



Article

Adaptive Fusion Positioning Based on Gaussian Mixture Model for GNSS-RTK and Stereo Camera in Arboretum Environments

Shenghao Liang , Wenfeng Zhao *, Nuanchen Lin  and Yuanjue Huang

College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China

* Correspondence: zwf555@scau.edu.cn; Tel.: +86-1353-899-8771

Abstract: The integration of Global Navigation Satellite System (GNSS) Real-Time Kinematics (RTK) can provide high-precision, real-time, and global coverage of location information in open areas. But in arboretum environment, the ability to achieve continuous high-precision positioning using global positioning technology is limited due to various sources of interference, such as multi-path effects, signal obstruction, and environmental noise. In order to achieve precise navigation in challenging GNSS signal environments, visual SLAM systems are widely used due to their ability to adapt to different environmental features. Therefore, this paper proposes an optimized solution that integrates the measurements from GNSS-RTK and stereo cameras. The presented approach aligns the coordinates between the two sensors, and then employs an adaptive sliding window approach, which dynamically adjusts the window size and optimizes the pose within the sliding window. At the same time, to address the variations and uncertainties of GNSS signals in non-ideal environments, this paper proposes a solution that utilizes a Gaussian Mixture Model (GMM) to model the potential noise in GNSS signals. Furthermore, it employs a Variational Bayesian Inference-based (VBI) method to estimate the parameters of the GMM model online. The integration of this model with an optimization-based approach enhances the positioning accuracy and robustness even further. The evaluation results of real vehicle tests show that in challenging GNSS arboretum environments, GMM applied to GNSS/VO integration has higher accuracy and better robustness.



Citation: Liang, S.; Zhao, W.; Lin, N.; Huang, Y. Adaptive Fusion Positioning Based on Gaussian Mixture Model for GNSS-RTK and Stereo Camera in Arboretum Environments. *Agronomy* **2023**, *13*, 1982. <https://doi.org/10.3390/agronomy13081982>

Academic Editor: Miguel-Ángel Muñoz-García

Received: 28 June 2023

Revised: 19 July 2023

Accepted: 24 July 2023

Published: 27 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: sensor fusion; global navigation satellite system (GNSS); vision SLAM; state estimation; Gaussian mixture models; arboretum

1. Introduction

In the field of robotics, localization refers to the process by which a robot determines its own position and orientation in an environment. Precise and reliable positioning is essential for autonomous navigation of robots [1]. Different types of navigation sensors can provide different information. Traditional Global Navigation Satellite Systems (GNSS) can provide global positioning information with an average accuracy of about 10 m [2], which can be improved to centimeter-level accuracy through Real-Time Kinematic (RTK) techniques [3]. However, in complex environments like arboretums, the reliability and accuracy of GNSS signals are limited due to factors such as signal obstruction, reflection, and multi-path effects [4]. These factors can cause signal loss or deformation, which hinders the effectiveness of GNSS positioning in such environments. Lidar is widely used in various fields such as autonomous driving, vehicle trajectory prediction [5], vehicle perception [6], and precision agriculture [7]. However, its high cost and demanding environmental requirements limit its widespread adoption in certain application areas. Camera sensors are widely adopted in various applications due to their low cost, small size, light weight, and high environmental adaptability [8,9]. Camera sensors can capture image information from the environment and extract visual features of objects. Indeed, there has been a significant amount of academic research on visual SLAM (Simultaneous Localization and

Mapping) [10–13]. Due to the nature of visual SLAM systems, which rely on historical observations for estimation and prediction, there is a tendency for cumulative errors to occur over long-duration operation. These accumulated errors can result in an increase in positioning errors over time. However, due to the unique advantages and limitations of each sensor, it is common to combine different sensors to leverage their complementary strengths. Research studies have indicated [14] that the complementary fusion of GNSS and vision can effectively integrate the advantages of both sensors.

To fully leverage the complementarity between different sensors and achieve multi-sensor fusion localization, it is essential to choose an appropriate fusion framework to enhance the robustness of the entire positioning system. In the early stages of research, most studies employed filtering-based approaches for sensor fusion [15]. In [16–19], the authors employed a GNSS-IMU fusion system based on the Kalman filter to perform motion estimation of the sideslip angle and attitude for autonomous driving vehicles. Ref. [20] proposed a loosely-coupled GPS/VO fusion navigation system and validates its observability. In [21], the authors propose a vehicle localization method in urban environments that integrates stereo vision and (GNSS-RTK) using the Extended Kalman Filter (EKF). This method effectively reduces visual drift errors. Ref. [22] propose an EKF method that tightly integrates measurements from GNSS RTK, MEMS-IMU, and monocular camera. The performance of the proposed method is validated through real-world vehicle experiments. However, filtering-based methods can only consider measurements at the current time step and may not perform well in handling nonlinear systems, which limits their applicability.

In recent years, with the significant improvement in computational resources, new techniques based on optimization methods have been proposed. For example, the factor graph-based methods [23] have emerged, which represent the estimation states and observation equations as nodes and edges in a factor graph. Optimization solvers such as G2o [24] and Ceres solver [25] are then used to solve the optimization problem. Refs. [26,27] proposed an optimization-based multi-sensor fusion framework that combines local VIO results with GNSS measurements. Ref. [28] proposed a GPS visual inertial odometer system based on tightly coupled optimization, which solves the problem of trajectory drift during long-term operation. Ref. [29] provides self-motion estimation by tightly coupling GNSS pseudorange measurements and camera feature points. These optimization-based methods offer flexibility in handling nonlinear problems and allow for simultaneous optimization of multiple variables. In [30], the study explores the integration of GNSS and INS using the Factor Graph Optimization (FGO) method and demonstrates its superior performance compared to the EKF-based approach.

However, the current optimization-based methods, while exhibiting some robustness to outliers, still rely on the Gaussian assumption. Especially in complex environments such as arboretums with tree obstructions, GNSS signals can be affected, and the noise distribution often deviates from a Gaussian distribution [31]. Therefore, existing robustness methods are difficult to meet the practical requirements, which remains a significant challenge for achieving accurate positioning. Ref. [32] proposes an adaptive fusion of GNSS and visual-inertial odometry, enabling positioning in intermittent degradation of GNSS. Ref. [33] combines a sky-facing camera with GNSS to mitigate the impact of GNSS outliers by performing sky region segmentation. However, these methods still cannot effectively cope with the non-Gaussian characteristics in GNSS signals, which also limits the application of traditional optimization-based methods.

In order to describe the non-Gaussian characteristics of GNSS errors, it is necessary to consider more complex error models and incorporate their true distribution into the state estimation equation. Among them, Gaussian Mixture Model (GMM) [34] is currently one of the most commonly used models that can effectively describe the characteristics of non-Gaussian error distributions. Currently, several algorithms [35–37] have attempted to address the issue of non-Gaussian noise distributions, and their research results indicate that modeling sensor noise with non-Gaussian distributions is feasible. However, in practical

applications, further research is needed to explore how to fully exploit the potential of non-Gaussian noise modeling in multi-sensor fusion applications. Among them, the authors of [38–41] conducted in-depth research on non-Gaussian measurement errors in robot sensor fusion algorithms and provided a series of advanced algorithms. In [38], a dynamic covariance estimation algorithm is proposed, which includes covariance in the optimization problem to better approximate the true attributes of the sensor. In [39], the authors use Maximum Mixture (MM) as a probability model and achieve good application results in online positioning with GNSS. In [40], the authors use a GMM to describe GNSS pseudorange errors, then estimate the mixture model through Expectation Maximization (EM), and finally combine nonlinear least squares optimization to reduce estimation errors caused by Non-Line-of-Sight (NLOS) measurements. However, it should be noted that the model used is fixed in terms of the number of mixture components. In recent work [41], the authors replace the EM algorithm with a Variational Bayesian Inference (VBI) method. The proposed algorithm is capable of automatically adjusting the number of mixture components in the GMM. Compared to the EM algorithm, the GMM estimation method using VBI is more robust. Therefore, this study will adopt the Variational Bayesian Inference method proposed in [41].

