Vision Integrated Inertial Navigation System

*Ankur Jain[a], A R Rekha[a] , J Krishnakumar[a] and Dr. P P Mohanlal[a]

[a]*ISRO Inertial Systems Unit, Thiruvananthapuram-695013, India*

Abstract

We combine a visual measurement system with an aided inertial navigation filter to produce a precise and robust navigation system that does not rely on external infrastructure. Incremental structure from motion using a stereo camera provides real-time highly accurate pose estimates of the sensor which are combined with six degree-of-freedom inertial measurements in an Extended Kalman Filter. The filter is structured to neatly handle the incremental and local nature of the visual odometry measurements and to handle uncertainties in the system in a principled manner.

Keywords: Vision, GPS, Kalman Filter, INS.

1. Introduction

A motion estimation device is one of the basic requirements for building an autonomous legged, wheeled or aerial robot. Besides autonomous navigation, other applications such as mapping and planetary landing also require accurate motion estimation [6], [7]. Such devices generally combine several sensors, which can be divided into two categories: *exteroceptive* and *proprioceptive*. Combining both types of sensors is an attractive solution because, roughly speaking, they have opposed strengths and weaknesses. The former estimate motion based on external observations such as images [8] or range data [9]. As a result, error accumulation or drift is essentially proportional to the length of the trajectory, although, it is also dependent on the geometry of the environment. The latter, by their nature, measure their own motion, and can operate in any kind of environment. The drawback is that error accumulation is a function of time rather than distance which is why they require some form of aiding.

With the recent advances in the manufacturing of micro electro mechanical based inertial sensors (MEMs) and CCD sensors, it is possible to build inexpensive and reliable inertial measurement units (IMU) and cameras. As a result, vision-aided inertial navigation systems are increasingly popular. State of the art systems currently augment inertial measurements with visual odometry (VO) [10], [11], [12].

Vision sensors (e.g., such as camera, hyper-spectral sensors, laser scanners etc.) are mainly used for mapping and environments detection, and usually geo-referenced by other sensors. Terrain Aided Navigation System (TANS) typically makes use of onboard sensors and a preloaded terrain database. Simultaneous Localization And Mapping (SLAM) algorithm can navigate vehicles or robots in an unknown environment. As the onboard vision sensors detect landmarks from the environments, the SLAM estimator augments the landmark locations to a map and estimates the vehicle position with successive observations. SLAM has been applied to field robot and air.

2. Frame work for vision aided INS

In this paper, we present a novel vision-aided navigation system which is based on an IMU and a stereo camera. Motion estimation is performed in real time and the integration of the sensors is robust.

By *robust*, we imply that VO is allowed to fail and be restarted at any moment. This sometimes happens in real life situations, for example, if the camera is directly pointed at the sun, or when illumination quickly changes. We rely on a delayed state Extended Kalman Filter (EKF) allowing a loose coupling of the two sensors. The delayed states simply correspond to the position and orientation of the vehicle pose at the last VO key frame. This effectively allows the VO output to the EKF to be a relative pose update. This has important practical advantages:

- The camera trajectory does not need to be registered within the global coordinate system;
- In case of failure, the update is set to have infinite uncertainty and the VO is simply restarted;
- Expensive uncertainty propagation over time [8], [11] is avoided since the uncertainty estimation is only required for the motion update.

Our visual-odometry is based on incremental structure from motion with key frame selection and sparse local bundle adjustment [13]. It can be used on any kind of robot or vehicle since it does not rely on a specific motion model, *i.e.* it estimates a six degree of freedom pose of the camera and does not use any kind of smoothing. In addition, it does not make any assumption about the geometry of the scene such as a flat ground plane [14].

Autonomous flight through a unknown territory is an extremely challenging problem. In general successful

* Corresponding author: e-mail: ankur_jain@vssc.gov.in

operation of a UAV (in any environment, cluttered or clear) involves three basic tasks:

1. The vehicle must maintain controlled flight while avoiding collisions with obstacles (the vehicle must aviate). This requires a means to determine the state of the vehicle and to detect and localize obstacles with enough accuracy that appropriate action can be taken.

