# MSCKF(Multi-State Constraint Kalman Filter) 论文总结 (一)

论文: Mourikis A I, Roumeliotis S I. A Multi-State Constraint Kalman Filter for Vision-Aided Inertial Navigation[J]. 2007.

下面主要对该论文的重点内容进行记录、总结,可能会涉及部分自己的理解。

# 摘要

作者提出了一种基于扩展卡尔曼(EKF)的滤波算法来实现实时的视觉辅助惯性导航。文章的主要贡献工作是推导了多相机位姿观测同一个静态特征时形成的地理约束的测量模型。该测量模型不需要在EKF的状态向量里包含3D特征点位置,而且在线性化误差范围里是最优的。该视觉惯性导航算法的时间复杂度只与特征点的数量成线性关系,而且能够在大尺度实际场景中提供高精度位姿估计。作者在实际城市区域进行了大量的实验对算法进行验证。

# 1 状态估计算法

文章提出的基于EKF的估计器目标是跟踪IMU固连坐标系(I系)相对全局参考坐标系(I6系)的3D位 姿。为了简化地球自转角速度对IMU测量的影响,文章中的全局坐标系选为地心固定坐标系(ECEF系)。算法的整体介绍见I8AI8GOrithm 1。

Algorithm 1 Multi-State Constraint Filter

**Propagation**: For each IMU measurement received, propagate the filter state and covariance (cf. Section 3.2).

Image registration: Every time a new image is recorded,

- augment the state and covariance matrix with a copy of the current camera pose estimate (cf. Section 3.3).
- image processing module begins operation.

**Update**: When the feature measurements of a given image become available, perform an EKF update (cf. Sections 3.4 and 3.5).

如上图所见,IMU的测量每次都只要可用便即可处理,在EKF里基于此进行状态量和协方差的预测更新(见1.2)。另一方面,每次当一张图像被记录时,在状态向量里增加当前的相机位姿估计量(见1.3)。对状态量进行增广对于处理视觉特征测量非常有必要,因为当EKF更新时,每一个跟踪到的视觉特征测量都会被用来在观测到这些视觉特征的所有相机位姿间建立约束。因此,在任何时刻,EKF的状态向量包含 1)IMU状态, $X_{IMU}$ ; 2)最大 $N_{max}$ 个过去的相机位姿。

# 1.1 EKF状态向量的结构

IMU的状态向量展开如下所示:

$$\mathbf{X}_{\mathrm{IMU}} = \begin{bmatrix} {}^{I}_{G}\bar{q}^{T} & \mathbf{b}_{g}{}^{T} & {}^{G}\mathbf{v}_{I}{}^{T} & \mathbf{b}_{a}{}^{T} & {}^{G}\mathbf{p}_{I}^{T} \end{bmatrix}^{T}$$

上式中,q是单位四元数,描述了G系到I系得旋转, $p_I^G$ 和 $v_I^G$ 表示IMU相对于G系的位置和速度。 $b_g$ 和 $b_a$ 是3\*1的向量,分别表示影响陀螺仪和加速度计测量的零偏。IMU的零偏建模为随机游走过程,通过高斯白噪声向量 $n_{va}$ 和 $n_{va}$ 进行推导。同时,可得IMU的误差状态可定义为:

$$\widetilde{\mathbf{X}}_{\mathrm{IMU}} = \begin{bmatrix} \delta \boldsymbol{\theta}_I^T & \widetilde{\mathbf{b}}_g^T & {}^G \widetilde{\mathbf{v}}_I^T & \widetilde{\mathbf{b}}_a^T & {}^G \widetilde{\mathbf{p}}_I^T \end{bmatrix}^T$$