In order to overcome the limitations of GNSS in complex environments, this paper proposes a novel integration system of GNSS-RTK and stereo Visual Odometry (VO). The system integrates the global data from the GNSS with the local data from binocular VO through nonlinear optimization, resulting in more accurate motion estimation. The proposed system addresses the issues of accumulated drift and trajectory deviation from the real world in Visual SLAM systems. Moreover, in challenging environments such as tree arboretums, this system effectively tackles issues like GNSS signal obstruction and multi-path effects by modeling GNSS noise using the GMM model. Furthermore, extensive experimental validations have been conducted in this study to demonstrate the superiority and robustness of the proposed system. These results provide new insights and methods for the further development of GNSS-based positioning and navigation technologies.

The rest of this paper is structured as follows. Section 2 presents the framework of the system and introduces the coordinate symbols used. In Section 3, a typical visual SLAM system will be reviewed first, and then the GNSS-VO initialization method will be introduced. In Section 4, it is explained how GNSS positioning and GMM are combined with graph optimization, and a method for dynamically estimating GMM parameters is introduced. Section 5 provides a detailed description of the global optimization layer in the GNSS-VO fusion system. Section 6 provides experimental content and evaluation. Finally, in Section 7, the experimental results and the limitations of the current study are discussed, along with the future directions, and followed by a conclusion in Section 8.

2. System Overview and Notation

2.1. System Overview

The system structure proposed in this paper is shown in Figure 1, which is mainly composed of two modules: a local preprocessing front-end and a global optimization back-end. The front-end processing of our system starts with the visual tracking module, which uses Structure from Motion (SfM) to estimate the pose of each frame. This visual component is developed based on ORB-SLAM2 [12], which provides a solid foundation for our research. In this paper, the front-end processing is divided into three stages based on the initialization state of the system: uninitialized, initialization, and initialized. In the uninitialized stage, the system relies solely on visual self-motion estimation for continuous localization. In the initialization stage, this paper proposes an initialization sliding window, where once the required number of GNSS measurements is available within the window, the transformation relationship between the local coordinate system and the global coordinate system is computed. After initialization, the system aligns the GNSS measurements with the local VO estimates. Finally, the estimated GNSS and VO information are propagated to the adaptive sliding window for state estimation.

In the back-end optimization, local pose estimation and GNSS measurements are used as inputs. An adaptive sliding window strategy is proposed to model GNSS errors within the window using GMM and estimate model parameters using the VBI algorithm. Additionally, when there are no GNSS measurements within the window or the system is in the uninitialized state, only visual information is optimized. After completing one round of window optimization, it is necessary to update the correspondence between the local system and the global system to ensure their synchronization. The proposed estimator addresses the issues of local accumulation drift and GNSS outliers, and it significantly improves the accuracy compared to existing methods such as VINS-Fusion and ORB-SLAM2.

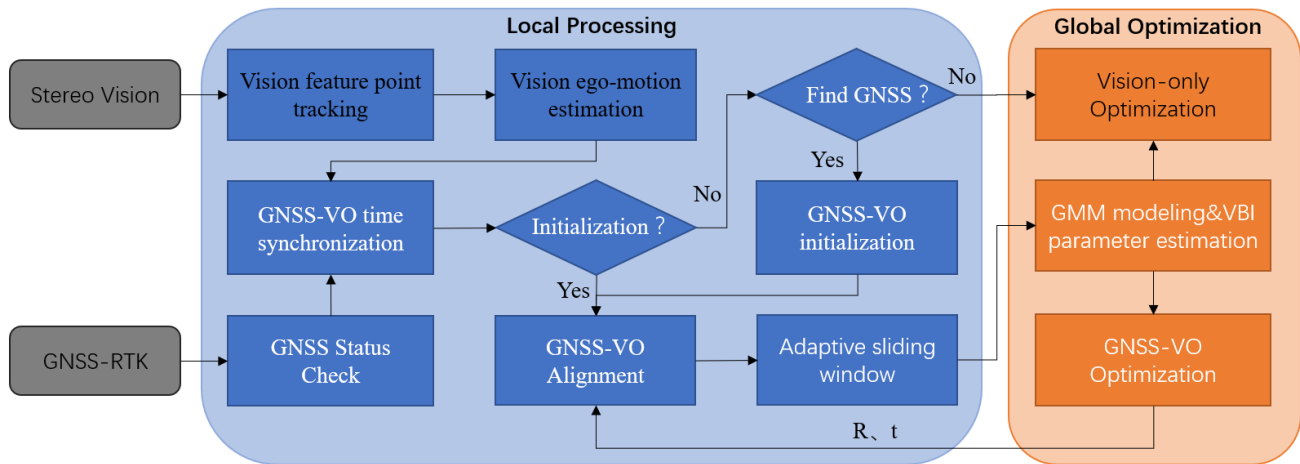


Figure 1. System framework diagram. Our system pipeline is an extension of ORB-SLAM2 [12]. Compared to ORB-SLAM2, our system incorporates several key improvements, which are highlighted in white font in the flowchart.

2.2. Notation

This paper introduces all the coordinate systems involved in this work, including the camera coordinate system F^c , the local coordinate system F^l , and the world coordinate system (ENU) F^w , as shown in Figure 2. Camera coordinate system F^c : The coordinate system is fixed on the vehicle, with the camera sensor as the origin, and it is used to describe the vehicle’s attitude and motion state. Local coordinate system F^l : The coordinate system is established with the first frame’s camera pose as the origin. Before the vehicle starts moving, the coordinate frames F^c and F^l coincide with each other. World coordinate system F^w : The coordinate system is defined with the first frame’s GNSS position information as the origin, and the positive directions are set as east, north, and up, respectively.

The conversion between different coordinate systems can be represented using a homogeneous transformation matrix, and the general coordinate transformation equation is as follows:

$$\begin{bmatrix} p' \\ 1 \end{bmatrix} = T \begin{bmatrix} p \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} p \\ 1 \end{bmatrix}, \tag{1}$$

where [42] adding 1 at the end of a three-dimensional vector is referred to as the homogeneous coordinate representation of a point. p' and p represent points in two different coordinate systems, T represents the transformation matrix of the two coordinate systems, R is a 3×3 rotation matrix, and t is a 3×1 translation vector. Therefore, assuming there is a point $p^c = [p_x^c \ p_y^c \ p_z^c]^T$, the conversion from the camera coordinate system to the world coordinate system can be expressed as:

$$\begin{bmatrix} p^w \\ 1 \end{bmatrix} = T_l^w T_c^l \begin{bmatrix} p^c \\ 1 \end{bmatrix} = \begin{bmatrix} R_l^w & t_l^w \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_c^l & t_c^l \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p^c \\ 1 \end{bmatrix}, \tag{2}$$

where T_c^l and T_l^w are the transformation matrices from the camera coordinate system to the local coordinate system and from the local coordinate system to the world coordinate system, respectively. T_c^l can be obtained from the local visual self-motion estimation in Section 3. T_l^w is the result of aligning the camera frame with the GNSS trajectory, which will be introduced in the GNSS-VO initialization stage in Section 3.

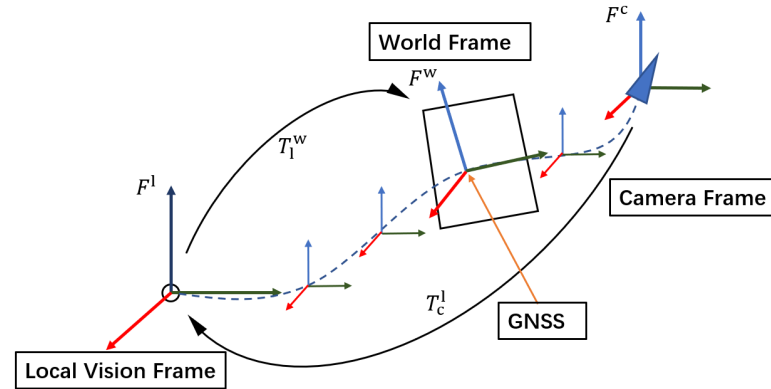


Figure 2. Coordinates involved in vehicle motion and the transformation matrices between them.

