2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Inertial Navigation Aided by Monocular Camera Observations of Unknown Features

Michael George and Salah Sukkarieh

Abstract— **This paper presents an algorithm which can effectively constrain inertial navigation drift using monocular camera data. It is capable of operating in unknown and large scale environments and assumes no prior knowledge of the size, appearance or location of potential environmental features. Low cost inertial navigation units are found on most autonomous vehicles and a large number of smaller robots. Depending on the grade of the sensor, when used alone, inertial data for control and navigation will only be reliable for a matter of seconds or minutes. An algorithm is presented that simultaneously estimates relative feature location in sensor space and inertial position, velocity and attitude in world coordinates. Feature locations are maintained in sensor space to ensure measurement linearity. Image depth is represented by an inverse function which permits un-delayed feature initialization and improves linearity and convergence. It is shown that the resulting navigation solution is able to be constrained, providing results comparable to inertial-GPS systems. Results are presented for an autonomous aircraft operating in a large semi-structured environment.**

I. INTRODUCTION

The problem of aiding inertial navigation sensors has received a substantial amount of attention, particularly in the aerospace literature where inertial sensors found their first application. In the most common scenario, inertial sensors are integrated with GPS measurements providing a solution that exploits the complementarity in these two sensors. In the robotics community efforts have focused on vehicle model and camera aiding, due in part to a requirement for operation without GPS sensors. However, there are no published methods of camera aided inertial navigation in the most general sense. By this we mean suitable for autonomous vehicles with purely monocular data, providing a constraint on position, velocity and orientation in three dimensions, requiring no separate depth sensors and making no prior assumptions about environmental landmarks or conditions. This paper presents such a formulation. It is demonstrated using autonomous aircraft data but is well suited to many other applications including urban, underwater, indoor and interplanetary environments.

Inertial navigation sensors provide a number of desirable attributes to autonomous vehicles. They provide six-degree-of-freedom data at high frequency for vehicle control, they are self-contained and modern electro-mechanical varieties are small, light weight and rugged. The principal drawback of such sensors is their dead-reckoning nature and the fact that navigation variables are time integrals of their measurements. This leads to unconstrained drift in position, velocity and attitude if some secondary sensor is not employed to periodically calibrate or constrain the inertial sensors.

Similarly, monocular vision is a particularly favored solution to the perception problem in the robotics community, owing to its availability, cost, and in the visual spectrum, its human friendly format. The two sensors therefore, are natural candidates for the approach presented here.

This paper is organized as follows. Section II presents a review of relevant literature and defines this papers contribution, section III sets out the mathematical formulation of the problem including the estimation structure, section IV presents results from a test implementation on autonomous aircraft data and section V concludes with future directions.

M. George and S. Sukkarieh are with the Australian Centre for Field Robotics, Department of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, NSW 2006, Australia.
{m.george,salah}@acfr.usyd.edu.au

II. PREVIOUS WORK

There are a number of fields where relevant work is being carried out with cross fertilization starting to occur more often. The computer vision and robotics literature both contain examples of camera aided inertial sensing systems and we also take inspiration from a number of developments in the SLAM literature. The sub-headings we use below reflect content as we judge it and not necessarily publication type or author affiliation.

A. Computer Vision

In [1], a stereo vision technique with inertial sensors and a known target is presented. The stereo vision sensors eliminate the image depth ambiguity for their application in small indoor scenes. Camera translations, on the order of centimeters and larger rotations are compensated with inertial measurements. In [2] image contours are tracked relative to an initial template with camera motion estimates compensated by inertial sensors. The template depth is manually entered and a scheme for automatic calculation is deferred to future work. In [3] inertial sensors are used as a vertical reference for a stereo vision system. The gravity vector defines the reference from which images can be segmented and focal length calibrated. In [4] and [5], an epipolar constraint filter is presented, which adds a constraint to the possible camera motion based on consecutive epipolar segments. The author's state that a deliberate bias is introduced for each feature used to constrain the inertial drift. This is done as a trade off for computational efficiency but is justified by their statement that a large number of well spaced features will mitigate the total effect of these combined biases. In addition, the technique does not constrain vehicle attitude.

1-4244-0602-1/07/$20.00 ©2007 IEEE.

