This article has been accepted for publication in IEEE Sensors Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2023.3294553

Sensor Applications

HINNet + HeadSLAM: Robust inertial navigation with machine learning for long-term stable tracking

Xinyu Hou[1], Jeroen Bergmann[1]

[1] Department of Engineering Science, University of Oxford. Oxford, UK, OX1 3PJ

Manuscript received June 7, 2017; revised June 21, 2017; accepted July 6, 2017. Date of publication July 12, 2017; date of current version July 12, 2017.

Abstract—In recent years, human position tracking with wearable sensors has been rapidly developed and shown great potential for applications within healthcare, smart homes, sports, and emergency services. Unlike tracking researches with sensors on the foot, human positioning studies with head-mounted sensors are fewer and still remain problems that have not been solved. We have proposed two studies solve part of the problems separately: HINNet is able to track people with free head rotations; HeadSLAM allows long-term tracking with stable errors. In this paper, to allow free head rotations meanwhile support long-term tracking, HINNet is combined with HeadSLAM and tested. The result shows that the combination could effectively distinguish head rotations and keep a low and stable position error in long-term tracking, with an absolute trajectory error (ATE) of 2.69m and relative trajectory error (RTE) of 3.52m.

Index Terms—Machine learning , Inertial Navigation, Pedestrian Dead Reckoning, Deep Neural Network , Inertial Measurement Unit, Wearable sensors, SLAM

I. INTRODUCTION

There has been a rapid development in technology and algorithms that allow for real-time human position tracking. The maturation of this technology has brought with it many possibilities that could substantially change our modern way of life. The real-world applicability of monitoring is nonetheless governed by the performance of these systems. However, certain scenarios require robust and accurate information, even when complex environmental constrains are in place. The environment can lead to a range of behavioural responses that influence our motor outcomes [1]. This indicates that people can move in unpredictable ways, as they navigate and interact within their environment. This is particular important to consider when we are exploring solutions for areas such as security, first responders or healthcare. The location of people might need to be tracked accurately as their safety and lives might depend on it. They can e.g. move in and out of buildings with unknown layouts and the monitoring system will need to be able to deal with that. Furthermore, in environmentally complex environments it is unlikely that any infrastructure is either available or operational under those conditions (disaster areas are a good example of this) . Normally additional infrastructure (such as Wi-Fi) can be leveraged for positional tracking, but it should be clear that there is no certainty of this in the aforementioned situation. The solution should thus be infrastructure agnostic.

The infrastructure issues also applies to solutions that are considered "globally" applicable. The best known positioning solution is after all the Global Navigation Satellite System (GNSS). It is widely used in outdoor environments, but requires signals from satellites, which could be blocked or suffer from rapid deterioration, in addition to multi-path effects in indoor environments. Other modalities that rely on more local infrastructure, like leveraging information from Received Signal Strength (RSS) of Wireless Local Area Network (WLAN) [2] / Bluetooth low energy (BLE) beacons [3] / ultra wide band (UWB)[4], radio frequency identification (RFID) [5] and ultrasound[6] are also limited in application for extreme environmental and behavioural conditions. All these rely on external aiding signals, information, or infrastructure, and thus, they are not applicable in scenarios where these signals are severely affected or when there is no signal at all. Inertial navigation does not rely on any infrastructure, which makes it possible to be used flexibly during a wider range of behavioral scenarios. Inertial navigation is now becoming one of the most popular tracking methods, and this is further propelled by the rise of smart devices. It offers a self-contained navigation technique and only requires inertial measurement units (IMUs) to be worn by the user. This offers a great opportunity to create low-cost devices that are more ecological valid under certain complex behavioral and environmental conditions.

According to a recent systematic literature review [7], most human positioning studies put inertial sensors on the foot, only limited papers adopted head-mounted inertial sensors. However, the head is a very suitable location for small, discreet, and unobtrusive sensors, as the whole body is working on stabilising our visual center during motions. Furthermore, there are a lot of everyday objects, like glasses, earphones, mouthguards, hearing aids, helmet, etc, which can be used to integrate these sensors inconspicuously. Recently, we have proposed two tracking systems for head-mounted IMUs: HeadSLAM [8] and HINNet [9].

