A Fast and Accurate Block Compression Solution for Spatiotemporal Kernel Density Visualization

Yue Zhong
Shenzhen University
2310273008@email.szu.edu.cn

Tsz Nam Chan
Shenzhen University
edisonchan@szu.edu.cn

Leong Hou U
University of Macau
ryanlhu@um.edu.mo

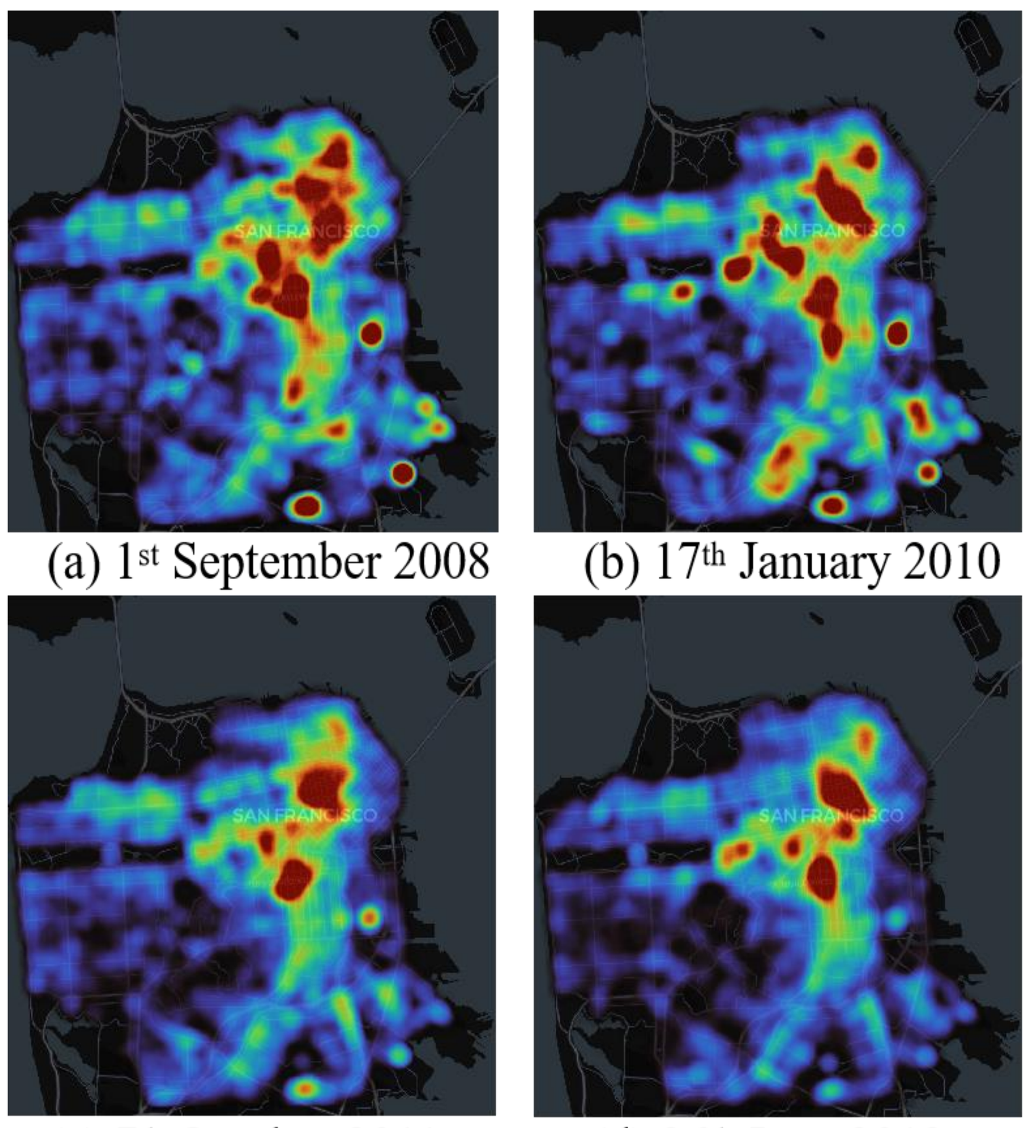
Dingming Wu
Shenzhen University
dingming@szu.edu.cn