2. It must find its way from the starting point to a goal location in a finite amount of time (the vehicle must navigate). This requires a means to localize the vehicle relative to the goal.

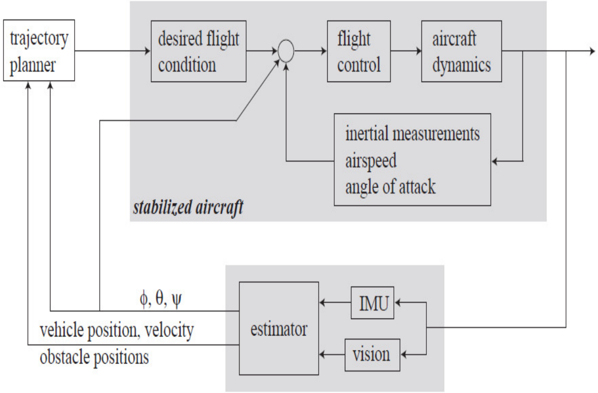
3. It must convey information about the environment to a human operator or other robots in the team (the vehicle must communicate). This requires a means of presenting data in a useful way to human operators or other robots in the team.

These tasks are complicated by the payload limitations imposed by small vehicles (both mass and dimensions of the total sensing payload are constrained) and by the environments where the vehicle operates. The unavailability of GPS in cluttered environments means that direct measurements of vehicle position are unavailable. Furthermore, the environment is unsurveyed; hence obstacle positions are initially unknown.

The task of aviation could be accomplished by flying reactively: the vehicle maintains heading until an obstacle is detected the vehicle maneuvers to avoid the obstacle and then attempts to reacquire the desired heading. However, while this reactive flight is adequate for small numbers of well-spaced obstacles, intuition suggests that the limited field of view of most sensors will cause this approach to fail in more complex environments with densely packed obstacles. Some means of accounting for obstacles which are outside of the field of view must be provided to plan safe maneuvers. While it is certainly aviating, purely reactive flight can hardly be said to be navigation: without knowledge of aircraft position there is no guarantee of reaching the goal. Thus in order to navigate in an obstacle-strewn environment some means of obtaining the position of the vehicle must be provided.

Estimation is the process of extracting information about variables of interest from measurements that are noisy and may be related to these variables through complex mathematical models. In this application, the variables of interest include: vehicle orientation and velocity (to maintain controlled flight); Obstacle relative position (to avoid collisions); and vehicle position (to enable navigation to a goal). Planning involves finding a safe, dynamically feasible path through cluttered environments to a goal location, and finally control involves both stabilizing the vehicle and following the path computed by the planning algorithm.

The problem of state estimation is directly tied to enabling a small UAV to aviate and navigate through the environment and to communicate its acquired knowledge. Nonlinearities in the system models (both vehicle kinematics and the vision model) coupled with potentially large uncertainties in system states makes this a particularly difficult estimation problem. This is further exacerbated by the lack of observability in the system: a monocular vision system provides only bearings to obstacles, making multiple measurements from different vantage points necessary to localize it.



A stabilized aircraft is an aircraft that can maintain a desired flight condition. This may require measurements such as angular rates, angle of attack, sideslip angle and airspeed.

2.1 Vision Based Navigation and Structure from Motion

Vision has been extensively studied for use as a sensor in estimation related applications. Examples include structure from motion, vision augmented inertial navigation, real time benthic navigation and relative position estimation.

Structure from motion attempts to reconstruct the trajectory of the video camera and an unknown scene. An example of an application is given in [1], which describes reconstruction of archaeological sites using video from a hand-carried camera. However, structure from motion algorithms are typically formulated as batch processes, analyzing and processing all images in the sequence simultaneously. While this will give the greatest accuracy of both the reconstructed scene and camera path, it does not lend itself to real-time operation.

Research into vision augmented inertial navigation [2, 3, 4] is primarily concerned with estimating the vehicle state by fusing inertial measurements either with bearings to known fiducially or data from optical flow algorithms.

