The Battle for Filter Supremacy: A Retrospective: The Movie: The Game: The Paper

Lee Clement[1] and Valentin Peretroukhin[1]

Abstract— **Monkey Style Chinese Kung Fu**

I. INTRODUCTION

The combination of visual and inertial sensors is a powerful tool for autonomous navigation in unknown environments. Indeed, cameras and inertial measurement units (IMUs) are complementary in several respects. Since an IMU directly measures accelerations and rotational velocities, these values must be integrated to arrive at a new pose estimate. However, the noise inherent in the IMU's measurements is included in the integration as well, and consequently the pose estimates can drift unbounded over time. The addition of a camera is an excellent way to bound this cumulative drift error since the camera's signal-to-noise ratio is highest when the camera is moving slowly. On the other hand, cameras are not robust to motion blur induced by large accelerations. In these cases, the strength of the IMU's signal far exceeds its baseline noise and can be relied upon more heavily in estimating pose changes.

The question, then, is how best to fuse measurements from these two sensor types to arrive at an accurate estimate of a vehicle's motion over time. A complicating factor in the general form of this problem is the absence of a known map of landmarks from which the camera can generate measurements. Any solution must therefore solve a Simultaneous Localization and Mapping (SLAM) problem, although the importance placed on the mapping component may vary from algorithm to algorithm. What follows is a discussion of two common solutions, the Extended Kalman Filter (EKF) and the Sliding Window Filter (SWF), as well as a third hybrid solution, the Multi-State Constraint Kalman Filter (MSCKF), that combines the strengths of both.

II. EXTENDED KALMAN FILTER

In the Extended Kalman Filter (EKF) solution, vehicle poses and landmark positions are simultaneously estimated at each time step by augmenting the filter state with landmark positions. This technique, sometimes referred to as EKF-SLAM, attempts to track pose changes and create a globally consistent map of landmarks by recursively updating the state as new measurements become available.

[EKF EQUATIONS HERE?]

Although the recursive nature of EKF-SLAM allows it to operate online, the computational cost of the filter grows cubically with map size. This behavior is due to the fact that the dimension of the state grows linearly with the number of landmarks, and the computational cost of inverting the state covariance matrix while computing the Kalman gain is cubic in the dimension of the state. Consequently, the spatial extent over which EKF-SLAM can be used online is limited by the necessarily finite compute envelope available to it.

Another limitation of EKF-SLAM is that it is forgetful. Because the filter state includes only the most recent vehicle pose, a given update step can never modify past poses even if later landmark measurements ought to constrain them. By locking in past poses, the EKF-SLAM formulation condemns itself to sub-optimally estimating both vehicle motion and landmark positions.

III. SLIDING WINDOW FILTER

In contrast to EKF-SLAM, the aim of the Sliding Window Filter (SWF) is not to construct a globally consistent map, but rather to estimate a vehicle's motion by optimizing a sliding window of vehicle poses and landmark positions. The optimization problem in the SWF is typically solved as a non-linear least squares problem using Gauss-Newton optimization or some other algorithm.

[BATCH MATH GOES HERE?]

An important advantage of the SWF is that its computational cost depends on the number of landmarks in the current window rather than the number of landmarks in the entire map. By varying the spatial or temporal extent of the sliding window, the computational cost of the algorithm can be tailored to fit a given compute envelope, which makes the algorithm suitable for online operation over paths of arbitrary length.

However, the hard cut-off of the SWF may result in only some measurements of a particular feature contributing to the optimization. As a result, the filter may not maximally constrain some vehicle poses, and hence localization may be less accurate than we could expect from the full batch solution.

IV. MULTI-STATE CONSTRAINT KALMAN FILTER

The Multi-State Constraint Kalman Filter (MSCKF) [1] can be thought of as a hybrid of EKF-SLAM and the SWF. The key idea of the MSCKF is to maintain a sliding window of vehicle poses and to simultaneously update each pose in the window using batch-optimized estimates of landmarks that are visible across the entire window. This update step typically occurs when a tracked landmark goes out of view of the camera, but it may also be triggered if the number of vehicle states in the window exceeds some preset threshold.

[1]Institute for Aerospace Studies, University of Toronto, Toronto, ON, Canada {lee.clement, valentin.peretroukhin} @robotics.utias.utoronto.ca

A. MSCKF State Parametrization

We evaluated the MSCKF using a dataset in which the IMU ‘measures’ gravity-corrected linear velocities rather than raw linear accelerations (see Section V). In order to accommodate this alternative state parametrization, the mathematical framework described in this section differs slightly from that described in [1].