B. Robotics and Autonomous Vehicles

We focus particularly here on outdoor applications as the robotics literature dealing with indoor applications is substantially consistent with the computer vision literature. The first application of camera aided inertial navigation was developed for the aiding of strategic missiles in their boost phase [6]. The inertial systems in use here were highly specialized and this technique required a single sighting of a known star to correct for initial alignment errors. It is of historical interest but provides no direct technical insight into our work. In [7] a low flying unmanned helicopter integrates motion estimates derived from stereo vision optical flow and inertial sensors. In [8] and [9] a NASA application is presented where the horizontal components of interplanetary lander descent velocity are deduced from inertial sensors and vision. This method relies on laser altimetry to estimate the depth in the image. Applications in environments containing landmarks with known size and location have been reported in [10] and [11]. These scenarios require the vehicle to observe and associate features that it has complete prior knowledge of, which severely limits this technique's applicability in new or unstructured environments. In [12] and [13] a scheme is presented for using bearing only observations of unknown landmarks to aid inertial sensors. The scheme, however relies of sophisticated stadiametric optics mounted on gimbals and requires a constant vehicle velocity vector for up to a minute at a time, restricting it to specific military applications in manned vehicles. No results are presented and the formulation is left as a mathematical concept.

C. Other Aided INS Approaches

It is worth briefly mentioning the other non-GPS inertial aiding approaches in the literature. In [14] an aircraft vehicle model is used to aid inertial sensors. Conceptually this approach bounds the allowable inertial dynamics to the vehicle flight envelope and in this way is able to constrain inertial drift. A similar scheme for ground vehicle applications was demonstrated in [15]. A ground vehicle model aided INS for periods of interrupted GPS was reported to be an important component of the recent DARPA Grand Challenge winner [16].

D. SLAM

Two additional publications that have influenced the current work are [17] and [18]. The first presents a SLAM formulation with an inverse depth parametrization for relative feature location. This formulation of the SLAM problem gives improved linearization characteristics and allows undelayed feature initialization in a bearings only setting. In [18] a methodology for relative position sensing using inertial and monocular camera data in grasping applications is developed. The authors present an argument, which we adopt, for sensor centric feature representation that places all non-linearities in the prediction model and subsequent unscented filtering. The inverse depth characterization is also presented here but without elaboration. The work in [19] presents two approaches, with different feature parameterizations but otherwise similar in state structure to [18]. The first uses a Levenberg-Marquardt batch optimization and the second an IEKF recursive formulation. In the recursive case, scene features are added to the state in a delayed fashion and are parameterized directly in cartesian coordinates by using the batch optimization over the set of stored measurements.

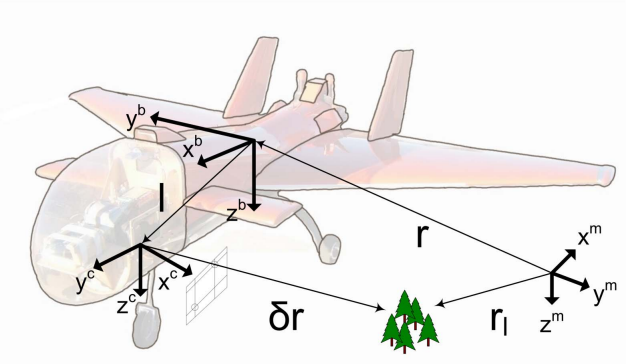

Fig. 1. Problem Geometry and Reference Frame Definitions

E. Contribution

We aim for a technique that allows aiding of an inertial navigation system (INS) especially during periods when GPS is unavailable. In our application, a UAV conducting a tracking mission, this occurs when the aircraft is banking and turning steeply and the GPS antennae are shaded by the fuselage. The technique should take advantage of well established real-time estimation techniques. In practice, this means a solution that is suited to some variation of the Kalman Filter. In this work we use the Unscented Filter [20]. We take a feature based approach to the vision processing and store feature information in our estimated state. Feature numbers can be scaled on-line to suit processing ability. Future work will look at the observability of vehicle dynamics given the number of features available. In generating the theory, we make no assumption about what types of features are used or what environmental conditions the vehicle encounters. In practice we use a template matching feature extraction technique but this is generalizable to any method that returns feature centroids in the image. We take advantage of the inverse depth method from [17] and we extend the work in [18] to incorporate a globally referenced position and multiple features. Unlike [18] and [19] we take the approach of previous inertial-GPS systems and use the inertial measurement for prediction rather than update in the estimation.

III. MATHEMATICAL FORMULATION

A. Problem Geometry

We take as our starting point a locally level frame with axes triplet corresponding to north, east and down directions relative to the local surface. This frame is referred to as the mechanization frame and denoted with superscript 'm'. We approximate an inertial frame with this mechanization frame.

