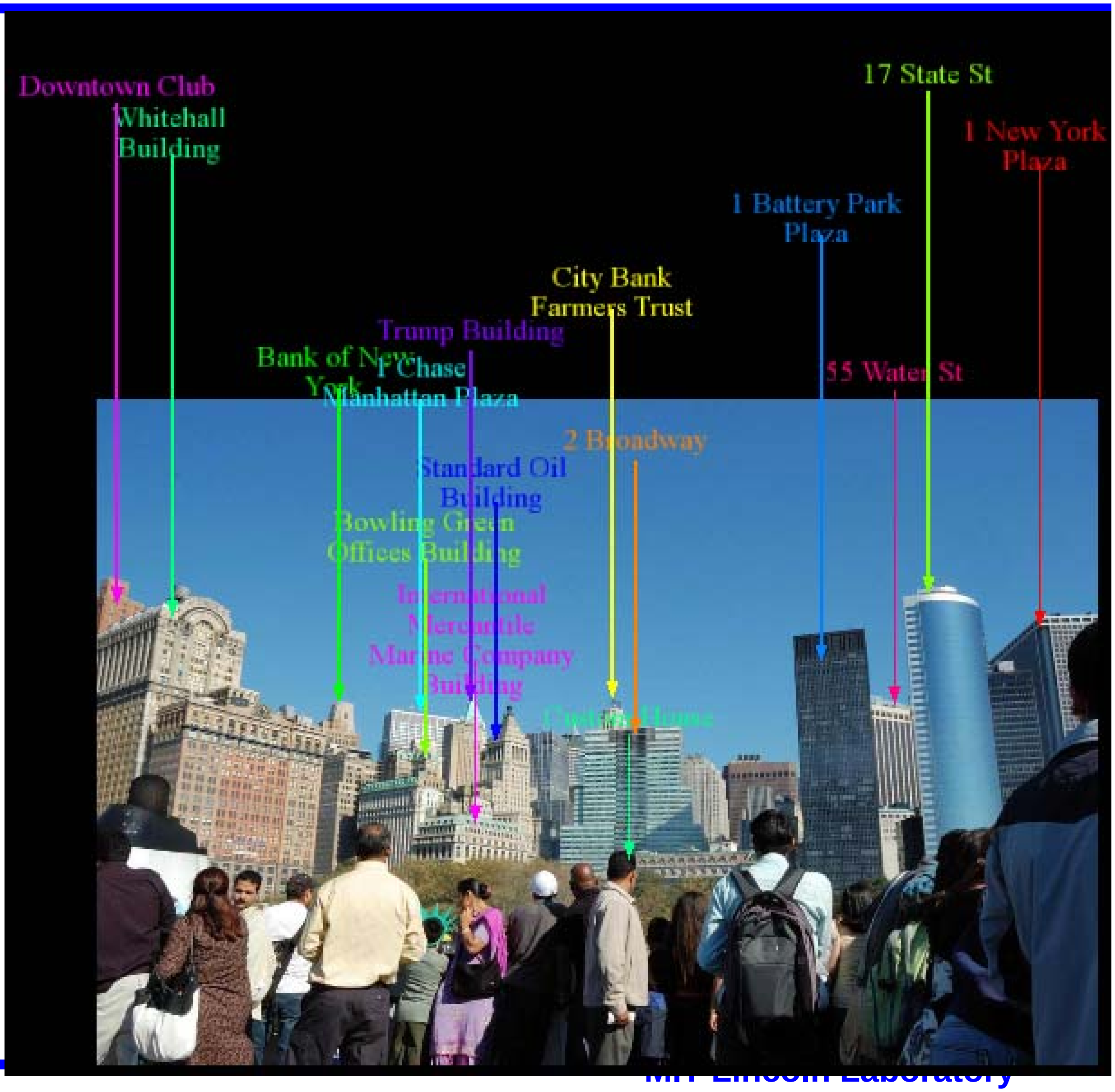
Automatic Feature Annotation

- **Project high-level knowledge from 3D world-space into georegistered 2D image planes to automatically label pixel clusters**
  - **Absolute geopositions**
  - **Names for buildings & streets**
  - **Landmarks & regions of interest**