对于位置、速度、和零偏,使用标准的相加误差定义(误差 $\check{x}=\mathfrak{g}\, \pm x-\mathrm{d}\, \pm \, \hat{a}\, \hat{x}$ )。但是,对于四元数误差定义是不同的。特殊的是, $q=\delta q \bigotimes \hat{q}$ , q是四元数, $\delta q$ 是四元数的估计值误差,  $\hat{q}$ 是四元数的估计值。该表达中, $\bigotimes$ 表示四元数乘法。误差四元数可定义为"

$$\delta \bar{q} \simeq \begin{bmatrix} \frac{1}{2} \delta \theta^T & 1 \end{bmatrix}^T$$

直观可见,四元数误差表示了真实姿态到估计的姿态间的微小旋转。由于姿态对应三个自由度,使用 $\delta\theta$ 可以对姿态误差进行最小维度描述。

假设在k时刻, 有N个相机位姿包含在EKF的状态向量里,则该向量具有下面的格式:

$$\hat{\mathbf{X}}_k = \begin{bmatrix} \hat{\mathbf{X}}_{\mathrm{IMU}_k}^T & {}^{C_1}\hat{\mathbf{q}}^T & {}^{G}\hat{\mathbf{p}}_{C_1}^T & \dots & {}^{C_N}\hat{\mathbf{q}}^T & {}^{G}\hat{\mathbf{p}}_{C_N}^T \end{bmatrix}^T$$

上式中, $\hat{q}_{\pi}\hat{p}$ 分别表示i=1...N个序列的相机姿态和位置估计值。则EKF的误差状态向量同时可以表示为:

$$\widetilde{\mathbf{X}}_k = \begin{bmatrix} \widetilde{\mathbf{X}}_{\mathrm{IMU}_k}^T & \delta\boldsymbol{\theta}_{C_1}^T & {}^{G}\widetilde{\mathbf{p}}_{C_1}^T & \dots & \delta\boldsymbol{\theta}_{C_N}^T & {}^{G}\widetilde{\mathbf{p}}_{C_N}^T \end{bmatrix}^T$$

# 1.2 状态预测/递推

通过下述IMU系统的连续时间模型推导滤波器的状态预测/递推模型:

#### 1.2.1 连续时间系统模型

IMU状态的连续时间变化方程如下:

$${}_{G}^{I}\dot{\bar{q}}(t) = \frac{1}{2}\Omega(\omega(t))_{G}^{I}\bar{q}(t), \quad \dot{\mathbf{b}}_{g}(t) = \mathbf{n}_{wg}(t)$$
$${}_{G}\dot{\mathbf{v}}_{I}(t) = {}_{G}\mathbf{a}(t), \quad \dot{\mathbf{b}}_{a}(t) = \mathbf{n}_{wa}(t), \quad {}_{G}\dot{\mathbf{p}}_{I}(t) = {}_{G}\mathbf{v}_{I}(t)$$

上式中, $a^G$ 表示全局坐标系下的载体运动加速度, $\omega=[\omega_x,\omega_y,\omega_z]^T$ 是IMU坐标系下的旋转角速度,且:

$$\Omega(\omega) = egin{bmatrix} -\lfloor \omega imes 
floor & \omega \ -\omega^T & 0 \end{bmatrix}, \quad \lfloor \omega imes 
floor = egin{bmatrix} 0 & -\omega_z & \omega_y \ \omega_z & 0 & -\omega_x \ -\omega_y & \omega_x & 0 \end{bmatrix}$$

 $\omega_m$ 和 $a_m$ 分别表示陀螺仪和加速度计的测量值,其可以表示为:

$$\omega_m = \omega + \mathbf{C}({}_{G}^{I}\bar{q})\omega_G + \mathbf{b}_g + \mathbf{n}_g$$
  
$$\mathbf{a}_m = \mathbf{C}({}_{G}^{I}\bar{q})({}^{G}\mathbf{a} - {}^{G}\mathbf{g} + 2\lfloor\omega_G \times\rfloor^G\mathbf{v}_I + \lfloor\omega_G \times\rfloor^2 {}^{G}\mathbf{p}_I) + \mathbf{b}_a + \mathbf{n}_a$$