Wei Tu
Ruisheng Wang
Shenzhen University
{tuwei,ruiswang}@szu.edu.cn

Joshua Zhexue Huang
Shenzhen University
zx.huang@szu.edu.cn

Overview of Spatiotemporal Kernel Density Visualization (STKDV)

(a) San Francisco 311 calls



(a) 1st September 2008 (b) 17th January 2010 (c) 7th October 2011 (d) 26th June 2013

Detect time-dependent hotspots using the location data of San Francisco 311 calls.

Color each pixel-timestamp $(\mathbf{q}, t_i)$ pair based on the spatiotemporal kernel density function $\mathcal{F}_{\hat{P}}(\mathbf{q}, t_i)$, where

$$\mathcal{F}_{\hat{P}}(\mathbf{q}, t_i) = \frac{1}{n} \sum_{(\mathbf{p}, t_\mathbf{p}) \in \hat{P}} K_{\text{space}}^{(b_\sigma)}(\mathbf{q}, \mathbf{p}) \cdot K_{\text{time}}^{(b_\tau)}(t_i, t_\mathbf{p})$$

Commonly used spatial kernel functions and temporal kernel functions.

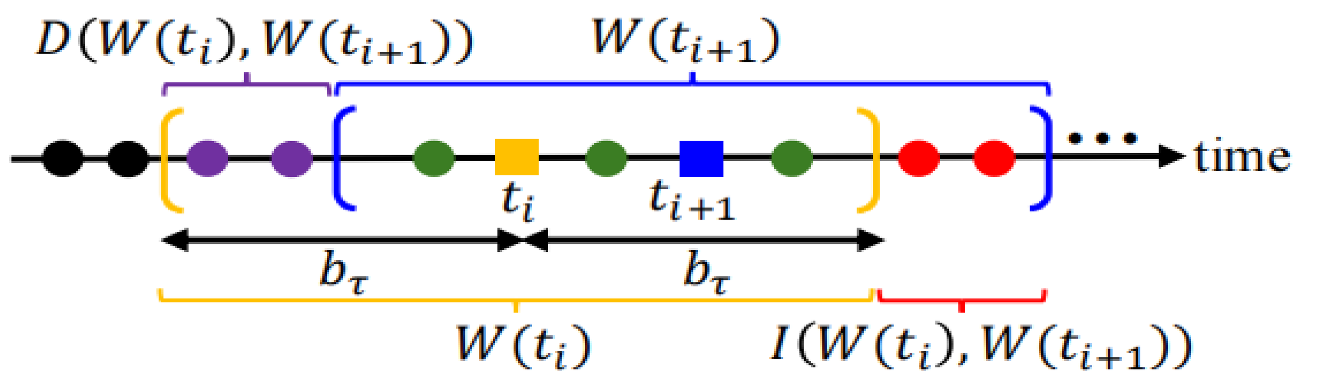
Kernel	$K_{\text{space}}(\mathbf{q}, \mathbf{p})$	$K_{\text{time}}(t_i, t_\mathbf{p})$
Triangular	$\begin{cases} 1 - \frac{1}{b_\sigma}\lVert\mathbf{q}-\mathbf{p}\rVert_2 & \text{if } \lVert\mathbf{q}-\mathbf{p}\rVert_2 \le b_\sigma \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 - \frac{1}{b_\tau}\lvert t_i - t_\mathbf{p}\rvert & \text{if } \lvert t_i - t_\mathbf{p}\rvert \le b_\tau \\ 0 & \text{otherwise} \end{cases}$
Epanechnikov	$\begin{cases} \frac{3}{4}\cdot\left(1 - \frac{1}{b_\sigma^2}\lVert\mathbf{q}-\mathbf{p}\rVert_2^2\right) & \text{if } \lVert\mathbf{q}-\mathbf{p}\rVert_2 \le b_\sigma \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \frac{3}{4}\cdot\left(1 - \frac{1}{b_\tau^2}(t_i - t_\mathbf{p})^2\right) & \text{if } \lvert t_i - t_\mathbf{p}\rvert \le b_\tau \\ 0 & \text{otherwise} \end{cases}$
Quartic	$\begin{cases} \frac{15}{16}\cdot\left(1 - \frac{1}{b_\sigma^2}\lVert\mathbf{q}-\mathbf{p}\rVert_2^2\right)^2 & \text{if } \lVert\mathbf{q}-\mathbf{p}\rVert_2 \le b_\sigma \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \frac{15}{16}\cdot\left(1 - \frac{1}{b_\tau^2}(t_i - t_\mathbf{p})^2\right)^2 & \text{if } \lvert t_i - t_\mathbf{p}\rvert \le b_\tau \\ 0 & \text{otherwise} \end{cases}$

