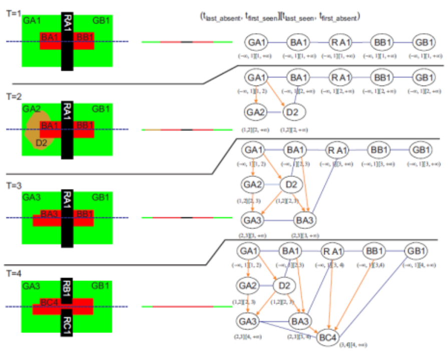
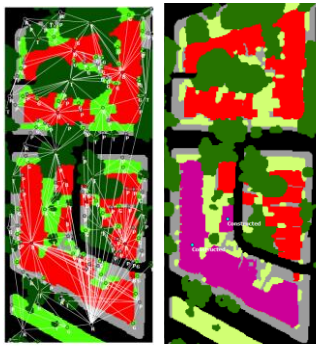
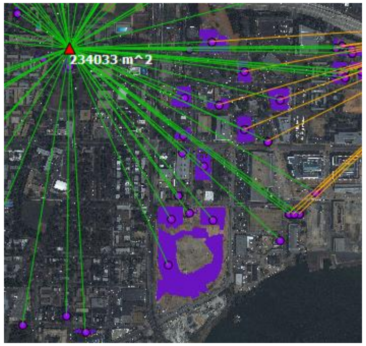
Exceptional service in the national interest

Sandia National Laboratories





UVM



DigitalGlobe©2006

Geospatial-Temporal Semantic Graphs for Wide-Area Search

Lawrence Livermore National Laboratory

Los Alamos NATIONAL LABORATORY EST.1943

Sandia National Laboratories

Randy C. Brost

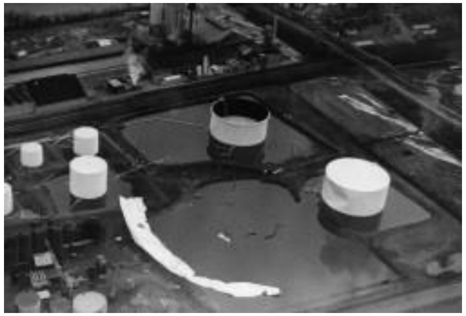
March 19, 2018

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

The Monongahela River Oil Spill

January 2, 1988:



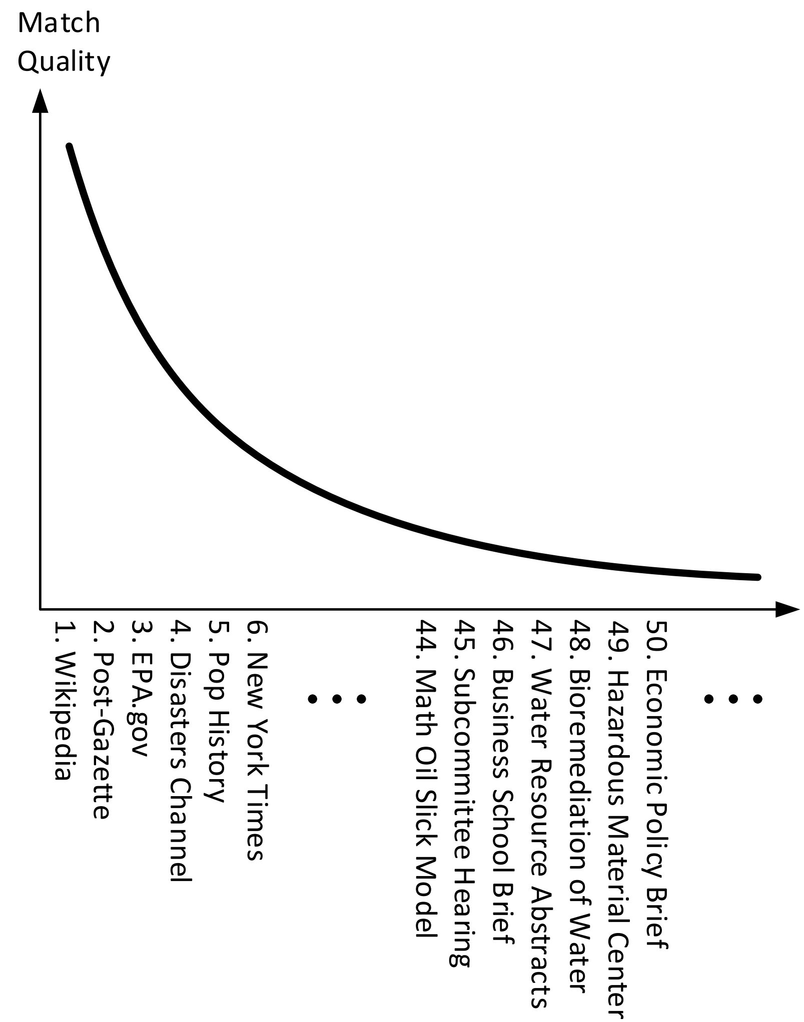
Google search terms “monongahela river oil spill”*

1. Ashland Oil Spill, Wikipedia.
2. The 1988 Monongahela Oil Spill, Pittsburgh Post-Gazette.
3. Pennsylvania Oil Spill Emergency Response, U.S. EPA.
4. Ashland Oil Spill Disaster Video, Disasters Channel.
5. “Disaster at Pittsburgh” 1988 Oil Tank Collapse, Pop History.
6. Monongahela River Oil Spill Threatens Water, New York Times.

⋮

44. Mathematical Model for Oil Slick Transport and Mixing in Rivers.
45. Monongahela River Oil Spill: Hearing Before the Subcommittee…
46. Harvard Business School Brief on Ashland Oil Spill and Response.
47. Selected Water Resource Abstracts, page 737.
48. Bioremediation of Water Areas Due to Oil Spills.
49. Ctr. for Haz. Matl. Research –Oil Spill Into the Monongahela River.
50. Economic and Policy Implications of the Ashland Oil Tank Collapse.

⋮



* Some titles paraphrased to fit.

† “The 1988 Monongahela oil spill,” Pittsburgh Post-Gazette, April 9, 2014. https://newsinteractive.post-gazette.com/thedigs/2014/04/09/monongahela-oil-spill

Problem Statement

Problem: Overhead imagery data is huge, overwhelming, and growing rapidly.

Goal: Construct a search engine for overhead image data.

- Allow user to specify desired goal.
- Automatic search.
- Return potential matches, ranked best first.

Hypothesis 1: A graph representing basic objects (buildings, trees,…), along with their attributes and relationships, can be used as a general-purpose search index.

Hypothesis 2: Given suitable overhead imagery, image processing can automatically find the required basic objects (buildings, water, trees,…).

Result Overview

Hypothesis 1: Graph can enable "search engine" for overhead imagery?

- Yes, very good results for some questions.
- Other questions limited.

Hypothesis 2: Automatic image processing can find basic objects (buildings,…)?

- Partially. Fully automated quality is reduced, and varies with feature type.
- Quality vs. temporal resolution issues.
- Rapid progress being made by other groups.

Scope:

- Several cases demonstrated, substantial range seems possible.
- Many will be out of scope.

Full details:

Brost, et al, "Geospatial-Temporal Semantic Graphs for Automated Wide-Area Search." Sandia Report SAND2017-8687, Sandia National Laboratories, August 2017.

Thanks

Tri-Lab PIs:

- Randy Roberts, LLNL
- Randy Brost, SNL
- Paul Pope, LANL
- Danny Rintoul, SNL

Universities:

- Jarlath O-Neil-Dunne, UVM
- Sean MacFaden, UVM
- Taylor Engel, UVM
- Mike Richardson, RIT

Industry:

- Brett Bader, DigitalGlobe
- Milan Karspeck, DigitalGlobe

Support:

- DOE NA-22 Office of Non-Proliferation
- Sandia LDRD Office

Team:

- Craig Blackhart, LANL
- Michelle Carroll, SNL
- Debbie Dennison, LLNL (now consulting)
- Luci Gaines, LANL (now at DigitalGlobe)
- John Goforth, LLNL
- Will McLendon, SNL
- Dan Morrow, SNL
- Jo Ellen Neuman, LLNL
- Ojas Parekh, SNL
- Andrew J. Patterson, SNL
- David Perkins, SNL
- Lakshman Prasad, LANL
- Jorge Roman, LANL (now retired)
- Shelly Spearing, LANL (now consulting)
- David Strip, SNL (now retired)
- Diane Woodbridge, SNL (now at USF)

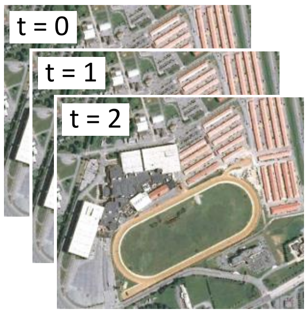
Overview

- Motivation.
- System Description.
- Examples.
- Evaluation.
- Summary.

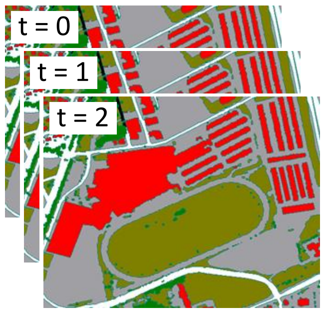
Overview

- Motivation.
- System Description.
  - Data.
  - Image Processing.
  - Geospatial-Temporal Graph Search.
- Examples.
- Evaluation.
- Summary.

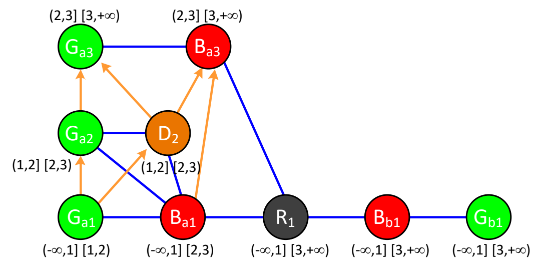
Geospatial-Temporal Graph Data Flow

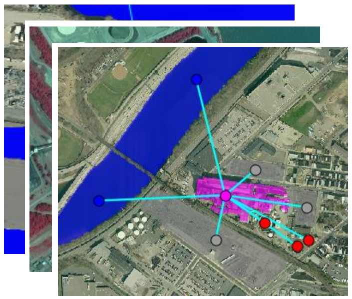


UVM



UVM





UVM/SNL

Remote Sensing → Region Classification → Spatiotemporal Graph → Matches

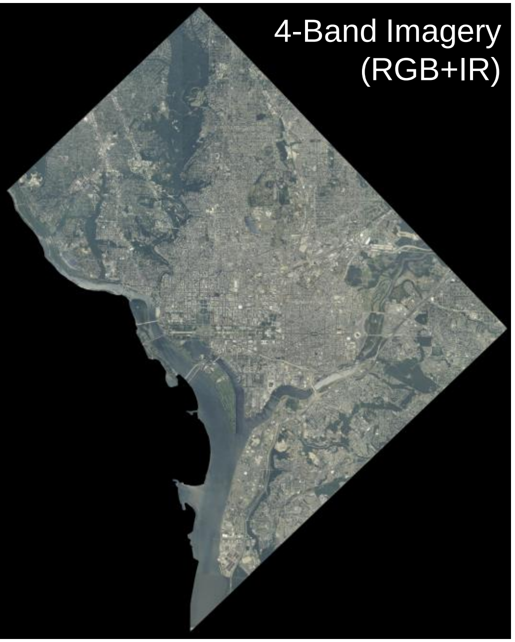
Spatial / Temporal / Spatiotemporal — Matches

Change Detection

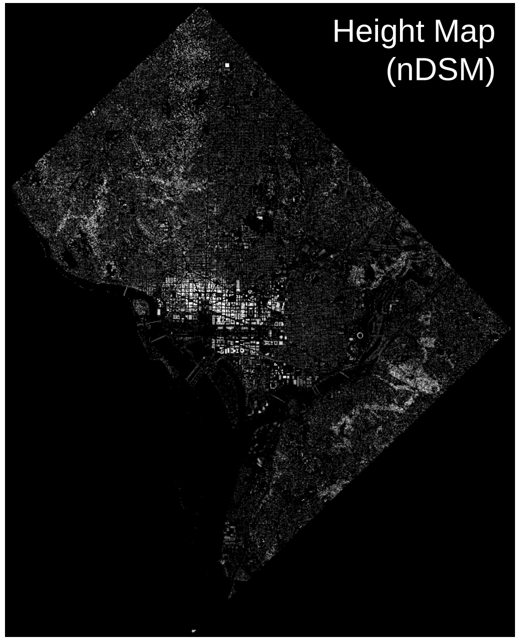
Query:
- Spatial parameters
- Temporal parameters
- Subgraph pattern
- Algorithm control

Ontology
(LANL, led by Paul Pope)

Input Data



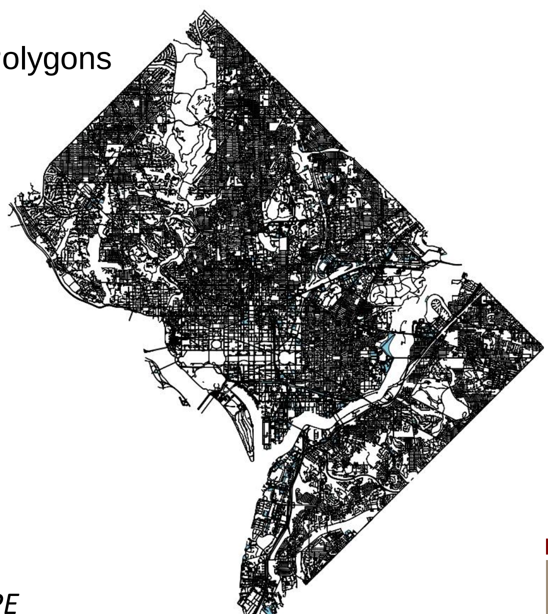
UVM



UVM

UVM

GIS Road Polygons
(optional)



Data provided by University of Vermont Spatial Analysis Laboratory.

Systematic Land Cover Classification

Steps:

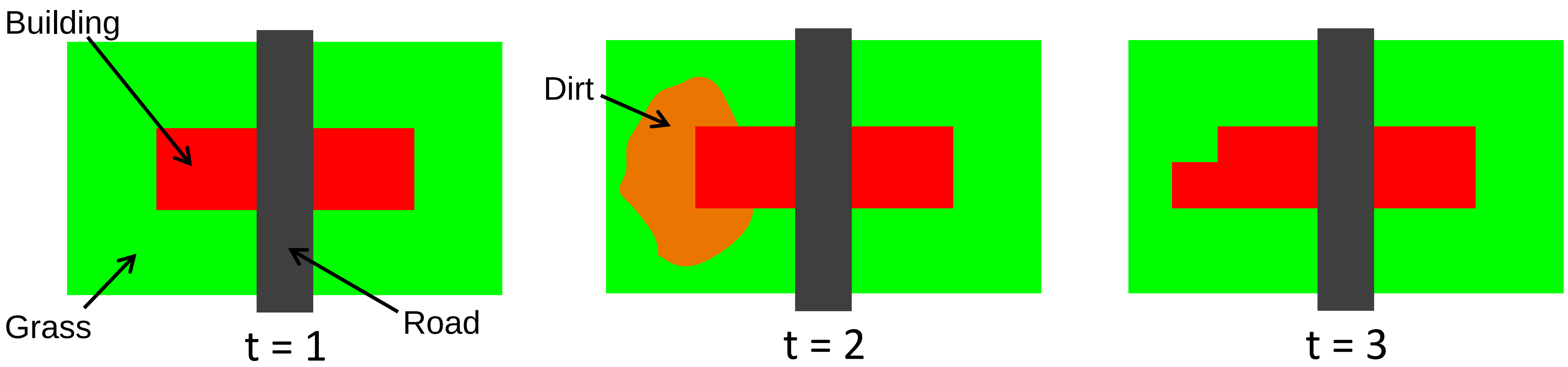
1. Assemble input:
   - RGB+IR imagery, near nadir.
   - LiDAR data.
   - GIS road centerlines or polygons (if available).
2. Construct normalized digital surface model (nDSM).
3. Using nDSM, split image into “Tall” vs. “Short.”
4. Using RGB+IR, split out “Tree” from Tall.
5. Using geometric morphology and LiDAR texture, split out “Building” from Tall.
6. Using morphology and LiDAR, split remaining Tall into “Wall” and “Other Structure.”
7. Using GIS road data, split out “Road” from Short.
8. Using RGB+IR, split remaining Short into
   “Grass/Shrub,” “Dirt,” “Water,” and “Other Paved.”
9. Output posterized file.

UUR SAND 2018-2803 PE

This is a refinement of O'Neil-Dunne, et al, “An Object-Based System for LiDAR Data Fusion and Feature Extraction,” Geocarto 2012.

Example Scenario

Data sequence:

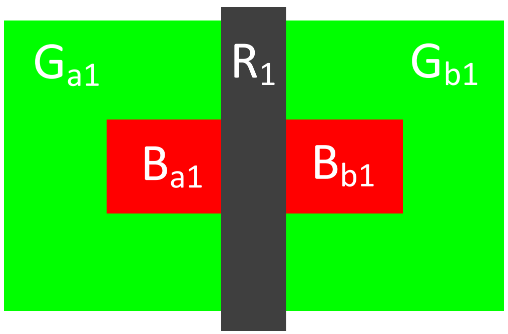


Data semantics:

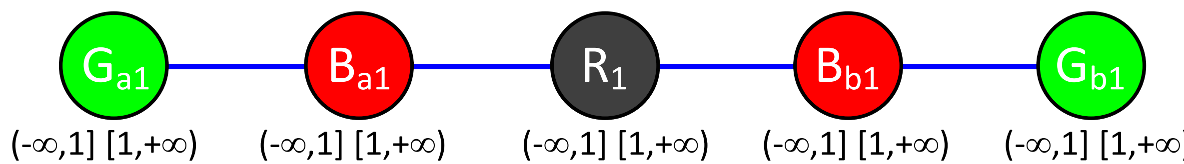
Building	B
Grass	G
Dirt	D
Road	R

StoredGraph Construction

Start with land cover #1:



t = 1



Each node has several attributes:
area, centroid, moments,
eccentricity, orientation,
bounding box, etc...

Legend:

G Durable nodes

Adjacency edges

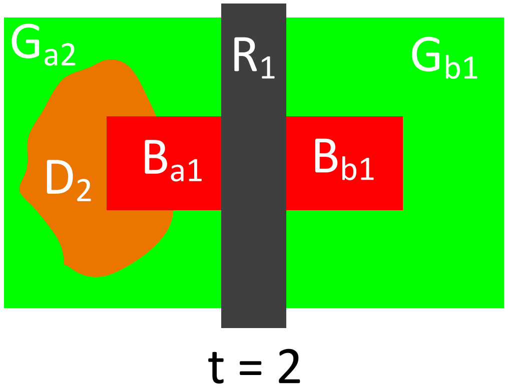
Data semantics:

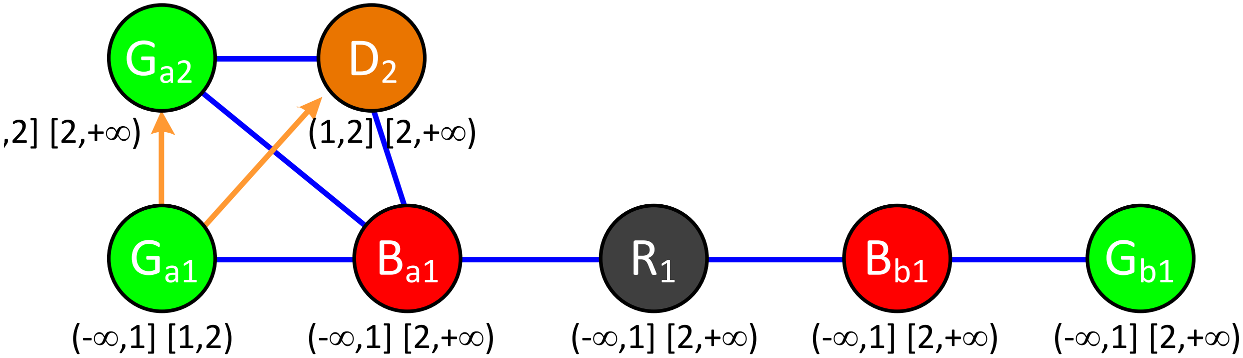
Building	B
Grass	G
Dirt	D
Road	R

* Road↔Grass edges omitted for clarity throughout.

