16th International Conference on Information Fusion
Istanbul, Turkey, July 9-12, 2013

Loosely Coupled Kalman Filtering for Fusion of Visual Odometry and Inertial Navigation

Salim Sırtkaya and Burak Seymen
ASELSAN Inc.
Microelectronics Guidance and Electro-optics Division
Ankara, Turkey
Email: sirtkaya@aselsan.com.tr
bseymen@aselsan.com.tr

A. Aydın Alatan
Middle East Technical University
Electrical and Electronics Engineering Department
Ankara, Turkey
Email: alatan@eee.metu.edu.tr

***Abstract*—Visual Odometry (VO) is the process of estimating the motion of a system using single or stereo cameras. Performance of VO is comparable to that of wheel odometers and GPS under certain conditions; therefore it is an accepted choice for integration with inertial navigation systems especially in GPS denied environments. In general, VO is integrated with the inertial sensors in a state estimation framework. Despite the various instances of estimation filters, the underlying concepts remain the same, an assumed kinematic model of the system is combined with measurements of the states of that system. The drawback of using kinematic models for state transition is that the state estimate will only be as good as the precision of the model used in the filter. A common approach in navigation community is to use an error propagation model of the navigation solution using inertial sensor instead of an assumed dynamical model. High rate IMU will trace the dynamic better than an assumed model. In this paper, we propose a loosely coupled indirect feedback Kalman filter integration for visual odometry and inertial navigation system that is based on error propagation model and takes into account different characteristics of individual sensors for optimum performance, reliability and robustness. Two measurement models are derived for the accumulated and incremental visual odometry measurements. A practical measurement model approach is proposed for the delta position and attitude change measurements that inherently includes delayed-state. The non-Gaussian, non-stationary and correlated error characteristics of VO, that is not suitable to model in a standard Kalman filter, is tackled with averaging the measurements over a Kalman period and utilizing a sigma-test within the filter.**

I. Introduction

Being the primary source of navigation for most of the living creatures, vision has been adopted as a navigation modality for manmade machines for the last three decades. Utilizing vision as a navigation modality has attracted robotics and computer vision societies before navigation society. Inertial sensors and GPS lie in the heart of navigation solution for most of the aerospace and military systems; however these sensors are not feasible in most of the missions for the robotics and computer vision societies. Inertial sensors were considered as expensive and not easy to reach sensors (up to advances in MEMS sensors) and GPS sensors are considered not feasible for indoor and urban environments. These challenges forced the robotics and computer vision societies to use camera as the main navigation sensor. The advances in vision based navigation have shown great potential for the navigation society as well, especially as an alternative to GPS in integrated inertial navigation. The research community has exploited the benefits of integrating visual and inertial sensors in GPS denied environments, and proposed solutions for several missions such as navigation, pin-point landing and flight control.

There are several ways to utilize visual measurements for motion estimation. Structure from Motion (SfM), Simultaneous Localization and Mapping (SLAM) and Visual Odometry (VO) are among the popular methodologies. In SfM and SLAM, motion estimation is formulated together with 3D structure estimation of the captured environment. VO, in contrast, focuses solely on motion estimation.

For visual-inertial integration, visual data is generally incorporated as 2D feature (identifier picture element) matches. Various types of estimation filters are designed to fuse visual and inertial data, Kalman Filter being the most preferred. However, the error characteristic of visual data is not well established in most of these approaches. The error is generally modeled as a white Gaussian noise at different stages of visual algorithms: feature detection, feature matching, feature tracking and visual motion estimation. However, VO algorithms induce non-Gaussian, non-stationary and correlated error on pose estimates due to the nature of the image processing algorithms, effects of optimization steps (such as outlier rejection and bundle adjustment) and error characteristics dependency on scene conditions.

It is required to establish a practical way to overcome the difficulty of modeling the visual motion estimation error in the integration scheme and provide reliable integrated motion estimate under different error characteristics. In the proposed method, a stereo-camera is explored as a sensor modality for inertial navigation aiding in the context of a loosely coupled indirect feedback Kalman Filter. Visual Odometry is utilized for this purpose, and the inertial navigation is supported by camera rotation and velocity estimations. The non-Gaussian behavior of VO is tackled with averaging the measurements over a Kalman period. Robustness to spurious VO measurements is achieved by utilizing a sigma-test within the filter. The main contribution of this paper is the formulation of a loosely coupled indirect feedback Kalman filter with two visual odometry measurement models, in which the filter inherently solves the delayed state measurement problem which is common in delta position sensor integration to inertial navigation.

978-605-86311-1-3

II. Related Work

Although the term Visual Odometry was coined by Nister in his seminal paper [1] at 2004, the history of vision based motion estimation dates back to 80s. A detailed survey is given as a tutorial in [2] which describes the progress within last 30 years. For the integration of VO measurement and inertial navigation system, an accurate representation of the VO output's uncertainty is as critical as an accurate measurement. Hence the VO literature will be reviewed from a visual-inertial integration perspective. The foremost problem of VO in terms of integration to inertial system is error modeling. Either tightly coupled or loosely coupled integrations require a neatly described error model since the outputs of VO algorithms do not conform to the traditional model of the true measurement plus a Gaussian white noise process [3]. In addition, unavoidable non-linear characteristics of visual measurements need to be treated within the filter frameworks. Regardless of the selected algorithm, VO depends on feature matches between the sequential images. Spurious matches (outliers) are the main cause of errors for most of the time. Removing outliers in early steps of VO leads a plausible error behavior at the output. In general, VO algorithms utilize RANSAC [4] as the outlier rejection scheme [1] [5] [6] [7]. Alternative schemes such as mean-shift based rejection [8], parallax-rigidity based rejection [9] try to improve results over RANSAC. In [3], the error caused by outliers is modeled, instead of rejection, to associate an accurate uncertainty at the output. In their approach, a convex combination model is proposed, in which the inlier error is modeled as a Gaussian distribution and outlier is modeled as a uniform distribution. Jiang et al. [10] modeled the visual odometry error as a composition of an unbounded deterministic part with unknown constant parameters, and a first-order Gauss-Markov process. Being mathematically correct, these attempts lack complete representation of the underlying dynamics that cause the error at the output of VO algorithms. For instance, error characteristic changes caused by scene conditions that effect feature inlier quality (in [1], a comparative error analysis for different scene conditions - such as wood, meadow- is given) are not considered. This spatial correlation creates time correlation in dynamic scenarios.