STKDV is computationally expensive, which takes $O(XYTn)$ time. Example:

- The resolution size ($X \times Y$): 1280×960
- The number of timestamps (T): 32
- The total number of data points (n): 1.83 million
- The total cost is: **71.96 trillion operations** ☹

Overview of Existing Solution : SWS and PREFIX

SWS:



Only those data points $(\mathbf{p}, t_\mathbf{p})$ in $W(t_i)$ can contribute to $\mathcal{F}_{\hat{P}}(\mathbf{q}, t_i)$.

Efficiently update from $S_{W(t_i)}^{(u)}(\mathbf{q})$ to $S_{W(t_{i+1})}^{(u)}(\mathbf{q})$.

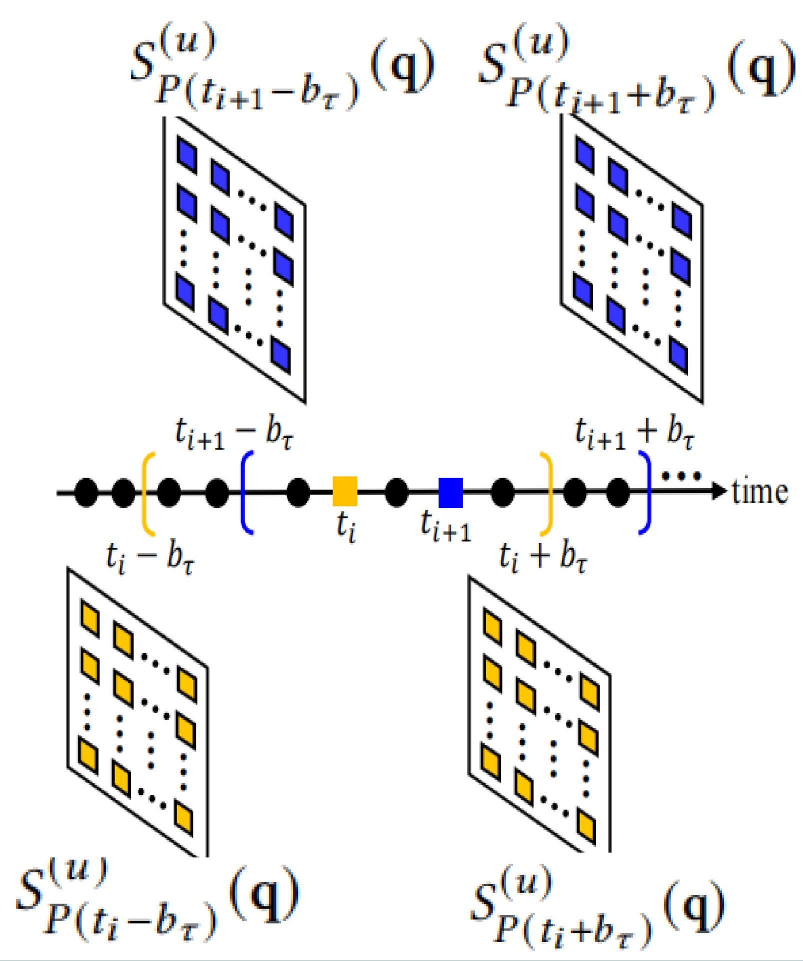
$$\mathcal{F}_{\hat{P}}(\mathbf{q}, t_i) = \sum_{(\mathbf{p}, t_\mathbf{p}) \in \hat{P}} w \cdot K_{\text{space}}^{(b_\sigma)}(\mathbf{q}, \mathbf{p}) \cdot \left(1 - \frac{1}{b_\tau^2} dist(t_i, t_\mathbf{p})^2\right)$$