HeadSLAM was proposed to improve the tracking accuracy during a longer tracking duration. Traditional pedestrian dead reckoning (PDR) methods suffer from error accumulations, because of the lack of calibration methods. HeadSLAM uses estimated trajectories at the earlier stage, which was proved to be more reliable, to calibrate

Corresponding author: Jeroen Bergmann (e-mail: jeroen.bergmann@eng.ox.ac.uk).
Associate Editor: .
Digital Object Identifier 10.1109/LSENS.2017.0000000

estimated trajectories in the later stage that tended to have larger accumulation errors. HeadSLAM could reach an average Root Mean Square Error (RMSE) of 0.34 m indoors and 0.83 m outdoors during 10 min walks in a 20 hour dataset. This showed a significant improvement compared to the PDR method. However, HeadSLAM in the original study still used odometry from a basic PDR method [10], which has two drawbacks; (i) parameters need to be optimized for each individual and (ii) it does not allow free head rotations during walking.

HINNet is a pedestrian inertial navigation system allowing free head movements with head-mounted IMUs by applying a deep neural network (DNN)[9]. It could effectively distinguish head rotations and changes in walking direction. It was shown that it had a relative trajectory error of 5.57m. Although, it solved the problem of head rotations, the estimation errors got larger as the testing time got longer. The underlying reason for this was because there is no efficient calibration taking place.

In this paper, the above two methods are neatly combined to solve both the head rotation problem and the long term estimation error accumulation. The average trajectory error (ATE) and relative trajectory error (RTE) were used as performance measurements. A performance comparison with just the HINNet will be given.

II. METHODS

The whole system is summarized in Figure 1.

A. HINNet

HINNet was proposed to solve the problem of differentiating between head rotations and changes in walking direction by applying the following three procedures.

1) Roll and Pitch compensation: The raw IMU sensor data was first transformed into a normalized coordinate system in which z-axis is aligned with the gravity direction, whilst there is no gravity component on the normalized x-axis and y-axis, by Equation (1) and (3), where a and ω are $3 * length_{data}$ vectors representing the accelerometer data and gyroscope data.

$$a_{norm} = R_a^{-1} \cdot a_{raw} \tag{1}$$

$$R_a = R_x(\phi)R_y(\theta) = \begin{bmatrix} cos\theta & 0 & -sin\theta \\ sin\phi sin\theta & cos\phi & sin\phi cos\theta \\ cos\phi sin\theta & -sin\phi & cos\phi cos\theta \end{bmatrix} \tag{2}$$

$$\omega_{norm} = R_\omega^{-1} \cdot \omega_{raw} \tag{3}$$

$$R_\omega = \begin{bmatrix} 1 & 0 & -sin\theta \\ 0 & cos\phi & sin\phi cos\theta \\ 0 & -sin\phi & cos\phi cos\theta \end{bmatrix} \tag{4}$$

2) Peak ratio feature: Two kinds of obvious body movements in walking generate regular acceleration waves in different directions: (i) body swinging from left to right, (ii) stepping generating linear acceleration front-to-back[11]. Figure 2 shows that when facing forward, the acceleration variation wave from stepping could be fully projected to x axis, while the wave from side swing has no projection to x axis. However, when facing sideways, the acceleration

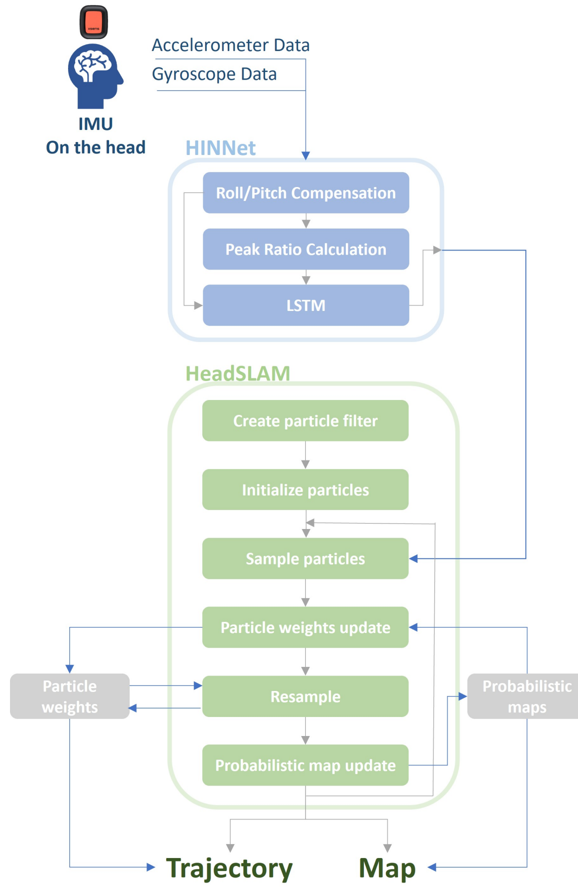

Fig. 1: Overview of the HINNet-HeadSLAM system. HINNet receives raw accelerometer and gyroscope data from the IMU on the head, and output the odometry to HeadSLAM. HeadSLAM calibrates and estimates the final trajectory. Long short-term memory is abbreviated by LSTM.

variation wave from both stepping and swinging will be projected on to x axis, and the magnitudes of these two projections will change when head rotation angle changes. Thus the head rotation could be detected when the magnitudes ratio changed.

