Mobile Sensing of Spatial Fields
Challenges and Opportunities

Jayakrishnan Unnikrishnan

Audiovisual Communications Laboratory
Ecole Polytechnique Fédérale de Lausanne

Collaborators: Martin Vetterli, José Luis Romero, Karlheinz Gröchenig, Farid M. Naini

Qualcomm
Feb. 14, 2014
Life in a networked world

Communication and computing devices are everywhere
Life in a networked world

- Communication and computing devices are everywhere

- Sensors are everywhere
Life in a networked world

- Communication and computing devices are everywhere
- Sensors are everywhere
- And getting smarter

Berkeley motes: “Smart Dust”
Environmental monitoring like never before

Traffic density, New York

Surface temperature, EPFL, [Nadeau et al., '09]
Environmental monitoring like never before

Traffic density, New York

Emerging paradigm:
- Sensing spatial fields with mobile sensors

Surface temperature, EPFL, [Nadeau et al., '09]
Environmental monitoring like never before

Traffic density, New York

Emerging paradigm:
- Sensing spatial fields with mobile sensors
- Offers unique advantages over static sensing

Surface temperature, EPFL, [Nadeau et al., '09]
Advantages of mobile sensing

- Mobile sensor can sample at arbitrarily high resolutions along path
Advantages of mobile sensing

- Mobile sensor can sample at arbitrarily high resolutions along path

- Mobile sensors can implement spatial anti-aliasing via time-domain filtering (See later)
Advantages of mobile sensing

- Mobile sensor can sample at arbitrarily high resolutions along path

- Mobile sensors can implement spatial anti-aliasing via time-domain filtering (See later)

- Single mobile sensor can cover a wide area of interest
  - Potentially cost-effective and more practical - e.g. pollution monitoring in a city
Pollution monitoring in Lausanne

[Image of a bus and a map of the area]
Mobile radiation sampling in Japan
Mobile radiation sampling in Japan

Mobile sensing on roads
Citizen sensing

Images from http://www.urban-atmospheres.net/CitizenScience/
Sensing for robotics

Image from http://users.ece.gatech.edu/magnus/
Scanning trajectories for MRI

Image from http://lapmal.epfl.ch
Outline of the talk

1. Designing Sensor Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories

2. Spatial Anti-aliasing via Mobile Sensing

3. Privacy of Mobility traces

4. Recap
Given 1-D bandlimited signal

$$X(\omega)$$
Classical sampling [WNKWRGSS, 1915 - 1949]

- Given 1-D bandlimited signal

\[ X(\omega) \]

Perfect recovery via uniform sampling provided \( \Delta \leq \frac{\pi}{W} \)

- Jayakrishnan Unnikrishnan (EPFL)
Classical sampling in higher dimensions

- Given: spatially bandlimited field $f : \mathbb{R}^d \mapsto \mathbb{R}$

$$\mathcal{F}(\omega) := \int f(r) e^{-j\langle \omega, r \rangle} dr = 0 \text{ for } \omega \notin \Omega$$

Spectrum $|\mathcal{F}(\omega_x, \omega_y)|$

Support of spectrum $\Omega$
Sampling on a lattice

Sampling lattice
Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice

Sampling lattice

Original spectrum
Sampling on a lattice

No aliasing in sampled spectrum for $X = Y \leq \frac{\pi}{R}$
Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice

Sampling lattice

Perfect recovery of original spectrum
Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice

\[
\frac{2\pi}{X} \quad \text{and} \quad \frac{2\pi}{Y}
\]

Aliased sampled spectrum

- Lattice should be fine enough \( \equiv \) Nyquist criterion in \( \mathbb{R}^d \)

Perfect recovery impossible
Classical sampling

Static sensors record field values at points
Classical sampling vs Mobile sampling

**Classical sampling**
Static sensors record field values at **points**

![Diagram showing static sensor points](image)

**Mobile sampling**
Mobile sensors record field values on **trajectories**

![Diagram showing mobile sensor trajectories](image)
Classical sampling
Static sensors record field values at points (e.g. a lattice)

Mobile sampling
Mobile sensors record field values on trajectories
Classical sampling vs Mobile sampling