In our implementation, we parametrize the IMU state at time k as the 13-dimensional vector

$$\mathbf{x}_{I,k} = \begin{bmatrix}\mathbf{q}_{IG,k}^T & \mathbf{b}_{\boldsymbol{\omega},k}^T & \mathbf{b}_{\mathbf{v},k}^T & \mathbf{p}_{G,k}^{IG\,T}\end{bmatrix}^T \tag{1}$$

where $\mathbf{q}_{IG,k}$ is the unit quaternion representing the rotation from the global frame $\mathcal{F}_G$ to the IMU frame $\mathcal{F}_I$, $\mathbf{b}_{\boldsymbol{\omega},k}$ is the bias on the gyro measurements $\boldsymbol{\omega}_m$, $\mathbf{b}_{\mathbf{v},k}$ is the bias on the velocity measurements $\mathbf{v}_m$, and $\mathbf{p}_{G,k}^{IG}$ is the vector from the origin of $\mathcal{F}_G$ to the origin of $\mathcal{F}_I$ expressed in $\mathcal{F}_G$ (i.e., the position of the IMU in the global frame).

At time k, the full state of the MSCKF consists of the current IMU state estimate, and estimates of N 7-dimensional past camera poses in which active feature tracks were visible:

$$\hat{\mathbf{x}}_k = \begin{bmatrix}\hat{\mathbf{x}}_{I,k}^T & \hat{\mathbf{q}}_{C_1G}^T & \hat{\mathbf{p}}_G^{C_1G\,T} & \dots & \hat{\mathbf{q}}_{C_NG}^T & \hat{\mathbf{p}}_G^{C_NG\,T}\end{bmatrix}^T \tag{2}$$

We can also define the MSCKF “error state” at time k:

$$\widetilde{\mathbf{x}}_k = \begin{bmatrix}\widetilde{\mathbf{x}}_{I,k}^T & \boldsymbol{\delta\theta}_{C_1}^T & \widetilde{\mathbf{p}}_G^{C_1G\,T} & \dots & \boldsymbol{\delta\theta}_{C_N}^T & \widetilde{\mathbf{p}}_G^{C_NG\,T}\end{bmatrix}^T \tag{3}$$

where

$$\widetilde{\mathbf{x}}_{I,k} = \begin{bmatrix}\boldsymbol{\delta\theta}_I^T & \widetilde{\mathbf{b}}_{\boldsymbol{\omega},k}^T & \widetilde{\mathbf{b}}_{\mathbf{v},k}^T & \widetilde{\mathbf{p}}_{G,k}^{IG\,T}\end{bmatrix}^T \tag{4}$$

is the 12-dimensional IMU error state. In the above, $\widetilde{x}$ denotes the difference between the true value and the estimated value of the quantity x. The rotational errors $\boldsymbol{\delta\theta}$ are defined according to

$$\boldsymbol{\delta}\mathbf{q} = \hat{\mathbf{q}}^{-1} \otimes \mathbf{q} \simeq \begin{bmatrix}\frac{1}{2}\boldsymbol{\delta\theta}^T & 1\end{bmatrix}^T. \tag{5}$$

Accordingly, the MSCKF state covariance $\hat{\mathbf{P}}_k$ is a $(12+6N) \times (12+6N)$ matrix that may be partitioned as

$$\hat{\mathbf{P}}_k = \begin{bmatrix}\hat{\mathbf{P}}_{II,k} & \hat{\mathbf{P}}_{IC,k} \\ \hat{\mathbf{P}}_{IC,k}^T & \hat{\mathbf{P}}_{CC,k}\end{bmatrix} \tag{6}$$

where $\hat{\mathbf{P}}_{II,k}$ is the 12×12 covariance matrix of the current IMU state, $\hat{\mathbf{P}}_{CC,k}$ is the $6N \times 6N$ covariance matrix of the camera poses, and $\hat{\mathbf{P}}_{IC,k}$ is the $12 \times 6N$ cross-correlation between the current IMU state and the past camera poses.

B. MSCKF State Augmentation

When a new camera image becomes available, the MSCKF state must be augmented with the current camera pose. We obtain the camera pose by applying the known transformation $\left(\mathbf{q}_{CI}, \mathbf{p}_I^{CI}\right)$ to a copy of the current IMU pose:

$$\hat{\mathbf{q}}_{C_{N+1}G} = \mathbf{q}_{CI} \otimes \hat{\mathbf{q}}_{IG,k} \tag{7}$$

$$\hat{\mathbf{p}}_G^{C_{N+1}G} = \hat{\mathbf{p}}_G^{IG} + \hat{\mathbf{C}}_{IG,k}^T \hat{\mathbf{p}}_I^{CI} \tag{8}$$

where $\hat{\mathbf{C}}_{IG,k}$ is the rotation matrix corresponding to $\hat{\mathbf{q}}_{IG,k}$ and $\otimes$ denotes quaternion multiplication.