StoredGraph Construction

Add land cover #2:





Legend:

Durable nodes

Adjacency edges

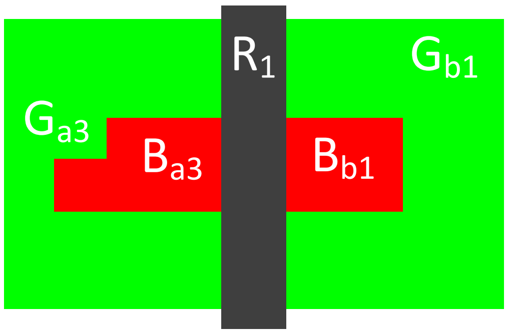
Change edges

Data semantics:

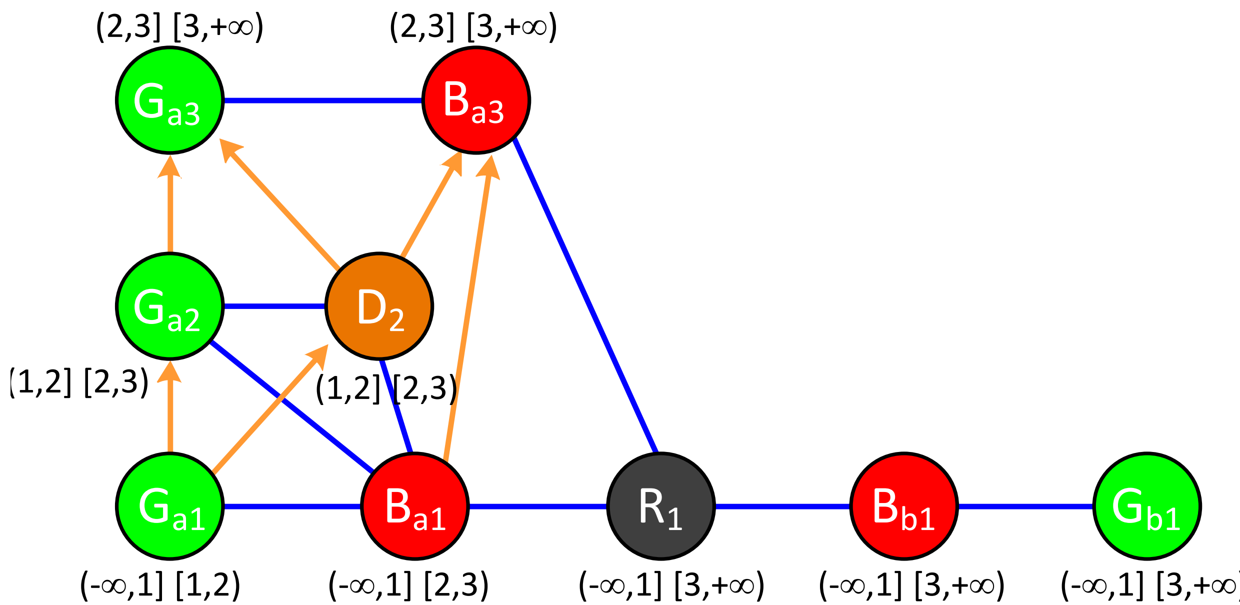
Building	B
Grass	G
Dirt	D
Road	R

StoredGraph Construction

Add land cover #3:



t = 3



Legend:

- G Durable nodes
- Adjacency edges
- Change edges

Data semantics:

Building	B
Grass	G
Dirt	D
Road	R

Query Definition

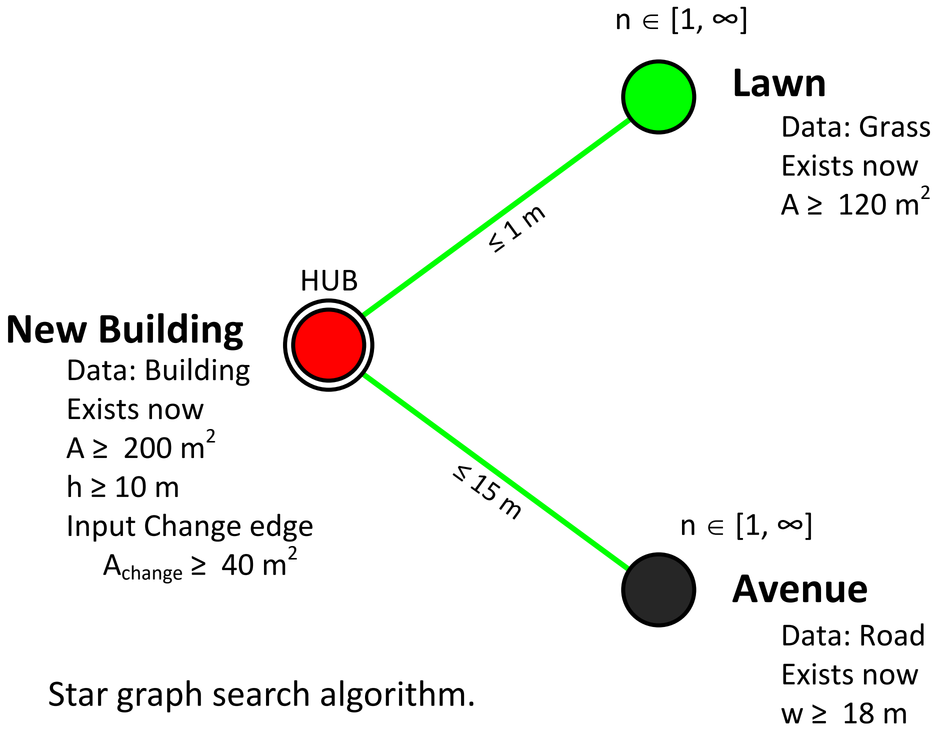
Question: Where is a new luxury office building?

Question semantics:

New Building	N
Lawn	L
Avenue	A

Data semantics:

Building	B
Grass	G
Dirt	D
Road	R

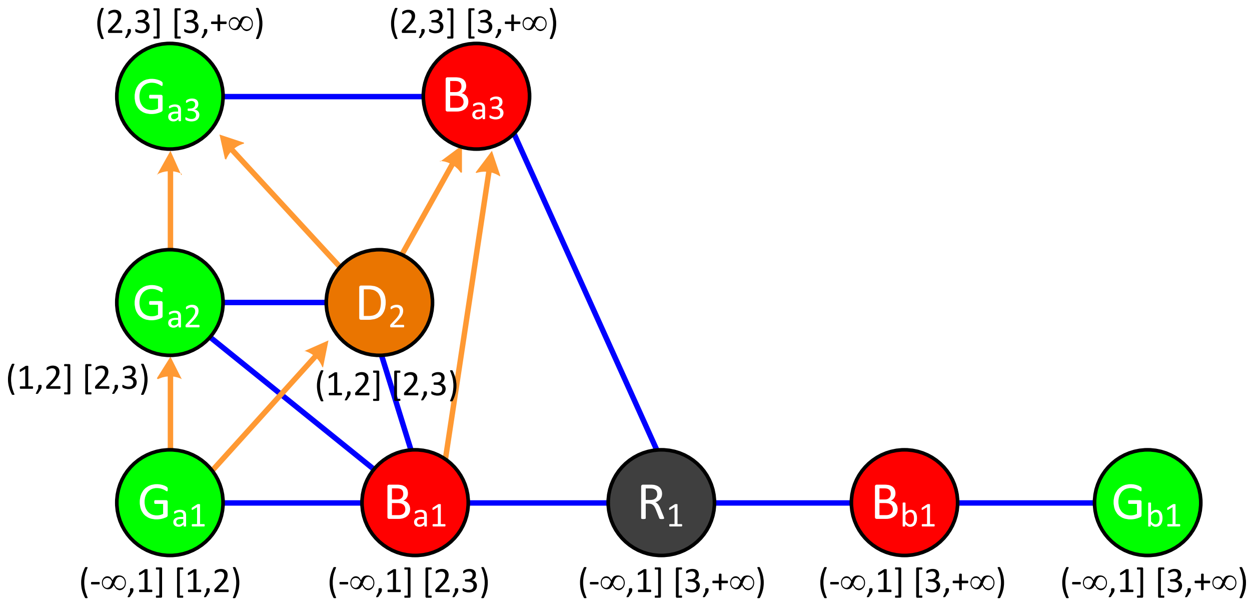


Star graph search algorithm.

This configuration is consistent with a luxury office building.

Search Graph Source: StoredGraph

Available StoredGraph:



Legend:

- G — Durable nodes
- Adjacency edges
- Change edges

Data semantics:

Building	B
Grass	G
Dirt	D
Road	R

Constructed SearchGraph

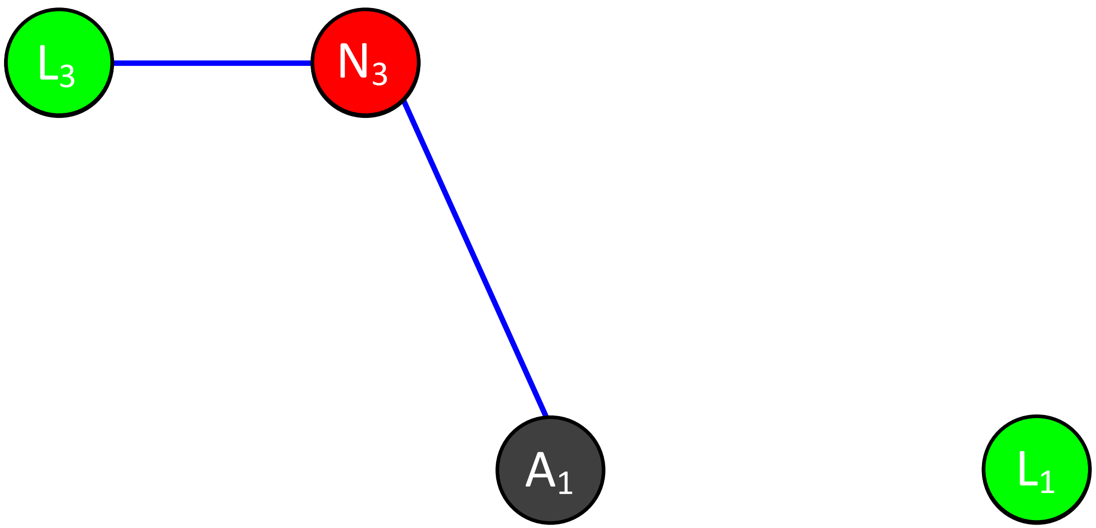
After adding edges:

Signature nodes:

- New Building
- Lawn
- Avenue

Signature edges:

- New Building ↔ Lawn
- New Building ↔ Avenue

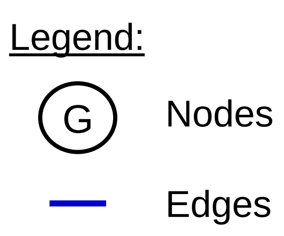


Question semantics:

New Building	N
Lawn	L
Avenue	A

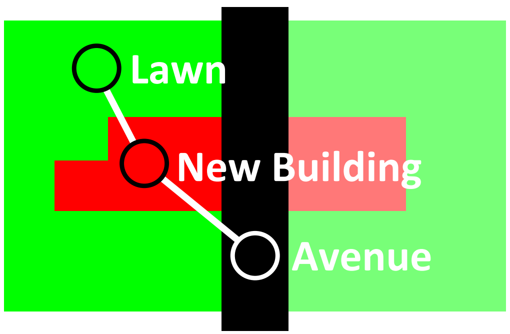
SearchGraph only contains elements relevant to the question.

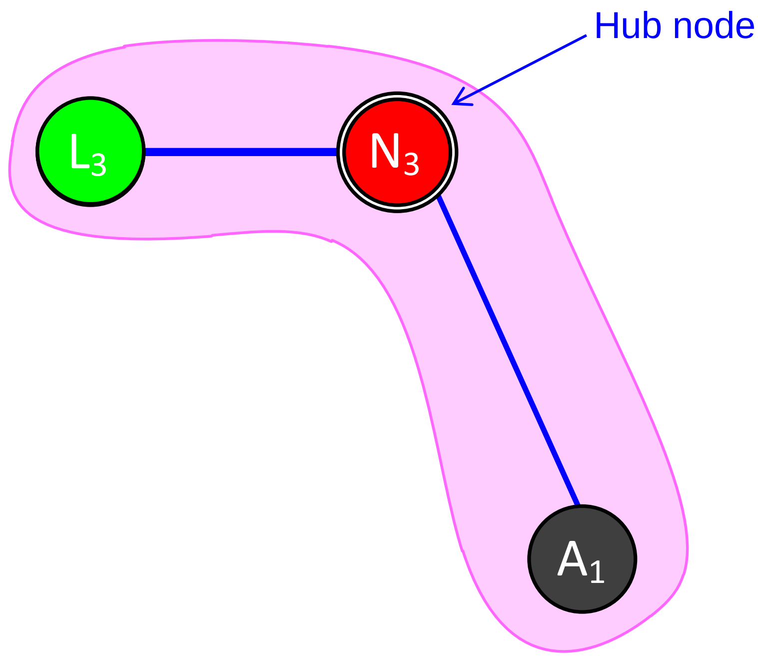
Semantics are now in terms of the question.



Search Result

Matches found:



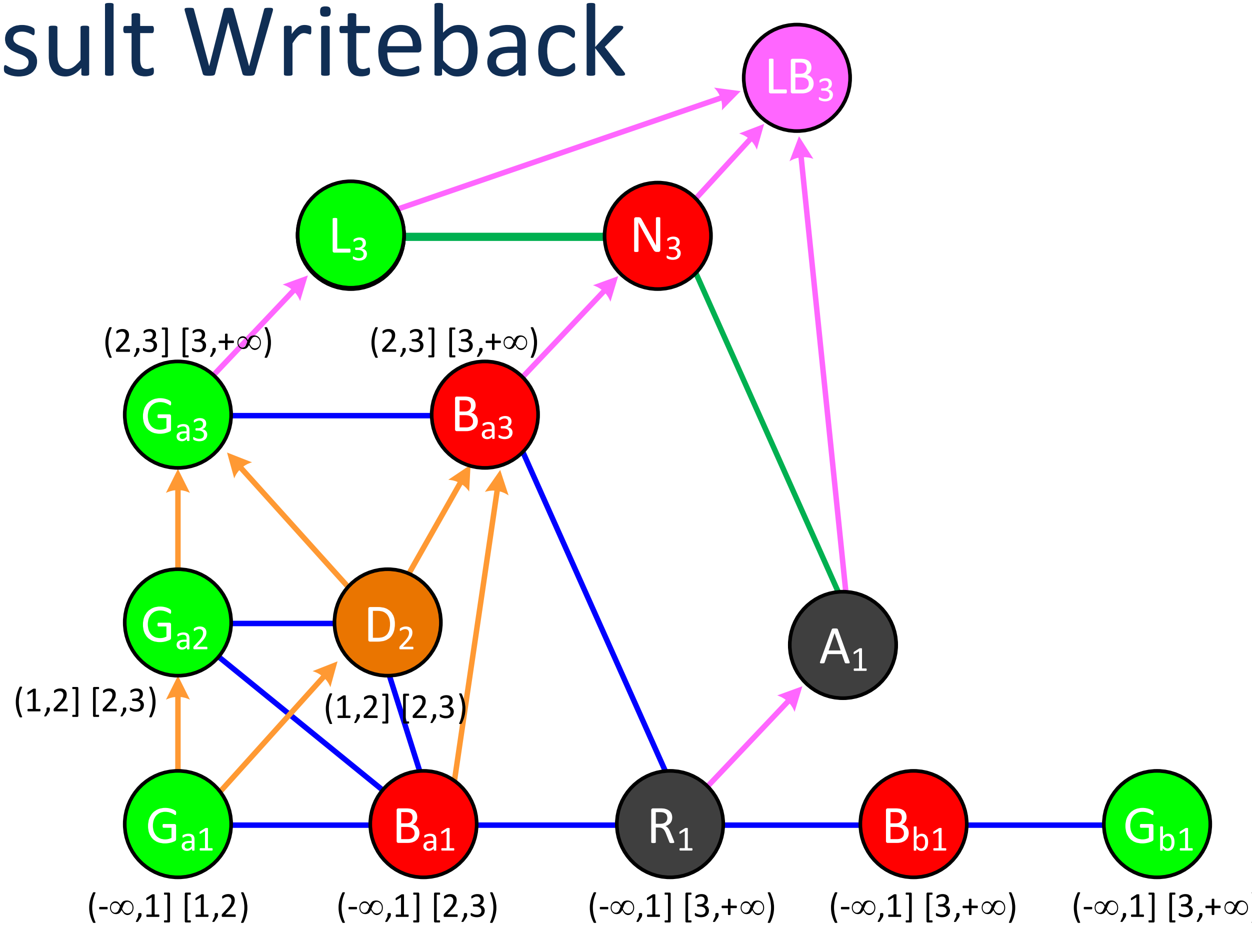


Remarks:

- This match is consistent with hypothesized signature.
- It does not "prove" this is a luxury office building --- this template might also match a movie star's mansion.

Search Result Writeback

Final StoredGraph:



Extended semantics:

New Building	N
Lawn	L
Avenue	A
Luxury Building	LB

Data semantics:

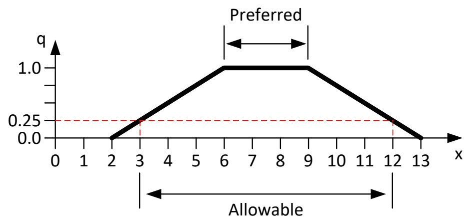
Building	B
Grass	G
Dirt	D
Road	R

Legend:

- G Durable nodes
- Adjacency edges
- Change edges
- Component-Of edges
- Within-Match edges

Quality Score Calculation

Attribute quality:



$$d_{ij} = \frac{1}{q_{ij}} - 1$$

Node quality:

$$D_i = \frac{\sum_j d_{ij}}{n_i}$$

$$q_i = \frac{1}{1 + D_i}$$

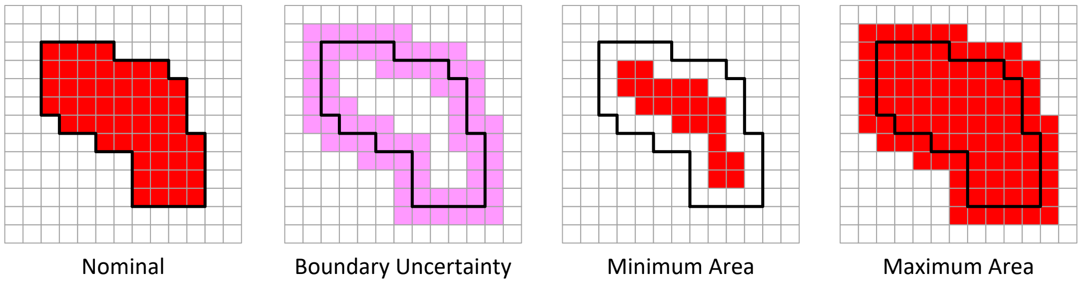
Match quality:

(If all nodes required*)

$$D_{req} = \frac{\sum_{req\ i} D_i}{n_{req}}$$

$$q_{match} = \frac{1}{1 + D_{req}}$$

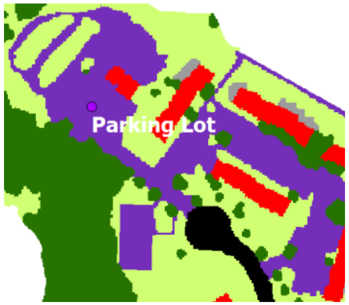
Variability model:



* Optional elements, variable topology require special handling.
See Brost, et al, "Geospatial-Temporal Semantic Graphs for Automated Wide-Area Search." Sandia Report SAND2017-8687, August 2017.

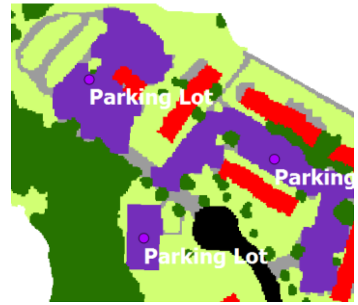
Segment Transformation

Undersegmentation:



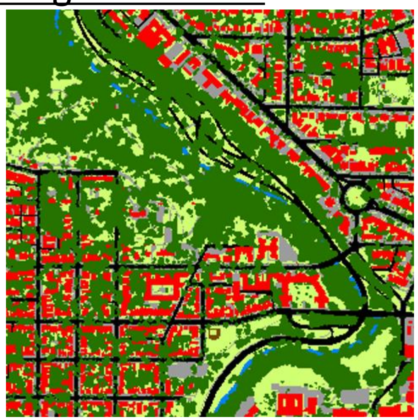
UVM/SNL

Solution: “Chunk” by shrink/grow



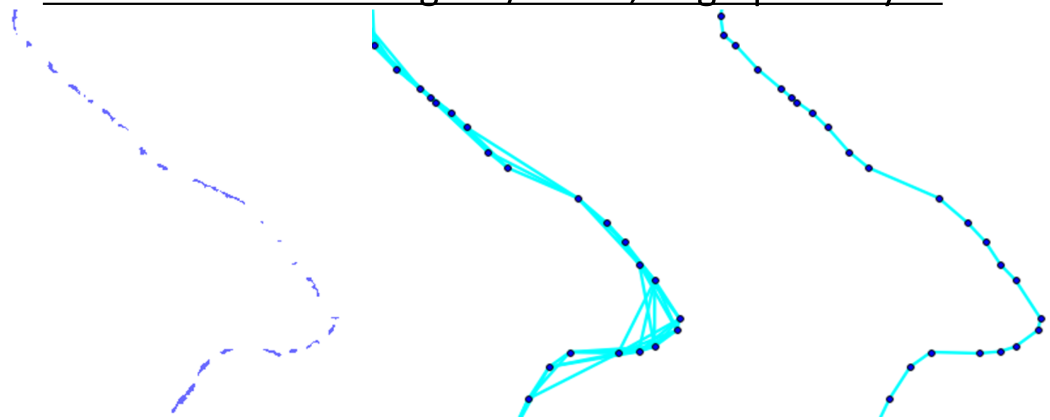
UVM/SNL

Oversegmentation:



UVM

Solution: Connect via grow/shrink, or graph analysis



Path routing:

UVM/SNL

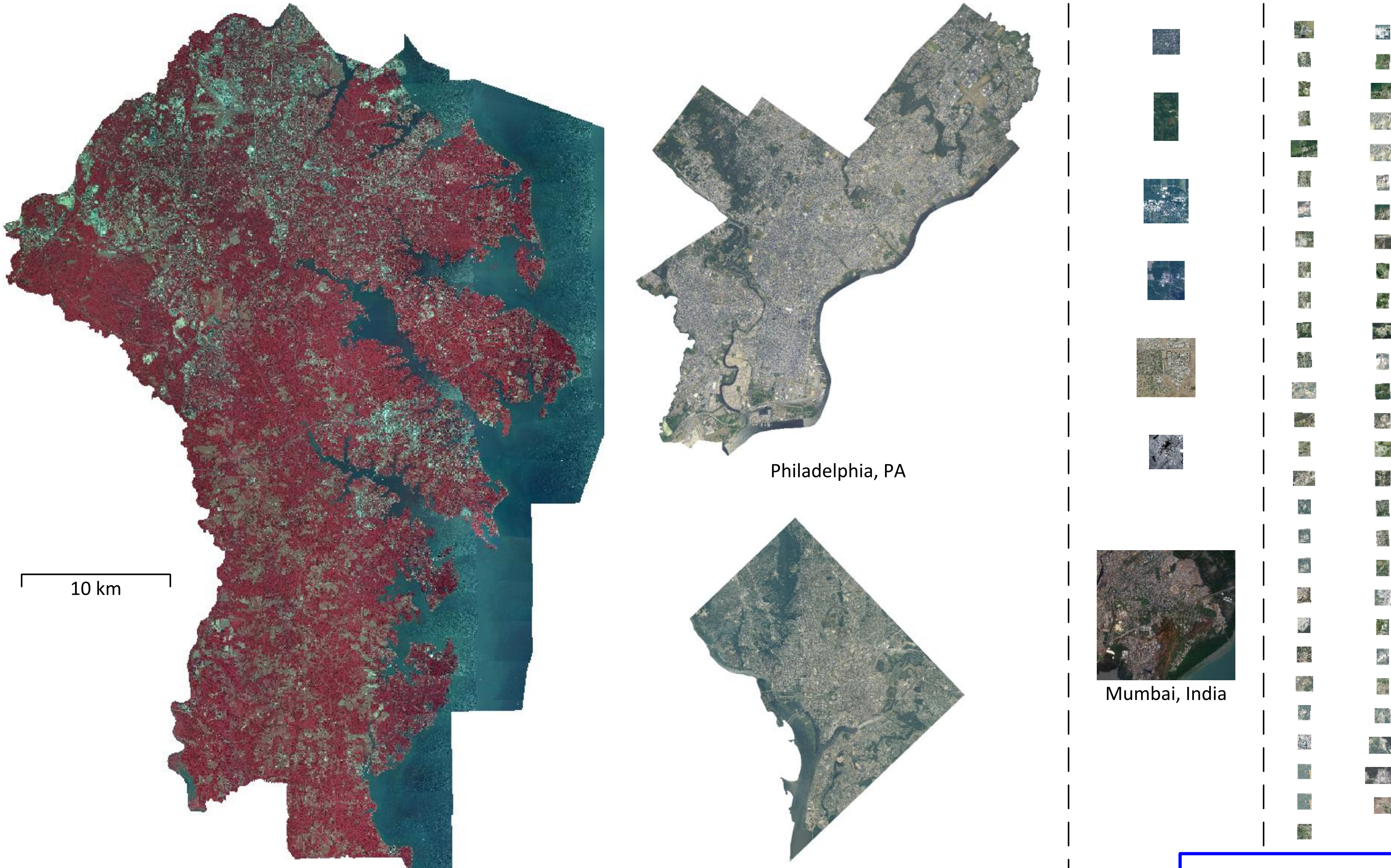
Solution: Path partitioning, path search

UVM/SNL

Overview

- Motivation.
- System Description.
- Examples.
  - Data.
  - Image Processing.
  - Geospatial-Temporal Graph Search.
- Evaluation.
- Summary.

