

Transformer-Based Smartphone PDR for Multiple Carrying Modes

Jooan Kim

*School of Electronic Engineering
Soongsil University
Seoul, South Korea
jooan0102@soongsil.ac.kr*

Yoan Shin*

*School of Electronic Engineering
Soongsil University
Seoul, South Korea
yashin@ssu.ac.kr
(*Corresponding author)*

Abstract—In smartphone inertial measurement unit (IMU)-based pedestrian dead reckoning, trajectory reconstruction performance is affected by smartphone carrying modes because sensor signal patterns vary with the user's carrying posture. In this paper, we propose a transformer encoder-based trajectory reconstruction model that estimates walking speed and heading changes from 9-axis IMU time-series data without manually selecting sensor axes for each carrying mode. The proposed model uses a 2 s IMU window as input and predicts walking speed and heading changes at 0.5 s intervals, which are accumulated to reconstruct a 2D pedestrian trajectory. The model was trained using approximately 16 h of IMU data collected in text, swing, and call carrying modes, and the learning targets were generated from real-time kinematic receiver speed and true-north-referenced heading information. Experimental results on nine test trajectories showed that the proposed model can reconstruct pedestrian trajectories under different carrying modes using a single integrated model, indicating the feasibility of smartphone IMU-based trajectory reconstruction without separate pose-specific models.

Index Terms—smartphone IMU, pedestrian dead reckoning, transformer encoder, carrying modes, trajectory reconstruction

I. INTRODUCTION

In smartphone inertial measurement unit (IMU)-based pedestrian dead reckoning (PDR), trajectory reconstruction performance is strongly affected by smartphone carrying modes. Although smartphone accelerometers and gyroscopes reflect the user's walking motion, their signal magnitude, pattern, and axis-dependent characteristics vary across carrying modes such as text, swing, and call, even along the same walking path [1]. Therefore, a PDR model trained or designed for a specific carrying mode may not provide stable trajectory reconstruction performance in practical environments with different carrying modes.

In practical walking environments, users commonly carry smartphones in various modes rather than in a single fixed carrying mode. Conventional PDR methods estimate movement and directional changes using step detection [2], Weinberg-based step-length estimation [3], and gyroscope- or magnetometer-based heading estimation [4], [5]. However, because the relationship between sensor axes and walking direction can vary depending on the carrying mode, these methods may require manual axis selection or calibration for each carrying mode [1]. Therefore, this study focuses on

estimating walking speed and heading changes directly from 9-axis IMU time-series inputs without explicitly performing such axis selection.

In this paper, we train a single PDR model based on the transformer encoder architecture [6] using smartphone 9-axis IMU data collected from text, swing, and call carrying modes. The model estimates walking speed and heading changes from a 2 s IMU window and reconstructs a two-dimensional pedestrian trajectory by accumulating the predicted values. The trajectory reconstruction performance was evaluated against the same ground truth (GT) trajectory using absolute trajectory error (ATE).

II. PROPOSED METHOD

A deep learning regression model based on the transformer encoder architecture was trained to estimate walking speed and heading changes from smartphone 9-axis IMU time-series data. The dataset consists of approximately 16 h of IMU data collected in the text, swing, and call carrying modes. These carrying modes were selected to reflect common smartphone usage conditions during walking, where the sensor orientation and motion patterns can vary depending on how the smartphone is held. During data collection, smartphone IMU data and real-time kinematic (RTK) data were recorded simultaneously so that the IMU measurements could be matched with accurate speed and heading information.

The input features consist of three-axis linear acceleration, three-axis gyroscope, and three-axis gravity vectors. All sensor data were resampled to 80 Hz to construct a uniform time-series input. Each input sample was generated using a 2 s sliding window, resulting in a 160×9 sequence. This window length was used to include sufficient temporal motion patterns for estimating walking speed and heading changes.

The learning targets were generated using the speed and true-north-referenced heading information obtained from the National Marine Electronics Association (NMEA) Recommended Minimum Specific GNSS Data (RMC) messages of the RTK receiver [7]. The RTK data were collected at 8 Hz and aligned with the smartphone IMU data resampled at 80 Hz based on a common time axis. For each input window, the RTK speed at the final time step t was used as the speed target, and the heading difference over the interval from $t - 0.5$ s to t was