上述公式中,C(.)表示一个旋转矩阵,描述了两个坐标系间的旋转关系,  $n_a$ 和 $n_g$ 分别表示加速度计和陀螺仪的测量噪声,此处建模为零均值的高斯白噪声模型。需要注意的是上述IMU的测量模型中包含了地球自转( $\omega_G$ )的影响。而且,局部坐标系下的加速度计测量值中也包含了重力加速度 $g^G$ 。

将期望操作符(-)用在IMU的状态传播模型中,可得连续IMU状态估计值的递推等式:

为了简化计算,C、 $\hat{a}$ 和 $\hat{\omega}$ 的表达可以用下式:

$$\mathbf{C}_{\hat{q}} = \mathbf{C}({}_{G}^{I}\hat{q}), \ \hat{\mathbf{a}} = \mathbf{a}_{m} - \hat{\mathbf{b}}_{a} \ \text{and} \ \hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_{m} - \hat{\mathbf{b}}_{g} - \mathbf{C}_{\hat{q}}\boldsymbol{\omega}_{G}.$$

所以, IMU误差状态的线性连续时变模型为:

$$\dot{\tilde{X}}_{IMU} = F\tilde{X}_{IMU} + Gn_{IMU}$$

上述方程中 $n_{IMU}=[n_g^T,n_{\omega g}^T,n_a^T,n_{\omega a}^T]$ 分别是系统噪声。 $n_{IMU}$ 和 $Q_{IMU}$ 依赖于传感器标定阶段离线计算得到的IMU噪声特性。最终,F和G可表示为:

$$\mathbf{F} = \begin{bmatrix} -\lfloor \hat{\omega} \times \rfloor & -\mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ -\mathbf{C}_{\hat{q}}^T \lfloor \hat{\mathbf{a}} \times \rfloor & \mathbf{0}_{3 \times 3} & -2 \lfloor \omega_G \times \rfloor & -\mathbf{C}_{\hat{q}}^T & -\lfloor \omega_G \times \rfloor^2 \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix}$$

 $I_3$ 为3x3的单位矩阵,且G为:

$$\mathbf{G} = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\mathbf{C}_{\hat{q}}^T & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix}$$

#### 1.2.2 离散时间模型

IMU以时间周期T采样\* $\omega_m$ 和 $a_m$ ,然后这些测量值用来在EKF里进行状态预测更新。每次接收到1个新得IMU测量值时,IMU得状态估计采用五阶龙格-库塔法进行数值积分。而且,EKF协方差矩阵也会同步更新。基于此,采用下面的分块矩阵描述和计算协方差:

$$\mathbf{P}_{k|k} = \begin{bmatrix} \mathbf{P}_{II_{k|k}} & \mathbf{P}_{IC_{k|k}} \\ \mathbf{P}_{IC_{k|k}}^T & \mathbf{P}_{CC_{k|k}} \end{bmatrix}$$

上式中 $P_{II_{k/k}}$ 是IMU状态的15×15维协方差矩阵, $P_{CC_{k/k}}$ 是相机位姿估计状态量 $6N\times6N$ 的协方差矩阵, $P_{IC_{k/k}}$ 是IMU状态和相机位姿估计量误差间的协方差矩阵。基于以上定义,整个协方差预测更新方程为:

$$\mathbf{P}_{k+1|k} = \begin{bmatrix} \mathbf{P}_{II_{k+1|k}} & \boldsymbol{\Phi}(t_k+T,t_k) \mathbf{P}_{IC_{k|k}} \\ \mathbf{P}_{IC_{k|k}}^T \boldsymbol{\Phi}(t_k+T,t_k)^T & \mathbf{P}_{CC_{k|k}} \end{bmatrix}$$

