NIN-DSO:Neural Inertial Navigation Aided Direct Sparse Visual-Inertial Odometry

Yilin Zhao[1], Yuhang Gao[1], Kun Wu[1] and Long Zhao[1,*]

[1] *School of Automation Science and Electrical Engineering, Beihang University, 37 Xueyuan Road, Beijing, 100191, China*

Abstract

In indoor localization, visual-inertial odometry (VIO) is widely used due to its low cost and effective environmental perception capabilities. However, feature based VIO performs poorly in indoor pedestrian localization challenges. The feature points are prone to occlusion or tracking loss during complex pedestrian movements, and IMU is difficult to perform dead reckoning due to irregular or excessively large anomalous measurements. To address these issues, this paper proposes a VIO approach using the direct method assisted by a neural inertial network. The proposed method replaces the feature based method in front-end in VIO with the direct method to mitigate issues related to occlusion and tracking loss of feature points. It utilizes a neural inertial network to supply initial values for pixel tracking within the direct method and and integrates dead reckoning results as constraints during position optimization. Experimental results demonstrate that the method proposed in this paper exhibits higher positioning accuracy and robustness compared to existing methods in indoor pedestrian localization scenarios.

Keywords

Indoor pedestrian positioning, Neural inertial navigation, Direct method, Visual-inertial odometry

1. Introduction

As a component of navigation, positioning, and timing technologies, indoor positioning technology is widely applied in fields such as indoor robot navigation, augmented reality, the Internet of Things, and indoor rescue. However, unlike outdoor positioning, GNSS signals are unavailable in indoor scenarios, making it more difficult to obtain accurate and robust positioning results. Most indoor positioning solutions typically use Bluetooth, UWB, or WiFi [1, 2] to replace satellites for absolute positioning which require external signals from the carrier. However, the accuracy of these methods depends on the quality of the signal and the location of the base station. Additionally, some indoor localization methods utilize LiDAR [3, 4] or cameras [5, 6] for more accurate relative localization, among which VIO is gaining attention and undergoing rapid development due to its affordability and effective environmental perception capabilities.

However, VIO does not perform optimally in indoor pedestrian localization. Most VIO systems currently employ feature points in visual processing to recover the relative position transformation between image frames. Nevertheless, when pedestrians carry the camera in motion, large changes in the image field of view or occlusions can lead to feature point loss, resulting in system failure. In the inertial component, consumer-grade IMUs are plagued by biases, noise, and thermal drift, resulting in significant measurement errors. Additionally, due to the highly irregular nature of pedestrian movement, position estimation based on kinematics or gait tracking is also suboptimal.

Therefore, to address the problems in indoor pedestrian localization, we propose a VIO system assisted by a neural inertial network to enhance the accuracy and robustness. The method employs a direct method within the VIO framework based on DM-VIO [7], which directly uses sparse pixel gray scale changes to determine the relative position transformation between frames. Directly processing pixels can avoid the problem of feature point loss during indoor pedestrian localization, but its initialization also requires more accurate IMU data. On this basis, we train and deploy a neural inertial network

Proceedings of the Work-in-Progress Papers at the 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN-WiP 2024)

[*] Corresponding author.

zyl314@buaa.edu.cn (Y. Zhao); BY2003125@buaa.edu.cn (Y. Gao); wukun0627@buaa.edu.cn (K. Wu); buaa_dnc@buaa.edu.cn (L. Zhao)

[8] to improve the accuracy of IMU-based position estimation in indoor pedestrian localization and supplements the direct VIO with the prediction results of the network. The main contributions of this work are as follows:

- The prediction results of the neural inertial network are utilized to assist in the initialization between frames for the direct method, thereby improving the accuracy of inter-frame tracking.
- The proposed method is the first direct VIO aided by neural inertial navigation. The method uses graph optimization to tightly couple direct VIO with the neural network navigation, thereby enhancing the solution accuracy and robustness of the system.