Assuming the MSCKF state has already been augmented by N camera poses, we add the $(N+1)^{\text{th}}$ camera pose to the state as follows:

$$\hat{\mathbf{x}}_k \leftarrow \begin{bmatrix}\hat{\mathbf{x}}_k^T & \hat{\mathbf{q}}_{C_{N+1}G}^T & \hat{\mathbf{p}}_G^{C_{N+1}G\,T}\end{bmatrix}^T. \tag{9}$$

We must also augment the MSCKF state covariance:

$$\hat{\mathbf{P}}_k \leftarrow \begin{bmatrix}\mathbf{1}_{12+6N} \\ \mathbf{J}_k\end{bmatrix} \hat{\mathbf{P}}_k \begin{bmatrix}\mathbf{1}_{12+6N} \\ \mathbf{J}_k\end{bmatrix}^T \tag{10}$$

where the Jacobian $\mathbf{J}_k$ is given by

$$\mathbf{J}_k = \begin{bmatrix}\hat{\mathbf{C}}_{CI,k} & \mathbf{0}_{3\times6} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times6N} \\ \left(\hat{\mathbf{C}}_{IG,k}^T \mathbf{p}_{I,k}^{CI}\right)^\times & \mathbf{0}_{3\times6} & \mathbf{1}_3 & \mathbf{0}_{3\times6N}\end{bmatrix}. \tag{11}$$

C. IMU State Estimate Propagation

The evolution of the mean estimated IMU state $\hat{\mathbf{x}}_I$ over time is described by the following differential equations:

$$\dot{\hat{\mathbf{q}}}_{IG} = \frac{1}{2}\boldsymbol{\Omega}\left(\hat{\boldsymbol{\omega}}\right)\hat{\mathbf{q}}_{IG} \tag{12}$$

$$\dot{\hat{\mathbf{b}}}_{\boldsymbol{\omega}} = \mathbf{0}_{3\times1} \tag{13}$$

$$\dot{\hat{\mathbf{b}}}_{\mathbf{v}} = \mathbf{0}_{3\times1} \tag{14}$$

$$\dot{\hat{\mathbf{p}}}_G^{IG} = \hat{\mathbf{C}}_{IG}^T\left(\mathbf{v}_m - \hat{\mathbf{b}}_{\mathbf{v}}\right) \tag{15}$$

where $\hat{\mathbf{C}}_{IG}$ is the rotation matrix corresponding to $\hat{\mathbf{q}}_{IG}$ and

$$\boldsymbol{\Omega}\left(\hat{\boldsymbol{\omega}}\right) = \begin{bmatrix}-\hat{\boldsymbol{\omega}}^\times & \hat{\boldsymbol{\omega}} \\ -\hat{\boldsymbol{\omega}}^T & 0\end{bmatrix}$$

with

$$\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_m - \hat{\mathbf{b}}_{\boldsymbol{\omega}}$$

and

$$\hat{\boldsymbol{\omega}}^\times = \begin{bmatrix}0 & -\hat{\omega}_3 & \hat{\omega}_2 \\ \hat{\omega}_3 & 0 & -\hat{\omega}_1 \\ -\hat{\omega}_2 & \hat{\omega}_1 & 0\end{bmatrix}.$$

In our implementation we use a simple forward-Euler integration rather than the fifth-order Runge-Kutta procedure used in [1].

D. MSCKF State Covariance Propagation

With reference to the partitions defined in Equation 6, we compute the predicted camera-camera and IMU-camera state covariances as follows:

$$\hat{\mathbf{P}}_{CC,k+1}^- = \hat{\mathbf{P}}_{CC,k} \tag{16}$$

$$\hat{\mathbf{P}}_{IC,k+1}^- = \boldsymbol{\Phi}\left(t_k + T, t_k\right)\hat{\mathbf{P}}_{IC,k} \tag{17}$$

where T is the IMU sampling period, and the state transition matrix $\boldsymbol{\Phi}\left(t_k + T, t_k\right)$ is obtained by integrating

$$\dot{\boldsymbol{\Phi}}\left(t_k + \tau, t_k\right) = \mathbf{F}\boldsymbol{\Phi}\left(t_k + \tau, t_k\right), \tau \in [0, T] \tag{18}$$

with the initial condition $\boldsymbol{\Phi}\left(t_k, t_k\right) = \mathbf{1}_{12}$.