3. Frontend Data Preprocessing and GNSS-VO Initialization

3.1. Visual SLAM

The local visual system takes stereo visual image data as input and preprocesses it by computing the disparity between the two images to obtain depth information and extract the positions of salient feature points. Then, the current camera pose and the position of the feature points are used as variables, and the reprojection error is used as the objective function to optimize the camera pose by minimizing the objective function. This part of the work builds upon the ORB-SLAM2 algorithm as a development foundation. For the completeness of the system, it will be briefly introduced in this section. The specific working principle can be referred to [12].

The optimization formula of only visual BA algorithm can be expressed as minimizing reprojection error, and the specific formula is as follows:

$$X = \arg \max_X \sum_i \rho \left(\left\| x_{p_i}^l - \pi_s(P_i, X) \right\|_{\Sigma}^2 \right), \tag{3}$$

where $X = \{R, t, x^i\}$ represents the variables to be optimized, which include the current camera pose and the positions of the 3D feature points; i represents the number of feature; ρ is the robust Huber cost function; P_i represents the position information of the landmark point matched with the i -th feature point; $x_{p_i}^l$ represents the actual image position of the i -th feature point; $\pi_s(P_i, X)$ represents the projection model, which refers to the projected position of the landmark point P_i on the camera image plane. Σ is a covariance matrix related to the scale of the feature points. Then, a nonlinear optimization approach is used to iteratively optimize the camera pose and the positions of the 3D feature points, aiming to obtain the optimal pose estimation for the current camera frame. Finally, select key frames for the next step of global fusion.

3.2. GNSS Data Preprocessing

In GNSS data processing, the RTK receiver obtains latitude and longitude information and altitude information. The observed positions of the satellites are in the Latitude-Longitude-Altitude (LLA) coordinate system. Therefore, it is necessary to convert these positions to the East-North-Up (ENU) coordinate system. First, it is necessary to determine the latitude, longitude, and altitude information of the observation point. Then, these coordinates need to be converted into position vectors in the Earth-Centered Earth-Fixed (ECEF) coordinate system. Finally, the position vector in the ECEF coordinate system

needs to be transformed into a position vector in the local ENU coordinate system. The transformation formula is as follows:

$$\begin{bmatrix} x_{ENU} \\ y_{ENU} \\ z_{ENU} \end{bmatrix} = R_{\phi,\theta,\psi} \begin{bmatrix} x_{ECEF} - x_{ref} \\ y_{ECEF} - y_{ref} \\ z_{ECEF} - z_{ref} \end{bmatrix}, \tag{4}$$

where $x_{ref}, y_{ref}, z_{ref}$ represent the reference point of the ECEF coordinate system; $R_{\phi,\theta,\psi}$ is the rotation matrix; ϕ is the geographic latitude, θ is the geographic longitude, and ψ is the angle between the true north direction and ENU. In [43], the form of the rotation matrix is derived as follows:

$$R_{\phi,\theta,\psi} = \begin{bmatrix} -\sin \theta & \cos \theta & 0 \\ -\sin \phi \cos \theta & -\sin \phi \sin \theta & \cos \phi \\ \cos \phi \cos \theta & \cos \phi \sin \theta & \sin \phi \end{bmatrix}, \tag{5}$$

3.3. GNSS-VO Time Alignment

In a multi-sensor system, time synchronization of different sensors is a crucial step for data fusion and processing. The data acquisition frequency and processing methods may vary among different sensors, leading to deviations or drift in the timestamps of the data. To achieve time alignment between GNSS and Visual Odometry (VO), this paper adopts a timestamp-based approximate matching for time synchronization. By comparing the time difference between the GNSS signal and the local VO, if the difference is smaller than a certain threshold, the GNSS signal timestamp is considered equal to the keyframe timestamp; otherwise, the signal will be discarded.

3.4. GNSS-VO Initialization

In order to fuse GNSS with visual information, it is necessary to first ensure that the local coordinate system is converted to the global coordinate system. Due to various factors such as signal obstruction or multi-path effects, the accuracy of GNSS measurements may be limited. When aligning with a relatively short trajectory length, the alignment process can be more susceptible to errors due to the limited number of data points available. To obtain more accurate transformation estimation, this study utilizes a longer trajectory length for the alignment operation. This is achieved by constructing a data set with an extended sliding window, which includes a set of keyframes and synchronized GNSS position measurements. Once the desired trajectory length is reached, a nonlinear optimization problem is formulated to align the local VO trajectory with the global ENU frame. By incorporating a larger number of synchronized measurements over an extended trajectory, the alignment process can achieve improved accuracy in the transformation estimation.

Given a set of GNSS position measurements $P^w = \{p_1^w, p_2^w, \dots, p_n^w\}$ in the ENU frame within the sliding window, and the local VO frame $P^l = \{p_1^l, p_2^l, \dots, p_n^l\}$, at this time, it is necessary to find a rotation matrix R_l^w and translation vector t_l^w , to align the local VO trajectory with the ENU frame. By constructing a least squares problem, the sum of squared errors is minimized:

$$E(R_l^w, t_l^w) = \frac{1}{n} \sum_{i=0}^n \|p_i^w - (R_l^w p_i^l + t_l^w)\|^2, \tag{6}$$

According to [44], first, use the SVD principle to solve the rotation matrix R, and define the centroid of the two groups of points as:

$$p^l = \frac{1}{n} \sum_{i=1}^n p_i^l, \quad p^w = \frac{1}{n} \sum_{i=1}^n p_i^w, \tag{7}$$

Then construct a 3×3 matrix W such that:

$$W = \sum_{i=1}^n (p_i^w - p^w) (p_i^l - p^l)^T, \quad (8)$$

And perform SVD decomposition on W to obtain:

$$W = U\Sigma V^T, \quad (9)$$

where Σ is a diagonal matrix composed of singular values, and its diagonal elements are arranged from large to small; while U and V are diagonal matrices. When W is full rank, the solution formula for R_i^w is:

$$R_i^w = UV^T, \quad (10)$$

The final transformation t_i^s can be obtained using the following equation:

$$t_i^w = p^w - R_i^w p^l, \quad (11)$$

After system initialization, the transformation of the entire map's keyframes to the world coordinate system can be achieved through the rotation matrix R_i^w and the translation matrix t_i^w , to further adjust the GNSS and visual positioning results.

4. Gaussian Mixture Models and Online Parameters Estimation

4.1. Multivariate Gaussian Distribution and Factor Graph

Factor Graph is a graph structure used to represent probabilistic models and is an extension of probabilistic graphical models. In a factor graph, an optimization problem is described, aiming to find the optimal set of states X that best explains the observed measurements Z . It can usually be regarded as a probability distribution function for a set of conditional variables, which can be expressed as:

$$P(X|Z), \quad (12)$$

In order to estimate the conditional distribution of state variables, the probability distribution function can be expressed using Bayes' theorem as [23]:

$$P(X|Z) = \frac{P(Z|X)P(X)}{P(Z)} \propto P(Z|X)P(X), \quad (13)$$

In the formula, the left side $P(X|Z)$ is called the posterior probability, which represents the probability of the state set X given the measurement value Z ; the right side $P(Z|X)$ is called the likelihood, and $P(X)$ is called the prior probability, representing the probability of the state set X without considering any measurements. According to Bayes' theorem, solving for the maximum posterior probability is equivalent to maximizing the product of likelihood and prior. However, in general, prior information is unknown, so there is no prior. In order to obtain an optimal estimate that maximizes the posterior probability, that is:

$$X^* = \arg \max_X P(X|Z) \propto \arg \max_X P(Z|X), \quad (14)$$

By taking the negative logarithm, the general optimization problem can be rewritten as:

$$X^* = \arg \min_X \sum_n -\ln(P(z_n|x_n)), \quad (15)$$

The subscript n represents the index of all factors in the factor graph model. For a model with k high-dimensional Gaussian components, the normal distribution is represented

as $\mathcal{N}(\mu_k, \Sigma_k)$, where μ_k and Σ_k correspond to the mean and covariance matrix of the k -th Gaussian variable, respectively. The expanded form of its probability density function is:

$$P(z_n|x_n) = c_k \cdot \exp\left(-\frac{1}{2}\left\|\Gamma_k^{\frac{1}{2}}(e_n - \mu_k)\right\|^2\right) \tag{16}$$

with $c_k = \det\left(\Gamma_k^{\frac{1}{2}}\right)$ and $\Gamma_k^{\frac{1}{2}} = \Sigma_k^{-\frac{1}{2}}$

where e_n is the residual of the measurement value z_n , and \mathcal{L}_k is the information matrix corresponding to the measurement. By substituting Equation (16) into Equation (15), we obtain the general formula for the least squares problem of multivariate Gaussian distribution:

$$X^* = \arg \min_X \sum_n \frac{1}{2} \left\|\Gamma_k^{\frac{1}{2}}(e_n - \mu_k)\right\|^2, \tag{17}$$

As mentioned earlier, in orchard environments, GNSS measurements are affected by factors such as multi-path effects, making the GNSS noise multi-modal. This means that the GNSS noise does not follow the assumption of a Gaussian distribution. Therefore, it is crucial to use a more advanced distribution to describe the GNSS noise.

4.2. Gaussian Mixture Model and Factor Graph

To mitigate the impact of GNSS outliers, this paper chooses Gaussian Mixture Model (GMM) to describe these outliers. GMM is capable of flexibly modeling various types of data distributions, including multi-modal distributions and asymmetric distributions. By combining multiple Gaussian distributions, GMM can approximate any complex probability distribution. In [40], the authors demonstrated the feasibility of using a multi-modal Gaussian mixture model to fit the GNSS error distribution. This section will show how GMM fits the least squares problem. Define a k -component GMM:

$$P \sim \sum_k w_k \cdot \mathcal{N}(\mu_k, \Sigma_k) \quad \text{with} \quad \sum_{k=1}^k w_k = 1, \tag{18}$$

where w_k represents the weight of each Gaussian component. Based on the work in [45], a method for applying any non-Gaussian distribution to the least squares problem is shown:

$$X^* = \arg \min_X \sum_n \left\|\sqrt{-\ln\left(\frac{P(z_n|x_n)}{\gamma_s}\right)}\right\|^2 \tag{19}$$

with $\gamma_s \geq \max_n((z_n|x_n))$

γ_s is a normalization constant used to keep the negative log-likelihood positive. Therefore, the probability function of a k -component GMM can be expressed as:

$$P(z_n|x_n) = \sum_k c_k \cdot \exp\left(-\frac{1}{2}\left\|\Gamma_k^{\frac{1}{2}}(e_n - \mu_k)\right\|^2\right) \tag{20}$$

with $c_k = w_k \cdot \det\left(\Gamma_k^{\frac{1}{2}}\right)$

Here, the normalization constant can be defined as $\gamma_s = \sum_k c_k$. From this, we obtain the general formula for the least squares problem of the GMM:

$$X^* = \arg \min_x \sum_n \left\|\sqrt{-\ln\left(\frac{\sum_k c_k \cdot \exp(\dots)}{\gamma_s}\right)}\right\|^2, \tag{21}$$

Therefore, the formula (21) allows us to describe GNSS errors as non-Gaussian distributions and apply this non-Gaussian characteristic to the factor graph state estimation problem.

4.3. Parameter Estimation Online

In the estimation problem, a GMM with k components is used to describe the measurement error $e(x, z)$ and approximate the true distribution of the measurements. While estimating the system state, it is also necessary to estimate the parameters $\Theta_k = \{w_k, \mu_k, \Gamma_k\}$ of the k -component Gaussian Mixture Model. The parameter set Θ_k includes the weight w_k , mean μ_k , and information matrix Γ_k for each component k .

$$\Theta^* = \arg \max_{\Theta} P(\Theta|e), \quad (22)$$

In nonlinear optimization, applying a GMM to describe residuals and performing online estimation of model parameters is a novel robust method. Variational Bayesian Inference (VBI) is a method used for approximate inference in probabilistic models and is commonly applied to handle complex probabilistic graphical models. The goal of VBI is to approximate the true posterior distribution by finding an approximate distribution. In Bayesian inference, we want to calculate the posterior probability distribution $P(\Theta|e)$, where Θ a set of hidden variables and e the error value. However, in general, the posterior distribution cannot be calculated directly, so variational inference is used to approximate the posterior. The key idea of VBI is to introduce a parameterized distribution $Q(\Theta)$ to approximate the true posterior distribution. By minimizing the difference between the two distributions, an approximate posterior distribution can be obtained. For more detailed introduction, please refer to paper [46].

In the work of [41], the authors introduced an algorithm that combines VBI with factor graphs. Similarly, by performing alternating iterations of the E-step and M-step, it can be used to estimate the parameters of a probabilistic model with hidden variables. In the E-step, the VBI algorithm is used to calculate the posterior probability of hidden variables based on the current measurement error e . The goal of the E-step is to calculate the expectation of the hidden variables $E(\Theta_t^*)$ for subsequent parameter updates.

$$\Theta_t^* = \arg \max_{\Theta} P(\Theta|e_{t-1}), \quad (23)$$

In the M-step, based on the expectation of the hidden variables calculated in the E-step, the model parameters are updated to maximize the likelihood function of the complete data:

$$\begin{aligned} X_t^* &= \arg \max_X P(X|Z_t, E(\Theta_t^*)) \\ e_t^* &= e(X_t^*, Z_t) \end{aligned} \quad (24)$$

Performing alternating iterations of the E-step and the M-step continues until a convergence criterion is met. Typically, the convergence criterion can be that the parameters exhibit minimal changes or that the gain of the auxiliary function is small.

5. Global Optimization

In global optimization, an adaptive sliding window is used, which takes two sets of observation data as input: $Z^L = \{p_i^w, q_i^w\}$ from the local system and $Z^G = \{p_i^w\}$ from the global system. Both sets of data are time-aligned and transformed to the world coordinate system (ENU), denoted as L and G respectively. The adaptive sliding window in this paper can adjust the window size according to the number of valid GNSS measurements. When there is sufficient data within the window, a global optimization pose graph is constructed as shown in Figure 3. The pose graph is composed of nodes and edges, where nodes

represent poses and edges represent constraint relationships between poses. Each node typically represents the system’s pose at a specific time, as shown below:

$$X = \{p_0, p_1, p_2, \dots, p_n\}$$

$$p_n = [p_n^w, q_n^w]$$
(25)

where X represents the state set within the sliding window. p_i represents the node, and p_n^w and q_n^w represent the system’s position and orientation (represented by quaternion) in the world coordinate system, respectively. Then, a cost function is constructed, and the optimal state set is estimated by minimizing the cost function:

$$X^* = \arg \min_X \sum_{i \in n} \left\{ \left\| r_i^L(\hat{z}_{i-1}, x) \right\|^2 + \left\| r_i^G(\hat{z}_i, x) \right\|^2 \right\},$$
(26)

where r^L and r^G are the residuals of the local and global systems, respectively; \hat{z} represents the sensor measurement value. Specifically, the relationship between pose nodes is described through local relative constraints and global absolute constraints. For a better explanation, we recommend readers to refer to [47].

