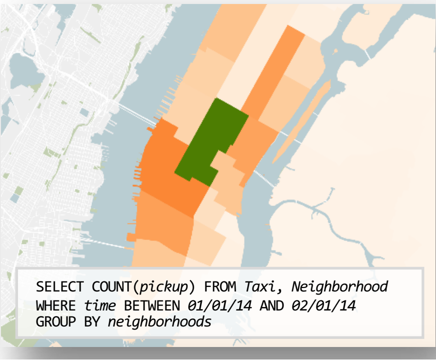
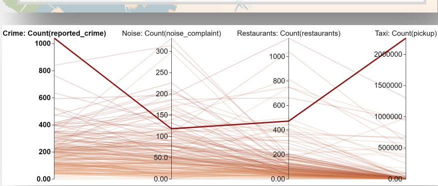
Interactive Visual Exploration of Spatio-Temporal Urban Data Sets using Urbane

Harish Doraiswamy, Eleni Tzirita Zacharatou, Fabio Miranda,
Marcos Lage, Anastasia Ailamaki, Claudio Silva, Juliana Freire

Visual Exploration using Urbane





Urbane

- Help architects in city planning.
- Visualize urban data.
- Visually compare multiple data.
- Multiple space and time resolutions.

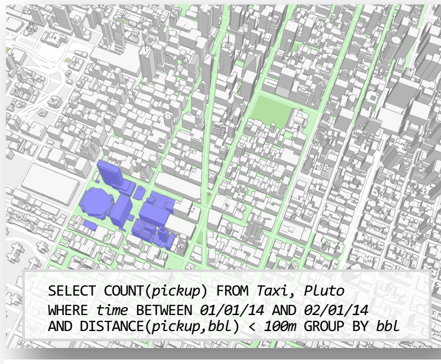
Spatial Aggregation Queries

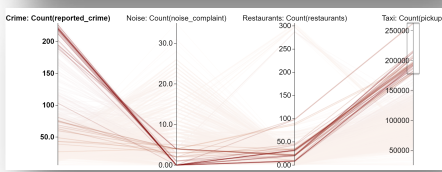
- Spatial join between points and polygons.
- Computationally-intensive Point-in-Polygon tests.

Challenges

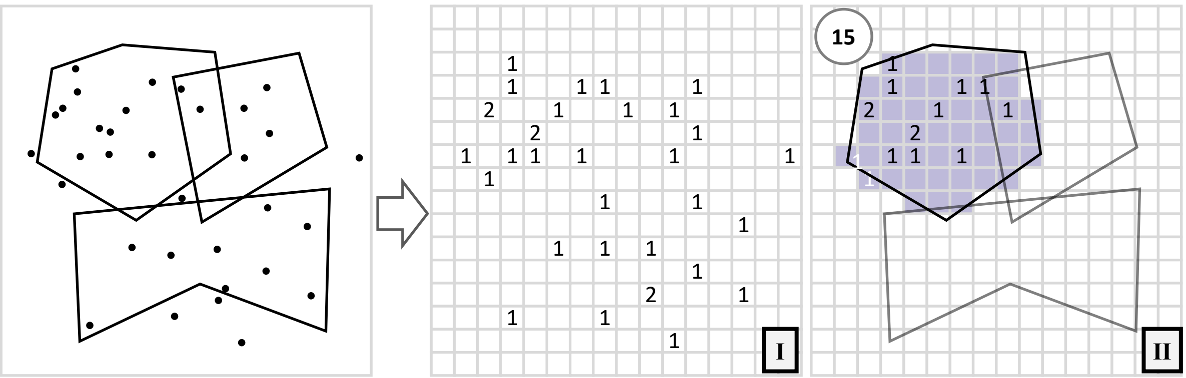
- Hundreds of millions of points.
- Arbitrary polygons.
- Query parameters change interactively.