2. Related Work

2.1. Visual-Inertial Odometry

Visual-inertial navigation systems (VINS) are typically derived from visual SLAM methods integrated with inertial sensors. The incorporation of inertial measurement data introduces stable scale information to SLAM, thereby effectively enhancing the accuracy of image inter-frame matching. Most current methods achieve tightly coupled VIO by fusing raw image and IMU data. The most classic approach is the Multi-State Constraint Kalman Filter (MSCKF) [9], which uses an Extended Kalman Filter (EKF) to add multiple camera poses to the state vector and employs a sliding window to maintain constraints between image frames and the IMU for position solving. Some systems achieve data fusion through optimization methods, with most utilizing feature point matching as constraints. The two most classic methods are ORB-SLAM [6, 10] and VINS-Mono [5]. ORB-SLAM2 [6] uses FAST corners and BRIEF descriptors to form ORB features, implementing monocular VO by using feature point reprojection error as the cost function. In VINS-Mono [5], the algorithm employs Harris corner points as feature points and utilizes the L-K optical flow method to track the feature points, establishing visual part constraints. Additionally, the algorithm adopts a marginalization strategy that introduces a priori marginalization error to improve the computational efficiency and robustness.

With the improvement in computational performance, VO methods can process more pixel data in real-time, leading to the emergence of direct VIO methods, which differ from feature point based approaches. Direct methods operate directly on the image pixels captured by the camera, thus eliminating the need for stable extraction and matching of feature points in the environment. The Direct Sparse Odometry (DSO) algorithm [11] integrates pixel association, photometric error optimization, position estimation, and sparse point cloud generation into a unified nonlinear optimization problem, discarding the traditional separation between front-end and back-end in VO. Building on this, the VI-DSO [12] algorithm combines the DSO with IMU data, while DM-VIO [7] incorporates a delayed marginalization strategy into VI-DSO to enhance the robustness and accuracy of position estimation in direct method based VIO. However, since these methods track pixels solely through photometric invariance, they require better initial values of relative position to achieve stable and accurate position estimation.

2.2. Neural Inertial Navigation

The emergence of neural inertial network methods enables deep analysis of IMU measurement data, extracting underlying motion patterns in complex movements. The RIDI [13] algorithm uses a Support Vector Machine (SVM) model to classify the motion states of pedestrians and a Support Vector Regression (SVR) model to regress the IMU data, thereby correcting low-frequency errors in IMU measurements. IONet [14] treats inertial localization as a time series learning problem, utilizing Long Short Term Memory (LSTM) to process inertial data within a time window to estimate the carrier's position and trajectory in a polar coordinate system. Ronin [8], based on ResNet, LSTM, and Temporal Convolutional Network (TCN) architectures, directly estimates positions through a neural inertial network. However, the relative position estimation of these networks lacks precise, necessitating the addition of accurate constraints.

2.3. Neural Inertial Navigation Aided VIO

In VIO methods, the processing of IMU data still adopts the method of integration using kinematic models in traditional navigation methods. In order to decrease the computational load and effectively improve the accuracy of inter-frame relative position estimation, the integration is typically transformed into an inter-frame pre-integration that remains unaffected by the initial state. However, these kinematic models perform poorly in complex motion scenarios, as irregular or excessive IMU measurements can cause the system state to gradually diverge.

Some algorithms have incorporated position prediction results from inertial neural networks into combined navigation systems, but fewer of them add inertial neural networks to VIO. TLIO [15] uses a neural network to learn the 3D displacement transformation and covariance directly from a time series of IMU data, then applies the results obtained from the network predictions to a Kalman filter to correct state quantities such as direction, velocity, position, and IMU deviation. RNIN-VIO [16] firstly utilizes a neural inertial navigation to assist feature point based VIO, incorporating network predictions into the optimization model in the back-end, to enhance the accuracy and robustness of position estimation.

However, it is worth noting that all of the above methods only involve the neural network inference results as additional constraints added to the feature point based VIO. This structure leads to neural network inference results with less impact on such systems. In contrast, the direct methods, although more accurate, require higher IMU data quality. Bad IMU data will cause the direct method to fail to initialize properly and have poorer results, which has a stronger dependence on IMU data.

Therefore, we implement a neural inertial navigation aided method VIO. The proposed method is the first method to combine inertial neural networks with direct method VIO, and it combines inertial neural networks more tightly than the existing methods. The method effectively improves the accuracy and robustness of the direct VIO initialization and solution.