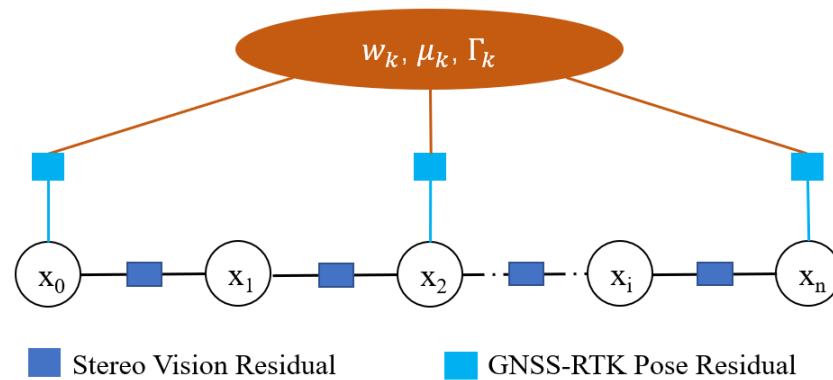


Figure 3. The factor graph model for GNSS/VO integration based on GMM is depicted below. The circles represent states, and the squares represent factors.

5.1. Local Visual Factor

This paper uses a binocular vision sensor as a local measurement, which can provide 6-DoF pose information. Since local visual estimation is accurate within a small interval, the camera motion between two adjacent moments is measured by visual odometry to obtain the relative pose difference. These relative pose differences can be used to construct local relative constraint edges of the pose graph:

$$r_i^L(\hat{z}_{i-1}, x) = \begin{bmatrix} p_i^w - p_{i-1}^w - q_{i-1}^w p_{i-1}^k \\ (q_{i-1}^w)^{-1} q_i^w \ominus q_{i-1}^i \end{bmatrix},$$
(27)

where p_{i-1}^i and q_{i-1}^i represent the pose relationship between the $i-1$ -th and i -th frames in the local system. \ominus is the negative operation on the error state of the quaternion.

5.2. GNSS-RTK Factor

In this paper, GNSS-RTK is used as an absolute sensor that provides absolute position information. Therefore, it can be used as a global reference to constrain the nodes in the pose graph. Due to the challenging GNSS environment, global measurements may be missing, which means that not all state vectors can be constrained by global constraints. Since the lower confidence level of GNSS measurements compared to visual odometry, this

paper only utilizes the latitude and longitude information from GNSS. Modeling GNSS measurement errors using GMM is typically represented as follows:

$$r_i^G(\hat{z}_i, x) = \sqrt{-\ln \left(\frac{\sum_{k=1}^N c_k \cdot \exp \left(-\frac{1}{2} \left\| \Gamma_k^{\frac{1}{2}} (e_i^G - \mu_k) \right\|^2 \right)}{\gamma_s} \right)} \quad (28)$$

with $e_i^G = z_i - h(x_i)$, $c_k = w_k \cdot \det \left(\Gamma_k^{\frac{1}{2}} \right)$

In the process of estimating the pose, the GMM model parameters $\theta_k = \{w_k, \mu_k, \Gamma_k\}$ are estimated simultaneously using the VBI algorithm. The pose graph is effectively established by integrating local relative constraints and GNSS absolute constraints, and the optimal solution of the pose nodes is estimated using the Ceres Solver [25] optimizer.

6. Results

6.1. Experiment Setup

The system is built upon the ORB-SLAM2 system with several improvements. A GNSS processing module is incorporated into the system, and GNSS nodes are inserted into an optimized graph with higher degrees of freedom. Additionally, GNSS noise is modeled using GMM, and GMM parameters are estimated using VBI. The implementation and parameter estimation of GMM are done using lisRSF [48]. Furthermore, to achieve visualization and debugging of robot systems, the Robot Operating System (ROS) is utilized. ROS provides a flexible communication mechanism that allows different components to integrate and collaborate.

To verify the performance of the system, we conducted a real-world vehicle tests in a arboretum of South China Agricultural University. The experimental platform used is an Akron robot based on ROS. The robot is equipped with a ZED2i stereo camera and a GNSS receiver operating in differential mode, please refer to Appendix A for the main parameters of the two sensors. Both sensors were mounted on the robot, and their relative poses remained fixed during the experiments, as shown in Figure 4a,b. Before the experiment, the coordinate systems between all sensors were calibrated. During the experiment, the ZED2i stereo camera was used to acquire left and right image information at a frequency of 30 Hz, and the GNSS receiver was used to collect differential GNSS measurements at a frequency of 20 Hz. Due to the severe obstruction and reflection of GNSS signals by surrounding trees and leaves, the accuracy of differential GNSS has significantly decreased, as shown in Figure 4c,d. The average GNSS residual reached 3.1042 m, and the maximum residual was 7.5285 m. Therefore, in this test, the differential GNSS measurements are not considered as ground truth but rather inserted as global observations into the optimization graph. For this purpose, 15 control points were set up, and an average of 10 min of differential measurement was performed for each point. Then, by specifying the trajectories of the control points, we tested the system's adaptability in a arboretum environment. All the algorithms in this test were performed on a ROS robot equipped with Jetson Xavier NX, and all the tested systems were running in real-time.

6.2. Positioning Performance

Considering that the positioning before loop closure is a good indicator of accuracy, this system uses GNSS absolute position correction instead of loop closure. To validate the performance of the proposed method, we also tested ORB-SLAM2 [12], vins-fusion [26], and a version of GNSS/VO without GMM modeling. The resulting trajectory comparison is shown in Figure 5a, the error comparison of all control points is shown in Figure 5b, and the statistical error comparison of all control points is shown in Table 1.

