2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Field-VIO: Stereo Visual-Inertial Odometry Based on Quantitative Windows in Agricultural Open Fields

Jianjing Sun[1,2], Shuang Wu[2], Jun Dong[2] and Junming He[2]

***Abstract*— In agricultural open fields, accurate autonomous localization of robots requires long-term data correlation to reduce cumulative error. Our article presents a Stereo Visual-Inertial Odometry (VIO) system based on ORB-SLAM3 to address the malfunction of the Loop Closure Detection (LCD) methods in this environment. In this method, we first propose a concept of quantitative windows to describe the robot's trajectory along the crop rows. We design a driving state quantification algorithm and accurately separate the quantitative windows between the crop rows. Our system constructs spatial constraints according to the parallelism between the quantitative windows. We apply an anomaly correction method to maintain the constructed parallel matching relationship and implement holistic pose correction for keyframes within abnormal quantitative windows. Our system demonstrated excellent performance over long distances in experiments on the Rosario dataset, verifying its effectiveness in reducing cumulative positioning error in agricultural open fields.**

I. INTRODUCTION

In the realm of autonomous robotics for agriculture, Visual-Inertial Odometry (VIO) systems are garnering increasing attention with their rich environmental information input and cost-effectiveness [1]. However, owing to the homogeneity, illumination variations, bumpy roads, and unstable features in agricultural scenarios, VIO systems struggle to complete accurate autonomous positioning. Several VIO systems based on optical flow or direct methods, such as VINS-Fusion [2], SVO 2.0 [3], ROVIO [4], have exposed severe performance degradation in recent experimental evaluation [5]. The difficulty of supporting the photometric invariance assumption is one of the primary reasons. In feature-based VIO systems, feature-matching strategies based on descriptors exhibit more robust performance in environments with changing lighting. Nevertheless, faced with typical challenges in the open fields, feature-based Visual Odometry (VO) might risk tracking failure. The system's recovery strategies, such as relocalization and multi-map system [6], are almost ineffective in this environments. A tightly coupled strategy of Inertial Measurement Units (IMU) and visual sensors can enhance the system's robustness.

Some improved methods [7], [8], [9] propose specialized recovery strategies targeting challenges in agricultural scenarios. However, the motion assumptions of robots these recovery strategies rely on can only cover part of the driving process in the fields. Triggering recovery strategies may pose risks in special driving stages, like the U-turn at the end of crop rows. Hence, it is significant to accurately judge the robots' driving state by combining extra clues. Precise driving state description provides valuable references for subsequent trajectory optimization.

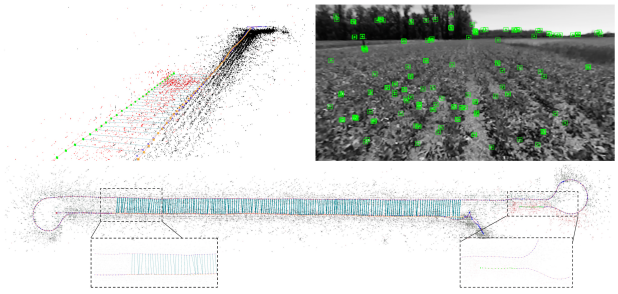


Fig. 1: Output of Field-VIO on the Rosario dataset [10]: Trajectory (red curve), current quantitative window (green chain), reference quantitative window (yellow chain) and parallel matching relationship (cyan line).

To alleviate the cumulative error of the odometry, incremental VO/VIO methods based on features often adopt Loop Closure Detection (LCD) strategy to establish the long-term correlation of pose and observation [11], [12], [13]. However, in the scenarios of homogenization, this traditional algorithms based on Bag of Words (BoWs) often fails. Furthermore, when a robot traverses along the crop rows and passes a consequential location for loop closure again, the viewing angle of scene observation is often thoroughly opposite. Such driving habits will further impede the application of LCD strategy. To our knowledge, there is currently no effective method to alleviate the cumulative error of VO/VIO in open field scenarios.

Following the above analysis, our paper presents a Stereo VIO system based on ORB-SLAM3 [14]. Fig. 1 displays the visual effects of our system. The primary contributions of our work are as follows:

- To our knowledge, our Stereo VIO system is the first attempt to reduce the cumulative error of the VIO system in agricultural open fields.
- We abstract the robot's trajectory along the crop rows by quantitative windows and present a driving state quantification algorithm for separating the quantitative windows accurately.
- We present a construction method of spatial parallel constraints and an anomaly correction mechanism to achieve long-term data association in this scenarios.

[1]Institutes of Physical Science and Information Technology, Anhui University, Hefei, 230601

[2]Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, 230031. dong.jun@iim.ac.cn

ACKNOWLEDGMENT: This work was supported in part by The National Key Research and Development Program of China grant number 2022YFD2001404.

II. RELATED WORK

A. Visual-Inertial SLAM/Odometry Systems

Visual-Inertial Simultaneous Localization and Mapping (SLAM) systems and VIO have received significant attention and research in the past decade. The combination of visual and inertial sensors provides robustness against poor texture, motion blur, and occlusions for the system. Many systems [15], [16], [17], [18] have extended the approaches used in VO systems, either directly or indirectly, and have further adopted tightly coupled methods to integrate IMU predictions. Among these, ORB-SLAM-VI [18] is the first to propose using maps with short-term, medium-term, and long-term data associations in a Visual-Inertial SLAM system. Building upon [18], ORB-SLAM3 [14] integrates a fast initialization method and multi-map systems, making it an extremely excellent Visual-Inertial SLAM system in accuracy and robustness.

