Machine Learning-Based GPS Multipath Detection Method Using Dual Antennas

Sanghyun Kim  
School of Integrated Technology  
Yonsei University  
Incheon, Korea  
sanghyun.kim@yonsei.ac.kr

Jungyun Byun  
School of Integrated Technology  
Yonsei University  
Incheon, Korea  
jungyun.byun@yonsei.ac.kr

Kwansik Park  
Korea Aerospace Research Institute  
Daejeon, Korea  
kspark6469@kari.re.kr

Abstract—In urban areas, global navigation satellite system (GNSS) signals are often reflected or blocked by buildings, thus resulting in large positioning errors. In this study, we proposed a machine learning approach for global positioning system (GPS) multipath detection that uses dual antennas. A machine learning model that could classify GPS signal reception conditions was trained with several GPS measurements selected as suggested features. We applied five features for machine learning, including a feature obtained from the dual antennas, and evaluated the classification performance of the model, after applying four machine learning algorithms: gradient boosting decision tree (GBDT), random forest, decision tree, and K-nearest neighbor (KNN). It was found that a classification accuracy of 82%–96% was achieved when the test data set was collected at the same locations as those of the training data set. However, when the test data set was collected at locations different from those of the training data, a classification accuracy of 44%–77% was obtained.

Index Terms—multipath detection, global positioning system, machine learning, dual antennas

I. INTRODUCTION

Global navigation satellite systems (GNSSs), particularly the global positioning system (GPS) of the United States of America, play an essential role in various aspects of our lives [1]–[19]. Recently, as autonomous vehicles, such as unmanned aerial vehicles (UAVs), have emerged, the use of GNSS has increased further [20]–[34]. However, in urban areas, it is difficult to achieve high positioning accuracy with the GNSS because the signals get reflected by buildings in many cases [35]–[42]. Using the reflected GNSS signals for positioning could result in a positioning error greater than 100 m [43]. From the perspective of signal reception, the GNSS signal can be received in the following three cases: when only line-of-sight (LOS) signal is received (that is, LOS-only condition), only non-line-of-sight (NLOS) signal is received (that is, NLOS-only condition), and both LOS and NLOS signals are simultaneously received (that is, LOS+NLOS condition). Among these conditions, the performance degrades only in the NLOS-only, and LOS+NLOS conditions. Therefore, various methods for detecting multipath signals (that is, NLOS-only or LOS+NLOS signals) and mitigating the corresponding GNSS errors are being actively studied [44]–[46].

The traditional method of detecting multipath signals uses the carrier-to-noise-density ratio (C/N₀) measurement of the GNSS signal [47]. By defining a threshold for the C/N₀ values, a signal with C/N₀ greater than the threshold is classified as an LOS-only signal, and a signal with C/N₀ lower than the threshold is classified as an NLOS-only or LOS+NLOS signal. Similarly, there are numerous software-based detection algorithms for GNSS measurements, such as the Doppler shift and code-minus-carrier values [48], [49]. Additionally, there are hardware-based detection methods that use a special antenna [50], a sky-pointing camera [51], and a method that uses a three-dimensional (3D) model of a city and ray-tracing techniques [52]. Furthermore, GNSS multipath signal detection techniques using machine learning approaches have been emerging recently [53]–[55].

Various machine learning techniques have been used for classifying GNSS signal reception conditions, and their multipath detection performances were better than those of traditional methods [56]–[58]. The performance of a machine learning technique is greatly influenced by the features that are used, and the algorithms that are applied to it. Guermah et al. [59] proposed a method that used the C/N₀ difference value obtained by a dual-polarized antenna, as a feature for the machine learning model. Sun et al. [57] classified LOS-only, NLOS-only, and LOS+NLOS signals by applying the gradient boosting decision tree (GBDT) algorithm, and used C/N₀ values, pseudorange residuals, and satellite elevation angles as features for machine learning.

