GPS Multipath Detection and Mitigation

Shiwen Zhang
Department of Aeronautics and Astronautics
Stanford University
Email: shiwenzh@stanford.edu

***Abstract*—This paper examines the performance of multiple unsupervised learning algorithms in detecting and mitigating multipath error in the Global Positioning System (GPS). The algorithms detect outliers in GPS measurements from each satellite and exclude the outliers from being used to calculate the navigation solutions. Four outlier detection algorithms were implemented. Position errors for each algorithm were computed and compared against the no exclusion algorithm. Experimental results show that the k-means cluster-based algorithm achieves the best performance. Error analysis shows that failing to exclude one or more outliers and excluding too many signals can both cause increase in position error.**

I. Introduction

The Global Positioning System (GPS) is a satellite-based navigation system that provides worldwide positioning, navigation, and timing services to users. GPS consists of 24 or more satellites broadcasting radio signals that carry satellite locations, status, and precise time from on-board atomic clock. A GPS receiver uses these information to calculate its distance to each satellite in view. With more than four satellites in view, the receiver can estimate its location in three dimensions using geometry.

GPS multipath refers to the phenomenon when the satellite signals are reflected by buildings before reaching the user receiver. Such reflections can cause significant errors in the user navigation solutions. Multipath is a significant source of error in urban environment. Identifying and reducing the effect of multipath would enable GPS and the other satellite navigation constellations to contribute to high integrity railway control and autonomous vehicle operating in urban environments.

This paper explores the use of unsupervised learning algorithms to detect and remove GPS multipath outliers from the received signals. The input of the algorithm are the GPS data collected by a stationary receiver. The algorithm was applied to data from each satellite separately. The algorithm then labels the outliers in the data. Position solution is then calculated by using the set of outlier-free measurements as predicted by the algorithm. Four outlier detection algorithms were implemented and performance of each algorithm was compared against a baseline algorithm which uses all received signals to calculate position solutions.

II. Related Work

There are many studies done on GPS multipath detection based on satellite exclusion. Meguro [1] described a detection technique that uses a omni-directional infrared camera to exclude invisible satellites. Obst [2] used a 3D environmental map and ray-tracing to predict and exclude multipath signals. On the other hand, Phan [3] presented a method to estimate multipath error using a nonlinear support vector regression model. But all these detection approaches require additional information from various sensors or models. This paper attempts to use unsupervised machine learning techniques to detect multipath signals purely based on unlabeled GPS measurements.

There exist many unsupervised anomaly detection algorithms and a comparative evaluation of these algorithms was given by Goldstein [4]. These algorithms can be categorized into nearest-neighbor based methods, cluster-based methods, and statistical methods. A nearest-neighbor based global anomaly detection algorithm was introduced in [5] that detects outliers based on the sum of the distances from the k-nearest neighbors for each data point. A modification of this algorithm was presented in [6] that used only the k^{th}-nearest neighbor to assign outlier scores. The Local Outlier Factor (LOF) algorithm can detect local anomalies by computing the local density of the k-nearest neighbors for each data point [7]. The Cluster-Based Local Outlier Factor (CBLOF) algorithm used a clustering algorithm to determine dense areas in the data and assigned outlier scores based on the distance of each instance to its cluster centroid and the cluster density [8]. The Histogram-based Outlier Score (HBOS) method used a statistical approach to estimate the probability of each feature assuming independent feature [9].

III. Dataset and Features

GPS signals include ranging signals and navigation messages. Each satellite transmits a unique ranging signal and a receiver can decode the ranging signals to obtain code-phase measurements, carrier-phase measurements, Doppler measurements, and signal-to-noise ratios, for all GPS satellites in view. Some satellites also transmit radio signals on a second frequency and the receiver will decode ranging measurements on both frequencies. Navigation messages contain satellite ephemeris data, which include orbital information to calculate satellite positions, as well as other information about the time and status of the constellation.