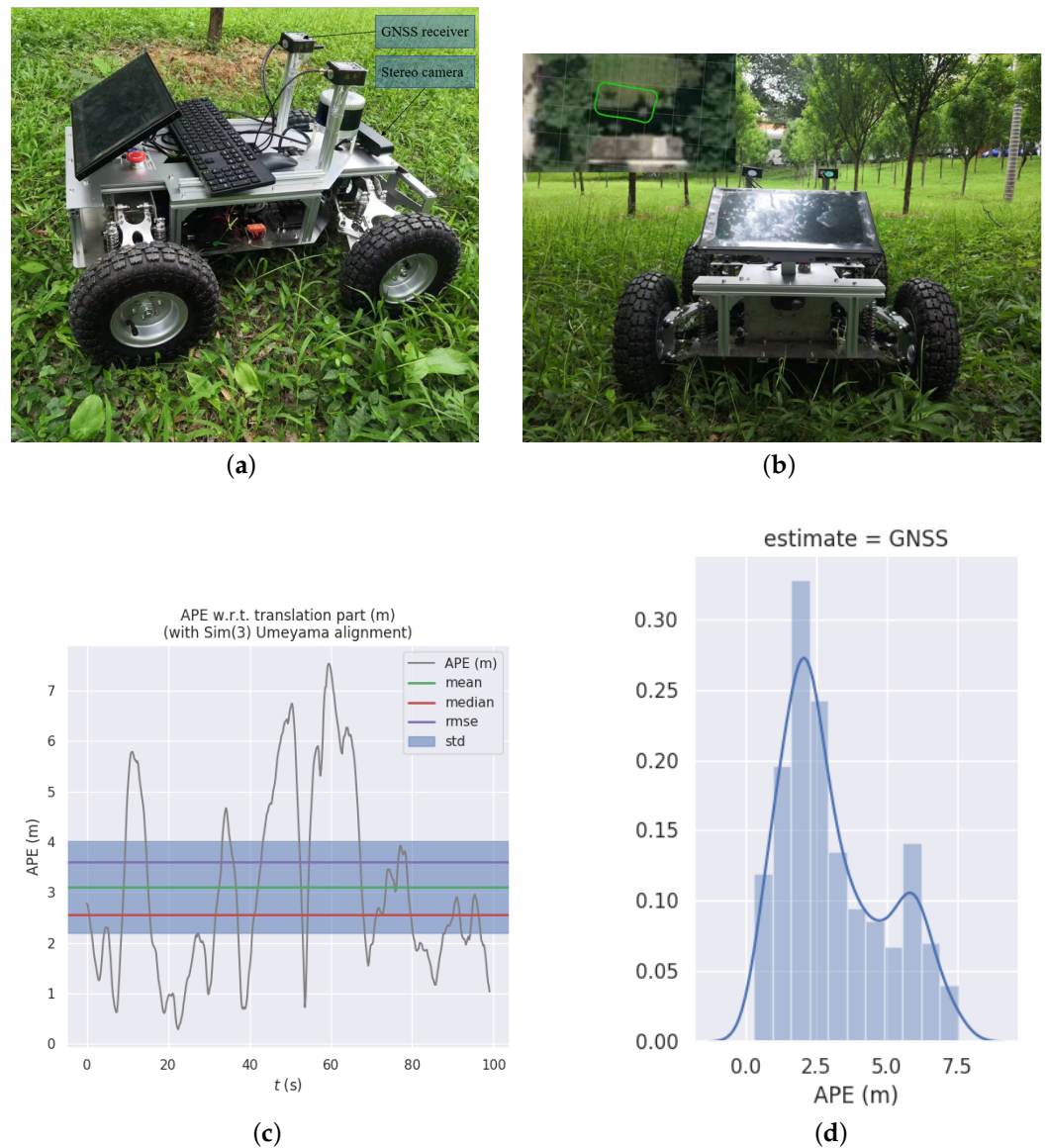


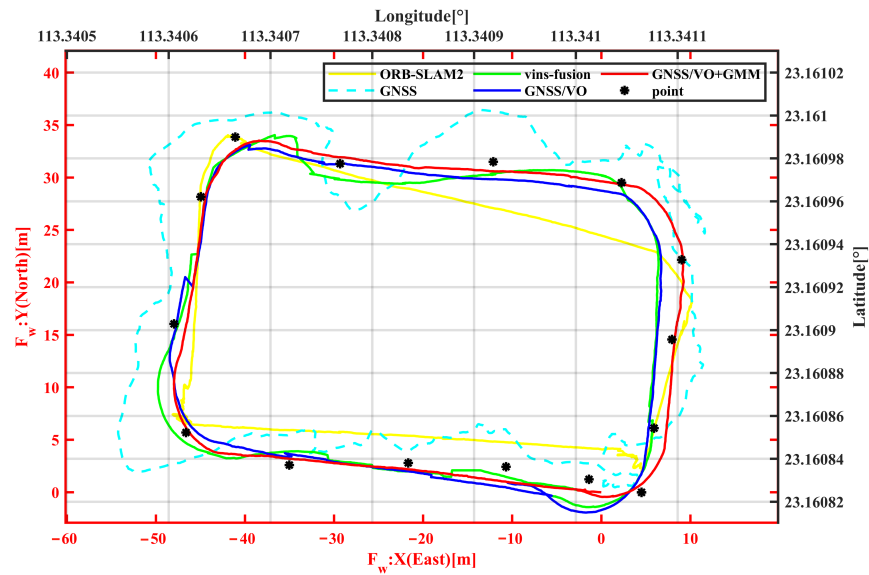
Figure 4. (a,b) The ground robot used in the experiment, as well as the real environment and trajectory in arboretum experiment. (c,d) The residual plot of GNSS-RTK positioning in the test and the proportional relationship of residuals.

From Table 1, it can be observed that due to signal obstruction and reflection, the average error and max error of the differential GNSS measurements are relatively large, with values of 2.1582 m and 4.7775 m, respectively. Interestingly, the performance of the pure visual system is slightly worse than GNSS, with an max error of 5.4348 m and Root Mean Square Error (RMSE) of 1.5579 m. We will discuss the reasons for this in the next section. In comparison, the overall performance of the GNSS and visual fusion method is better. Among them, vins-fusion has an average error of 1.1755 m and an RMSE of 0.7721 m, while GNSS/VO has an average error of 1.0869 m and an RMSE of 0.6112 m. By modeling GNSS errors with GMM, the adaptive GMM system further reduces the average error to 0.7380 m and the RMSE error to 0.4245 m.

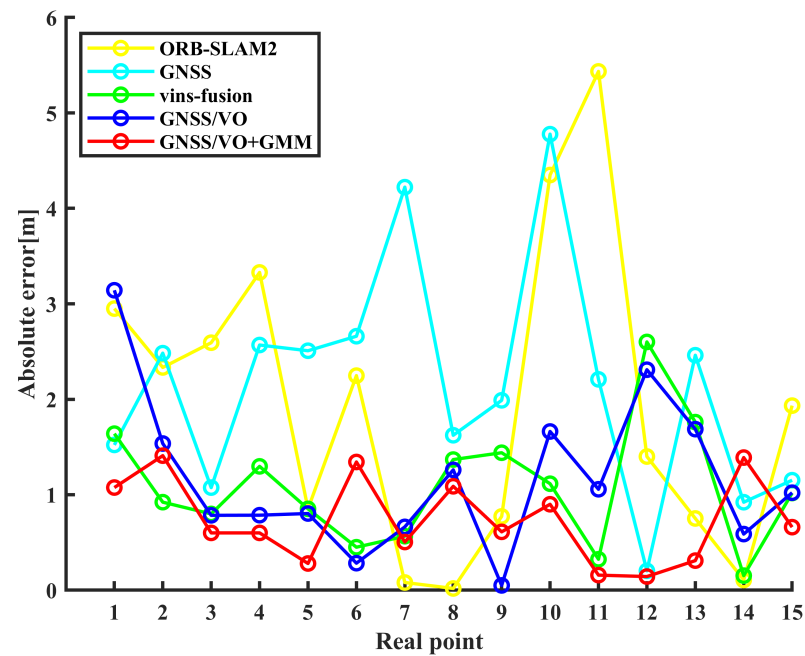
From Figure 5a,b, it can be seen that compared with the traditional GNSS/VO integration, the GNSS/VO system based on GMM modeling has lower positioning errors at most trajectory points. Moreover, the fusion result of GNSS and visual VO is smoother and closer to the real trajectory points. This effect is mainly due to the relative smooth pose constraint of vision, which can correct GNSS position measurements with obvious deviations.

Table 1. Comparison of localization errors for different algorithms at control points in real-world tests.

	ORB-SLAM2	GNSS	VINS-Fusion	GNSS/VO	GNSS/VO + GMM
RMSE	1.5579	1.1583	0.7721	0.6112	0.4245
Mean Error	1.9434	2.1582	1.1755	1.0869	0.7380
Max Error	5.4348	4.7775	3.1419	2.6007	1.4092



(a)



(b)

Figure 5. (a,b) Comparison of trajectories and positioning errors at control points for different algorithms.

6.3. Localization Performance

Considering that in the application, the camera will not access the same place twice, loop closure is not available. In addition, in many cases, we require real-time localization performance rather than localization after loop closure. Therefore, we use GNSS absolute

position measurements to correct visual accumulated errors instead of relying on loop closure. At the same time, we also tested the real-time positioning performance of our system with GNSS measurement constraints compared to ORB-SLAM2 without loop closure. This test selects the positioning results of the 45th and 65th seconds of the above trajectory for comparison, as shown in Figure 6a,b. Whenever the robot makes a turn, there is a significant directional offset, mainly due to the slow speed of the robot and the unevenness of the ground. This poses a significant challenge for localization. Therefore, when no loop closures are detected, the localization results of ORB-SLAM2 gradually deviate from the true path. However, in such cases, this algorithm demonstrates superior performance. With the constraints provided by GNSS, the system achieves high-precision and real-time localization. This further emphasizes that the improvements made to ORB-SLAM2 in this algorithm exhibit better and more stable performance.