3. Method

3.1. Overview

The proposed method in this paper is a direct visual-inertial odometry, which improves upon the DM-VIO framework, as illustrated in Fig. 1. After the visual part completes initialization and the neural inertial navigation accumulates sufficient IMU measurement data, the network will estimate the position of carrier. The method utilizes the network's predicion results to set an initial relative position between image frames, providing a better initial photometric error for the visual part. In the joint optimization phase, the prediction results are also integrated into the factor graph as a constraint for position estimation. Compared to DM-VIO, the improvements of our method are highlighted in red boxes in Fig. 1.

3.2. DM-VIO

In direct VO [11], the initial relative position is determined by kinematic modeling. Specifically, various photometric errors are calculated based on assumptions such as stationary, constant velocity, or constant acceleration motions. The minimum among these errors is selected as the initial value for iterative photometric error minimization, with the corresponding relative position used as the initial position.

In DM-VIO [7], the introduction of IMU measurements enables the derivation of a coarse relative position estimate through pre-integration. The relationship serves as the initial value, and the photometric error calculated from this value is used as the initial error for iteration, as shown in (1).

$$E^{\boldsymbol{p}}_{IMU} = \sum_{\boldsymbol{p}\in N_{\boldsymbol{p}}} \omega_{\boldsymbol{p}} ||(I_j\left[\boldsymbol{p}'_{IMU}\right] - b_j) - (I_i\left[\boldsymbol{p}\right] - b_i)t_j e^{a_j}/t_i e^{a_i}|| \tag{1}$$

where $N_{\boldsymbol{p}}$ represents a small neighborhood around point $\boldsymbol{p}$. I is the image frame. i and j are the sequence numbers of the image frames. t represents the exposure time. a and b are the factors for

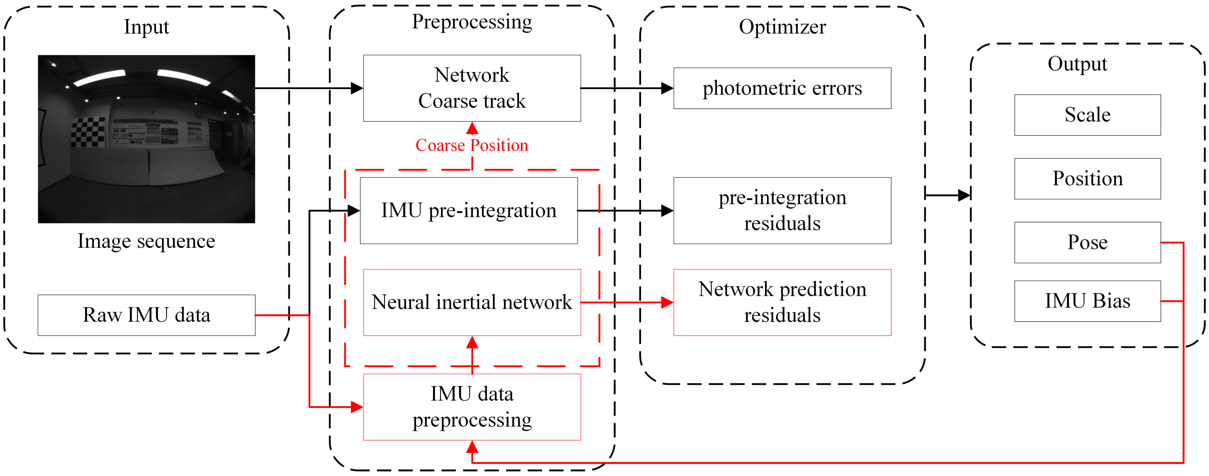


Figure 1: Structure of NIN-DSO.

correcting the affine brightness transformation. $\omega_{\boldsymbol{p}}$ is the weight related to the gradient. $\boldsymbol{p}'_{IMU}$ is the projected point obtained by pre-integration results.

DM-VIO incorporates photometric errors and IMU residuals for optimization process. And its energy function is

$$E_{total} = \lambda E_{photo} + E_{inertial} \tag{2}$$

which consists of the photometric error E_{photo}, pre-integration residuals $E_{inertial}$. And λ is the weight of the photometric error.

