Calibrating Dead Reckoning with Deep Reinforcement Learning

Sangmin Lee and Hwangnam Kim
Department of Electrical Engineering
Korea University
Seoul, Republic of Korea
lsm5505, hnkim@korea.ac.kr

***Abstract*—In general, as the moving distance increases, positioning accuracy degradation due to cumulative error is pointed out as a major problem in DR (Dead Reckoning). The distance-based positioning system has the advantage of high positioning accuracy depending on the sensor used. However, in environments where the use of infrastructure is restricted, such as security facilities or disaster environments, its utilization is limited, and Line of Sight (LOS) between devices must be guaranteed. To compensate for those limitations, we propose a framework for resetting the initial point of an object in DR. In order to reduce the position error that increases in proportion to the accumulated distance of DR, a method of resetting the starting point of DR to the position derived by distance-based trilateration was adopted. However, the process of resetting every moment when the trilateration-based position is derived can increase the system overhead and cause a positioning delay. Therefore, we derived the optimal starting point reset position and frequency by using Deep Reinforcement Learning (DRL). Through simulation, we verified that the proposed system improves the positioning accuracy compared to conventional DR through low resource consumption.**

***Index Terms*—Dead Reckoning, Deep Reinforcement Learning, Trilateration, Indoor Positioning System**

I. INTRODUCTION

The field of indoor positioning technology, which has been consistently spotlighted and studied, is classified into various technologies depending on whether infrastructure exists and whether distances and angles are used. In addition, it is classified in various ways according to the types of sensors and radio signals used, and positioning methods. Among indoor positioning technologies, Dead Reckoning (DR) based on Inertial Measurement Unit (IMU) is used in various environments. In particular, it is a positioning technology with high potential because it can estimate the location of an object with only a single device in an environment where communication infrastructure is limited.

However, DR is rarely used as a single mechanism. This is because DR is an offline tracking structure, unlike the online structure that can derive the current location in real time. Since it is a method of estimating the position by calculating the direction and speed of an object from a known starting position by using accumulated data, it has a serious problem of reducing positioning accuracy due to accumulated error. In particular, if a large error occurs in the direction or speed at the beginning of the path, the final estimated position is derived significantly different from the actual position. In this cumulative error problem, the final position error increases in proportion to the moving distance of the object. To solve this problem, DR usually forms a form of using a mixture of two or more positioning systems. Therefore, DR is generally utilized by fusion of two or more positioning systems [1].

In this paper, we propose a calibration system that resets the starting point using minimal infrastructure and resources to solve the accumulated error problem of dead reckoning. Deep Reinforcement Learning (DRL) and distance-based trilateration is utilized to reset the starting point. Depending on the required communication coverage or signal type, the wireless signal and sensor used for trilateration can be changed flexibly.

II. MODELING AND IMPLEMENTATION

A. System Overview

This paper proposes a system to reset the starting point of DR to the trilateration result at a specific point while an object is moving. When comparing single systems, the positioning accuracy of trilateration is generally better than that of DR. However, because DR-based positioning is used in an environment where the infrastructure is limited in whole or in part, trilateration cannot be applied in all locations. In addition, excessively frequent starting point resets may increase the system overhead and result in positioning delay. We adopted deep reinforcement learning to derive the optimal reset points to obtain high positioning accuracy while solving the overhead problem.

B. DRL Modeling

The proposed system utilizes Proximal Policy Optimization (PPO) [2] based DRL. We adopted the PPO algorithm because it has low computational and implementation complexity and is stable by preventing abrupt changes in policy by using a clipped surrogate objective function.

1) State/Action statement: The state, which is the input array of the deep neural network, consists of the position estimated at the current step, the difference between the estimated position and the actual position, and the system accumulated resource value in the current episode.

The action space of the learning model consists of 0 and 1. In case of value 0 is output, IMU-based DR is performed in

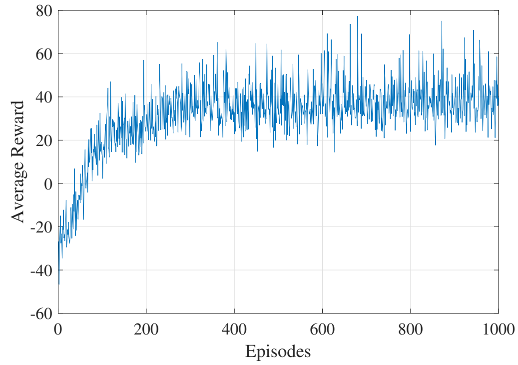


Fig. 1: Average rewards of 1000 episodes.

current step. In case of value 1, the DR starting point is reset to the distance-based positioning value with the pre-deployed anchor nodes and proceeds to the next step.