This approximation is appropriate to local movement on the order of kilometers [21]. We define a vehicle body frame centered at the inertial sensor cluster within the vehicle, denoted by superscript 'b'. A sensor (camera) frame, denoted by superscript 'c', is defined at the camera focal point offset from the vehicle body frame by a constant lever arm, $\mathbf{l}$. These three frames are depicted in Fig. 1. A direction cosine matrix (DCM) may be defined to transform vector components in one frame into any other frame. Such a construct is denoted $\mathbf{C}_i^j$ if transforming components from frame 'i' into frame 'j'. Free vectors are denoted without superscripts but where appropriate for concrete realization, a superscript will be added to define which frame the vector is decomposed in.

B. Filter Structure

The simplest mechanization frame for inertial navigation is the local level frame discussed above but it is expected that the resulting formulations could be extended to other inertial mechanizations. The equations of navigation (See [21] for a derivation) for this frame are

$$\mathbf{r} = \iint \dot{\mathbf{v}} + \mathbf{r}_0 \tag{1}$$

$$\dot{\mathbf{v}} = \mathbf{C}_b^m(\mathbf{f}^b - \mathbf{\Delta f}^b) - (2\mathbf{\Omega}_{ie}) \times \mathbf{v} + \mathbf{g}_l \tag{2}$$

$$\mathbf{g}_l = \mathbf{g}_m - \mathbf{\Omega}_{ie} \times (\mathbf{\Omega}_{ie} \times \mathbf{r}) \tag{3}$$

$$\mathbf{\Psi} = \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} \tag{4}$$

$$\dot{\mathbf{\Psi}} = E(\mathbf{\Psi})(\omega^b - \mathbf{\Delta}\omega^b) \tag{5}$$

$$E(\mathbf{\Psi}) = \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi\sec\theta & \cos\phi\sec\theta \end{bmatrix} \tag{6}$$

$$\dot{\mathbf{\Delta f}}^b = 0 \tag{7}$$

$$\dot{\mathbf{\Delta}\omega}^b = 0 \tag{8}$$

Where $\mathbf{v}$ is the time derivative of position, $\mathbf{\Omega}_{ie}$ is the constant Earth rotation rate vector, $\mathbf{g}_m$ is the mass gravitation vector, $\mathbf{g}_l$ is a 'local' gravitation vector incorporating centripetal acceleration effects from the Earth's rotation, $\mathbf{\Psi}$ is an array of angles representing vehicle roll, pitch and yaw respectively (collectively termed Euler angles), $\mathbf{C}_b^m$ is a function of $\mathbf{\Psi}$, $\mathbf{f}^b$ and ω^b are the inertial sensor measurements of specific force and angular velocity, respectively, both made in the vehicle body frame and $\mathbf{\Delta f}^b$, $\mathbf{\Delta}\omega^b$ are the inertial sensor measurement biases which much be estimated online for fast drifting sensors or operations over long periods of time. These represent the traditional non-linear continuous time processes of inertial navigation. We will alter them in a limited fashion to give an intuitive understanding of the approach this work implements.

First we assume a stationary landmark and ignore aerodynamic flexibility to write a relative velocity equation between the vehicle and any features it may be observing

$$\begin{aligned} \delta\mathbf{v} &= \frac{d}{dt}\delta\mathbf{r} && (9)\\ &= \frac{d}{dt}(\mathbf{r}_l - \mathbf{r} - \mathbf{C}_b^m\mathbf{l}^b) && (10)\\ &= -\mathbf{v} - \mathbf{C}_b^m[\omega^b\times]\mathbf{l}^b && (11) \end{aligned}$$

The notation, $[\mathbf{n}\times]$ represents the skew-symmetric cross product matrix of vector $\mathbf{n}$. Taking the next derivative

$$\begin{aligned} \dot{\delta\mathbf{v}} &= -\dot{\mathbf{v}} - \mathbf{C}_b^m[\dot{\omega}^b\times]\mathbf{l}^b - \mathbf{C}_b^m[\omega^b\times]^2\mathbf{l}^b && (12)\\ &= -\mathbf{C}_b^m(\mathbf{f}^b - \mathbf{\Delta f}^b) + (2\mathbf{\Omega}_{ie}) \times \dot{\mathbf{r}} - \mathbf{g}_l && (13)\\ &\quad -\mathbf{C}_b^m[\dot{\omega}^b\times]\mathbf{l}^b - \mathbf{C}_b^m[\omega^b\times]^2\mathbf{l}^b && (14) \end{aligned}$$

Rearranging (11) also gives

$$\dot{\mathbf{r}} = -\delta\mathbf{v} - \mathbf{C}_b^m[\omega^b\times]\mathbf{l}^b \tag{15}$$