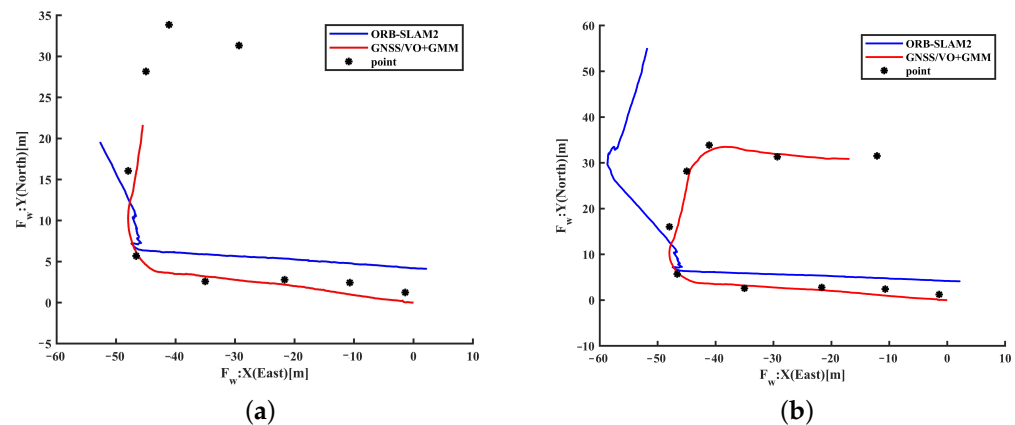


Figure 6. (a,b) Comparison of the localization trajectories between this algorithm and ORB-SLAM2 at the 45th and 65th seconds.

7. Discussion

In the previous section, we evaluated our algorithm based on real vehicle testing and conducted an analysis of the results. The experimental findings indicate that significant noise errors still exist when using RTK in environments with GNSS challenges. Additionally, when ORB-SLAM2 fails to detect loop closures, the positioning results gradually deviate from the actual path. However, the integration of GNSS with VO significantly improves the positioning accuracy. Compared to vins-fusion, GNSS/VO performs slightly better, benefiting from the optimization of the visual processing component and the construction of a global framework in our algorithm. Furthermore, by introducing GMM, GNSS/VO + GMM further enhances the positioning accuracy and exhibits improved performance. These results demonstrate that the integration of GNSS and visual odometry, along with the application of GMM graph optimization, leads to significant improvements in positioning accuracy.

In addition, the complex terrain in arboretums can cause rapid and drastic changes in targets, making time synchronization and state prediction between sensors more challenging and potentially leading to increased positioning errors. To overcome these limitations and challenges, we consider incorporating other sensors such as LiDAR and Inertial Measurement Units (IMU) to improve the robustness and accuracy of the positioning system. This is also one of our future directions.

8. Conclusions

Accurate localization in arboretums is an important prerequisite for autonomous navigation of agricultural robots. In this paper, an integrated GNSS and vision SLAM system is presented. In this system, binocular cameras and differential mode GNSS are applied and investigated. The fusion of GNSS and vision not only corrects the visual cumulative errors, but also smoothes the positioning results, effectively improving the

long-term positioning accuracy. In addition, this paper explores the combination of factor map and GMM, which is of great importance for sensor fusion. By combining GMM with GNSS/VO, the system benefits from the smoothing constraints provided by visual ranging and the accurate absolute positioning measurements of GNSS. This combination leads to improved accuracy and robustness in challenging environments, such as arboretums where signal obstructions and reflections can affect GNSS measurements. The results show that the proposed fusion method can significantly improve the positioning accuracy of the system with better performance. The integration of GMM with GNSS/VO in a tree garden environment provides higher accuracy and better robustness relative to other GNSS/VO fusion systems. In addition, the flexibility of the system allows the incorporation of other sensors, such as LiDAR or IMU, to further improve positioning and mapping capabilities. This allows the system to adapt to different environments and handle different types of terrain or obstacles.

Author Contributions: All named authors initially contributed a significant part to the paper. S.L. conceptualized the experiment, selected the algorithms, collected and analyzed the data, and wrote the manuscript. N.L. and Y.H. collected and analyzed data, and wrote the manuscript. W.Z. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: The open competition program of top ten critical priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (2022SDZG03) and Guangdong Provincial Science and Technology Innovation Strategy Special Funds Project (pdjh2022b0078).

Data Availability Statement: No applicable.

Acknowledgments: S.L. thank N.L. for supporting and helping, W.Z. for supervision, and Y.H. for revising the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GNSS	Global Navigation Satellite System
RTK	Real-Time Kinematics
GMM	Gaussian Mixture Model
VBI	Variational Bayesian Inference
SLAM	Simultaneous Localization and Mapping
EKF	Extended Kalman Filter
FGO	Factor Graph Optimization
MM	Maximum Mixture
EM	Expectation Maximization
NLoS	Non-Line-of-Sight
SfM	Structure from Motion
LLA	Latitude-Longitude-Altitude
ENU	East-North-Up
ECEF	Earth-Centered Earth-Fixed
SVD	Singular Value Decomposition
ROS	Robot Operating System
RMSE	Root Mean Square Error

Appendix A

In the experimental setup, the main parameters of different sensors are provided. These parameters include but are not limited to sampling frequency, camera resolution, and so on. By providing these key parameters, a better understanding of the experimental conditions and data synchronization between sensors can be obtained, which helps evaluate the performance and accuracy of the positioning system.

Table A1. Main parameters of GNSS-RTK sensor.

Sensors	GNSS-RTK
Sampling frequency	20 HZ
Operating temperature	−30° to +70°
Signal frequency	GPS: LIC/A,L2C BDS: BIL,B2L GLONASS: L1OF,L20F QZSS: LIC/A,L2C
NMEA-0183 protocol	Default GGA 115,200 baud rate

Table A2. Main parameters of stereo camera sensor.

Sensors	GNSS-RTK
Sampling frequency	30 HZ
Operating temperature	−10° to +50°
Output Resolution	1344 × 376 (WVGA)
Attitude drift	Translation: 0.35%; rotation: 0.005°/m

References

- Xu, Q.; Wang, M.; Du, Z.; Zhang, Y. A positioning algorithm of autonomous car based on map-matching and environmental perception. In Proceedings of the 33rd Chinese Control Conference, Nanjing, China, 28–30 July 2014; IEEE: Manhattan, NY, USA, 2014; pp. 707–712.
- Guan, Q.; Fan, C.; Chen, G.; Chen, C. Performance Evaluation of BDS/GPS Combined Single Point Positioning with Low-cost Single-Frequency Receiver. *J. Indian Soc. Remote Sens.* **2021**, *49*, 2847–2861. [[CrossRef](#)]
- Ferreira, A.; Matias, B.; Almeida, J.; Silva, E. Real-time GNSS precise positioning: RTKLIB for ROS. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1729881420904526. [[CrossRef](#)]
- Guvenc, I.; Chong, C.C. A survey on TOA based wireless localization and NLOS mitigation techniques. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 107–124. [[CrossRef](#)]
- Xia, X.; Meng, Z.; Han, X.; Li, H.; Tsukiji, T.; Xu, R.; Zheng, Z.; Ma, J. An automated driving systems data acquisition and analytics platform. *Transp. Res. Part C: Emerg. Technol.* **2023**, *151*, 104120. [[CrossRef](#)]
- Meng, Z.; Xia, X.; Xu, R.; Liu, W.; Ma, J. HYDRO-3D: Hybrid Object Detection and Tracking for Cooperative Perception Using 3D LiDAR. *IEEE Trans. Intell. Veh.* **2023**, *Early Acces.* [[CrossRef](#)]
- Liu, W.; Quijano, K.; Crawford, M.M. YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8085–8094.
- Sun, M.; Hua, C.; Pan, R. Binocular Ranging Based on Improved ORB-RANSAC. In Proceedings of the 2022 41st Chinese Control Conference (CCC), Hefei, China, 25–27 July 2022; IEEE: Manhattan, NY, USA, 2022; pp. 6216–6221.
- Lin, N.; Zhao, W.; Liang, S.; Zhong, M. Real-Time Segmentation of Unstructured Environments by Combining Domain Generalization and Attention Mechanisms. *Sensors* **2023**, *23*, 6008. [[PubMed](#)]
- Davison, A.J.; Reid, I.D.; Molton, N.D.; Stasse, O. MonoSLAM: Real-time single camera SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1052–1067. [[CrossRef](#)] [[PubMed](#)]
- Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [[CrossRef](#)]
- Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262.
- Wang, R.; Schworer, M.; Cremers, D. Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017, pp. 3903–3911.
- Vetrella, A.R.; Fasano, G.; Accardo, D.; Moccia, A. Differential GNSS and vision-based tracking to improve navigation performance in cooperative multi-UAV systems. *Sensors* **2016**, *16*, 2164. [[CrossRef](#)]
- Thrun, S. Probabilistic algorithms in robotics. *AI Mag.* **2000**, *21*, 93–93.
- Xia, X.; Hashemi, E.; Xiong, L.; Khajepour, A. Autonomous vehicle kinematics and dynamics synthesis for sideslip angle estimation based on consensus kalman filter. *IEEE Trans. Control Syst. Technol.* **2022**, *31*, 179–192. [[CrossRef](#)]
- Xia, X.; Xiong, L.; Lu, Y.; Gao, L.; Yu, Z. Vehicle sideslip angle estimation by fusing inertial measurement unit and global navigation satellite system with heading alignment. *Mech. Syst. Signal Process.* **2021**, *150*, 107290. [[CrossRef](#)]
- Xiong, L.; Xia, X.; Lu, Y.; Liu, W.; Gao, L.; Song, S.; Yu, Z. IMU-based automated vehicle body sideslip angle and attitude estimation aided by GNSS using parallel adaptive Kalman filters. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10668–10680. [[CrossRef](#)]
- Liu, W.; Xia, X.; Xiong, L.; Lu, Y.; Gao, L.; Yu, Z. Automated vehicle sideslip angle estimation considering signal measurement characteristic. *IEEE Sens. J.* **2021**, *21*, 21675–21687.
- Dusha, D.; Mejias, L. Error analysis and attitude observability of a monocular GPS/visual odometry integrated navigation filter. *Int. J. Robot. Res.* **2012**, *31*, 714–737. [[CrossRef](#)]