The use of vision for aiding UAV navigation has become an active area of research. In many cases vision is not the primary navigation/control sensor but is used in conjunction with inertial navigation systems and GPS to increase situation awareness.

3. The state estimation

This section defines the estimation problem. It has three purposes: (a) define the state estimation problem; (b) develop equations for plant and sensor models; (c) provide some justification for applying a Kalman Filter to this estimation problem.

The choice of variables used to describe the state of the vehicle and its environment is an important factor in the design of a solution and its eventual complexity. The state variables must be sufficient to enable control of the vehicle, avoid obstacles and allow navigation to a goal. At the same time the choice of state variables has a strong effect on the complexity of the models used to describe the system. For example, a particular choice of state variables may lead to a very simple model for the vision system but complex models for vehicle and landmark dynamics. This trade off must be made in consideration of the limitations imposed by real-time operation of the resulting estimator.

In addition to vehicle position, orientation and speed, low cost IMUs are subject to scale factor and bias errors that can drift with time, thus estimates of scale factor and bias are also required. The vehicle state vector is

$$X_V = [x \; y \; z \; z \; i \; \} \; u \; v \; w \; a^T \; b_a{}^T \; b_w{}^T]$$

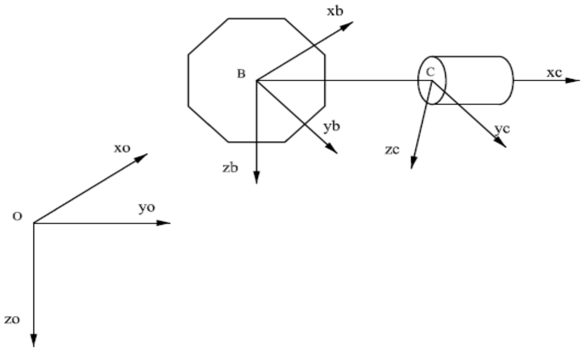
(x y z) represents position in the inertial frame,(ϕ θ ψ) represent Euler angles with respect to the inertial frame, (u v w) represents velocity expressed in the body frame, $\alpha^{\mathbf{T}}$ represents IMU scale factor error, $b^T{}_a$ represents accelerometer bias, and finally $b^T{}_\omega$ represents rate gyro bias.

Description	Purpose	Source
Angular Rate	Flight Control	Measured by IMU
Orientation	Flight Control & navigation	Estimated
Speed	Flight Control & navigation	Estimated
Vehicle Position	Navigation	Estimated
Obstacle Relative position	Collision Avoidance	Estimated

3.1 Sensor and System Models

3.1.1 Coordinate frames:

Navigation is done with respect to an inertial North-East-Down (NED) coordinate frame O. Sensors are fixed to the vehicle with known position and angular offsets with respect to a body-fixed frame B. Acceleration and angular rate are measured using a strap down inertial measurement unit in the body frame B, bearings to landmarks are obtained in a camera frame C.



Frame O is an inertial NED frame. B is the vehicle body-fixed frame, the matrix T defines the transformation of a vector in O to its representation in B. Frame C is the camera-fixed frame, with the camera's optical axis aligned with xc. The transformation T_{cam} between the camera frame and the body frame B is assumed known and the axes of the inertial measurement unit are assumed to be aligned perfectly with the body frame B.

Transformation matrices T and T_{cam} define the transformation of a vector expressed in O to B and a vector expressed in B to C, respectively. Coordinate frames are shown schematically in Figure above.

3.1.2 Vehicle Kinematic Model:

A dynamic model requires knowledge of all inputs, including disturbances. For small UAVs there is a very high degree of uncertainty associated with disturbances which act on the vehicle. In this case a standard technique is to use a kinematic model driven by inertial measurements as a process model.