Automatic Feature Annotation

- **Project high-level knowledge from 3D world-space into georegistered 2D image planes to automatically label pixel clusters**
  - **Absolute geopositions**
  - **Names for buildings & streets**
  - **Landmarks & regions of interest**



Automatic Feature Annotation

- **Project high-level knowledge from 3D world-space into georegistered 2D image planes to automatically label pixel clusters**
  - **Absolute geopositions**
  - **Names for buildings & streets**
  - **Landmarks & regions of interest**

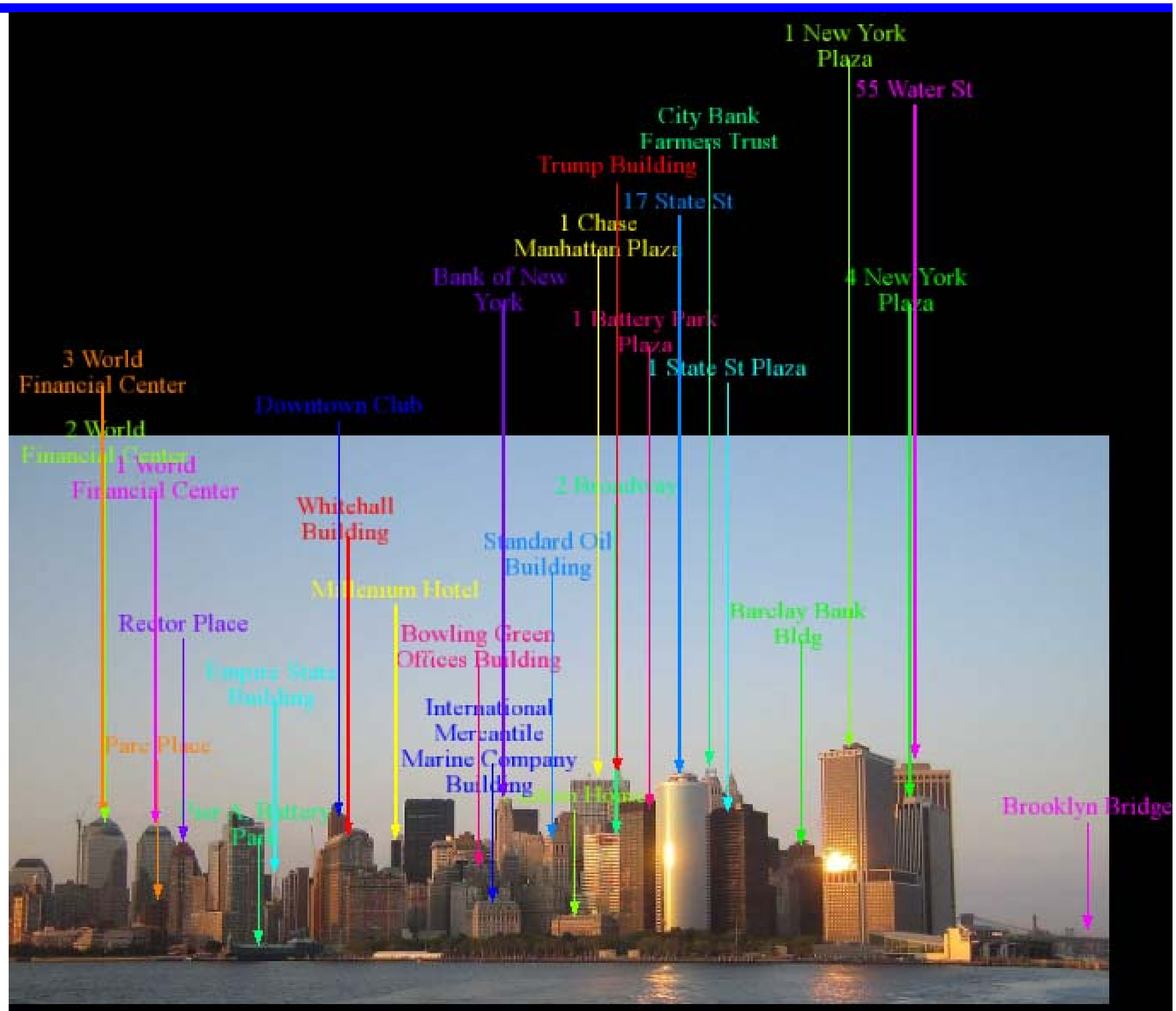
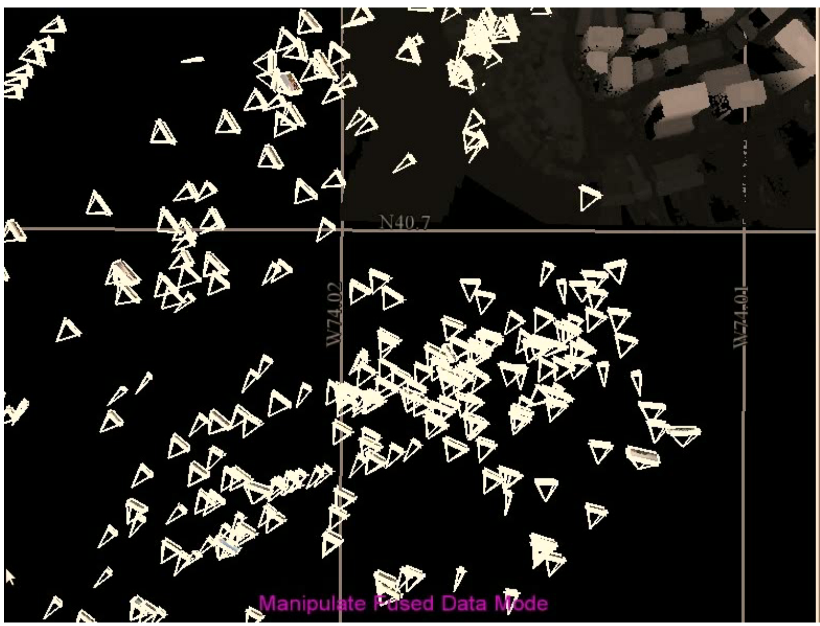