Data Set



Totals:
$2{,}289\ km^2$
23 billion pixels

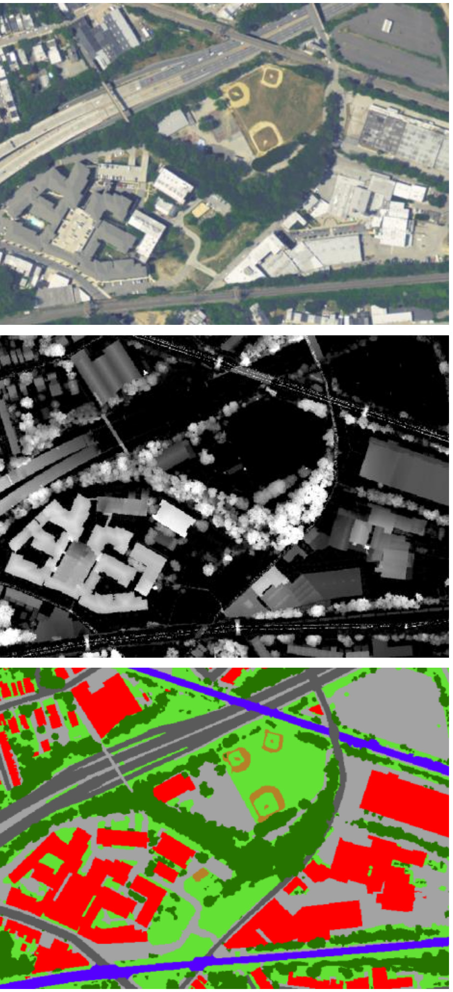
RGB+IR, height maps, and land cover for all locations.

Image data provided by UVM, LLNL, LANL, DigitalGlobe.
All image data from sources available to the public.

UUR SAND 2018-2803 PE

Image Processing Results

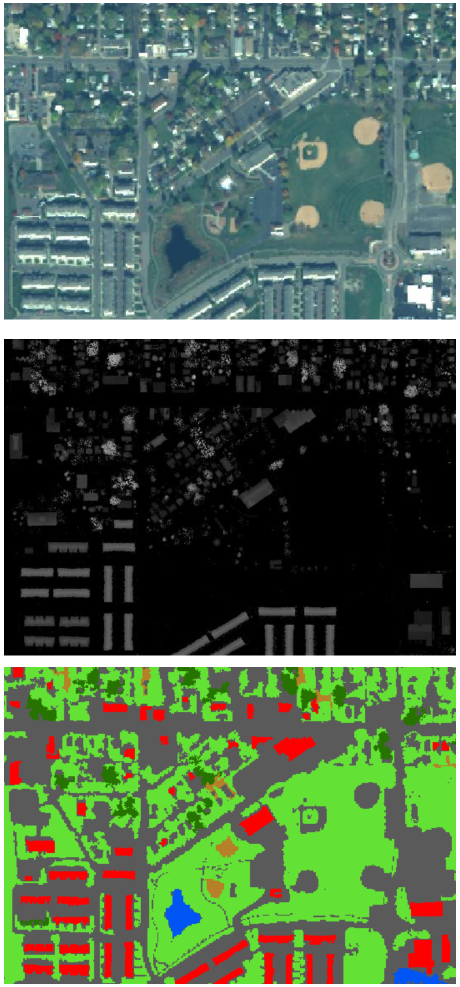
Highest Quality



UVM

RGB+IR, LiDAR, GIS, manual corrections

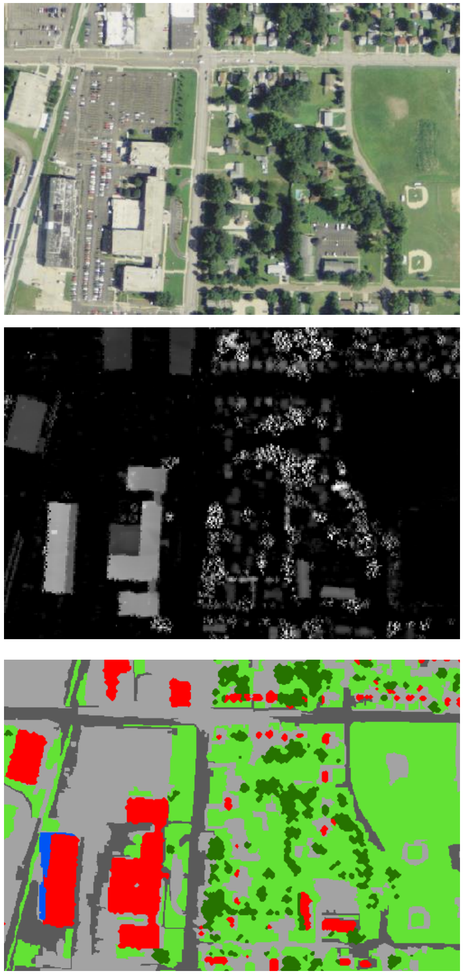
Intermediate



DigitalGlobe©2016/USGS/LLNL

RGB+IR, LiDAR, manual tuning

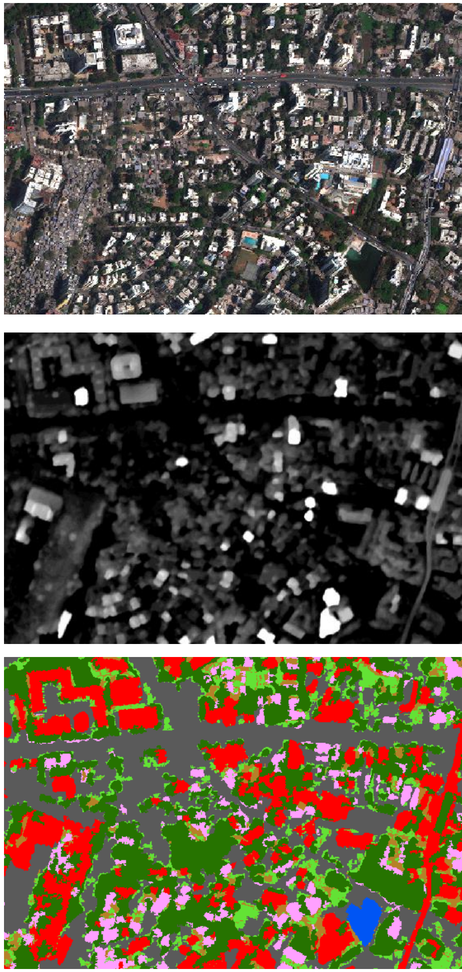
Fully Automated



LLNL+LANL/LLNL

RGB+IR, LiDAR

Global



DigitalGlobe©2015/Vricon©2015/LANL/LLNL

Satellite RGB+IR, Satellite Height Map

Power Plant Search Template

Question: *Where are fossil-burning electricity generation plants?*

Components:

Heat Building
Type: Building, with Height
$2{,}500\ m^2 \le A \le 70{,}000\ m^2$ (prefer $A \ge 4{,}000\ m^2$)
$Perimeter^2/4\pi A \le 100$
Circularity ≤ 0.85
$h_{97.5} \ge 25$ m (prefer $h_{97.5} \ge 30$ m)
$h_{max} / h_{median} \ge 1.6$ (prefer $h_{max} / h_{median} \ge 2.5$)

Body of Water
Type: Water Segment
$A \ge 15{,}000\ m^2$ (prefer $A \ge 17{,}250\ m^2$)

Coal Pile
Type: Dirt, with RGB
$A \ge 30{,}000\ m^2$ (prefer $37{,}500\ m^2 \le A \le 50{,}000\ m^2$)
Eccentricity ≤ 3.0 (prefer Eccentricity ≤ 2.25)
$[R,G,B]_{median} \le 90$ (prefer $[R,G,B]_{median} \le 60$)

Evaporation Pond
Type: Water
$1{,}000\ m^2 \le A \le 20{,}000\ m^2$ (prefer $1{,}750\ m^2 \le A \le 9{,}500\ m^2$)
Eccentricity ≤ 6.0 (prefer Eccentricity ≤ 4.5)

Tank_Landcover
Type: Building, with Height
Circularity ≥ 0.85
$50\ m^2 \le A \le 3{,}000\ m^2$ (prefer $A \ge 80\ m^2$)

Transformer
Type: Paved Chunk
$A \ge 1{,}000\ m^2$ (prefer $3{,}000\ m^2 \le A \le 32{,}741\ m^2$)
1.0 ≤ Eccentricity ≤ 4.0 (prefer 1.0 ≤ Eccentricity ≤ 2.75)
$h_{median} \le 5$ m (prefer $h_{median} \le 2.6$ m)
$4\ m \le h_{97.5} \le 25$ m (prefer $h_{97.5} \ge 5$ m)
$h_{max} \ge 10$ m (prefer $h_{max} \ge 12$ m)
At least $200\ m^2$ above 4 m

Conveyor
Type: Conveyor

Other Structure
Type: Other Structure

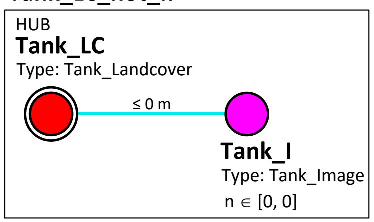
Pipe
Type: Pipe

Stack
Type: Stack

Tall Slender
Type: Tall Slender

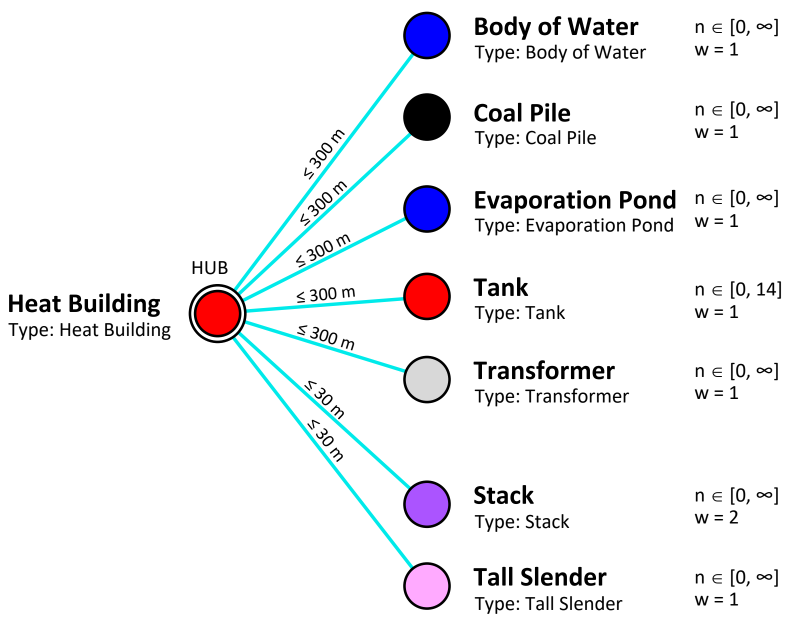
Tank_Image
Type: Tank_Solid OR Tank_Floatroof

Tank_LC_not_I:



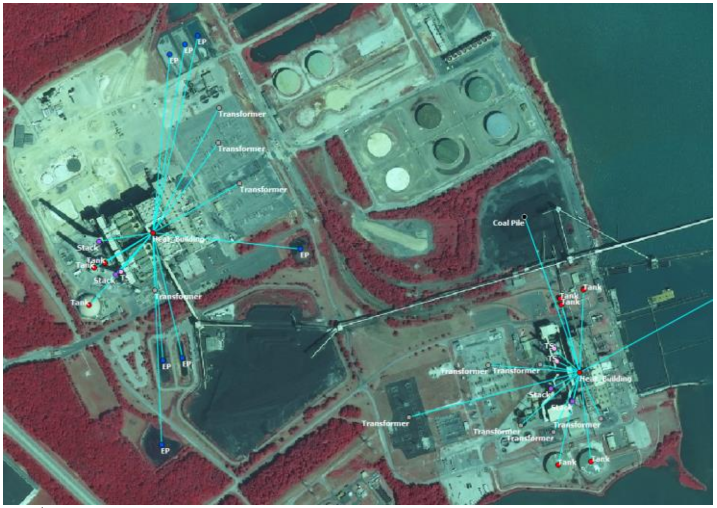
Tank
Type: Tank_LC_not_I OR Tank_Image

Ensemble:

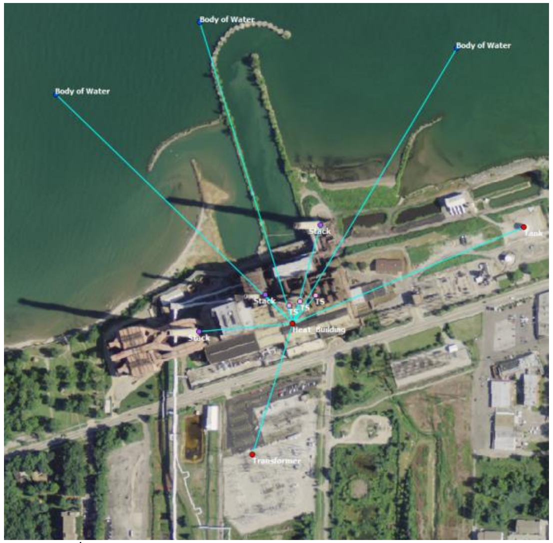


* Batch templates not shown.

