SAFE: A Share-and-Aggregate Bandwidth Exploration Framework for Kernel Density Visualization

Tsz Nam Chan
Hong Kong Baptist University
edisonchan@comp.hkbu.edu.hk

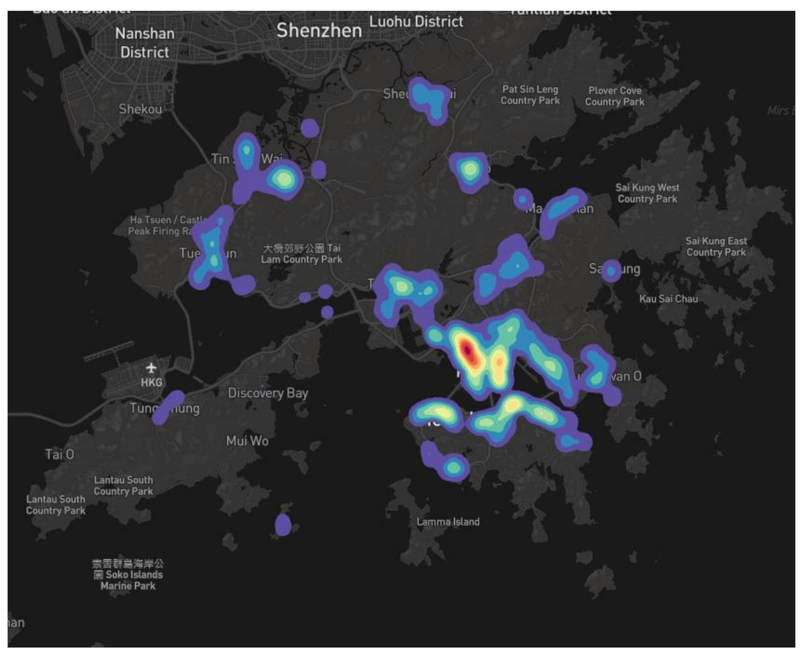
Pak Lon Ip
University of Macau
SKL of Internet of Things for Smart City
paklonip@um.edu.mo

Leong Hou U
University of Macau
SKL of Internet of Things for Smart City
ryanlhu@um.edu.mo

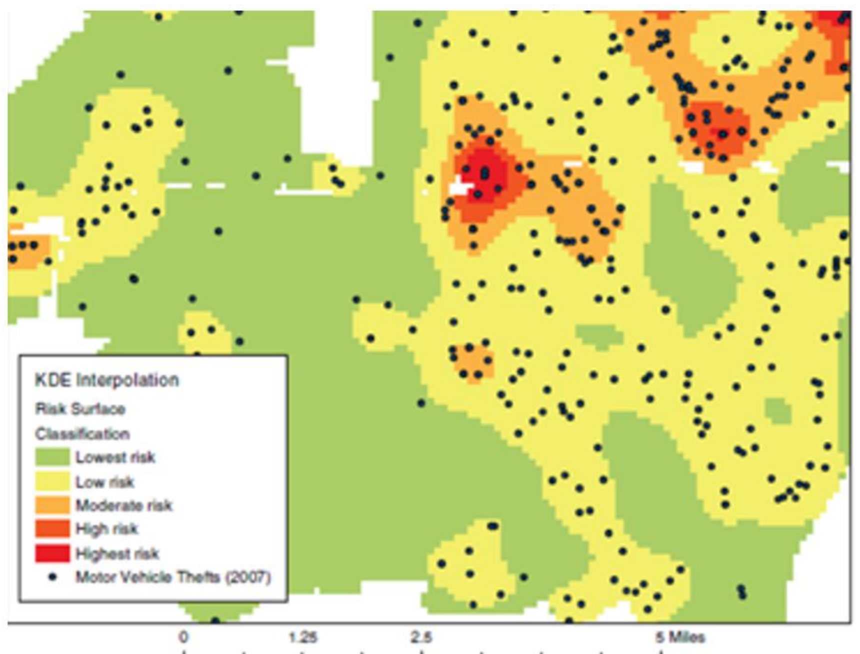
Byron Choi
Hong Kong Baptist University
bchoi@comp.hkbu.edu.hk

Jianliang Xu
Hong Kong Baptist University
xujl@comp.hkbu.edu.hk

Overview of Kernel Density Visualization (KDV)



(a) COVID-19 hotspot map (in Hong Kong)