$$= w\left(1 - \frac{t_i^2}{b_\tau^2}\right) S_{W(t_i)}^{(0)}(\mathbf{q}) + \frac{2wt_i}{b_\tau^2} S_{W(t_i)}^{(1)}(\mathbf{q}) - \frac{w}{b_\tau^2} S_{W(t_i)}^{(2)}(\mathbf{q})$$

where $S_{W(t_i)}^{(u)}(\mathbf{q}) = \sum_{(\mathbf{p}, t_\mathbf{p}) \in W(t_i)} t_\mathbf{p}^u \cdot K_{\text{space}}^{(b_\sigma)}(\mathbf{q}, \mathbf{p})$

The time complexity of SWS is $O(XY(T+n))$.

PREFIX:



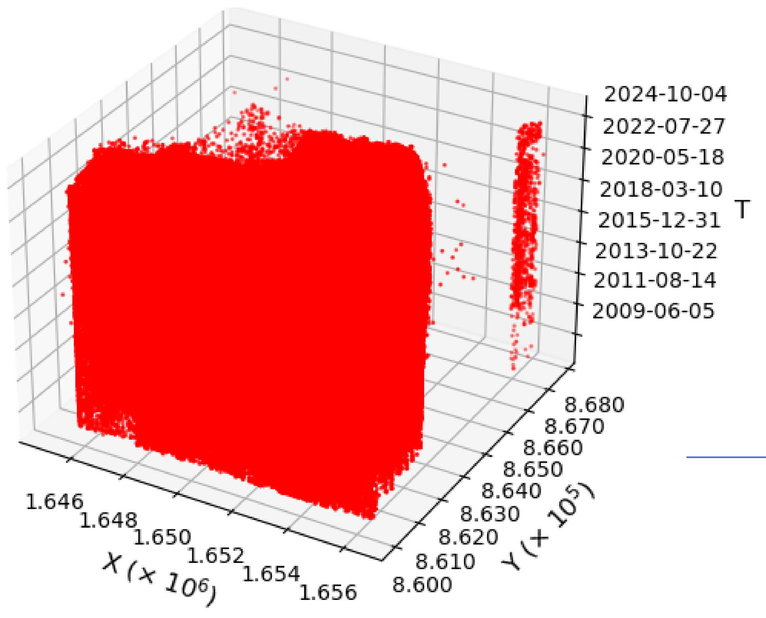
First maintain the prefix-matrix with respect to all X×Y pixels for each end point in the time axis.

Once they have all prefix-matrices, they then evaluate each $S_{W(t_i)}^{(u)}(\mathbf{q})$ based on the following equation and compute $\mathcal{F}_{\hat{P}}(\mathbf{q}, t_i)$ for all pixels $\mathbf{q}$.

$$S_{W(t_i)}^{(u)}(\mathbf{q}) = S_{P(t_i+b_\tau)}^{(u)}(\mathbf{q}) - S_{P(t_i-b_\tau)}^{(u)}(\mathbf{q})$$