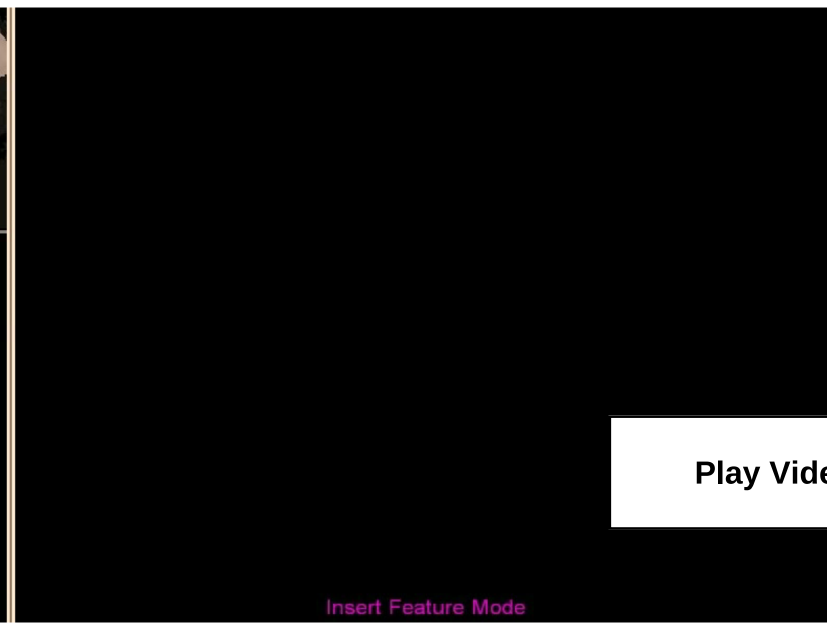
Image Based Querying

- **Compute 3D voxel corresponding to user-selected 2D photo pixel**
  - **Assume skyscraper walls missing from ladar data are well approximated by rooftop extrusion**
  - **Take into account line-of-sight occlusion by walls when ray tracing**
- **Reproject voxel back onto all photos in which it is visible**
- **Return urban features' ranges & altitudes above sea-level**

