OLD DOMINION UNIVERSITY

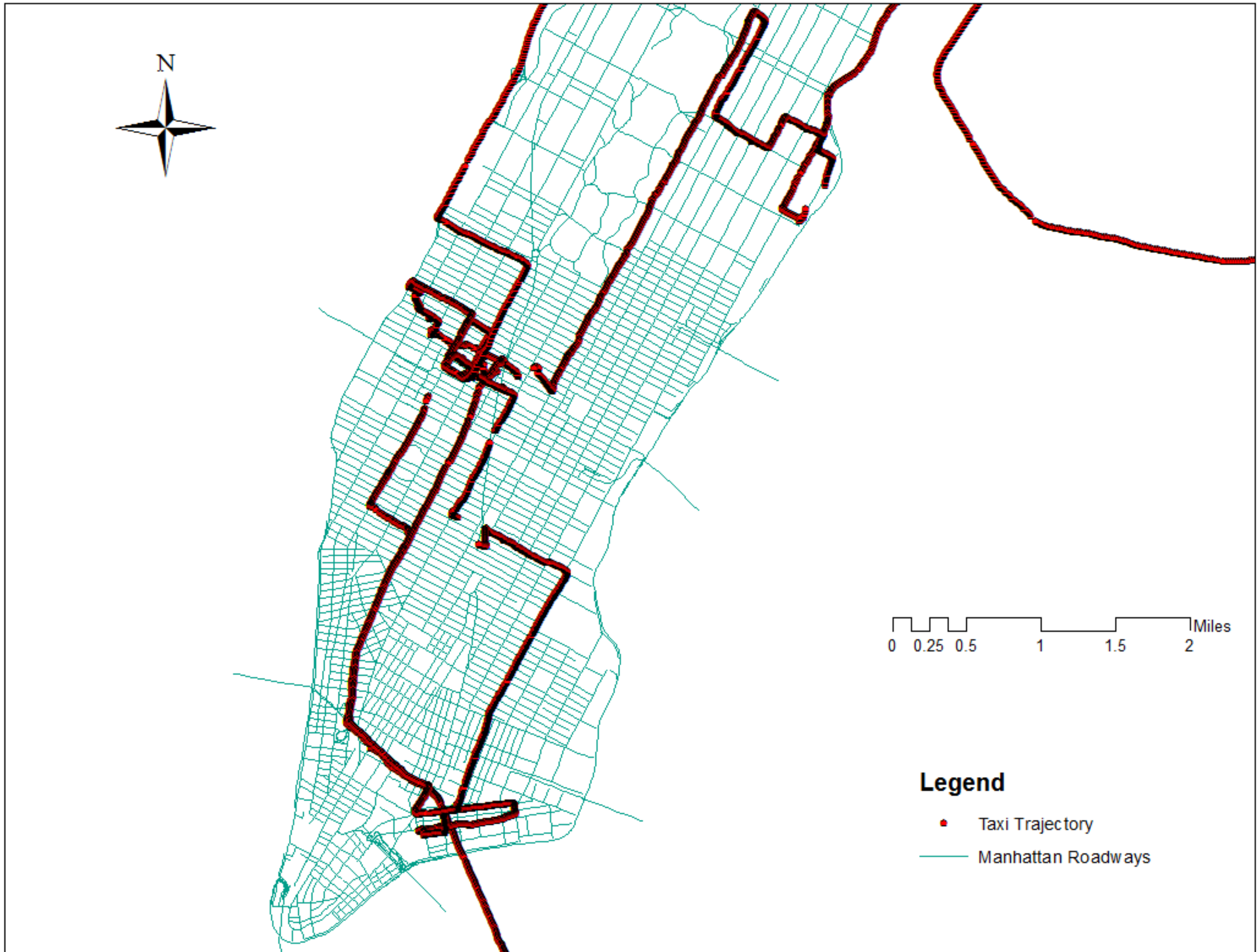
IDEA FUSION

Sparse GPS Trajectory Data Compression and Recovery based on Compressed Sensing

Hong Yang, Zhenyu Wang
Department of Modeling, Simulation and Visualization Engineering
Old Dominion University