B. Mapping and Localization in Agricultural Open Fields

In recent years, some reviews [19], [20], [21] of agricultural robots have demonstrated a strong interest in the integration of SLAM systems. Faced with the significant challenges of agricultural scenarios, SLAM systems have been extensively researched from various angles. The utilization of crop features has given birth to some specialized VO systems for specific planting scenarios [22], [23], [24]. Authors in [22] extracted grapes' semantic features and depth information and achieved a short-distance, centimeter-level localization system using short-term data correlation. Similarly, a loop closure detection method based on grape stem semantic segmentation proposed in [23] further optimized the long-term data correlation problem of VIO in vineyards. To avoid interference or specific dependencies related to crops, authors in [25], [26], [27], [28] have used artificial landmarks, such as squared fiducial markers, placed in the environment to help tracking and relocalization. However, this method is not feasible for open fields in agriculture. From the perspective of universality, researchers attempt to study the performance degradation of VSLAM in agricultural scenarios and make adaptive improvements [8], [9], [29]. The authors in [7] based on Stereo-DSO [30], proposed a set of improvement strategies for agricultural scenarios such as depth suppression, tracking failure and recovery, feature selection. This method can accomplish positioning tasks without relying on specific crop features or fiducial markers, but the direct method limits its association with long-term data. [31] published the first effective monocular SLAM in the agricultural field scene and simulated the system input of RGB-D SLAM through unsupervised depth estimation. In addition, SLAM based on multi-sensor fusion has been widely studied in agricultural scenarios due to its superior accuracy and robustness. Authors in [32] proposed a multi-camera visual system with non-overlapping views, completing the construction of corn crop rows containing scale and semantic information. Authors in [33] proposed a global pose estimation framework based on a multi-module fusion strategy. However, this system will face high hardware costs and complicated calibration in the promotion process.

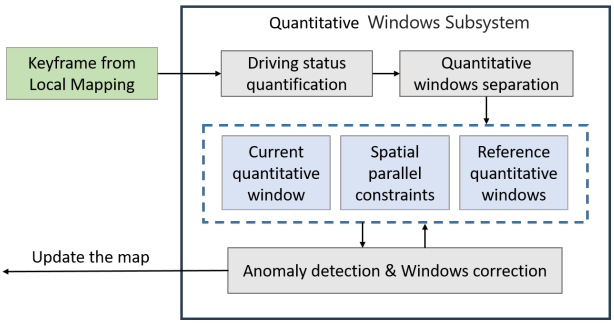


Fig. 2: Core subsystem components of Field-VIO.

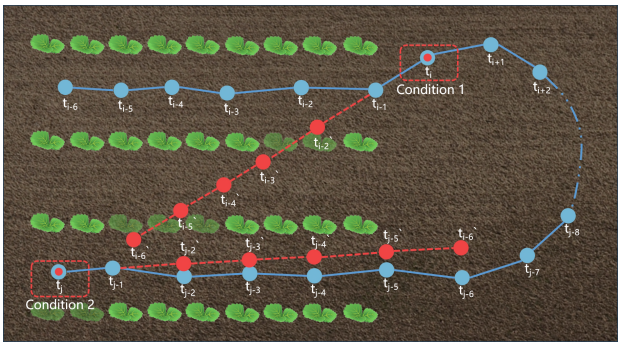


Fig. 3: The driving state quantification algorithm.

III. SYSTEM DESCRIPTION

In our system, the visual front-end retains the initialization strategy of ORB-SLAM3 [14] and realizes robust tracking. We shield the original LCD task and design a subsystem. Fig. 2 depicts the primary architecture of this subsystem. We utilize the driving state quantification module of the agricultural robot to receive the latest keyframes provided by the local mapping thread. Subsequently, the quantification module separates quantitative windows. These separated elements can be specifically divided into current quantitative window, reference quantitative windows, and spatial parallel constraint relationships (shown in blue boxes of Fig. 2). Finally, the system will correct the current quantitative window and update the map if abnormal spatial parallelism is detected between the windows. Tab. I explains some symbols that will be used shortly.

A. Quantitative Index of Driving State

Within the original loop and map merging thread, we have maintained a queue of the recent keyframes with a length of N_{cur}. We simulate the theoretical positions of the recent N_{cur} keyframes based on the structural features of crop rows in the field. (1) represents the reverse recursion of the theoretical positions of each keyframe in the queue:

$$\left[\boldsymbol{t}'_{i-2} - \boldsymbol{t}'_{i-1}\right]_{xy} = \frac{dis(\boldsymbol{t}_{i-2}, \boldsymbol{t}_{i-1})}{dis(\boldsymbol{t}_{i-1}, \boldsymbol{t}_{i})} \left[\boldsymbol{t}'_{i-1} - \boldsymbol{t}'_{i}\right]_{xy}, i > N_{cur} \tag{1}$$

where $\boldsymbol{t}$ represents the estimated positions of keyframes from the odometry, corresponding to the blue points in Fig. 3,

