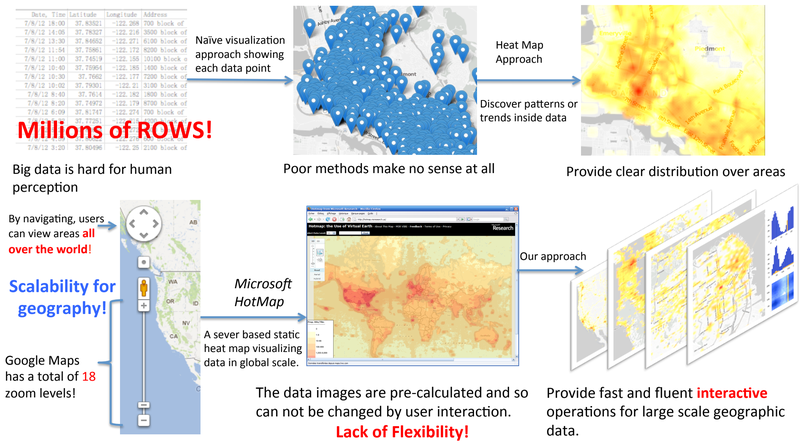
Interactive Visual Analysis of Large-Scale Geographic Data using WebGL

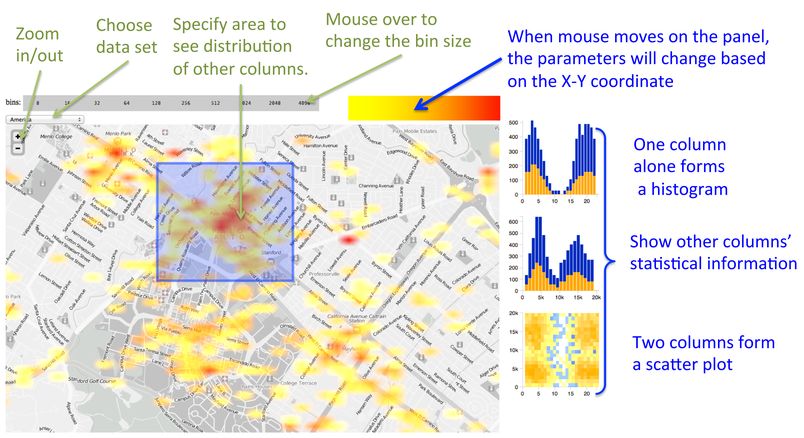
Biye Jiang[1,2], Zhicheng Liu[1] and Jeffrey Heer[1]
[1]Visualization Group, Department of Computer Science, Stanford University
[2]Department of Computer Science & Technology, Tsinghua University

1. Overview

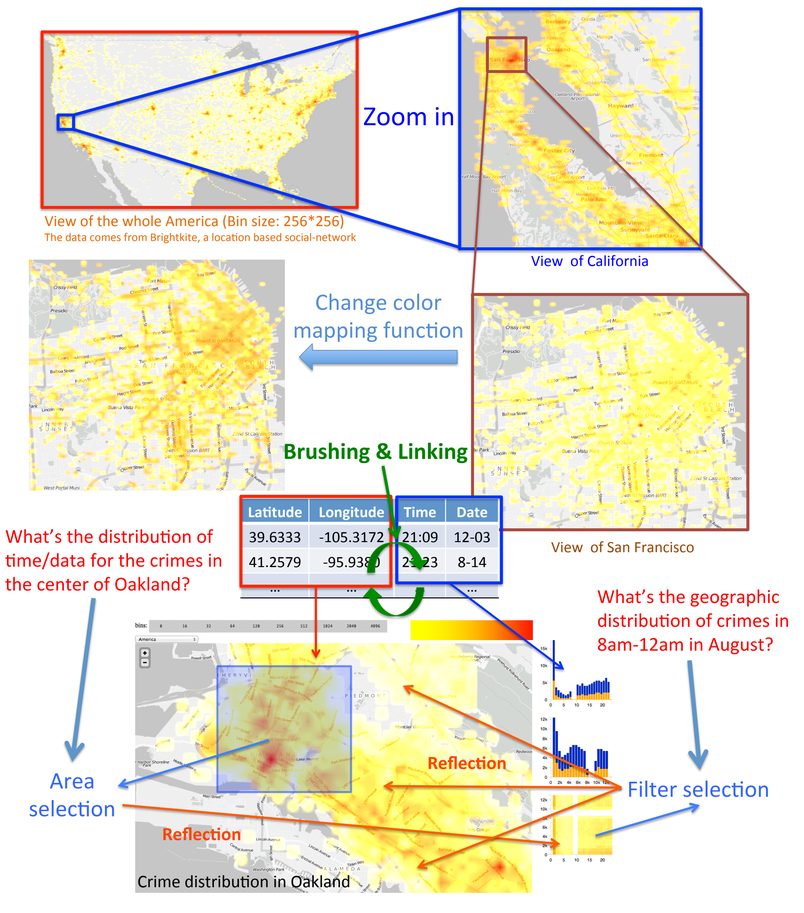


Our novel approach applies WebGL-based parallel computation to enable rapid interaction in the browser.