GPS data were collected by a stationary receiver on the roof of Durand Building at Stanford University on October 29, 2017. The receiver recorded ranging measurements for each satellite at a rate of 1 Hz over a period of 24 hours. Each row of the dataset contains GPS time in seconds, the satellite identification number, and the ranging measurements. The true position of the receiver is known, but the dataset is considered as unlabeled as no information was given regarding which satellite signals are corrupted by multipath.

The dataset was then split into several subsets, one for each satellite. Features directly extracted from the dataset are the code-phase measurements and the signal-to-noise ratios. Features derived from input data were azimuth and elevation angles of each satellite. Dimension of the features for each satellite depends on whether dual frequency signals are available for that specific satellite.

IV. Methods

Four unsupervised outlier detection algorithms were implemented and tested on the dataset.

A. *K-means Cluster-based Detection Algorithm*

The k-means clustering algorithm was used to classify input data 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. The scores are then normalized to values between 0 and 1. The cutoff value for outliers is set to be 0.58. Any data point with an outlier score higher than the cutoff value was determined by the algorithm as an outlier.

B. *K-Nearest Neighbors Detection Algorithm (k-NN)*

Features were first normalized to values between 0 and 1. The k-nearest neighbor search was then performed on the normalized input and an outlier score was assigned to each data point based on the average distance to the k-nearest neighbors. $k = 20$ was chosen for this algorithm. Smoothing and peak detection were then applied to detect sharp variation in the outlier scores with respect to GPS time. Outliers were determined based on the absolution height of the peaks and the peak prominence.

C. *K^{th}-Nearest Neighbor Detection Algorithm (k^{th}-NN)*

The same procedure as k-NN was used for this algorithm. But unlike k-NN, the outlier score was based on the distance to the k^{th}-nearest neighbor instead of based on the average distance of all k neighbors. The same value of k was chosen for this algorithm.

D. *Local Outlier Factor Detection Algorithm (LOF)*

The k-nearest neighbor search was performed on the normalized features with $k = 100$. And the local reachability density (LRD) was computed for each data point.

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

where $d_k(x, o)$ is the Euclidean distance.

Then the LOF score for each data point was computed by comparing the LRD of the data point to the LRD's of its k-nearest neighbors.

$$LOF_k(x) = \frac{\sum_{o \in N_k(x)} \frac{LRD_k(o)}{LRD_k(x)}}{|N_k(x)|} \tag{2}$$

LOF value cutoff for outliers were set to be 1.5 and any data point with a LOF score higher than 1.5 was considered as an outlier.

V. Experimental Results

At each time step, the position solution of the receiver was computed using a set of received satellite signals. At least four satellites are needed to get a unique position solution. The detection algorithms output the set of signals that are outlier-free to be used in calculating position solutions. The baseline algorithm does not exclude any received signals and uses all the received signals to calculate position solutions. Position error at each time step was then computed for the four detection algorithms and the baseline algorithm. Error corrections were applied to the measurements before computing position solution so that all the major errors were removed and position error were mostly due to