**Classical sampling**
Static sensors record field values at points (e.g. a lattice)

**Mobile sampling**
Mobile sensors record field values on trajectories

Focus on time-invariant fields
Sampling trajectories

- A trajectory set $p$ is a countable collection of paths $p_i$:

$$p = \{p_i : i \in \mathbb{I}\}$$
Sampling trajectories

- **A trajectory set** $p$ is a countable collection of paths $p_i$:

$$p = \{p_i : i \in \mathbb{I}\}$$

- **Path density** of $p$: Total path-length per unit spatial volume

$$\ell(p) := \lim_{a \to \infty} \frac{\mathcal{D}^p(a)}{\text{Vol}_d(a)}$$
A trajectory set $p$ is a countable collection of paths $p_i$:

$$p = \{ p_i : i \in I \}$$

Path density of $p$: Total path-length per unit spatial volume

$$\ell(p) := \lim_{a \to \infty} \frac{D^p(a)}{\text{Vol}_d(a)}$$
Sampling trajectories

- A trajectory set $p$ is a countable collection of paths $p_i$:
  \[ p = \{ p_i : i \in \mathbb{I} \} \]

- Path density of $p$: Total path-length per unit spatial volume
  \[ \ell(p) := \lim_{a \to \infty} \frac{D^p(a)}{\text{Vol}_d(a)} \]
$\mathcal{N}_\Omega$: collection of **Nyquist trajectory sets** $p$ for fields $f$ bandlimited to $\Omega$
Sampling trajectories for bandlimited fields

\[ \mathcal{N}_\Omega: \text{collection of Nyquist trajectory sets } p \text{ for fields } f \text{ bandlimited to } \Omega \]

Requirements:

- Field \( f(\cdot) \) can be reconstructed stably from values on trajectories
Sampling trajectories for bandlimited fields

- $N_{\Omega}$: collection of Nyquist trajectory sets $p$ for fields $f$ bandlimited to $\Omega$

- Requirements:
  - Field $f(.)$ can be reconstructed stably from values on trajectories
  - Regularity conditions
Main contributions

- Examples of trajectory sets in $\mathcal{N}_\Omega$
  - Some new results
Main contributions

- Examples of trajectory sets in $\mathcal{N}_\Omega$
  - Some new results

- Designing trajectory-sets that are minimal in path density
  - New formulation
  - Optimality results from restricted classes
  - Fundamental limits on path density similar to Nyquist rate
Scanning trajectories for MRI

Trajectories in Fourier space indicate how to vary magnetic field in time
Related work

- Scanning trajectories for MRI
  - Trajectories in Fourier space indicate how to vary magnetic field in time

- Reconstructing bandlimited fields from readings on circles [Tewfik, Levy, Willsky '88], [Myridis, Chamzas '98] and spirals [Benedetto, Wu '00]
Related work

- Scanning trajectories for MRI
  - Trajectories in Fourier space indicate how to vary magnetic field in time

- Reconstructing bandlimited fields from readings on circles [Tewfik, Levy, Willsky ’88], [Myridis, Chamzas ’98] and spirals [Benedetto, Wu ’00]

- Adaptive path-planning in mobile sensor networks etc.
A Uniform trajectory set in $\mathbb{R}^2$
Proof - Uniform set

Original field spectrum
Proof - Uniform set

Sampled field spectrum
Proof - Uniform set

Sampled field spectrum as $\epsilon \to 0$
Perfect recovery provided $D \leq \frac{2\pi}{\Omega^-_y - \Omega^+_y}$
i.e. $p \in \mathcal{N}_\Omega$
Union of Uniform sets in $\mathbb{R}^2$
Can identify exact conditions on $D_i, v_i$ to ensure $p \in \mathcal{N}_\Omega$
Example: Orthogonal trajectories and Isotropic field

Orthogonal sets of trajectories

Sampled spectra from the two sets
Example: Orthogonal trajectories and Isotropic field

Orthogonal sets of trajectories

Critical sampling
Implications for isotropic fields

\[ \frac{\sqrt{2\pi}}{R} \]

Parallel lines have lower path density than Manhattan sampling grid!
Implications for isotropic fields

Parallel lines have lower path density than Manhattan sampling grid!

