Large-scale Spatiotemporal Kernel Density Visualization

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

Pak Lon Ip
University of Macau
paklonip@um.edu.mo

Bojian Zhu
Hong Kong Baptist University
csbjzhu@comp.hkbu.edu.hk

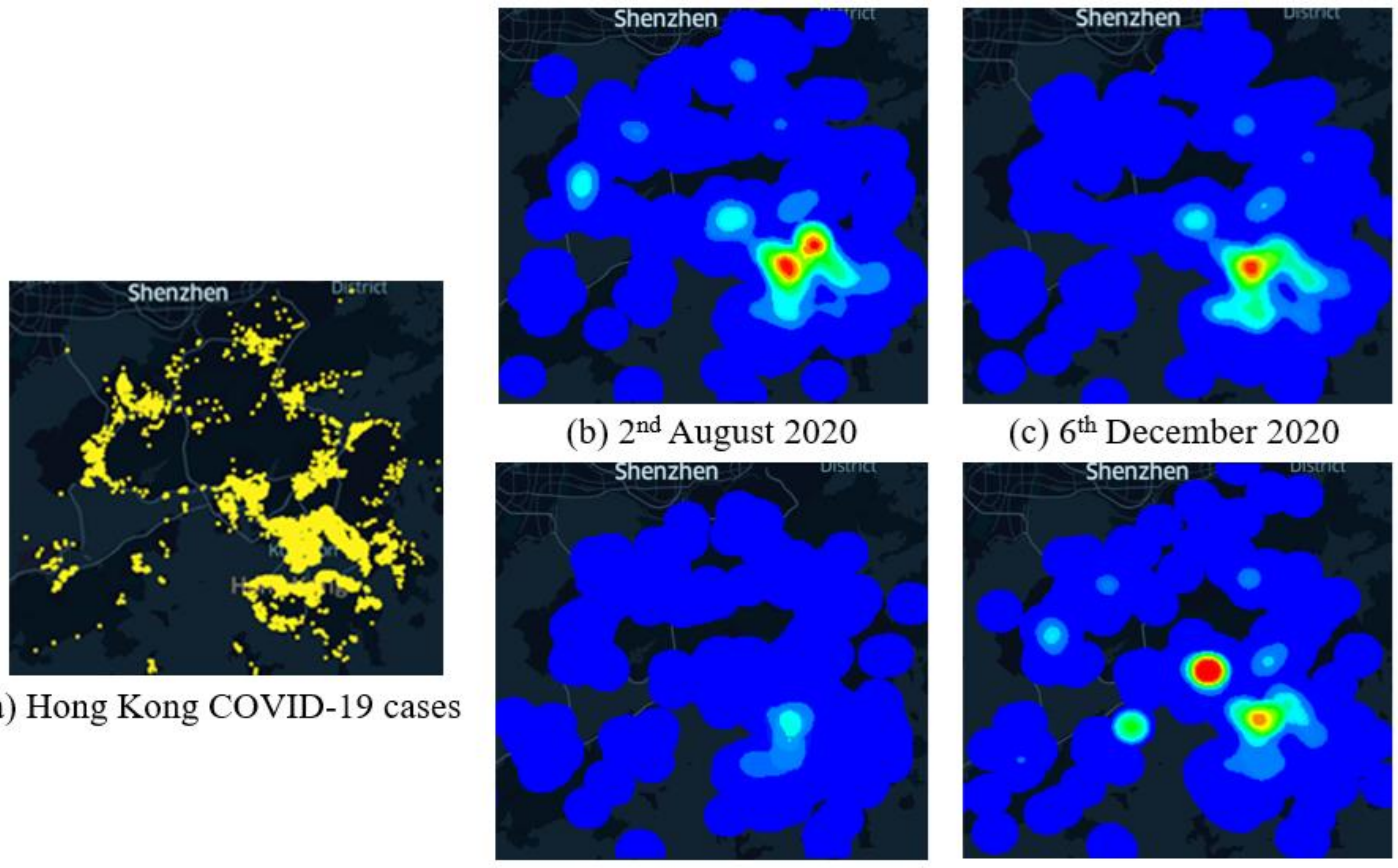
Leong Hou U
University of Macau
ryanlhu@um.edu.mo

Dingming Wu
Shenzhen University
dingming@szu.edu.cn

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

Christian S. Jensen
Aalborg University
csj@cs.aau.dk

What is Spatiotemporal Kernel Density Visualization (STKDV)?