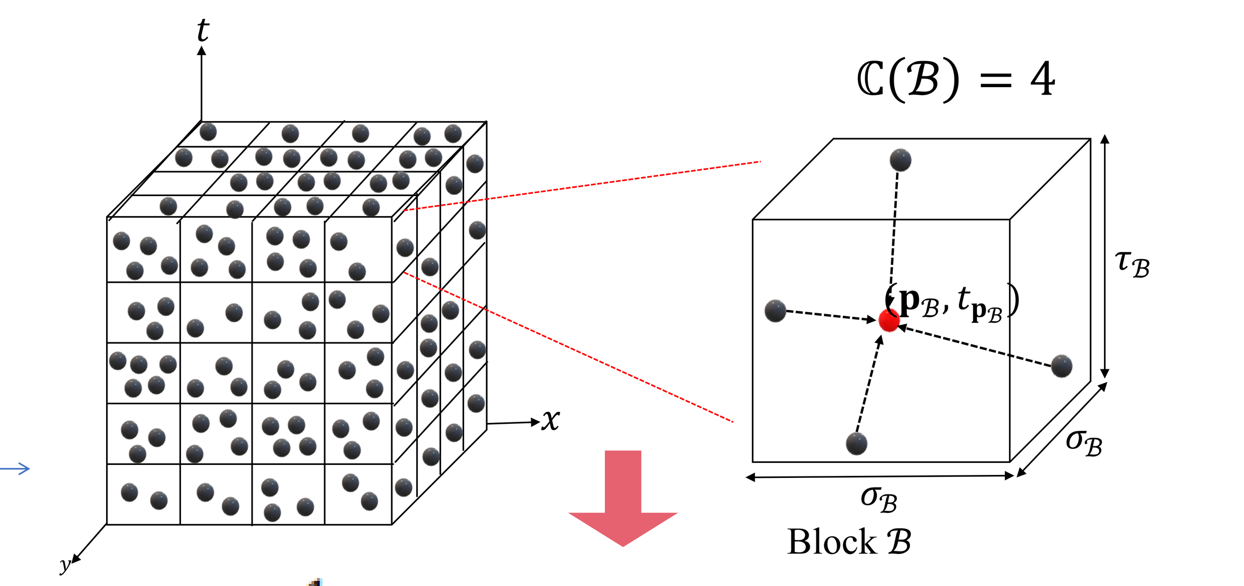
The time complexity of PREFIX is $O(XYT + Yn)$.

Core ideas of Our Solution (COMP)



The distribution of San Francisco 311-call data points.

Real data often contain spatiotemporally close points.



$$A_\mathcal{S}(\mathbf{q}, t_i) = \frac{1}{n} \sum_{\mathcal{B} \in \mathcal{S}} \mathbb{C}(\mathcal{B}) \cdot K_{\text{space}}(\mathbf{q}, \mathbf{p}_\mathcal{B}) \cdot K_{\text{time}}(t_i, t_{\mathbf{p}_\mathcal{B}})$$

COMP reduces data by merging nearby spatiotemporal points into block centers with weights.

How to determine the block size of COMP to ensure that the following absolute error guarantee holds?

$$|A_\mathcal{S}(\mathbf{q}, t_i) - \mathcal{F}_P(\mathbf{q}, t_i)| \le \epsilon$$