Integration of inertial sensors with vision sensors for navigation is an active research problem. Various derivatives of Kalman Filter are utilized for this purpose. The integration schemes can be divided as tightly coupled and loosely coupled. In tightly coupled integration, the early steps of visual algorithms are embedded to the estimation process. A common choice is to support feature matching or tracking steps with inertial navigation system [11] [12]. In general, feature locations are utilized as visual measurements in these approaches. The features need not to be augmented to the state vector as opposed to SLAM based estimation for practical purposes. Epipolar constraints are also enforced in some applications for real-time implementation considerations [13] [14].

On the other hand, in the loosely coupled integration the output of visual motion estimation stage is used as the measurement. Rotational and translational components of visual motion estimates are used separately or together depending on construction of the integration filter. The filter requires only the uncertainty of the visual motion estimate in this case. In [15] and [16], visual odometry pose estimates are integrated with inertial navigation system in a delayed state Kalman Filter framework.

III. Visual Odometry

Visual Odometry is described as calculation of position and orientation of a camera throughout a sequence. The visual odometry term was chosen for its similarity to wheel odometry, which incrementally estimates the motion of a vehicle by integrating the number of turns of its wheels over time. The VO process is memoryless, meaning that the estimation is conducted on sequential image pairs. VO is denoted as a particular case of Structure from Motion, in which the focus is on estimating the 3-D motion of the camera sequentially as a new frame arrives in real time. VO can be conducted on stereo image sequences as well as monocular image sequences. In monocular case, only bearing information is available. The disadvantage is that motion can only be recovered up to a scale factor. The absolute scale can then be determined from direct measurements (e.g., measuring the size of an element in the scene, the height of the camera from ground level) or from the integration with other sensors, such as IMU, barometer, and range sensors. Over the years, monocular and stereo VOs have almost progressed as two independent lines of research, however most of the research done in VO has been produced using stereo cameras [2]. The interest in monocular methods is due to the observation that stereo VO can degenerate to the monocular case when the distance to the scene is much larger than the stereo baseline. It has been demonstrated that VO provides accurate trajectory estimates with relative position error ranging from 0.1 to 2 percent of distance travelled [2] under certain conditions.

This capability makes VO an interesting supplement to navigation algorithms. The motion estimation approach depends on whether the feature correspondences are specified in two or three dimensions. For a 2D-to-2D motion estimation, the Essential matrix [17] is estimated first from the epipolar constraint and the rotation and translation of the camera is extracted from this matrix. However, the scale ambiguity is a major problem in this type of estimation. Only relative scale between successive frames can be estimated. The 3D-to-3D motion estimation operates on stereo vision. The solution can be formulated as the transformation between successive camera coordinates that minimizes the L2 distance between the two 3D feature set. To compute the transformation, it is also possible to avoid the triangulation of the 3D points in the stereo camera and use trifocal or quadrifocal constraints instead. As pointed out by Nister et al. [1], motion estimation from 3-D-to-2-D correspondences is more accurate than from 3D-to-3D correspondences because it minimizes the image reprojection error.

The work of Geiger et al [6] [18] is followed to get reliable odometry results from a stereo sequence. Their stereo based VO scheme proved to be useful in urban environments. The visual motion estimation framework is based on the trifocal tensors. Trifocal tensors describe the relationship between three images of the same static scene. It encapsulates the projective geometry between the different viewpoints and is independent from the structure of the scene. This approach can be classified as a 3D-to-2D motion estimation algorithm.

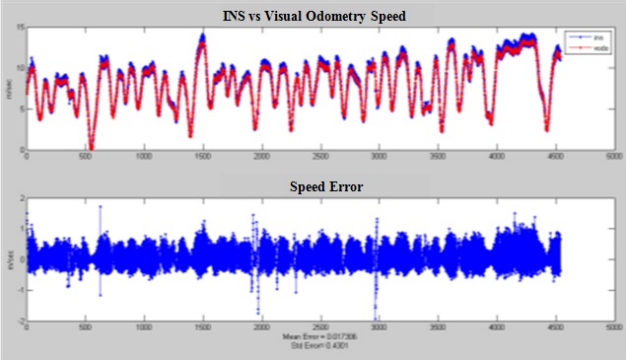

Fig. 1. Ground Truth versus Visual Odometry Speed

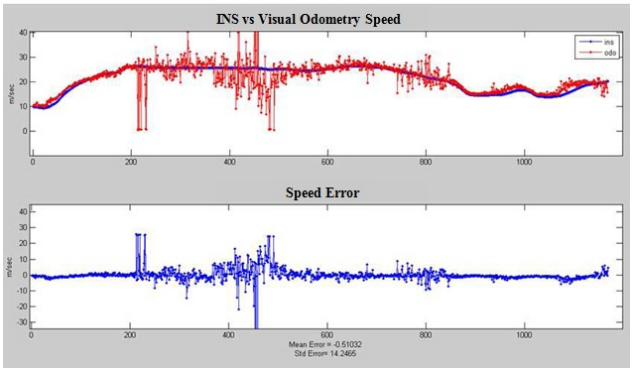

Fig. 2. Ground Truth versus Visual Odometry Speed from a different run. Note that the VO result is erroneous at certain parts of the route

A. Visual Odometry Error Characteristics

The error characteristic of visual odometry should be established well in order to construct a reliable visual odometry based navigation scheme. In general, visual odometry error is denoted as offset ratio in the literature [1] [2] [6], i.e. the ratio of the final drift value to the traveled distance. The offset ratio is rough and does not represent the heterogeneous character of the error. The error characteristic is correlated with the utilized algorithms for individual steps of visual odometry process. As a new and promising sensor, visual odometry needs a methodology for systematic and comparative analysis of its drift, in order to quantify the performance of various algorithms. Drift does not increase linearly by the distance traveled. Moreover, running the same algorithms on the same dataset repeatedly produces quite different offset ratios. The reason is that visual odometry position error is a random process, and it may not always increase, as errors in different motion vectors will compensate each other to some extent. Thus, using end-point values (the final drift values, and the final traveled distances) is incapable to model the whole random process. In addition, there may be occasions where the visual odometry process totally fails. These erroneous conditions occur due to unreliable feature matches especially when the camera is directed to the sky or there are independently moving objects in the scene. In Figures 1 and 2, the visual odometry result is depicted for two different runs of the same sensor setup and algorithm [6]. In the lower figure, the scene was cluttered by independently moving objects. It is evident from these figures that the visual odometry error is non-stationary and requires neat handling when integrated to other sensors.