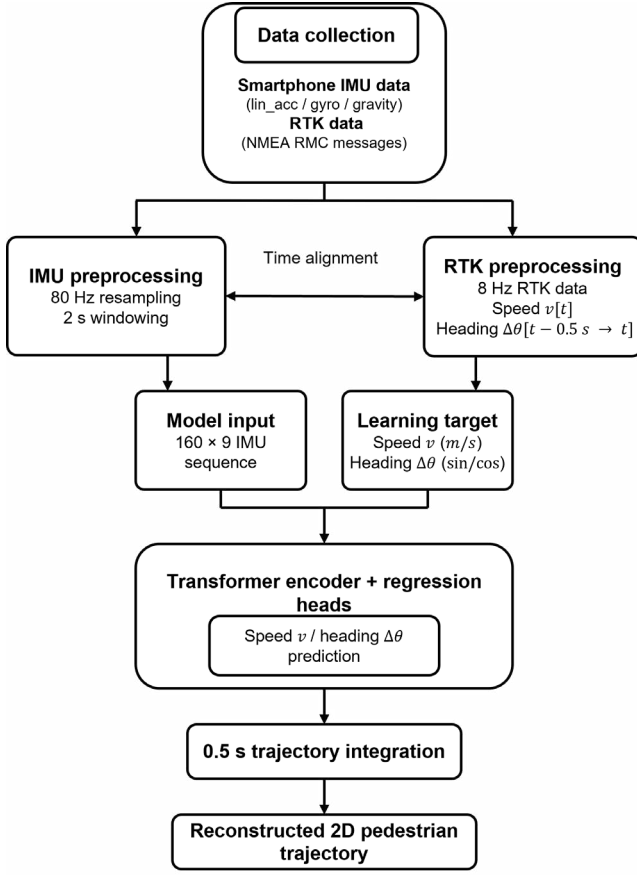


Fig. 1. Overall structure of the proposed smartphone IMU-based PDR system.

used as the heading change target. To reduce the discontinuity problem that may occur at angular boundaries, the heading change was converted into sine and cosine components for training.

The proposed model estimates walking speed and heading changes in an end-to-end manner from a 2 s 9-axis IMU input. Specifically, the model is trained to learn IMU time-series patterns from various carrying modes using a single set of model parameters, without manually selecting sensor axes for heading estimation or constructing separate models for the text, swing, and call carrying modes. This design enables a single inference pipeline to be used across different carrying modes.

Fig. 1 shows the overall structure of the proposed smartphone IMU-based PDR system. In the proposed framework, the IMU and RTK data have different roles. The IMU data are used as the model input after 80 Hz resampling and 2 s windowing. In contrast, the RTK data are used to generate the learning targets, as described above. Therefore, the RTK data provide supervision during training, but they are not used as input to the model during trajectory reconstruction.

Based on the preprocessed IMU window, the transformer encoder extracts temporal motion features related to walking dynamics. The extracted features are then passed to two regression heads that separately estimate walking speed and

heading changes. Finally, the estimated values are accumulated at 0.5 s intervals to reconstruct a two-dimensional pedestrian trajectory.

In the regression stage, the model was trained using the AdamW optimizer with a learning rate of 0.0001 and a batch size of 256. The loss function combines mean squared error (MSE) loss for speed prediction and cosine loss for learning the direction of heading changes. Since the heading change is represented as a sine/cosine vector, the angle loss is defined using cosine similarity as in (1).

$$\mathcal{L}_\theta = 1 - \frac{\hat{\mathbf{y}}_\theta \cdot \mathbf{y}_\theta}{\|\hat{\mathbf{y}}_\theta\| \|\mathbf{y}_\theta\|}. \quad (1)$$

Here, $\hat{\mathbf{y}}_\theta$ denotes the sine/cosine vector of the heading change predicted by the model, $\mathbf{y}_\theta$ denotes the GT sine/cosine vector generated from the RTK heading, and $\|\hat{\mathbf{y}}_\theta\|$ and $\|\mathbf{y}_\theta\|$ denote their Euclidean norms.

III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed model, test trajectories collected in three smartphone carrying modes, text, swing, and call, were used. For each carrying mode, three test trajectories were collected, resulting in a total of nine test trajectories for evaluation.

Absolute trajectory error (ATE) was used as the evaluation metric to measure the average positional error between the predicted and GT trajectories. ATE is calculated as the mean Euclidean distance between the GT and predicted positions over all trajectory points, as defined in (2).

$$\text{ATE} = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i^{GT} - \hat{x}_i)^2 + (y_i^{GT} - \hat{y}_i)^2}. \quad (2)$$

Here, N denotes the number of trajectory points, (x_i^{GT}, y_i^{GT}) denotes the i -th GT position, and $(\hat{x}_i, \hat{y}_i)$ denotes the i -th predicted position.

Table I presents the ATE results for the text, swing, and call carrying modes. The proposed model reconstructed walking trajectories in all three carrying modes using a single integrated model without separate pose-specific models. These results show that a single transformer encoder-based regression model can estimate walking speed and heading changes across carrying modes with different IMU signal characteristics.

TABLE I
ATE RESULTS FOR EACH SMARTPHONE CARRYING MODE [M]

Mode	Test 1	Test 2	Test 3	Avg.
Text	0.627	0.598	0.923	0.716
Swing	0.582	1.113	1.027	0.907
Call	0.548	0.742	0.460	0.583

In addition to the quantitative evaluation, all nine test trajectories were visualized in Fig. 2 to examine the trajectory reconstruction results across the three carrying modes. As shown in Fig. 2, the predicted trajectories closely followed the overall shape and movement trend of the GT trajectories in the text, swing, and call carrying modes.