3.3. Neural Inertial Network

We utilize the neural inertial network and proposed in [16] to supplement the visual-inertial odometry. And the data used for network training was also taken from [16]. The network comprises a ResNet18 structure, an LSTM structure, and fully connected layers in a cascade. The ResNet18 structure primarily facilitates supervised learning of pedestrian motion patterns and estimation of position changes. The LSTM structure analyzes the hidden states of pedestrian over a period of time, which may implicitly influence the current state, to avoid the divergence problem in dead reckoning due to abnormal IMU measurements. During the network operation, the IMU data must be transformed from the IMU coordinate system to the VIO coordinate system, with gravity and bias effects removed. After preprocessing the IMU measurements, they serve as the input to the network, enabling the direct acquisition of relative position estimation of pedestrian and their corresponding covariance, as shown in (3).

$$\begin{aligned}\left(\Delta \boldsymbol{t}_{v_i}^{\prime v_j}, \boldsymbol{\sigma}_{v_i}^{\prime v_j}\right) &= f\left((\boldsymbol{a}_{n-N}^{v}, \boldsymbol{w}_{n-N}^{v}), \ldots, (\boldsymbol{a}_{n}^{v}, \boldsymbol{w}_{n}^{v}), \boldsymbol{h}_{n-N}\right) \\ \boldsymbol{a}_n^v &= \boldsymbol{R}_n^v\left(\boldsymbol{a} - \boldsymbol{b}_a\right) - \boldsymbol{g}^{\boldsymbol{v}} \\ \boldsymbol{w}_n^v &= \boldsymbol{R}_n^v\left(\boldsymbol{w} - \boldsymbol{b}_g\right)\end{aligned} \tag{3}$$

where $f(\bullet)$ is the function obtained by neural network fitting. $\boldsymbol{a}$ and $\boldsymbol{w}$ are the raw acceleration and angular velocity measured by the IMU sensors. $\boldsymbol{b}_a$ and $\boldsymbol{b}_g$ are the bias obtained from the VIO system. $\boldsymbol{g}^{\boldsymbol{v}}$ is the gravity vector. v is the VIO coordinate system while n is the IMU coordinate system. $\boldsymbol{a}_n^v$ and $\boldsymbol{w}_n^v$ are the acceleration and angular velocity at the nth IMU time step in VIO coordinate system, respectively. $\boldsymbol{h}$ is the hidden state produced by LSTM at the last time step. The relative displacement $\Delta \boldsymbol{t}_{v_i}^{\prime v_j}$ and its covariance $\boldsymbol{\sigma}_{v_i}^{\prime v_j}$ are the outputs of the network.

3.4. NIN-DSO

3.4.1. Time Synchronization

To ensure that the predicion results of network facilitate effective convergence of the VIO position estimation, we bundle these predictions to the image inputs for approximate temporal synchronization. Indeed, predictions from the neural inertial network resemble IMU pre-integration, serving as a type of constraint between images. Therefore, the proposed method adjusts the prediction output frequency to match the images input frequency. Additionally, in order to minimize the temporal discrepancy between the images and the predition results, we interpolate the network inference results at the image input moments. The temporal relationship of IMU data, image data, and network prediction results is illustrated in Fig. 2.

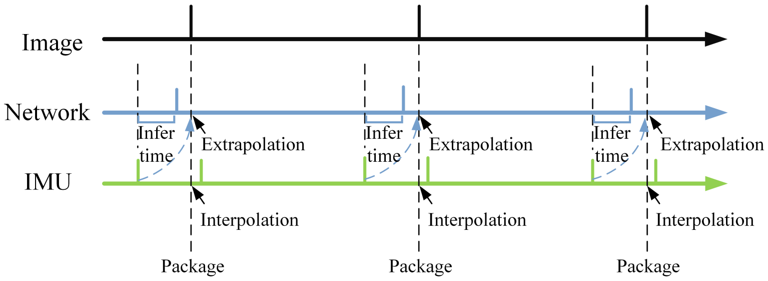


Figure 2: Temporal relationship of tree types of data.