3D map



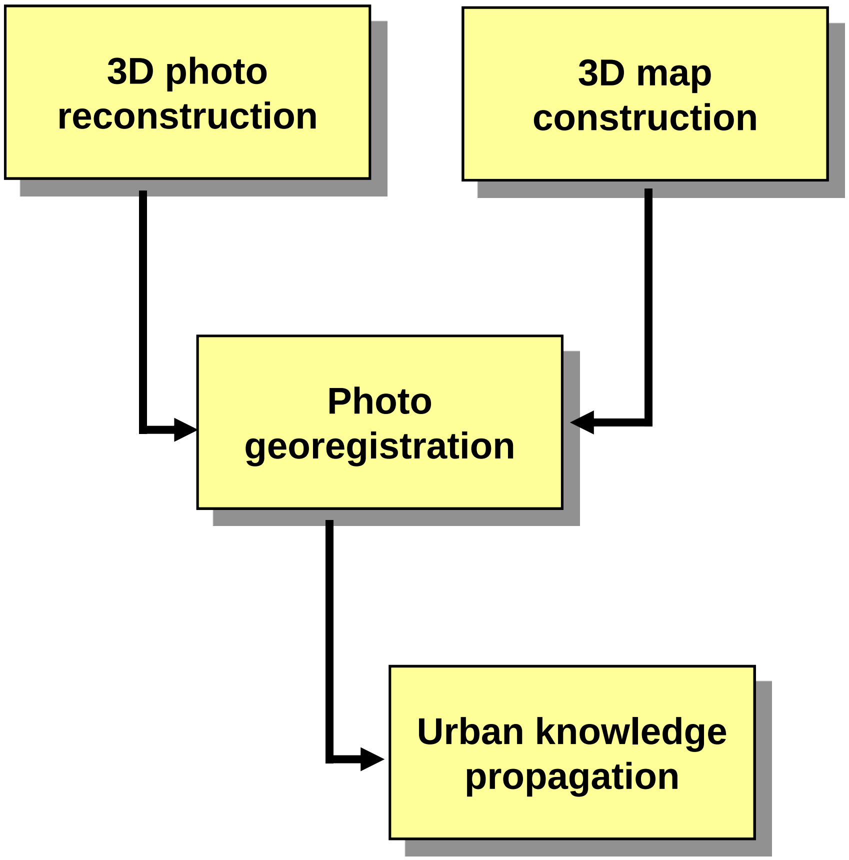
2D photo



Play Video

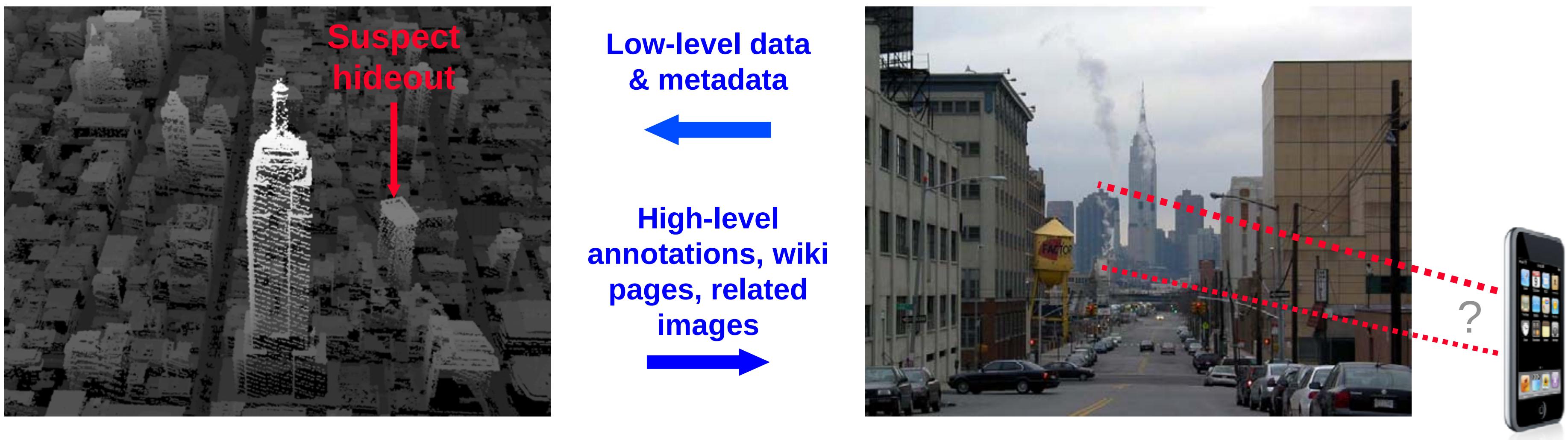
Outline

- 3D photo reconstruction
- 3D map construction
- Photo georegistration