(a) Hong Kong COVID-19 cases

(b) 2nd August 2020

(c) 6th December 2020

(d) 28th February 2021

(e) 28th January 2022

Detect the disease outbreak based on the Hong Kong COVID-19 location dataset.

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

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

Commonly used spatial kernel functions and temporal kernel functions.

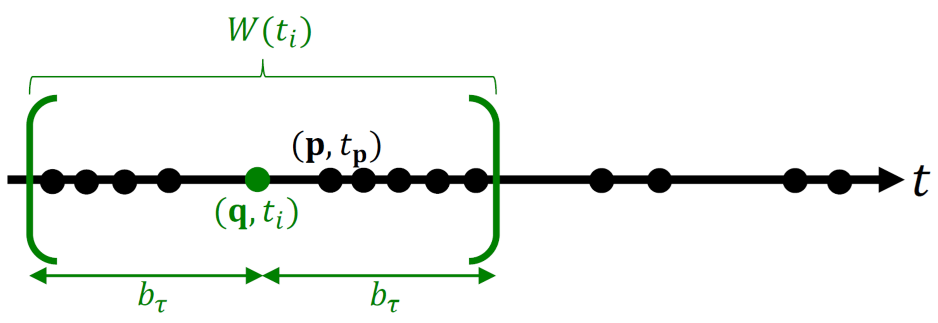
Kernel	$K_{\text{space}}^{(b_\sigma)}(\mathbf{q}, \mathbf{p})$	$K_{\text{time}}^{(b_\tau)}(t_i, t_\mathbf{p})$
Uniform	$\begin{cases} \frac{1}{b_\sigma} & \text{if } dist(\mathbf{q},\mathbf{p}) \le b_\sigma \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \frac{1}{b_\tau} & \text{if } dist(t_i,t_\mathbf{p}) \le b_\tau \\ 0 & \text{otherwise} \end{cases}$
Epanechnikov	$\begin{cases} 1-\frac{1}{b_\sigma^2}dist(\mathbf{q},\mathbf{p})^2 & \text{if } dist(\mathbf{q},\mathbf{p}) \le b_\sigma \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1-\frac{1}{b_\tau^2}dist(t_i,t_\mathbf{p})^2 & \text{if } dist(t_i,t_\mathbf{p}) \le b_\tau \\ 0 & \text{otherwise} \end{cases}$
Quartic	$\begin{cases} \left(1-\frac{1}{b_\sigma^2}dist(\mathbf{q},\mathbf{p})^2\right)^2 & \text{if } dist(\mathbf{q},\mathbf{p}) \le b_\sigma \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \left(1-\frac{1}{b_\tau^2}dist(t_i,t_\mathbf{p})^2\right)^2 & \text{if } dist(t_i,t_\mathbf{p}) \le b_\tau \\ 0 & \text{otherwise} \end{cases}$

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

Example:

- The resolution size ($X \times Y$): 128 × 128
- The number of timestamps (T): 128
- The total number of data points (n): 1,000,000
- The total cost is: **2.09 trillion operations** ☹

Overview of Existing Solution (SWS)

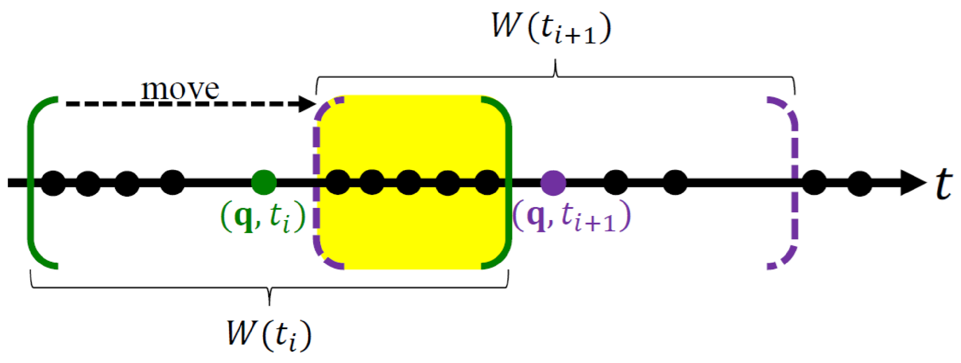


Only those data points $(\mathbf{p}, t_\mathbf{p})$ in $W(t_i)$ can contribute to $\mathcal{F}_{\hat{P}}(\mathbf{q}, t_i)$.