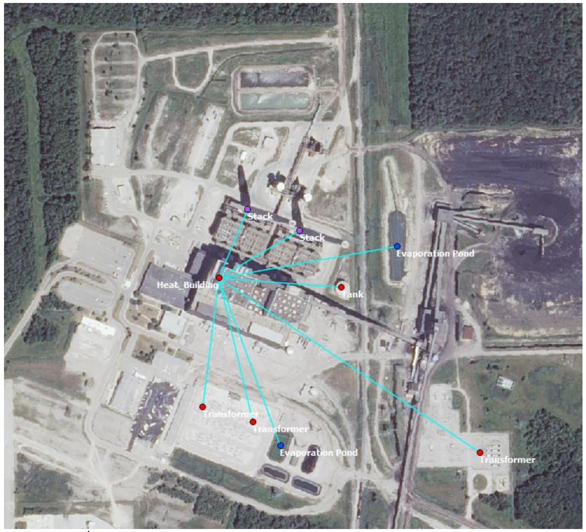
Power Plant Search Results 1-6



UVM/SNL **1. FFe Power Plants A, B**

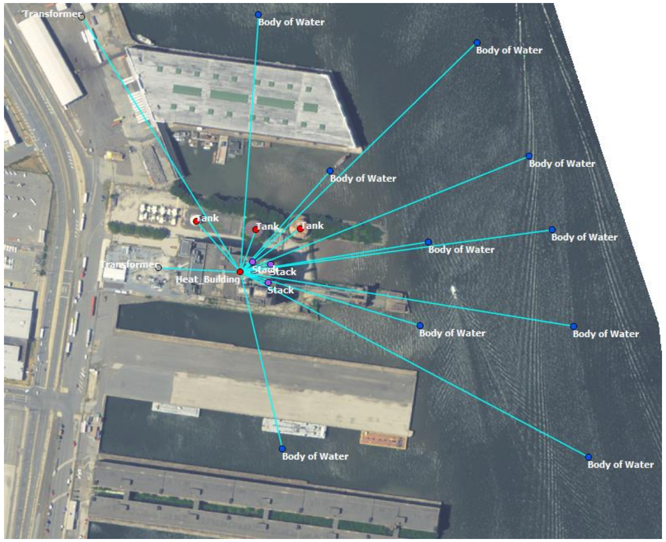


LLNL+LANL/SNL **2. FFe Power Plant 6**

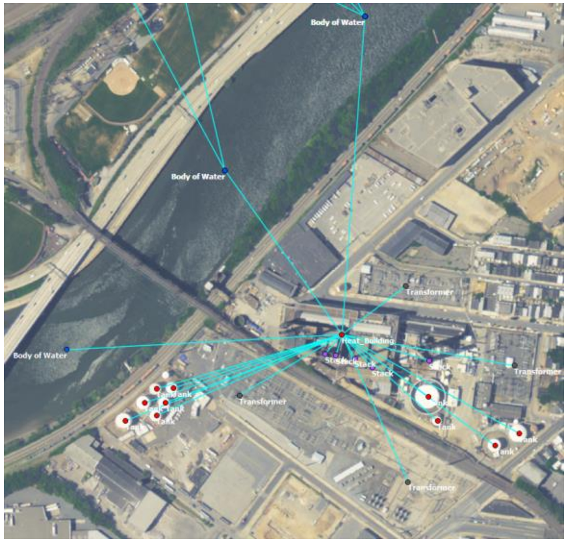


LLNL+LANL /SNL **3. FFe Power Plant 8**

LLNL+LANL/SNL **4. FFe Power Plant 1**

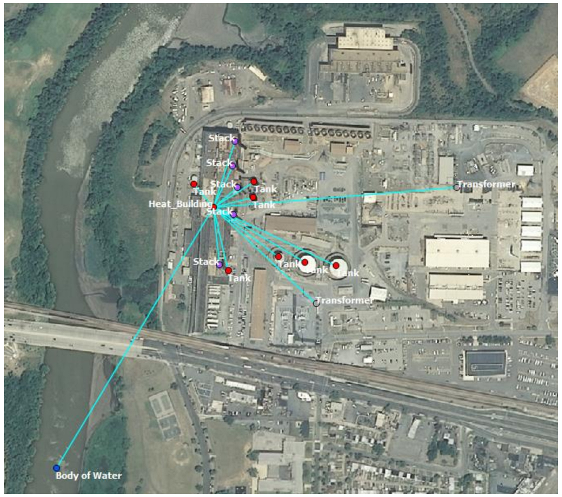


UVM/SNL **5. FFe Power Plant F**



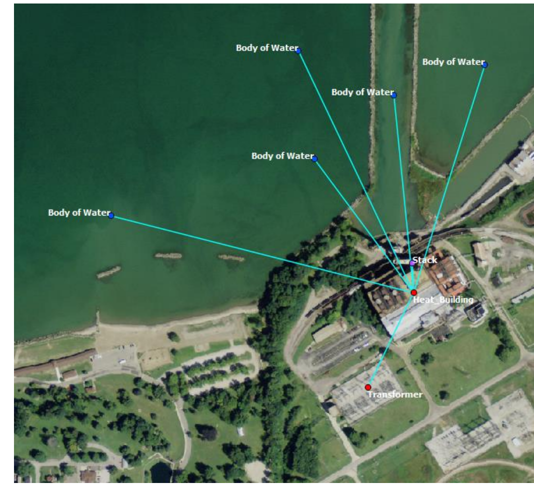
UVM/SNL **6. FFe Power Plant E**

Power Plant Search Results 7-12

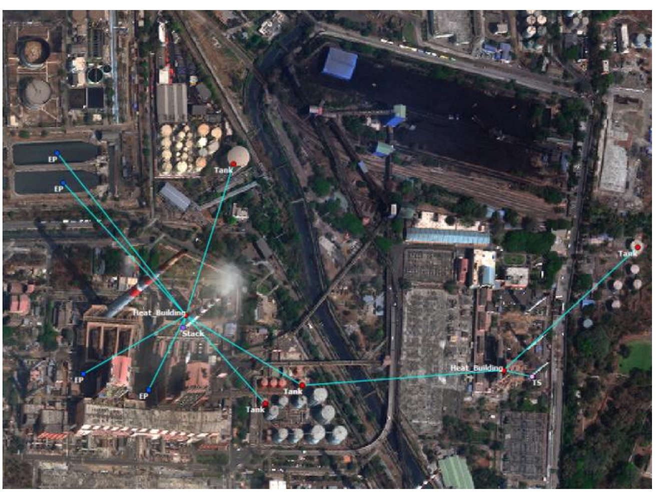


UVM/SNL 7 FFe Power Plant G

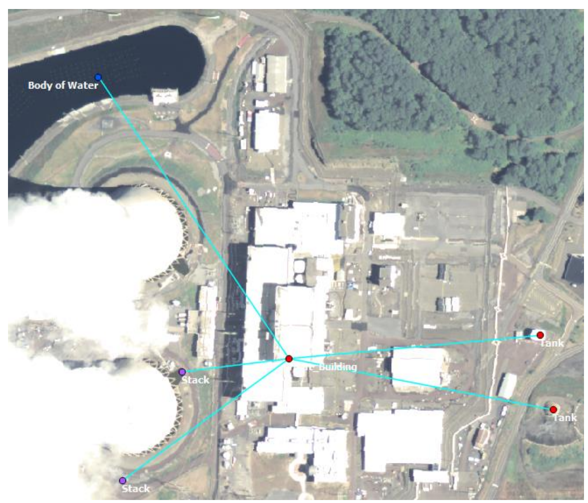
UVM/SNL 8. FFe Power Plant D



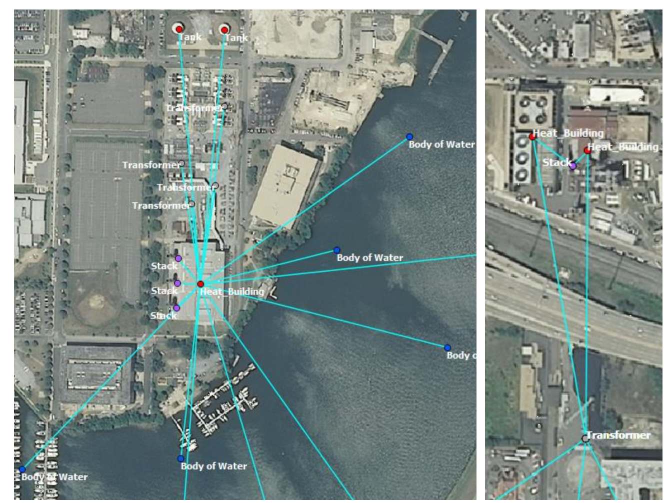
LLNL+LANL/SNL 9. FFe Power Plant 5



DigitalGlobe©2015/SNL 10. FFe Power Plants 9, 10

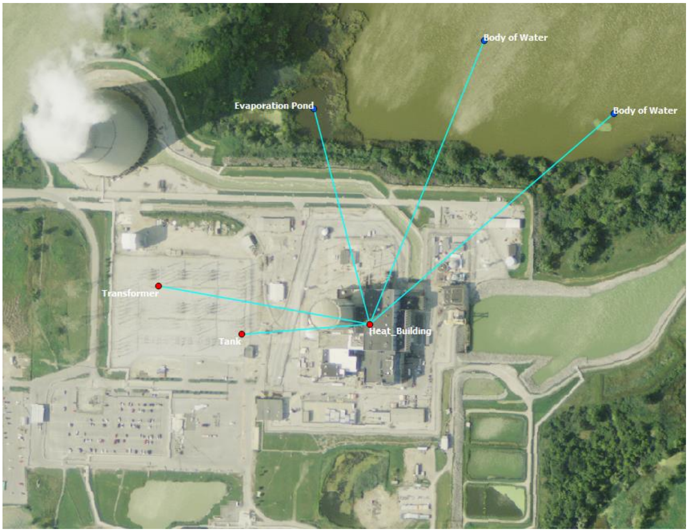


LLNL+LANL /SNL 11. Nuclear Power Plant 3

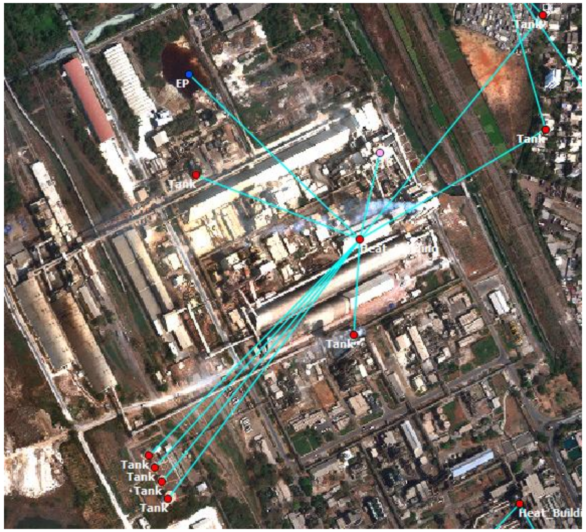


UVM/SNL 12. FFe Power Plant H, FF Power Plant 3 UVM/SNL

Power Plant Search Results 13-18

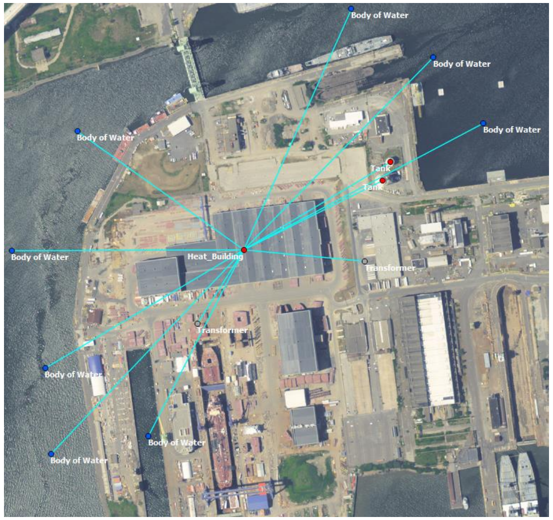


LLNL+LANL/SNL 13. Nuclear Power Plant 4



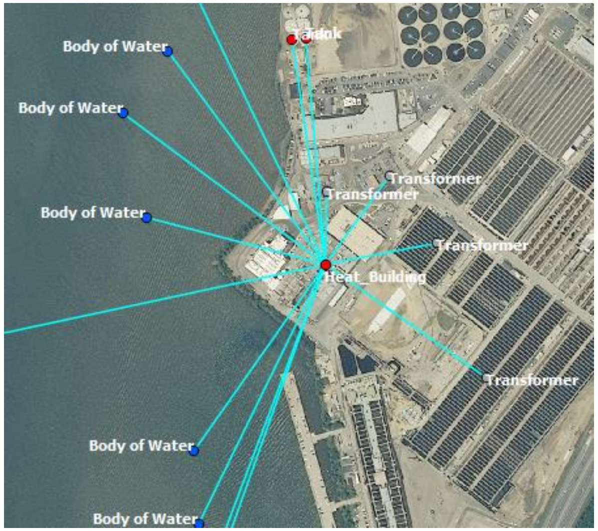
DigitalGlobe©2015/SNL 14. Industrial Site

UVM/SNL 15. FFe Power Plant C



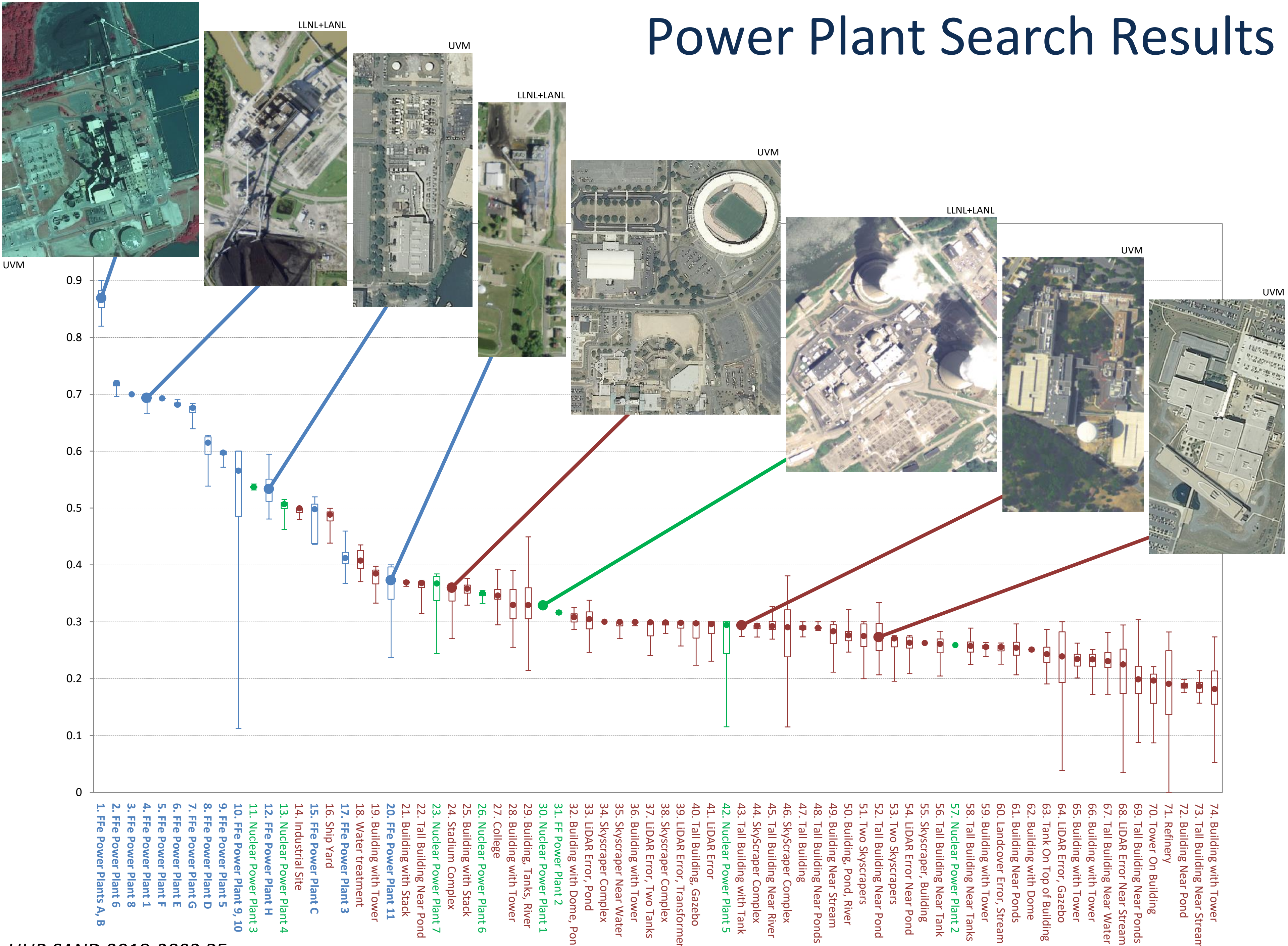
UVM/SNL 16. Ship Yard

LLNL+LANL/SNL 17. FFe Power Plant 3



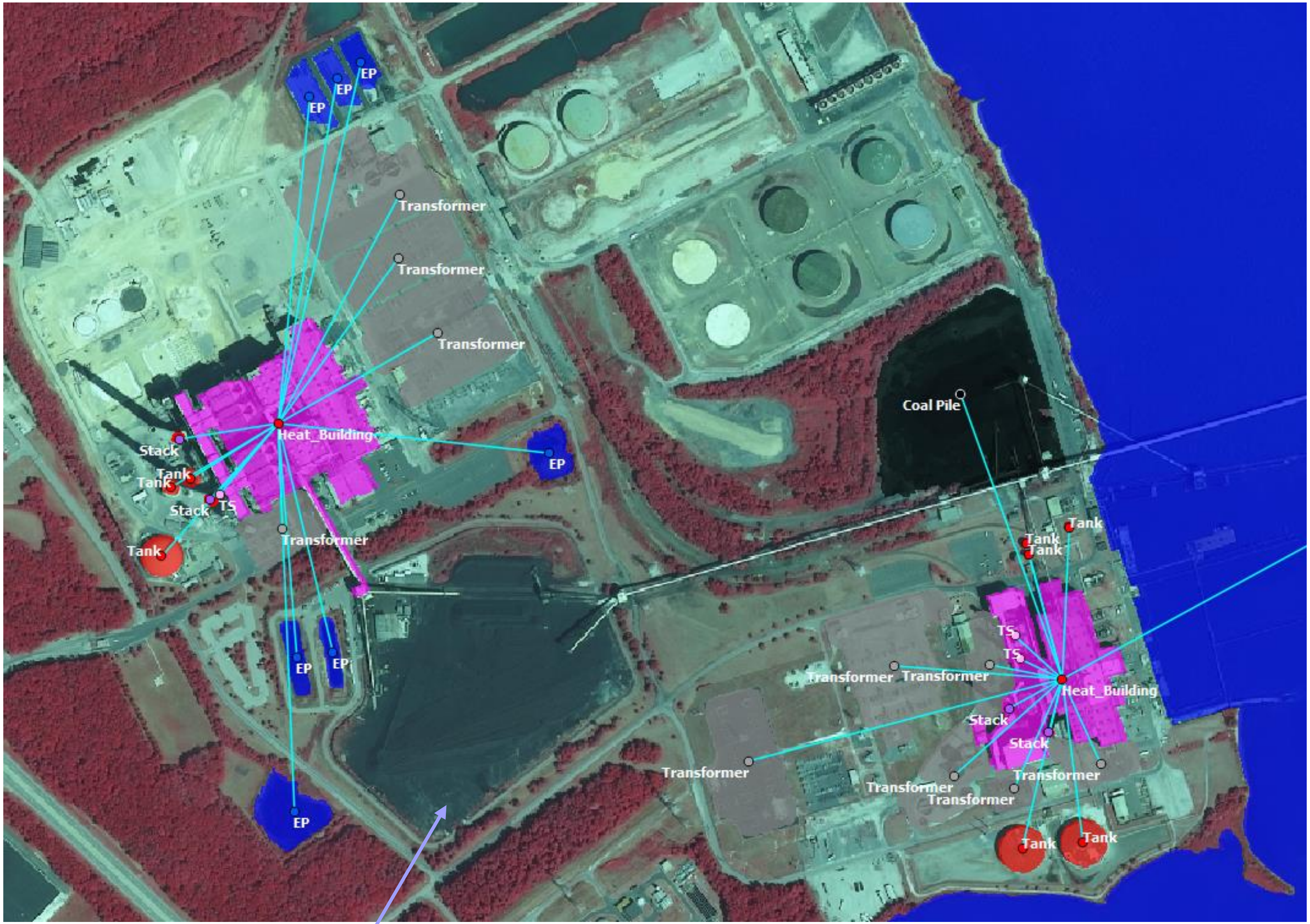
UVM/SNL 18. Water Treatment

Power Plant Search Results



UUR SAND 2018-2803 PE

Power Plant Match #1 (True Positive)



UVM/SNL

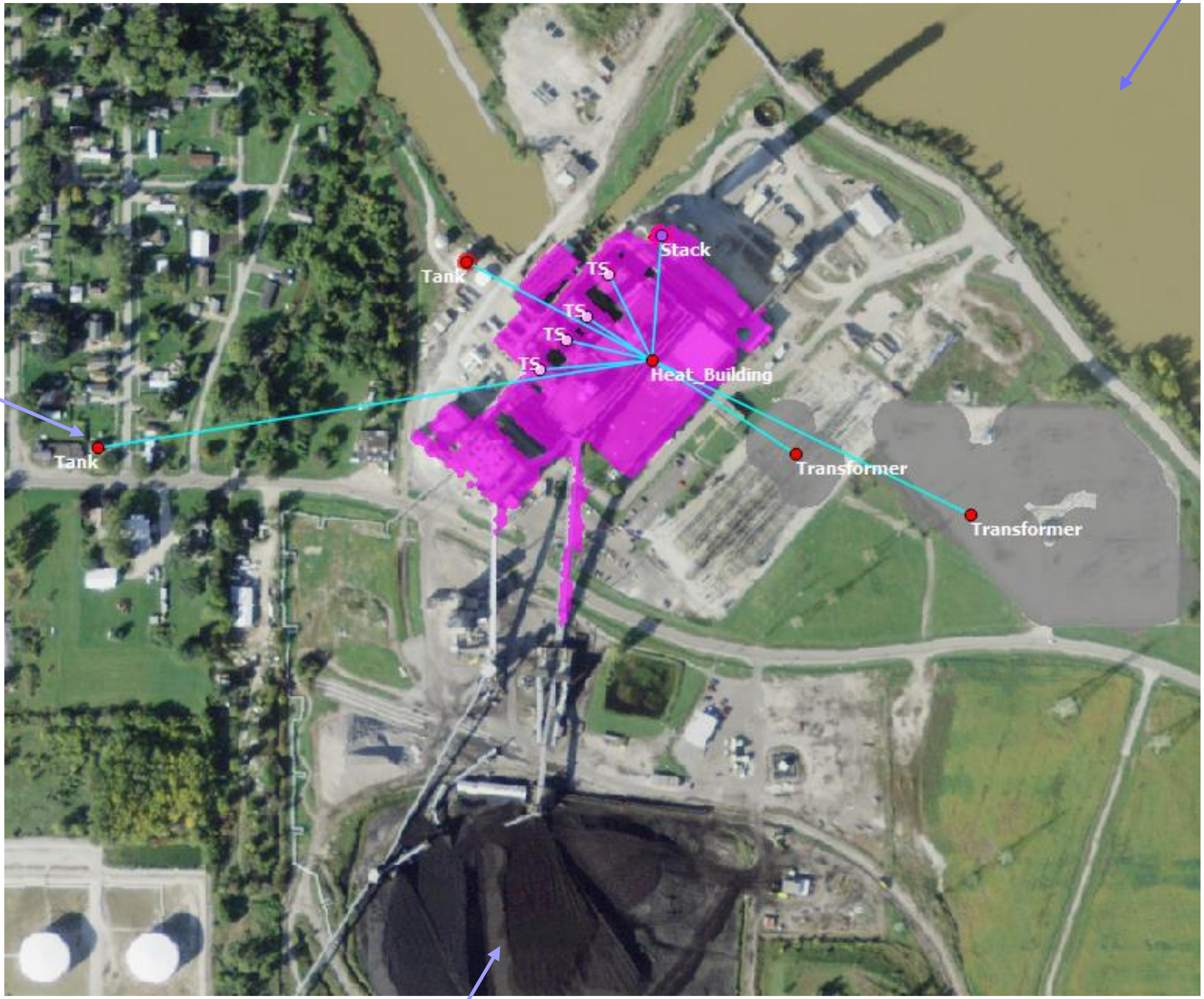
Coal pile missed, due to image processing error.

q = 0.870

Match #4 (True Positive)

Water missed by image processing.

Spurious tank found.

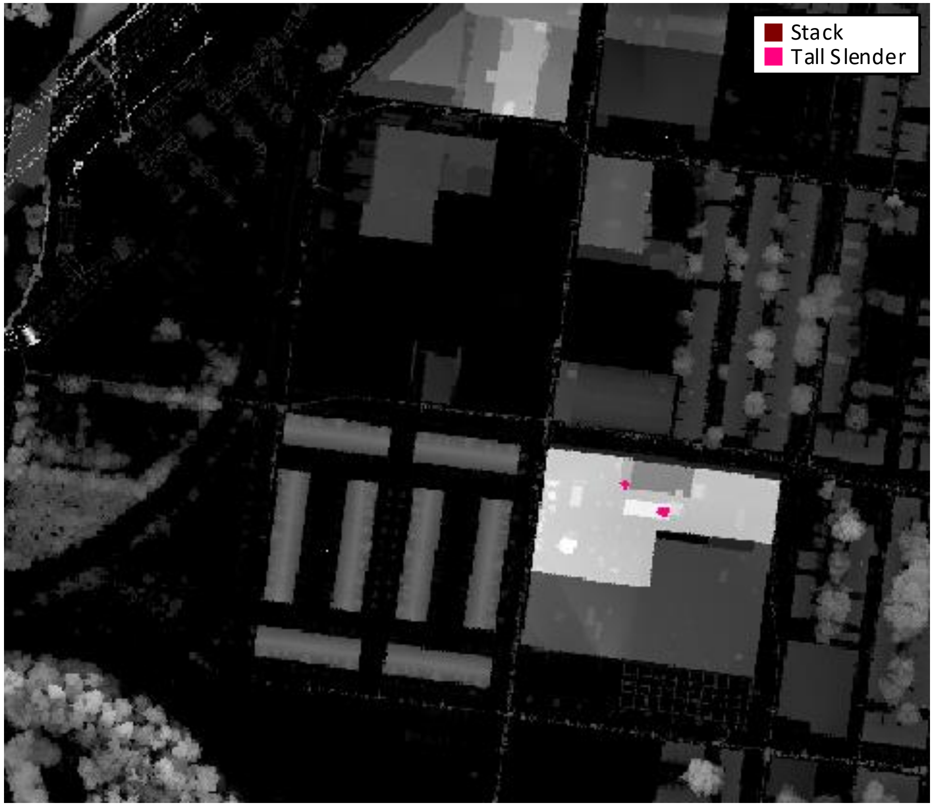


LLNL+LANL/SNL

Coal pile missed, due to image processing error.

q = 0.694

Match #21 (False Positive)



UVM/SNL

Special Features

UVM/SNL

Match

Match has multiple features consistent with a power plant, but is not a power plant.

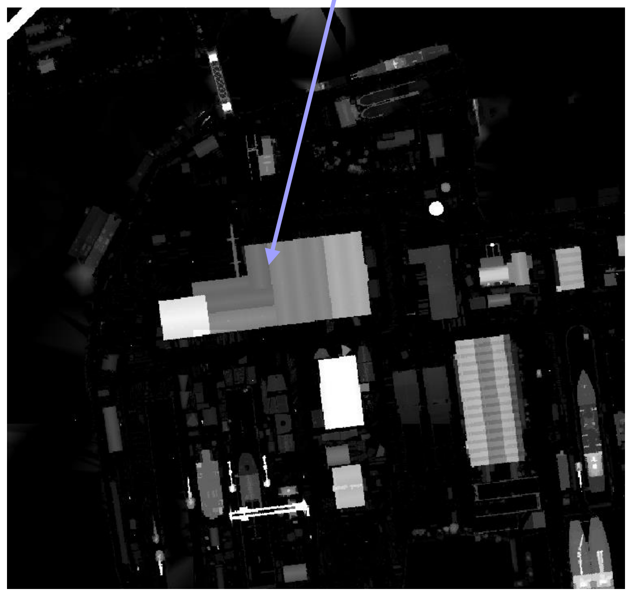
q = 0.369

Match #16 (False Positive)

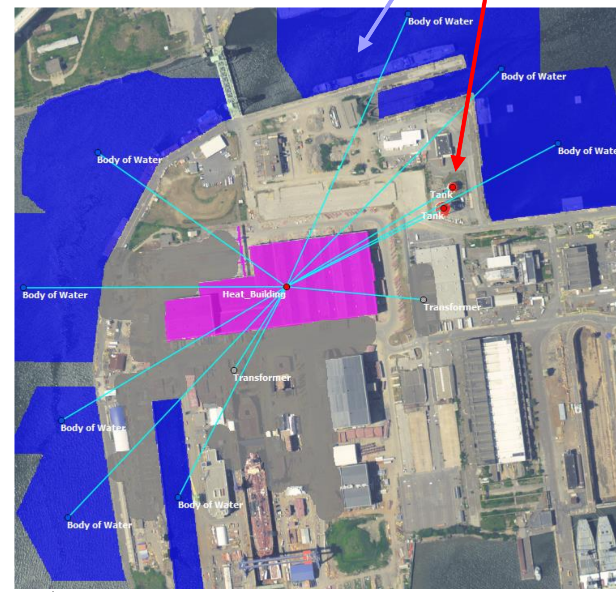
Large building, with tall subset, matches heat building criteria.