IV. CONCLUSION

In this paper, a single PDR model based on the transformer encoder architecture was trained using smartphone IMU data collected in the text, swing, and call carrying modes, and the feasibility of using a single model for trajectory reconstruction across multiple carrying modes was evaluated. The proposed model estimates walking speed and heading changes without manually selecting sensor axes or constructing separate models for each carrying mode, and reconstructs a two-dimensional pedestrian trajectory by accumulating the estimated values. Future work will extend the evaluation to special carrying environments such as pocket carrying and investigate a trajectory reconstruction model that is more robust to accumulated errors and drift in long-distance walking.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-16071705).

REFERENCES

- [1] O. Asraf, F. Shama, and I. Klein, "PDRNet: A deep-learning pedestrian dead reckoning framework," *IEEE Sensors Jour.*, vol. 22, no. 6, pp. 4932–4939, Mar. 2022.
- [2] J. W. Kim, H. J. Jang, D. H. Hwang, and C. Park, "A step, stride and heading determination for the pedestrian navigation system," *Jour. Global Positioning Syst.*, vol. 3, no. 1-2, pp. 273–279, 2004.
- [3] H. Weinberg, "AN-602: Using the ADXL202 in pedometer and personal navigation applications," *Analog Devices Tech. Rep.*, 2002.
- [4] K.-S. Kim and Y. Shin, "Deep learning-based PDR scheme that fuses smartphone sensors and GPS location changes," *IEEE Access*, vol. 9, no. 1, pp. 158616–158631, Dec. 2021.
- [5] X. Li and Y. Wang, "Evaluation of AHRS algorithms for foot-mounted inertial-based indoor navigation systems," *Open Geosci.*, vol. 11, pp. 48–63, Mar. 2019.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Proc. NIPS 2017*, pp. 6000–6010, Long Beach, USA, Dec. 2017.
- [7] R. B. Langley, "NMEA 0183: A GPS receiver interface standard," *GPS World*, vol. 6, no. 7, pp. 54–57, July 1995.

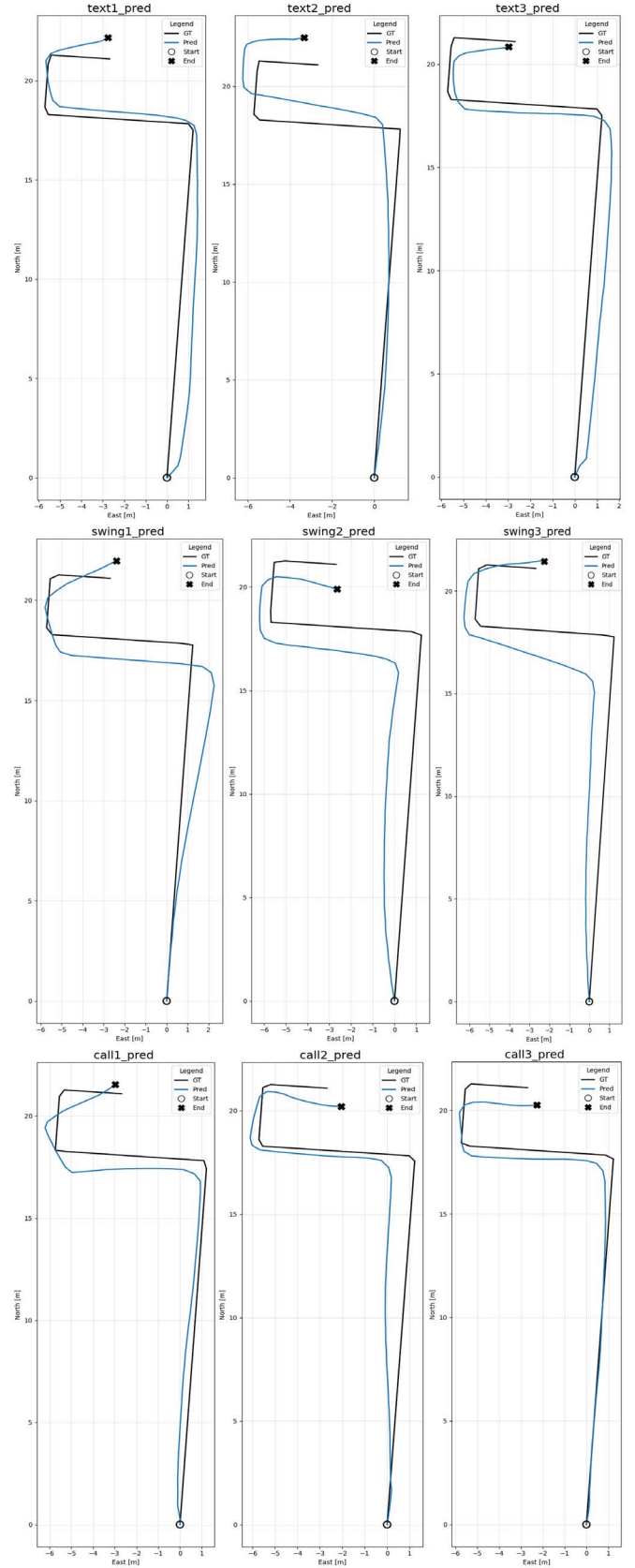


Fig. 2. GT and predicted trajectories for all nine test cases. Rows correspond to text, swing, and call carrying modes.