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

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

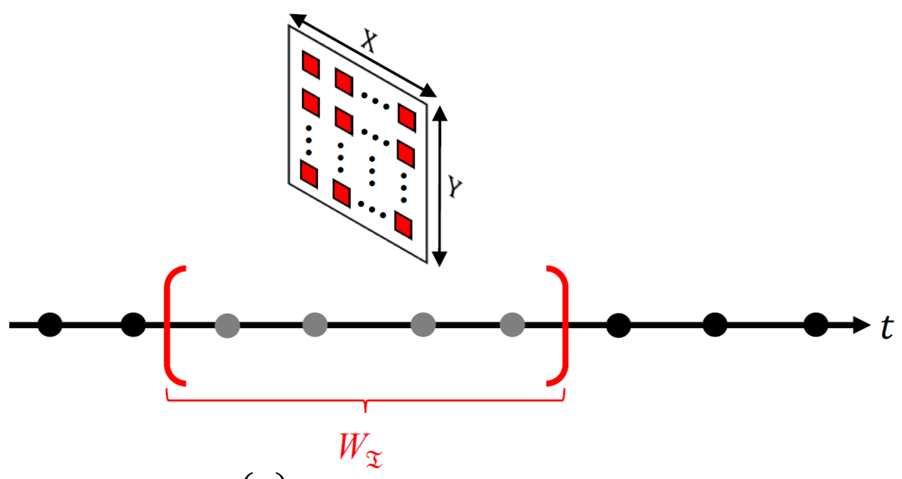
where $S_{W(t_i)}^{(u)}(\mathbf{q}) = \sum_{(\mathbf{p}, t_\mathbf{p}) \in W(t_i)} t_\mathbf{p}^u \cdot K_{\text{space}}^{(b_\sigma)}(\mathbf{q}, \mathbf{p})$



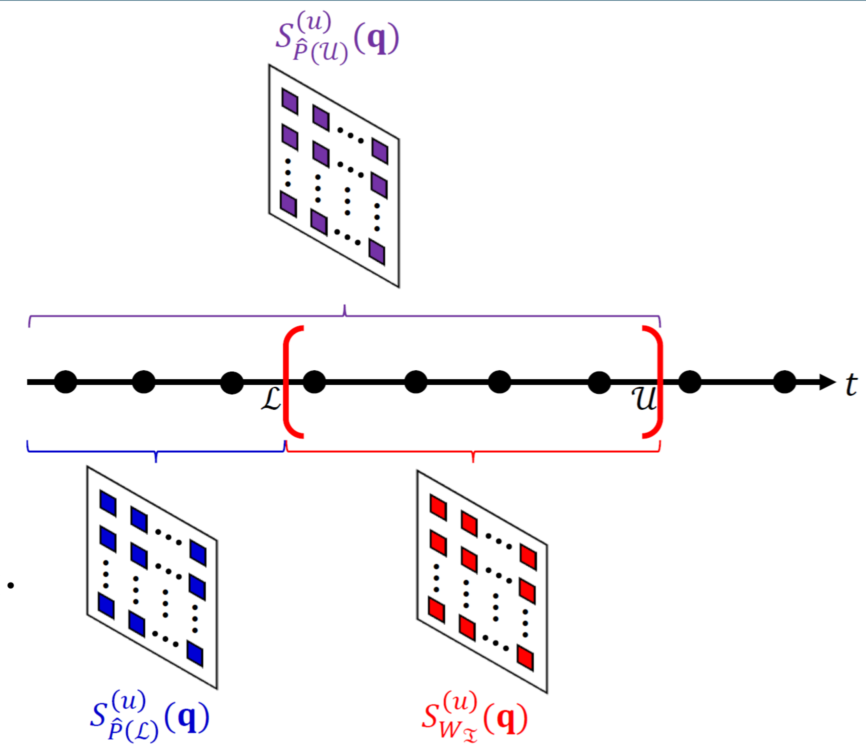
Efficiently update from $S_{W(t_i)}^{(u)}(\mathbf{q})$ to $S_{W(t_{i+1})}^{(u)}(\mathbf{q})$.