- In fact, lower than any union of uniform sets
Other trajectory configurations

Equispaced concentric circles
Other trajectory configurations

Interleaved spirals
Non-uniform parallel trajectory sets for $\mathbb{R}^d$

- Can identify reconstruction conditions and optimal configurations
Classical sampling vs Mobile sampling

**Classical sampling**

- Minimum sampling density
  - $\propto \text{Vol}(\Omega)$ [Landau '67]

**Sampling on parallel lines**

- Minimum path density
  - $\propto \text{Min. section}(\Omega)$
Outline of the talk

1. Designing Sensor Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories

2. Spatial Anti-aliasing via Mobile Sensing

3. Privacy of Mobility traces

4. Recap
Anti-aliasing in one dimension

Bandlimited signal spectrum
Anti-aliasing in one dimension

\[ X_s(\omega) \]

Sampled spectrum
Anti-aliasing in one dimension

Imperfectly bandlimited signal

\[ X(\omega) \]
Anti-aliasing in one dimension

\[ X_s(\omega) \]

Sampled spectrum

\[ \frac{2\pi}{\Delta} \]
Anti-aliasing in one dimension

\[ \hat{X}(\omega) \]

Aliasing in reconstructed spectrum
Anti-aliasing in one dimension

\[ X(\omega) \quad H_{aa}(\omega) \]

\[ x(t) \xrightarrow{h_{aa}(.)} nT \xrightarrow{\text{anti-aliasing}} x[n] \xrightarrow{h_r(.)} \hat{x}(t) \]

Anti-aliasing filtering
Anti-aliasing in one dimension

\[ \hat{X}(\omega) \]

\[ x(t) \quad h_{aa}(\cdot) \quad nT \quad x[n] \quad h_r(\cdot) \quad \hat{x}(t) \]

No aliasing in reconstruction
Spatial anti-aliasing

- Impossible with static sensors
  - Cannot integrate over continuous space
Spatial anti-aliasing

- Impossible with static sensors
  - Cannot integrate over continuous space
- Mobile sensor sees field as function of time
  - \( s(t) = f(r(t)) \)
  - If constant velocity then \( s(t) \) is bandlimited
Spatial anti-aliasing

- Impossible with static sensors
  - Cannot integrate over continuous space

- Mobile sensor sees field as function of time
  - \( s(t) = f(r(t)) \)
  - If constant velocity then \( s(t) \) is bandlimited

- In presence of noise time-domain filtering can suppress spatial aliasing along direction of motion
Spatial anti-aliasing filter
Spatial anti-aliasing filter
Spatial anti-aliasing filter
Spatial anti-aliasing filter

\[ h(x, y) \]

Jayakrishnan Unnikrishnan (EPFL)
Spatial anti-aliasing filter
Spatial anti-aliasing filter

\[ h(x,y) \]

\[ H_{aa}(\omega) \]
Spatial anti-aliasing illustrated

Static sampling

Original spectrum

Sampled spectrum
Spatial anti-aliasing illustrated

Static sampling

Original spectrum with noise

Sampled spectrum aliased in all directions
Spatial anti-aliasing illustrated

Static sampling

Original spectrum with noise

Reconstructed spectrum aliased in all directions
Spatial anti-aliasing illustrated

Mobile sampling

Spatially filtered field with noise

Sampled spectrum aliased only in two directions
Spatial anti-aliasing illustrated

Mobile sampling

Spatially filtered field with noise

Reconstructed spectrum aliased in two directions
A practical pre-filter

Sensor measures average in chamber
A practical pre-filter

\[ v \propto h(nT - t) \]
A practical pre-filter

$$s(nT) = \int f(vt)h(nT - t)dt$$
A practical pre-filter

\[ s(nT) = \int f(vt)h(nT - t)dt \]

- Clear chamber before next sample is taken
- For long impulse responses, use multiple such chambers
An example: Campus temperature

T (°C)

8.5 9 9.5 10 10.5 11

meters

0 50 100 150 200 250 300
## Table: Reconstruction errors

<table>
<thead>
<tr>
<th>Data type</th>
<th>Static sensing</th>
<th>Mobile sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>0.53%</td>
<td>0.45%</td>
</tr>
<tr>
<td>BL in noise</td>
<td>9.9%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>
## Performance comparison

### Table: Reconstruction errors

<table>
<thead>
<tr>
<th>Data type</th>
<th>Static sensing</th>
<th>Mobile sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>0.53%</td>
<td>0.45%</td>
</tr>
<tr>
<td>BL in noise</td>
<td>9.9%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

- Improvement in SNR can be quantified analytically
## Performance comparison

### Table: Reconstruction errors

<table>
<thead>
<tr>
<th>Data type</th>
<th>Static sensing</th>
<th>Mobile sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>0.53%</td>
<td>0.45%</td>
</tr>
<tr>
<td>BL in noise</td>
<td>9.9%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

- **Improvement in SNR** can be quantified analytically
- **Conclusion**: Mobile sensing enables spatial anti-aliasing filtering
Outline of the talk