3.4.2. Coarse Tracking

In this paper, we use the relative position transformation predicted by neural inertial networks as an alternative for initializing the photometric error, providing greater robustness than IMU pre-integration, as shown in (4).

$$E^{\boldsymbol{p}}_{Network} = \sum_{\boldsymbol{p}\in N_{\boldsymbol{p}}} \omega_{\boldsymbol{p}} ||(I_j\left[\boldsymbol{p}'_{Network}\right] - b_j) - (I_i\left[\boldsymbol{p}\right] - b_i) t_j e^{a_j}/t_i e^{a_i}|| \tag{4}$$

where $\boldsymbol{p}'_{Network}$ is the projected point obtained by prediction results of network. And the remaining items are consistent with (1).

3.4.3. Visual-Inertial Optimization

The VIO proposed in this paper achieves position estimation by minimizing the energy function , which comprises photometric error, pre-integration residuals, and network prediction position residuals. Unlike the feature point based method, our approach simultaneously optimizes the position, sensor error, scale and 3D structure, resulting in higher accuracy. The factor graph for the proposed method is shown in Fig. 3.

Based on DM-VIO, the method further incorporates the position predictions from the neural inertial network as additional constraints. And its energy function is

$$E_{total} = \lambda E_{photo} + E_{inertial} + E_{network} \tag{5}$$

which consists network prediction position residuals $E_{network}$. The photometric error is the sum of the individual pixel photometric erros, and can be determined by (6).

$$E_{photo} = \sum_{i\in F} \sum_{\boldsymbol{p}\in \boldsymbol{P}_i} \sum_{\boldsymbol{p}_k \in key(\boldsymbol{p})} E^{\boldsymbol{p}_k}_{Network} \tag{6}$$

where F is the set of keyframes i. $\boldsymbol{P}_i$ is the set of points in keyframes. And $key(\boldsymbol{p})$ is the set of points that are jointly observed in keyframes.

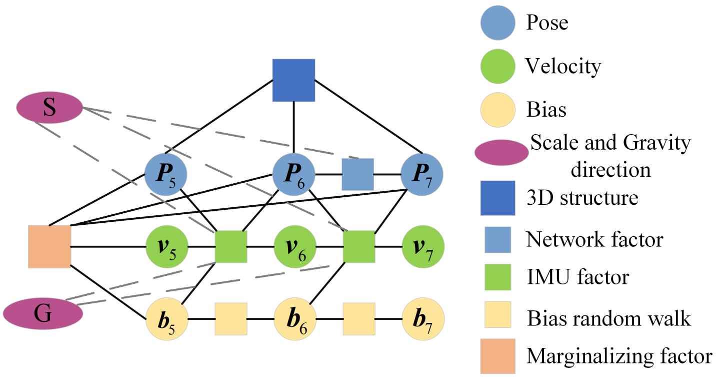


Figure 3: Factor graph of NIN-DSO.

The pre-integration residuals are

$$E_{inertial} = \left(\boldsymbol{s}^j_{IMU} \odot \hat{\boldsymbol{s}}^j\right)^T * \sum\nolimits^{-1}_{\boldsymbol{s}_{IMU},j} * \left(\boldsymbol{s}^j_{IMU} \odot \hat{\boldsymbol{s}}^j\right) \tag{7}$$

where $\boldsymbol{s}^j_{IMU}$ is the state obtained from pre-integration. $\hat{\boldsymbol{s}}^j$ is the estimated state of VIO. $\sum_{\boldsymbol{s}_{IMU},j}$ is the corresponding covariance of the state. And $\odot$ is the subtraction operation in Lie algebra.

The IMU measurements are preprocessed before being input into the network in our method. As a result, the relative position transformation predicted by the network remains consistent with the VIO coordinate system. Based on this characteristic, in the part of neural inertial navigation, we use the predictions as direct constraints between image frames rather than the constraints in [16], as shown in (8).

$$E_{Network} = \left(\boldsymbol{s}^j_{Network} \odot \hat{\boldsymbol{s}}^j\right)^T * \sum\nolimits^{-1}_{\boldsymbol{s}_{Network},j} * \left(\boldsymbol{s}^j_{Network} \odot \hat{\boldsymbol{s}}^j\right) \tag{8}$$

where $\boldsymbol{s}^j_{Network}$ is the state obtained from Neural inertial network prediction results. $\sum_{\boldsymbol{s}_{Network},j}$ is the corresponding covariance of the state.