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

Core ideas of Our Solution (PREFIX)

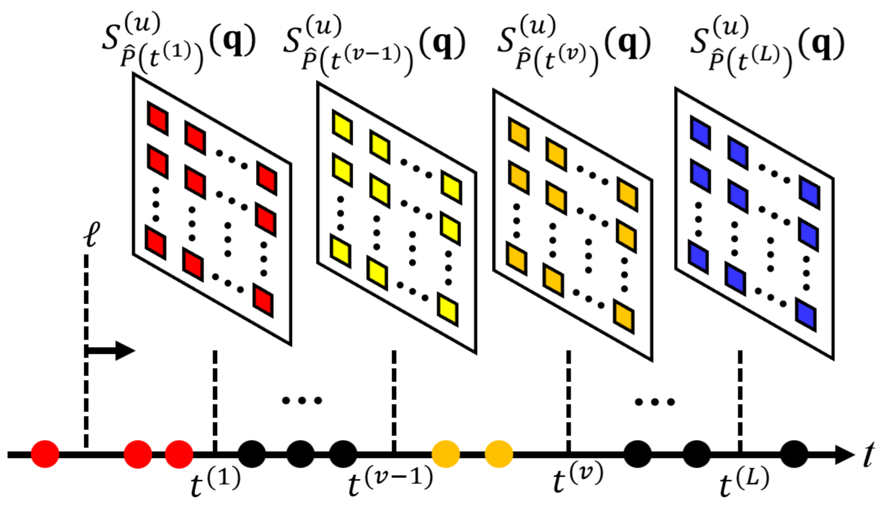


Core idea 1: $S_{W_\mathfrak{T}}^{(u)}(\mathbf{q})$ can be computed in $O(Y(X+|W_\mathfrak{T}|))$ time for all pixels $\mathbf{q}$ with any $W_\mathfrak{T}$.



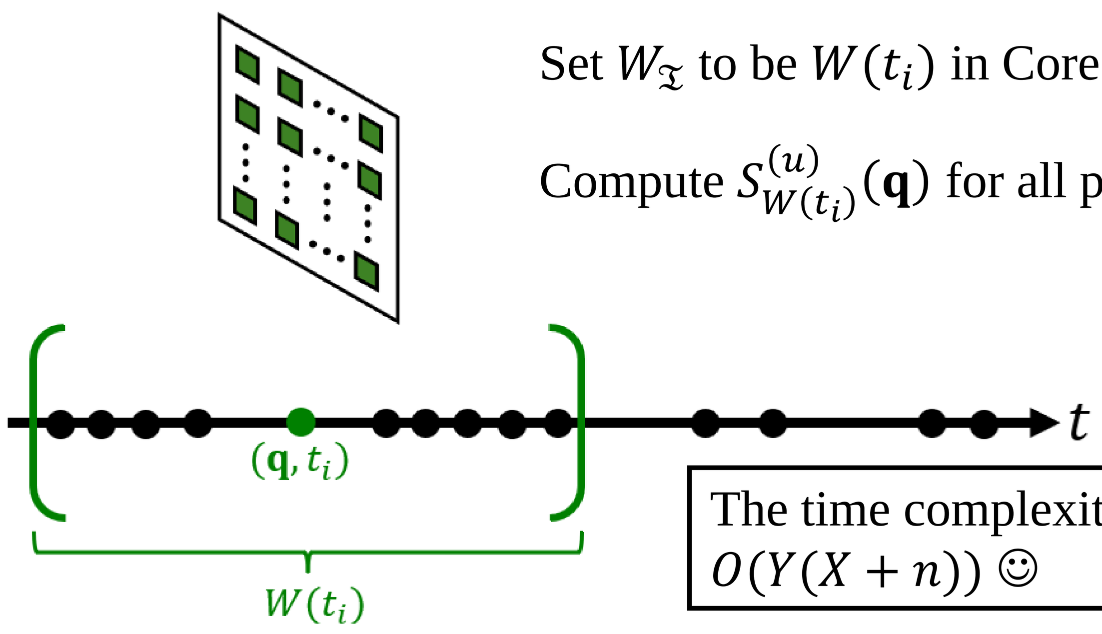
Core idea 2: Takes $O(XY)$ time to obtain $S_{W_\mathfrak{T}}^{(u)}(\mathbf{q})$, where