We define an estimation problem with state and initial covariance

$$\mathbf{x} = \begin{bmatrix} \mathbf{r} \\ \delta\mathbf{v} \\ \mathbf{\Psi} \\ \mathbf{\Delta f}^b \\ \mathbf{\Delta}\omega^b \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} \sigma^2_{\mathbf{r}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma^2_{\delta\mathbf{v}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2_{\mathbf{\Psi}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2_{\mathbf{\Delta f}^b} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2_{\mathbf{\Delta}\omega^b} \end{bmatrix} \tag{16}$$

The state derivatives are defined by equations (15), (14), (5), (7) and (8) which are discretized using a first order model at 400 Hz (the inertial sensor frequency) and estimated using the unscented filter [20] driven by white noise on the measurements $\mathbf{f}^b$ and ω^b as defined by

$$\mathbf{Q} = \begin{bmatrix} \sigma^2_{\mathbf{f}^b} & \mathbf{0} \\ \mathbf{0} & \sigma^2_{\omega^b} \end{bmatrix} \tag{17}$$

Initial covariances are determined from calibration sensitivities for velocity, attitude and sensor biases and a GPS fix on runway location for the position.

C. Camera Model

Let $\delta\mathbf{r}^c$ be the realization of vector $\delta\mathbf{r}$ in the camera reference frame resulting from a transformation from the mechanized frame according to $\mathbf{C}_m^c$. That is

$$\begin{aligned} \delta\mathbf{r}^c &= \mathbf{C}_m^c\delta\mathbf{r}^m && (18)\\ &= \mathbf{C}_b^c\mathbf{C}_m^b\delta\mathbf{r}^m && (19)\\ &= \mathbf{C}_b^c\mathbf{C}_b^{mT}\delta\mathbf{r}^m && (20) \end{aligned}$$

where $\mathbf{C}_b^c$ is a fixed calibrated transformation. The projection of this vector onto the image plane results in the camera measurement. We use a modified pinhole model which defines pixel coordinates by

$$\begin{aligned} u &= f_u\frac{\delta r_y^c}{\delta r_x^c} && (21)\\ v &= f_v\frac{\delta r_z^c}{\sqrt{(\delta r_x^c)^2 + (\delta r_y^c)^2}} \cdot \sqrt{1 + \frac{(\delta r_y^c)^2}{(\delta r_x^c)^2}} && (22) \end{aligned}$$

where f_u and f_v are the camera focal lengths and the subscript denotes the respective component of the vector. In [17] and [18] an inverse depth formulation of the bearings only SLAM problem is presented wherein linearity and improved estimation properties result. Following those publications we define

$$\rho = \frac{1}{\delta r_x^c} \tag{23}$$

and rewrite (21) and (22) as

$$u = f_u \rho \delta r_y^c \tag{24}$$

$$v = f_v \frac{\delta r_z^c}{\sqrt{\frac{1}{\rho^2} + (\delta r_y^c)^2}} \cdot \sqrt{1 + \rho^2 (\delta r_y^c)^2} \tag{25}$$

D. Feature Initialization

The methods presented are applicable to any feature extraction technique that returns the centroid of a feature in image coordinates. In our implementation we have used a template matching technique that extracts small plastic targets, cars and buildings in the environment (Fig. 2).
Given a set of *p* features from an image processing algorithm we augment the unscented filter process with pixel coordinates u, v and initial inverse depth ρ such that

$$\mathbf{x}_{\text{aug}} = \begin{bmatrix} \mathbf{x} \\ u_1 \\ v_1 \\ \rho_1 \\ \vdots \\ u_i \\ v_i \\ \rho_i \end{bmatrix} \quad \mathbf{P}_{\text{aug}} = \begin{bmatrix} \mathbf{P} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_1 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{P}_i \end{bmatrix} \quad i \in 1:p \tag{26}$$

$$\mathbf{P}_i = \begin{bmatrix} \sigma_{u_i}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{v_i}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{\rho_i}^2 \end{bmatrix} \tag{27}$$

The image covariances σ_{u_i} and σ_{v_i} are expressed in pixels as defined by camera resolution and chosen feature extraction technique and σ_ρ is chosen according to the heuristic algorithm presented in [17]. The initial inverse depth ρ is a somewhat arbitrary selection and we make the choice on very approximate prior knowledge of the environment. It will be shown later however that the estimation scheme is robust to enormous variations in this initial condition (order of meters to kilometers), making the algorithm capable of operating in environments with very large and very small scales simultaneously. The number of augmented features, *p*, is scalable and in the results here was set to a maximum of six. Features are deleted when they are predicted to be outside the scene and new ones, if available, are added. In this way, computational burden can be controlled.