4. Experiments and Results

This section presents the test results on various datasets to demonstrate the effectiveness of our method. All the experiments were conducted on an Intel Core I5-10510U CPU@1.80GHz×8 computer equipped with Ubuntu 20.04 and NVIDIA GeForce RTX3060 12GB graphics card.

4.1. TUM-VI Dataset Experiments

In this subsection, we evaluate the performance of the proposed method using the TUM-VI dataset [17]. The dataset is a widely used public dataset collected by a pedestrian. However, since only the room scenario offers full ground truth, we present only the experimental results for this data sequence. The absolute trajectory error (ATE) in this paper were obtained from the EVO toolbox [18], but the APE values are larger than ATE reported in [17] because we did not use EVO to fix the scale of the algorithm's results. The comparison of ATE between VINS-Mono, NIN-VINS [19], DM-VIO, and our proposed method NIN-DSO is shown in TABLE 1 below.

From the data in TABLE 1, it is evident that the position estimation of direct method based VIO outperform feature point based method, regardless of whether an neural inertial network is added. This indicates that direct method VIO is more suitable for indoor pedestrian positioning scenarios. Our method achieved the best results across all data in the room sequence, reducing the average ATE by 60.2% compared to DM-VIO. Notably, in the fourth data set, the ATE was reduced by 84.2%, demonstrating that the addition of neural network predictions effectively addresses the issue of DM-VIO's inability to track and converge effectively when IMU data quality is poor.

Table 1
ATE in TUM-VI.(Unit:m)

Datasets	VINS-Mono	NIN-VINS	DM-VIO	NIN-DSO
Room1	1.52	1.56	0.57	**0.26**
Room2	1.53	1.52	0.44	**0.32**
Room3	1.83	1.80	0.48	**0.15**
Room4	1.61	1.62	1.27	**0.20**
Room5	1.80	1.83	0.14	**0.12**
Room6	1.48	1.51	0.22	**0.19**
Average	1.63	1.64	0.52	**0.21**

4.2. Real Experiment

Compared to the room sequence of the TUM-VI dataset, indoor pedestrian positioning typically involves faster movement speeds, more complex lighting environments, and larger movement scales. Additionally, VIO tends to accumulate significant errors over long periods of operation. Therefore, we collected a custom dataset to supplement the experiments. This dataset was collected using a Flir camera and a Livox Avia LiDAR as the acquisition sensors, with IMU measurements provided by the Avia. The frequency of image data is 10Hz and the resolution is 1440×1080. The frequency of IMU data is 200Hz. And the two sensors are synchronized by hardware trigger. The pseudo ground truth was obtained from a higher precision Lidar-Visual-Inertial Navigation System to evaluate the performance of the method. In this dataset, we carried the device around the interior of the New Main Building at Beihang University. The total distance covered is 513.56 meters, and the duration of the data collection lasts 368.41 seconds. The data contains typical pedestrian positioning characteristics, and an example of large attitude change and complex lighting environment from the data is shown in Fig. 4.

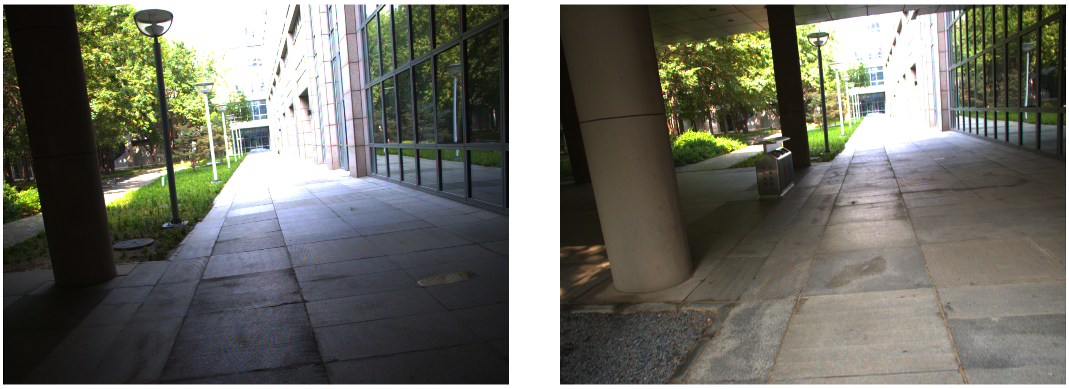


Figure 4: The same example scene under different exposures.