In this study, we proposed a machine learning-based GPS multipath detection method that uses dual antennas. Using dual antennas instead of a single antenna helps to obtain more diverse GPS measurements, and utilize them as features for machine learning. Resultantly, it is expected that the performance of classifying signal reception conditions would be improved. Using more than two antennas might be more effective, but it has disadvantage in terms of the size and cost.
We used four features (namely, $C/N_0$, temporal difference of $C/N_0$, difference between delta pseudorange and pseudorange rate, and satellite elevation angle) that could be obtained from a single antenna, and one feature (namely, double-difference pseudorange residual) that could be obtained from dual antennas. Four algorithms were applied as the machine learning algorithms in this study, and their performances were compared. There has been a study on GNSS multipath detection using dual antennas [59]. However, unlike our study, it was based on a statistical approach, and not a machine learning approach. Compared to statistical methods, machine learning methods can identify dependencies in data sets that cannot be modeled mathematically [60]. We classified signal reception conditions by a machine learning method that utilizes five features, whereas the existing method [59] classified the conditions by a statistical method that uses a single metric.

II. GPS SIGNAL COLLECTION AND LABELING

A. GPS Signal Collection

Fig. 1 shows the GPS signal acquisition hardware that was set up to collect the large amount of GPS data for machine learning. It consisted of two right-hand circularly polarized (RHCP) antennas produced by Antcom, two NovAtel PwrPak7 GNSS receivers, and a laptop for storing raw GPS data. The distance between the two antennas was maintained at 19 cm, which was equal to the wavelength of the GPS L1 signal. We collected GPS L1, L2, and L5 signals at intervals of 1 s, in a static environment.

B. Data Labeling Based on Ray-Tracing Simulation

It is essential to label the true signal reception conditions on the GPS data collected for machine learning. In this study, labeling was performed through ray-tracing simulation, using a commercial ray-tracing software called Wireless InSite [61]. The ray-tracing simulator displays the propagation paths of the signal from the GPS satellite to the receiver, based on the 3D building information around the receiver. We used a commercial 3D building map provided by 3dbuildings [62], and calculated the locations of the GPS satellites based on almanac data.

III. FEATURE SELECTION AND MACHINE LEARNING ALGORITHMS

A. Features to Classify the Signal Reception Conditions

The features used in this study were as follows: $C/N_0$, temporal difference of $C/N_0$, difference between delta pseudorange and pseudorange rate, satellite elevation angle, and double-difference pseudorange residual.

- $C/N_0$: $C/N_0$ is defined as the carrier-to-noise density ratio. It is a parameter containing received signal strength information, and its unit is dB-Hz. Generally, since the propagation loss increases when a signal is reflected by an obstacle, the $C/N_0$ value of the LOS-only signal is larger than that of the NLOS-only signal.

- Temporal difference of $C/N_0$: This parameter is denoted as $\Delta C/N_0$, which indicates the change in $C/N_0$ between two consecutive epochs. In a static environment, $C/N_0$ values of NLOS-only or LOS+NLOS signals can increase.

| GPS data set       | T0  | T1  | T2  | T3  |
|--------------------|-----|-----|-----|-----|
| Total samples      | 7500| 7500| 6311| 11575|
| NLOS-only samples  | 2500| 2500| 2038| 3608|
| LOS-only samples   | 2500| 2500| 2195| 4033|
| LOS+NLOS samples   | 2500| 2500| 2078| 3934|

The data collected at the remaining two locations (that is, D and E) were used only for testing. T2 and T3 in Table I represent the data set collected at locations D and E, respectively.
with respect to time, based on the characteristics of the receiver tracking loop [58].

- Difference between delta pseudorange and pseudorange rate: This parameter represents the consistency between the pseudorange measurement and the Doppler shift. It has been proposed as a feature in [58], and can be calculated as follows:

\[ \Delta \rho - \dot{\rho} \cdot \Delta t \]  

where \( \Delta \rho \) is the change in pseudorange between two epochs; \( \dot{\rho} \) is the pseudorange rate, which is related to the Doppler shift; and \( \Delta t \) is the time difference between the two epochs.

- Satellite elevation angle: Generally, the higher the elevation angle of the satellite, the lower is the probability that the signal is reflected or blocked by a building.