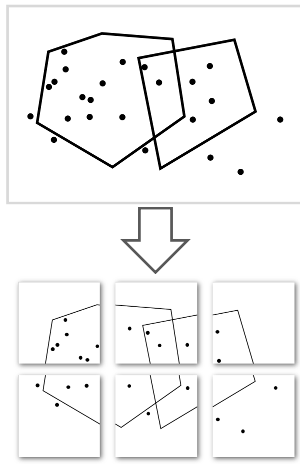
Existing solutions don't support interactive responses.





Raster Join: A GPU Rasterization-based Approach





Key Ideas

- Decompose spatial operators into graphics primitives.
- Couple join and aggregation.
- Trade off accuracy for interactivity.

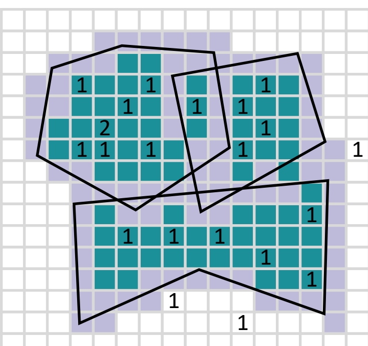
I. Render Points: Aggregate points within each pixel, and store result in pixel color channels.
II. Render Polygons: Aggregate pixel values inside the polygons.

Accuracy Bound

- Bound the Hausdorff distance between the input and the pixel-approximated polygons.
- Smaller pixel size → higher accuracy.

- GPU resolution might be insufficient for given accuracy bound.
- Split canvas to increase accuracy.

Accurate Variant



- Point-in-Polygon tests only at the boundary.

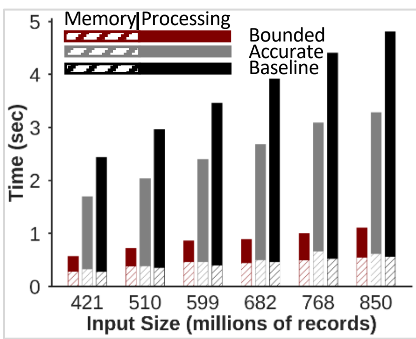
Experimental Evaluation

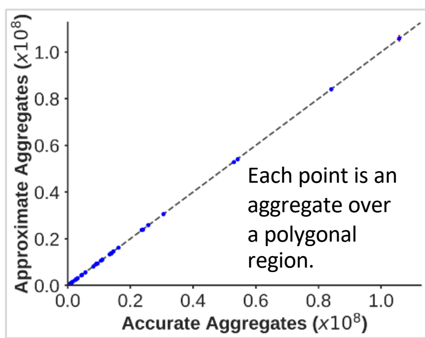
Hardware

Intel Core i7 Quad-Core @2.8 GHz, 16GB RAM. NVIDIA GTX 1060 GPU, 6GB VRAM (usage limited to 3GB). OpenGL implementation.

Data Sets

NYC Taxi data (over 868 million points), 260 NYC neighborhood polygons.





Performance

- Only 1.1 seconds for 850 million points.

Accuracy

- All points close to the diagonal → negligible errors.

Data Sets used in the Demonstration

Point Data Sets

Name	# Points	# Attributes
Taxi	380,633,852	6
Restaurants	24,957	2
Sky Exposure	379,387	4
Schools	1,817	3
Pluto	42,638	8
Crime	939,526	3
Subway	470	2
Noise	274,155	3

Regions

Name	# Polygons	Avg. size of Polygon
Lots	42,638	11.6
Neigh-borhoods	357	587.9
Zip Codes	263	1,061.9

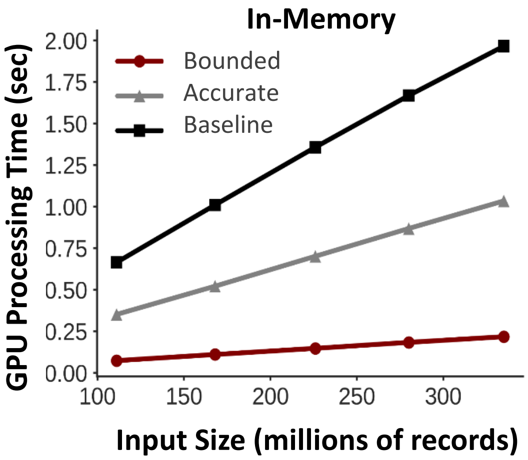
Street Network: Graph with 379,387 nodes

