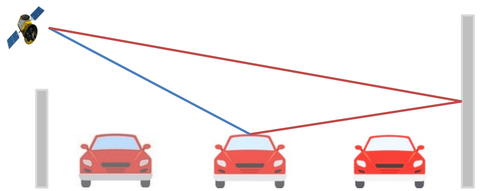
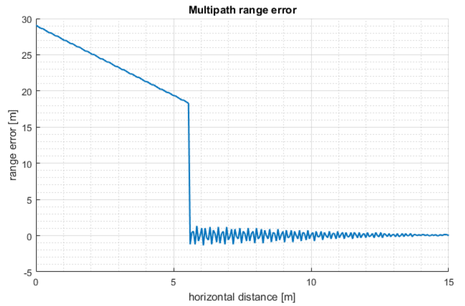
GPS Multipath Detection and Mitigation

Shiwen Zhang, shiwenzh@stanford.edu
Department of Aeronautics & Astronautics, Stanford University

Motivation





Background: GPS multipath refers to the phenomenon where satellite signals are reflected by buildings before reaching the user receiver. Such reflections can cause significant errors in the user navigation solutions.
Goal: Detect and remove GPS multipath outliers to improve position accuracy.
Method: Use unsupervised outlier detection algorithms on GPS data collected at the receiver to detect and remove satellite signals that are corrupted by multipath.

Data

- GPS measurements were collected by a stationary receiver over a period of 24 hours on the roof of Durand Building at Stanford University
- Raw data from the receiver include observation data and navigation messages
  - Observation data include code-phase and carrier-phase measurements, and signal-to-noise ratios, recorded at a rate of 1 Hz for each satellite in view
  - Navigation messages include information to calculate satellite position and satellite clock error
- The true position of the receiver is known, but data are considered as unlabeled as no information was given regarding which signals contain multipath

Features

- Features directly extracted from raw input data are the code-phase measurements and the signal-to-noise ratios at two different frequencies for each satellite
- Features derived from input data were azimuth and elevation angles of each satellite
- Dimension of the features depends on whether dual frequency signals are available for that satellite

Algorithms

Cluster-based detection algorithm

- k-means: Input examples were first classified into 2 clusters. An outlier score was then assigned to each data point based on its distance to the centroid of the cluster with higher average signal-to-noise ratio.

Nearest neighbor based detection algorithms

- k-nearest neighbor (k-NN): k-nearest neighbor search was performed on the inputs and an outlier score was assigned to each data point based on the average distance to the k nearest neighbors. Smoothing and peak detection were then performed to detection sharp variation in the outlier scores with respect to time.
- k^{th}-NN: The same procedure as k-NN was used in this algorithm, but unlike k-NN, the outlier score was based on the distance to the k^{th} nearest neighbor.
- Local outlier factor (LOF): k-nearest neighbor search was first performed. And the local reachability density (LRD) was computed for each data point.

$$LRD_k(x) = 1/\left(\frac{\sum_{o\in N_k(x)} d_k(x,o)}{|N_k(x)|}\right)$$

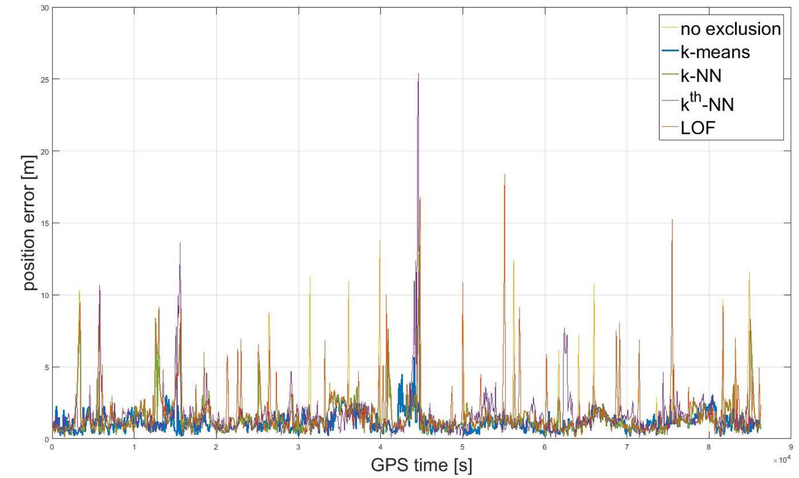
Finally the LOF score is computed.

$$LOF(x) = \frac{\sum_{o\in N_k(x)} \frac{LRD_k(o)}{LRD_k(x)}}{|N_k(x)|}$$

Reference

- Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, *11*(4), e0152173.
- Hautamaki, V., Karkkainen, I., & Franti, P. (2004, August). Outlier detection using k-nearest neighbour graph. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 3, pp. 430-433). IEEE.
- Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May). LOF: identifying density-based local outliers. In *ACM sigmod record* (Vol. 29, No. 2, pp. 93-104). ACM.

Results and Discussion



Algorithm	False positive rate	False negative rate	Position error improvement ratio	Average position error improvement
K-means	28.59%	1.24%	71.17%	0.7360 m
K-NN	9.49%	9.43%	34.97%	0.4585 m
K^{th}-NN	47.45%	2.11%	42.94%	0.3513 m
LOF	18.98%	15.75%	1.84%	0.0926 m

Discussion and Conclusion

- The baseline algorithm uses all the available GPS measurements to calculate position solution. The detection algorithms predict whether a measurement is an outlier or not, and use only the set of outlier-free measurements to estimate the position of the receiver. Algorithm performance was evaluated by comparing position error at each time step against position error of the baseline.
- The false positive rate for all four algorithms are relatively high because the algorithms have no way to distinguish between multipath error and other propagation errors, which can be estimated and removed from the measurements.
- The LOF algorithm performs poorly because it does not take into account the correlation of measurements with respect to time.

Future Work

- Explore more features such as carrier-phase measurements and Doppler measurements.
- Explore how detection results generalize for different multipath environments and for different days of the year.
- Investigate how the size of the dataset affects detection accuracy.