上式中, $P_{II_{k+1/k}}$ 可以通过**李雅普诺夫方程的数值积分?** 进行计算:

$$\dot{\mathbf{P}}_{II} = \mathbf{F}\mathbf{P}_{II} + \mathbf{P}_{II}\mathbf{F}^T + \mathbf{G}\mathbf{Q}_{\mathrm{IMU}}\mathbf{G}^T$$

基于 $P_{II_{k/k}}$ 的初始值,在时间区间 $(t_k, t_k + T)$ 里进行数值积分。状态转移矩阵 $\Phi(t_k, t_k + T)$ 可以通过差分方程的数值积分同样计算:

$$\dot{\Phi}(t_k + \tau, t_k) = \mathbf{F}\Phi(t_k + \tau, t_k), \quad \tau \in [0, T]$$

### 1.3 状态增广

当记录一帧新的图片时,基于IMU的位姿估计可以计算相机的位姿估计量:

$${}^{C}_{G}\hat{\bar{q}} = {}^{C}_{I}\bar{q} \otimes {}^{I}_{G}\hat{\bar{q}}, \quad \text{and} \quad {}^{G}\hat{\mathbf{p}}_{C} = {}^{G}\hat{\mathbf{p}}_{I} + \mathbf{C}^{T}_{\hat{q}}{}^{I}\mathbf{p}_{C}$$

上式中 $q_I^C$ 时I系和相机系间的旋转四元数, $P_C^I$ 是相机相对I系的位置关系,这两个量都是已知的,可以通过标定得到。当前的相机位姿估计量添加到状态量里时,对应的EKF协方差矩阵也需要增广:

$$\mathbf{P}_{k|k} \leftarrow \begin{bmatrix} \mathbf{I}_{6N+15} \\ \mathbf{J} \end{bmatrix} \mathbf{P}_{k|k} \begin{bmatrix} \mathbf{I}_{6N+15} \\ \mathbf{J} \end{bmatrix}^T$$

雅各比J可由IMU和相机间的位姿转换关系等式得到:

$$\mathbf{J} = \begin{bmatrix} \mathbf{C} \begin{pmatrix} C & \bar{q} \\ I & \bar{q} \end{pmatrix} & \mathbf{0}_{3 \times 9} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 6N} \\ \lfloor \mathbf{C}_{\hat{g}}^{TI} \mathbf{p}_{C} \times \rfloor & \mathbf{0}_{3 \times 9} & \mathbf{I}_{3} & \mathbf{0}_{3 \times 6N} \end{bmatrix}$$

# 1.4 测量模型

接着是被用来进行状态估计的系统测量模型,这是本篇论文的主要贡献工作。由于使用EKF进行状态估计,为了构造一个测量模型,首先需要定义个残差 $\mathbf{r}$ ,该残差与状态误差 $\tilde{X}$ 成正比,一般形式为:

$$r = H\widetilde{X} + noise$$

在该表达式中,**H**式测量的雅各比矩阵,噪声项需要时零均值,与状态误差无关的白噪声,这是EKF的算法框架要求的。

作者受从多个相机位姿视角观测同一个静态特征时对这些位姿形成的约束关系启发,从而推导了该测量模型。在本篇论文中,作者根据每个跟踪到的特征将多个相机测量打包为一组观测,而不是使用每个相机作为观测(后者这种方法中需计算两两位姿之间的约束关系)。所有这些相同3D特征点的相机测量值被用来定义一个约束等式,将所有相机位姿联系起来。这就实现了在滤波器状态向量里不包含特征点位置。

作者提出的测量模型描述的情况为: 一组 $M_j$ 个相机位姿( $\bar{q}_G^{C_i}$ ,  $p_{C_i}^G$ ),  $i\in S_j$ 观测到同一个单独的特征 $f_i$ 。每一个特征的 $M_j$ 观测可以通过下面模型描述:

$$\mathbf{z}_{i}^{(j)} = \frac{1}{C_{i}Z_{j}} \begin{bmatrix} C_{i}X_{j} \\ C_{i}Y_{j} \end{bmatrix} + \mathbf{n}_{i}^{(j)}, \quad i \in \mathcal{S}_{j}$$