Body of water, tanks, correctly found

UVM Optical Image



UVM nDSM

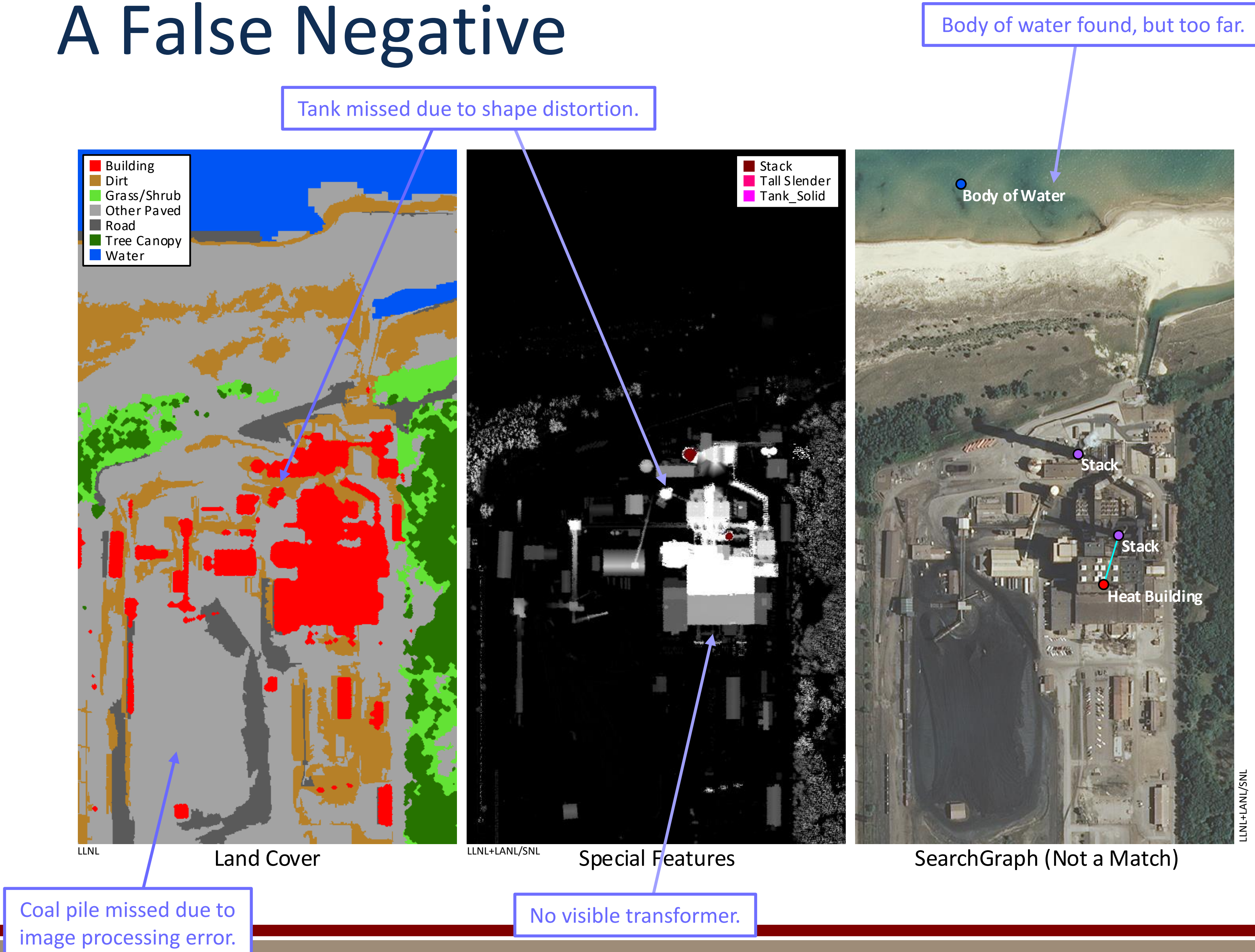


UVM/SNL Match

Paved areas with clutter meet transformer criteria.

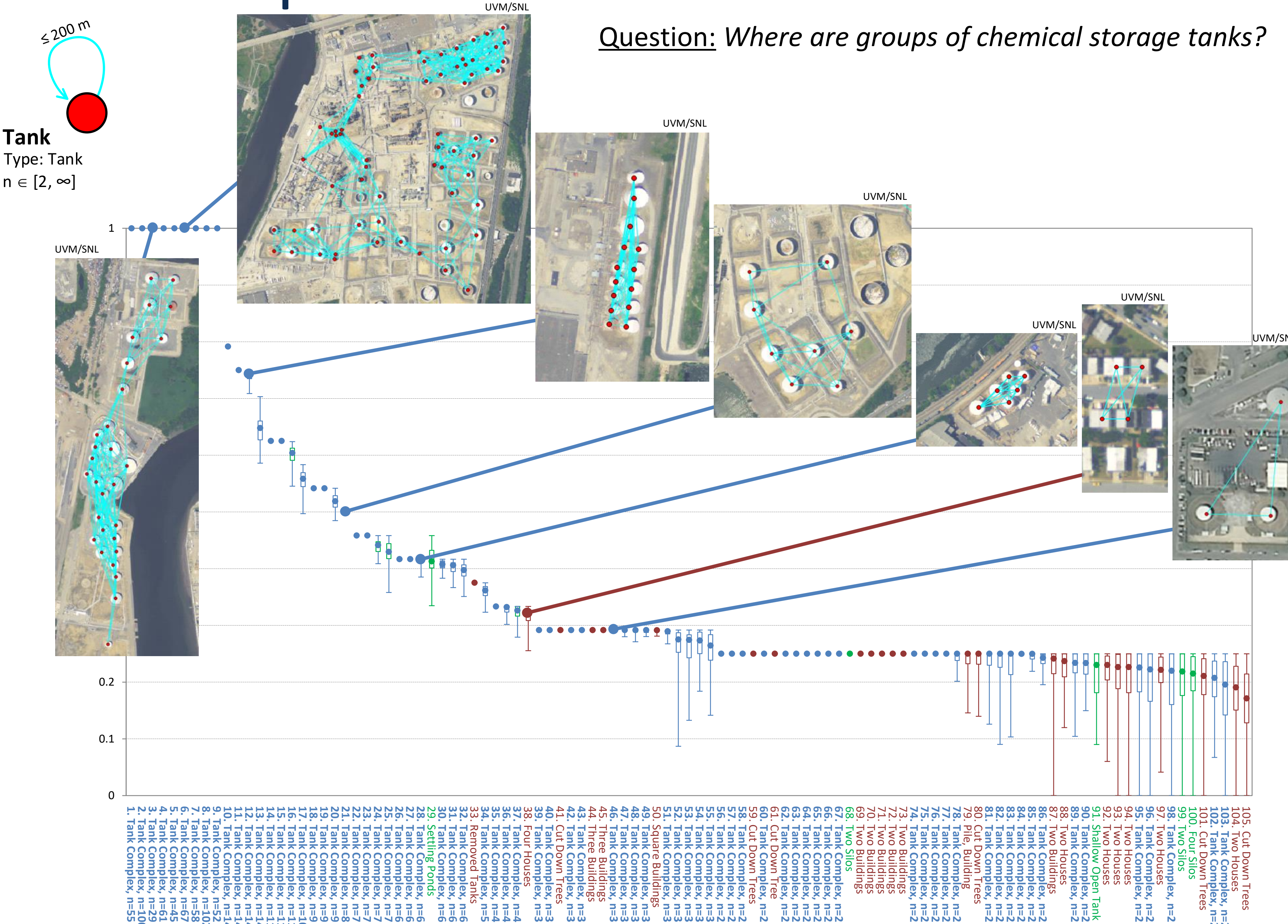
q = 0.488

A False Negative



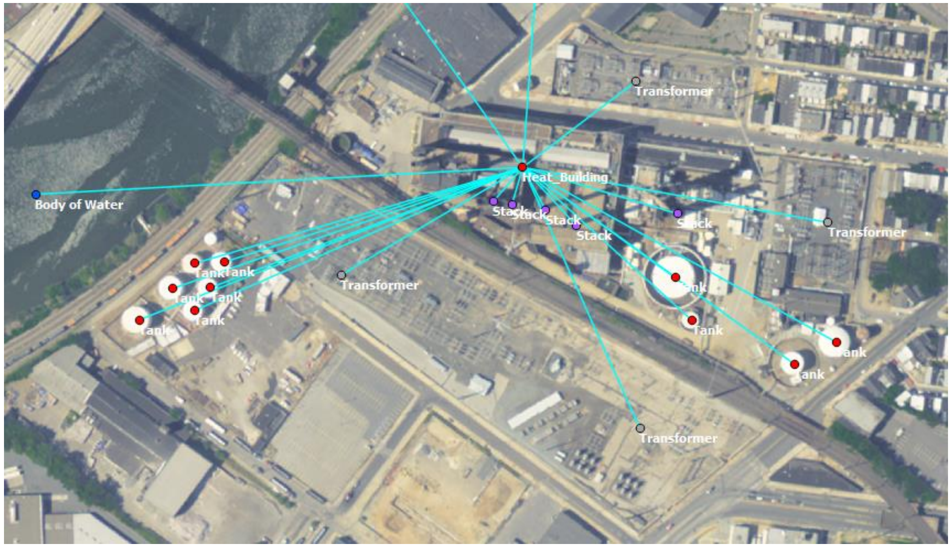
Tank Complex Search

Question: *Where are groups of chemical storage tanks?*



Multiple Matches for the Same Site

Power Plant Search Result



UVM/SNL

Tank Complex Search Result

UVM/SNL

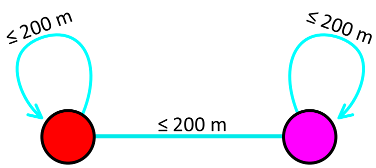
Reporting alternative explanations is important for user understanding.

A final system should clarify these.

Refinery Search

<u>Question:</u> *Where are fossil fuel and chemical refineries?*

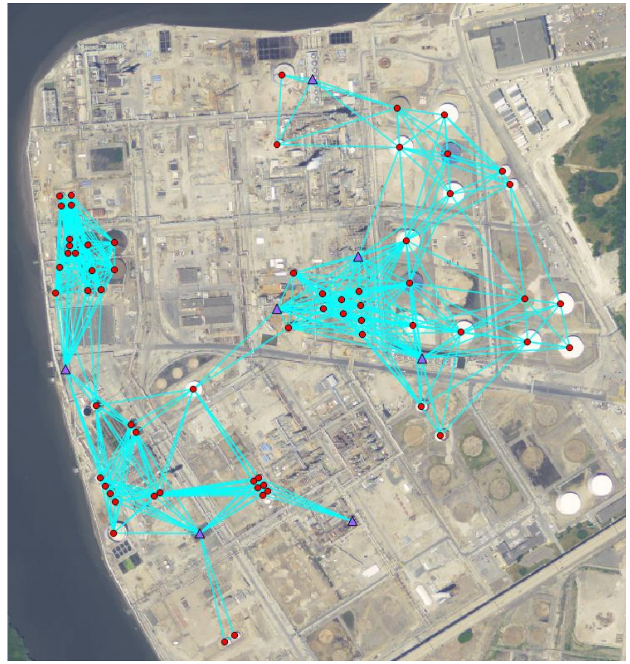
<u>Large refineries:</u>



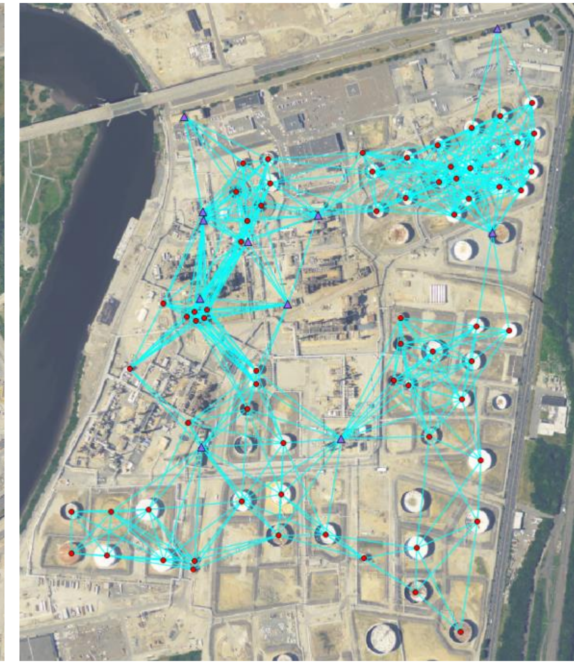
Tank
Type: Tank
n ∈ [10, ∞]

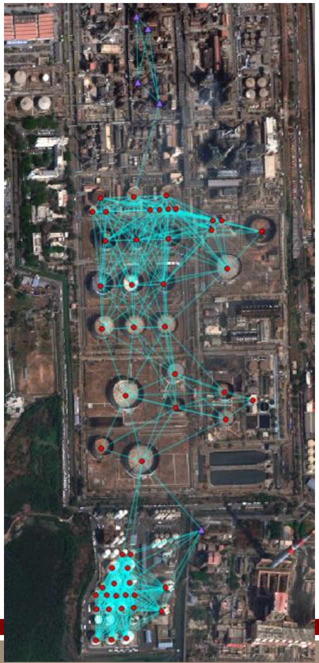
Processing Tower
Type: Building, with Details
Not circular
$20\ m^2 \le A \le 2{,}500\ m^2$
$h_{max} \ge 20$ m
n ∈ [5, ∞]

UVM/SNL



UVM/SNL



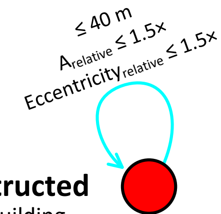


DigitalGlobe©2015/SNL

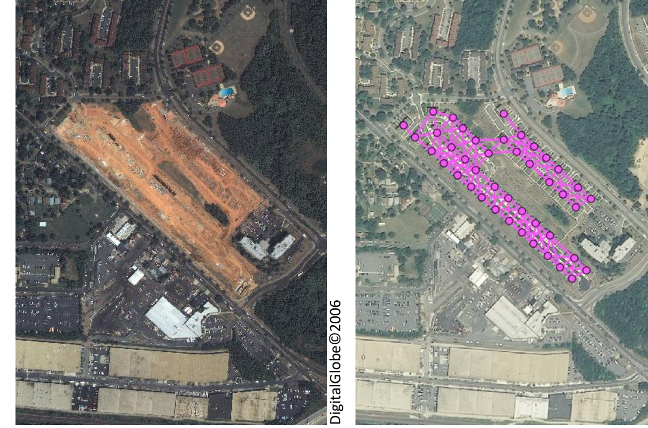
UVM/SNL

New Complexes

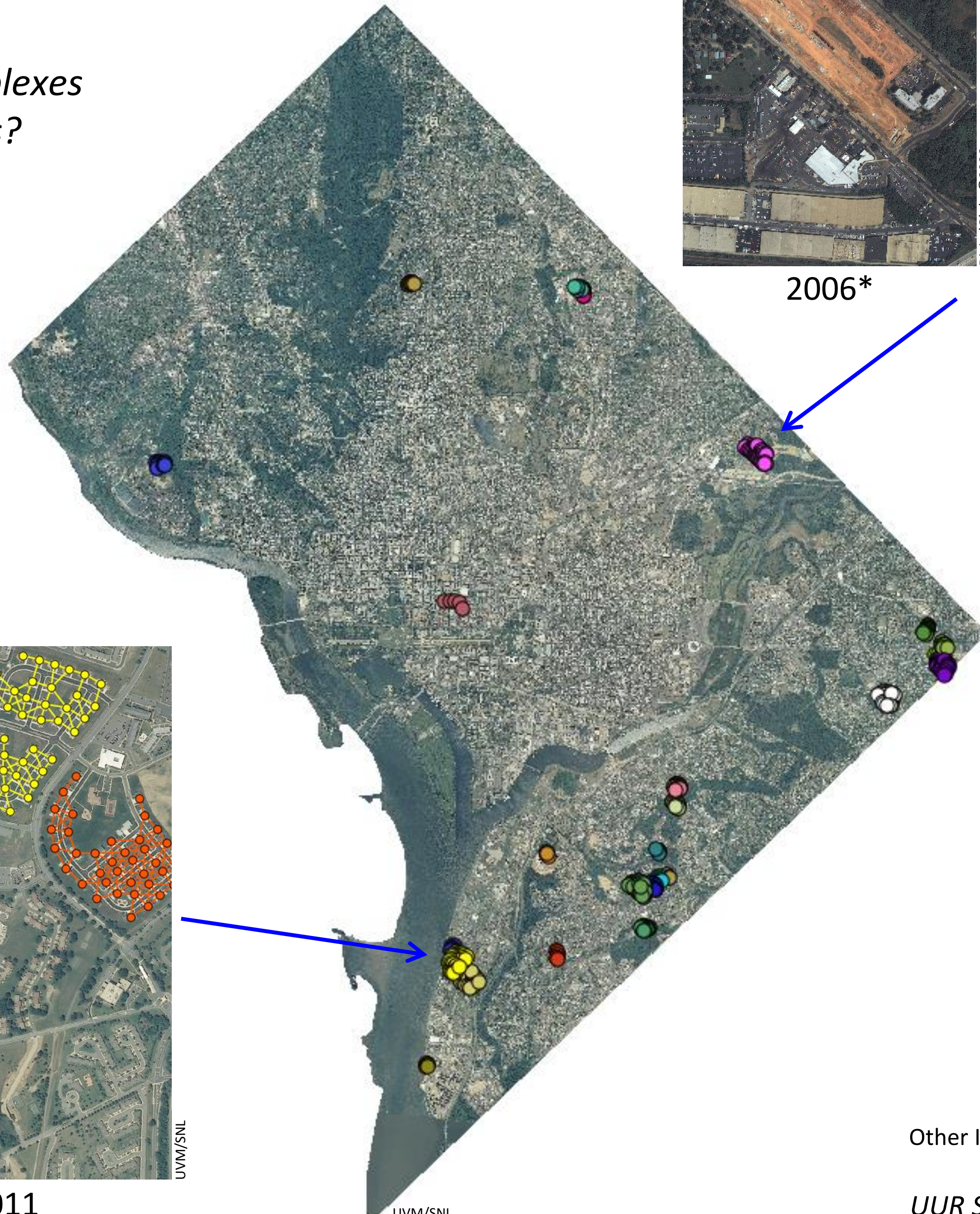
<u>Question:</u> *Where are complexes of new buildings?*

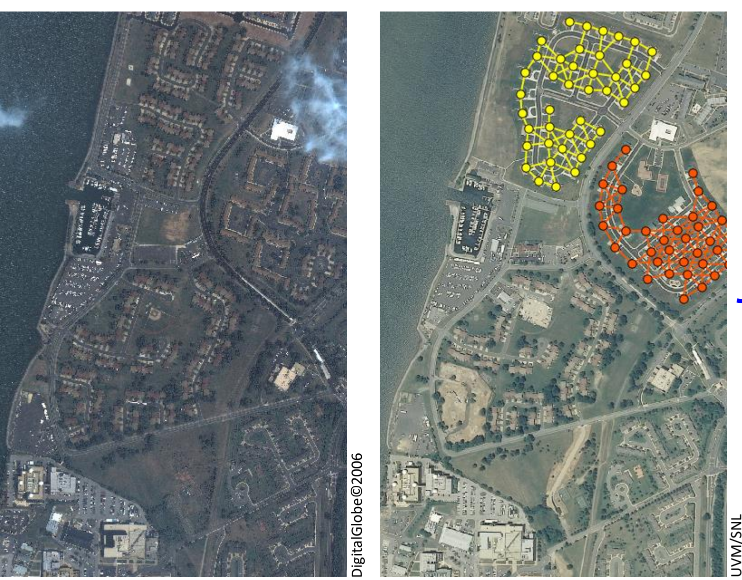


Constructed
Data: Building
Exists now
A ≥ 100 m^2
New, Extended, Changed
n ∈ [5, ∞]



2006* 2011





2006* 2011

* Image from DigitalGlobe.
Other Image data provided by UVM.

Desirable House Search

<u>Question:</u> *Where are houses with multiple trees, a pleasant walk to a park, not near traffic?*

<u>Search:</u>

Batch:

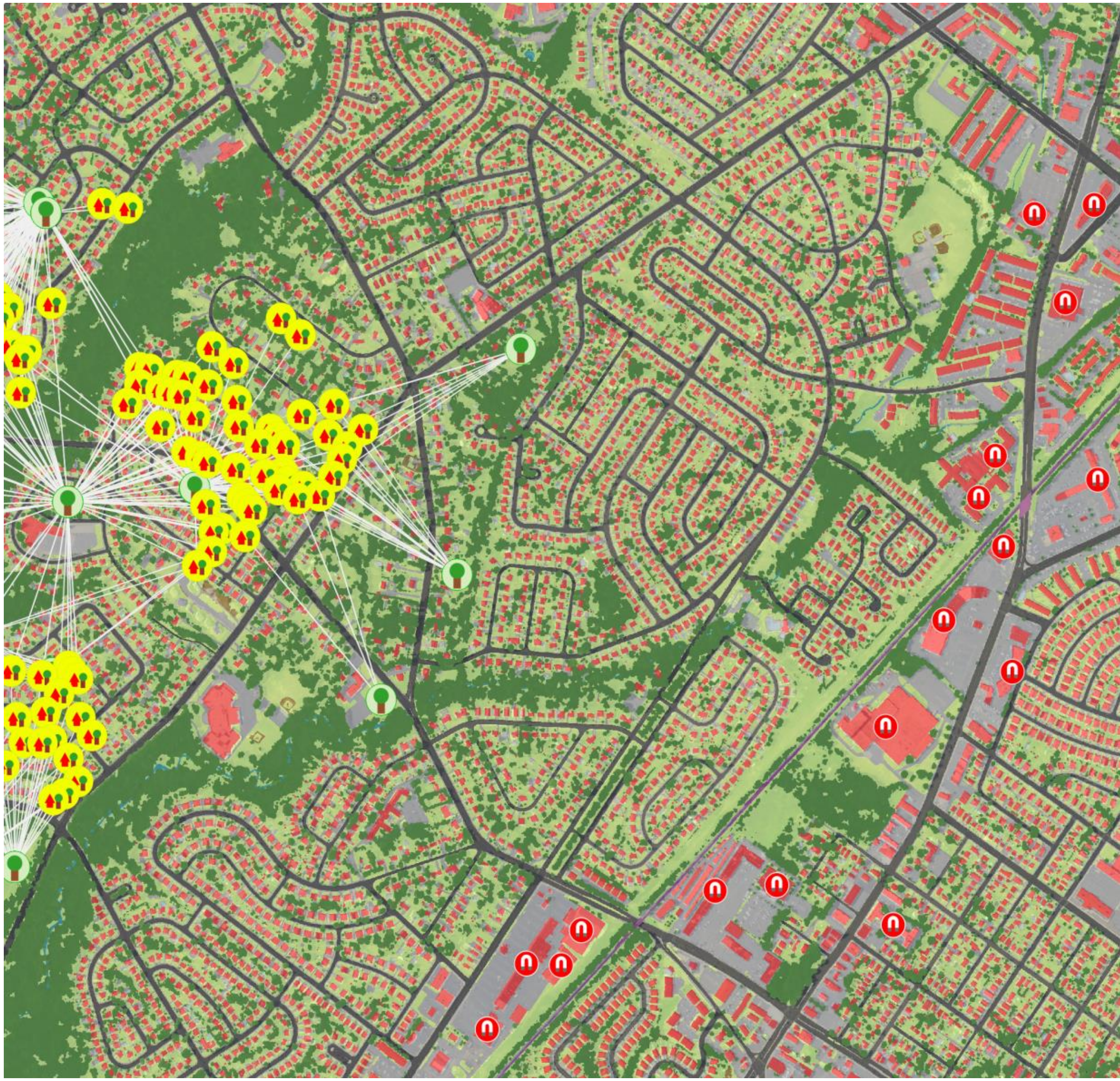
- Road path segment.
- Building, with height.
- Paved chunk.