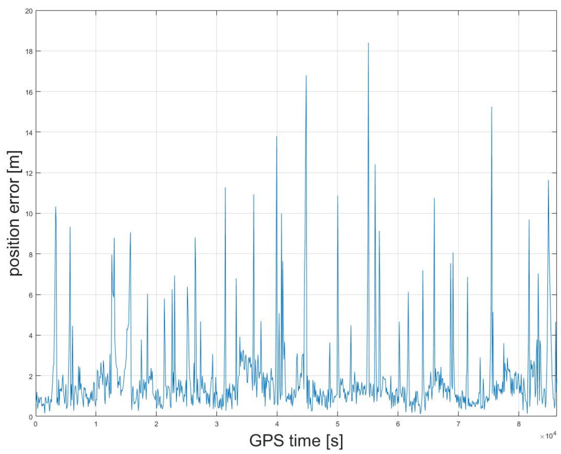


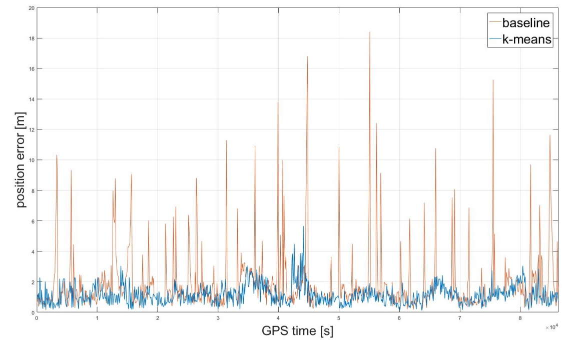
Fig. 1. Position error for the baseline algorithm.

signal multipath. Figure 1 is the position error at each time step for the baseline algorithm. The nominal 3D position accuracies for GPS after error corrections are typically around 2 to 3 meters. The spikes in Fig.1 indicate that there are one or more outliers present in the signals used for position solution at the time of the spikes.

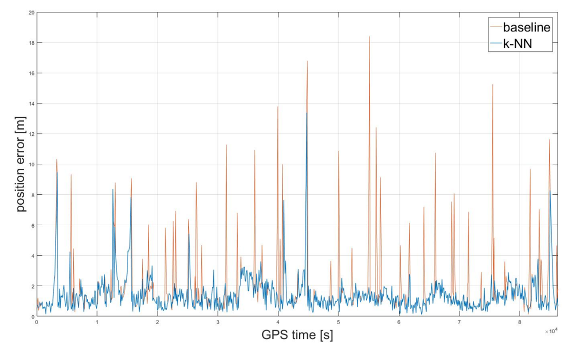
Position errors for each of the four detection algorithms were calculated and compared with the errors for the baseline algorithm. Figure 2 shows the comparison results. According to the plots, the k-means algorithm achieved the best overall performance. The algorithm was able to suppress all the spikes due to outliers. But there occurred in some instances increase in position error comparing to the baseline algorithm. Position accuracy suffered in these instances because the k-means algorithm incorrectly classified some of the nominal signals as outliers.

GPS position solution depends on both the accuracy of the ranging measurements and the geometry of the satellites in view. Excluding satellites before calculating the position solution results in worse satellite geometry. Position solutions improve when multipath error in the ranging measurements is greater than the loss in geometry by excluding satellites. But when multipath error is not present in the signals, excluding the normal signals will result in worse geometry and therefore worse position accuracy.

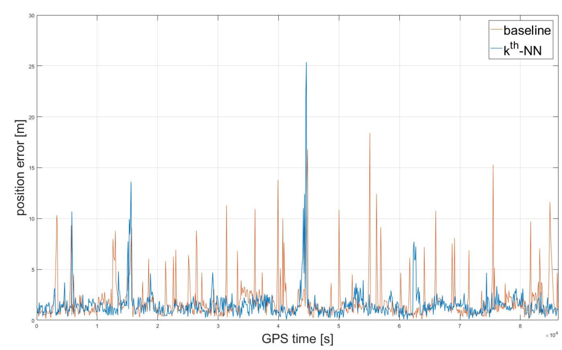
Therefore, it is important to look at the false alarm rate and miss detection rate when evaluating the performance for each of the algorithm. Four metrics were used to evaluate the performance: false alarm rate, miss



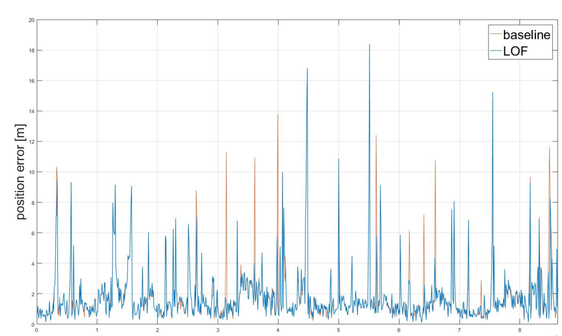
(a)



(b)



