

Article

Monocular Camera/IMU/GNSS Integration for Ground Vehicle Navigation in Challenging GNSS Environments

Tianxing Chu ^{1,*}, Ningyan Guo ², Staffan Backén ³ and Dennis Akos ³

¹ School of Earth and Space Sciences, Peking University, No. 5 Yiheyuan Road, Haidian District, Beijing 100871, China

² School of Instrumentation Science and Optoelectronics Engineering, Beihang University, Xueyuan Road No.37, Haidian District, Beijing 100191, China; E-Mail: ningyan.guo@gmail.com

³ Department of Aerospace Engineering Sciences, University of Colorado at Boulder, Boulder, CO 80309, USA; E-Mails: staffan.backen@ltu.se (S.B.); dma@colorado.edu (D.A.)

* Author to whom correspondence should be addressed; E-Mail: tianxing.chu@colorado.edu; Tel.: +1-303-492-4462.

Received: 27 January 2012; in revised form: 22 February 2012 / Accepted: 22 February 2012 / Published: 7 March 2012

Abstract: Low-cost MEMS-based IMUs, video cameras and portable GNSS devices are commercially available for automotive applications and some manufacturers have already integrated such facilities into their vehicle systems. GNSS provides positioning, navigation and timing solutions to users worldwide. However, signal attenuation, reflections or blockages may give rise to positioning difficulties. As opposed to GNSS, a generic IMU, which is independent of electromagnetic wave reception, can calculate a high-bandwidth navigation solution, however the output from a self-contained IMU accumulates errors over time. In addition, video cameras also possess great potential as alternate sensors in the navigation community, particularly in challenging GNSS environments and are becoming more common as options in vehicles. Aiming at taking advantage of these existing onboard technologies for ground vehicle navigation in challenging environments, this paper develops an integrated camera/IMU/GNSS system based on the extended Kalman filter (EKF). Our proposed integration architecture is examined using a live dataset collected in an operational traffic environment. The experimental results demonstrate that the proposed integrated system provides accurate estimations and potentially outperforms the tightly coupled GNSS/IMU integration in challenging environments with sparse GNSS observations.

Keywords: sensor integration; extended Kalman filter; GNSS; strapdown mechanization; computer vision; tightly coupled integration

1. Introduction

GNSS systems have been the prominent technology to fulfill the demanding requirements for road navigation. However, GNSS signals are subject to attenuation, reflections, or even blockages as a result of driving in difficult signal reception areas, such as urban canyons, enclosed parking garages and underground concrete tunnels. RF interference (RFI), whether deliberate or not, can misguide the user towards an erroneous position solution [1]. The use of augmentation technologies is capable of mitigating these disadvantages of standalone GNSS solutions and providing improved availability and accuracy. By coupling the measurements from both GNSS and IMU sensors, integration techniques are commercially employed to calculate a high-bandwidth navigation solution. This effectively enables the calibration of the IMU sensors and the control of the error growth of inertial navigation, and also enables higher data output compared to a standalone GNSS solution. A number of researchers have presented detailed realization methods and performance analyses of the loosely/tightly coupled GNSS/IMU implementations [2–4].

Apart from the traditional navigation technologies, the progress in computer vision has motivated a variety of location-based applications, such as simultaneous localization and mapping (SLAM) [5], and provided potential capabilities for the navigation community. Visual sensors usually incorporate other sensors, such as GNSS, IMU and laser scanners, to achieve a system level integration. Soloviev [6] provided an integration of GPS carrier phase and vision-based measurements for navigation in GPS challenged environments. The GPS carrier phase measurements were used to resolve the scale ambiguity of the unknown change in position vector. The algorithm was validated via simulations as well as live experiments.

In more common cases, a vision sensor is incorporated with an IMU sensor for motion estimation with high frequency. Corke [7] presented a tutorial overview on inertial and visual sensors and gave the essential principles of integrating them to obtain robust estimates of sensor motion and 3D reconstruction. The integration strategy is categorized into the loose and tight coupling. A detailed performance comparison of the loosely/tightly coupled camera/IMU integration was given by Chu [8]. The tightly coupled integration might yield higher accuracy, but it is prone to divergence under certain conditions and *vice versa*.

By utilizing strapdown inertial sensors and a fisheye lens, Schlaile [9] demonstrated a navigation system for an indoor vertical-take-off-and-landing (VTOL) micro aerial vehicle (MAV) based on the Kalman filter. The inverse depth method was applied to determine the movement of the camera. The scale factor ambiguity was recovered by simulating the laser ranging as additional source of information and measuring the flight height over ground. Apart from using a monocular camera, Soloviev [10] proposed a multi-aperture approach integrated with IMU measurements. By observing the feature points over a wide range, the vision-related navigation uncertainty could be significantly

reduced. The unit sphere on which the feature points were mapped was presented and used to model the sensor fusion algorithm.

As opposed to utilizing a forward-looking camera, Hide [11] used a camera with a ground-facing orientation to accomplish the camera/IMU integration for pedestrian navigation. Given the estimated height to the ground, the camera velocity was retrieved from the translation information. The residual between IMU and camera velocities then served as the Kalman filter's measurements. With a similar IMU/camera mounting method, Huang [12] provided a tightly coupled integration by using the residual between the predicted and actual feature location as the filter's input, which was expressed in the image coordinate frame.

Nowadays, low-cost MEMS-based IMUs, video cameras and portable GNSS devices are commercially available for automotive applications and some manufacturers have already integrated such facilities into their vehicle systems. This research aims at robust ground vehicle navigation based on camera/IMU/GNSS integration against challenging GNSS environments where the blocked sky can significantly obstruct the view of low-elevation-angle satellites and hinder a user's navigation performance.

Three key contributions are addressed herein: first, the proposed camera/IMU/GNSS integration architecture complements each of the sensors and potentially overcomes the disadvantage of traditional GNSS/IMU integration. For example, the number of line-of-sight (LOS) GNSS pseudoranges remains between two and three for a while such as wandering through the narrow streets in a city's downtown area. Although the tightly coupled GNSS/IMU integration allows the filter to obtain the user's position solution with sparse measurements, the overall solution is prone to divergence over time due to the lack of redundant constraints. On the other hand, with the aid of more than two LOS pseudorange observations from high-elevation-angle satellites, the baseline magnitude of position change within an allocated time interval can be extracted. This GNSS-derived baseline magnitude in turn enables a monocular camera to resolve its scale factor ambiguity if the data rates of GNSS and camera sensors are synchronized.

Second, although significant effort has been spent on developing visual odometry algorithms, the efficiency and performance regarding image frame rate has yet to be evaluated. The navigation information derived from the computer vision algorithms is, ideally, independent of frame rate, provided two adjacent image frames overlap each other. However, in reality the actual situation does not provide this independence due to the vehicle dynamics and sufficiency of the accurate image feature correspondences. In an effort to balance the tradeoff between practical performance and computational cost, we deliberately reduce the image frame rate and examine the positioning accuracy of our integrated scheme.

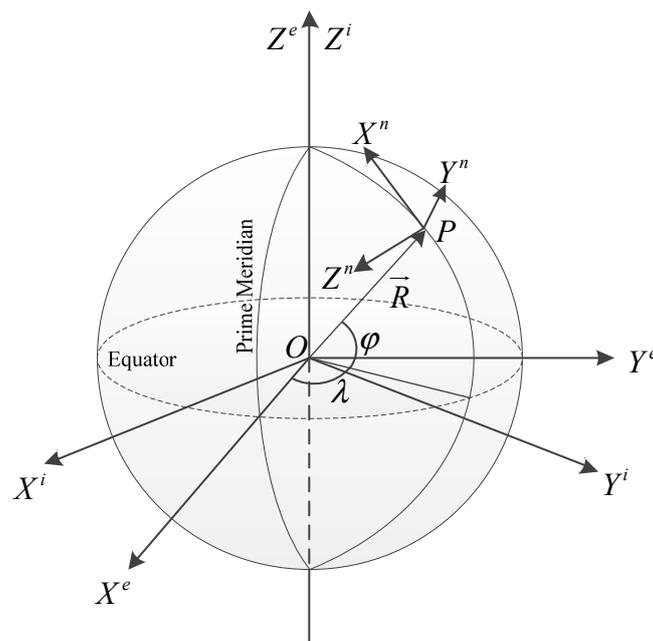
Third, the integrated system is validated through a live dataset which was collected in an operational traffic environment with available centimeter-accuracy truth reference. Although there are previous usable datasets for SLAM-based analysis, few of them were collected in the actual environment with available high-accuracy truth reference [13,14]. The camera images we used were captured in an operational traffic environment with vehicles passing by and overtaking the host vehicle during the test period. The high-accuracy truth reference was generated by leveraging three RTK GPS receivers on the vehicle roof top, which facilitates a fair and reliable comparison between algorithms under assessment. The only simulation in this paper is the GNSS pseudorange measurements.

The remainder of the paper proceeds as follows: we first introduce the coordinate frames utilized in this paper. The inertial navigation and its error model are then summarized. Thirdly, the principle of computer vision-based motion estimation is presented. GNSS simulation and differential processing are discussed afterwards. Based on the above sensor description, the proposed EKF model and integrated architecture are established, followed by the live test validation. The paper concludes with the results as well as planned future work.