IV. Loosely Coupled Visual Odometry - Inertial Navigation Integration

Inertial navigation solution, which is obtained by the integration of the acceleration in a reference frame (navigation frame) maintained by the high rate gyroscopes and accelerometers, supplies high dynamic information of the platform. More explicitly, the inertial navigation output is obtained by solution of the following differential equations,

$$\begin{aligned} \dot{C}_B^N &= C_B^N \ \Omega_{NB}^B \\ \dot{\bar{v}}^N &= C_B^N \bar{f}^B + \bar{g}^N - (2\bar{w}_{IE}^N + \bar{w}_{EN}^N) \times \bar{v}^N \\ \dot{\bar{r}}^N &= \bar{v}^N \end{aligned} \tag{1}$$

where C_B^N is the coordinate transformation matrix from body frame to navigation frame, which keeps the attitude information of the body. Ω_{NB}^B is the skew symmetric matrix form the body angular rate vector with respect to navigation frame expressed in body frame coordinates. $\bar{v}^N$ is the velocity of the body with respect to earth frame expressed in navigation frame coordinates. $\bar{f}^B$ is the specific force measured by accelerometer triad, $\bar{g}^N$ is the local gravity vector expressed in navigation frame coordinates. $(2\bar{w}_{IE}^N + \bar{w}_{EN}^N) \times \bar{v}^N$ term represents the apparent acceleration (coriolis) that arises from earth rate and transport rate due to motion in a rotating frame. However, due to the inevitable inertial measurement and initialization errors, the solution drifts in time.

On the other hand, the visual odometry can generate relatively bounded but noisy delta-position and attitude change information for short time periods. Instead of the direct fusion of the high rate inertial sensors and relatively low rate VO output which will cause a loss in high dynamic information obtained by inertial sensors (due to low-pass characteristics of Kalman filter), a traditional indirect (error-state) Kalman Filtering method [19] is proposed to utilize the complementary characteristics of inertial sensors and visual odometry. In this method, the errors of the navigation states are estimated by using the error state measurements constructed by the difference of the VO solution and navigation solution rather than directly estimating the navigation states (attitude, velocity, position). In addition to the navigation and inertial sensor errors (bias, scale factor), systematic visual odometry errors such as scale factor and bias are also augmented to the state vector of the indirect Kalman Filter. The estimated errors are then feedback to the inertial navigation integration (instead of feed-forward) in order not to violate the linearity assumption of error dynamics by eliminating the error growths in navigation variables. The structure of the proposed indirect feedback Kalman Filter is depicted in Figure 3.

A. Total System Error Model

In this section, a generic truth error model is constructed. The total error sources that may affect the system performance are modeled by a stochastic linear dynamics. The corresponding error state vector is given as,

$$\delta\bar{x}^T = [\delta\bar{\phi}^N \ \delta\bar{v}^N \ \delta\bar{r}^N \ \delta\bar{f}^B \ \delta\bar{\omega}^B \ \delta\bar{m}_{sf} \ \delta\bar{m}_b \ \delta\bar{m}_{bs} \ \delta\bar{m}_{la}] \tag{2}$$

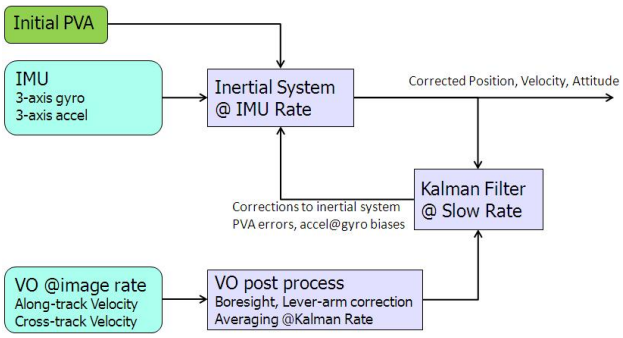

Fig. 3. Loosely Coupled Indirect Feedback Kalman Filter structure

Here $\delta\bar{\phi}^N$ is the perturbed attitude tilt error, $\delta\bar{v}^N$ is the perturbed velocity error and $\delta\bar{r}^N$ is the perturbed position error. $\delta\bar{f}^B$, $\delta\bar{w}^B$ are the accelerometer and gyro error vectors respectively. Each of these vectors includes the systematical errors of the inertial sensor such as bias, scale factor, misalignment errors. $\delta\bar{m}_{sf}$ is the visual odometry scale factor error and $\delta\bar{m}_b$ is the visual odometry correlated bias error that is induced by the visual odometry algorithms under different scene conditions. $\delta\bar{m}_{bs}$, $\delta\bar{m}_{la}$ are the angular (boresight) and translational (lever-arm) displacement between the camera and the IMU, respectively. The INS error state dynamics is well studied subject in inertial navigation community. The perturbation form of the error state dynamics can be expressed in a compact form as [20],

$$\begin{aligned}
\delta\dot{\bar{\phi}}^N &= -(\times\bar{\omega}^N_{IN})\delta\bar{\phi}^N + \delta\bar{\omega}^N_{IN} - C^N_B\delta\bar{\omega}^B \\
\delta\dot{\bar{v}}^N &= -(\times C^N_B\bar{f}^B)\delta\bar{\phi}^N - (2\bar{\omega}^N_{IE} + \bar{\omega}^N_{EN}) \times \delta\bar{v}^N - \\
&\quad (2\delta\bar{\omega}^N_{IE} + \bar{\omega}^N_{EN}) \times \bar{v}^N + C^N_B\delta\bar{f}^B + \delta\bar{g}^N \\
\delta\dot{\bar{r}}^N &= \delta\bar{v}^N
\end{aligned} \tag{3}$$