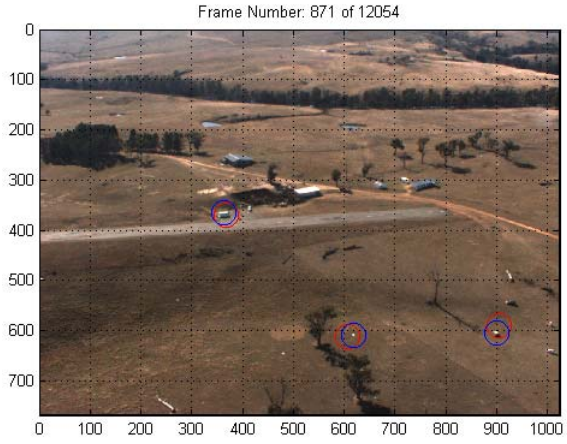

Fig. 2. Example vision frame. Small, plastic targets are placed in the environment and along with cars and buildings are extracted using a template matching algorithm. Blue circles represent feature extraction coordinates (measurements), red circles represent feature coordinates from the localization filter (predictions).

E. Measurement Equation

Given a set of q initialized and associated features the measurement equation is the linear mapping

$$\mathbf{z} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_i \\ v_i \end{bmatrix} \quad i \in 1:q \tag{28}$$

This formulation eliminates linearization about uncertain range estimates which adds significant robustness over other approaches to bearing only estimation. Features are associated with their predictions using the joint compatibility test described in [22]. This provides a robust means of eliminating spurious matches and minimizes any 'flickering' in the associations of closely spaced features which are unreliably extracted.

F. Augmented State Prediction

Given the augmented states described above, which are kept in sensor centric coordinates it is necessary to derive a prediction equation. This process will describe the constraint applied to the inertial drift by the camera measurements. As a reference we firstly calculate

$$\frac{d}{dt}\delta\mathbf{r}^c = \dot{\mathbf{C}}_b^c \mathbf{C}_b^{mT} \delta\mathbf{r} + \mathbf{C}_b^c \dot{\mathbf{C}}_b^{mT} \delta\mathbf{r} + \mathbf{C}_m^c \delta\dot{\mathbf{r}} \tag{29}$$

$$= \mathbf{C}_b^c (\mathbf{C}_b^m [\omega^b \times])^T \delta\mathbf{r} + \mathbf{C}_m^c \delta\dot{\mathbf{r}} \tag{30}$$

$$= -\mathbf{C}_b^c [\omega^b \times] \mathbf{C}_b^{mT} \delta\mathbf{r} + \mathbf{C}_m^c \delta\dot{\mathbf{r}} \tag{31}$$

Now from (23) and (24) we have

$$\dot{\rho} = -\rho^2 \dot{\delta r_x^c} \tag{32}$$

$$\dot{u} = f_u \dot{\rho} \delta r_y^c + f_u \rho \dot{\delta r_y^c} \tag{33}$$

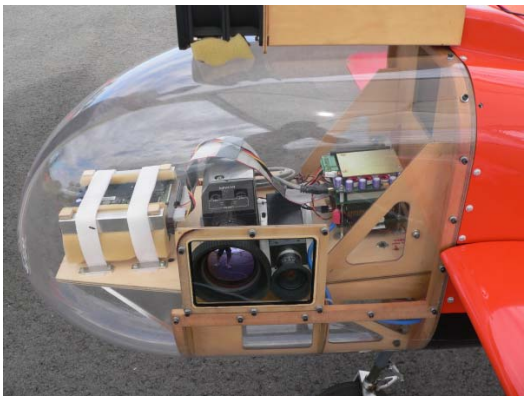

Fig. 3. Vehicle close up. An large infra-red camera (not used in this work) is visible on the left hand side. The smaller camera on the right is the visual spectrum camera used in this work. A processing stack is visible on the far right and a hard drive for logging is visible on the far left.