2. Coordinate Frames

Figure 1 illustrates three major coordinate frames on the Earth ellipsoid. The inertial coordinate frame (i frame) $O-X^iY^iZ^i$ is the fundamental reference frame for inertial navigation and an IMU enclosure usually measures the specific force as well as the angular rate relative to the i frame. Its origin is on the Earth's center of mass O and it is stationary relative to the fixed stars. We usually express the ultimate position solution, latitude φ , longitude λ and height h , in the Earth frame (e frame) $O-X^eY^eZ^e$. Its origin is the Earth's center of mass O , with Z^e axis forming along polar axis and X^e axis lying on the intersection of the equatorial plane and prime meridian plane. The Y^e axis is also on the equatorial plane and satisfies the right-hand rule. The origin and Z -axes of the i and e frames are coincident, respectively. In this paper, the e frame-based position solutions are based on the WGS-84 ellipsoid. To describe the velocity as well as the orientation of a ground vehicle, the right-handed navigation frame (n frame) $P-X^nY^nZ^n$ is utilized and locally placed at the user's position P . The Z^n axis is collinear with the local normal line of the reference ellipsoid pointing downwards while X^n and Y^n axes indicate the local north and east direction, respectively.

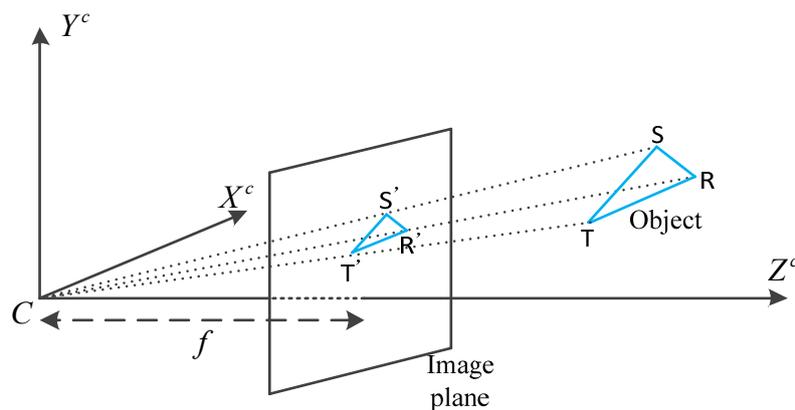
Figure 1. Inertial frame, earth frame and navigation frame representation on the Earth ellipsoid.



The specific force f_{ib}^b and angular rate ω_{ib}^b from an IMU are usually measured relative to the i frame and expressed in the body frame (b frame). The b frame has its origin at the center of the IMU

enclosure. If the IMU is mounted parallel to the vehicle frame, we allocate the X^b axis pointing forward, in view of the vehicle, and Z^b axis aligning with local gravity direction. The Y^b axis satisfies the right-hand rule and indicates the right-side direction of the vehicle. Based on the pinhole projection model in our research, the image feature points are expressed in the camera frame (c frame) which is extended from the 2D image plane. The c frame representation is shown in Figure 2. The c frame's origin is at the camera center C , and X^c as well as Y^c axes point towards left and upward directions of the image plane, correspondingly. The Z^c axis lies along the principal axis and is orthogonal to the image plane. f denotes the focal length of the camera.

Figure 2. Representation of the camera frame with respect to the pinhole projection model.



In this paper, the vehicle dynamics model is established in terms of the b frame instead of other frames. According to Hol [15], the b frame based motion estimate has less modeling complexity and is prone to eliminating the influence of angular and position installation errors.

3. Inertial Navigation and IMU Error Model

A six-degree-of-freedom (6-DOF) strapdown IMU provides the user high-bandwidth position, velocity and orientation estimation with complete independence of the reception of electromagnetic waves as following differential equations [16]:

$$\dot{C}_b^n = C_b^n \cdot (\omega_{ib}^b \times) - [(\omega_{ie}^n \times) + (\omega_{en}^n \times)] \cdot C_b^n \quad (1)$$

$$\dot{V}^n = C_b^n f_{ib}^b - [2(\omega_{ie}^n \times) + (\omega_{en}^n \times)] \cdot V^n + g^n(\varphi, h) \quad (2)$$

$$\dot{P}^e = \text{diag}\left(\frac{1}{R_M + h}, \frac{1}{(R_N + h)\cos\varphi}, -1\right) \cdot V^n \quad (3)$$

In Equation (1), C_b^n is the direction cosine matrix representing the rotation between the b and n frames. The variables ω_{ie}^n and ω_{en}^n are the Earth's rotation rate and body transport rate, respectively. The symbol \times denotes the skew symmetric matrix of a specific angular rate term. The calculated C_b^n is then used to estimate the acceleration in terms of the n frame by multiplication with measured specific force f_{ib}^b . As described in Equation (2), Coriolis force, body transport rate and local gravity terms are compensated in the estimation. Although the local gravity changes with latitude φ and height h , the insensitivity can be ignored for a ground vehicle. In Equation (3), the position update is expressed in

the e frame and therefore needs R_M and R_N , radii of curvatures in the meridian and prime vertical, to be estimated.

Although a strapdown IMU propagates the high-frequency inertial navigation mathematically derived from Equations (1) to (3), the MEMS accelerometer and gyroscope sensors inevitably drift and give rise to quadratic and cubic accumulated mechanization errors. Researchers have developed error propagation models in various approaches for practical applications such as [17–19]. To describe a ground vehicle in low dynamics with well-calibrated a priori information, the Ψ angle model [20] is introduced in this paper on account of simplicity:

$$\delta \dot{P}^n = -(\omega_{en}^n \times) \delta P^n + \delta V^n \quad (4)$$

$$\begin{aligned} \delta \dot{V}^n = & \text{diag}\left[-\frac{g}{R_e} \quad -\frac{g}{R_e} \quad \frac{2g}{R_e}\right] \delta P^n - [(2\omega_{ie}^n + \omega_{en}^n) \times] \delta V^n \\ & + [(C_b^n f_{ib}^b) \times] \delta \Psi^n + C_b^n \delta f^b \end{aligned} \quad (5)$$

$$\delta \dot{\Psi}^n = -(\omega_{in}^n \times) \delta \Psi^n - C_b^n \delta \omega^b \quad (6)$$

In the above equations, R_e is the Earth radius; δP^n , δV^n , $\delta \Psi^n$, δf^b and $\delta \omega^b$ are the perturbation terms of position, velocity, orientation and sensor biases, respectively.

As noted earlier, this paper aims to efficiently integrate data from camera, IMU and GNSS sensors to achieve high accuracy ground vehicle navigation in challenging GNSS environments. Therefore, a model for reasonably calibrating the IMU sensor biases needs to be established considering the practical situation. A commonly cited first-order Gauss-Markov model regards the IMU errors as 3 independent components: a random constant, a Gauss-Markov variable and a white noise term and therein the accelerometer/gyroscope error propagations are written as following equations [16,18]:

$$\delta f = \delta f_c + \delta f_m + w_f \quad \delta \omega = \delta \omega_c + \delta \omega_m + w_\omega \quad (7)$$

The subscripts of c and m indicate the random constant and first-order Gauss-Markov terms, and their derivatives yield:

$$\delta \dot{f}_c = 0 \quad \delta \dot{\omega}_c = 0 \quad (8)$$

$$\delta \dot{f}_m = -\frac{\delta f_m}{T_f} + \mu_f \quad \delta \dot{\omega}_m = -\frac{\delta \omega_m}{T_\omega} + \mu_\omega \quad (9)$$

where the constants T_f and T_ω are the correlation times; w_f , w_ω , μ_f and μ_ω are assumed uncorrelated and correlated white Gaussian noise (WGN) for inertial sensors. Equations (4) to (9) construct the fundamental linearized dynamic system of the EKF integrated system which will be further discussed in Section 6.

4. Motion Estimation in Computer Vision

The camera requires calibration prior to the computer vision routine for motion estimation. To effectively estimate the intrinsic camera parameters as well as distortion characteristics of the images, researchers have developed practical calibration methods such as [21–23] and there exist open source toolkits for calibration-related tasks. Normally, the camera calibration parameters do not change

rapidly over time, moreover, the run-to-run calibration differences are even negligible within a limited period of time, provided the camera is not reconfigured. In addition, as opposed to an IMU, a temperature-independent CMOS/CCD sensor makes the camera resistant to ambient environmental change. Therefore, compared to the inertial sensors, the camera exhibits different calibration mechanism and characteristics coupled with a diverse model for long-term stability.

As mentioned previously, the perspective projection with a pinhole camera model is used. To estimate the motion in successive image frames, the computer vision module executes three main functions: (1) feature extraction and matching; (2) outlier removal and discrimination of the moving features; (3) motion estimation.

4.1. Feature Extraction and Matching

Feature extraction is the essential process for subsequent motion estimation. As an outdoor environment usually has a high-contrast texture in the image scene, image points with distinct structural information are selected as features. In computer vision community, the scale-invariant feature transform (SIFT) algorithm [24] is a well-accepted approach to describe and match the feature points which are usually invariant to distortion, scaling, orientation, affine transformation, and illumination changes. Although several faster algorithms, such as PCA [25] and SURF [26], facilitate reducing the computational load, the point location accuracy can be in turn degraded [27]. We, consequently, use the SIFT algorithm to locate the feature points and create the corresponding scale-invariant descriptor vectors by measuring the gradient magnitude and orientation in the pre-defined pixel neighborhood. To provide sufficient dimensionalities in the feature space for favorable matching result, a 128-dimensional descriptor:

$$\vec{d} = (d_1, d_2, \dots, d_{128})^T \quad (10)$$

is used for each pre-selected feature point. In adjacent images, the similarity of two candidate feature points, A and B , is given by the Euclidean distance:

$$l_{AB} = \|\vec{d}_A - \vec{d}_B\| = \sqrt{\sum_{i=1}^{128} \delta d_i^2} \quad (11)$$

By setting the empirically examined accept-reject threshold, the feature matching process is evaluated using the ratio of the minimum distance to the next minimum [24]. Ideally, if an image scene contains a non-uniform texture, there should be a large amount of SIFT features which guarantee desirable redundancy for optimized motion estimation. However, there inevitably exist outliers due to the limited descriptor dimensionality and the imperfect matching determination criteria. Therefore, an outlier removal scheme is undertaken prior to the motion determination process. In order to remove the outlier correspondences which potentially give rise to undesirable estimation, the random sample consensus (RANSAC) algorithm [28] is taken into consideration in this paper.