One full wave of swing requires two steps but one full "stepping" wave requires one step. The frequency difference makes it possible to distinguish these two motions in the frequency domain. After applying Fast Fourier Transform (FFT) to normalised x-axis acceleration, the first two peaks on the frequency spectrum represent the swings and steps accordingly. The ratio of these two peak values is the peak ratio feature P_{ratio}(Equation 5)). The peak ratios in $2s$ before and after each sample were calculated and recorded as two different input features for the neural network.

$$P_{ratio} = \frac{P_{swing}}{P_{stepping}} \tag{5}$$

3) Deep neural network framework: HINNet uses 2-layer bidirectional long short-term memory (LSTM) [12].

$$(a, \omega, P_{ratio})_{8*60} \xrightarrow{F_\theta} (\Delta l, \Delta\psi)_{2*1} \tag{6}$$

The 8-dimension input is composed of normalised accelerometer data a, gyroscope data ω and two peak ratio features P_{ratio}. And outputs are walking distance Δl and rotation $\Delta\psi$. A window length

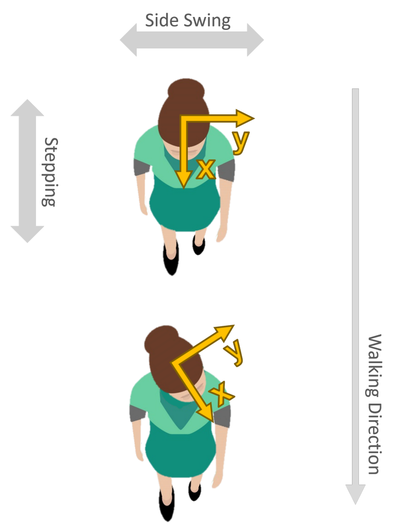

Fig. 2: Different magnitudes of side swing and "stepping "are measured by the x and y axis depending on the orientation of the head.

of 60 frames (1 s) was used in this study to generate a smoother trajectory in HeadSLAM.

B. HeadSLAM

The odometry generated in HINNet is subsequently fed into HeadSLAM for trajectory calibration.

HeadSLAM uses Rao Blackwellized Particle filter (RBPF) from FastSLAM algorithm [13]. The two-dimensional space is first divided into a grid of adjacent hexagons of a given radius. The SLAM problem can be decomposed into a problem of pedestrian localization and mapping conditioned on the pedestrian's position (pose), with a posterior simplified as:

$$p(\mathbf{P}_{0:k}, \mathbf{M}|\mathbf{Z}_{1:k}) = p(\mathbf{M}|\mathbf{P}_{0:k}) \cdot p(\mathbf{P}_{0:k}|\mathbf{Z}_{1:k}) \tag{7}$$

where $\mathbf{P}$ and $\mathbf{M}$ represent the pose and the map, $\mathbf{Z}_k$ is a noisy measurement of the difference between $\mathbf{P}_{k-1}$ and $\mathbf{P}_k$, which is the step vector estimated from the previous PDR layer. The pose could be estimated recursively:

$$p(\mathbf{P}_{0:k}|\mathbf{Z}_{1:k}) \propto p(\mathbf{Z}_k|\mathbf{P}_{k-1:k}) \cdot p(\mathbf{P}_k|\mathbf{P}_{0:k-1}) \cdot p(\mathbf{P}_{0:k-1}|\mathbf{Z}_{1:k-1}) \tag{8}$$