TABLE I: Summary of Notations

w_{cur}	Current quantitative window of keyframes (shown in Fig. 4).
w_{ref}	Reference quantitative window for the establishment of spatial parallel constraints.
N_{cur}	Length of temporary window in w_{cur} (shown in Fig. 4).
N_{ref}	Length threshold associated with w_{ref}.
c_{QIDS}	In Sec. III-A, the latest quantitative index by the driving state quantification algorithm.

where $\boldsymbol{t}^{'}$ represents the theoretical positions, corresponding to the red points in Fig. 3. $dis(\cdot)$ means the horizontal distance between two positions and $[\cdot]_{xy}$ denotes the initial two dimensions of the matrix. In the initial condition of this recursion, the two latest keyframes' theoretical and actual estimated positions are identical. In Fig. 3, the two conditions illustrate the discrepancy between the actual trajectory (blue polyline) and the theoretical trajectory (red polyline) under different driving states (U-turn motion and straight motion). Inspired by the Absolute Trajectory Error (ATE) evaluation metric [34], we quantify the motion state of the agricultural robot by calculating the trajectory error between the actual keyframe sequence and the simulated keyframe sequence. The formula for the Quantitative Index of Driving State (QIDS) is as follows:

$$QIDS = \sqrt{\frac{1}{N_{cur}} \sum_{i=1}^{N_{cur}} \left\| \text{trans}\left(T_{esti,i}^{-1} \cdot T_{\text{theo},i}\right)_{\text{xy}} \right\|_2^2} \quad (2)$$

where $T_{esti,i}$ and $T_{\text{theo},i}$ respectively denote the estimated pose in the real and the theoretical pose in the simulated. $\text{trans}(\cdot)_{xy}$ signifies the projection of the pose matrix's translation vector on the XY plane. $\|\cdot\|_2^2$ represents the square of the vector's Euclidean norm. It is noteworthy that we have not considered the position information of the keyframe in the direction of gravity, and the two primary reasons are as follows:

- The driving state quantification algorithm primarily aims to describe the change in the horizontal heading of the agricultural robot as accurately as possible.
- In agricultural fields, the component in the direction of gravity often becomes the primary source of the entire trajectory error due to the robot's bumpy ride. It could lead to frequent false positives of the quantification algorithm, thereby affecting the reliability of quantitative windows separation.

B. Quantitative Windows and Spatial Parallel Constraints

Our system perceives segments of the robot's straight-line trajectory along the crop rows as quantitative windows. Some windows that meet the requirements are regarded as reference quantitative windows. Under the premise of avoiding false positive judgments, the spatial parallel constraints between windows will be constructed.

Quantitative windows separation process is shown in Alg. 1. We take the completion of the system's first inertial initialization as the start flag for this process (line 1 of Alg. 1). In the control loop, we obtain first the c_{QIDS} of each keyframe that has just arrived from the local mapping thread (line 4). We set a threshold α to distinguish the apparent change of driving state. If c_{QIDS} is less than α (line 5), it is considered that the robot maintains a straight-line driving state along the crop row, and the latest keyframe is added to w_{cur} (line 6). Conversely, it is considered that the driving state of the current frame may have changed significantly. The algorithm will mark the current frame (line 8) and analyze the recent temporary window (shown in Fig. 4) in w_{cur}. If the marked keyframes in this temporary window meet the requirements of time continuity, then w_{cur} will be reset (line 12).

Algorithm 1 Quantitative Windows Separation Algorithm

```
1:  INERTIALINITIALIZATION
2:  repeat
3:    KF_cur ← GETKEYFRAMEFROMLM()
4:    c_QIDS ← COMPUTEQIDS(KF_cur)
5:    if c_QIDS < α then
6:      APPENDINTOCURRENTWIN(KF_cur)
7:    else
8:      SETUPLABLE(KF_cur)
9:      if CONTINUITYDISCERNMENT(w_cur,N_cur) then
10:       if COUNTUNMATCH() and SIZEOFSUB > N_ref
          then
11:       w_ref ← ESTABLISHREFWIN()
12:       RESET(w_cur)
13: until SYSTEMSHUTDOWN
```

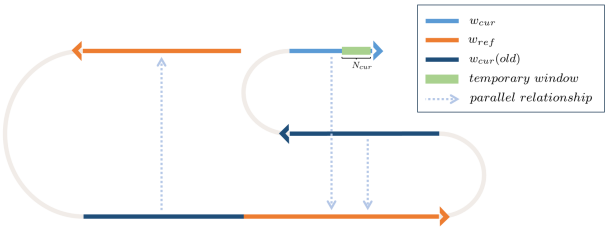


Fig. 4: Construction of spatial parallel constraints.

Spatial parallel constraints is constructed in Fig. 4. Parallel constraint matching is performed for each keyframe in w_{cur}, meeting the following two criteria: 1) The keyframe in w_{ref} with the minimal distance to the current keyframe's optical center is selected as the matching frame. 2) The matching frame should not be positioned at the edge of w_{ref}. Notably, the first keyframe in w_{cur} that completes parallel constraint matching will serve as the starting keyframe for the subsequent anomaly correction method.