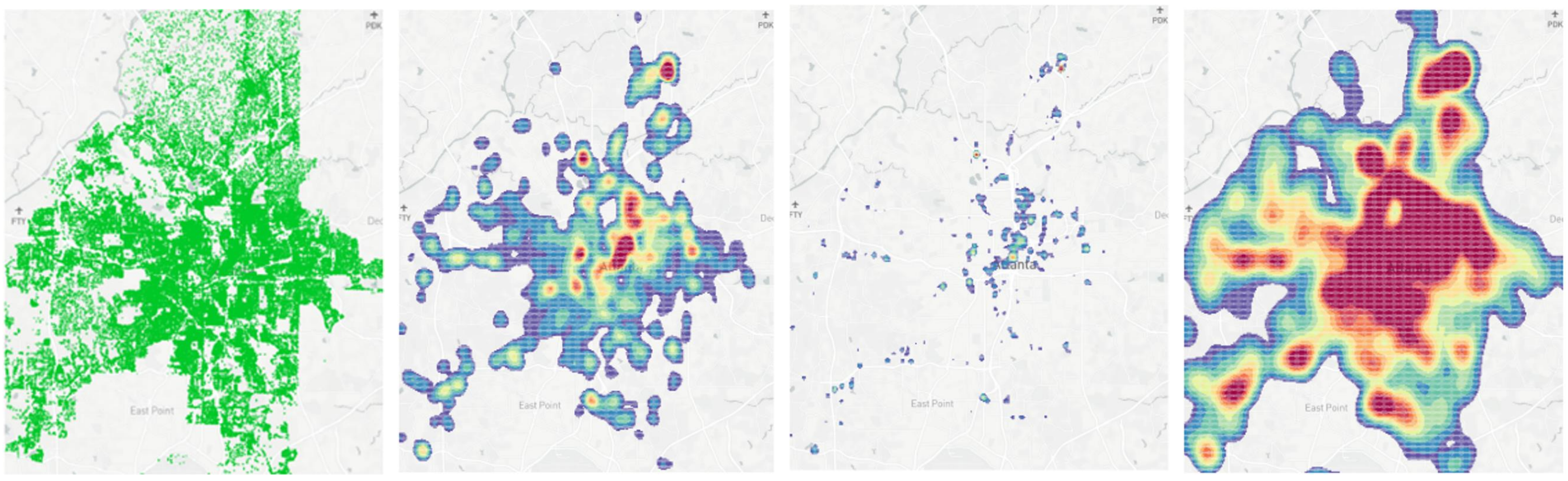
(b) Crime hotspot map (in Arlington, Taxes)

Given a location dataset P (e.g., black dots in (b)), we need to color each pixel $\mathbf{q}$ based on the kernel density function $\mathcal{F}_P^{(b)}(\mathbf{q})$.

$$\mathcal{F}_P^{(b)}(\mathbf{q}) = \sum_{\mathbf{p}\in P} w \cdot \begin{cases} 1 - \frac{1}{b^2} dist(\mathbf{q},\mathbf{p})^2 & \text{If } dist(\mathbf{q},\mathbf{p}) \le b \\ 0 & \text{Otherwise} \end{cases}$$

(2D pixel: $\mathbf{q}$; weighting: w; Euclidean distance: $dist(\mathbf{q},\mathbf{p})$; dataset: P; bandwidth: b)

KDV is Sensitive to the Bandwidth Parameter b



Crime location data (in Atlanta) | b is small (undersmoothing) | b is moderate | b is large (oversmoothing)

Domain experts choose multiple bandwidths, $b_1, b_2, \ldots, b_L$, specify the resolution size of the plane, $X \times Y$, and provide the location dataset $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ with size n.

Our goal: Generate L KDVs with respect to these bandwidth parameters.