Here $\bar{\omega}^N_{IN}$ is the angular velocity of the navigation frame with respect to inertial frame and $\bar{\omega}^N_{EN}$ is the angular velocity of the navigation frame with respect to the Earth frame and $\delta\bar{\omega}^N_{IN}, \delta\bar{\omega}^N_{EN}$ are the perturbation errors of those quantities. The IMU errors are generally modeled by first-order Gauss-Markov process and/or random walk process. The visual odometry errors scale factor, bias and lever-arm errors can be modeled by random walk processes with a suitable noise density.

B. Visual Odometry Measurement Model

Visual odometry calculates the delta position vector and attitude change between two consecutive camera frames at image rate. Raw visual odometry measurement cannot guarantee that the measurement noise is normally distributed given the complexity of the algorithm and dynamic scene conditions. Without these guarantees, the Kalman filter is suboptimal at best. Tardif et al. [15] proposed a method that utilizes central limit theorem to overcome this problem. The visual odometry measurements are accumulated over the Kalman Filter interval before incorporated to the filter. By this type of pre-filtering, the effect of relatively high frequency correlated noises (such as oscillatory disturbances due to vehicle engine) is also eliminated. In this manner, the total error will tend to satisfy the Kalman Filter assumptions for a sufficiently large accumulation time. This results in a slower rate than the image rate for the Kalman measurement update which is computationally efficient especially for large state filters. In addition, the indirect feedback form of the Kalman Filter encourages slow rate measurement updates due to relatively low bandwidth of the error dynamics.

1) Accumulated Delta Position Measurement Models: There can be two different types of accumulation for the delta position vectors. One possible way is the accumulation of the delta position vectors expressed in the instant camera coordinate system or the accumulation of the vectors expressed in the initial or the final time of the accumulation interval. In [15] the accumulation of the vectors in the initial time camera coordinate system is proposed.

Model I

The accumulated delta position vector of visual odometry, $\Delta\bar{r}^{\emptyset_k}_{VO,k}$, at Kalman cycle (k) defined in the final time camera optical coordinate axis, $\emptyset_k$, for the accumulation interval $(k-1)$ can be expressed as,

$$\begin{aligned}
\Delta\bar{r}^{\emptyset_k}_{VO,k} &= \bar{r}^{\emptyset_k}_{VO,k} - \bar{r}^{\emptyset_k}_{VO,k-1} \\
&= [C^{\emptyset_k}_{\emptyset_{k-1}}]_{VO} \sum_{m=1}^{N} ([C^{\emptyset_{k-1}}_{\emptyset_{k-1[m-1]}}]_{VO}(M_{SF}\Delta\bar{r}^{\emptyset_{k-1[m-1]}}_{VO,k-1[m]} + \\
&\bar{m}_{b,k-1[m]} + \bar{\nu}_{k-1[m]}))
\end{aligned} \tag{4}$$

where $\Delta\bar{r}^{\emptyset_{k-1[m-1]}}_{VO,k-1[m]}$ is the delta position vector of visual odometry at the minor interval (m) of the $(k-1)^{th}$ major interval expressed in the previous minor interval $(m-1)$ camera coordinate system and $\bar{\nu}_{k-1[m]}$ is the measurement noise vector on the minor interval's delta position vector. M_{SF} is the visual odometry scaling diagonal matrix and $\bar{m}_{b,k-1[m]}$ is the visual odometry correlated bias term. Note that, N corresponds to the number of images in the Kalman interval. $[C^{\emptyset_{k-1}}_{\emptyset_{k-1[m-1]}}]_{VO}$ is calculated by the incremental attitude change measurements recursively as,

$$C^{\emptyset_{k-1}}_{\emptyset_{k-1[m-1]}} = C^{\emptyset_{k-1}}_{\emptyset_{k-1[1]}} \cdot C^{\emptyset_{k-1[1]}}_{\emptyset_{k-1[2]}} \cdots C^{\emptyset_{k-1[m-2]}}_{\emptyset_{k-1[m-1]}} \tag{5}$$

On the other hand, the delta position measurement at the camera optical center can be calculated in terms of the INS computed navigation quantities as,

$$\begin{aligned}
&\Delta\bar{r}^{\emptyset_k}_{INS,k} = \\
&\tilde{C}^{\emptyset}_B(\tilde{C}^N_{B_k})^T(\tilde{\bar{r}}^N_{INS,k} - \tilde{\bar{r}}^N_{INS,k-1}) + \tilde{C}^{\emptyset}_B(I - \tilde{C}^{B_k}_{B_{k-1}})\tilde{\bar{m}}^B_{la}
\end{aligned} \tag{6}$$