Inspired by the multi-map system in [6], we have adopted a strategy of multiple reference windows and stipulate the following: 1) The system can possess multiple reference windows, which do not overlap with each other; 2) w_{cur} can establish a frame-to-frame parallel constraint relationship with only one w_{ref}. Due to the delay in resetting w_{cur} caused by the discernment of temporal continuity, we prioritize trimming off several frames newly added to w_{cur}. When the system resets w_{cur}, if there exists a sub-window in w_{cur} that

has not completed the parallel constraint matching and the length of the sub-window reaches N_{ref}, the sub-window will be set as a w_{ref} (line 10 of Alg. 1). If the current keyframe in w_{cur} matches a w_{ref} different from the past, the separation of w_{cur} ends, and a new w_{cur} is immediately created.

C. Anomaly Correction

Trigger of correction. The anomaly correction mechanism will detect abnormalities in the current parallel matching relationship. When the length of the sub-window with parallel constraint matching relationship in w_{cur} is adequate, the anomaly correction mechanism will be activated. The parallel constraint discrimination formula is as follows:

$$dis_{start} * (1 - \lambda) < dis_{cur} < dis_{start} * (1 + \lambda) \quad (3)$$

where dis_{start} is the horizontal distance of optical center between the starting keyframe and the matching keyframe. dis_{cur} has a similar meaning to dis_{start} but represents the current keyframe. λ is the threshold for controlling the sensitivity of discrimination. When the above inequality does not hold, the system will further perform orientation angle verification of quantitative windows to prevent the occurrence of false positives:

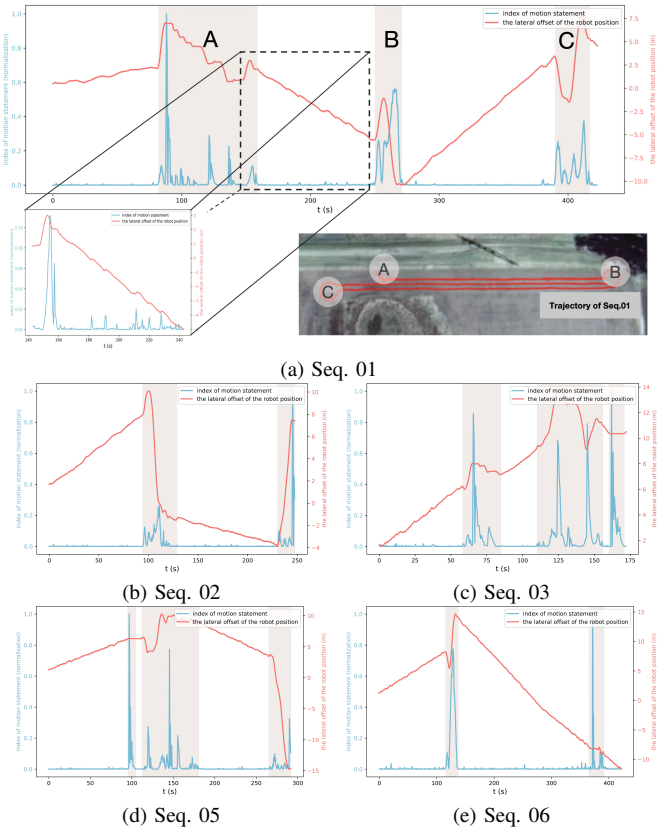
$$arccos(\boldsymbol{v}_{cur} \cdot \boldsymbol{v}_{ref}) < \Theta \quad (4)$$

where $\boldsymbol{v}_{cur}$ represents the unit vector of the overall orientation of w_{cur}, and $\boldsymbol{v}_{ref}$ represents the matched w_{ref}. In addition, there is a delay in the reset of w_{cur} in Sec. III-B, and positive judgment during this period may not necessarily be caused by anomalies of parallelism. Significant changes in the driving state may also trigger the aforementioned dual discrimination. So it is necessary to ensure that the frequency of the high QIDS appearing in the temporary window is low enough. The window correction algorithm will be carried out when all the above conditions are met.

Window correction algorithm. This algorithm is to rectify the sub-windows in w_{cur} that have already established the matching relationship. And the parallel spatial position relationship between w_{cur} and w_{ref} can be maintained again. In (5), the corrected position vector of the current keyframe is inferred as $\boldsymbol{t}_{pred}$ based on the relative position transformation vector $\Delta\boldsymbol{t}$ between the starting keyframe in w_{cur} and the keyframe matched in w_{ref}. $\boldsymbol{t}_{cur,match}$ denotes the position vector of the keyframe matched in w_{ref} for the current keyframe. The quantitative window rectification emphasizes eliminating the cumulative error of the yaw angle estimation. Hence, we retain the component in the direction of gravity in the original keyframe position vector $\boldsymbol{t}_{cur}$ and deduce the final vector $\boldsymbol{t}_{pred}^{'}$:

$$\begin{aligned} \boldsymbol{t}_{pred} &= \boldsymbol{t}_{cur,match} + \Delta\boldsymbol{t} \\ \boldsymbol{t}_{pred}^{'} &= \begin{bmatrix} \boldsymbol{t}_{pred,1} & \boldsymbol{t}_{pred,2} & \boldsymbol{t}_{cur,3} \end{bmatrix}^T \end{aligned} \quad (5)$$