- Double difference pseudorange residual: The above four features can be obtained from a single antenna, and they have been proposed in previous literature. However, dual antennas are necessary to obtain the double difference pseudorange residual, which was proposed as a feature for multipath detection for the first time in this study. This feature can be calculated as follows:

\[
\rho_{12,\text{residual}}^{kj} = \rho_{12,m}^{kj} - \rho_{12,e}^{kj} = \{(\rho_{1,m}^{kj} - \rho_{2,m}^{kj}) - (\rho_{1,e}^{kj} - \rho_{2,e}^{kj})\} - \{(\rho_{1,e}^{k} - \rho_{2,e}^{j}) - (\rho_{1,e}^{1} - \rho_{2,e}^{2})\} 
\]

where \( \rho_{12,m}^{kj} \) and \( \rho_{12,e}^{kj} \) are the measured and expected double difference pseudorange values between satellites \( k \) and \( j \), and receivers 1 and 2, respectively. The expected pseudorange is the distance between the satellite and the receiver, where the satellite and receiver positions are obtained based on the ephemeris data and the receiver position fix, respectively. Since this feature was calculated based on the signals from two satellites, the satellite with the highest elevation angle was selected as the reference satellite for the given epoch and the double difference pseudorange residual was calculated between each satellite and the reference satellite.

If the received signal was NLOS-only or LOS+NLOS, its propagation path from the satellite to each of the two antennas would be noticeably different, which caused the absolute value of the double difference pseudorange residual to increase. Therefore, this feature can be used for the classification of signal reception conditions.

**B. Machine Learning Algorithms**

The four machine learning algorithms used in this study were GBDT, random forest, decision tree, and K-nearest neighbor (KNN). We used the algorithms in a machine learning library called scikit-learn [63], and performed hyperparameter tuning using the GridSearchCV function included in scikit-learn.

**IV. RESULTS**

As shown in Fig. 2, the machine learning model was trained using the data set T0, and the trained model was tested using the data sets T1, T2, and T3. Fig. 3 shows the classification accuracy of each machine learning algorithm for the test data set T1, T2, and T3. When the trained model was tested using the data set T1, the classification accuracy was 82%–96%, which was significantly high. This was because the training and test data sets, T0 and T1, respectively, consisted of data collected at the same locations (that is, locations A, B, and C). However, when the trained model was tested using the test data sets T2 and T3, which were collected at locations different from the training data locations (that is, locations D and E), classification accuracies of 46%–77%, and 44%–55% were obtained, respectively. The decision tree algorithm performed better than other algorithms in this case.

Figs. 4–6 show the classification accuracy of each machine learning algorithm according to the signal reception conditions. For the data set T1, as shown in Fig. 4, an accuracy of 70% or greater was achieved. However, for the data sets T2 and T3, as shown in Figs. 5 and 6, respectively, the classification accuracy of the LOS+NLOS signals was only 18%–53%, which was significantly lower than that of the other cases. This result implied that the training data set should include more LOS+NLOS data samples to achieve a better training result. This was understandable because the LOS+NLOS environment had greater diversity than the LOS-only or NLOS-only environments, and thus required more training.

**V. CONCLUSION**

In this study, we proposed a machine learning-based GPS multipath detection method that uses dual antennas. We used five features for machine learning, comprising four features that could be obtained from a single antenna, and one feature that could be obtained from dual antennas. The performances of four machine learning algorithms (that is, GBDT, random forest, decision tree, and KNN) using the proposed features...
were evaluated. The classification accuracy for LOS+NLOS signals was found to be low. However, it is expected to improve if more LOS+NLOS data samples are used for the training, and this can be investigated in a future work.

REFERENCES

[1] K. Park, W. Kim, and J. Seo, “Effects of initial attitude estimation errors on loosely coupled smartphone GPS/IMU integration system,” in Proc. ICCAS, Oct. 2020, pp. 800–803.

[2] H. Yoon, H. Seok, C. Lim, and B. Park, “An online SBAS service to improve drone navigation performance in high-elevation masked areas,” Sensors, vol. 20, no. 11, 2020.

[3] S. Jeong, H. Lee, T. Kang, and J. Seo, “RSS-based LTE base station localization using single receiver in environment with unknown path-loss exponent,” in Proc. ICTC, Oct. 2020, pp. 958–961.

[4] K. Shamaei and Z. M. Kassas, “A joint TOA and DOA acquisition and tracking approach for positioning with LTE signals,” IEEE Trans. Signal Process., vol. 69, pp. 2680–2705, 2021.