2. User Interface



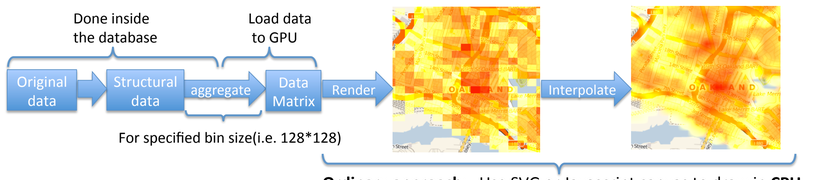
3. Dynamic Interaction



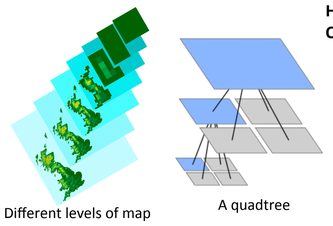
Latitude	Longitude	Time	Date
39.6333	-105.3172	21:09	12-03
41.2579	-95.9380	23	8-14
...	...		...

4. System Architecture

Visualization Pipeline



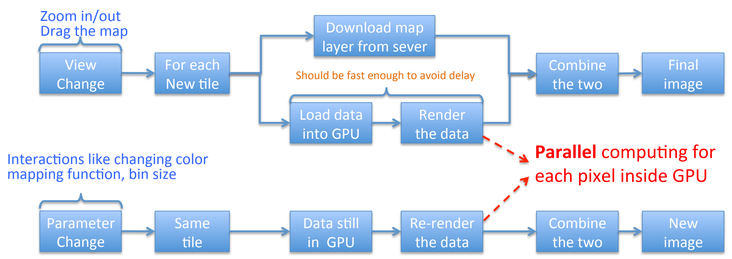
Ordinary approach: Use SVG or Javascript canvas to draw in **CPU**
HotMap approach: Pre-render and store the images on the **server**
Our approach: Use WebGL to render pixels parallelly in **GPU**


Different levels of map

A quadtree

Map layer:

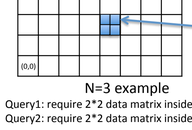
-We use the Open-Source JavaScript library *Leaflet* to build the map.
-*Leaflet* manages canvas and cache strategy in the concept of **tiles**. Tiles are organized hierarchically in a **quadtree.**



Goal: Loading and rendering must be fast enough to enable smooth interaction.

Database approach:

The database should return a data matrix with the resolution of R*R for a specified tile X.



N=3 example
Query1: require 2*2 data matrix inside(4,4,7,7)
Query2: require 2*2 data matrix inside(4,2,4,2)

Option1: Load the data from server. (Efficient but less flexible)
Option2: Load the data form browser memory. (We currently use this)

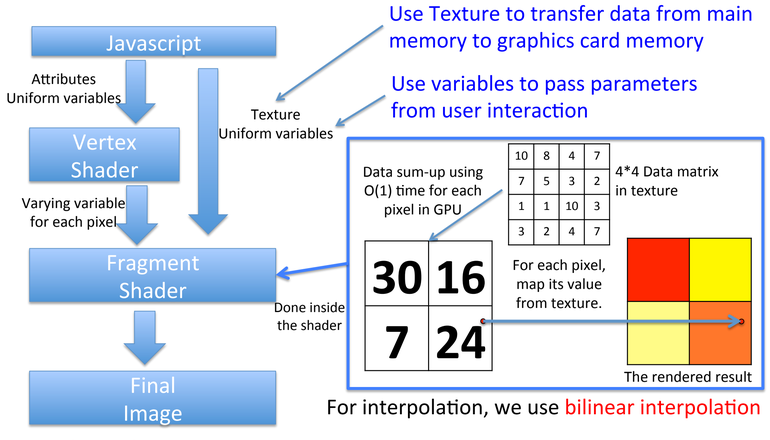
Query1: Return pre-calculated data

Query2: Re-load source data and rebuild

The complexity is $O(M+2^{n'}*2^{n'})$ if there are M data records in that area.
Set N=12 thus M should be small!