November 20, 2015

Motivations



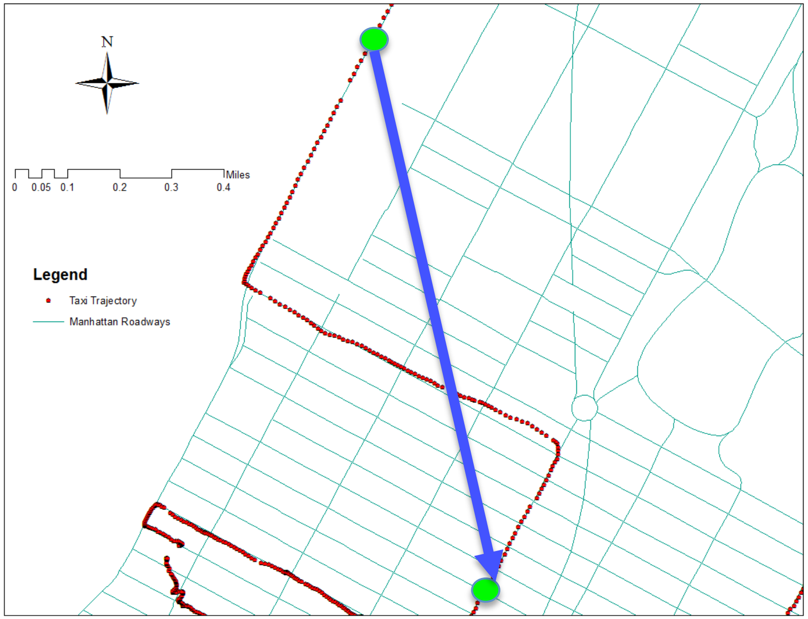
Massive GPS/ Smartphone trajectory data

Privacy concern

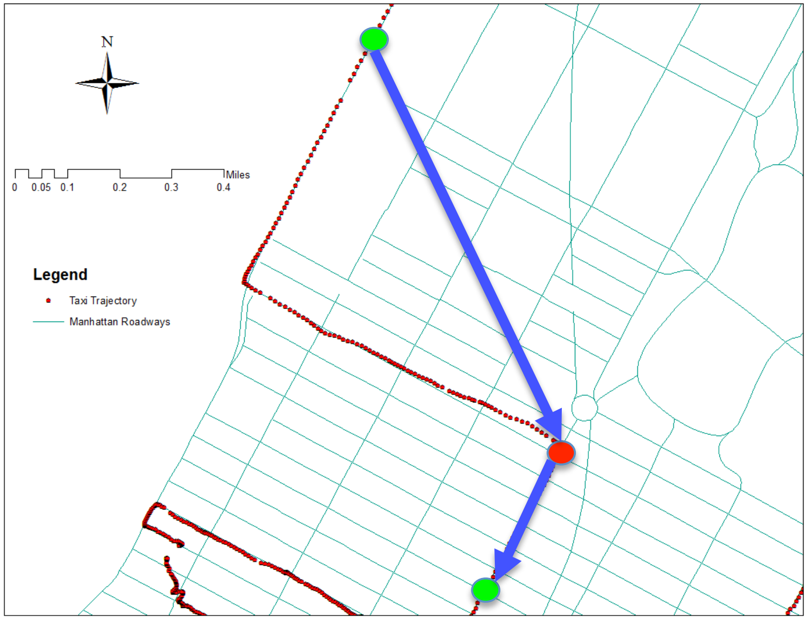
Data needs/ applications

Storage/processing issues

Challenges for Over-Compressed Data