Components:

- Short building. < 120 ft
- Tall building. ≥ 120 ft
- Residential housing. Small houses, no tall buildings
- Shaded house. ≥ 2 trees
- Grass+Trees complex.
- Park. Grass+Trees, near houses
- Traffic magnet. Large buildings near parking lots

Ensemble:

- Shaded house near Park, not near Traffic magnet.



UVM/SNL

Desirable House Search

Are these parks, or not?

UVM/SNL

Proper segmentation may depend on legal property boundaries, which are invisible in overhead imagery.

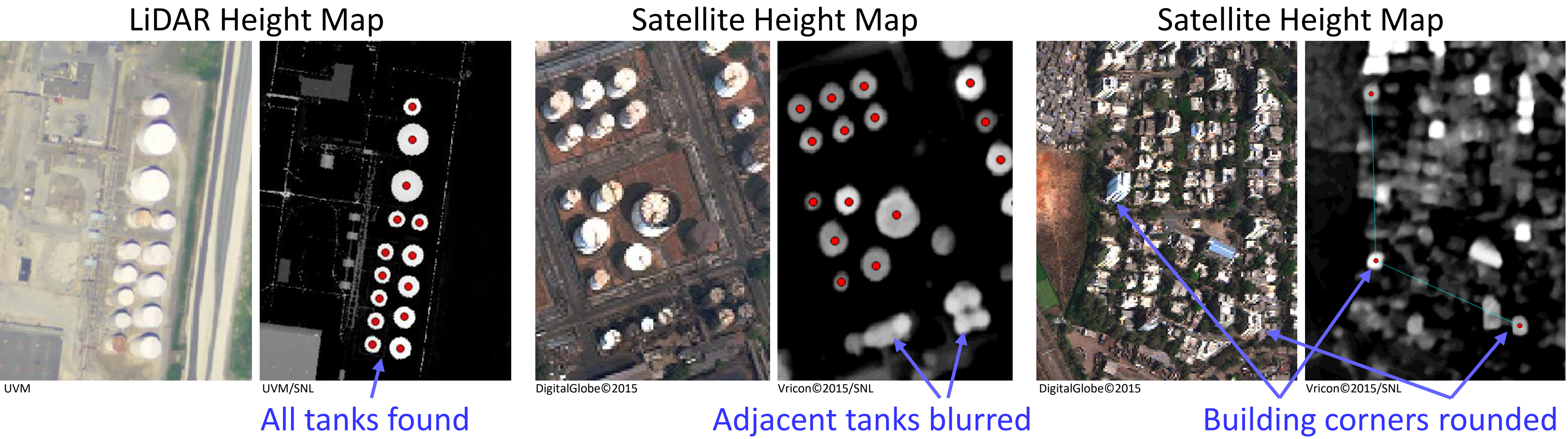
Overview

- Motivation.
- System Description.
- Examples.
- Evaluation.
- Summary.

Evaluation: Data

<u>Data corpus:</u>

- Systematic requirements.
- Commercially available.
- Anywhere in the world.
- Large area tested (2,289 km^2).
- Multiple types: Airborne vs. satellite, LIDAR vs. satellite.



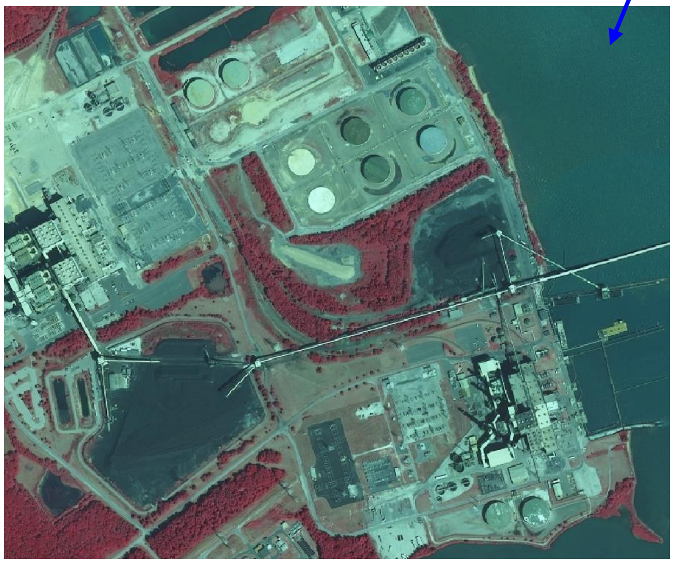
Evaluation: Image Processing

Comments on image processing:

- Successful.
- Systematic approach.
- Best use of each modality.
- Effective use of height data, IR.
- Quality varies with data, manual tuning.
- Quantitative accuracy estimated for some cases (LLNL, led by Randy Roberts).
- Several lessons learned. Examples:
  - Wide variation in surface spectral properties.
  - Wide variation in surface illumination/reflection.
  - Temporal aliasing.
  - (See following slides.)
- Others are making significant strides.
  Example: DigitalGlobe/PSMA continent-scale landcover for Australia.

Evaluation: Image Processing

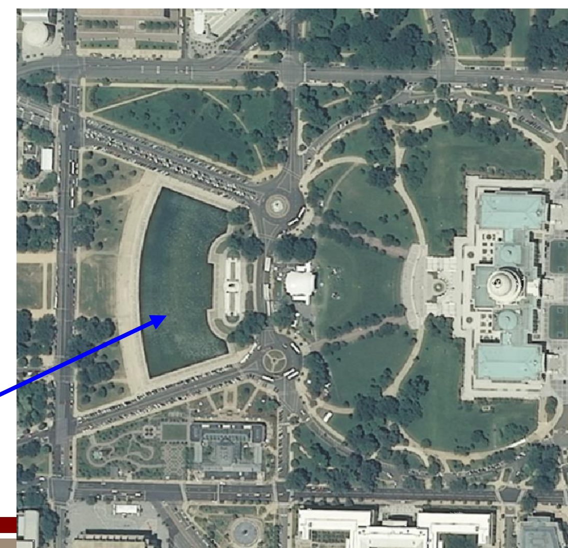
<u>Spectral variation in material:</u>



UVM

LLNL+LANL

LLNL+LANL



UVM

Evaluation: Image Processing

Spectral variation due to illumination:

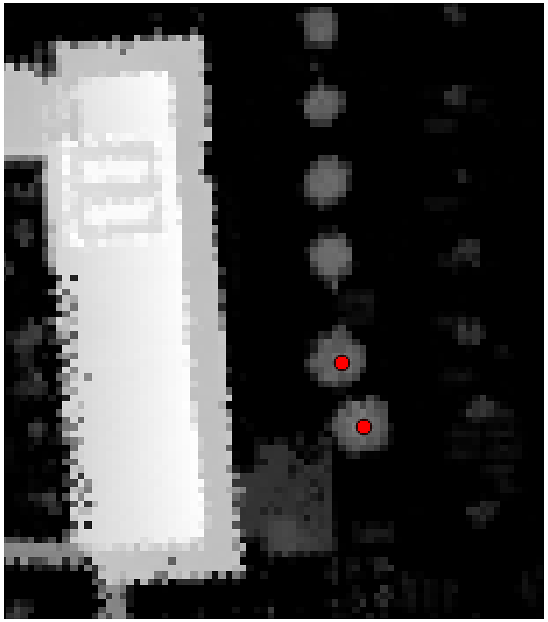
Variation along road

LLNL+LANL

Evaluation: Image Processing

Temporal aliasing:

LiDAR

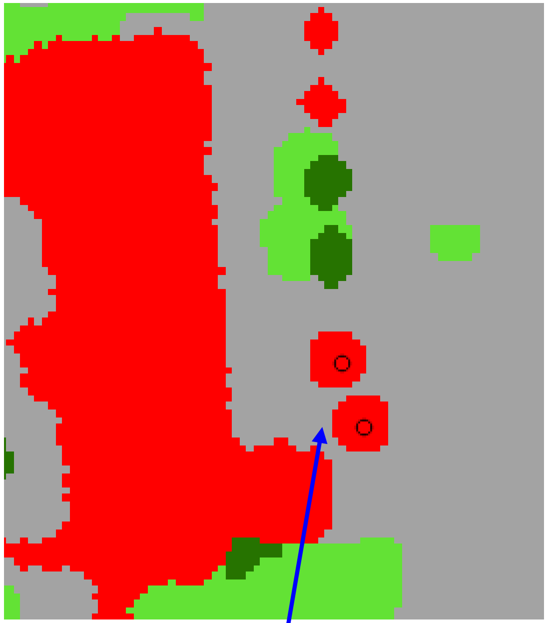


LLNL+LANL/SNL

Optical (Later Date)

LLNL+LANL/SNL

Resulting Landcover



UVM/SNL

Buildings seen in previous tree locations.

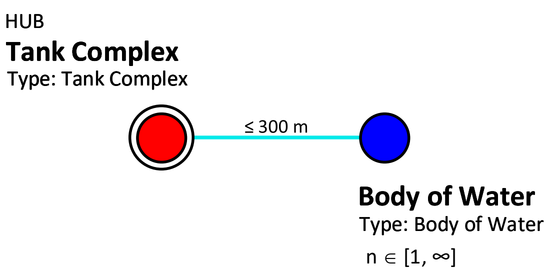
Evaluation: Template Preparation

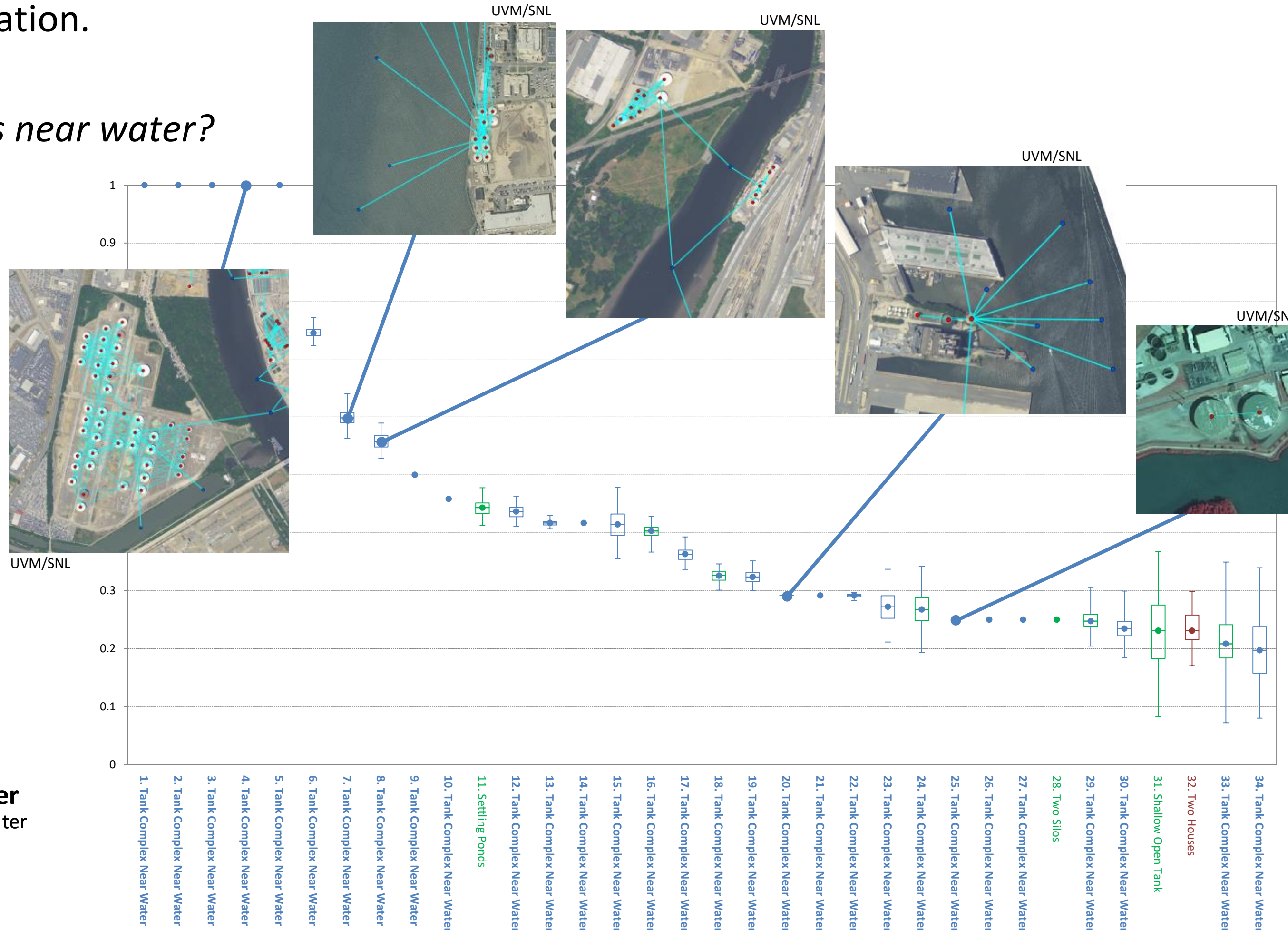
<u>How long does it take to design a template?</u>

- User studies of template design not performed yet.
- Power plant – developed over years, mixed with code development.
- Tank complex – quick, easy.
- Refinery – easy, some iteration.
- Measured example:
  Where are tank complexes near water?

Template design	15 min *
Run	3 min
Render and review	30 min *
Total	48 min

* Manual setup.
(no user interface)





UUR SAND 2018-2803 PE

Evaluation: Search

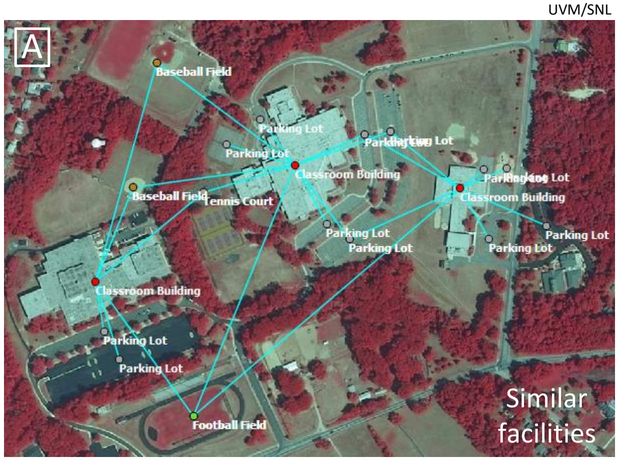
	Fossil Fuel Electric Power Plant		Tank Complex		Refinery	Tank Complex Near Water	
True Positives	16	24	76	81	4	30	33
False Positives	60	52	29	24	0	4	1
False Negatives	1		?		?	?	
Precision	21%	32%	72%	77%	100%	88%	97%
True positives before first false	10	13	28	32	4	10	31
False positives before last true	6 / 20	4 /20	27 / 103	22 / 103	0 / 4	4 / 34	1 / 34

Cautions:

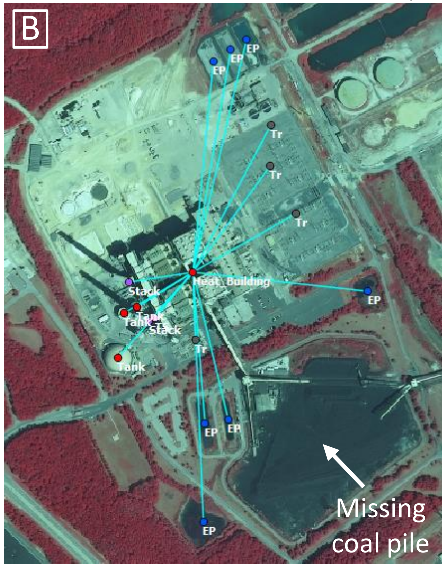
- Cueing system:
  - Groups treated as one match [A].
  - Within-match errors not counted [B].
- Small number of query experiments.
- A measure of the query, not the system.

Green = includes serendipitous

UVM/SNL



UVM/SNL



Evaluation: Run Time

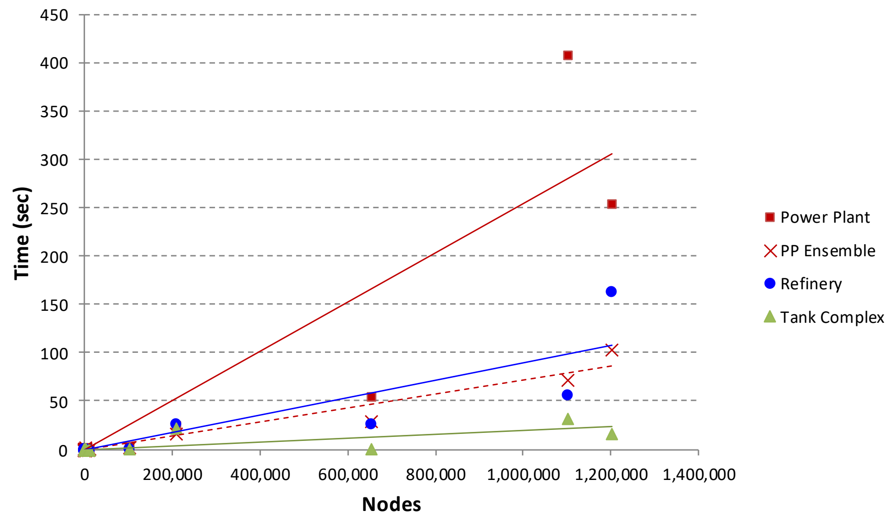
Problem Size

Number of Locations	62
Total Pixels	2.3×10^9
Total Area	2,289 km^2
Initial Graph Nodes	3,328,055
Initial Graph Edges	7,071,644
Objects/km^2	172 - 3,701
Number of Steps, All Searches	22
Total GeoSearch Runs	1,364

Total Run Time

	Total (sec)	Rate (sec/km^2)
Batch:		
Graph Construction	7,334	3.2
Batch Search	36,408	15.9
Interactive:		
Power Plant Search	784	0.34
Power Plant Search (ensemble only)	234	0.10
Tank Complex Search	80	0.03
Refinery Search	274	0.12
Tank Complex Near Water	358	0.16

Interactive Search Run Time



Evaluation: Scope

What range of problems can we solve?

Issues:

- *Feature Robustness.* Does search goal include difficult objects?
  Buildings are more reliable than ground-level features.
  Larger objects are typically more robust.
- *Data Semantic Adequacy.* Does data contain required information?
  Example: Building interior features are not visible.
- *Segment Semantic Relevance.* Does segmentation match question of interest?
  Undersegmentation: Paved region vs. parking lots, two towers connected by a breezeway.
  Oversegmentation: River interrupted by bridges.
  Lack of segment information: Invisible property boundaries needed.
- *Environment Difficulty.* How many similar, undesired items are in the environment?
- *Search Goal Variation.* Requires thoughtful template design.

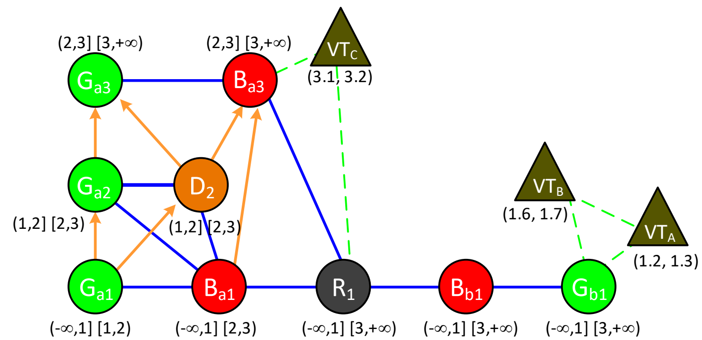
Best results for search goals:

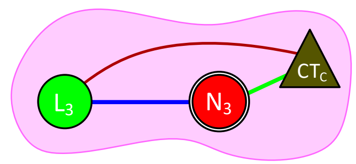
- Large components which do not change frequently.
- Classes robustly recognized by image processing.
- An environment with few visually similar elements.
- Improved if goal has multiple visible identifying components.
- Improved with special image processing for key features.

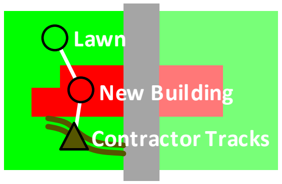
Evaluation: Other Applications

Ephemeral Activity Analysis:

- Ephemeral nodes.
- Spatiotemporal relationships.
- Activity pattern search.