Vehicle position x, y, z is expressed in the inertial frame, rotations are expressed as Euler angles ϕ, θ, ψ relative to the inertial frame and velocity u, v, w are in expressed in the body frame. The coordinate transform T projects a vector expressed in the inertial frame O into the body frame B. Vehicle kinematics are:

$$\begin{matrix} \text{SçW} & & u \\ \text{SçW} = & T^{-1} & v \\ \text{SçW} & & w \end{matrix}$$

The transformation matrix T is defined by the Euler angles of the aircraft with respect to the inertial frame. Following a roll-pitch-yaw convention,

$$T = T_z \; T_i \; T\}$$

Where

$$T_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos z & \sin z \\ 0 & -\sin z & \cos z \end{bmatrix}$$

$$T_j = \begin{bmatrix} \cos j & 0 & -\sin j \\ 0 & 1 & 0 \\ \sin j & 0 & \cos j \end{bmatrix}$$

$$T_\} = \begin{bmatrix} \cos \} & \sin \} & 0 \\ -\sin \} & \cos \} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Therefore

$$T = \begin{bmatrix} \cos j * \cos \} & \cos j * \sin \} & -\sin j \\ \sin z * \sin j * \cos \} - \cos z * \sin \} & \sin z * \sin j * \sin \} + \cos z * \cos \} & \sin z * \cos j \\ \cos z * \sin j * \cos \} + \sin z * \sin \} & \cos z * \sin j * \sin \} - \sin z * \cos \} & \cos z * \cos j \end{bmatrix}$$

Body angular rates can be expressed as Euler angle rates by:

$$\begin{bmatrix} \dot{z} \\ \dot{j} \\ \dot{\}} \end{bmatrix} = \begin{bmatrix} 1 & \sin z * \tan j & \cos z * \tan j \\ 0 & \cos z & -\sin z \\ 0 & \sin z / \cos j & \cos z / \cos j \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix}$$

By expanding these equations

$$\dot{x} = \cos j * \cos \} * u + \left(\sin z * \sin j * \cos \} - \cos z * \sin \}\right) * v + \left(\cos z * \sin j * \cos \} + \sin z * \sin \}\right) * w$$

$$\dot{y} = \cos j * \sin \} * u + \left(\sin z * \sin j * \sin \} + \cos z * \cos \}\right) * v + \left(\cos z * \sin j * \sin \} - \sin z * \cos \}\right) * w$$

$$\dot{z} = -\sin j * u + \sin z * \cos j * v + \cos z * \cos j * w$$

$$\dot{z} = p + \sin z * \tan j * q - \cos z * \tan j * r$$

$$\dot{j} = \cos z * q - \sin z * r$$

$$\dot{\}} = (\sin z / \cos j) * q + (\cos z / \cos j) * r$$

3.1.3 Inertial Measurement Model:

The inertial measurement unit includes accelerometers and rate gyros. The accelerometers measure specific force, which includes the acceleration of the vehicle and the projection of the acceleration due to gravity onto the body frame. The rate gyros measure the rotational velocity of the vehicle. Both sensors include sensor biases and zero mean Gaussian random noise.

3.1.4 Vision Model:

The camera is assumed to be fixed to the UAV with known offset from the CG and known angular offset from the body-fixed frame, defined by a transformation T_{cam}. The camera x-axis is perpendicular to the image plane.

A pinhole camera model describes the projection of a vector onto the image plane as

$$Z = \frac{f}{x} \begin{bmatrix} y \\ z \end{bmatrix}$$

Where f is the focal length and $[x\ y\ z]^T$ is the vector (expressed in the camera frame). The focal length f can be normalized without loss of generality.

For cameras with "standard" field of view (less than approximately 70^0) this model is sufficient. In wide field of view cameras (> 90^0) this model becomes problematic.

VO algorithm:

These are the main steps of VO algorithm:

Initialization:
1) Feature detection in left and right images
2) Sparse stereo matching
3) Feature triangulation

For each new image pair*:
1) Feature detection in left and right images.
2) Feature matching between previous and current left images.
3) Feature matching between previous and current right images using constraints from the left image.
4) Sparse stereo matching on the remaining features.
5) Pose estimation using the 3-point algorithm.
6) Local bundle adjustment on the last key frames.
7) Relative pose computation and uncertainty estimation.
8) Pose update to the EKF.
* If any step fails, send pose update of infinite uncertainty and start from the beginning.