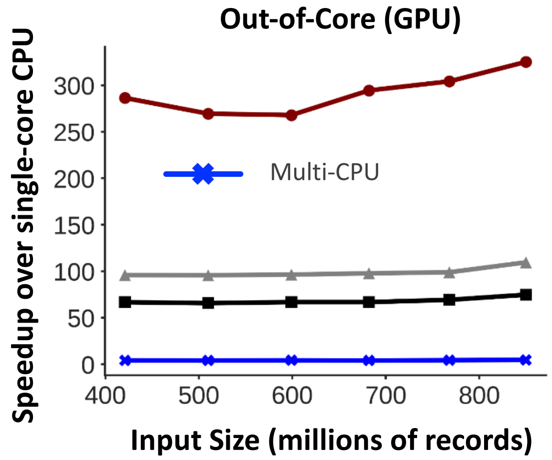
Additional Experimental Results

Databases don't support interactive responses:

- A join between only 10 neighborhood polygons and the taxi data took over 10 minutes.
- Bounded Raster Join takes only 1.1 seconds for 260 neighborhood polygons and 850 million points.

Performance

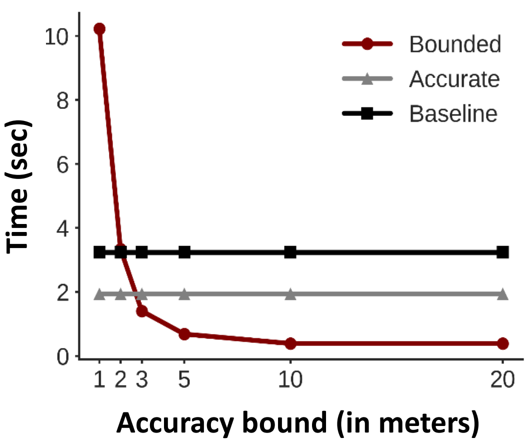


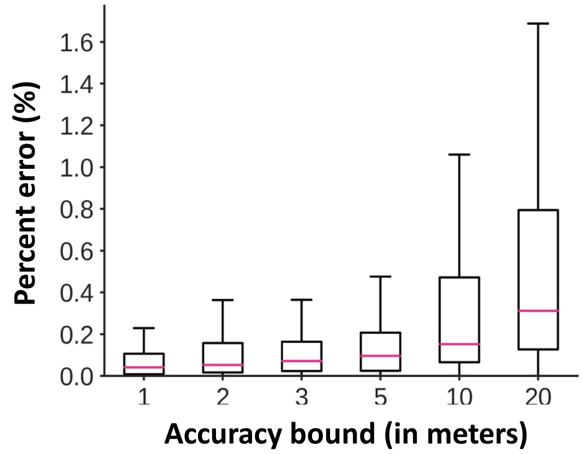


Bounded Raster Join:

- is over 4 times faster than the accurate versions.
- achieves speedup of over two orders of magnitude over a single-core CPU implementation.

Trading off accuracy for interactivity





As the accuracy bound becomes tighter:

- more rendering passes are required, increasing the query time.
- the approximate aggregate results converge towards the accurate values.

The Raster Join approach has been published at PVLDB 11, 3 (2017) (to be presented at VLDB 2018) under the title:
GPU Rasterization for Real-Time Spatial Aggregation over Arbitrary Polygons.

E. Tzirita Zacharatou (eleni.tziritazachatatou@epfl.ch), H. Doraiswamy (harishd@nyu.edu),
A. Ailamaki (anastasia.ailamaki@epfl.ch), C. T. Silva (csilva@nyu.edu), and J. Freire (juliana.freire@nyu.edu).

We have made the code available at: https://github.com/vida-nyu/raster-join