Trajectory Analysis:

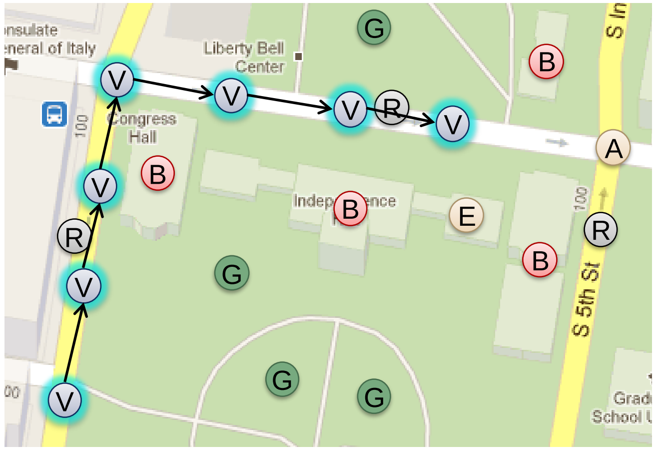
- Moveable object nodes.
- Trajectory nodes.
- Special distance calculations.
- Special trajectory subset handling.

Durable Nodes
Building B
Grass G
Road R

Ephemeral Nodes
Accident A
Exhibit E

Movable Nodes
Vehicle V

Google©2014/SNL



Brost, Perkins, Czuchlewski. Activity Analysis with Geospatial-Temporal Semantic Graphs. Sandia Report SAND2015-10455, December 2015.
Perkins, Ray, Brost. Geospatial-Temporal Semantic Graph Models of Trajectories. Sandia Report SAND2015-8592, November 2015 (OUO).

Evaluation: Relation to Other Systems

Geospatial-Temporal Semantic Graphs:

- *Representation Generality.* Multiple modalities, hierarchical semantics, geospatial and temporal relationships, variety of attributes and constraints.
- *Ease of Query Design.* Modular graphical queries vs. complex nested join statements. (Caveat: User studies not performed yet.)
- *Search Flexibility.* Variable topology, heterogenous elements, controllable cardinality. Star graph, connected component, path connectivity, nested structure.
- *User-Driven Focus.* Hierarchical semantics, edge construction follow search desires.
- *Problem Flexibility.* Wide-area search, Activity analysis, Trajectory analysis.

GeoGraphy code:

- *Multi-Modality Data.* Raster images and tables points. (Shape files not supported.)
- *Edge Semantics.* Distance, Adjacency, Change, ComponentOf, WithinMatch, multiple constraints.
- *Lazy-Evaluated Distance Edges.* Avoids $O(n^2)$ complexity.
- *Segment Transformation Operations.* Chunking, path network, grow/shrink.
- *Automatic Provenance Management.* Lookup of raw source information for regions.
- *Search Algorithms.* Star graph, connected component, source-to-sink path, disconnected star.
- *Storing Search Results Back Into Graph.* Includes semantic ComponentOf and WithinMatch edges.

Overview

- Motivation.
- System Description.
- Examples.
- Evaluation.
- Summary.

Key Ideas

- **Data → primitive objects → StoredGraph | query → relevant objects → SearchGraph → matches.**
- **Search based on semantic properties, not particular geometry or layout.**
- **Data semantics (StoredGraph) transform to user question semantics (SearchGraph).**
- **Segment transformation operations.** Chunking, agglomeration (grow/shrink), path network,…
- **Search writeback.** Modular queries, faster interaction, user-driven semantics, increased scope.
- **Quality scoring.**
- Base on simple objects.
- Use geospatial location and time as registration anchors.
- Mixed modality: regions, points, multiple sensors, durable, ephemeral, trajectories.
- Variable-topology templates.
- Simple graph search algorithms: Star graph, connected component, interrupted star.
- Dynamic multi-attribute relationship specification (e.g., relative time, area, etc).
- Lazy-evaluate edges, complex attributes ⇒ user-driven computation focus.
- Durable change representation, chronology.
- Ephemeral activity extension.
- Trajectory extension.
- Systematic approach to image capture (UVM, USGS, DigitalGlobe, Vricon).
- Systematic approach to image processing (UVM).
- Support images anywhere in the world.

Summary

Hypothesis 1: Graph can enable "search engine" for overhead imagery?

Yes, we can construct a general-purpose search capability.

+ Rank results by match quality to user goal.
+ Segment transformation can improve semantic adequacy.
+ Search writeback → complex semantics, simpler queries, faster response.
+ Results for several examples.
+ Graph offers improvements over basic database approaches.
+ Multiple uses, including: Wide-area search, Activity analysis, and Trajectory analysis.
– Ease of query writing not yet understood.
– Limited by semantic adequacy of objects, and confusers in the environment (question-specific).
– Limited by image processing (see Hypothesis 2).

Hypothesis 2: Automatic image processing can find basic objects (buildings, trees, water,…)?

Yes, we can construct a general-purpose image processing capability.

+ Clear image capture requirements.
+ Anywhere in the world, but with varying quality.
+ Fully automatic processing → Several successful search examples.
– Quality varies by feature type.
– Quality depends on 3-d height map.
– Quality with manual intervention much higher than fully automated.
– Subset of imagery usable.
– Quality vs. temporal resolution tradeoff.

Scope:

Best results for search goals with:

- Large components which do not change frequently.
- Classes robustly recognized by image processing.
- An environment with few visually similar elements.
- Improved if goal has multiple visible identifying components.
- Improved with special image processing for key features.

Notes:

- Image processing progress by several groups.
- Time analysis is an open frontier.
- User interface, code architecture work needed.
- See report SAND 2017-8687 for full details.

Image Credits

All of the overhead image data presented here are available to the public, either free from sources such as the U.S. Geological Survey (USGS), or for purchase from commercial vendors such as DigitalGlobe and Vricon. We also include Google Map images.

Images were provided by colleagues at the University of Vermont (UVM), Lawrence Livermore National Laboratory (LLNL), and Los Alamos National Laboratory (LANL), who in turn obtained the imagery from USGS, other public sources, and from DigitalGlobe and Vricon.

Image credits appear throughout the presentation, listing the data provider. If the image contains annotations, the annotation source is also listed. Thus "UVM/SNL" indicates the data was provided by UVM, and annotations were created by Sandia National Laboratories (SNL).

Images with data source marked "LLNL+LANL" are from the Benchmark Data Suite assembled by Lawrence Livermore and Los Alamos. For a description of this data set, see:

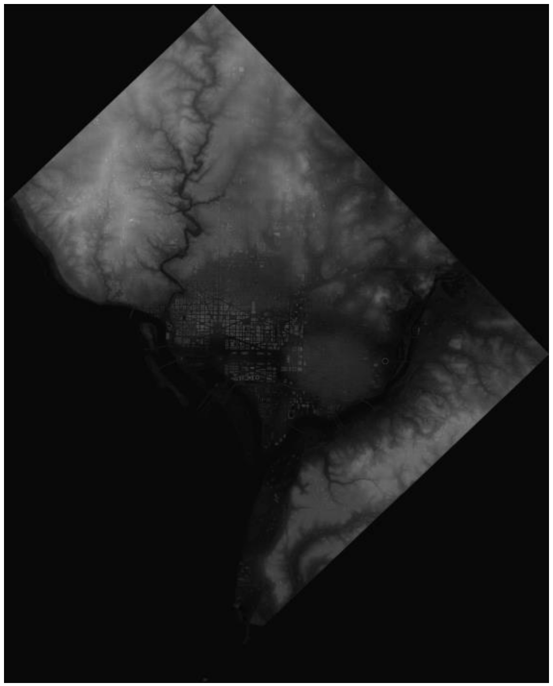
> White, et al. Benchmark imagery for assessing geospatial semantic content extraction algorithms. Technical Report LLNL-TR-645052, Lawrence Livermore National Laboratory, October 2013.

We gratefully thank our colleagues for their help obtaining these data.

Images of the Monongahela River oil spill courtesy the Pittsburgh Post-Gazette.

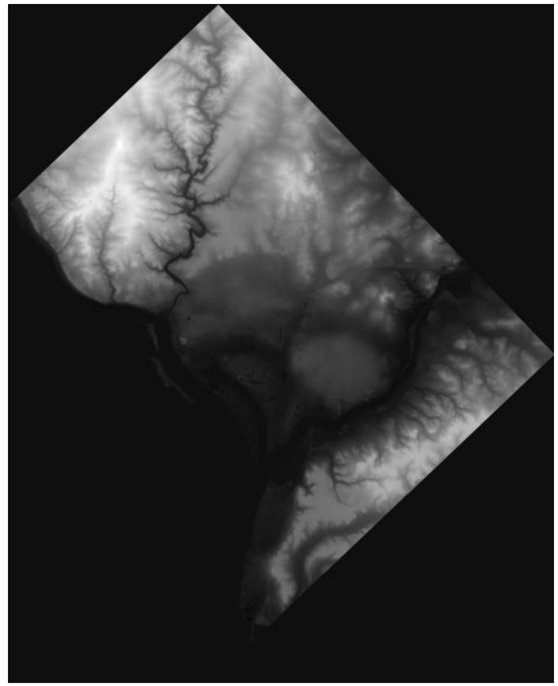
BACKUP SLIDES

Normalized Height Map Construction



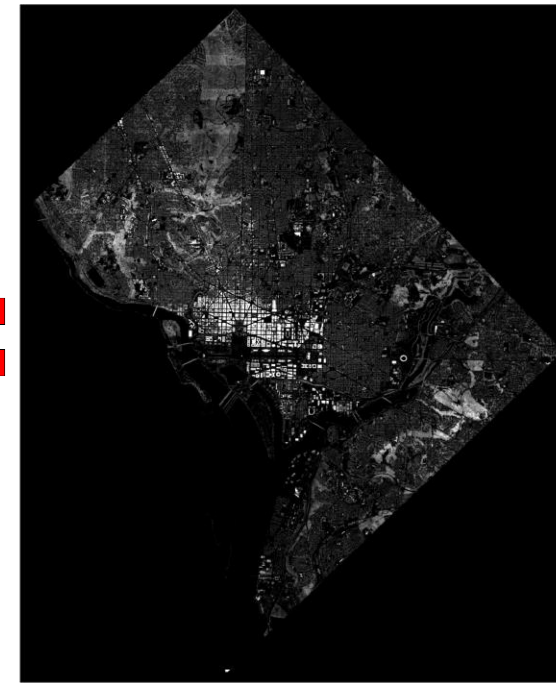
UVM

Digital Surface Model
(DSM)



UVM

Digital Elevation Model
(DEM)



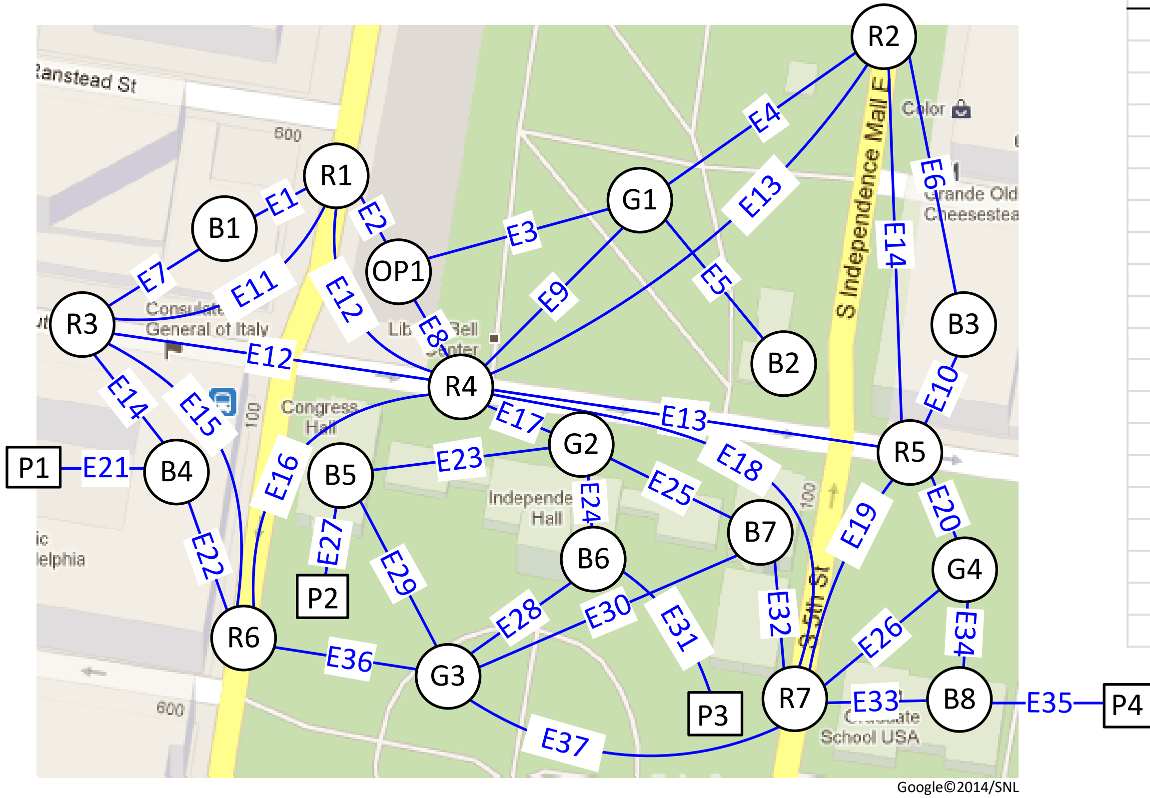
UVM

Normalized Digital Surface Model
(nDSM)

Data provided by Jarlath O'Neil-Dunne at the University of Vermont Spatial Analysis Lab.

Example Geospatial Semantic Graph

Independence Hall, Philadelphia:



Region node table:

id	type	area	centroid x	centroid y
B1	building	3200	-75.14900	39.94939
R1	road	1800	-75.14910	39.94949
OP1	paved	4700	-75.14935	39.94934
G1	grass	22000	-75.15010	39.94944
R2	road	1900	-75.15060	39.94999
R3	road	1100	-75.14885	39.94934
R4	road	2200	-75.14980	39.94924
B2	building	780	-75.15045	39.94931
B3	building	6000	-75.15075	39.94944
B4	building	12000	-75.14895	39.94884
B5	building	2100	-75.14920	39.94899
G2	grass	7700	-75.14990	39.94906
R5	road	870	-75.15065	39.94896
B6	building	2000	-75.15000	39.94889
B7	building	3150	-75.15040	39.94884
G4	grass	15300	-75.15080	39.94869
R6	road	1970	-75.14905	39.94844
G3	grass	25000	-75.14960	39.94829
R7	road	1810	-75.15050	39.94834
B8	building	2700	-75.15090	39.94819

Edge table:

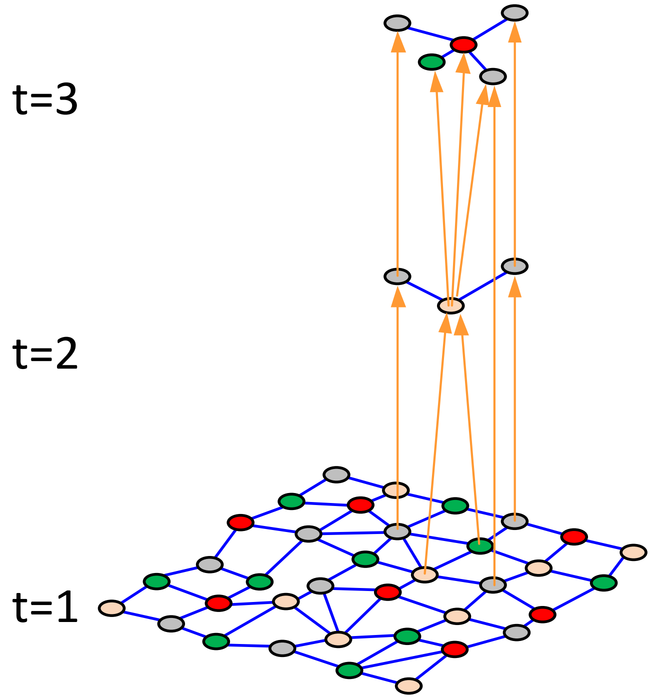
edge_id	node_1	node_2
E1	B1	R1
E2	R1	OP1
E3	OP1	G1
E4	G1	R2
E5	G1	B2
E6	R2	B3
E7	R3	B1
E8	OP1	R4
E9	R4	G1
.	.	.
.	.	.
.	.	.

Point node table:

id	Name	Address	Latitude	Longitude
P1	Consulate of Italy	150 S. Independent Mall West #1026	-75.14895	39.94884
P2	Congress Hall	41 N 6th Street	-75.14920	39.94899
P3	Independence Hall	520 Chestnut Street	-75.15000	39.94889
P4	Graduate School USA	150 S. Independence Mall West #674	-75.15090	39.94819

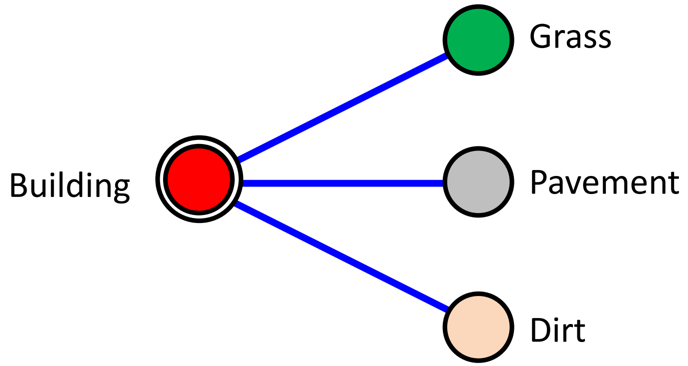
Representing Change Over Time

- Encode change:
  - Node attributes include duration seen.
  - Only construct new nodes for changes.
  - “Changed-to” arcs encode time evolution.
  - Graph complexity focuses on change areas.

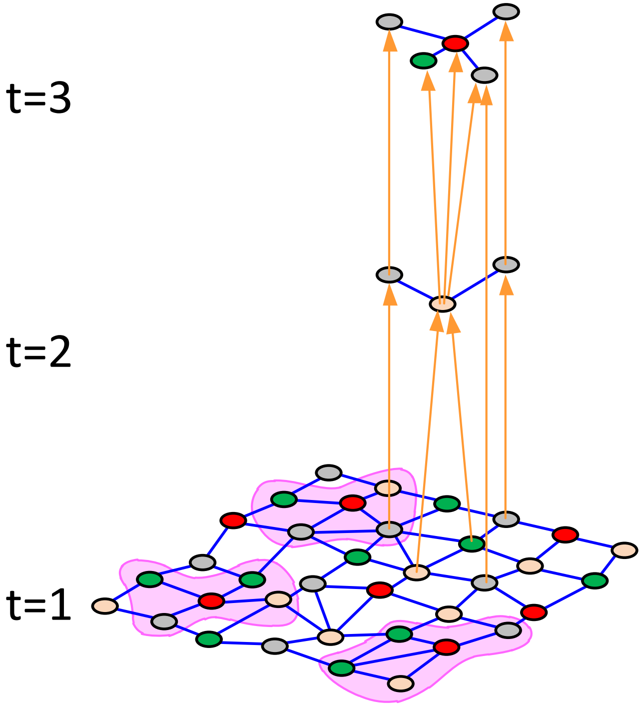


Signature Search

- A signature encodes a desired question.
- Example: “Where are buildings with nearby grass, pavement, and dirt?

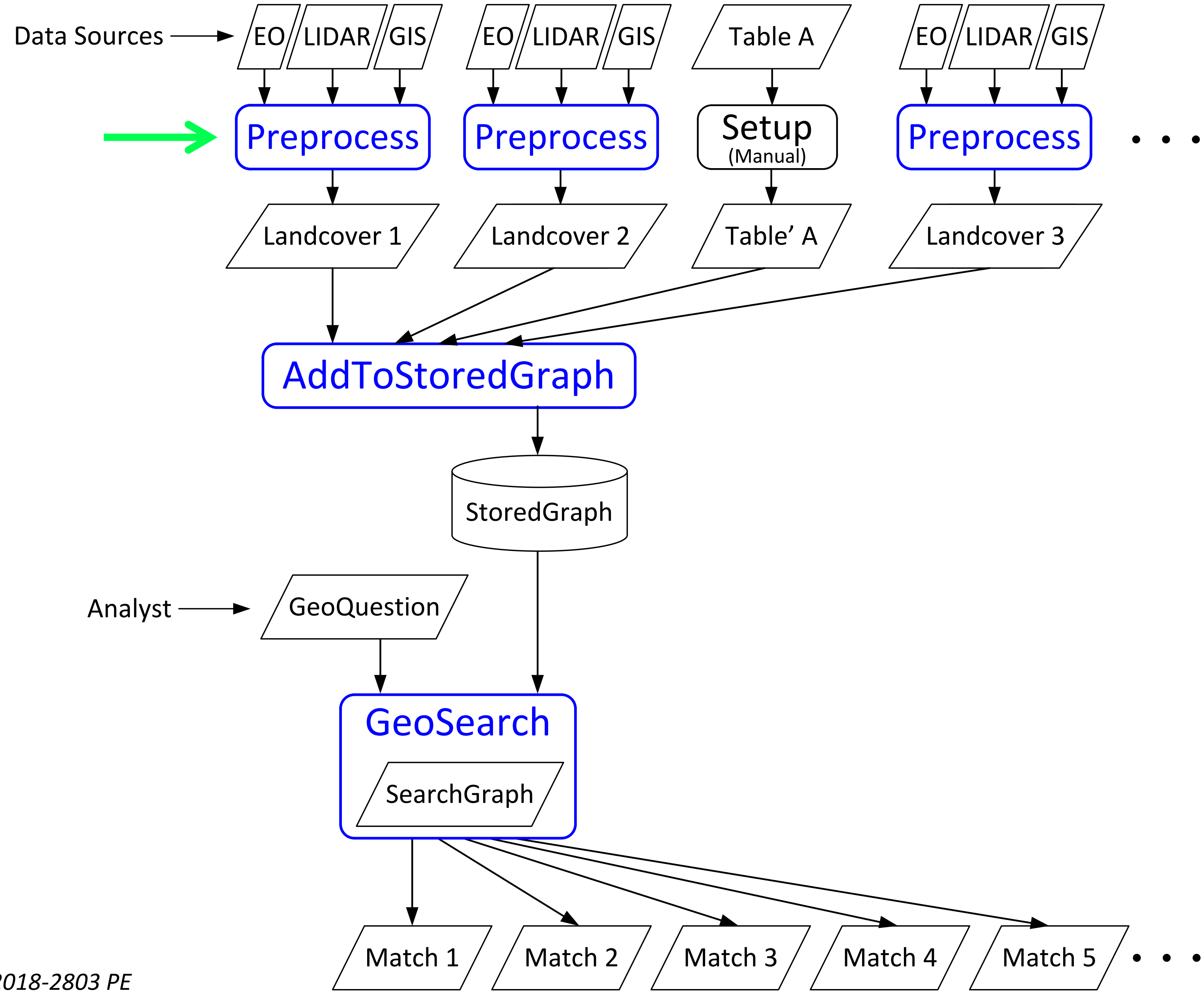


Query Template



Search Results

Primary Data Flow



Graph Search

Initial implementation [Watson 2010]:

- Subgraph isomporhism.
- NP-complete! [Cook 1971, Ullmann 1976]

Current approach:

- Identify relevant portion of StoredGraph.
- Lazy constraint, distance edge evaluation, with caching.
- Simple graph search algorithms:
  - Star graph.
  - Connected component.
  - Interrupted star, using transitive closure.
- Postprocessing calculations.