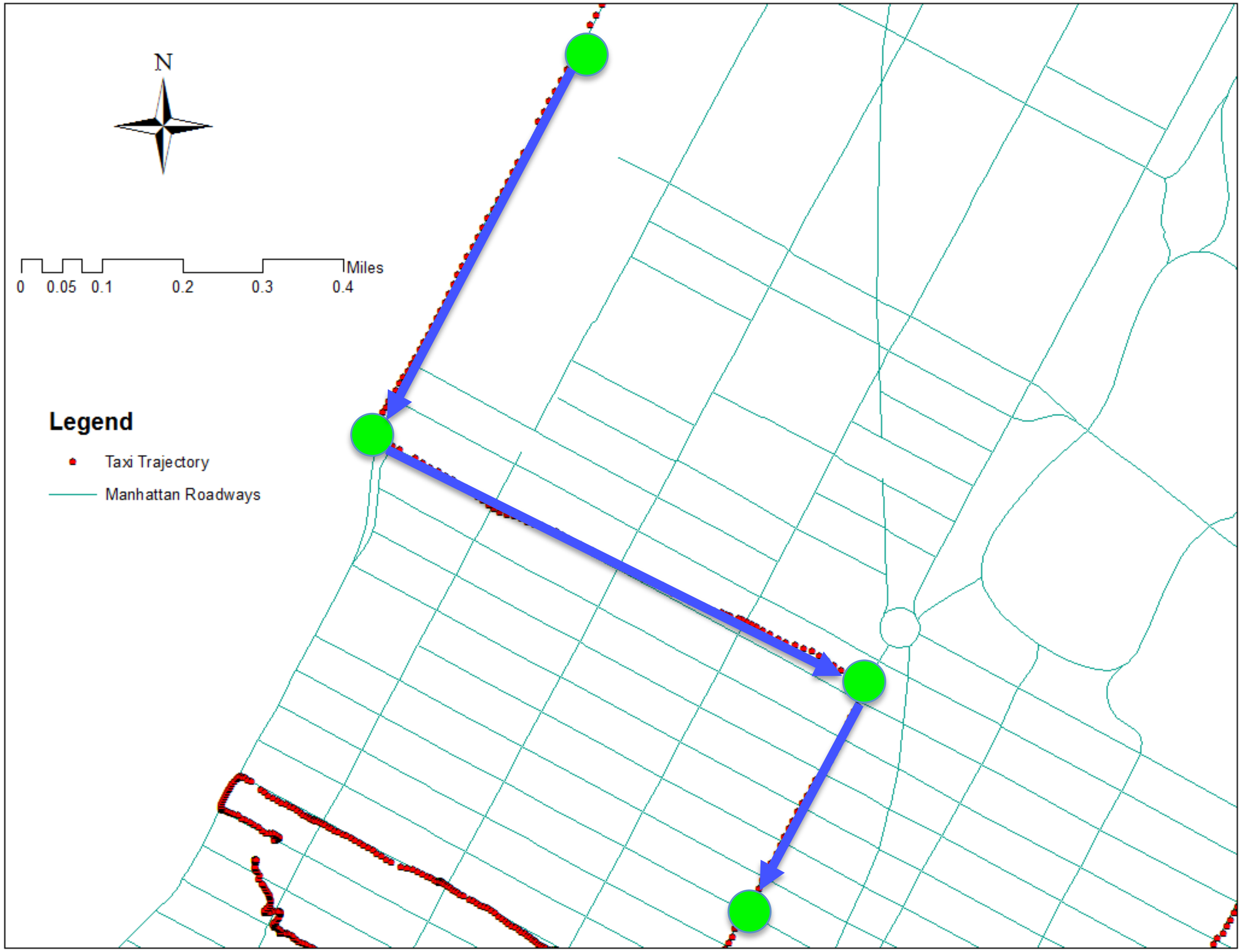


Only Origin and Destination are available. (i.e. current NYC Taxi trip data)



Origin, Destination and if given some points in between

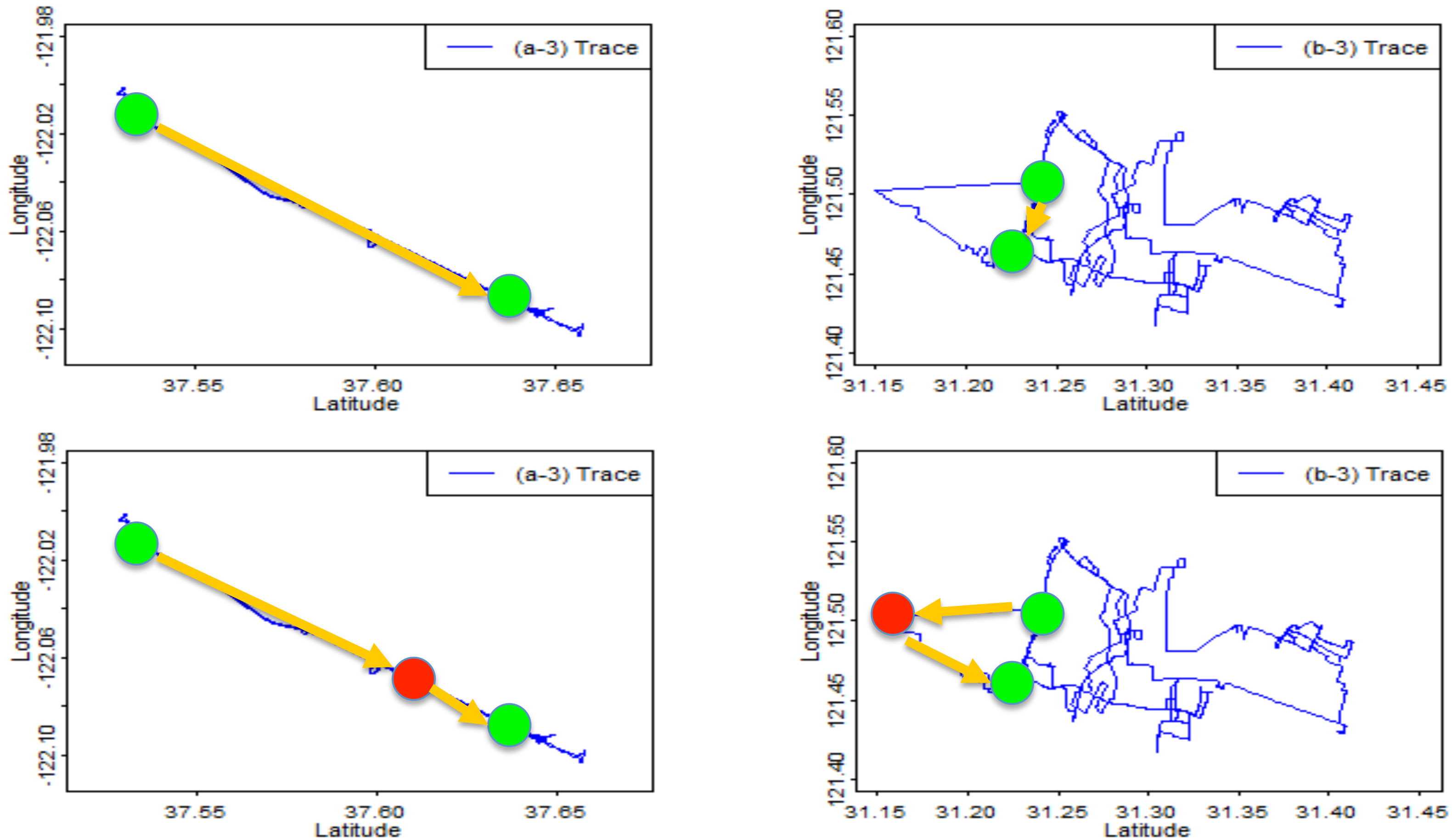
What if we provide more data points?