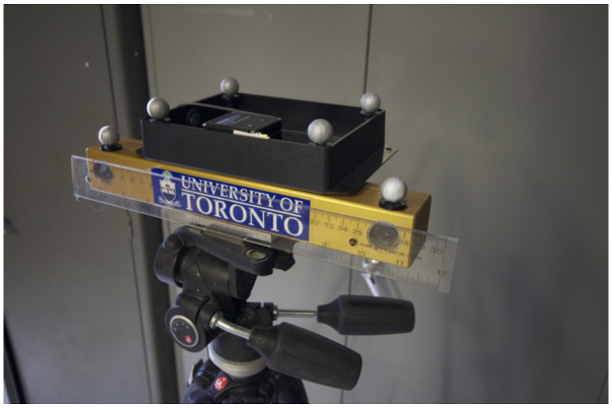

Fig. 1. The sensor head used in our experiments. The IMU measures translational and rotational velocities, while the stereo camera measures the positions of point landmarks. In our experiments, we artificially blinded the stereo camera by using measurements from the left camera only.

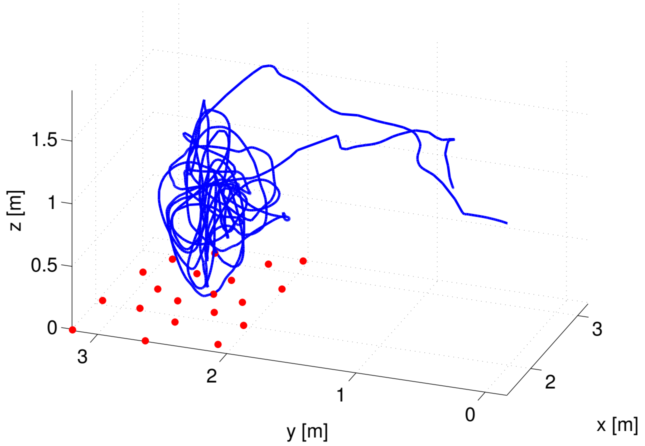

Fig. 2. Vicon ground truth for sensor head motion (blue) and landmark positions (red) in the "Starry Night" dataset.

We obtain the predicted IMU-IMU state covariance $\hat{\mathbf{P}}^{-}_{II,k+1}$ by integrating

$$\begin{aligned}\dot{\hat{\mathbf{P}}}_{II}\left(t_k, t_k+\tau\right) = {} & \mathbf{F}\hat{\mathbf{P}}_{II}\left(t_k, t_k+\tau\right) \\ & + \hat{\mathbf{P}}_{II}\left(t_k, t_k+\tau\right)\mathbf{F}^T \\ & + \mathbf{G}\mathbf{Q}_I\mathbf{G}^T, \tau \in [0, T] \end{aligned} \tag{19}$$

with the initial condition $\hat{\mathbf{P}}_{II}\left(t_k, t_k\right) = \hat{\mathbf{P}}_{II,k}$. $\mathbf{Q}_I$ is the covariance of the IMU measurement noise.

The Jacobians $\mathbf{F}$, $\mathbf{G}$ in Equations 18 and 19 are given by

$$\mathbf{F} = \begin{bmatrix} -\hat{\boldsymbol{\omega}}^{\times} & -\mathbf{1}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\hat{\mathbf{C}}_{IG}^T \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \tag{20}$$

$$\mathbf{G} = \begin{bmatrix} -\mathbf{1}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{1}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{1}_3 \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix}. \tag{21}$$

E. MSCKF State Correction Equations

[NULL SPACE TRICK, KALMAN GAIN, ETC.]

V. EXPERIMENTS

We conducted a comparative study of the MSCKF and SWF algorithms using the "Starry Night" dataset from the University of Toronto Institute for Aerospace Studies (UTIAS). The dataset consists of a rigidly attached stereo camera and IMU (Figure 1) observing a set of 20 landmarks while moving along an arbitrary 3D path. The dataset is well-suited to evaluating SLAM algorithms since accurate ground truth from a Vicon motion capture system is available for both the sensor head motion and the landmark positions (Figure 2). Since our algorithms are designed to make use of a monocular camera, we artificially blinded the stereo camera by using measurements from the left camera only.

Things quickly got out of hand when Matlab achieved sentience. We were forced to abandon our lab and have been surviving in the harsh environment of the MarsDome for weeks. If this ever gets published, please send help!

VI. CONCLUSIONS

REFERENCES

[1] 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*. IEEE, 2007, pp. 3565–3572.