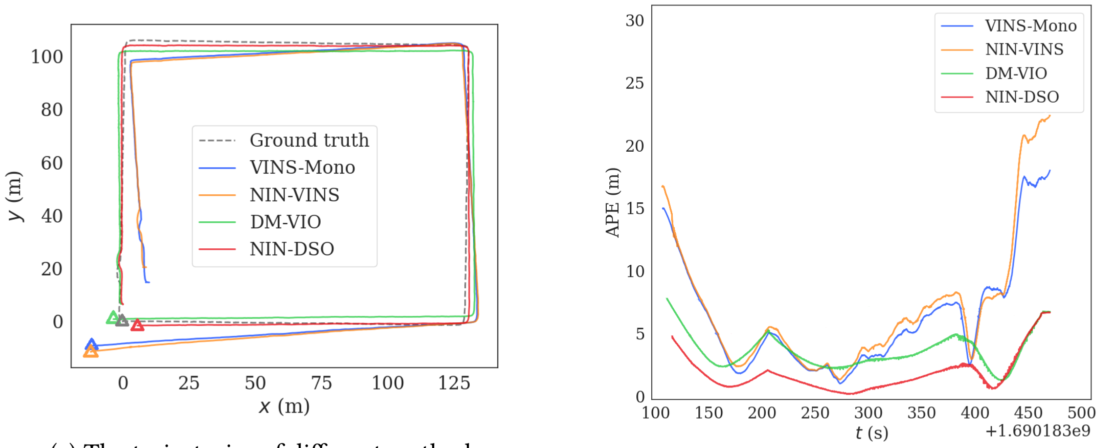
As with the experiments in the first subsection, the performance of the algorithms is still evaluated using ATE. The comparison of the results among VINS-Mono, NIN-VINS, DM-VIO, and NIN-DSO are shown in TABLE 2.

Table 2
ATE in custom data.(Unit:m)

Datasets	VINS-Mono	NIN-VINS	DM-VIO	NIN-DSO
Custom	6.22	6.98	3.58	**1.77**

From the TABLE 2, it can be seen that our method still performs well on this custom data, achieving a 50.7% reduction in ATE relative to the best method among the other three. The same conclusion can be drawn from experiments on the public dataset. Direct methods are more suitable for indoor pedestrian localization than feature point based methods. Additionally, incorporating prediction of neural inertial network can effectively address issues where the system fails to track and converge properly, mitigate cumulative errors during long term positioning, and consequently improve the accuracy and robustness of VIO.

The estimated trajectories and ATE curves for different method on this dataset are shown in Fig. 5, wiht the starting points of these trajectories marked with triangles. As can be seen in Fig. 5, the accuracy using the direct method is higher than the feature point method. In the case of starting positions with similar errors, our method has better scaling in the middle of the trajectory compared to DM-VIO.



(a) The trajectories of different method.

(b) The ATE curve of different method.

Figure 5: The result of different methods in real experiment.

5. Conclusion

This paper proposes a direct visual-inertial odometry with the assistance of a neural inertial network. The neural network is implemented and trained using PyTorch, and deployed within the VIO framework utilizing ONNX and TensorRT. The inclusion of neural inertial navigation provides better initial values for the image inter-frame tracking in direct method based VIO, leading to a more stable visual initialization process than DSO and DM-VIO. Furthermore, by incorporating the network's position prediction results into the energy function of the optimization model, in addition to photometric error and IMU residuals, the system reduces its reliance on the visual component in complex motion environments and mitigates errors and divergence of IMU data. Experimental results demonstrate that the proposed neural inertial navigation aided direct VIO achieves more accurate and robust position estimation in indoor pedestrian movement scenarios.