A better representation of the route?

What if we provide more data points?

- Different compressibility of the GPS trajectory may have different effects on different networks



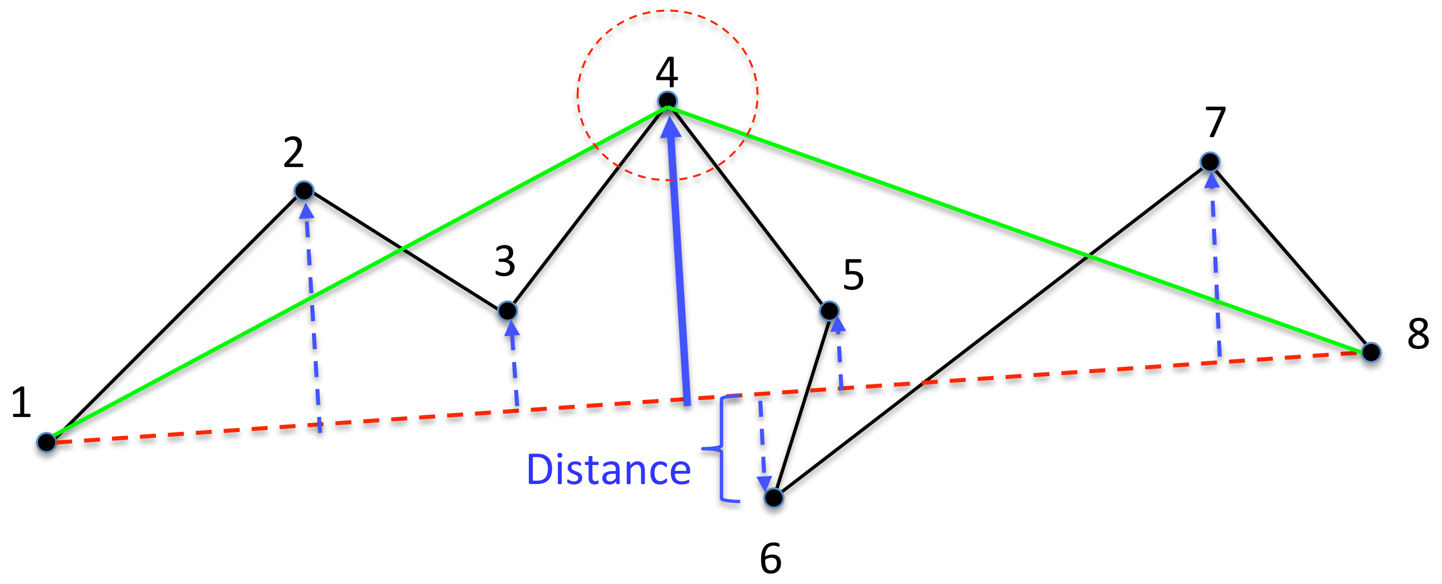
Existing Research / Practices

- Several classical methods using spatial and temporal dimensions to compress data
  - Uniform sampling
  - Douglas-Peucker algorithm
  - Bellman's algorithm
  - STTrace algorithm

- Other new methods using dimensions like sparsity and category
  - Greedy matching pursuit algorithm (GMP)
  - Compressed sensing (CS)
  - Coupled Hidden Markov Models

Methodology: Douglas-Peucker (DP) Algorithm

- DP - Using the spatial information

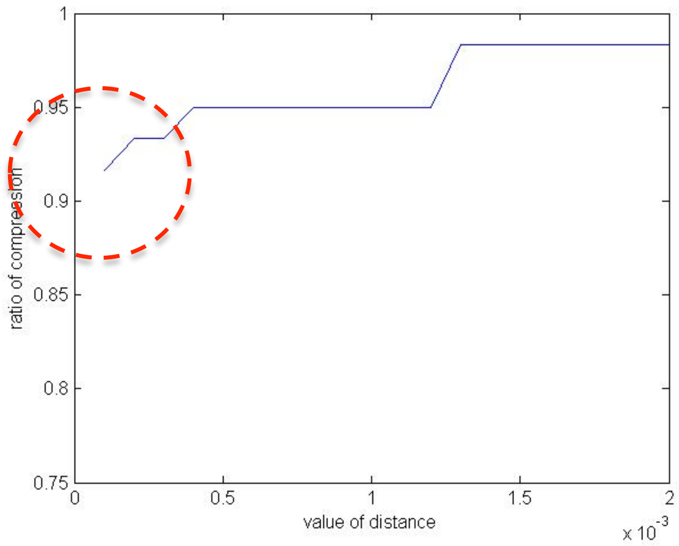


- Step 1: Link nodes 1 and 8
- Step 2: Identify node with maximum distance (Node 4)
- Step 3: Iteration