上式中, $n_i^{(j)}$ 是2×1的图像噪声向量,其协方差为 $R_i^{(j)}=\sigma_{im}^2I_2$ 。特征的位置表示在相机坐标系中为:

$${}^{C_i}\mathbf{p}_{f_j} = \begin{bmatrix} {}^{C_i}X_j \\ {}^{C_i}Y_j \\ {}^{C_i}Z_j \end{bmatrix} = \mathbf{C}({}^{C_i}_G\bar{q})({}^{G}\mathbf{p}_{f_j} - {}^{G}\mathbf{p}_{C_i})$$

上式中, $p_{f_i}^G$ 是3D特征点在全局坐标系下的位置,由于这是未知的,在算法的第一步使用最小二乘得到其估计值。这通过使用测量值 $z_i^{(j)}$ ,和对应时刻的相机位姿滤波估计值计算(详细见附录)

一旦得到特征位置的估计值,便可以计算测量的残差:

$$r_i^{(j)} = z_i^{(j)} - \hat{z}_i^{(j)} \tag{1}$$

上式中 $z_i^{(j)}$ 直接可以根据相机的测量值得到, $\hat{z}_i^{(j)}$ 可以根据下式得到:

$$\hat{\mathbf{z}}_{i}^{(j)} = \frac{1}{C_{i}\hat{Z}_{j}} \begin{bmatrix} C_{i}\hat{X}_{j} \\ C_{i}\hat{Y}_{j} \end{bmatrix} , \begin{bmatrix} C_{i}\hat{X}_{j} \\ C_{i}\hat{Y}_{j} \\ C_{i}\hat{Z}_{j} \end{bmatrix} = \mathbf{C} \begin{pmatrix} C_{i}\hat{q} \\ G \end{pmatrix} \begin{pmatrix} G \hat{\mathbf{p}}_{f_{j}} - G \hat{\mathbf{p}}_{C_{i}} \end{pmatrix}$$

线性化上述等式(1)中对相机位姿和特征位置的估计,可以得到:

$$\mathbf{r}_i^{(j)} \simeq \mathbf{H}_{\mathbf{X}_i}^{(j)} \widetilde{\mathbf{X}} + \mathbf{H}_{f_i}^{(j)G} \widetilde{\mathbf{p}}_{f_j} + \mathbf{n}_i^{(j)}$$

上式中, $H_{X_i}^{(j)}$  和 $H_{f_i}^{(j)}$  分别表示相机测量值相对状态量和特征位置的雅各比, $\tilde{p}_{f_i}^G$  表示特征 $f_i$  的位置估计误差。雅各比矩阵见下:

$$\mathbf{H}_{\mathbf{X}_{i}}^{(j)} = \begin{bmatrix} \mathbf{0}_{2\times15} & \mathbf{0}_{2\times6} & \dots & \underbrace{\mathbf{J}_{i}^{(j)}\lfloor^{C_{i}}\hat{\mathbf{X}}_{f_{j}}\times\rfloor}_{\text{Jacobian wrt pose i}} & -\mathbf{J}_{i}^{(j)}\mathbf{C}\binom{C_{i}}{G}\hat{q} \end{pmatrix} & \dots \end{bmatrix}$$

$$\mathbf{H}_{f_{i}}^{(j)} = \mathbf{J}_{i}^{(j)}\mathbf{C}\binom{C_{i}}{G}\hat{q}$$

$$\mathbf{J}_{i}^{(j)} = \nabla_{C_{i}\hat{\mathbf{p}}_{f_{j}}}\mathbf{z}_{i}^{(j)} = \frac{1}{C_{i}\hat{Z}_{j}} \begin{bmatrix} 1 \mid 0 & -\frac{C_{i}\hat{\mathbf{X}}_{j}}{C_{i}\hat{\mathbf{Z}}_{j}} \\ 0 & 1 & -\frac{C_{i}\hat{\mathbf{Y}}_{j}}{C_{i}\hat{\mathbf{Z}}_{j}} \end{bmatrix}$$