It should be noted that the position vectors mentioned above refer to the world coordinate system. While the subsequent algorithm steps refer to the coordinate system of the starting keyframe (the primary coordinate system) in



(a) Seq. 01

(b) Seq. 02 (c) Seq. 03

(d) Seq. 05 (e) Seq. 06

Fig. 5: Visual output and comparison of proposed driving state quantification algorithm.

w_{cur}. We can then obtain the unit vectors of $\boldsymbol{t}_{cur}$ and $\boldsymbol{t}_{pred}^{'}$ in the primary coordinate system:

$$\hat{\boldsymbol{v}}_{unc} = \frac{\boldsymbol{T}_{mw}\boldsymbol{t}_{cur}}{\|\boldsymbol{T}_{mw}\boldsymbol{t}_{cur}\|}, \hat{\boldsymbol{v}}_{c} = \frac{\boldsymbol{T}_{mw}\boldsymbol{t}_{pred}^{'}}{\|\boldsymbol{T}_{mw}\boldsymbol{t}_{pred}^{'}\|} \quad (6)$$

where $\boldsymbol{T}_{mw}$ represents the primary coordinate system. In (7), we solve for the axial angle $\boldsymbol{\phi}$, which is necessary for correcting the primary coordinate system, by utilizing $\hat{\boldsymbol{v}}_{unc}$ and $\hat{\boldsymbol{v}}_{c}$.

$$\begin{aligned} \theta &= arccos(\hat{\boldsymbol{v}}_{c} \cdot \hat{\boldsymbol{v}}_{unc}) \\ \hat{\boldsymbol{r}} &= \frac{\hat{\boldsymbol{v}}_{c} \times \hat{\boldsymbol{v}}_{unc}}{\|\hat{\boldsymbol{v}}_{c} \times \hat{\boldsymbol{v}}_{unc}\|} \\ \boldsymbol{\phi} &= \theta\hat{\boldsymbol{r}}, \theta \in [0, \pi] \end{aligned} \quad (7)$$

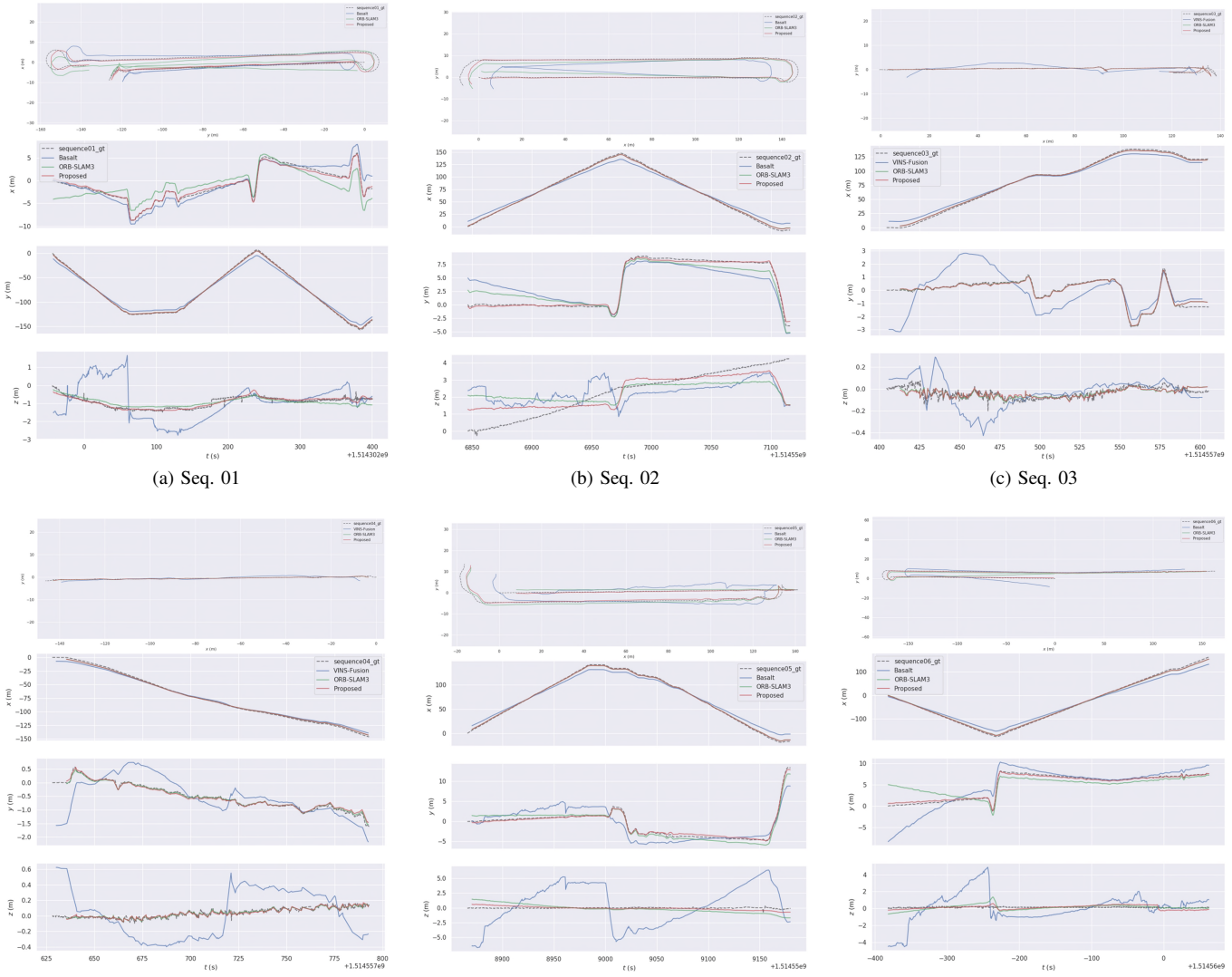
The pose transformation matrix $\boldsymbol{T}_{m'm}$ required for the correction is further obtained through $\boldsymbol{\phi}$:

$$\begin{aligned} \boldsymbol{R}_{m'm} &= \exp\left(\boldsymbol{\phi}^{\wedge}\right) \\ \boldsymbol{T}_{m'm} &= \begin{bmatrix} \boldsymbol{R}_{m'm} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \end{aligned} \quad (8)$$

Equation (8), we represent the exponential map as $\exp(\cdot)$, where $\boldsymbol{R}_{m'm} \in \mathbb{R}^{3\times3}$ is the rotation matrix required for correction. Then, all keyframe poses in sub-window in w_{cur} are corrected and propagated utilizing $\boldsymbol{T}_{m'm}$. Since the relative poses among the keyframes in the window do not change before and after the correction of the primary coordinate system, that is, $\boldsymbol{T}_{i'm'} = \boldsymbol{T}_{im} = \boldsymbol{T}_{iw}\boldsymbol{T}_{wm}$. The pose transformations of each rectified keyframe relative

TABLE II: The Mean and the Root-Mean-Square Error (RMSE) of the Absolute Trajectory Error (ATE)

Dataset Rosario		VINS-Fusion		OKVIS		SVO 2.0		Basalt		ORB-SLAM3		Proposed	
Seq	Length(m)	ATE	RMSE	ATE	RMSE	ATE	RMSE	ATE	RMSE	ATE	RMSE	ATE	RMSE
01	475	-	-	6.54	7.29	11.70	12.79	5.61	6.27	1.95	2.12	**1.13**	**1.27**
02	320	13.33	16.08	8.24	9.51	7.32	8.49	7.19	8.06	2.25	2.44	**1.72**	**1.98**
03	170	5.71	6.79	8.13	9.39	6.69	7.65	6.17	7.11	**1.60**	**1.84**	1.67	1.92
04	150	4.01	4.68	7.45	8.69	5.81(1)	6.74	6.00	6.92	**1.51**	**1.76**	**1.51**	1.77
05	330	7.04(3)	8.67	7.78(5)	9.00	10.53	12.15	6.98	8.02	1.54	1.72	**1.37**	**1.55**
06	530	13.47	14.92	15.81	17.81	17.58(4)	20.04	13.10	15.19	3.82	4.29	**3.07**	**3.72**



(a) Seq. 01 (b) Seq. 02 (c) Seq. 03

(d) Seq. 04 (e) Seq. 05 (f) Seq. 06

Fig. 6: Comparison of xy-plane trajectories and three-axis trajectory components on the Rosario dataset [10].

to the world coordinate system, $\boldsymbol{T}_{i'w}$, can be calculated as follows:

$$\boldsymbol{T}_{i'w} = \boldsymbol{T}_{iw}\boldsymbol{T}_{wm}\boldsymbol{T}_{m'm}\boldsymbol{T}_{wm}^{-1} \tag{9}$$

In a manner akin to the rectification of the pose, the transformation of the velocity vectors of keyframes with the primary coordinate system remains invariant, i.e., $\boldsymbol{V}_{m'i'} = \boldsymbol{V}_{mi}$. Thus, the corrected velocity vectors of each keyframe in the world coordinate system, $\boldsymbol{V}_{i'}$, can be obtained:

$$\boldsymbol{V}_{i'} = \boldsymbol{R}_{wm}\boldsymbol{R}_{mm'}\boldsymbol{R}_{wm}^{-1}\boldsymbol{V}_{i} \tag{10}$$

IV. EXPERIMENTAL EVALUATION

To evaluate our system, we have conducted experiments from two perspectives. We visualized the QIDSs and compared them intuitively with the estimated lateral offset of

TABLE III: Details of Quantitative Windows Correction

Sequence	01	02	03	04	05	06
U-turn	3	1	1	0	1	1
Traversed Row	3	2	1	1	2	2
Frequency	3.8	2	0	0	0.9	1.5
Reduction	42.1%	23.6%	0	0	11.0%	19.6%

the robot's location. Then, we compared the performance of the proposed system with the current state-of-the-art VIO systems. All experiments were conducted on the same computer equipped with Ubuntu 18.04 LTS, an Intel Core™ i7-12700KF Processor, and 32GB of RAM.