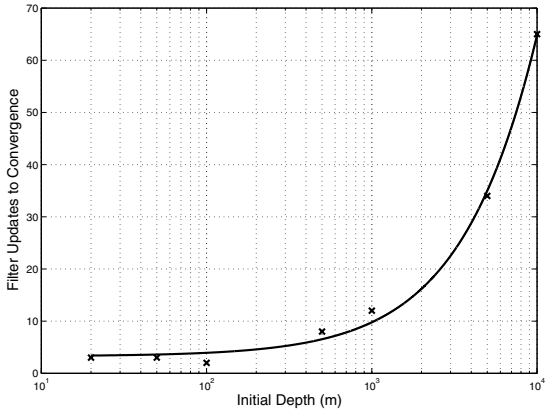

Fig. 4. Initial depth convergence properties. True feature depth is at 200 meters. This log plot shows convergence properties are approximately linear in initial depth error. Future work will investigate the convergence of far features from underestimated initial depths.

and from (25) we have

$$\dot{v} \quad = \quad f_v \cdot \frac{\rho^3 \sqrt{\frac{1}{\rho^2} + (\delta r_y^c)^2} (\frac{\dot{\delta r_z^c}}{\rho} - \delta r_z^c \dot{\delta r_x^c})}{\sqrt{\rho^2 (\delta r_y^c)^2 + 1}} \tag{34}$$

where the component rates of $\delta \mathbf{r}^c$ are taken row-wise from (31). The prediction of (23), (24) and (25) in the unscented filter is accomplished with discretized versions of (32) through (34).

IV. RESULTS

Inertial and image data sets are gathered aboard a UAV operating over a rural environment. Images are collected at 20 *fps* and are synchronized to inertial measurements with GPS time. A view of the sensors we used is found in Fig. 3. The results presented here are post processed from data logged during flight. For comparison we use an INS-DGPS baseline with standard deviation on the order of 2 meters in position, 0.2 meters/second in velocity and 0.5 degrees in attitude [23].

The data is generated from a tracking mission where the aircraft observes small plastic targets, which are used only to simplify the vision processing, cars and buildings. The aircraft flies a low altitude (approximately 125 meters above ground level) trajectory orbiting pre-programmed ground coordinates in a tracking mission. We switch off the GPS updates and observe the performance of the camera aided approach we have presented next to an inertial only solution in the same situation.

A. Feature Initialization and Filter Convergence

Figure 4 demonstrates the convergence of the depth estimate for a newly initialized feature. In this example the true feature location is at approximately 200 meters depth in the image. Initial depths from 20 meters to 10 kilometers are shown to converge. Convergence, even in the worst case scenario occurs within 65 filter updates, which is equivalent to approximately 3.25 seconds of flight. The number of required filter updates, or equivalently, time to depth convergence is approximately linear in the initial depth error. By selecting the initial depth estimate with reasonable prior environmental knowledge, in our case a known maximum altitude and a tracking mission, convergence occurs within 10 filter updates or 0.5 seconds for all encountered features.

B. Ground Observation Trajectory

Here we present a segment of trajectory with the camera aided INS in operation for approximately four minutes. Reliable operation over this time frame without absolute positioning would be a substantial improvement over current INS-GPS navigation when the GPS is unreliable. We commonly find when the aircraft is banking sharply and orbiting features that the GPS antennae are shaded by the fuselage and/or unable to maintain signal tracking given the dynamics. Figure 5 demonstrates velocity and attitude results from the trajectory segment. Velocity and attitude estimates are crucial for vehicle stabilization and control and it is typical to feed the control system high frequency estimates from the inertial sensors. In this application we supply inertial data to the control system at 50 Hz.

Velocity estimates from the inertial system alone display linear drift in time while velocity estimates are constrained around zero using the camera aided technique we have presented. At some points the east velocity results display a correlation between the INS only and camera aided methods. This occurs during periods when image features are lacking. Attitude estimate errors are less pronounced than velocity errors as they are described by random walk from integrated sensor noise. Velocity and position errors, in contrast are induced by sensor noise and compounded by incorrect gravity compensation and specific force resolution.

A two dimensional view of the trajectory is given in Fig. 6 and the corresponding altitude is depicted in Fig. 7. Horizontal positional control of the UAV is important in a mission specific context but is less important for vehicle safety than accurate altitude, velocity and attitude estimates.