通过将所有的 $M_j$ 个特征测量的残差堆起来,便可以得到单个特征对多个相机位姿的约束形成的残差观测方程:

$$\mathbf{r}^{(j)} \simeq \mathbf{H}_{\mathbf{X}}^{(j)} \widetilde{\mathbf{X}} + \mathbf{H}_f^{(j)G} \widetilde{\mathbf{p}}_{f_j} + \mathbf{n}^{(j)}$$

由于不同图像中的特征观测时独立的,所以协方差矩阵 $n^{(j)}$ 是 $R^{(j)}=\sigma^2I_{2M_i}$ 。

需要注意的是,由于状态估计量X被用来计算特征点位置估计(见附录),两者间建立了联系。所以特征点的位置误差 $\tilde{p}_{f_i}^G$ 间接被状态量的误差 $\tilde{X}$ 进行了修正。因此,上述的残差格式与前述的EKF中的残差观测格式不同,所以该残差方程还不能直接在EKF更新中使用。为了解决这个问题,作者定义了一个残差 $r_o^{(j)}$ ,将 $r^{(j)}$ 投影到 $H_f^{(j)}$ 的左零空间中。具体而言,定义一个矩阵A,表示U阵,其列构成 $H_f$ 左零空间的基向量。可得:

$$\begin{aligned} \mathbf{r}_o^{(j)} &= \mathbf{A}^T (\mathbf{z}^{(j)} - \hat{\mathbf{z}}^{(j)}) \simeq \mathbf{A}^T \mathbf{H}_{\mathbf{X}}^{(j)} \widetilde{\mathbf{X}} + \mathbf{A}^T \mathbf{n}^{(j)} \\ &= \mathbf{H}_o^{(j)} \widetilde{\mathbf{X}}^{(j)} + \mathbf{n}_o^{(j)} \end{aligned}$$

由于 $2M_j imes 3$ 的矩阵 $H_f^{(j)}$ 是列满秩的,所以其左零空间的维数是 $2M_j - 3$ 。所以 $r_o^{(j)}$ 是一个  $(2M_j - 3) imes 1$ 的向量。该误差独立于3D特征的坐标位置,而且因此EKF更新可以基于此进行。上式定义了一个在特征 $f_i$ 处观测到的所有相机位姿构成的线性约束关系。此表示所有 $M_j$ 状态的可用测量信息  $z_i^{(j)}$ ,和继而计算得到的EKF更新是最优的,处理由于线性化引起的精度损失。

此外,为了计算残差 $r_o^{(j)}$ 和观测矩阵 $H_o^{(j)}$ ,酉阵A不需要显式计算。相反,向量r和矩阵 $H_X^{(j)}$ 在 $H_f^{(j)}$ 的零空间的投影可以使用给定的旋转非常方便有效的进行计算[21],时间复杂度是 $O(M_j^2)$ 。另外,由于矩阵A是酉阵,则噪声向量 $n_o^{(j)}$ 的协方差为;

$$E\{n_o^{(j)}n_o^{(j)T}\} = \sigma^2 A^T A = \sigma^2 I_{2M_i-3}$$
 (2)

上述残差计算过程不是基于同一个静态特征观测多张照片产生的地理约束形成的唯一可能表达式。

另一个可以采用的方法是,比如对 $M_j(Mj-1)/2$ 对图像( $C_{M_j}^2$ )构造极线约束。但是,这  $M_j(Mj-1)/2$ 对极线约束同样只会产生 $2M_j-3$ 个独立约束,因为每个测量被使用了多次,有些等式是相关的。作者已经进行实验表明采用极线约束形式的线性化方法更加复杂,且相比于上述方法性能略 差。

# 1.5 EKF更新