21. Wei, L.; Cappelle, C.; Ruichek, Y.; Zann, F. Intelligent vehicle localization in urban environments using ekf-based visual odometry and gps fusion. *IFAC Proc. Vol.* **2011**, *44*, 13776–13781. [[CrossRef](#)]
22. Li, T.; Zhang, H.; Gao, Z.; Niu, X.; El-Sheimy, N. Tight fusion of a monocular camera, MEMS-IMU, and single-frequency multi-GNSS RTK for precise navigation in GNSS-challenged environments. *Remote Sens.* **2019**, *11*, 610. [[CrossRef](#)]
23. Dellaert, F.; Kaess, M. Factor graphs for robot perception. *Found. Trends[®] Robot.* **2017**, *6*, 1–139. [[CrossRef](#)]
24. Kümmerle, R.; Grisetti, G.; Strasdat, H.; Konolige, K.; Burgard, W. g 2 o: A general framework for graph optimization. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; IEEE: Manhattan, NY, USA, 2011; pp. 3607–3613.
25. Agarwal, S.; Mierle, K.; The Ceres Solver Team. Ceres Solver. 2012. Available online: <http://ceres-solver.org> (accessed on 26 July 2023).
26. Qin, T.; Cao, S.; Pan, J.; Shen, S. A general optimization-based framework for global pose estimation with multiple sensors. *arXiv* **2019**, arXiv:1901.03642.
27. Hou, Z.; Wang, R. A Loosely-Coupled GNSS-Visual-Inertial Fusion for State Estimation Based On Optimization. In Proceedings of the 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC), Greenville, SC, USA, 12–14 November 2021; IEEE: Manhattan, NY, USA, 2021; pp. 163–168.
28. Han, S.; Deng, F.; Li, T.; Pei, H. Tightly coupled optimization-based GPS-visual-inertial odometry with online calibration and initialization. *arXiv* **2022**, arXiv:2203.02677.
29. Gong, Z.; Ying, R.; Wen, F.; Qian, J.; Liu, P. Tightly coupled integration of GNSS and vision SLAM using 10-DoF optimization on manifold. *IEEE Sens. J.* **2019**, *19*, 12105–12117. [[CrossRef](#)]
30. Wen, W.; Pfeifer, T.; Bai, X.; Hsu, L.T. Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter. *NAVIGATION J. Inst. Navig.* **2021**, *68*, 315–331. [[CrossRef](#)]
31. Brach, M.; Zasada, M. The effect of mounting height on GNSS receiver positioning accuracy in forest conditions. *Croat. J. For. Eng. J. Theory Appl. For. Eng.* **2014**, *35*, 245–253.
32. Gong, Z.; Liu, P.; Wen, F.; Ying, R.; Ji, X.; Miao, R.; Xue, W. Graph-based adaptive fusion of GNSS and VIO under intermittent GNSS-degraded environment. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–16. [[CrossRef](#)]
33. Gakne, P.V.; O’Keefe, K. Tightly-coupled GNSS/vision using a sky-pointing camera for vehicle navigation in urban areas. *Sensors* **2018**, *18*, 1244. [[CrossRef](#)] [[PubMed](#)]
34. Reynolds, D.A. Gaussian mixture models. *Encycl. Biom.* **2009**, *741*, 659–663.
35. Sünderhauf, N.; Protzel, P. Switchable constraints for robust pose graph SLAM. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Algarve, Portugal, 7–12 October 2012; IEEE: Manhattan, NY, USA, 2012; pp. 1879–1884.
36. Agarwal, P.; Tipaldi, G.D.; Spinello, L.; Stachniss, C.; Burgard, W. Robust map optimization using dynamic covariance scaling. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; IEEE: Manhattan, NY, USA, 2013; pp. 62–69.
37. Agamennoni, G.; Furgale, P.; Siegwart, R. Self-tuning M-estimators. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; IEEE: Manhattan, NY, USA, 2015; pp. 4628–4635.
38. Pfeifer, T.; Lange, S.; Protzel, P. Dynamic Covariance Estimation—A parameter free approach to robust Sensor Fusion. In Proceedings of the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Daegu, Republic of Korea, 16–18 November 2017; IEEE: Manhattan, NY, USA, 2017; pp. 359–365.
39. Pfeifer, T.; Protzel, P. Robust sensor fusion with self-tuning mixture models. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Manhattan, NY, USA, 2018; pp. 3678–3685.
40. Pfeifer, T.; Protzel, P. Expectation-maximization for adaptive mixture models in graph optimization. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Macau, China, 3–8 November 2019; IEEE: Manhattan, NY, USA, 2019; pp. 3151–3157.
41. Pfeifer, T.; Protzel, P. Incrementally learned mixture models for GNSS localization. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; IEEE: Manhattan, NY, USA, 2019; pp. 1131–1138.
42. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
43. Zhu, J. Conversion of Earth-centered Earth-fixed coordinates to geodetic coordinates. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 957–961. [[CrossRef](#)]
44. Arun, K.S.; Huang, T.S.; Blostein, S.D. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *5*, 698–700. [[CrossRef](#)]
45. Rosen, D.M.; Kaess, M.; Leonard, J.J. Robust incremental online inference over sparse factor graphs: Beyond the Gaussian case. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; IEEE: Manhattan, NY, USA, 2013; pp. 1025–1032.
46. Corduneanu, A.; Bishop, C.M. Variational Bayesian model selection for mixture distributions. In *Proceedings of the Artificial Intelligence and Statistics*; Morgan Kaufmann: Waltham, MA, USA, 2001; Volume 2001, pp. 27–34.

47. Grisetti, G.; Kümmerle, R.; Stachniss, C.; Burgard, W. A tutorial on graph-based SLAM. *IEEE Intell. Transp. Syst. Mag.* **2010**, *2*, 31–43. [[CrossRef](#)]
48. Pfeifer, T. libRSF. Available online: <https://github.com/TUC-ProAut/libRSF> (accessed on 26 July 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.