4.2. Outlier Removal by RANSAC

Based on the primary principle of epipolar geometry, all corresponding image points satisfy the coplanarity equation:

$$x'^T F x = 0 \quad (12)$$

where F is a 3×3 fundamental matrix which geometrically describes the projection between two corresponding points in a pair of images representing the same 3D object; x and x' are the coordinates of the points in image planes, respectively. It is supposed that RANSAC algorithm estimates the fundamental matrix F which optimally fits all the inliers. However, RANSAC may miss out a few outliers and estimate a biased F matrix due to the improper number of iteration operations and a relatively large number of outliers. Consequently, a multilayer RANSAC scheme [29] is proposed and implemented as follows:

- (1) arbitrarily selecting a subset samples from all pre-selected matching features;
- (2) reconstructing epipolar geometry constraint satisfying Equation (12) and computing the fundamental matrix F based on the selected samples;
- (3) using the same constraint and the calculated matrix F to determine the error from all features;
- (4) with a proper threshold, categorizing all feature correspondences as inliers or outliers by appraising the error statistics;
- (5) iterating the first three steps multiple times (refer to [28] for a favorable number of iterations);
- (6) finding the best-fit matrix F which yields the optimal error estimates and removing the outliers discriminated by this matrix F ;
- (7) repeating the first six steps with a gradually shrunken inlier/outlier determination threshold.

To test the multilayer RANSAC scheme, we use a sequence of images (2.5 Hz rate) which will be elaborated in Section 7. Figure 3 shows the number of originally matched features as well as that after implementing multilayer RANSAC. It is clearly seen that after the fourth layer of RANSAC processing, there are still hundreds of matched features for every moment to maintain observable redundancy for motion estimation.

Figure 3. Number of matched features after implementing multilayer RANSAC.

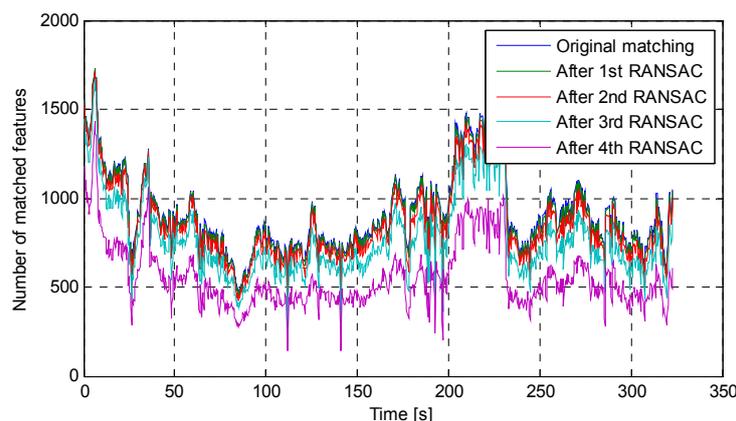
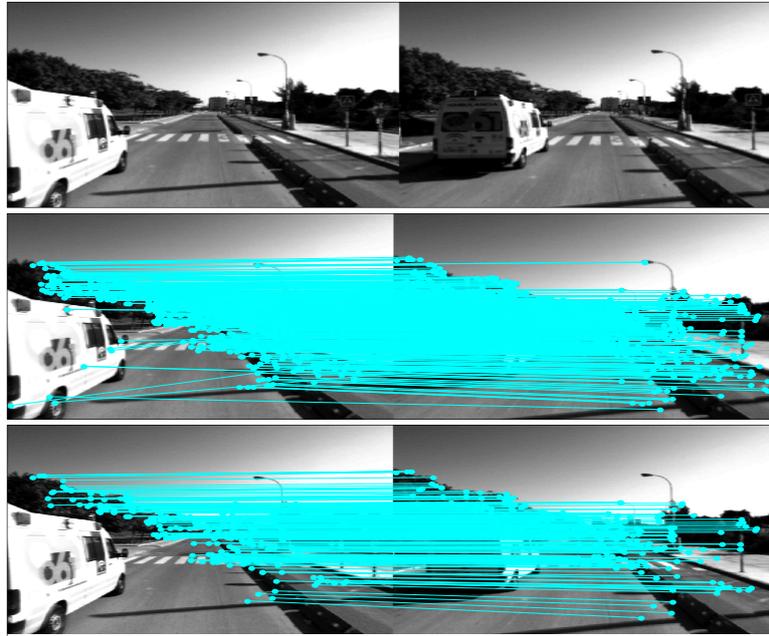


Figure 4 intuitively illustrates a demonstration of feature matching and outlier removal between two adjacent image frames. The upper two subfigures are the original images for analysis showing a van overtaking the host vehicle. The middle ones present the feature matching after SIFT processing, and each cyan line indicates a pair of feature correspondence. It can be noted that the false matching arises and the images thus require outlier removal algorithms in order to retain only the static inliers. The

lower subplots present the refined matching results by applying multilayer RANSAC scheme. As it has shown, all the remainder feature correspondences are recognized as inliers and no feature locates on the moving van any more.

Figure 4. Feature matching and outlier removal between two image frames.



4.3. Motion Estimation

After the refinement of feature matching by implementing the multilayer RANSAC scheme, we also compute the most reasonable estimate of the matrix F . According to the primary principles of epipolar geometry and the properties of fundamental matrix [30], we have the following relationship between the motion of the camera and fundamental matrix:

$$E = K^T F K = (\vec{l} \times) \cdot R \quad (13)$$

where $\vec{l} \times$ is the skew symmetric of camera translation; R is the rotation matrix; K contains the intrinsic calibration parameters and E is the essential matrix. By implementing the singular value decomposition (SVD), the camera rotation R and unit translation vector \vec{l} regarding two adjacent images can be finally retrieved and serve as the vision-based navigation parameters for integration purposes. It is worth noting that since the essential matrix E has only five degrees of freedom, the 3×1 vector \vec{l} is resolved up to scale and is accordingly subject to an ambiguity issue in terms of the translation magnitude. If the absolute magnitude of the translation between two successive images is extracted by other approaches, which we will elaborate in Section 5, the scale ambiguity of \vec{l} can be finally resolved.

5. Determination of Translation Magnitude from GNSS Measurements

In an effort to resolve the scale ambiguity issue of a monocular camera, there are several available approaches, such as using the wheel tick and inertial sensors [6]. Before resorting to the inertial-aided approach, a decorrelation solution between sensor error and inertial mechanization is desired. An

analogous situation applies to using a wheel tick sensor if its drift arises quickly. Conversely, as one of the most widely applied outdoor positioning technologies, GNSS does not yield accumulated ranging or positioning errors, and can provide the capability to resolve the scale ambiguity problem of a monocular camera by leveraging GNSS differential technique. If the acquired GNSS and image data are synchronized, the baseline of position change between two GNSS epochs coincides with the desired camera translation. The GNSS observation equation for pseudorange measurements is given by:

$$P^i = \rho^i + c \cdot t_r - c \cdot T^i - d_I^i + d_T^i + \varepsilon^i \quad i = 1, \dots, k \quad (14)$$

where:

P^i is the pseudorange measurement;

ρ^i is the true geometric range between the designated GNSS satellite and the vehicle;

t_r is the receiver clock error;

T^i is the satellite clock error;

c is the speed of light;

d_I^i and d_T^i are ionospheric and tropospheric corrections, respectively, along the signal propagation path;

ε^i denotes the unmodeled observation errors containing the thermal noise and so forth;

The superscript i indicates the satellite index and k represents the total tracked number of satellites.

Resorting to the concept of differential positioning technology, Equation (14) between two epochs yield differential expression as follows:

$$\delta P^i = \delta \rho^i + c \cdot \delta t_r + \delta \varepsilon^i \quad i = 1, \dots, k \quad (15)$$

where the notation δ depicts the differential operation for each term. Under low dynamic conditions with short time intervals, the differential operation adequately eliminates the ionospheric and tropospheric effects. For the differential term of the satellite clock error, it can be either calculated using the decoded ephemeris or simply ignored as the satellite clock drift can be neglected in a short temporal scale.

According to van Graas [31] and Soloviev [6], the true range differential $\delta \rho^i$ correlates the baseline magnitude of position change and the relationship satisfies:

$$\begin{aligned} \delta \rho^i &= \rho^i(t) - \rho^i(t-1) \\ &= \text{dot}(\vec{R}^i(t) - \vec{R}_r(t), \vec{u}(t)) - \text{dot}(\vec{R}^i(t-1) - \vec{R}_r(t-1), \vec{u}(t-1)) \\ &= \text{dot}(\vec{R}^i(t), \vec{u}(t)) - \text{dot}(\vec{R}^i(t-1), \vec{u}(t-1)) - \text{dot}(\delta \vec{R}_r(t), \vec{u}(t)) - \text{dot}(\vec{R}_r(t-1), \delta \vec{u}(t)) \end{aligned} \quad (16)$$

$i = 1, \dots, k$

where $\text{dot}(\cdot)$ defines the vector dot product; \vec{R}^i and \vec{R}_r denote the position vectors of the satellite and the receiver in the e frame; $\vec{u} = \frac{\vec{R}^i - \vec{R}_r}{\|\vec{R}^i - \vec{R}_r\|}$ is the unit vector along the signal reception path.