Where $\tilde{C}^{\emptyset}_B$ is the estimated boresight matrix between the camera axis and IMU axis where the true value of it is denoted as $C^{\emptyset}_B$. $\tilde{\bar{m}}^B_{la}$ is the estimated lever-arm vector emanating from the INS center of navigation to the camera center expressed in body frame. It should be noted that the change of the navigation frame during the Kalman filter cycle is assumed to be negligible for the error modeling. The delta-position error measurement is constructed by the difference of the visual odometry measurement and INS computed delta position

quantity as,

$$\delta\bar{y}_k = \Delta\bar{r}^{\emptyset_k}_{VO,k} - \Delta\bar{r}^{\emptyset_k}_{INS,k} = [-\tilde{C}^{\emptyset}_B(\tilde{C}^N_{B_k})^T(\times\Delta\bar{r}^N_{INS,k})]\delta\bar{\phi}^N_k + [C^{\emptyset}_B(\tilde{C}^N_{B_k})^T](\delta\bar{r}^N_k - \delta\bar{r}^N_{k-1}) + [C^{\emptyset}_B(\tilde{C}^N_{B_k})^T(\times\Delta\bar{r}^N_{INS,k}) + (\times\tilde{C}^{\emptyset}_B(I - \tilde{C}^{B_k}_{B_{k-1}})\bar{m}^B_{la})]\delta\bar{m}_{bs,k} + \tilde{C}^{\emptyset}_B(I - \tilde{C}^{B_k}_{B_{k-1}})\delta\bar{m}^B_{la} - diag(\Delta\bar{r}^{\emptyset_k}_{VO,k})\delta\bar{m}_{sf,k} - \delta\bar{m}_{b,k} - \bar{\nu}_{ACC,k} \quad (7)$$

where $\bar{\nu}_{ACC,k}$ is the measurement noise after the accumulation process.

In this model, the delta position error is made correlated with the attitude change measurement error due to coordinate transformation process. In the case of using both the delta distance and attitude measurements, the Kalman filter shall be modified for the correlated measurement noise.

Model II

In this case, the accumulated value of the incremental delta-position vector is obtained as,

$$\Delta\bar{r}_{VO,k} = \sum_{m=1}^{N}(\Delta\bar{r}^{\emptyset_{k-1[m-1]}}_{VO,k-1[m]} + \bar{\nu}_{k-1[m]}) \quad (8)$$

The corresponding delta position measurement at the camera center can be calculated in terms of the INS computed navigation quantities by assuming sufficiently small minor interval as,

$$\Delta\bar{r}_{INS,k} = \tilde{C}^{\emptyset}_B\int_{t_{k-1}}^{t_k}\left((C^N_{B(t)})^T\tilde{\tilde{v}}^N_{INS} + \bar{w}^B_{EB}\times\tilde{\tilde{m}}^B_{la}\right)dt \quad (9)$$

The measurement error vector, $\delta\bar{y}_k = \Delta\bar{r}_{VO,k} - \Delta\bar{r}_{INS,k}$, is constructed as,

$$\tilde{C}^{\emptyset}_B\int_{t_{k-1}}^{t_k}\left((C^N_{B(t)})^T[\times\tilde{\tilde{v}}^N_{INS}]\delta\bar{\phi}^N(t) - (C^N_{B(t)})^T\delta\bar{v}^N(t)\right)dt - \tilde{C}^{\emptyset}_B diag(\Delta\bar{r}_{INS,k})\delta\bar{m}_{bs,k} - \bar{\nu}_{ACC,k} \quad (10)$$

2) Attitude Change Measurement Model: The attitude change measurement over the Kalman cycle can be obtained by the available incremental attitude change measurements as,

$$[C^{C_k}_{C_{k-1}}]_{VO} = \prod_{m=1}^{N}[C^{C_{k-1[m-1]}}_{C_{k-1[m]}}]_{VO} \quad (11)$$

Then visual odometry attitude change error measurement can be constructed as follows,

$$[C^{\emptyset_k}_{\emptyset_{k-1}}]_{VO}[C^{\emptyset_k}_{\emptyset_{k-1}}]^T_{INS} \approx I_{3x3} + \Delta\Psi^{\emptyset_k}_{mea} + \Delta\Psi^{\emptyset_k}_{INS} \quad (12)$$

where $[C^{\emptyset_k}_{\emptyset_{k-1}}]_{INS} = \tilde{C}^{\emptyset}_B(\tilde{C}^N_{B_k})^T\tilde{C}^N_{B_{k-1}}(\tilde{C}^{\emptyset}_B)^T$ is the INS calculated accumulated attitude change measurement. Here, $\Delta\Psi^{\emptyset_k}_{mea} = [\times\Delta\psi^{\emptyset_k}_{mea}]$ is the VO attitude measurement error matrix. $\Delta\Psi^{C_k}_{INS} = [\times\Delta\psi^{\emptyset_k}_{INS}]$ where,

$$\Delta\Psi^{\emptyset_k}_{INS} = \tilde{C}^{\emptyset}_B(\tilde{C}^N_{B_k})^T(\delta\bar{\phi}^N_k - \delta\bar{\phi}^N_{k-1}) \quad (13)$$

C. Delayed-State Kalman Filtering

Note that the derived Model I delta-position and attitude change measurement models include (one sample) delayed elements of the error state vector which is not suitable for the direct implementation of the Kalman filter. Similarly, Model II delta position measurement includes time integral of the error state vector in the measurement interval. In [15] and [16], the delayed-state measurement case is handled by stochastic cloning method where a duplicate of the pose error states, that is the attitude and position error states, are used as the traditional states of the Kalman filter. The adaptation of the delayed-state measurements to Kalman filter is a known problem in estimation theory. This type of measurement model is handled by the widely used delayed-state measurement Kalman filter [19]. In this method, the measurement model is represented in the following form,

$$\delta\bar{y}_k = \Delta\bar{r}^{\emptyset_k}_{VO,k} - \Delta\bar{r}^{\emptyset_k}_{INS,k} = H_k\delta\bar{x}_k + J_k\delta\bar{x}_{k-1} + \bar{\nu}_k \quad (14)$$

where H_k and J_k are the measurement matrices that relates the measurement errors to the current and delayed states respectively. By using the linear error dynamics, the delayed error state is expressed in terms of the current error state as follows,

$$\delta\bar{x}_{k-1} = \Phi^{-1}_{k,k-1}\delta\bar{x}_k - \Phi^{-1}_{k,k-1}\bar{w}_k \quad (15)$$

Then the measurement model takes the following form,

$$\delta\bar{y}_k = (H_k + J_k\Phi^{-1}_{k,k-1})\delta\bar{x}_k + (-J_k\Phi^{-1}_{k,k-1}\bar{w}_k + \bar{\nu}_k) \quad (16)$$

Note that the measurement model includes the measurement noise that is correlated with the process noise. Using the Kalman filter equations that are derived for the correlated measurement and process noise, optimal state estimation equations can be derived for the delayed-state measurement [19].