1. Designing Sensor Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories

2. Spatial Anti-aliasing via Mobile Sensing

3. Privacy of Mobility traces

4. Recap
Privacy concerns

- Personal data being collected at unprecedented levels
Unique in the crowd

- Anonymized user data is not anonymous given auxiliary information
Anonymized user data is not anonymous given auxiliary information

Users are uniquely identifiable from small set of observations

- 87% Americans uniquely identified given ZIP code, birthdate, and sex
  [Sweeney 2000]

- 95% of mobile users in a country are uniquely identified from four spatio-temporal points [Montjoye et al 2013]
De-anonymization
De-anonymization
De-anonymization
De-anonymization

- Uniqueness of trajectories → Easily de-anonymized
Sometimes applications require only statistics of data, e.g.,
- Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
- # visits to particular restaurants (for popularity surveys)
- # tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)
Sometimes applications require only statistics of data, e.g.,
- Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
- \# visits to particular restaurants (for popularity surveys)
- \# tweets/blog comments on a particular topic/containing a particular word (for targeted ads)

Question: How private are anonymized statistics?
Privacy of Statistics

- Sometimes applications require only statistics of data, e.g.,
  - Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
  - \(\#\) visits to particular restaurants (for popularity surveys)
  - \(\#\) tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)

- **Question:** How private are anonymized statistics?
  - What is the optimal de-anonymization strategy given independent auxiliary information?
Sometimes applications require only statistics of data, e.g.,
- Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
- # visits to particular restaurants (for popularity surveys)
- # tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)

**Question:** How private are anonymized statistics?

- What is the optimal de-anonymization strategy given independent auxiliary information?
- Focus on histograms
Sometimes applications require only statistics of data, e.g.,
- Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
- \# visits to particular restaurants (for popularity surveys)
- \# tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)

**Question:** How private are anonymized statistics?
- What is the optimal de-anonymization strategy given independent auxiliary information?
- Focus on histograms
- How unique are user location histograms?
De-anonymizing as Matching

Given: Anonymized Statistics of $K$ users
Given: Anonymized Statistics of $K$ users’ data and Auxiliary Information about \textit{independently} generated data.
De-anonymizing as Matching

- Given: Anonymized Statistics of $K$ users’ data and Auxiliary Information about independently generated data

- Task: Match Auxiliary Information to the correct Anonymized Statistics
Matching Problem

Day 1 histograms (anon.)  Day 2 traces

$p_1$  $y_1$
$p_2$  $y_2$
$p_{K-1}$  $y_{K-1}$
$p_K$  $y_K$
Matching Problem

E.g., \( p_i = \left( \frac{t_{i1}}{T}, \frac{t_{i2}}{T}, \ldots, \frac{t_{iN}}{T} \right) \)

where \( T \) is the total time units in a day, and \( N \) is the total number of locations
Matching Problem

Task: Identify the correct matching
Task: Identify the correct matching

What is the right strategy?
Task: Identify the correct matching

What is the right strategy?

Solution: View as a hypothesis test over $K!$ hypotheses, one for each possible matching
Related problem: only one observation in second set; studied by Gutman (’89)
Related problem: only one observation in second set; studied by Gutman ('89)

Repeating this test $K$ times is suboptimal as constraint not used
Solution via Minimum Weight Bipartite Matching

Here \( q_i \) is histogram of \( y_i \)

\[
w_{ij} = D(p_i \parallel \frac{1}{2}(p_i + q_j)) + D(q_j \parallel \frac{1}{2}(p_i + q_j))
\]
Solution via Minimum Weight Bipartite Matching

Min-Weight Matching $\Rightarrow$ easy to compute
Solution via Minimum Weight Bipartite Matching

Day 1 histogram \quad \text{Day 2 histograms}

\begin{align*}
\ p_1 & \quad q_1 \\
\ p_2 & \quad q_2 \\
\ p_i & \quad q_j \\
\ p_{K-1} & \quad q_{K-1} \\
\ p_K & \quad q_K \\
\end{align*}

\[ w_{ij} = D(p_i \| \frac{1}{2}(p_i+q_j)) + D(q_j \| \frac{1}{2}(p_i+q_j)) \]

- **Min-Weight Matching** \( \Rightarrow \) easy to compute
- Is this optimal?
Solution via Minimum Weight Bipartite Matching

\[
\begin{align*}
    w_{ij} &= D(p_i \parallel \frac{1}{2} (p_i + q_j)) + D(q_j \parallel \frac{1}{2} (p_i + q_j)) \\
\end{align*}
\]