According to the commutative and distributive laws of the vector dot product, Equation (16) yields the baseline of position change $\delta \vec{R}_r$, as well as the change of the unit vector $\delta \vec{u}$. With the satellite broadcast ephemeris, the satellite position vector $\vec{R}^i(t-1)$ and $\vec{R}^i(t)$ can be easily estimated. In addition, the user's previous position vector $\vec{R}_r(t-1)$ has been optimally estimated by the filter of the

integrated system, and the present position vector $\vec{R}_r(t)$ in Equation (16) can be initially determined by inertial navigation given a short time interval. The $\vec{R}_r(t)$ at this stage is not for the ultimate position solution at the present time t but for the upcoming δR_r estimation. Note that after the dot product operation, the desired unknown becomes the scalar δR_r instead of $\delta \vec{R}_r$.

Combining the Equations (15) and (16) with rearranged known/unknown quantities, the revised differential equation is expressed as:

$$\delta P'^i = -\cos(\theta_r^i) \cdot \delta R_r + c \cdot \delta t_r + \delta \varepsilon^i \quad i = 1, \dots, k \quad (17)$$

where:

$\delta P'^i$ is the pseudorange differential term compensated by the known quantities of the dot product operation;

θ_r^i represents the included angle between the vectors of $\delta \vec{R}_r$ and \vec{u} .

With two or more available pseudorange observations, the unknown terms of δR_r and δt_r can be derived by implementing the least-square fit. It is worth noting that the LOS measurements play an important role in accurate δR_r estimation as a contaminated measurement does not purely satisfy the above-mentioned satellite/user relationship and will give rise to a biased estimate of δR_r . In an effort to conduct multipath detection and mitigation in challenging GNSS environments, previous researchers have applied several approaches [32,33] by leveraging pseudorange, carrier phase or signal-to-noise ratio (SNR) measurements based on the test statistics. However, difficulties arise to road applications as a result of undesirable detection performance and required hardware upgrades [34]. Another easier option uses a weighting scheme by evaluating the satellite elevation angle and carrier-to-noise ratio (C/No) to restrain the multipath effect. With the synchronized GNSS and camera observations, the camera scale ambiguity can be finally resolved by the estimation of the baseline magnitude of position change δR_r between two GNSS epochs.

6. EKF Modeling and Implementation

To optimally implement sensor integration, several forms of non-linear filtering techniques were previously developed and validated, such as the EKF, unscented Kalman filter (UKF) as well as particle filter (PF). An EKF design linearizes the system and measurement models by considering a first-order Taylor series expansion at the predicted state estimate. This first-order approximation enables the linearized models to implement standard Kalman filter processing. In a typical UKF design, particularly for high non-linear instances, the selected sigma points through the unscented transform estimate the mean and covariance of the state vector. Based on the Bayesian estimation theory, the PF design is supposed to achieve a better solution compared to the EKF and UKF implementations if it is properly established. Nevertheless, the intense computational burden of the UKF and PF processing, one of their prominent shortcomings, brings about severe time-lag estimation and fatally hinders real-time road applications. On the other hand, for low-dynamic applications such as consumer-grade ground vehicle navigation, a time-efficient EKF-based sensor integration scheme requires limited computation and can agree well with the actual situation in terms of the accuracy in statistics [35,36]. The EKF approach is thus adopted as the proposed filter design.

As described previously, Equations (4) to (9) serve as the EKF discretized continuous-time model which is summarized as:

$$\frac{dx(t)}{dt} = Fx(t) + Gw(t), \quad w(t) \sim N(0, Q(t)) \quad (18)$$

In Equation (18), F and G are the linearized dynamic matrix and input coefficient matrix with constant dimensionalities. Additive white Gaussian noise of the system is denoted as $w(t)$ with a power spectral density $Q(t)$ and given by:

$$w(t) = [w_f^T \quad w_\omega^T \quad \mu_f^T \quad \mu_\omega^T] \quad (19)$$

where the four triads denote uncorrelated and correlated noises of the sensor biases. The 15-element system state vector is given by:

$$x(t) = [\delta P_{3 \times 1}^n \quad \delta V_{3 \times 1}^n \quad \delta \Psi_{3 \times 1}^n \quad \delta f_{3 \times 1}^b \quad \delta \omega_{3 \times 1}^b]^T \quad (20)$$

where the triad perturbation terms of position, velocity, orientation and sensor biases are included.

As camera-derived navigation parameters, the rotation matrix $R_c(t)$ and translation unit vector $\vec{l}_c(t)$ require additional transformations to obtain the camera-based position and orientation under the e and n frames, correspondingly. Let $C_b^n(0)$ be the initial IMU orientation matrix and C_c^b be the constant rotation matrix between the c and b frames. Assuming the origins of the c and b frames coincide, the camera-based orientation $\tilde{C}_b^n(t)$ can be expressed as:

$$\tilde{C}_b^n(t) = C_b^n(0) \cdot C_c^b \cdot (V(t))^T \cdot (C_c^b)^T \quad (21)$$

where $V(t)$ is the accumulated 3×3 camera rotation from the first image and satisfies:

$$V(t) = R_c(t) \cdot V(t-1), \quad V(1) = R_c(1) \quad (22)$$

Assuming $\tilde{C}_b^n(t)$ contains only WGN, the orientation error term $\delta \Psi^n$ in Equation (20) correlates $\tilde{C}_b^n(t)$ as well as IMU-based orientation C_b^n with the relationship of:

$$\tilde{C}_b^n(t) = [I_{3 \times 3} + (\delta \Psi^n(t) \times)] \cdot C_b^n(t) + \Gamma_{3 \times 3}(t) \quad (23)$$

where matrix $\Gamma_{3 \times 3}$ consists of the measurement noises. Rearranging Equation (23) yields:

$$\delta \Psi^n(t) \times = (\tilde{C}_b^n(t) - C_b^n(t)) \cdot (C_b^n(t))^{-1} + \Gamma_{3 \times 3}(t) \quad (24)$$

Let $a_{3,2}$, $a_{1,3}$ and $a_{2,1}$ be the corresponding elements from the matrix of right-hand side multiplication and construct:

$$A = [a_{3,2} \quad a_{1,3} \quad a_{2,1}]^T \quad (25)$$

$$\eta = [\Gamma_{3,2} \quad \Gamma_{1,3} \quad \Gamma_{2,1}]^T \quad (26)$$

According to Equations (24) to (26), the linear observation equation with respect to the orientation is finally given by:

$$A(t) = \delta \Psi^n(t) + \eta(t) \quad (27)$$

On the other hand, a camera-derived translation unit vector $\vec{l}_c(t)$ is transformed as:

$$\vec{l}_n(t) = \hat{C}_b^n(t-1) \cdot C_c^b \cdot \vec{l}_c(t) \quad (28)$$

which means a unit vector expressed in the n frame. Note that the orientation matrix \hat{C}_b^n does not refer to the present moment but the instant of previous image update and the caret ^ indicates the EKF-based estimation. By obtaining the translation magnitude δR_r which has been resolved in Section 5 by using differential LOS GNSS pseudoranges over epochs, the camera-based position is expressed as:

$$\tilde{P}^e(t) = \hat{P}^e(t-1) + \delta R_r(t) \cdot l_n(t) \quad (29)$$

In the above equation, $\tilde{P}^e(t)$ depicts the camera-based position at the present moment, and $\hat{P}^e(t-1)$ indicates the previous EKF position update, both expressed in the e frame. By subtracting the IMU-based position $P^e(t)$ from $\tilde{P}^e(t)$, the observation equation with respect to the position is established by:

$$P^e(t) - \tilde{P}^e(t) = \delta P^n(t) + \zeta(t) \quad (30)$$

where $\zeta(t)$ denotes the measurement noise. Combining Equations (27) and (30) yields the EKF measurement model as below:

$$Y(t) = Hx(t) + v(t), \quad v(t) \sim N(0, R_{cov}(t)) \quad (31)$$

In Equation (31), the measurement and noise vectors are respectively given by:

$$Y(t) = [(P^e(t) - \tilde{P}^e(t))^T \quad A(t)^T]^T \quad (32)$$

$$v(t) = [\zeta(t)^T \quad \eta(t)^T]^T \quad (33)$$

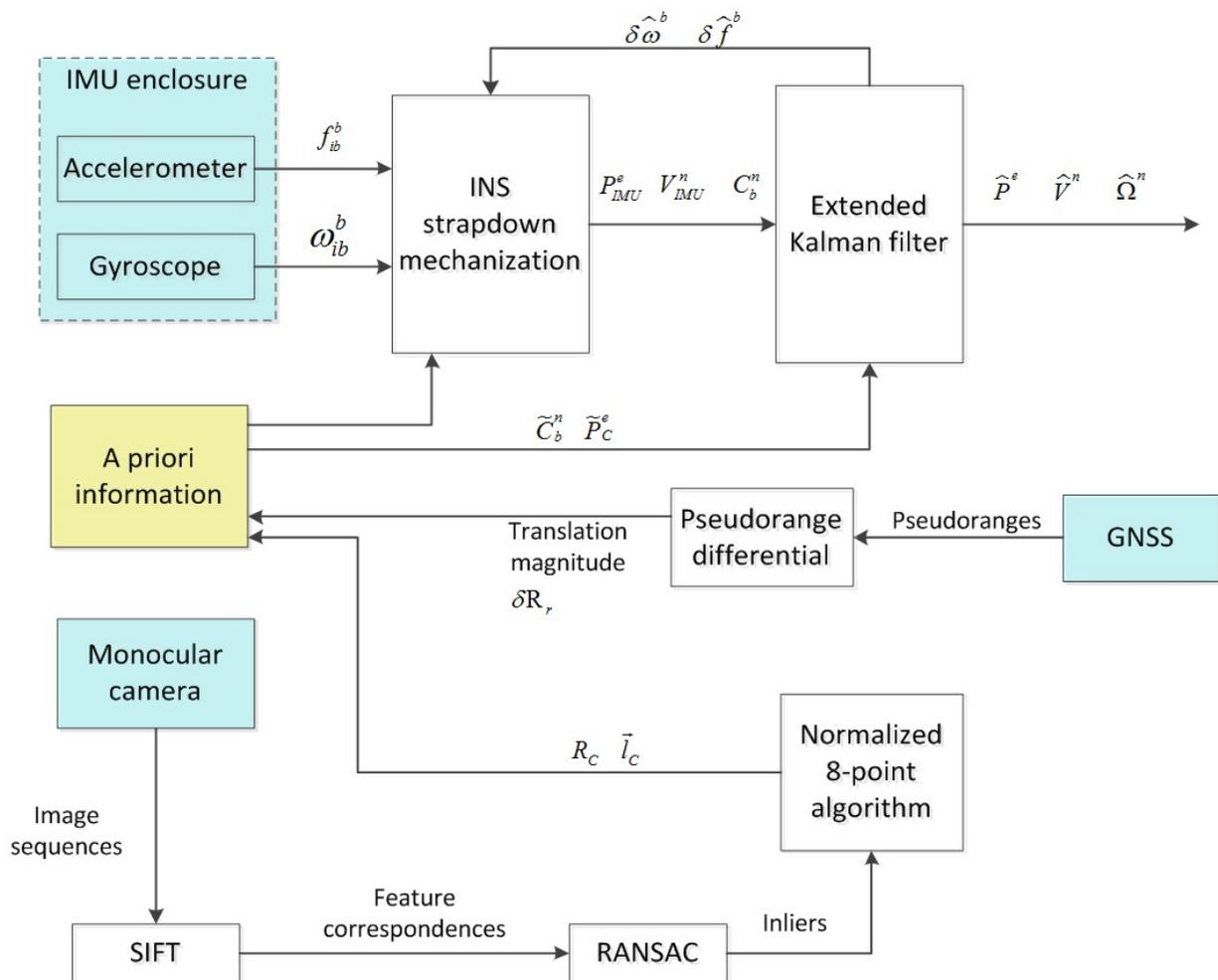
$v(t)$ is assumed as WGN vector with a measurement noise covariance $R_{cov}(t)$. The measurement model matrix H is given by:

$$H = \begin{bmatrix} I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \end{bmatrix} \quad (34)$$

Following the standard EKF procedure, the state vector as well as its covariance matrix can be estimated and applied to ground vehicle navigation. After each update cycle, the state vector is reset to zero and refreshed by the next update operation.

Figure 5 depicts an overall flowchart for integrating the camera, IMU and GNSS sensors through the EKF engine. The sensor units are colored cyan while white blocks indicate the utilized algorithms for intermediate parameter calculation. Between camera/GNSS data update, the vehicle motion is estimated by strapdown IMU mechanization. Once new camera/GNSS measurement data arrive, all derived navigation information is transformed in order to enable the EKF engine for accurate navigation solution and IMU sensor calibration. In such a method, the measurement model is straightforward to construct as relatively small/constant size of measurement elements are included in the EKF. The whole processing requires a relatively low computational load since covariance matrices have low dimensionality. Finally, a low-pass filter before implementing the EKF can be added to smooth the camera-derived rotation and translation parameters.

Figure 5. Flowchart of the proposed EKF-based architecture.



7. Experimental Description and Result Analysis

Our proposed integrated architecture is examined through a live test which was collected by a ground vehicle in an operational traffic environment in Málaga, Spain and is publicly available online [13]. The IMU and monocular forward-looking camera sensors were rigidly fixed on the vehicle rooftop. In order to provide a truth reference solution, three RTK GPS receivers were installed on the vehicle to generate centimeter-level positioning and orientation solutions. In total six drive segments have been accomplished, however the corresponding truth references were vulnerable to limited signal qualities. The *Campus_0L* segment had full range availability of truth reference and contains most common ground vehicle dynamics, such as acceleration, deceleration, stillness, forward driving, cornering, *etc.*, and it is therefore adopted in this paper to test our integrated design. Other sensors utilized during the drive test, such as laser scanners, are disregarded.

Table 1 presents some basic information of the overall drive test. In attempt to keep the recorded data from all sensors available during the whole test, we deliberately remove the beginning of the dataset when the vehicle was keeping static and the sensors were being powered on sequentially. The timespan in Table 1 is, consequently, approximately 20 s shorter than in the actual case.

Table 1. Basic information of the live drive test (Reproduced with permission from [13]).

CAMPUS-0L test details	
Data-collection time (UTC)	2008/11/09, 10:07:57
Distance of travel	1,139 m
Timespan	322 s
IMU make/model	Xsens MTi
IMU output rate	100 Hz
Camera make/model	AVT Marlin F-131C
Image frame rate	7.5 Hz
Image resolution	1,024 × 768
Truth output rate	100 Hz

Table 2 lists the individual IMU sensor specification of Xsens MTi enclosure [37]. Although this product enables 3D axes-based magnetometer to compensate for gyroscope drift, we ignore the magnetic field data and just make use of the 6-DOF raw angular rate and specific force measurements.

Table 2. Gyroscope and Accelerometer specification of Xsens MTi enclosure (Reproduced with permission from [37]).

	Gyroscope	Accelerometer
Drift rate	1 deg/s	0.02 m/s ²
Noise	0.05 deg/s/√Hz	0.002 m/s ² /√Hz
Bandwidth	40 Hz	30 Hz
Misalignment	0.1 deg	0.1 deg
Scale factor	—	0.03%

Table 3 summarizes the major intrinsic calibration parameters from the camera. There were actually two cameras of the same model mounted at both sides in front of the vehicle rooftop. Without loss of generality, only the left camera is employed in this paper and parameters in Table 3 are, accordingly, derived from the left one.

Table 3. Intrinsic camera parameters of the camera in use (Reproduced with permission from [13]).

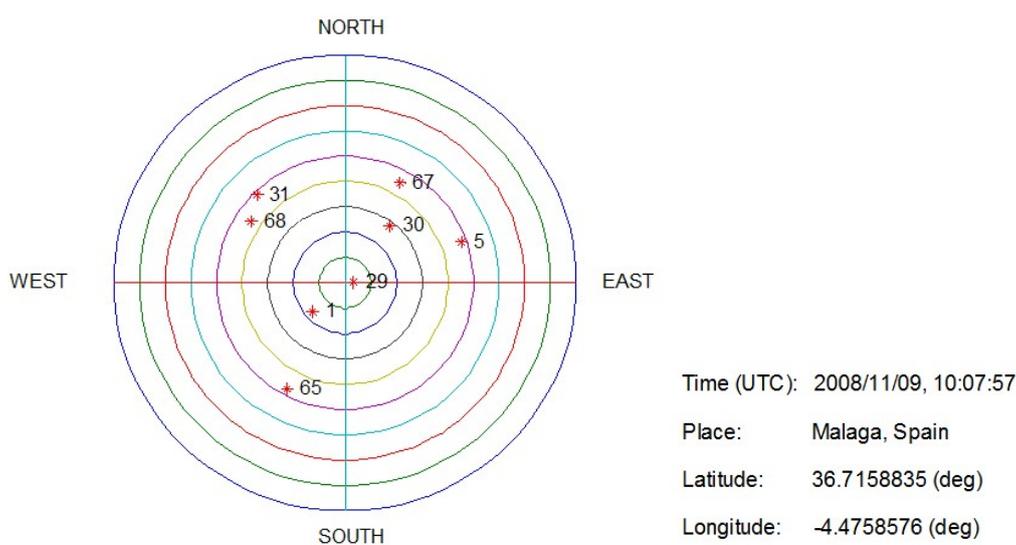
AVT Marlin F-131C camera	
f_x (pixel)	923.5295
f_y (pixel)	922.2418
c_x (pixel)	507.2222
c_y (pixel)	383.5822
k_1	-0.353754
k_2	0.162014
p_1	1.643379×10^{-3}
p_2	3.655471×10^{-4}

The parameters f_x as well as f_y denote the location of the principal point by means of pixel element on the image plane. The focal length is, similarly, defined in terms of pixel element in both x and y axes, and is hence depicted as c_x and c_y . The remainder of the parameters in Table 3 are radial and

tangential distortions, respectively. In the *Campus_0L* segment, the rectified image frames are utilized instead of the originally captured ones to avoid the image projection offset and obtain accurate motion estimation parameters for sensor integration.

Although the b and c frames do not necessarily coincide with each other as in our case, the camera and IMU sensors are usually rigidly fixed to the vehicle and their relative translation and rotation parameters are, therefore, constant. The rotation calibration between the b and c frames can be accomplished according to Lobo [38] while the lever arm vector needs more effort to be accurately estimated. The difficulties of lever arm vector estimation arise due to the sensitivity of the selected target position and the geometric configuration [38]. Normally, in low dynamic situations with moderate maneuvers, particularly for ground vehicle applications, a mild lever arm has little impact little on the navigation solution and it does not necessarily require an accurate estimation [15,39]. Therefore, although both rotation and translation calibration results between the b and c frames are provided in the dataset, we simply disregard the lever arm effect caused by the translation and take only the relative rotation C_c^b into consideration.