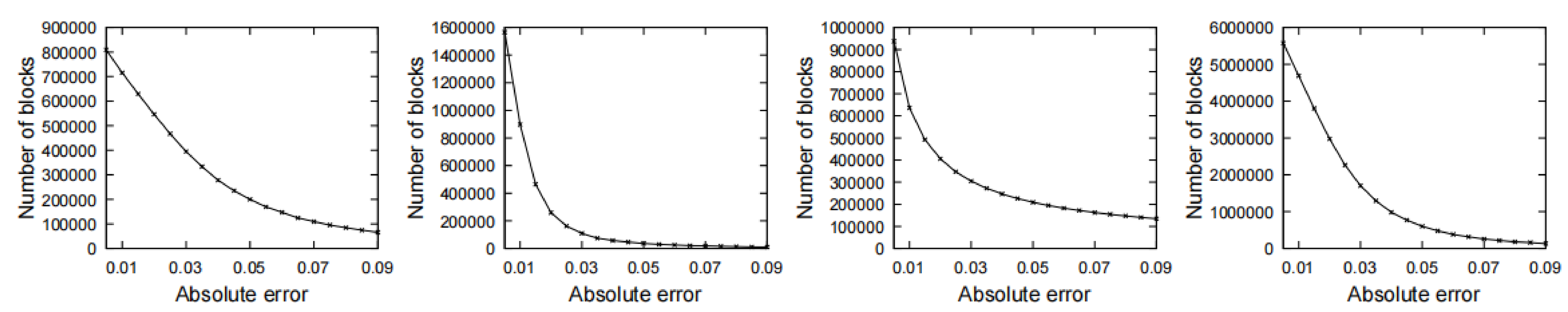
Theorem 1. *Given a location dataset P and an absolute error ϵ, we can achieve the absolute error guarantee ϵ if $\omega_\mathcal{B}$ and $\lambda_\mathcal{B}$ of each block $\mathcal{B}$ in $\mathcal{S}$ have the following settings.*

(1) $\omega_\mathcal{B} = \frac{\sqrt{2}\cdot\epsilon\cdot b_\sigma}{2}$ *and* $\lambda_\mathcal{B} = \epsilon \cdot b_\tau$ *using the triangular spatial and temporal kernels.*

(2) $\omega_\mathcal{B} = \frac{4\sqrt{2}\cdot\epsilon\cdot b_\sigma}{9}$ *and* $\lambda_\mathcal{B} = \frac{8\cdot\epsilon\cdot b_\tau}{9}$ *using the Epanechnikov spatial and temporal kernels.*

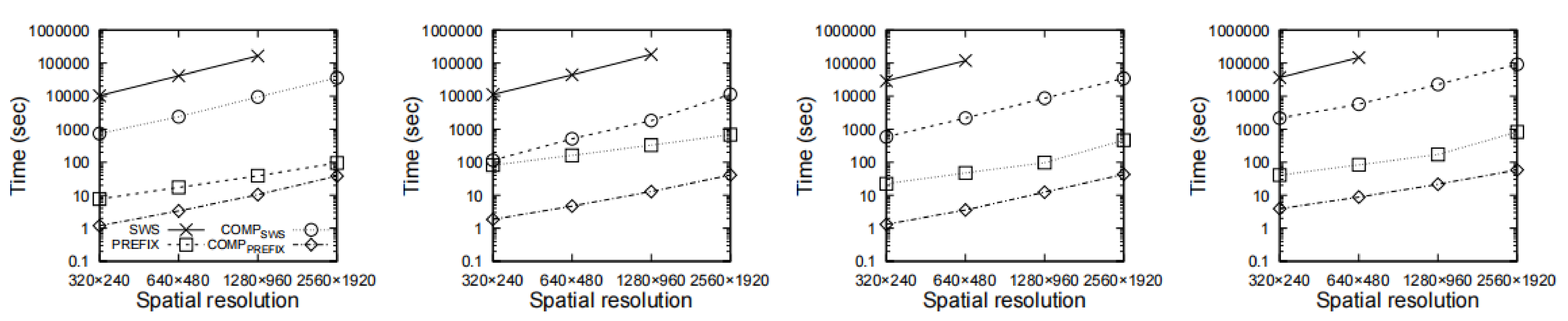
(3) $\omega_\mathcal{B} = \frac{16\sqrt{6}\cdot\epsilon\cdot b_\sigma}{75}$ *and* $\lambda_\mathcal{B} = \frac{32\sqrt{3}\cdot\epsilon\cdot b_\tau}{75}$ *using the quartic spatial and temporal kernels.*

Experimental Results



(a) Montgomery (b) New York (c) Chicago (d) San Francisco

The number of blocks, varying the absolute error ϵ.



(a) Montgomery (b) New York (c) Chicago (d) San Francisco

Response time for computing STKDV, varying the spatial resolution.