5. WebGL-based Rendering

WebGL is a 3D graphics API based on OpenGL ES 2.0 and includes the GLSL shader API. Its pipeline can be described as follows:

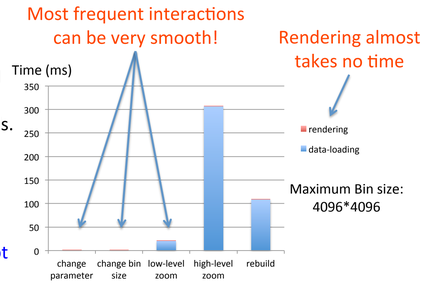


For interpolation, we use bilinear interpolation

Benchmark:

Platform:
MacBook Pro with i7 2.6GHz, 8GB, GT650M
Data set:
Brightkite in America. 2919613 data records.
Evaluating method:
Calculate the time used inside redraw tiles function of leaflet.

The result shows that parallel computation greatly improves performance and JavaScript data processing becomes the bottleneck.



6. Color Perception

Factors to consider in color scheme design:

1. Highlight areas that have very high values.
2. Distinguish data of different levels.

Approach 1:
$$color = k \cdot \log(data)$$
Approach 2:
$$color = (\frac{data}{Max\{data\}})^k$$
Approach 3:

Different parameters give different visual results when mapping 0~1 float number into RGBA color space. Use approach 3

$$color = \begin{cases} p + (\frac{data}{Max\{data\}} - \frac{average}{Max\{data\}})^k * (1-p) & (data > average) \\ (\frac{average}{Max\{data\}} - \frac{data}{Max\{data\}})^k * p & (data < average) \end{cases}$$

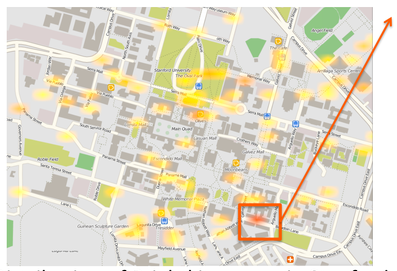
We generally set $k = \frac{1}{3}$

And one can use median instead of average as well

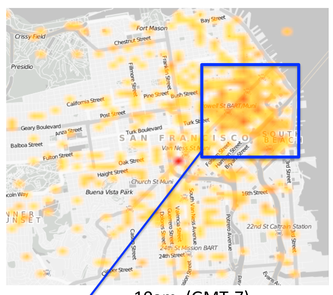
$$FinalColor = rgba(255, (1-color)*255, 0, 255*\sqrt{color})$$

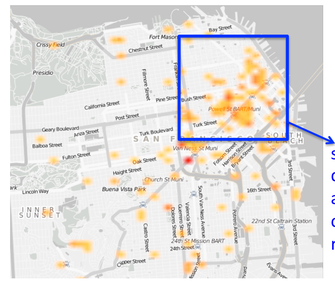
In practice, approach 3 can fit different distribution as users can change *p* to give an "average point" a reasonable color.

7. Applications

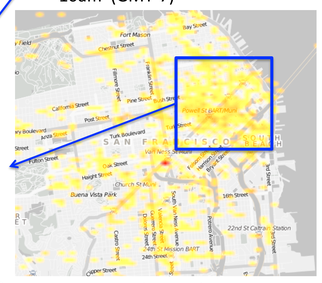


Distribution of Brightkite usage in Stanford.


10am (GMT-7)

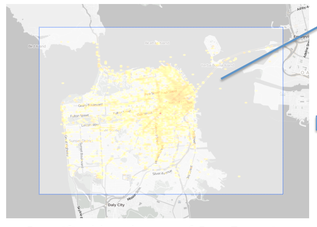

6am (GMT-7)

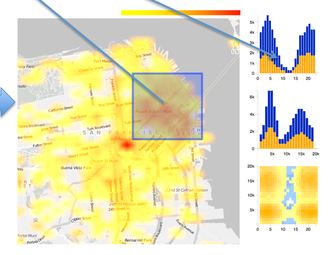
similar patterns in different periods of a day may indicate commercial or residential area


9pm (GMT-7)

$$Ratio\ of\ \frac{Blue\ area}{White\ area}$$

A high ratio of the blue area supports the inference above


Specified local area of San Francisco



8. Summary

-Our work provides an interactive approach for large scale geographic data analysis.
-Our system supports real-time interaction, enabling users to rapidly explore large location-based data sets.

Future work:

Incorporate back-end database server and explore caching/pre-fetching schemes
More interactive operations and visualization methods should be tried and applied.

Reference:

Danyel Fisher ***Hotmap: Looking at Geographic Attention*** (Microsoft)
Sean Kandel, et al. ***Profiler: Integrated Statistical Analysis and Visualization for Data Quality Assessment*** (Stanford University)

Acknowledgements:

Stanford Undergraduate Visiting Research (UGVR) program

Contact: byeah@stanford.edu