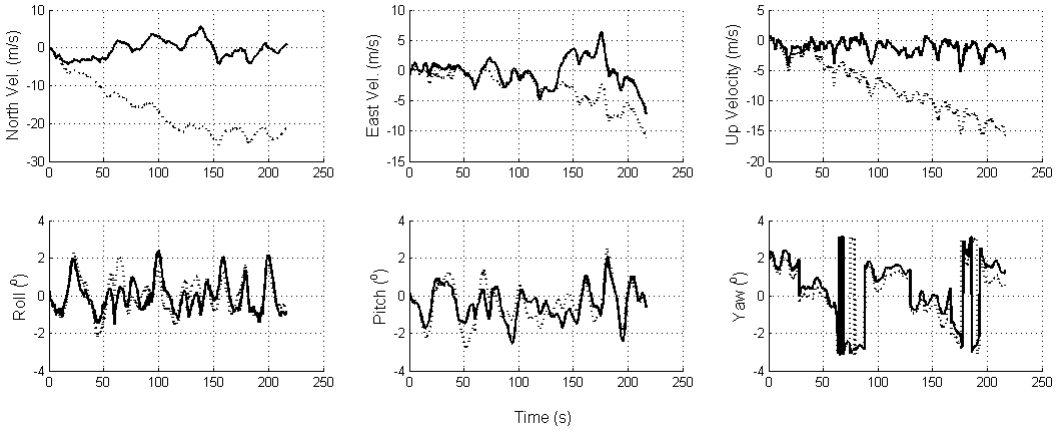

Fig. 5. Deviation of aircraft velocity and attitude parameters from INS-DGPS baseline. Solid lines indicate the camera aided INS algorithm described, dotted lines indicate INS only operation. Velocity components are presented in the top row, attitude components in the bottom row. The data is taken from a segment of the flight path orbiting a number of ground features for approximately four minutes. INS only operation shows substantial drift and becomes unusable early in the segment, in contrast to the proposed camera aided algorithm which is able to successfully constrain velocity and attitude errors during the segment. Accurate estimation of these parameters is essential for vehicle stabilization and control. On this time scale INS only attitude drift is not pronounced. Yaw error is defined in the traditional sense, clockwise from North, the vertical jumps indicate switching signs in the yaw angle deviation.

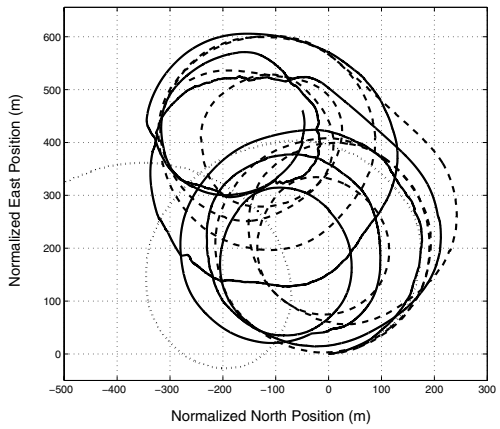

Fig. 6. 2D trajectory. The solid line indicates the camera aided INS algorithm, the dotted line is INS only and the dashed line is the INS-DGPS baseline. The INS only solution continues to drift off the scale while camera aided INS successfully constrains the trajectory. Table I defines the RMS error in position. Altitude is displayed in Fig. 7.

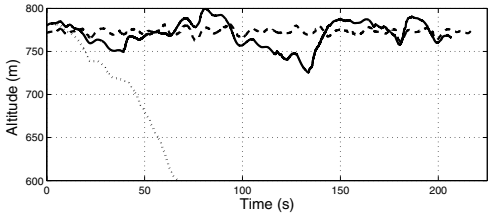

Fig. 7. Vehicle altitude. The solid line indicates the camera aided INS algorithm, the dotted line is INS only and the dashed line is the INS-DGPS baseline. The INS only solution is in error more than 100 meters after just 50 seconds of operation. RMS altitude error is displayed in Table I. The altitude scale is referenced to sea level, ground level is approximately 650 meters.

TABLE I
CAMERA AIDED RMS ERRORS DURING TEST FLIGHT SEGMENT

		RMS Error
Position	North	37.6 m
	East	36.1 m
	Altitude	14.8 m
Velocity	North	2.4 ms^{-1}
	East	2.2 ms^{-1}
	Up	1.1 ms^{-1}
Attitude	Roll	0.86 *deg*
	Pitch	0.87 *deg*
	Yaw	0.78 *deg*

Root mean square errors over the trajectory segment are summarized in Table I. The position and velocity errors are an order of magnitude above typical GPS performance specifications but are still significantly improved compared to the unbounded divergence of inertial only operation.

V. CONCLUSION AND FUTURE WORK

A new method of aiding inertial navigation sensors with monocular camera data was presented. The typical rapid drift in inertial navigation estimates is effectively constrained, adding robustness to periods of operation without other aiding sensors. The method borrows from recent advances in the SLAM literature to allow un-delayed initialization of features in a 3D world with significant scale variations. The algorithm is scalable to computational availability and exhibits naturally efficient data association properties. It is the first demonstration of camera aided inertial navigation capable of handling arbitrary environments with purely bearings only measurements. Future work will focus on extending the timeframe that this method is capable of operating for which at present is largely limited by the practical considerations of sensor placement, feature type and filter tuning. A system

analysis, including observability of bias drift and quantitative filter performance will also be looked at. In addition, the performance during trajectories with substantially steady level flight will be explored by extracting and tracking image features close to the horizon. The detection of moving targets which will corrupt the state estimates as they are currently formulated will also be necessary for operation in more fluid environments. A real time implementation will be considered for ground and air vehicles.