$$S_{W_\mathfrak{T}}^{(u)}(\mathbf{q}) = S_{\hat{P}(U)}^{(u)}(\mathbf{q}) - S_{\hat{P}(L)}^{(u)}(\mathbf{q})$$



Core idea 3: Takes $O(XYL + Yn)$ time for maintaining statistical matrices for L timestamps.

PREFIX (On-the-fly timestamps)

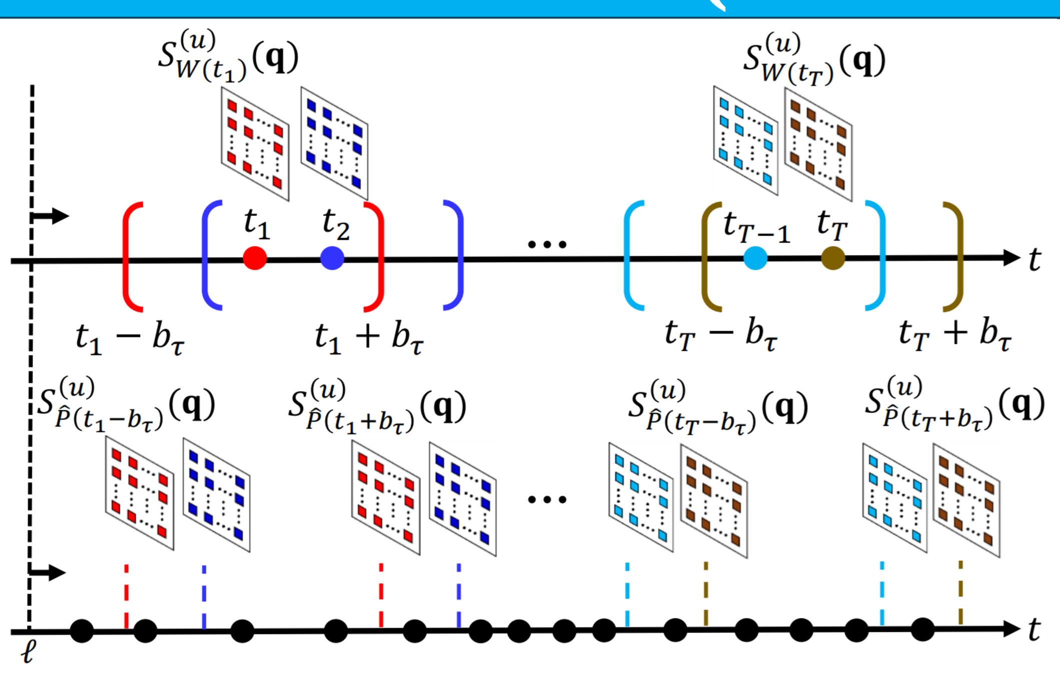


Set $W_\mathfrak{T}$ to be $W(t_i)$ in Core idea 1.

Compute $S_{W(t_i)}^{(u)}(\mathbf{q})$ for all pixels $\mathbf{q}$ in $O(Y(X+n))$ time ☺

The time complexity of PREFIX is $O(Y(X+n))$ ☺

PREFIX (Known timestamps)



Computing the statistical matrices for all $2T$ prefix sets is $O(XYT + Yn)$ time (by setting $L = 2T$ in Core idea 3).