3.2 Position Estimation from camera images

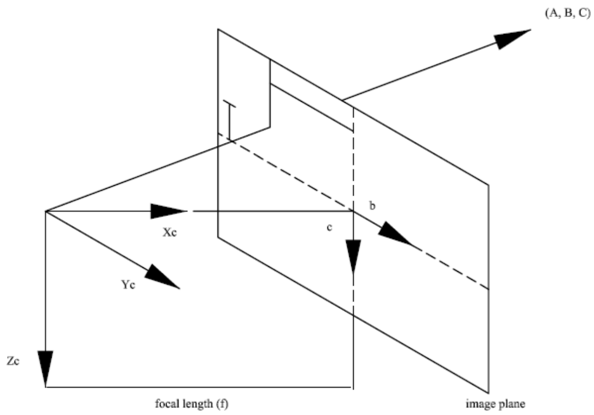
The EKF estimates vehicle position using three measurements of a target obtained from camera images. Since the position and orientation of the target are known, the vision-based estimator needs only to determine the pose of the target to resolve the UAV's position. This section first provides some fundamentals on relating 3D position to 2D images, and then describes the implementation of the EKF.

3.2.1 Relating 3D position to 2D Images:

A perspective projection model of a pinhole camera allows position in a 2D camera image to be inferred from 3D position as shown in Figure. The model projects an arbitrary point (*A, B, C*) to a pixel point (*b, c*) on the image plane (the camera image) according to the following relations:

$$b = f * B/A$$
$$c = f * C/A$$

where *f* is the focal length of the camera. The focal length depends solely on the image width in pixels of the camera image (*w*) and the angle of the field of view (*FOV*), both of which are characteristics of the camera, according to f = w/2* tan (FOV/2)



Camera perspective projection model used for relating 3D position to position in 2D images. The point (*A*, *B*, *C*) is projected onto the camera image plane to the point (*b*, *c*).

3.2.2 Reference frames

Three frames of reference are used in this paper. The inertial reference frame is a local inertial frame. The camera frame has its origin at the camera's principal point with the x_c axis along the camera's optical axis, whereas the body frame is the conventional aircraft body frame centered at the vehicle center of mass. Vector components in the different reference frames can be transformed using direction cosine matrix sequences as follows:

$$L_{cb} = \begin{bmatrix} \cos\theta_c & 0 & -\sin\theta_c \\ 0 & 1 & 0 \\ \sin\theta_c & 0 & \cos\theta_c \end{bmatrix} \begin{bmatrix} \cos\psi_c & \sin\psi_c & 0 \\ -\sin\psi_c & \cos\psi_c & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$L_{bi} = \begin{bmatrix} q_1^2+q_2^2-q_3^2-q_4^2 & 2(q_2*q_3+q_1*q_4) & 2(q_2*q_4-q_1*q_3) \\ 2(q_2*q_3-q_1*q_4) & q_1^2-q_2^2+q_3^2-q_4^2 & 2(q_3*q_4+q_1*q_2) \\ 2(q_2*q_4+q_1*q_3) & 2(q_3*q_4-q_1*q_2) & q_1^2-q_2^2-q_3^2+q_4^2 \end{bmatrix}$$

$$L_{ci} = L_{cb} * L_{bi}$$

L_{cb} is a rotation sequence that converts vectors from components in the body frame to components in the camera frame by using the pan (Ψ_c) and tilt (θ_c) angles of the camera. Note, however, that the transformation from the body to the camera frame accounts for only the orientation differences between the two frames. The fact that the camera frame is centered at the camera location, whereas the body frame is centered at the vehicle center of mass, is neglected. L_{bi} is a standard rotation matrix from the body to the local inertial frame expressed in quaternion.

3.3 Extended Kalman Filter

The vision-based EKF is a mixed continuous-discrete time filter. The process model uses a continuous-time linearized model of the platform or UAV dynamics to predict expected values for the state variables, whereas the measurement model runs at discrete time steps to correct the process model's estimates by using camera measurements. The EKF measurement vector consists of the horizontal and vertical Cartesian coordinates (in pixels) of the target center in the camera image, and the square root of the target area in pixels. By comparing predicted values of the measurement vector with actual measurements from the image processor, the EKF is able to maintain estimates of the platform's or UAV's position and attitude.