Figure 6. Sky plot of available GPS and GLONASS satellites with the elevation mask angle of 40 degree.



As described in Section 5, the scale factor ambiguity between every two successive image frames can be resolved using GNSS differential techniques, capable of obtaining the distance traveled over ground. In an effort to acquire more available GNSS measurements in potentially challenging environments, we take advantage of both GPS and GLONASS constellations for the pseudorange simulation. In typical challenging GNSS environments like urban downtown areas, low-elevation-angle signals are prone to severe multipath degradation or obstruction caused by nearby buildings. We, therefore, filter out those low-elevation-angle satellites from subsequent integration demonstrations. Figure 6 shows the sky plot of GPS/GLONASS satellites when the dataset was initially collected in Málaga with the elevation mask angle of 40 degree. The GPS pseudo random number (PRN) is between 1 and 32, while GLONASS orbital slot is designated within the range 51 to 74. Although the elevation mask angle is significantly higher than in most other applications, both constellations

contributed 8 space vehicles for the integration processing. According to the differential operation described in Equation (15), most of the common error sources can be eliminated from the satellite-to-user geometry.

It is simulated in this paper that the LOS pseudorange measurements contain thermal noise with 0.25 m root-mean-square error (RMSE) during the whole *Campus_OL* segment. Note that the two constellations do not share the same coordinate system as well as time reference. In order to avoid newly introduced unknown quantities, the transformation between their coordinate frames has been implemented by using the approach presented by Cook [40] and, similarly, the system time offset can be extracted from the ephemerides of all modernized GLONASS-M satellites [41].

All the camera, IMU and GNSS measurements are timestamped. Instead of using the original 7.5 Hz frame rate, we deliberately downsample the visual measurements to 2.5 Hz to reduce the computational load and maintain preferable alignment with the 100 Hz IMU measurements for the EKF state update estimation. Since the IMU measurements were not necessarily synchronized with the image frames, an extrapolation process of IMU measurements is expected to establish synchronization consistency. According to You [42], a second-order polynomial extrapolation is utilized to balance the computational load and the accuracy of extrapolation. In order to demonstrate and assess how different sensory-integration schemes perform in challenging environments, we gradually reduce the number of visible GPS/GLONASS satellites with lower elevation angles over time and compare the performance between the tightly coupled GNSS/IMU integration and our proposed camera/IMU/GNSS scheme.

A Google Earth visualization of 1 Hz navigation trajectories of different integration schemes is qualitatively given in Figure 7 with common start/end times. The drive test started at the lower left and ended at the upper middle with the upside indicating the north direction. The black trajectory serves as the truth reference which was obtained by leveraging three RTK GPS receivers. The tightly coupled GNSS/IMU solution is illustrated in red and the vision/IMU/GNSS integration is shown as cyan line. Two purple arrows directly denote the place where the number of available satellites changes. A more quantitative comparison of the horizontal accuracy is shown in Figure 8 to reveal how the different integrated architectures react in terms of reduced observability of GNSS measurements in challenging environments.

Based on Figures 7 and 8, the tight coupling of GNSS and IMU measurements yields an accurate solution when an adequate number of satellites are available (above four). Meanwhile, the gradually accumulated error of proposed camera/IMU/GNSS approach reaches a maximum of 5 m before the number of visible satellites drops to three. The small systematic error mainly stems from erosions of unexpected pixel positioning error, camera calibration error, numerical instability, and remnants of the outliers after RANSAC processing and so forth. After the number of satellites drops to three, the tightly coupled GNSS/IMU integration still enables a continuous navigation solution with insufficient observability, however the estimation tends to sharply drift away from the truth reference over time and its horizontal error exceeds 20 m at the end of the test. Contrarily, compared to previous navigation results, the camera/IMU/GNSS integration yields a slightly better accuracy with only three utilizable satellites. The reasons for this phenomenon are twofold. First, based on Equation (17) which contains two unknown quantities, three observation equations still maintain a redundant constraint provided that the measurements are derived from LOS directions. More importantly, the systematic error of the image processing module gives rise to the reversal of the positioning error during the

vehicle’s U-turn operation. As seen in Figure 7, the cyan line slightly tends to the left side of truth reference prior to the U-turn and therefore the same directionality of the systematic error makes the horizontal accuracy be partially compensated after the U-turn.

Figure 7. Google Earth visualization of the navigation trajectories (black line: truth trajectory, cyan line: camera/IMU/GNSS integration, red line: tightly coupled GNSS/IMU).



Figure 8. Comparison of horizontal error between the tightly coupled GNSS/IMU and camera/IMU/GNSS integrations in terms of reduced observability of GNSS measurements.

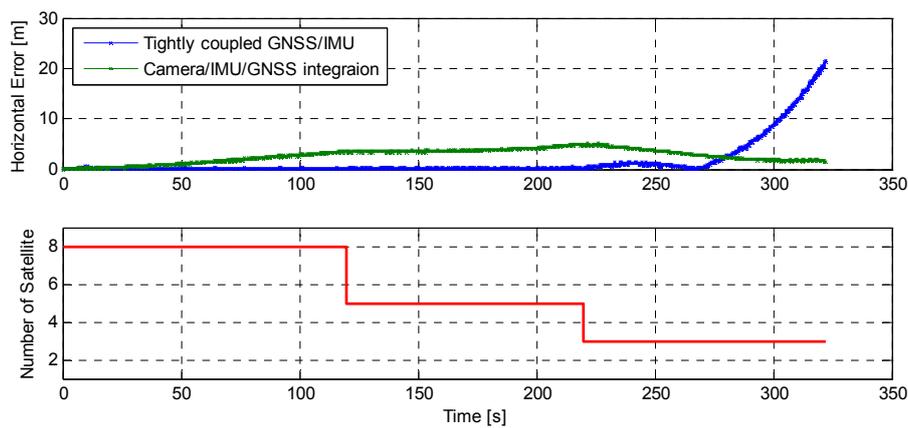
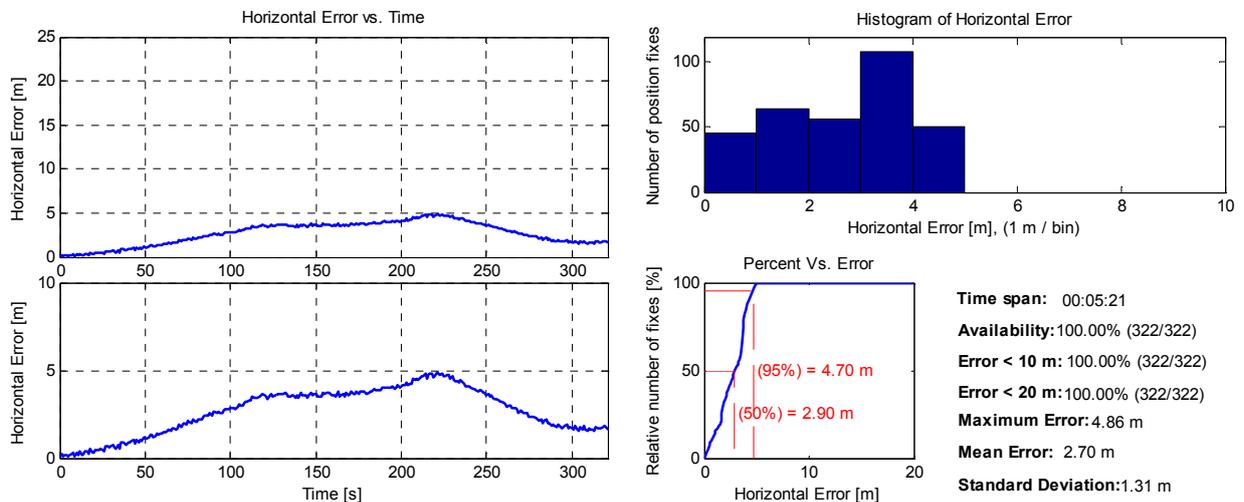


Figure 9. Horizontal positioning error statistics of the camera/IMU/GNSS integration.