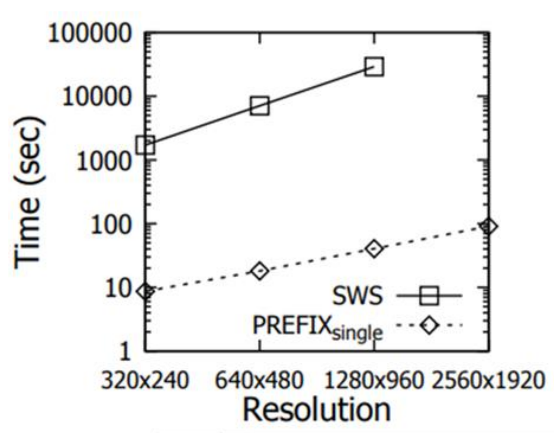
Since there are T windows, we can compute $S_{W(t_i)}^{(u)}(\mathbf{q})$ for all pixels $\mathbf{q}$ in $O(XYT)$ time. (based on Core idea 2).

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

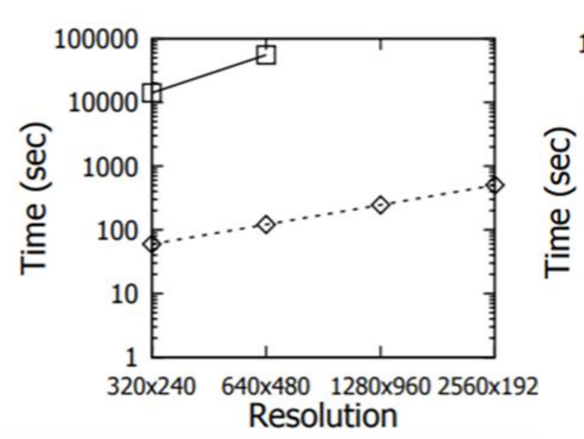
Theoretical Results

Problem	Method	Time complexity	Space complexity
STKDV (on-the-fly timestamp)	SWS	$O(XYn)$	$O(XY+n)$
	PREFIX$_{\text{single}}$ (Section IV-B)	$O(Y(X+n))$ (Theorem 1)	$O(XY+n)$ (Theorem 4)
STKDV (T known timestamps)	SWS	$O(XY(T+n))$	$O(XYT+n)$
	PREFIX$_{\text{multiple}}$ (Section IV-C)	$O(XYT+Yn)$ (Theorem 2)	$O(XYT+n)$ (Theorem 5)
Bandwidth tuning	SWS	$O(MNXY(T+n))$	$O(MNXYT+n)$
	PREFIX$_{\text{tuning}}$ (Section IV-D)	$O(M(NXYTN+Yn))$ (Theorem 3)	$O(MNXYT+n)$ (Theorem 6)

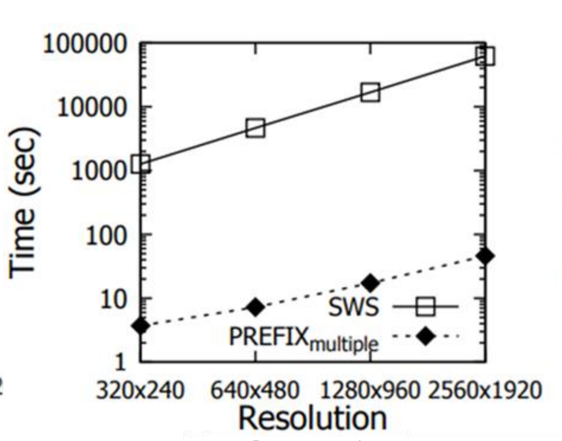
Experimental Results



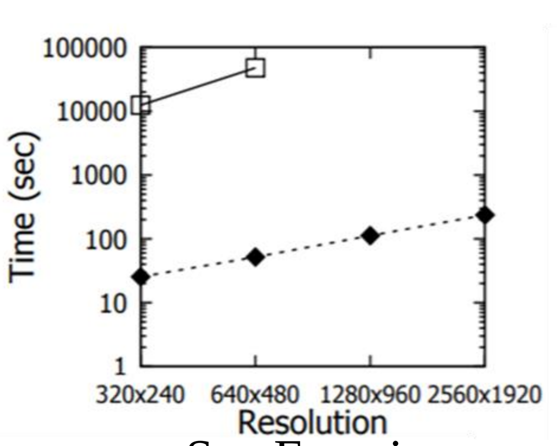
Ontario



San Francisco



Ontario



San Francisco