A. The Dataset and Evaluation Method

In agricultural open fields, there is a significant dearth of navigation datasets featuring real sensor readings collected by robots. We have chosen the Rosario dataset [10] composed of six sequences recorded in a soybean field. The dataset presents real and challenging scenarios: highly repetitive scenes, reflections, and overexposed images caused by direct sunlight and rough terrain. And it includes synchronized readings from a wheel odometer, IMU (140 Hz), stereo camera (672 × 376 px, 15 Hz), and a GPS - RTK system (5 Hz). We conducted exhaustive experiments in this dataset, running each VIO system ten times on each sequence. Given that the Rosario dataset provides ground truth for position (without orientation), we used the mean and the Root Mean Square Error (RMSE) of the Absolute Trajectory Error (ATE) as quantitative indices of system performance.

B. The Quantified Performance of Driving State

First, we normalize some of the data. In Fig. 5 (a), we provide a detailed analysis of Seq. 01 from the Rosario dataset as a representative case. The results of four additional sequences are presented in Fig. 5 (b, c, d, e). We omit Seq. 04, as it proceeds in a straight line along a single crop row, a situation that is covered by the other sequences.

In Fig. 5 (a), the estimated lateral offset of the robot's location (orange curve, without anomaly correction) indirectly reflects the changes in the robot's heading during its journey. During the three time periods (top, bottom right, translucent areas marked with the letter A, B, and C), it can be observed that the output (blue curve) of the driving state quantification algorithm undergoes dramatic changes, which correspond precisely to the robot's turning state at the end of the crop rows. Notably, as can be seen in the magnified area in Fig. 5 (a, bottom left), even when the robot makes minor adjustments to its heading during its straight-line motion along the row (orange curve), our algorithm can generate sensitive feedback. Furthermore, during the time periods marked in Fig. 5 (translucent areas), sudden stops of the robot occurred in the third period of Seq. 03, the first period of Seq. 05, and the second period of Seq. 06. Our algorithm effectively recognized these state changes and provided distinct output.

C. Odometry System Evaluations

We have evaluated the performance of the proposed system against several exemplary systems, which include VINS-Fusion [2], OKVIS [35], SVO 2.0 [3], ORB-SLAM3 [14], and Basalt [36]. In Tab. II, the bold values represent the best results for each sequence. If a system failure occurs during the experiment, the number of failures is recorded at the end of the APE value. There is a special symbol (-) to indicate that system always fails in sequence.

In Tab. II, it is evident that our approach has garnered an average APE of 1.13m, a decrease of 42.1% compared to the next best, ORB-SLAM3, on Seq. 01. Our methodology showed an improvement of 10% to 25% over the second-best approach in Seq. 02, 05, and 06 without malfunction. The aforementioned four sequences all belong to the longer trajectory sequences in the dataset, with Seq. 01 even encompassing three 180-degree end-of-row turnarounds. Notably, for Seq. 03 and Seq. 04, our system's localization performance is virtually identical to that of ORB-SLAM3 [14], as our correction method was never triggered in Seq.03 and Seq.04. In Tab. III, it is discernible that the extent (Reduction line) of performance enhancement built upon ORB-SLAM3 [14] bears a certain positive correlation with the frequency of quantitative windows correction (Frequency line). The correction frequency is directly linked to the number of U-turns and the crop rows traversed within the sequence.

In Fig. 6, we only exhibit the trajectory results of the top three methods that performed most impressively on each sequence. Our method yields trajectory outputs on the Rosario dataset [10] that are most proximate to the ground truth. Furthermore, from the lateral offset component output results of Seq.01, 02, 05, 06, it can be inferred that our odometry system is capable of more accurately estimating the yaw angle during the robot's travel. In summary, our VIO system demonstrates superior positioning accuracy and reliability in large agricultural field scenarios.

V. CONCLUSIONS

In this paper, we have introduced a novel Stereo VIO system customized for agricultural open field scenes. The system makes full use of the clues of spatial structure from semi-structured crop rows. The driving state quantification algorithm can accurately describe the driving state of robots and separate quantitative windows. Spatial parallel constraints are formulated at the keyframe level based on the identified quantified windows. The introduced anomaly detection mechanism continually monitors potential spatial relationship anomalies among windows. The quantitative window correction algorithm maintains the spatial parallelism and mitigates cumulative positioning errors through long-term information association. Our latest findings, obtained from the publicly available Rosario dataset [10], demonstrate that our method is effective in agricultural open fields. In the future, we will delve deeper into the issues of semantic reconstruction in agricultural settings.

REFERENCES

[1] Y. Bai, B. Zhang, N. Xu, J. Zhou, J. Shi, and Z. Diao, "Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review," *Computers and Electronics in Agriculture*, vol. 205, p. 107584, Feb. 2023.

[2] T. Qin, J. Pan, S. Cao, and S. Shen, "A General Optimization-based Framework for Local Odometry Estimation with Multiple Sensors," Jan. 2019.

[3] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, Apr. 2017.

[4] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept. 2015, pp. 298–304.

[5] J. Cremona, R. Comelli, and T. Pire, "Experimental evaluation of Visual-Inertial Odometry systems for arable farming," *Journal of Field Robotics*, vol. 39, no. 7, pp. 1121–1135, Oct. 2022.

[6] R. Elvira, J. D. Tardós, and J. Montiel, "ORBSLAM-Atlas: A robust and accurate multi-map system," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 6253–6259.