- **Min-Weight Matching** \(\Rightarrow\) easy to compute
  - A slight variant of this test is asymptotically optimal!
Generalizations

- Different sets of distinct users observed in two days
  - Can be handled provided \# common users known

![Diagram showing different sets of users and their overlap]
Generalizations

- Different sets of distinct users observed in two days
  - Can be handled provided $\#$ common users known

Other applications: browsing statistics; language statistics; matching usernames on different websites
Experiment: Mobility traces on EPFL campus

Obtained from Wi-Fi connections
Dataset

- Anonymized mobility traces of $K \approx 1000$ users on EPFL campus measured on Mondays for two weeks
Anonymized mobility traces of $K \approx 1000$ users on EPFL campus measured on Mondays for two weeks

Number of locations $N = 933$ access points
Anonymized mobility traces of $K \approx 1000$ users on EPFL campus measured on Mondays for two weeks

Number of locations $N = 933$ access points

Data-length $T = 28800$ seconds (8 hours)
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th># users (K)</th>
<th>Matching $K$ users in second week (fraction of correct matches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mondays</td>
<td>1154</td>
<td>52.9%</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th># users ($K$)</th>
<th>Matching $K$ users in second week (fraction of correct matches)</th>
<th>Matching 1 random user in second week (prob of success)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mondays</td>
<td>1154</td>
<td>52.9%</td>
<td>44.5%</td>
</tr>
</tbody>
</table>

- Info on more users on second day ⇒ Higher accuracy
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th># users ($K$)</th>
<th>Matching $K$ users in second week (fraction of correct matches)</th>
<th>Matching 1 random user in second week (prob of success)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mondays</td>
<td>1154</td>
<td>52.9%</td>
<td>44.5%</td>
</tr>
<tr>
<td>Mondays and Tuesdays</td>
<td>1047</td>
<td>70.5%</td>
<td>53.5%</td>
</tr>
</tbody>
</table>

- Info on more users on second day $\Rightarrow$ Higher accuracy
- More days $\Rightarrow$ Higher accuracy

**Conclusion:** Simple anonymization is not effective
Outline of the talk

1. Designing Sensor Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories

2. Spatial Anti-aliasing via Mobile Sensing

3. Privacy of Mobility traces

4. Recap
Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories
Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories

- Perfect reconstruction conditions; Shortest trajectories
  - Uniform set better than unions of Uniform sets for $\mathbb{R}^2$
  - Optimal design of parallel trajectory sets for $\mathbb{R}^d$
Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories

- Perfect reconstruction conditions; Shortest trajectories
  - Uniform set better than unions of Uniform sets for $\mathbb{R}^2$
  - Optimal design of parallel trajectory sets for $\mathbb{R}^d$

- Spatial anti-aliasing via time-domain filtering
Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories

- Perfect reconstruction conditions; Shortest trajectories
  - Uniform set better than unions of Uniform sets for $\mathbb{R}^2$
  - Optimal design of parallel trajectory sets for $\mathbb{R}^d$

- Spatial anti-aliasing via time-domain filtering

- Mobility statistics: Simple anonymization is ineffective
  - Optimal de-anonymization strategy can be identified
A new sampling theory
A new sampling theory

\[ \Delta \]

\[
\begin{array}{c}
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\end{array}
\]

\[
\begin{array}{c}
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\end{array}
\]

\[
\begin{array}{c}
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\end{array}
\]

\[
\begin{array}{c}
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\end{array}
\]

\[
\begin{array}{c}
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\end{array}
\]
A new sampling theory
Thank You!

Questions?
Mobile Sensing


Privacy of Mobile Traces

Uniform sets for $\mathbb{R}^3$
Uniform sets for $\mathbb{R}^3$
Uniform sets for $\mathbb{R}^3$

$p$

$u$

$\text{Lat}(p)$
Theorem: For convex and symmetric $\Omega$

$p \in \mathcal{N}_\Omega$ iff $\text{Lat}(p)$ forms sampling lattice for $\Omega \cap u_{\perp}$. Furthermore

$$\ell(p) = \text{Sampling density(\text{Lat}(p))}$$
Section of a set
Extensions

- Generalize trajectories further to manifolds
Extensions

- Generalize trajectories further to **manifolds**

- Sampling on arbitrary curves: Ill-posed but can be fixed
Extensions

- Generalize trajectories further to manifolds

- Sampling on arbitrary curves: Ill-posed but can be fixed

- Time-varying bandlimited fields e.g. spatial audio fields
  - Reduce sensor density by increasing temporal sampling rate