Although the proposed method does not require stable feature point extraction and matching as feature based methods, it relies on visual component as the foundation for system calculations, rendering it unable to estimate positions when the visual component fails. The current supplementation using the neural inertial network does not fundamentally solve the problem. We plan to further leverage the powerful data processing capabilities of neural networks to evaluate visual inter-frame position estimation results using IMU data. By using this approach, we aim to reduce the reliance of the direct VIO on the visual component, enabling the system to continue functioning normally even when the visual component fails.

Acknowledgments

The project is supported by the National Natural Science Foundation of China (Grant No. 42274037), the Aeronautical Science Foundation of China (Grant No. 2022Z022051001), and the National Key Research and Development Program of China (Grant No. 2020YFB0505804).

References

[1] Y. Tao, B. Huang, R. Yan, L. Zhao, W. Wang, Cbwf: A lightweight circular-boundary-based wifi fingerprinting localization system, IEEE Internet of Things Journal 11 (2024) 11508–11523. doi:10.1109/JIOT.2023.3329825.

[2] Y. Tao, L. Zhao, A Novel System for WiFi Radio Map Automatic Adaptation and Indoor Positioning 67 (2018) 10683–10692. doi:10.1109/TVT.2018.2867065.

[3] T. Shan, B. Englot, Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 4758–4765.

[4] W. Xu, F. Zhang, Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter, IEEE Robotics and Automation Letters 6 (2021) 3317–3324.

[5] T. Qin, P. Li, S. Shen, Vins-mono: A robust and versatile monocular visual-inertial state estimator, IEEE Transactions on Robotics 34 (2018) 1004–1020.

[6] R. Mur-Artal, J. D. Tardós, Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras, IEEE transactions on robotics 33 (2017) 1255–1262.

[7] L. Von Stumberg, D. Cremers, Dm-vio: Delayed marginalization visual-inertial odometry, IEEE Robotics and Automation Letters 7 (2022) 1408–1415.

[8] S. Herath, H. Yan, Y. Furukawa, Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods, in: 2020 IEEE international conference on robotics and automation (ICRA), IEEE, 2020, pp. 3146–3152.

[9] A. I. Mourikis, S. I. Roumeliotis, A multi-state constraint kalman filter for vision-aided inertial navigation, in: Proceedings 2007 IEEE international conference on robotics and automation, IEEE, 2007, pp. 3565–3572.

[10] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, J. D. Tardós, Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam, IEEE Transactions on Robotics 37 (2021) 1874–1890.

[11] J. Engel, V. Koltun, D. Cremers, Direct sparse odometry, IEEE transactions on pattern analysis and machine intelligence 40 (2017) 611–625.

[12] L. Von Stumberg, V. Usenko, D. Cremers, Direct sparse visual-inertial odometry using dynamic marginalization, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 2510–2517.

[13] H. Yan, Q. Shan, Y. Furukawa, Ridi: Robust imu double integration, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 621–636.

[14] C. Chen, X. Lu, A. Markham, N. Trigoni, Ionet: Learning to cure the curse of drift in inertial odometry, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.

[15] W. Liu, D. Caruso, E. Ilg, J. Dong, A. I. Mourikis, K. Daniilidis, V. Kumar, J. Engel, Tlio: Tight learned inertial odometry, IEEE Robotics and Automation Letters 5 (2020) 5653–5660.

[16] D. Chen, N. Wang, R. Xu, W. Xie, H. Bao, G. Zhang, Rnin-vio: Robust neural inertial navigation aided visual-inertial odometry in challenging scenes, in: 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, 2021, pp. 275–283.

[17] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stuckler, D. Cremers, The tum vi benchmark for evaluating visual-inertial odometry, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 1680–1687.

[18] M. Grupp, evo: Python package for the evaluation of odometry and slam., https://github.com/MichaelGrupp/evo, 2017.

[19] Y. Gao, L. Zhao, Nin-vins: Neural inertial navigation aided visual-inertial system for pedestrian dead reckoning, in: International Conference on Guidance, Navigation and Control, Springer, 2022, pp. 6868–6878.