[7] T. Yu, J. Zhou, L. Wang, and S. Xiong, "Accurate and Robust Stereo Direct Visual Odometry for Agricultural Environment," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE, May 2021, pp. 2480–2486.

[8] K. Song, J. Li, R. Qiu, and G. Yang, "Monocular Visual-Inertial Odometry for Agricultural Environments," *IEEE Access*, vol. 10, pp. 103 975–103 986, 2022.

[9] T. Yu, X. Yu, W. Liu, and S. Xiong, "Scale-aware stereo direct visual odometry with online photometric calibration for agricultural environment," *Advanced Robotics*, pp. 1–14, Nov. 2022.

[10] T. Pire, M. Mujica, J. Civera, and E. Kofman, "The Rosario dataset: Multisensor data for localization and mapping in agricultural environments," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 633–641, May 2019.

[11] Y. Latif, C. Cadena, and J. Neira, "Robust loop closing over time for pose graph SLAM," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1611–1626, Dec. 2013.

[12] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[13] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[14] C. Campos, R. Elvira, J. J. Gomez Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *Ieee Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.

[15] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. Rome, Italy: IEEE, Apr. 2007, pp. 3565–3572.

[16] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[17] L. Von Stumberg, V. Usenko, and D. Cremers, "Direct Sparse Visual-Inertial Odometry Using Dynamic Marginalization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, QLD: IEEE, May 2018, pp. 2510–2517.

[18] R. Mur-Artal and J. D. Tardos, "Visual-Inertial Monocular SLAM With Map Reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, Apr. 2017.

[19] B. Xie, Y. Jin, M. Faheem, W. Gao, J. Liu, H. Jiang, L. Cai, and Y. Li, "Research progress of autonomous navigation technology for multi-agricultural scenes," *Computers and Electronics in Agriculture*, vol. 211, p. 107963, Aug. 2023.

[20] H. Ding, B. Zhang, J. Zhou, Y. Yan, G. Tian, and B. Gu, "Recent developments and applications of simultaneous localization and mapping in agriculture," *Journal of Field Robotics*, vol. 39, no. 6, pp. 956–983, Sept. 2022.

[21] A. S. Aguiar, F. N. dos Santos, J. B. Cunha, H. Sobreira, and A. J. Sousa, "Localization and Mapping for Robots in Agriculture and Forestry: A Survey," *Robotics*, vol. 9, no. 4, p. 97, Nov. 2020.

[22] A. K. Nellithimaru and G. A. Kantor, "ROLS : Robust Object-Level SLAM for Grape Counting," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Long Beach, CA, USA: IEEE, June 2019, pp. 2648–2656.

[23] A. Papadimitriou, I. Kleitsiotis, I. Kostavelis, I. Mariolis, D. Giakoumis, S. Likothanassis, and D. Tzovaras, "Loop Closure Detection and SLAM in Vineyards with Deep Semantic Cues," in *2022 International Conference on Robotics and Automation (ICRA)*, May 2022, pp. 2251–2258.

[24] M. Qadri and G. Kantor, "Semantic Feature Matching for Robust Mapping in Agriculture," July 2021.

[25] W. Zhang, L. Gong, S. Huang, S. Wu, and C. Liu, "Factor graph-based high-precision visual positioning for agricultural robots with fiducial markers," *Computers and Electronics in Agriculture*, vol. 201, p. 107295, Oct. 2022.

[26] M. Kalaitzakis, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, "Experimental Comparison of Fiducial Markers for Pose Estimation," in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, Sept. 2020, pp. 781–789.

[27] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Tracking fiducial markers with discriminative correlation filters," *Image and Vision Computing*, vol. 107, p. 104094, Mar. 2021.

[28] R. Muñoz-Salinas and R. Medina-Carnicer, "UcoSLAM: Simultaneous localization and mapping by fusion of keypoints and squared planar markers," *Pattern Recognition*, vol. 101, p. 107193, May 2020.

[29] R. Islam, H. Habibullah, and T. Hossain, "AGRI-SLAM: A real-time stereo visual SLAM for agricultural environment," *Autonomous Robots*, vol. 47, no. 6, pp. 649–668, Aug. 2023.

[30] R. Wang, M. Schwörer, and D. Cremers, "Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 3923–3931.

[31] F. Shu, P. Lesur, Y. Xie, A. Pagani, and D. Stricker, "SLAM in the Field: An Evaluation of Monocular Mapping and Localization on Challenging Dynamic Agricultural Environment," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021, pp. 1760–1770.

[32] J. Yuan, J. Hong, J. Sattar, and V. Isler, "ROW-SLAM: Under-Canopy Cornfield Semantic SLAM," in *2022 International Conference on Robotics and Automation (ICRA)*, May 2022, pp. 2244–2250.

[33] M. Imperoli, C. Potena, D. Nardi, G. Grisetti, and A. Pretto, "An Effective Multi-Cue Positioning System for Agricultural Robotics," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3685–3692, Oct. 2018.

[34] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura-Algarve, Portugal: IEEE, Oct. 2012, pp. 573–580.

[35] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, Mar. 2015.

[36] V. Usenko, N. Demmel, D. Schubert, J. Stuckler, and D. Cremers, "Visual-Inertial Mapping With Non-Linear Factor Recovery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 422–429, Apr. 2020.