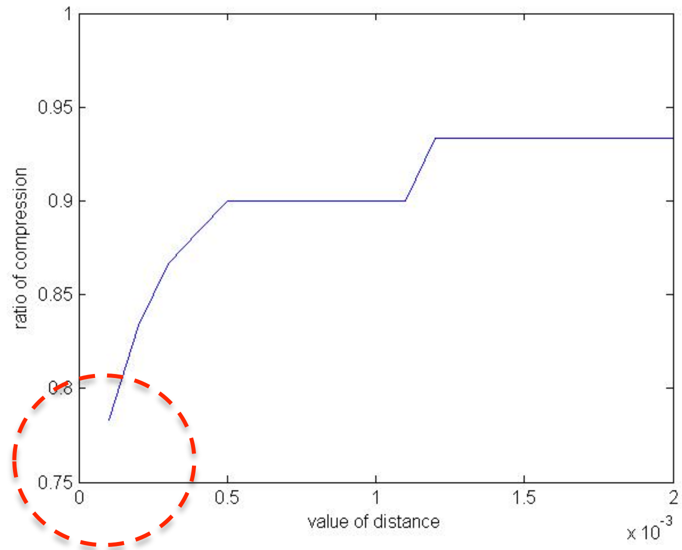
Issues of DP Algorithm

- Threshold of DP and the compression rate

Compression for highway data



Compression for local street data

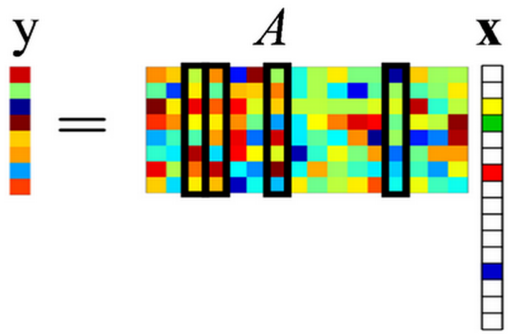


Some Issues:

- Need to sample all the data at the beginning
- Hard to deal with loop in the trajectory

Methodology: Compressed Sensing

- CS – using the sparsity information



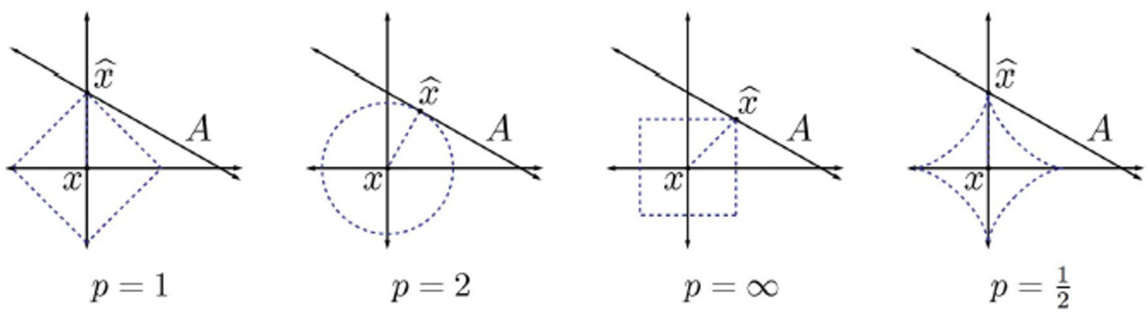
$$Ax = y$$

$$A = \begin{bmatrix} a_{11} & a_{12} & 1 & 1 & a_{15} \\ a_{21} & a_{22} & 1 & 1 & a_{25} \\ a_{31} & a_{32} & 1 & 1 & a_{35} \end{bmatrix},$$

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

The sparsity of the trajectory in the sparse basis is the key point sparse(A)>2k

The recovery of the 2-norm of x is a convex problem



Methodology: Compressed Sensing

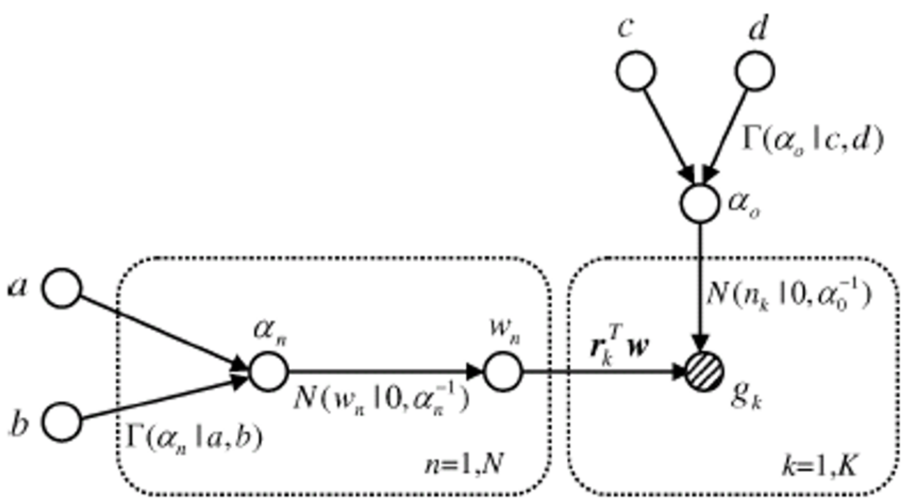
- $y = Ax = sample\_matrix \times basis\_matrix \times x$