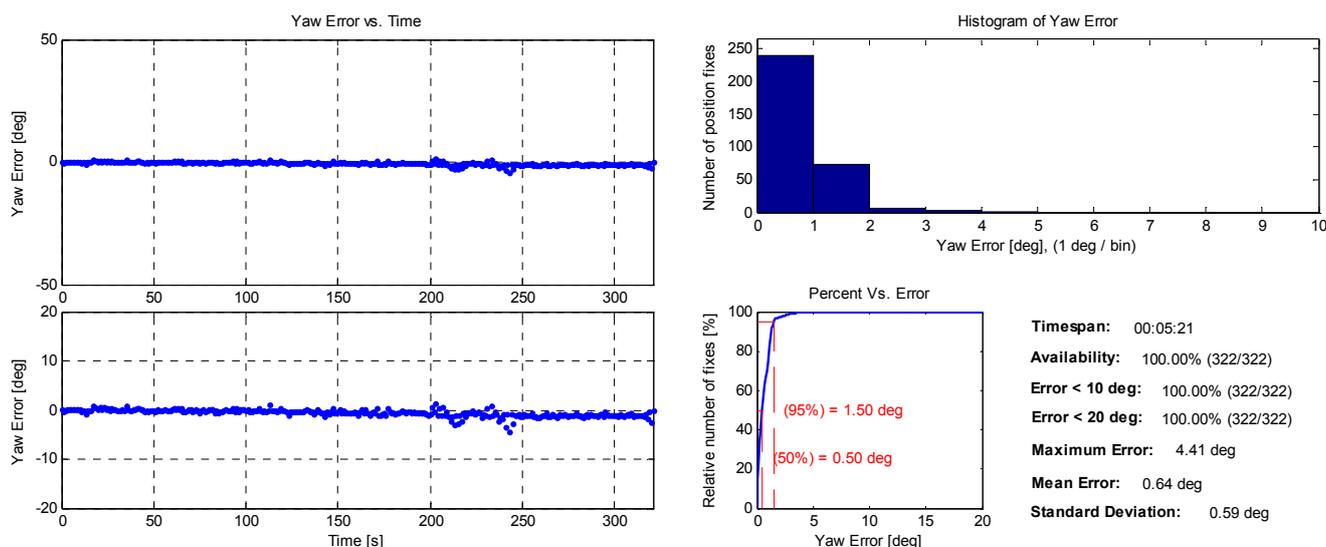


According to the above-mentioned facts, the camera/IMU/GNSS integration brings us a potential advantage compared with the traditional tightly coupled GNSS/IMU technique in challenging environments with sparse GNSS observations. A more detailed horizontal positioning error analysis of the camera/IMU/GNSS integration is presented in Figure 9 with the same hypothesis.

The left side shows the horizontal positioning error in two different zoom levels. The right side shows the histograms, cumulative distribution curves and other related statistics. All the subfigures and corresponding statistics are based on 1 Hz solution. Refer to [1,43] which define these statistics/plots in detail.

Researchers sometimes refer to quaternions, rather than Euler angles, for orientation representation, particularly for high dynamic applications [14,15]. The quaternion expression avoids the gimbal lock phenomenon when the pitch angle approaches ± 90 deg, and operates more efficiently compared with multiplications of direction cosine matrices. Whereas, since a ground vehicle is incapable of operating orthogonal to the ground plane, Euler angles do not suffer from the singularity problem and can provide an intuitive manner for the user to perceive the vehicle direction. We, therefore, choose roll, pitch and yaw angles for representing the vehicle's orientation based on Tait-Bryan convention. Figure 10 shows the yaw angle error statistics of the camera/IMU/GNSS integration. Yaw jitters occur as a result of the abrupt change in the centrifugal force when the vehicle underwent the U-turn operation. Although the angle residual is maintained within 1 deg in a majority of the time during the test segment, the relatively low level of systematic error still can be observed from the zoomed-in subfigure due to the imperfection of the computer vision module. This further proves the reversal of the positioning error during the vehicle's cornering as shown in Figure 8. It can be inferred that the accumulated position and orientation errors will gradually erode the navigation solution if the dataset is long enough without any other sources of corrections.

Figure 10. Yaw angle error statistics of the camera/IMU/GNSS integration.



In addition, the IMU sensor bias estimation is shown in Figure 11. Those large biases from both the accelerometer and gyroscope sensors inevitably cause erroneous IMU mechanization solutions in a matter of seconds without sensor calibration. The non-smoothly varying biases, particularly for the

accelerometer components, gave rise to the unpredictability of the sensor errors. Simply averaging the sensor biases within even a few seconds may yield an inaccurate navigation solution, particularly when the vehicle dynamics significantly change with velocity and cornering stiffness. As seen in Figure 11, during the vehicle's cornering, the accelerometer bias on x-axis dramatically dropped by approximately 200 milli-g. However, the actual sensor biases may be, to a certain extent, different from the estimation shown in Figure 11 in terms of the fidelities of the utilized filtering models [16].

Figure 11. Accelerometer and gyroscope estimation based on the camera/IMU/GNSS integration.

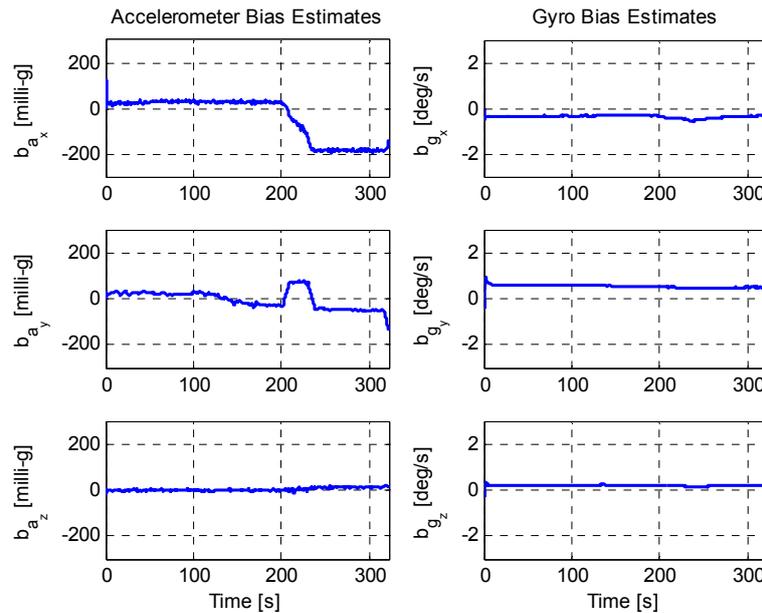
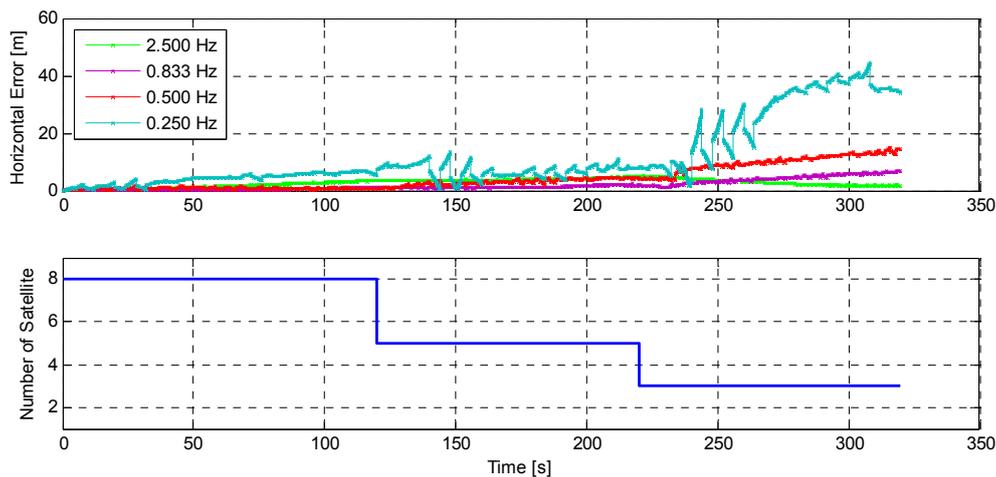


Figure 12. Horizontal positioning error comparison by using different image frame rate.



In an effort to balance the tradeoff between practical performance and computational cost, we deliberately reduce the image frame rate and examine the position solution of our integrated scheme by comparing with the truth reference. Figure 12 demonstrates a comparison of the horizontal positioning errors with different image frame rates based on the same GNSS observability as discussed previously.

As the interval of images is gradually enlarged, EKF filtering leans more on the inertial navigation alone which gives rise to larger location and orientation errors, and moreover, those fewer detected

feature correspondences tend to degrade the reliability of camera motion estimation. The sawtooth-shaped curves therefore arise with worse accuracy. During the U-turn operation when the images partially overlap each other, SIFT/RANSAC algorithms acquire much fewer feature correspondences with lower frame rate, which results in abrupt change in positioning accuracy as seen in Figure 12, particularly for the low frame rates of 0.25 and 0.50 Hz. The worst situation may occur when no overlap can be found in two consecutive images. Consequently, the strategy to determine an image frame rate is primarily dependent on the vehicle dynamics, IMU quality and fidelity of utilized filtering models. In our case, 0.25 Hz frame rate reaches the lower limit as during the U-turn the gaps of the image overlap arise for a frame interval of approximately 4 s. It is worth noting that the vision-based error characteristics are relevant to the processed images and the systematic error therefore varies with the frame rate as seen in Figure 12, particularly for the time period after U-turn. A higher image frame rate tends to yield better integration accuracy at the cost of more excessive computational load, whereas the acceptable processing speed inevitably depends on the available computational capability. In *Campus_0L* segment, we used a 64-bit desktop computer with a 3 GHz processor to process the 2.5 Hz rate images with $1,024 \times 768$ resolution based on a MATLAB platform. Our implementation requires a few seconds of processing for each pair of images and tens of milliseconds for all other filtering routines and, consequently, has yet to achieve real-time functionality. Moore's law, fortunately, has been appropriately indicating the rapid development of integrated circuits during the past decades and also has given us encouraging prediction on the exponentially growing processing capacity of semiconductor materials in the future. In addition, instead of using a CPU, a GPU module also facilitates speeding up the image processing task towards a real-time computation.