(c)



(d)

Fig. 2. Position error comparison between the baseline algorithm and (a) k-means; (b) k-NN; (c) k^{th}-NN; (d) LOF.

detection rate, position error improvement ratio, and average position error improvement. Since the ground truth of multipath signals is unknown, these quantities have to be defined differently on the position solution level rather than on ranging signal level. False alarm rate is defined as the percentage of times when the detection algorithm performed worse than the baseline algorithm. Miss detection rate is defined as the percentage of times when the detection algorithm performed worse than the baseline algorithm when outliers are present. And outliers are assumed to be present when the position error for the baseline algorithm is greater than 2.2627 meters. The position error improvement ratio is defined as the difference between the number of outlier instances for baseline and the number of outlier instances after applying the detection algorithm, divided by the number of outlier instances for baseline. And the average position error improvement is defined as the average differences between the position error in baseline and the position error for the detection algorithm for each time step.

Table I shows the summary of the evaluation results for the four detection algorithms. The false alarm rate for all four algorithms are relatively high because the algorithms have no way to distinguish between multipath error and other propagation and clock errors from the input features. The propagation errors and clock errors were estimated and removed from the ranging measurements and therefore did not contribute to the position error in the final solution. But if any of these errors are high in the ranging measurements, the detection algorithms will classify the measurements as outliers even though they are not multipath outliers. And since multipath error is highly correlated with time, the LOF algorithm performs poorly because it did not take into account this correlation in the detection process.

VI. Conclusion

The paper implemented, evaluated and compared four unsupervised outlier detection algorithms to detect GPS multipath outliers in ranging measurements. All algorithms achieved improvement in overall position accuracy in GPS solutions. But the performance for correctly detecting outliers varies greatly across different algorithms.

Future work will be done on exploring different features such as carrier-phase measurements and Doppler measurements. Statistical based detection algorithms will also be explored and implemented. The algorithms need to be tested on more and larger datasets to investigate how the detection results generalize for different multipath environments and for different days of the year. How the size of the dataset affects detection accuracy also requires further study if the algorithm need to be implemented in real-time.

TABLE I
SUMMARY OF DETECTION RESULTS

Algorithm	False alarm rate	Miss detection 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

REFERENCES

[1] Meguro, Jun-ichi, et al. ”GPS multipath mitigation for urban area using omnidirectional infrared camera.” *IEEE Transactions on Intelligent Transportation Systems* 10.1 (2009): 22-30.

[2] Obst, Marcus, Sven Bauer, and Gerd Wanielik. ”Urban multipath detection and mitigation with dynamic 3D maps for reliable land vehicle localization.” *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*. IEEE, 2012.

[3] Phan, Quoc-Huy, Su-Lim Tan, and Ian McLoughlin. ”GPS multipath mitigation: a nonlinear regression approach.” *GPS solutions* 17.3 (2013): 371-380.

[4] Goldstein, Markus, and Seiichi Uchida. ”A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data.” *PloS one* 11.4 (2016): e0152173.

[5] Angiulli, Fabrizio, and Clara Pizzuti. ”Fast outlier detection in high dimensional spaces.” *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer, Berlin, Heidelberg, 2002.

[6] Ramaswamy, Sridhar, Rajeev Rastogi, and Kyuseok Shim. ”Efficient algorithms for mining outliers from large data sets.” *ACM Sigmod Record*. Vol. 29. No. 2. ACM, 2000.

[7] Breunig, Markus M., et al. ”LOF: identifying density-based local outliers.” *ACM sigmod record*. Vol. 29. No. 2. ACM, 2000.

[8] Amer, Mennatallah, and Markus Goldstein. ”Nearest-neighbor and clustering based anomaly detection algorithms for rapidminer.” Proc. of the 3rd RapidMiner Community Meeting and Conference (RCOMM 2012). 2012.

[9] Goldstein, Markus, and Andreas Dengel. ”Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm.” KI-2012: Poster and Demo Track (2012): 59-63.