Fusion:

Algorithm Outline
We combine the inertial and VO data in an Extended Kalman Filter. A summary of the algorithm is given here.

- Initialization:

1) Determine initial alignment of IMU-camera frame to global navigation frame.
2) Calibrate IMU from static measurements.
3) Append estimates of initial position and orientation as delayed states.

- Inertial Prediction:

1) For each IMU measurement integrate state and uncertainty estimates.

- Vision Measurement:

1) For each VO measurement, form measurement prediction from current states.
2) Compute and apply state corrections using extended Kalman filter formulations with VO predicted measurement uncertainties if uncertainty is not infinite.
3) Reform state by appending updated estimates of current position and orientation as delayed states.
4) Return to Inertial Prediction step.

4. Challenges:

This paper proposed an approach for Vision Integrated inertial Navigation System using vision information. However, in outdoor environment this integration is a multi-disciplinary issues, there are several aspects of the future need to strengthen.

- Due to complexity of the outdoor environment, how to extract useful visual information is very complex when the UAV avoids obstacle or turns.
- Some special causes or too many obstacles will let UAV move out of its course and lose itself in the environment. In this case, how to position itself is a big challenge.

References:

1. M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, and J. Tops. *3D capture of archaeology and architecture with a hand-held camera. In ISPRS workshop on Vision Techniques for Digital Architectural and Archaeological Archives*, pages 262–267, Ancona, Italy.

2. Stergios I. Roumeliotis, Andrew E. Johnson, and James F. Montgomery. *Augmenting inertial navigation with image-based motion estimation. In IEEE International Conference on Robotics and Automation (ICRA),* Washington, DC, 2002. IEEE.

3. Jorge Lobo and Jorge Dias. *Integration of inertial information with vision. In 24th Conference of IEEE Industrial Electronics Society (IECON),* pages 1263–1267, Aachen, Germany, 1998. IEEE.

4. Thomas Netter and Nicolas Franceschini. *A robotic aircraft that follows terrain using a neuromorphic eye. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Lausanne, Switzerland, 2002.

5. *State estimation for autonomous flight in cluttered environment,* phD thesis, Stanford University, March 2006.

6. A. I Mourikis, N. Trawny, S. I Roumeliotis, A. Johnson, and L. Matthies. *Vision aided inertial navigation for precise planetary landing: Analysis and experiments. In Proc. Robotics Systems and Science Conference, 2007.*

7. S. I Roumeliotis, A. E Johnson, and J. F Montgomery. *Augmenting inertial navigation with image-based motion estimation. In Proceedings-IEEE International Conference on Robotics and Automation, volume 4, page 43264333, 2002.*

8. A. J. Davison, I. D. Reid, N. Molton, and O. Stasse. MonoSLAM: *Real-Time single camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6):1052–1067, 2007.*

9. F. Lu and E. Milios. *Globally consistent range scan alignment for environment mapping. Autonomous Robots, 4(4):333349, 1997.*

10. K. Konolige, M. Agrawal, and J. Sola. *Large scale visual odometry for rough terrain. In International Symposium on Robotics Research, 2007.*

11. M. Maimone, Y. Cheng, and L. Matthies. *Two years of visual odometry on the mars exploration rovers. Journal of Field Robotics, 24(3):169186, 2007.*

12. A. I Mourikis and S. I Roumeliotis. *A multi-state constraint kalman filter for vision-aided inertial navigation. In IEEE International Conference on Robotics and Automation, page 35653572, 2007.*

13. B. Triggs, P. F McLauchlan, R. I Hartley, and A.WFitzgibbon. *Bundle adjustment-a modern synthesis. Lecture Notes in Computer Science, page 298372, 1999.*

14. R. Garca-Garca, M. A Sotelo, I. Parra, D. Fernndez, J. E Naranjo, and M. Gaviln. *3D visual odometry for road vehicles. Journal of Intelligent and Robotic Systems, 51(1):113134, 2008.*