2) Reward function: The reward function $R(action)$ of the model is configured differently according to the output action value as shown in the following equation.

$$R(action) = \begin{cases} 1/Error, & \text{if } action = 0, \\ 1/Error - Resource, & \text{if } action = 1. \end{cases}$$

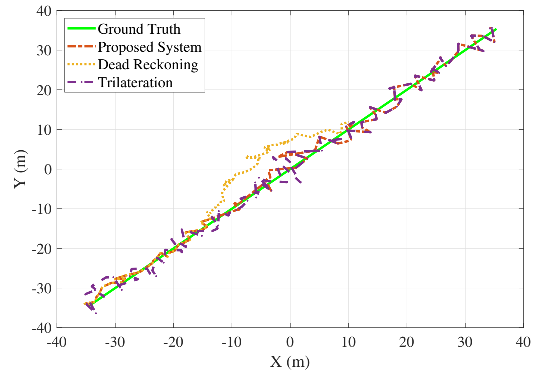
When an action of 0 is output, the reward function is formed in inverse proportion to the difference between the current estimated position and the actual position, that is, the positioning error. When an action of 1 is output, the reward is derived by subtracting the reset consumption resource from the inverse proportion of the positioning error. At this time, the value corresponding to the resource scale parameter is subtracted and the scale parameter can be adjusted appropriately for the sensor or radio signal type used for trilateration.

III. Performance Evaluation

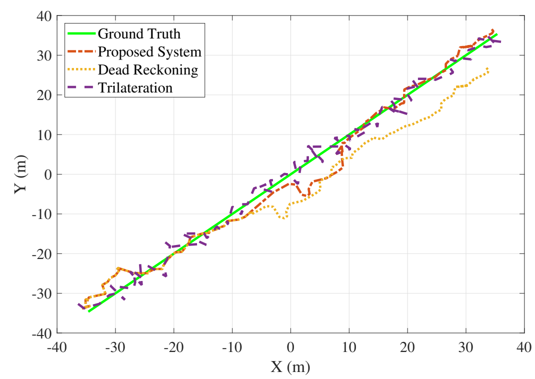
We verified the superior performance of the proposed system through simulation. In the simulation, the object moves at a constant velocity of $1m/s$ in a constant direction of positive 45 degrees at every step. The object travels a distance of 100 meters from $(-25\sqrt{2}, -25\sqrt{2}, 1)$ to $(+25\sqrt{2}, +25\sqrt{2}, 1)$ for a total of 100 steps within the episode. In addition, it is assumed that the object moves at a fixed altitude and has a resource scale factor of 2.0.

Fig. 1 shows the average rewards of each episode during the learning progress. After about 200 episodes, the reward showed convergence, and after about 400 episodes, it can be judged that the optimal DR calibration policy network is obtained as the convergence is actually completed.

Fig. 2 shows the object path of each system in episode 100 and 900. It can be seen that the proposed system in episode 100 resets the starting point at many times out of 100 steps because the learning process is not completed. Although positioning accuracy is better than DR, system overhead may increase by resetting the starting point a total of 48 times. In the path in episode 900, where it is judged that learning is complete, the proposed system shows the highest positioning accuracy



(a) Path of episode 100.



(b) Path of episode 900.

Fig. 2: Each system path according to the learning progress.

among all systems. In addition, in this case, there is little loss in terms of resources by performing reset only 5 times.

IV. Conclusion

In this paper, we propose an optimal DR calibration system that uses DRL to timely reset the DR starting point to a distance-based positioning value. In addition, the superiority of the proposed system was verified through simulation. The proposed DRL-based DR calibration system is expected to be utilized in a variety of environments because it is suitable for implementing an accurate DR system with low resource consumption.

Acknowledgment

This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-0-01835) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) and in part by the National Research Foundation of Korea funded by the Korean Government under Grant 2020R1A2C1012389.

References

[1] S. Qiu, Z. Wang, H. Zhao, K. Qin, Z. Li, and H. Hu, "Inertial/magnetic sensors based pedestrian dead reckoning by means of multi-sensor fusion," *Information Fusion*, vol. 39, pp. 108–119, 2018.

[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.