[5] W. Kim, P.-W. Son, J. Rhee, and J. Seo, “Development of record and management software for GPS/Loran measurements,” in Proc. ICCAS, Oct. 2020, pp. 796–799.

[6] M. Maaref and Z. M. Kassas, “Ground vehicle navigation in GNSS-challenged environments using signals of opportunity and a closed-loop map-matching approach,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 7, pp. 2723–2738, 2020.

[7] J. Park, P.-W. Son, W. Kim, J. Rhee, and J. Seo, “Effect of outlier removal from temporal ASF corrections on multichain Loran positioning accuracy,” in Proc. ICCAS, Oct. 2020, pp. 824–826.

[8] P.-W. Son, J. Rhee, J. Hwang, and J. Seo, “Universal kriging for Loran ASF map generation,” IEEE Trans. Aerosp. Electron. Syst., vol. 55, no. 4, pp. 1828–1842, Oct. 2019.

[9] S. Han, T. Kang, and J. Seo, “Smartphone application to estimate distances from LTE base stations based on received signal strength measurements,” in Proc. ITC-CSCC, Jun. 2019.

[10] J. H. Rhee, S. Kim, P.-W. Son, and J. Seo, “Enhanced accuracy simulator for a future Korean nationwide eLoran system,” IEEE Access, vol. 9, pp. 115042–115052, Aug. 2021.

[11] E. Schmidt, N. Gatisis, and D. Akopian, “A GPS spoofing detection and classification correlator-based technique using the LASSO,” IEEE Trans. Aerosp. Electron. Syst., vol. 56, no. 6, pp. 4224–4237, 2020.

[12] K. Park and J. Seo, “Single-antenna-based GPS anti-jamming method exploiting polarization diversity,” IEEE Trans. Aerosp. Electron. Syst., vol. 57, no. 2, pp. 919–934, Apr. 2021.

[13] M. Braasch and A. Dempster, “Tutorial: GPS receiver architectures, front-end and baseband signal processing,” IEEE Aerosp. Electron. Syst. Mag., vol. 34, no. 2, pp. 20–37, 2019.

[14] J. Lee, D. Lee, and J. Seo, “Dual-polarized GPS antenna array algorithm to adaptively mitigate a large number of interference signals,” Aerosp. Sci. Technol., vol. 78, pp. 387–396, Jul. 2018.

[15] S. Kim, K. Park, and J. Seo, “Mitigation of GPS chirp jammer using a transversal FIR filter and LMS algorithm,” in Proc. ICTC-CSCC, Jun. 2019.

[16] V. L. Knoop, P. F. de Bakker, C. C. J. M. Tiberius, and B. van Arem, “Lane determination with GPS precise point positioning,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 9, pp. 2503–2513, 2017.

[17] T. Kang and Y. Shin, “Indoor navigation algorithm based on a smartphone inertial measurement unit and map matching,” in Proc. ICTC, Oct. 2021, pp. 1421–1424.

[18] A. K. Sun, H. Chang, S. Pullen, H. Kil, J. Seo, J. Y. J. Morton, and J. Lee, “Matkov chain-based stochastic modeling of deep signal fading: Availability assessment of dual-frequency GNSS-based aviation under ionospheric scintillation,” Space Weather, vol. 19, no. 9, pp. 1–19, Sep. 2021.

[19] H. Lee, S. Pullen, J. Lee, B. Park, M. Yoon, and J. Seo, “Optimal parameter inflation to enhance the availability of single-frequency GBAS for intelligent air transportation,” IEEE Trans. Intell. Transp. Syst., in press.

[20] S. Kim, J. Park, J.-K. Yun, and J. Seo, “Motion planning by reinforcement learning for an unmanned aerial vehicle in virtual open space with static obstacles,” in Proc. ICCAS, Oct. 2020, pp. 784–787.

[21] K. Sun, H. Chang, J. Lee, J. Seo, Y. Jade Morton, and S. Pullen, “Performance benefit from dual-frequency GNSS-based aviation applications under ionospheric scintillation: A new approach to fading process modeling,” in Proc. ION ITM, Jan. 2020, pp. 889–899.