We propose to utilize a more practical method, as suggested in [21] and [22] for INS-GPS integration, using the linear system error dynamics by ignoring the effect of the correlation between the process and measurement noise within the time interval. They used this technique for processing the time-differenced carrier phase measurements of GPS in INS-GPS Kalman filter. In this regard, the delta-position measurement vector for Model I can be re-expressed as,

$$\delta\bar{y}_k = [-\tilde{C}^{\emptyset}_B(\tilde{C}^N_{B_k})^T(\times\Delta\bar{r}^N_{INS,k})]\delta\bar{\phi}^N_k + C^{\emptyset}_B(\tilde{C}^N_{B_k})^T\int_{t_{k-1}}^{t_k}\delta\bar{v}^N(t)dt + [\tilde{C}^{\emptyset}_B(\tilde{C}^N_{B_k})^T(\times\Delta\bar{r}^N_{INS,k}) + (\times\tilde{C}^{\emptyset}_B(I - \tilde{C}^{B_k}_{B_{k-1}})\bar{m}^B_{la})]\delta\bar{m}_{bs,k} + \tilde{C}^{\emptyset}_B(I - \tilde{C}^{B_k}_{B_{k-1}})\delta\bar{m}^B_{la} - diag(\bar{r}^{\emptyset_{k-1}}_{VO,k})\delta\bar{m}_{bs,k} - \delta\bar{m}_{b,k} - \bar{\nu}_{ACC,k} \quad (17)$$

Velocity error vector at a given time t in the time interval $[t_{k-1}, t_k]$ can be expressed in terms of the final error state vector as,

$$\delta\bar{v}^N(t) = H(t)\delta\bar{x}(t) \approx H(t)\Phi(t, t_{k-1})\Phi^{-1}_{k,k-1}\delta\bar{x}(t_k) \quad (18)$$

where $\Phi(t, t_{k-1})$ is the state transition matrix that maps the the state vector at time $k-1$ to the time t. Then the error state

vector at the final time can be extracted in the time interval by utilizing the state-transition matrix as defined in [22],

$$\int_{t_{k-1}}^{t_k} \delta\bar{v}^N(t))dt = \left(\int_{t_{k-1}}^{t_k} H(t)\Phi(t,t_{k-1})dt\right)\Phi_{k,k-1}^{-1}\delta\bar{x}(t_k) \tag{19}$$

where the integral term $\int_{t_{k-1}}^{t_k} H(t)\Phi(t,t_{k-1})dt$ can be calculated during the measurement time interval with the increments Δt recursively as defined in [22],

$$\begin{aligned}\int_{t_{k-1}}^{t_{k-1}+i\Delta t} & H(t)\Phi(t,t_{k-1})dt \cong \\ & H(t_{k-1}+i\Delta t)\Phi(t_{k-1}+i\Delta t,t_{k-1})\Delta t \\ & + \int_{t_{k-1}}^{t_{k-1}+(i-1)\Delta t} H(t)\Phi(t,t_{k-1})dt \end{aligned} \tag{20}$$

The same approach can also be applied for Model II delta position and attitude measurements as well.

D. Robustness to Visual Odometry Outliers

Visual odometry, which has a satisfactory performance in most conditions, is not totally a reliable sensor due to its performance dependency on spurious data such as independently moving objects and sudden scene condition changes. Thus, robustness of the VO-INS integration filter is necessary. However, it is known that the Kalman filter is susceptible to the outliers due its linear structure. A single outlier can deteriorate the performance of Kalman filter significantly. Different approaches have been proposed in literature that makes the Kalman filter more robust to outliers. A common technique is performing a sigma-check to the measurement residual. If the current measurement residual is greater than the predicted covariance of the residual, then the measurement is discarded. Another way would be using a kind of robust M-estimate [23] that uses a nonlinear response function rather than linear one.

V. EXPERIMENTAL RESULTS

The VO-INS integration experiments are performed on KITTI Vision Benchmark Suite [18]. The dataset consists of stereo image sequences, raw IMU data, ground truth INS-GPS data and Camera-to-IMU rotation/translation parameters. The selected sequence is captured in a residential area and its duration is 520 seconds. The stereo camera directs to the front of the car. The image sequence, IMU and ground truth data is time stamped for proper synchronization. The stereo sequence is captured at 10 Hz and the IMU data is collected at 100 Hz. The IMU uses three solid-state MEMS angular rate sensors and three servo (force-feedback) accelerometers. The gyroscope and accelerometer bias specifications are announced as $36deg/hr, 1\sigma$ and $10mm/s^2(1milig), 1\sigma$ respectively. The stereo images are grayscale and image size is 1226x370 pixels. The ground truth, VO only and VO-INS integrated 2D position results are displayed in 4; corresponding 2D position errors are displayed in Figure 5. Note the position error deviation around 100^{th} second in Figure 5. This deviation is due to loss of GPS satellites in the INS-GPS (ground-truth) data. As expected, VO-IMU or VO data is not affected from this

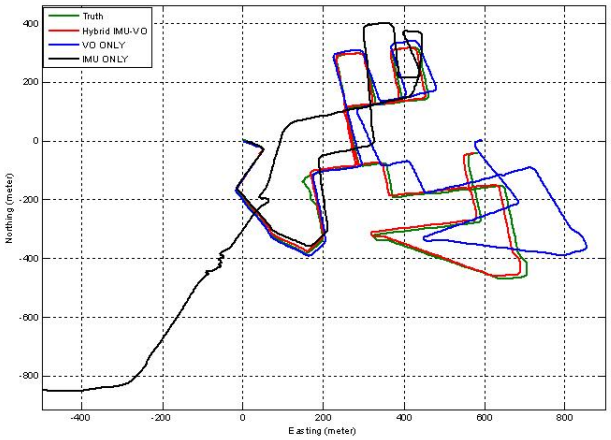

Fig. 4. 2D Position Comparison between ground truth INS-GPS, IMU Only, VO Only and Hybrid IMU-VO data