8. Conclusions

Aiming at taking advantage of the existing onboard technologies for ground vehicle navigation in challenging environments, this paper has developed an integrated camera/IMU/GNSS system based on the EKF design. The estimated motion parameters from successive image frames were utilized to compensate the drifting IMU sensor biases and maintain the accurate strapdown mechanization between the update of image data. For resolving the scale ambiguity problem of a monocular camera, GNSS differential technique provided an effective manner to extract the baseline magnitude of position change. Based on the EKF design, the proposed integrated system is able to provide 15-state high-bandwidth navigation solutions. Our implementation has been validated through a 5-minute long drive test which provided raw camera and IMU data as well as centimeter-accuracy truth reference. For a reasonable simulation of challenging GNSS environments, we gradually decreased the number of visible satellites with lower elevation angles over time. The experiment has provided results with high accuracies, a majority to the sub-degree level in the yaw angle and meter level in the horizontal plane. Furthermore, it was demonstrated that under the condition of sparse GNSS observations, the camera/IMU/GNSS integration potentially outperformed the tightly-coupled GNSS/IMU scheme which has yielded diverging navigation solutions. The camera/IMU/GNSS integration has also been tested with different image frame rates and the results revealed that the accuracy deteriorates with decreasing frame rates. Future work will further integrate GNSS measurements into the filtering

architecture for better error growth control and automatically enable the loosely/tightly coupled GNSS/IMU integration in case relatively favorable GNSS conditions arise. Future work also includes a target objective of real-time functionality with revised computer vision algorithms on the GPU platform.

Acknowledgments

The invaluable assistance from Sihao Zhao of China Academy of Space Technology and Liang Zhao of Peking University for insightful suggestions is appreciated. The authors are also grateful to MAPIR Lab of University of Málaga for providing a powerful collection of robotic datasets in their MRPT project. The first author's visiting stay to conduct his research at the University of Colorado is financially funded by China Scholarship Council, File No. 2009601147.

References

1. Chu, T.; Akos, D. Assisted GNSS—Performance Results of Multiplexed Measurements, Limited Bandwidth, and a Vectorized Implementation. In *Proceedings of the 2011 International Technical Meeting of the Institute of Navigation*, San Diego, CA, USA, 24–26 January 2011; pp. 1007–1018.
2. Park, M.; Gao, Y. Error and performance analysis of MEMS-based inertial sensors with a low-cost GPS receiver. *Sensors* **2008**, *9*, 2240–2261.
3. Godha, S.; Cannon, M.E. GPS/MEMS INS integrated system for navigation in urban areas. *GPS Solut.* **2007**, *11*, 193–203.
4. Bijker, J.; Steyn, W. Kalman filter configurations for a low-cost loosely integrated inertial navigation system on an airship. *Control Eng. Pract.* **2008**, *16*, 1509–1518.
5. Durrant-Whyte, H.; Bailey, T. Simultaneous localisation and mapping (SLAM): Part I the essential algorithms. *Robot. Autom. Mag.* **2006**, *13*, 99–110.
6. Soloviev, A.; Venable, D. Integration of GPS and Vision Measurements for Navigation in GPS-Challenged Environments. In *Proceedings of IEEE/ION PLANS 2010*, Indian Wells, CA, USA, 4–6 May 2010; pp. 826–833.
7. Corke, P.; Lobo, J.; Dias, J. An introduction to inertial and visual sensing. *Int. J. Robot. Res.* **2007**, *26*, 519–535.
8. Chu, C.; Lie, F.; Lemay, L.; Gebre-Egziabher, D. Performance Comparison of Tight and Loose INS-Camera Integration. In *Proceedings of the 24rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2011)*, Portland, OR, USA, 19–23 September 2011; pp. 3516–3526.
9. Schlaile, C.; Meister, O.; Frietsch, N. Using natural features for vision based navigation of an indoor-VTOL MAV. *Aero. Sci. Technol.* **2009**, *13*, 349–357.
10. Soloviev, A.; Touma, J.; Klausutis, T.; Miller, M.; Rutkowski, A.; Fontaine, K. Integrated Multi-Aperture Sensor and Navigation Fusion. In *Proceedings of the 22nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2009)*, Savannah, GA, USA, 22–25 September 2009; pp. 759–766.
11. Hide, C.; Moore, T.; Botterill, T. Low cost vision aided IMU for pedestrian navigation. *J. Glob. Position. Syst.* **2011**, *10*, 3–10.

12. Huang, B.; Du, S.; Gao, Y. An Integrated MEMS IMU/Camera System for Pedestrian Navigation in GPS-Denied Environment. In *Proceedings of the 24rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2011)*, Portland, OR, USA, 19–23 September 2011; pp. 2373–2380.
13. Blanco, J.L.; Moreno, F.A.; Gonzalez, J. A collection of outdoor robotic datasets with centimeter-accuracy ground truth. *Auton. Robot.* **2009**, *27*, 327–351.
14. Randeniya, D.I.B.; Sarkar, S.; Gunaratne, M. Vision-IMU integration using a slow-frame-rate monocular vision system in an actual roadway setting. *IEEE Trans. Intell. Transp.* **2010**, *11*, 256–266.
15. Hol, J.D. Pose estimation and calibration algorithms for vision and inertial sensors. M.S. Thesis, Linköping University, Linköping, Sweden, May 2008; pp. 46–47.
16. Gleason, S.; Gebre-Egziabher, D. *GNSS Applications and Methods*; Artech House: Norwood, MA, USA, 2009.
17. Kong, X. Inertial navigation system algorithm for low cost IMU. Ph.D. Thesis, The University of Sydney, Sydney, NSW, Australia, 2000; pp. 12–22.
18. Groves, P.D. *Principles of GNSS, Inertial, and Multi-Sensor Integrated Navigation Systems*; Artech House: Norwood, MA, USA, 2008.
19. Shin, E. Estimation techniques for low-cost inertial navigation. Ph.D. Thesis, University of Calgary, Calgary, AB, Canada, May 2005; pp. 46–56.
20. Benson, D.O. A comparison of two approaches to pure-inertial and Doppler-inertial error analysis. *IEEE Trans. Aero. Electron. Syst.* **1975**, *11*, 447–455.
21. Zhang, Z. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, Greece, 20–27 September 1999; pp. 666–673.
22. Clarke, T.A.; Fryer, J.G. The development of camera calibration methods and models. *Photogramm. Rec.* **1998**, *16*, 51–66.
23. Heikkilä, J.; Silvén, O. A Four-Step Camera Calibration Procedure with Implicit Image Correction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, San Juan, Puerto Rico, 17–19 June 1997; pp. 1106–1112.
24. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
25. Ke, Y.; Sukthankar, R. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2004; Volume I, pp. 511–517.
26. Bay, H.; Ess, A.; Tuytelaars, T.; van Gool, L. Speeded-up robust features (SURF). *J. Comput. Vis. Image Underst.* **2004**, *110*, 346–359.
27. Lingua, A.; Marenchino, D.; Nex, F. Performance analysis of the SIFT operator for automatic feature extraction and matching in photogrammetric applications. *Sensors* **2009**, *9*, 3745–3766.
28. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395.

29. Zhao, L.; Huang, S.; Yan, L.; Wang, J.; Hu, G.; Dissanayake, G. Large-Scale Monocular SLAM by Local Bundle Adjustment and Map Joining. In *Proceedings of the 11th International Conference on Control, Automation, Robotics & Vision (ICARCV)*, Singapore, 5–8 December 2010; pp. 431–436.
30. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2003.
31. Van Graas, F.; Soloviev, A. Precise velocity estimation using a stand-alone GPS receiver. *J. Inst. Navigat.* **2004**, *51*, 283–292.
32. Jülg, T. Evaluation of Multipath Error and Signal Propagation in a Complex Scenario for GPS Multipath Identification. In *Proceedings of the 4th IEEE International Symposium on Spread Spectrum Techniques and Applications*, Mainz, Germany, 22–25 September 1996; pp. 872–876.
33. Comp, C.J.; Axelrad, P. Adaptive SNR-based carrier phase multipath mitigation technique. *IEEE Trans. Aero. Electron. Syst.* **1998**, *34*, 264–276.
34. Lee, H.K.; Lee, J.G.; Jee, G.I. GPS multipath detection based on sequence of successive-time double-differences. *IEEE Signal Proc. Let.* **2004**, *11*, 316–319.
35. Armesto, L.; Tornero, J.; Vincze, M. Fast ego-motion estimation with multi-rate fusion of inertial and vision. *Int. J. Robot. Res.* **2007**, *26*, 577–589.
36. Duong, T.; Chiang, K. Non-Linear, Non-Gaussian Estimation for INS/GPS Integration. In *Proceedings of the 24rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2011)*, Portland, GA, USA, 19–23 September 2011; pp. 946–953.
37. Sensor Performance for Xsens MTi. Available online: <http://www.xsens.com/en/general/mti> (accessed on 4 January 2012).
38. Lobo, J.; Dias, J. Relative pose calibration between visual and inertial sensors. *Int. J. Robot. Res.* **2007**, *26*, 561–575.
39. Hol, J.D.; Schön, T.B.; Luinge, H.; Slycke, P.J.; Gustafsson, F. Robust real-time tracking by fusing measurements from inertial and vision sensors. *J. RT. Image Process.* **2007**, *2*, 149–160.
40. Cook, G.L. Critical GPS-GLONASS Interoperability Issues. In *Proceedings of the 1997 National Technical Meeting of The Institute of Navigation*, Santa Monica, CA, USA, 14–16 January 1997; pp. 183–193.
41. *Global Navigation Satellite System-GLONASS*; Interface Control Document (Edition 5.1); Russian Institute of Space Device Engineering: Moscow, Russia, 2008.
42. You, W.; Jiang, F. Data synchronization technology of INS/GPS integrated navigation system. *J. Chin. Inert. Technol.* **2003**, *11*, 20–22.
43. Chu, T.; Akos, D. Assisted GNSS—Traditional and Vectorized: Implementation and Performance Results. In *Proceedings of the 23rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2010)*, Portland, OR, USA, 21–24 September 2010; pp. 1605–1614.