- Urban knowledge propagation
- **Ongoing & future work**



Mobile Exploitation of Urban Imagery Archives

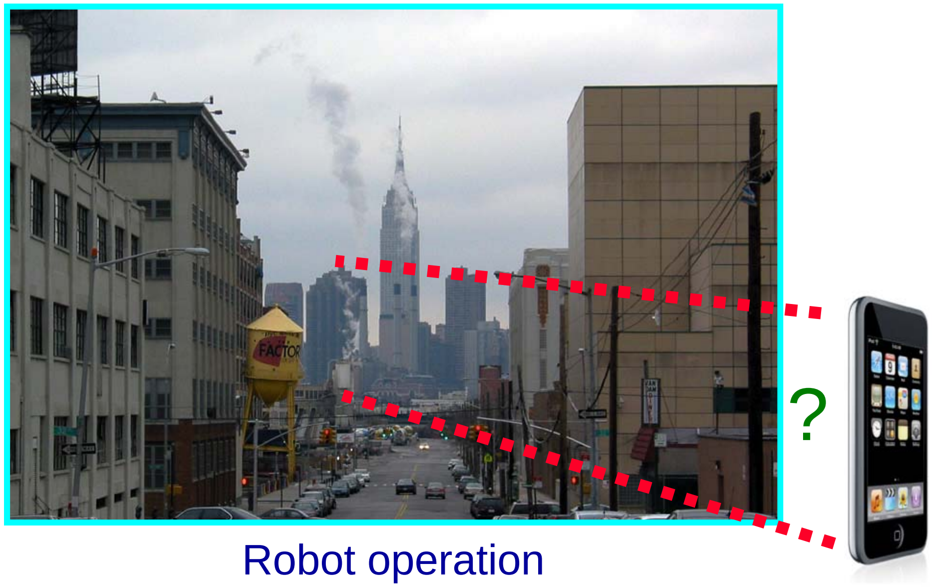
Server with virtual city map | Mobile clients probing real city