VI. ACKNOWLEDGMENTS

This work is supported in part by the ARC Center of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

REFERENCES

[1] S. G. Chroust and M. Vincze, "Fusion of vision and inertial data for motion and structure estimation," *Journal of Robotic Systems*, vol. 21, no. 2, pp. 73–83, 2004.

[2] G. Alenya, E. Martinez, and C. Torras, "Fusing visual and inertial sensing to recover robot ego-motion," *Journal of Robotic Systems*, vol. 21, no. 1, pp. 23–32, 2004.

[3] J. Lobo and J. Dias, "Vision and inertial sensor cooperation using gravity as a vertical reference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1597–1608, 2003.

[4] D. Diel, P. DeBitetto, and S. Teller, "Epipolar constraints for vision-aided inertial navigation," in *IEEE Workshop on Motion and Video Computing*. Breckenridge, Colorado: IEEE, 2005.

[5] D. D. Diel, "Stochastic constraints for vision-aided inertial navigation," M.S., MIT, 2005.

[6] S. S. Rounds and G. Marmar, "Stellar inertial guidance capabilities for advanced icbm," in *AIAA Guidance and Control Conference*. AIAA, 1983, aIAA Paper: 83-2297.

[7] P. Corke, "An inertial and visual sensing system for a small autonomous helicopter," *Journal of Robotic Systems*, vol. 21, no. 2, pp. 43–51, 2004.

[8] S. I. Roumeliotis, A. E. Johnson, and J. F. Montgomery, "Augmenting inertial navigation with image-based motion estimation," in *International Conference on Robotics and Automation*. Washington D.C.: IEEE, 2002.

[9] A. Johnson, R. Willson, J. Goguen, J. Alexander, and D. Meller, "Field testing of the mars exploration rovers descent image motion estimation system," in *International Conference on Robotics and Automation*. Barcelona, Spain: IEEE, 2005.

[10] S. Graovac, "Principles of fusion of inertial navigation and dynamic vision," *Journal of Robotic Systems*, vol. 21, no. 1, pp. 13–22, 2004.

[11] J. Kim and S. Sukkarieh, "Autonomous airborne navigation in unknown terrain environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 1031–1045, 2004.

[12] M. Pachter, "Ins aiding using optical flow - theory," in *11th Mediterranean Conference on Control and Automation*. Rhodes, Greece: IEEE, 2003.

[13] M. Pachter and A. Porter, "Bearings-only measurements for ins aiding: The three dimensional case," in *American Control Conference*. Boston, Massachussetts: IEEE, 2004.

[14] M. Koifman and I. Y. Bar-Itzhack, "Inertial navigation system aided by aircraft dynamics," *IEEE Transactions on Control Systems Technology*, vol. 7, no. 4, pp. 487–493, 1999.

[15] G. Dissanayake, S. Sukkarieh, E. Nebot, and H. Durrant-Whyte, "The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 5, 2001.

[16] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. v. Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Winning the darpa grand challenge," *Journal of Field Robotics*, 2006, accepted for publication.

[17] J. Montiel, J. Civera, and A. Davison, "Unified inverse depth parametrization for monocular slam," in *Robotics Science and Sytems*, Philadelphia, 2006.

[18] A. Huster and S. M. Rock, "Relative position sensing by fusing monocular vision and inertial rate sensors," in *International Conference on Advanced Robotics*. Coimbra, Portugal: IEEE, 2003.

[19] D. Strelow, "Motion estimation from image and inertial measurements," Ph.D. dissertation, Carnegie Mellon University, 2004.

[20] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, ser. Intelligent Robotics and Autonomous Agents. Cambridge: The MIT Press, 2005.

[21] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*, 2nd ed., ser. IEE Radar, Sonar and Navigation Series. Stevenage: The Institution of Electrical Engineers, 2004, vol. 17.

[22] J. Neira and J. Tardos, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 890–897, 2001.

[23] J. Kim and S. Sukkarieh, "Flight test results of gps/ins navigation loop for an autonomous unmanned aerial vehicle (uav)," in *The 15th International Technical Meeting of the Satellite Division of the Institute of Navigation*. Portland, OR: ION, 2002, pp. 510–517.