$= sample\_matrix \times Tr$

y: $sampled\ data\ (m \times 1)$ $sample\_matrix$: $\mathrm{m} \times n$

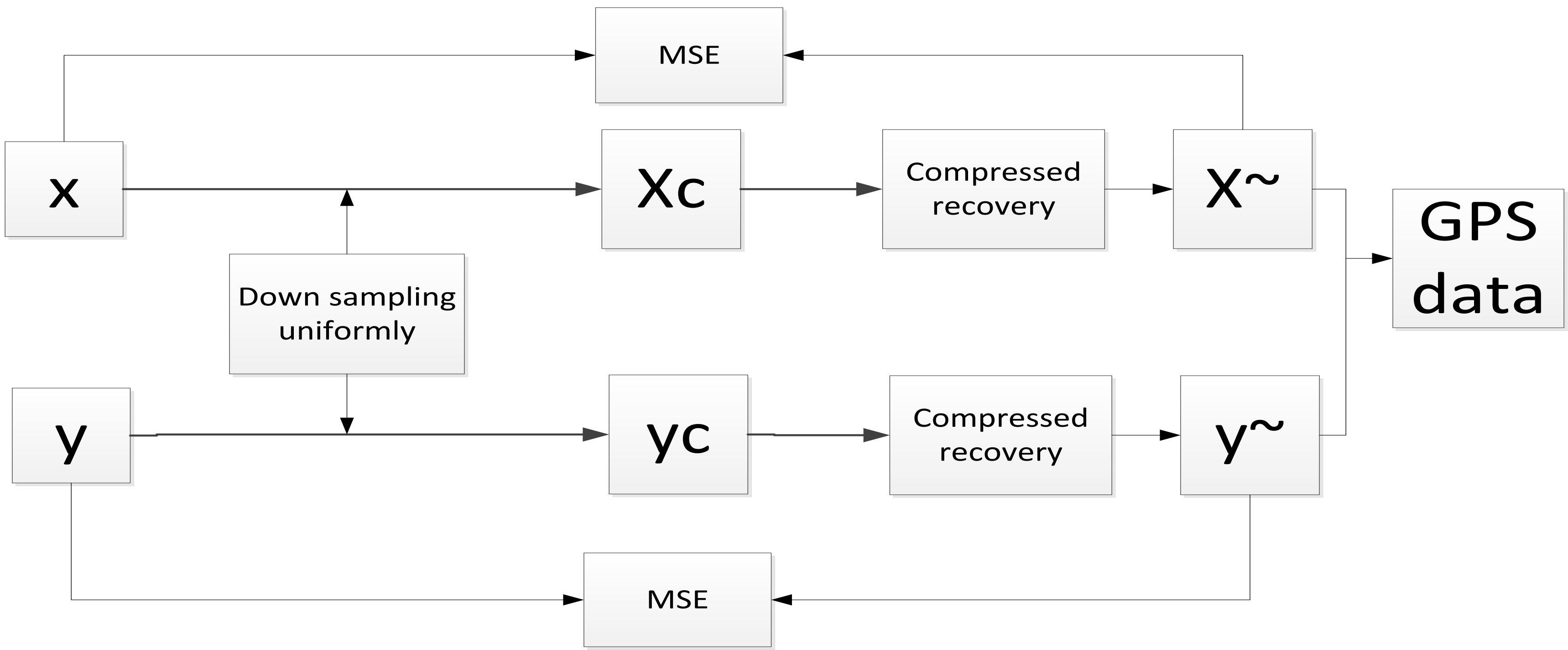
x: $n \times 1$ $basis\_matrix$: $n \times n$ Tr: $raw\ data\ (n \times 1)$

- Choose suitable $sample\_matrix$ and $basis\_matrix$ to get a sparse representation x of y
- The mode (BCS)



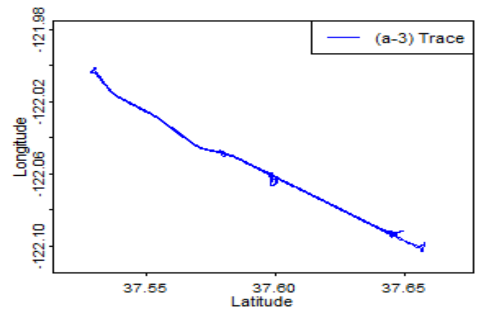
Methodology: Compressed Sensing (CS)