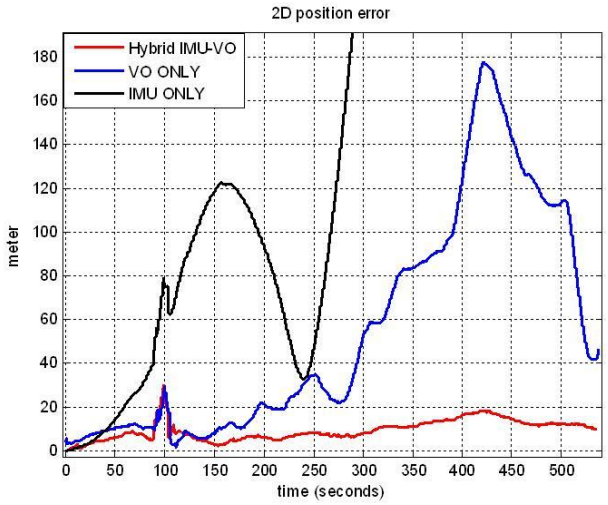

Fig. 5. 2D Position Errors of Hybrid IMU-VO, VO Only and IMU Only solutions

jump. This scenario is selected on purpose to support the proposition that GPS may not be feasible in cluttered urban environments and visual aiding is a feasible alternative to GPS. The results show that, it is possible to get admissible position performance ($\sim 0.0025 \times DistanceTraveled$) by integration of visual odometry to inertial navigation, even with a low performance inertial measurement unit. As expected integration of visual odometry cannot eliminate the drift in inertial navigation totally since VO is not an absolute sensor. However, as shown in the results, it can significantly suppress the drift.

In Figures 6 and 7, Kalman Filter estimates for gyroscope bias and accelerometer bias for all three axes are depicted respectively. The bias estimates are coherent with the specified values in IMU datasheet.

Visual odometry scale factor is modeled as a random walk process in the filter. The scale factor emanates mainly from the stereo baseline ambiguity and camera calibration error. In Figure 8, the scale factor estimate throughout the run is given. It can be observed that the scale factor converges to a constant

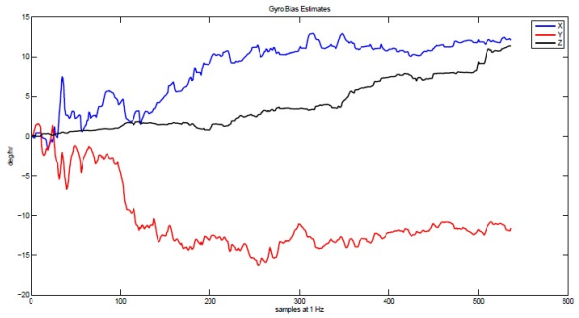

Fig. 6. Kalman Filter Gyroscope Bias Estimates (deg/hr)

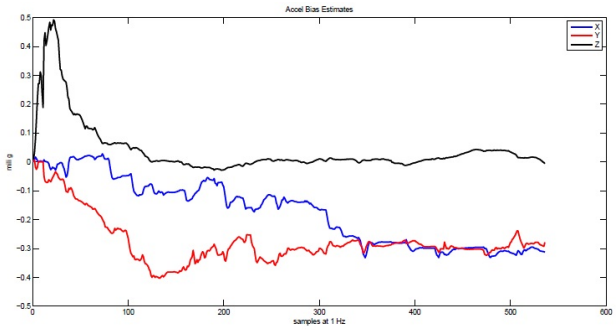

Fig. 7. Kalman Filter Accelerometer Bias Estimates (milli-g)

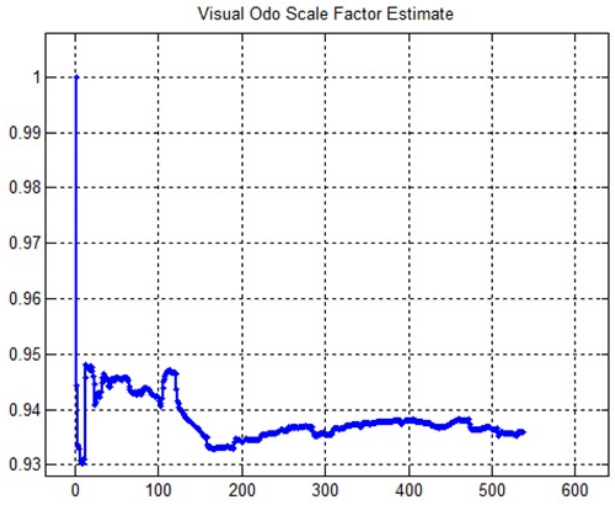

Fig. 8. Kalman Filter Visual Odometry Scale Factor Estimate

value; this value is verified with ad-hoc methods.

In order to test the robustness of the filter, errors are injected to the visual odometry data manually. For this specific dataset, visual odometry data is reliable. However, sudden jumps or correlated errors are probable in visual odometry, especially when the camera is directed to sky or the number of feature inliers are small and/or unreliable. The 3σ test is supposed to handle erroneous data in visual odometry. In Figure 9, the error injected visual odometry 3D speed is shown together with ground truth and hybrid VO-IMU speed. It can be observed that the 3σ test avoids the use of erroneous visual odometry data and the filter continues with inertial

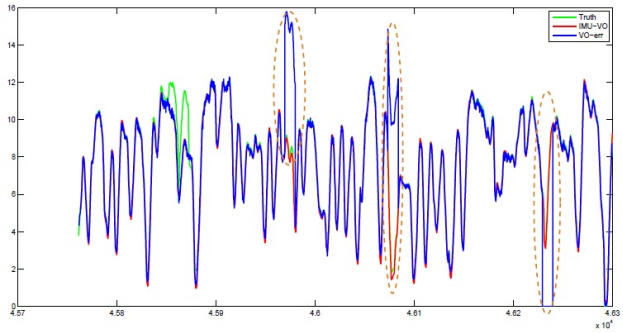

Fig. 9. 2D Speed (m/sec) Comparison between ground truth INS-GPS, error injected VO and Hybrid IMU-VO solutions