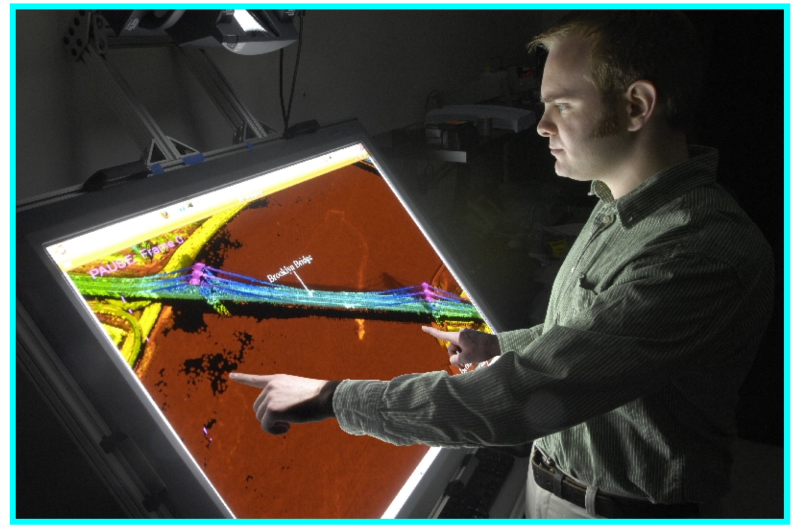
- **Mark up outputs from mobile cameras operating in complex cities**
  - **Automatically project information about buildings & streets into instantaneous image planes as users walk, drive or fly through cities**
  - **Input photo regions rather than text strings into database queries**
- **LLGrid reconstruction & registration of 30,000+ photos shot around MIT campus is currently underway**
  - **Data set includes significant urban canyon occlusion**
  - **Data set includes outdoor & indoor imagery**

Future Applications

Augmented reality

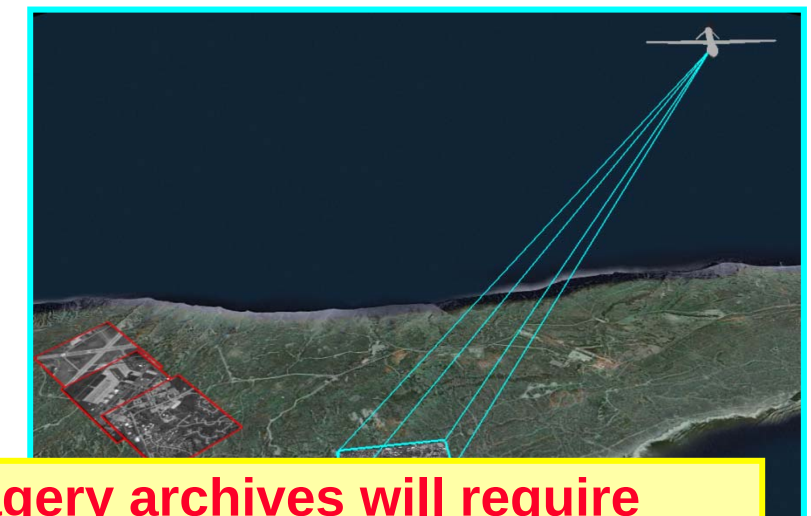


Urban mission planning



Robot operation

UAV video monitoring



Mobile exploitation of urban imagery archives will require onboard embedded computing

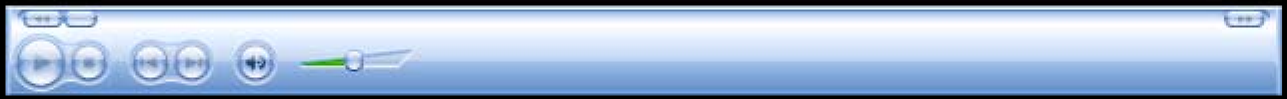
Summary

- 3D geometry provides an organizing principle for 2D urban imagery
  - 3D photo reconstruction yields structured output from unstructured input
  - Ladar map registration endows photos with absolute geo-coordinates
  - Geometrical organization enables intuitive navigation of large archives

- Urban-sized photo collections can only be processed on parallelized clusters
  - SIFT feature detection & matching are embarrassingly parallel
  - Structure from motion requires nonlinear optimization over millions of parameters
  - 1000+ NYC skyline & statue photos reconstructed in 4 hours on LLGrid

- 3D exploitation of large urban photo archives is just beginning and its directions to explore are growing
  - Incorporating sensor metadata into reconstructions
  - Identifying coverage gaps & skeletal sets via graph techniques
  - Extending current text-only searching to image-based querying

Fused 3D NYC Map [Cho, 2007]



Play Video