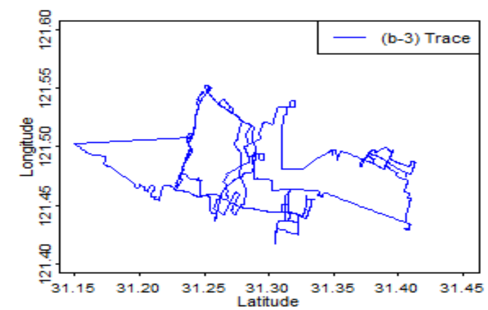
- Framework of CS in GPS data compression



Simulation Results

- Test Scenarios





- Data and Parameters

Parameter	Description	Value
τ_1	The interval of trajectory observation (highway)	3s
τ_2	The interval of trajectory observation (local street)	10s
N	The length of the processed sequence	60
n	The number of Monte Carlo simulations	100
б	The dB value of Gaussian noise	15

- Performance measure

$$\mathrm{Error}(x, y, x^{\sim}, y^{\sim}) = \frac{\left\|(x^{\sim}, y^{\sim}) - (x, y)\right\|_2}{\left\|(x, y)\right\|_2}$$

Simulation Results

- Performance of DP vs. CS with compression rate = 0.5

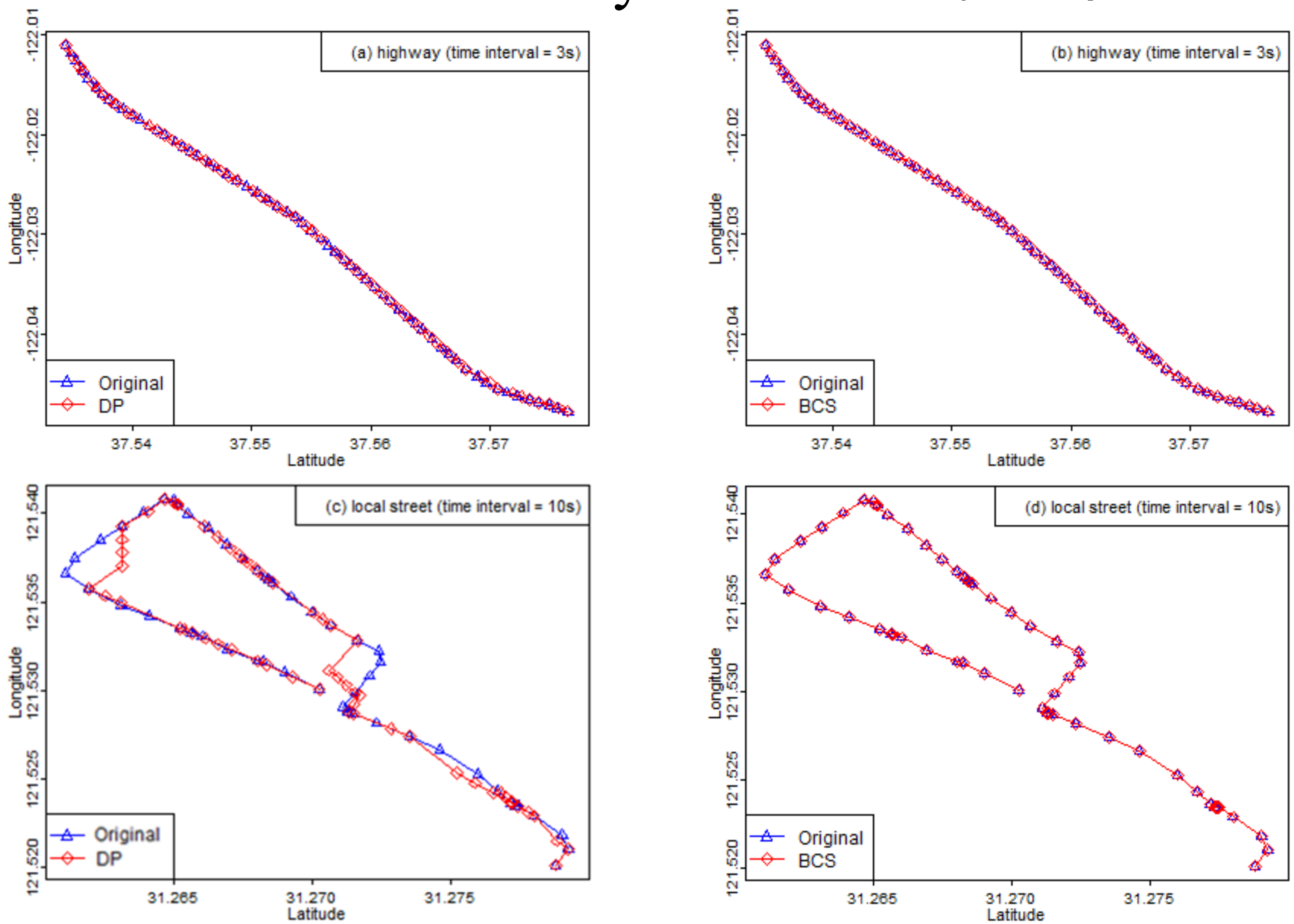
Data	Algorithm	Error
Vehicle 1 (highway)	DP	$2.7*10^{-6}$
Vehicle 1 (highway)	BCS	$3.4*10^{-11}$
Vehicle 2 (local street)	DP	0.1291
Vehicle 2 (local street)	BCS	$9.1*10^{-10}$
Vehicle 3 (highway)	DP	$3.9*10^{-6}$
Vehicle 3 (highway)	BCS	$1.1*10^{-11}$
Vehicle 4 (local street)	DP	0.1825
Vehicle 4 (local street)	BCS	$7.7*10^{-10}$
Vehicle 5 (highway)	DP	$3.3*10^{-6}$
Vehicle 5 (highway)	BCS	$1.06*10^{-11}$
Vehicle 6 (local street)	DP	0.2132
Vehicle 6 (local street)	BCS	$4.3*10^{-10}$