Time complexity: $O(LXYn)$ time ☹

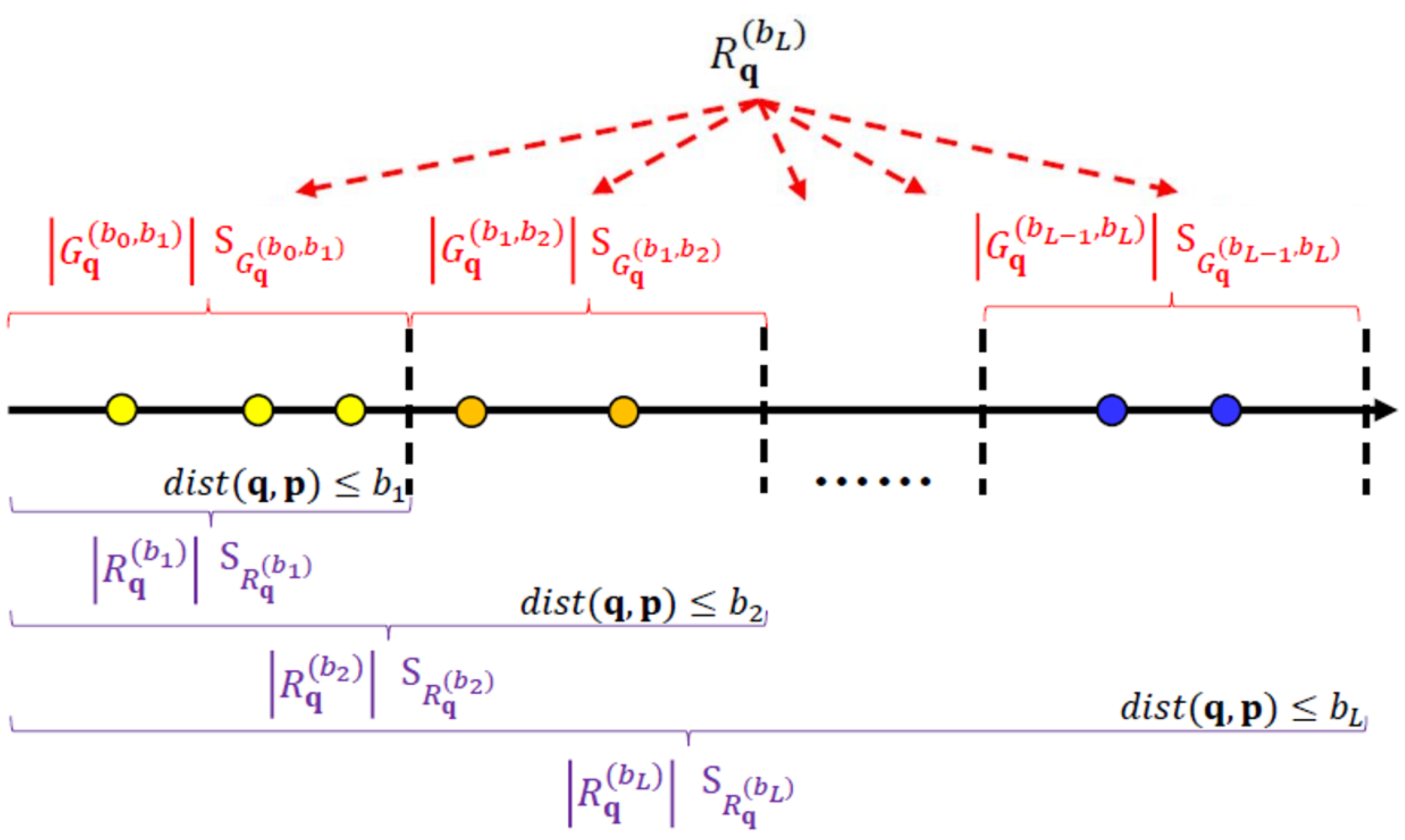
Share-and-Aggregate Framework (SAFE)

$$\mathcal{F}_P^{(b)}(\mathbf{q}) = \sum_{\mathbf{p}\in P} w \cdot \begin{cases} 1 - \frac{1}{b^2} dist(\mathbf{q},\mathbf{p})^2 & \text{If } dist(\mathbf{q},\mathbf{p}) \le b \\ 0 & \text{Otherwise} \end{cases}$$

$$= \sum_{\mathbf{p}\in R_{\mathbf{q}}^{(b)}} w \cdot \left(1 - \frac{1}{b^2} dist(\mathbf{q},\mathbf{p})^2\right)$$

$$= w\left|R_{\mathbf{q}}^{(b)}\right| - \frac{w}{b^2} S_{R_{\mathbf{q}}^{(b)}}$$

where $S_{R_{\mathbf{q}}^{(b)}} = \sum_{\mathbf{p}\in R_{\mathbf{q}}^{(b)}} dist(\mathbf{q},\mathbf{p})^2$



Share
$O(n \log L + L)$ time for each pixel $\mathbf{q}$

Aggregate
$O(n + L)$ time for each pixel $\mathbf{q}$

Theoretical Results

Method	Time complexity	Space complexity	Quality	Bandwidth properties
Baseline (cf. Section 2.2)	$O(LXYn)$	$O(XYL + n)$	Exact	Known in advance On-the-fly
SAFE (cf. Section 3.2)	$O(XY(n \log L + L))$ (cf. Theorem 1)	$O(XYL + n)$ (cf. Theorem 3)	Exact	Known in advance
SAFE_{all} (cf. Section 4.1)	$O(XY(n + L) \log n)$ (cf. Theorem 4)	$O(XY(n + L))$ (cf. Theorem 5)	Exact	On-the-fly
SAFE_{exp} (cf. Section 4.2)	$O(XY(n \log n + L \log(\log n)))$ (cf. Theorem 8)	$O(XY(\log n + L) + n)$ (cf. Theorem 7)	2-approximation (cf. Theorem 6)	On-the-fly

Experimental Results