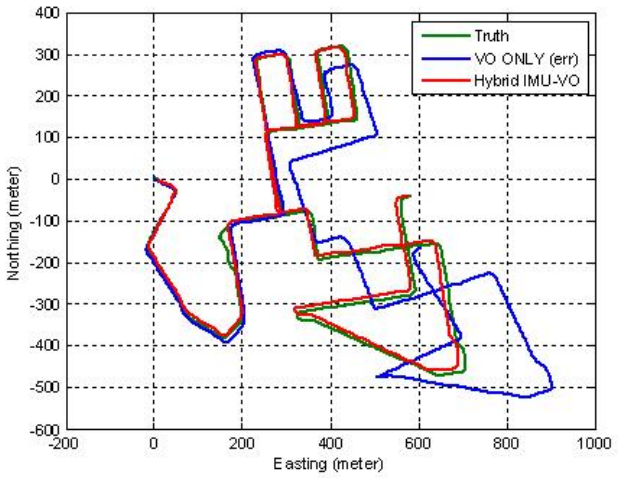

Fig. 10. 2D Position Comparison between ground truth INS-GPS, Hybrid IMU-VO and error injected VO-IMU solutions

measurements during this period.

In Figure 10, 2D position results of ground truth INS-GPS, Hybrid VO-IMU and error injected Hybrid VO-IMU are displayed. It can be observed that the error injected VO does not affect the result since the 3σ test avoids the use of erroneous visual odometry data in the filter.

VI. Conclusion

In this paper, a loosely coupled indirect feedback Kalman Filter is presented for integration of visual odometry and inertial navigation system which exploits the complementary characteristics of the visual and inertial sensors. The non-Gaussian, non-stationary and correlated error characteristics of the visual odometry motion estimate is explored and practical methods are proposed to tackle the problems induced by these error characteristics. Visual odometry pre-filtering and measurement sigma-test in the Kalman filter are introduced for this purpose. Two measurement models are derived for the accumulated and incremental visual odometry measurements. A practical measurement model approach is proposed for the delta position and attitude change measurements that inherently includes delayed-state. 6-dof visual odometry motion estimates (rotational and translational) are used as the measurement. Experimental results are presented for a real dataset

collected with a land vehicle in an urban environment. The results show that it is possible to get admissible navigation performance even with a low performance inertial sensor by integration of visual odometry to inertial navigation.

REFERENCES

[1] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 652–659, 2004.

[2] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80 –92, dec. 2011.

[3] C. N. Taylor, "Improved fusion of visual measurements through explicit modeling of outliers," *Position Location and Navigation Symposium (PLANS), IEEE/ION*, 2012.

[4] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/358669.358692

[5] M. Agrawal and K. Konolige, "Rough terrain visual odometry," *Proceedings of the International Conference on Advanced Robotics (ICAR)*, August 2007.

[6] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," in *IEEE Intelligent Vehicles Symposium*, San Diego, USA, June 2010.

[7] K. Konolige, M. Agrawal, and J. Sol, "Large scale visual odometry for rough terrain," in *In Proc. International Symposium on Robotics Research*, 2007.

[8] R. Subbarao, Y. Genc, and P. Meer, "Nonlinear mean shift for robust pose estimation," in *Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision*, ser. WACV '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 6–. [Online]. Available: http://dx.doi.org/10.1109/WACV.2007.44

[9] E. Tola and A. A. Alatan, "Fast outlier rejection by using parallax-based rigidity constraint for epipolar geometry estimation," in *Proceedings of the 2006 international conference on Multimedia Content Representation, Classification and Security*, ser. MRCS'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 578–585.

[10] R. Jiang, R. Klette, and S. Wang, "Modeling of unbounded long-range drift in visual odometry," in *Proceedings of the 2010 Fourth Pacific-Rim Symposium on Image and Video Technology*, ser. PSIVT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 121–126. [Online]. Available: http://dx.doi.org/10.1109/PSIVT.2010.27

[11] M. Veth and J. Raquet, "Fusion of low-cost imaging and inertial sensors for navigation," Air Force Institute of Technology, Tech. Rep., 2007.

[12] J. Jurado, K. Fisher, and M. Veth, "Inertial and imaging sensor fusion for image-aided navigation with affine distortion prediction," in *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, april 2012, pp. 518 –526.

[13] J.-O. Nilsson, D. Zachariah, M. Jansson, and P. Handel, "Realtime implementation of visual-aided inertial navigation using epipolar constraints," in *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, april 2012, pp. 711 –718.

[14] D. D. Diel, P. DeBitetto, and S. Teller, "Epipolar constraints for vision-aided inertial navigation," in *In EEE Workshop on Motion and Video Computing*, 2005, pp. 221–228.

[15] J.-P. Tardif, M. D. George, M. Laverne, A. Kelly, and A. Stentz, "A new approach to vision-aided inertial navigation." in *IROS*. IEEE, 2010, pp. 4161–4168.

[16] S. Roumeliotis, A. Johnson, and J. Montgomery, "Augmenting inertial navigation with image-based motion estimation," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 4, 2002, pp. 4326 – 4333 vol.4.

[17] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[18] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[19] R. G. Brown and P. Y. C. Hwang, *Introduction to random signals and applied kalman filtering*. New York, NY: Wiley, 1997.

[20] D. Titterton, J. Weston, and I. of Electrical Engineers, *Strapdown Inertial Navigation Technology*, ser. Iee Radar Series. Institution of Electrical Engineers, 2004. [Online]. Available: http://books.google.com.tr/books?id=WwrCrn54n5cC

[21] J. Farell, "Carrier phase processing without integers," in *ION 57th Annual Meeting*, 2001.

[22] J. Wendel and G. F. Trommer, "Tightly coupled gps/ins integration for missile applications," *Aerospace Science and Technology*, vol. 8, pp. 627–634, 2004.

[23] P. Huber, *Robust statistics*. John Wiley & Sons, 2003.