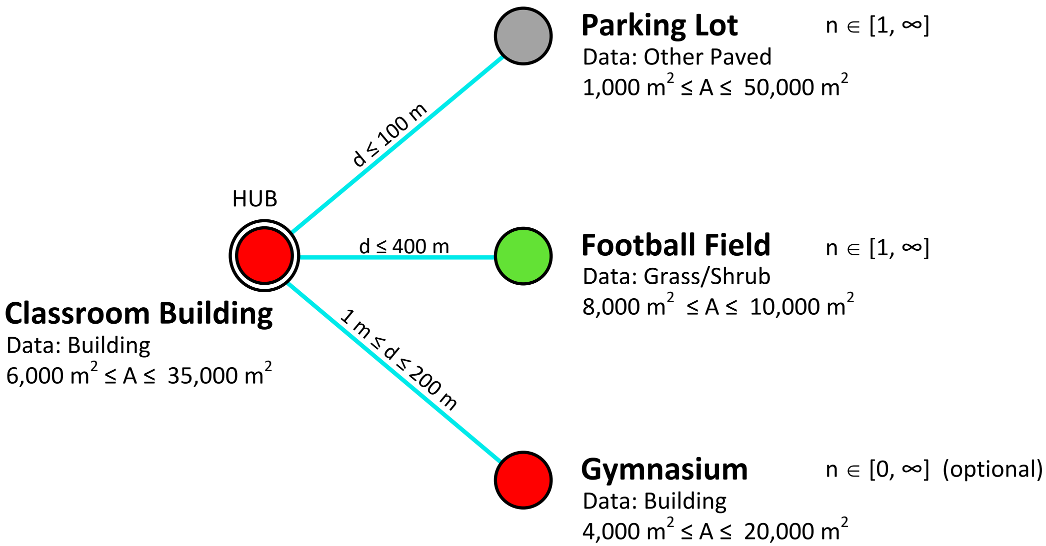
[Cook 1971] S. A. Cook, "The Complexity of Theorem-Proving Procedures," 3rd ACM Symposium on Theory of Computing, pp. 151–158, 1971.

[Ullmann 1976] J. R. Ullmann, "An Algorithm for Subgraph Isomorphism," Journal of the ACM 23(1), pp. 31–42, 1976.

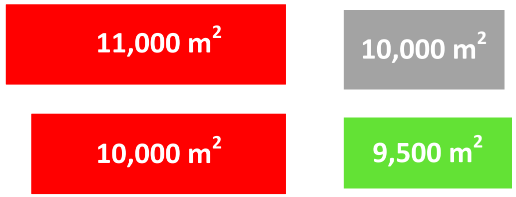
[Watson 2010] J. P. Watson, “Complex Signature Detection Using Geospatial/Temporal Semantic Graphs,” Simulations, Algorithms, and Modeling Program Review Meeting (SAM2010), April 2010.

Match Ambiguity

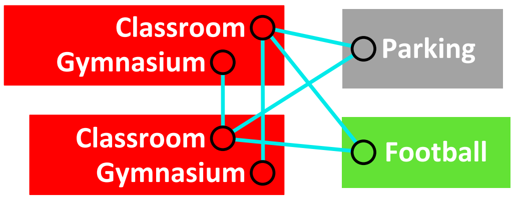
Example high school search:



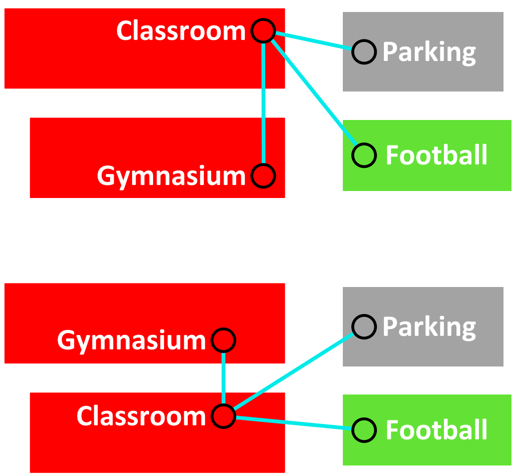
StoredGraph:



SearchGraph:



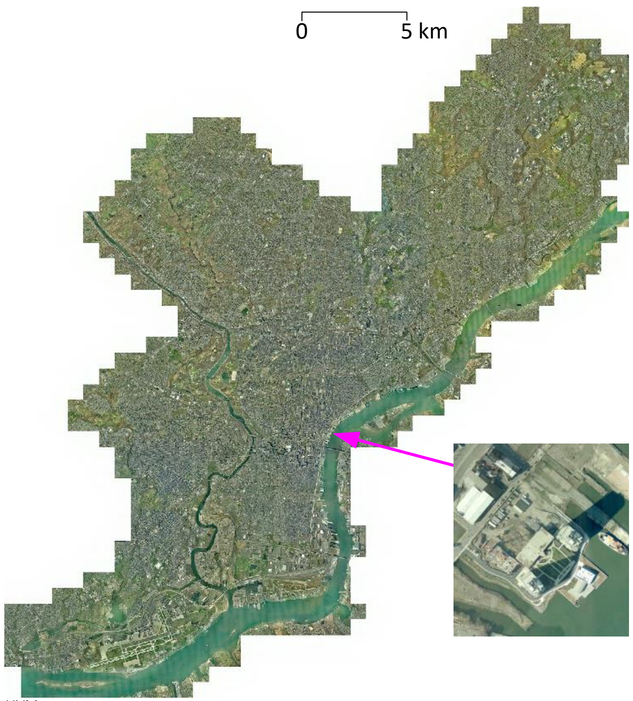
Matches:



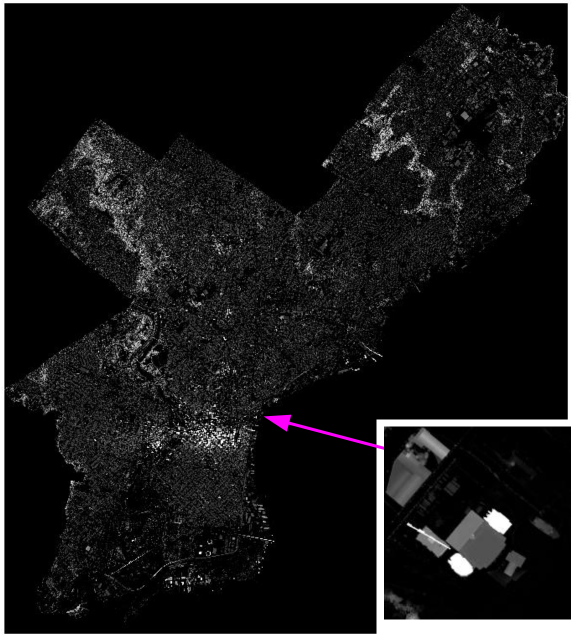
Match grouping provides a practical solution.

Philadelphia 2008

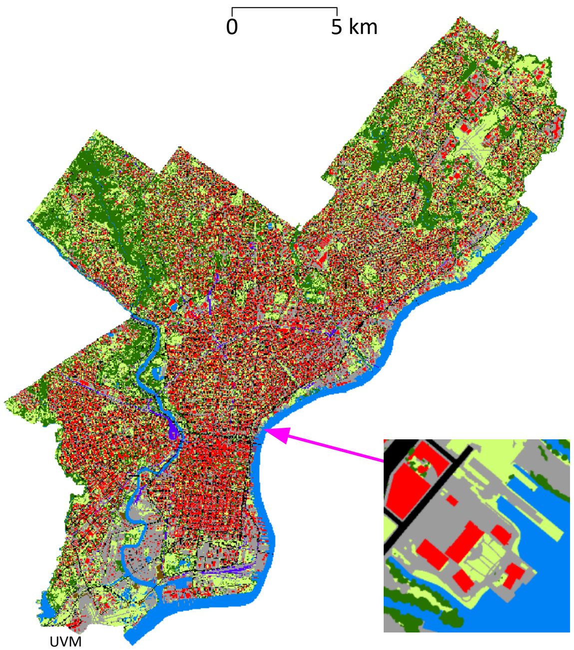
Primary input:



Optical Image



LiDAR nDSM



Land Cover

Pixel size 0.1 m
307,531 × 330,033 pixels
(101.5 Gpix)
7,669 MB

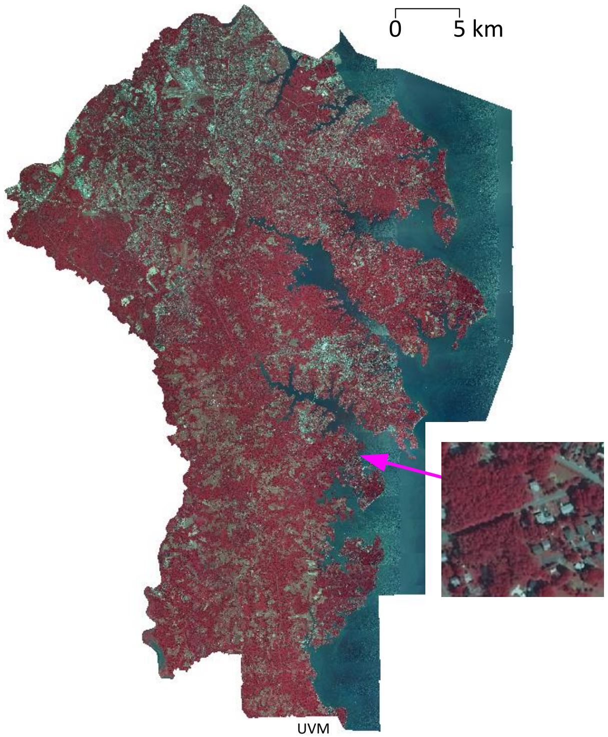
Pixel size 0.3 m
89,540 × 100,294 pixels
(9.0 Gpix)
2,084 MB

Pixel size 0.3 m
89,548 × 100,303 pixels
(9.0 Gpix)
8,775 MB

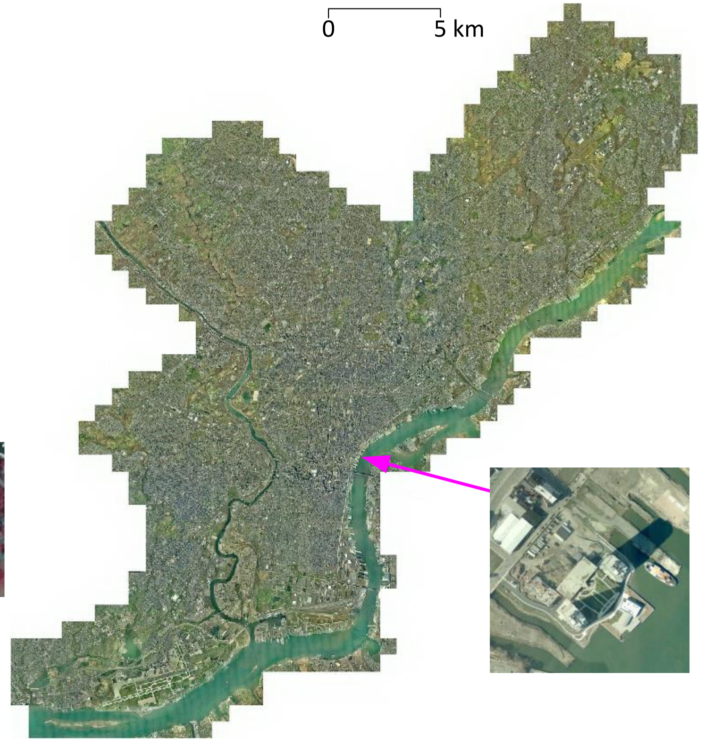
Image data provided by the University of Vermont.

Three Data Regions

Largest input images:



Anne Arundel County, MD



Philadelphia, PA

Washington, DC

Total:

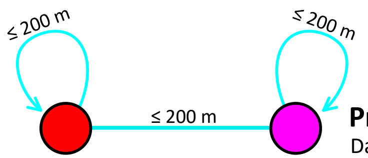
2,067	km^2 total area
135 billion	Pixels
3.6 million	Features

Total file size was about 88 GB.

Image data provided by the University of Vermont.

Large Refinery Search

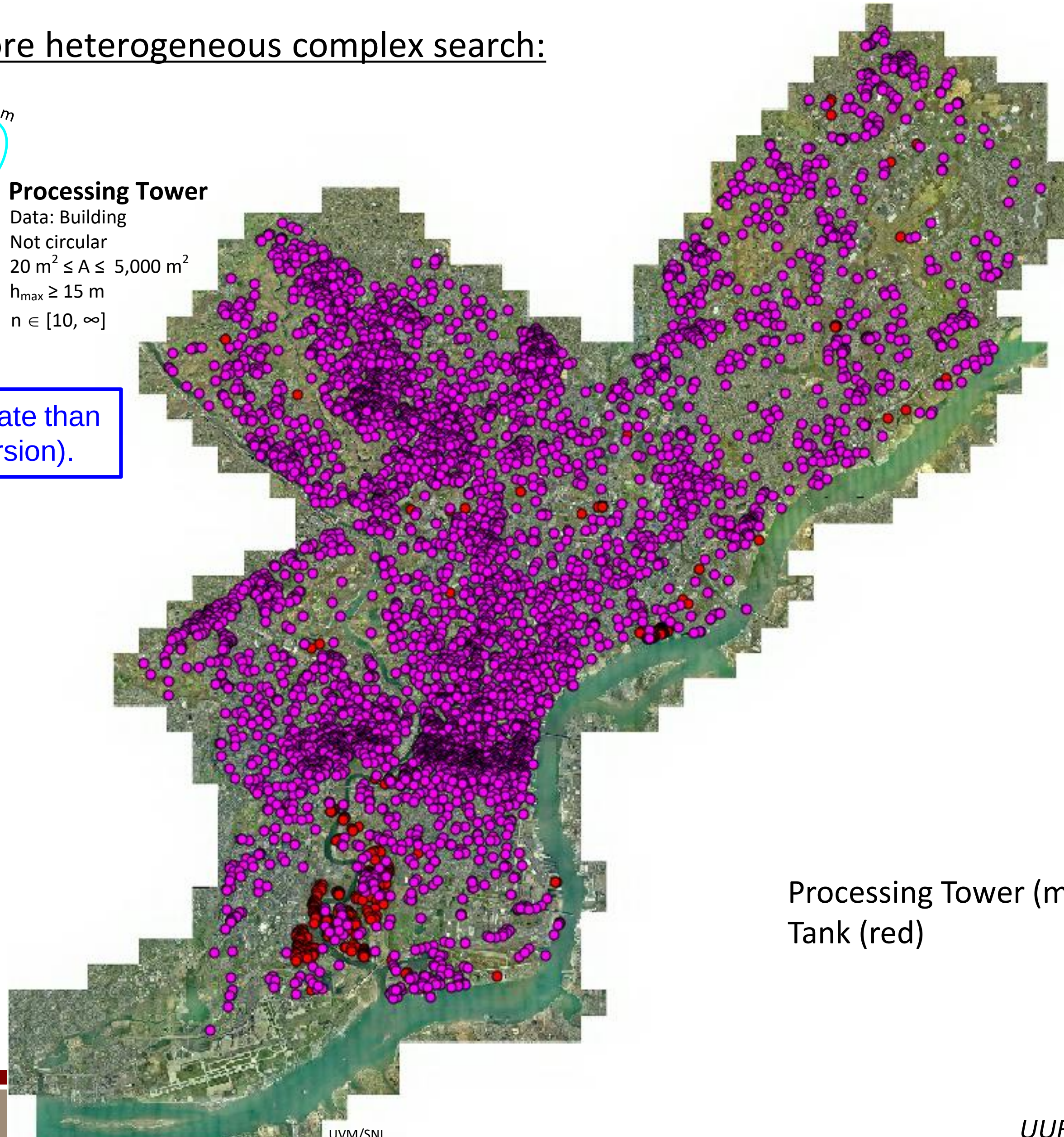
SearchGraph, before heterogeneous complex search:



Tank
Data: Building
Circular
$150\ m^2 \leq A \leq 3{,}000\ m^2$
$n \in [10, \infty]$

Processing Tower
Data: Building
Not circular
$20\ m^2 \leq A \leq 5{,}000\ m^2$
$h_{max} \geq 15$ m
$n \in [10, \infty]$

Slightly different template than above (an earlier version).

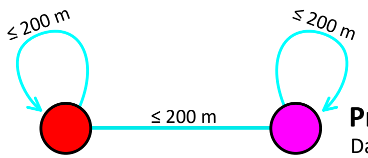


Processing Tower (magenta)	4,909
Tank (red)	371

Image data provided by UVM.

Large Refinery Search

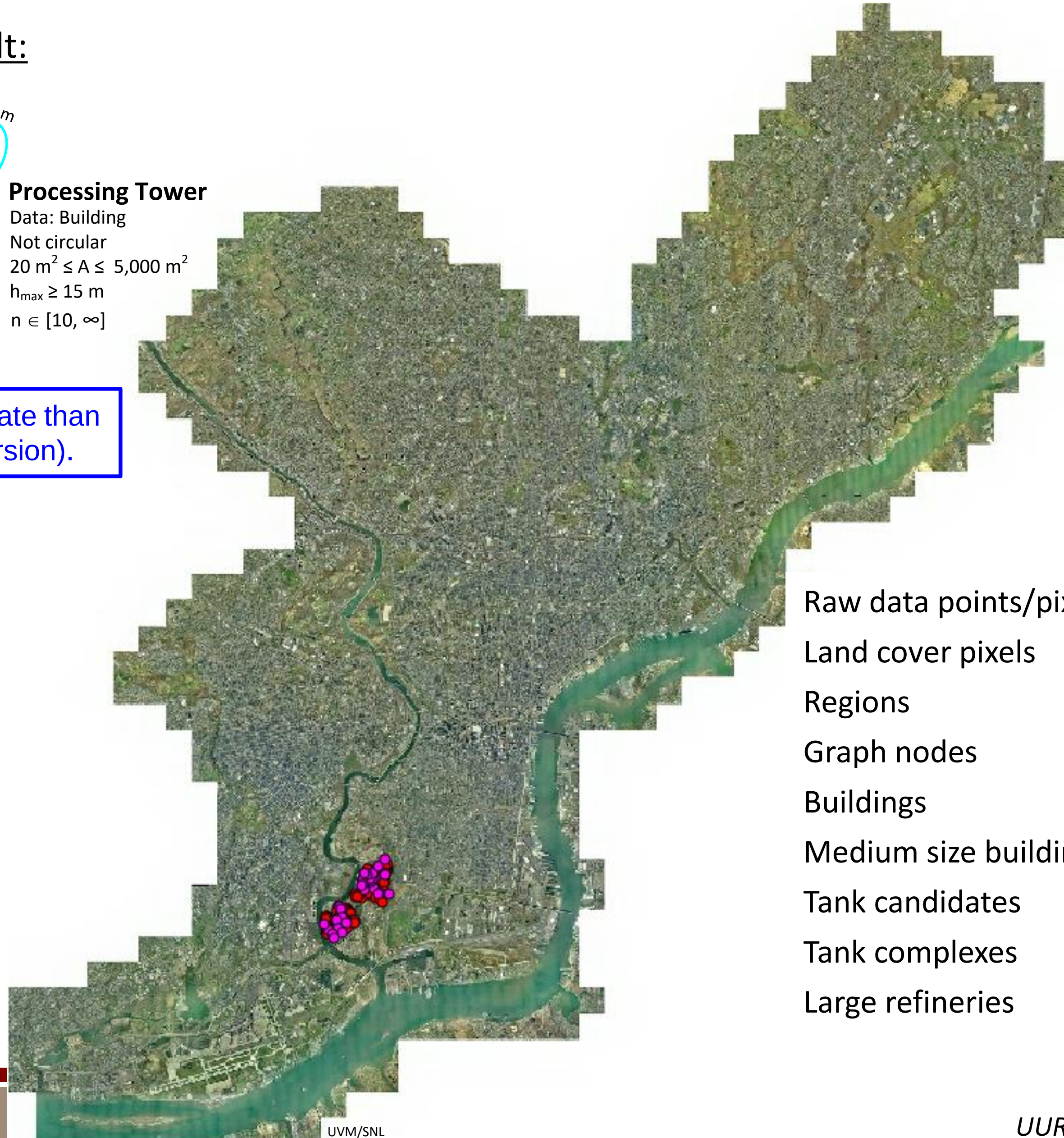
Graph search result:



Tank
Data: Building
Circular
150 m^2 ≤ A ≤ 3,000 m^2
n ∈ [10, ∞]

Processing Tower
Data: Building
Not circular
20 m^2 ≤ A ≤ 5,000 m^2
h_{max} ≥ 15 m
n ∈ [10, ∞]

Slightly different template than above (an earlier version).



0 False positives
0 False negatives

Raw data points/pixels	101,495,378,523
Land cover pixels	8,981,933,044
Regions	1,133,822
Graph nodes	1,133,822
Buildings	154,062
Medium size buildings	87,170
Tank candidates	371
Tank complexes	28
Large refineries	2

Image data provided by UVM.

Recap

We have demonstrated:

- Imagery + LiDAR + GIS $\Rightarrow$ Land cover [O'Neil-Dunne 2012].
- Sequence $\Rightarrow$ Geospatial-temporal graph.
- Spatial search: power plants, refineries, high schools…
- Spatial-temporal search: change, construction complexes…
- Over a wide area (2,067 km^2, 135 billion pixels, 3.6 million graph nodes).
- Multi-modality.
- Quality scoring.

We have NOT shown:

- Continent-scale robust image supply and pre-processing.
- Recognition scope.
- Complex temporal analysis.
- Open issues:
  Multiple hypotheses, match ambiguity, scale.

UUR SAND 2018-2803 PE

[O'Neil-Dunne 2012] O'Neil-Dunne, et al, An object-based system for LiDAR data fusion and feature extraction, Geocarto (28), pp. 227–242, 2012.