$p(\mathbf{Z}_k|\mathbf{P}_{k-1:k})$ is the likelihood function, which adopts a normal distribution to draw possible poses after each step. The pose transition function $p(\mathbf{P}_k|\mathbf{P}_{0:k-1})$ is computed by marginalizing over the map. Integrating it yields:

$$I^i \propto \frac{N_{\tilde{h}}^{\tilde{e}} + \alpha_{\tilde{h}}^{\tilde{e}}}{N_{\tilde{h}} + \alpha_{\tilde{h}}} \tag{9}$$

where $N_{\tilde{h}}^{\tilde{e}}$ is the number of times the i-th particle crossed edge $\tilde{e}$ (edges of hexagons), $N_{\tilde{h}}$ is the sum of the crossed times of all edges of the hexagon in this particle's map counters. $\alpha_{\tilde{h}}^{\tilde{e}}$ and $\alpha_{\tilde{h}} = \sum_{e=0}^{5} \alpha_{\tilde{h}}^{\tilde{e}}$ are the prior counts. The result is used in the particle weight update:

$$w_k^i \propto w_{k-1}^i \cdot I^i \tag{10}$$

where w_k^i denotes the weight of the i-th particle at step k. If a particle crossed an edge which has been crossed more frequently than the other edges of the last hexagon from which the particle comes from, it tends to have more weight. Thus a consistent walking pattern would be generated.

Each particle contains information about the previous track and the probability of transitions from each hexagon to its adjacent hexagons, which is captured by a probabilistic map. The final result is the best map from the particle with the highest weight.

III. RESULTS

This paper uses data from HINNet [9], in which the IMU data was collected at 60Hz from a head-mounted XSens Dot whilst the groundtruth was collected from a chest-mounted phone which used visual inertial odometry (VIO). It includes 79 datasets with a total time of around 528 minutes. The dataset was collected outside in three different scenarios with different lengths and paths to ensure a greater external validity. The trajectories include straight routes, curves, and turns at different angles, which extends the complexity and applicability of the tests.

Figure 3 shows the results from HINNet with HeadSLAM comparing to original HINNet on three different tracks.

Three metrics were utilized for quantitative analysis:

ATE (m): Absolute trajectory error. ATE is the root mean square error (RMSE) between the whole ground truth trajectory and the estimated trajectory.

RTE (m in Δt): Relative trajectory error. RTE is defined as the average RMSE over a fixed time interval (1 minute in this study) with alignments of the initial states.

RTE and ATE are standard position evaluation metrics in navigation [14].

Distance error rate (%): Drift of the estimated total distance.

RTE, ATE and percentage error of total distance of each method were summarized in Table 1.

Table 1: Relative trajectory error (RTE), absolute trajectory error (ATE), and percentage error of total distances of HINNet and HINNet+HeadSLAM.

Methods	RTE (m)	ATE (m)	Distance Error(%)
HINNet	3.69	7.13	1.15
HINNet + HeadSLAM	3.52	2.69	2.19

IV. CONCLUSION

HINNet solved the problem of the confusion between pure head rotations and walking direction changing when using head-mounted IMUs. HeadSLAM solved the problem of error accumulations in long-term tracking. The combination of them could solve both problems: allowing free head rotations and meanwhile supporting long-term tracking. ATE is the RMSE of the whole trajectory (8 - 12 minutes in this study). RTE could be recognised as ATE in one minute. Both methods have similar RTEs. But the ATE of combined method is significantly smaller than that of original HINNet, which proved the HINNet+HeadSLAM is able to maintain a stable and consistent error in long-term tracking. It has also been proved in Figure 3, the trajectories generated by HINNet are gradually getting farther and farther away from ground truth because of the error accumulations and the lack of calibration. However, the trajectories estimated by HINNet+HeadSLAM still keep close to the ground truth as time goes on.