- Performance with Gaussian noise of 15dB in highway

Compression rate	Scenario	DP Error	BCS Error
0.5	highway	0.1040	0.0697
0.2	highway	0.1192	0.1041

Simulation Results

- Trace and recovered trace by DP & BCS (compression = 0.5)



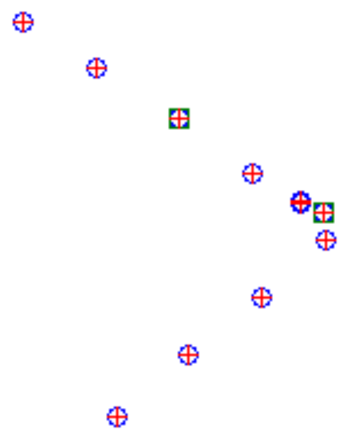
Discussion

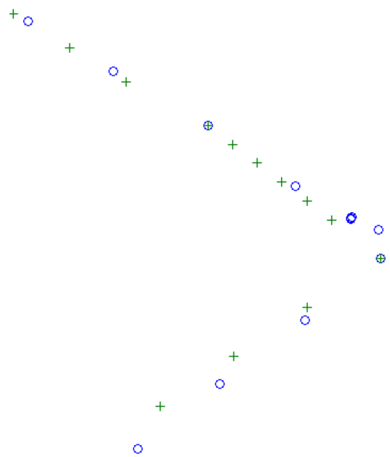
- Loss/Distortion of information (acceleration, speed, travel time, etc.)

Labels:
Circle: original points
Plus: recovered points

10 real points on local street

CS Approach

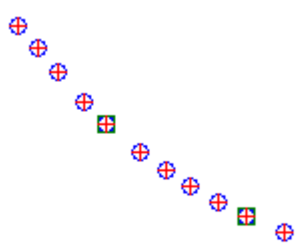


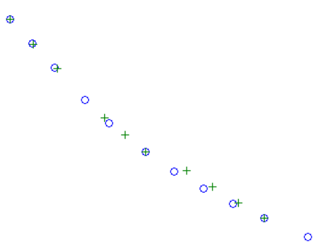


DP Approach

10 real points on highway

CS Approach





DP Approach

Concluding Remarks

- Raw GPS data can be represented relatively well by using appropriate compression techniques
- The proposed BCS approach can achieve relatively higher compression rate but maintain a better performance
- Despite the complexity, BCS approach does not require to store all raw data before sampling
- Compression means information loss/distortion (Consider trade off between compression rate and information change)

References

- Tobler, W. R. Numerical map generalization, In Nystuen, J.D., ed.: IMaGe Discussion Papers. Michigan Interuniversity Community of Mathematical Geographers. University of Michigan, Ann Arbor, Mi, USA, 1966.
- Douglas, D. H., and T. K. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Cartographica: *The International Journal for Geographic Information and Geovisualization,* Vol. 10, No. 2, 1973. pp. 112–122
- Bellman, R. On the approximation of curves by line segments using dynamic programming, *Commun. ACM*, Vol. 4, No. 6, 1961, pp. 284,
- Potaminas, M., K. Patroumpas, and T. Sellis. Sampling trajectory streams with spatiotemporal criteria. *In 18th International Conference on Scientific and Statistical Database Management*, 2006, pp. 275-284
- Zhang, B., X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, Sparse target counting and localization in sensor networks based on compressive sensing, in *Proc. IEEE INFOCOM*, 2011
- Xu, D., Y. Wang, H. Li, et al. The measurement of road traffic states under high data loss rate, *Measurement* Vol. 69, 2015, pp.134-145.
- Herring, R., A. Hofleitner, P. Abbeel, and A. Bayen. Estimating arterial traffic conditions using sparse probe data. 2010, *13th International IEEE Annual Conference on Intelligent Transportation Systems.*

Thank You Very Much!

Hong Yang, Ph.D.
Assistant Professor
Dept. of Modeling, Simulation & Visualization Engineering
Old Dominion University
Norfolk, VA
Email: hyang@odu.edu