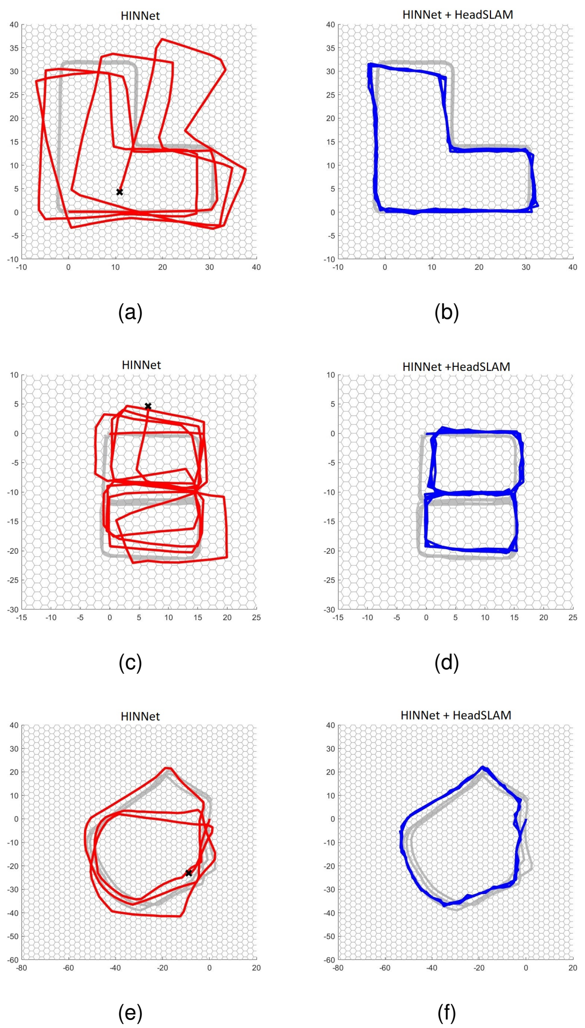

Fig. 3: Estimated trajectories in meters. Light gray lines are ground truth. Red lines in (a), (c), (e) represent trajectories generated from HINNet only. Blue lines in (b), (d), (f) shows the results from the combination of HINNet and HeadSLAM.

Although the proposed method has the advantage from both HINNet and HeadSLAM, it also obtained some of the limitations from them. It could differentiate pure head rotations from walking direction changing as long as they do not happen at the same time. Overlaps of head and body rotations lead to a confusion in heading estimation and information from other sensors or new features are needed to maintain a correct heading direction. It should also be noted that, just like HeadSLAM, the effectiveness of this combined method only exists when walking repeatedly on a restricted, predefined path, such as walking through indoor corridors or outdoor tracks for several laps. This is because the calibration of HeadSLAM depends on the probability map which is updated in overlaps. Thus the combined method is applicable for scenarios in which people cover the same path multiple times.

Besides the above limitations, there are also other possible future research directions. In real world scenarios, people not just rotate their heads and walk with constant pace, they may run, jump, turn, slide or stumble in daily activities. Users will have different body data and movement pattern. Sensors may also have different accuracy or other parameters. If a tracking system seek extensive use in daily real scenarios, datasets with a larger scale and variety should be essential for the generalization and robustness of the system.

ACKNOWLEDGMENT

This research was partly financially supported by the EPSRC Impact Acceleration AccountAward EP/R511742/1.

REFERENCES

[1] J. H. Bergmann, P. M. Langdon, R. E. Mayagoitia, and N. Howard, "Exploring the use of sensors to measure behavioral interactions: an experimental evaluation of using hand trajectories," *PloS one*, vol. 9, no. 2, p. e88080, 2014.

[2] B. Ferris, D. Fox, and N. D. Lawrence, "Wifi-slam using gaussian process latent variable models." in *IJCAI*, vol. 7, no. 1, 2007, pp. 2480–2485.

[3] R. K. Yadav, B. Bhattarai, H.-S. Gang, and J.-Y. Pyun, "Trusted k nearest bayesian estimation for indoor positioning system," *IEEE Access*, vol. 7, pp. 51 484–51 498, 2019.

[4] J. A. Corrales, F. A. Candelas, and F. Torres, "Hybrid tracking of human operators using imu/uwb data fusion by a kalman filter," in *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2008, pp. 193–200.

[5] J. Huang, X. Yu, Y. Wang, and X. Xiao, "An integrated wireless wearable sensor system for posture recognition and indoor localization," *Sensors*, vol. 16, no. 11, p. 1825, 2016.

[6] G. Zizzo and L. Ren, "Position tracking during human walking using an integrated wearable sensing system," *Sensors*, vol. 17, no. 12, p. 2866, 2017.

[7] X. Hou and J. Bergmann, "Pedestrian dead reckoning with wearable sensors: A systematic review," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 143–152, 2020.

[8] ——, "Headslam: pedestrian slam with head-mounted sensors," *Sensors*, vol. 22, no. 4, p. 1593, 2022.

[9] ——, "Hinnet: inertial navigation with head-mounted sensors using a neural network," *Engineering Applications of Artificial Intelligence*, 2023.

[10] ——, "A pedestrian dead reckoning method for head-mounted sensors," *Sensors*, vol. 20, no. 21, p. 6349, 2020.

[11] J. Windau and L. Itti, "Walking compass with head-mounted imu sensor," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 5542–5547.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit *et al.*, "Fastslam: A factored solution to the simultaneous localization and mapping problem," *Aaai/iaai*, vol. 593598, 2002.

[14] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7244–7251.