MSCKF(Multi-State Constraint Kalman Filter) 论文要点总结（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（*G*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**Algorithm 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_{wg}和n_{wa}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$\breve{x}$ = 变量x − 估计值$\hat{x}$）。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"

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

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：

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

上式中，$\hat{q}$和$\hat{p}$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：

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

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：

$$
\begin{gathered}
{}^{I}_{G}\dot{\bar{q}}(t) = \frac{1}{2}\boldsymbol{\Omega}\big(\boldsymbol{\omega}(t)\big){}^{I}_{G}\bar{q}(t), \quad \dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg}(t) \\
{}^{G}\dot{\mathbf{v}}_I(t) = {}^{G}\mathbf{a}(t), \quad \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t), \quad {}^{G}\dot{\mathbf{p}}_I(t) = {}^{G}\mathbf{v}_I(t)
\end{gathered}
$$

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：

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix}-\lfloor\boldsymbol{\omega}\times\rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0\end{bmatrix}, \quad \lfloor\boldsymbol{\omega}\times\rfloor = \begin{bmatrix}0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0\end{bmatrix}$$

ω_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：

$$
\begin{aligned}
\boldsymbol{\omega}_m &= \boldsymbol{\omega} + \mathbf{C}({}^{I}_{G}\bar{q})\boldsymbol{\omega}_G + \mathbf{b}_g + \mathbf{n}_g \\
\mathbf{a}_m &= \mathbf{C}({}^{I}_{G}\bar{q})({}^{G}\mathbf{a} - {}^{G}\mathbf{g} + 2\lfloor\boldsymbol{\omega}_G\times\rfloor{}^{G}\mathbf{v}_I + \lfloor\boldsymbol{\omega}_G\times\rfloor^2\,{}^{G}\mathbf{p}_I) + \mathbf{b}_a + \mathbf{n}_a
\end{aligned}
$$

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（ω_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：

$$
\begin{gathered}
{}^{I}_{G}\dot{\hat{\bar{q}}} = \frac{1}{2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}){}^{I}_{G}\hat{\bar{q}}, \quad \dot{\hat{\mathbf{b}}}_g = \mathbf{0}_{3\times 1}, \\
{}^{G}\dot{\hat{\mathbf{v}}}_I = \mathbf{C}_{\hat{q}}^T\hat{\mathbf{a}} - 2\lfloor\boldsymbol{\omega}_G\times\rfloor{}^{G}\hat{\mathbf{v}}_I - \lfloor\boldsymbol{\omega}_G\times\rfloor^2\,{}^{G}\hat{\mathbf{p}}_I + {}^{G}\mathbf{g} \\
\dot{\hat{\mathbf{b}}}_a = \mathbf{0}_{3\times 1}, \quad {}^{G}\dot{\hat{\mathbf{p}}}_I = {}^{G}\hat{\mathbf{v}}_I
\end{gathered}
$$

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

$$\mathbf{C}_{\hat{q}} = \mathbf{C}({}^I_G\hat{\bar{q}}),\ \hat{\mathbf{a}} = \mathbf{a}_m - \hat{\mathbf{b}}_a \text{ and } \hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_m - \hat{\mathbf{b}}_g - \mathbf{C}_{\hat{q}}\boldsymbol{\omega}_G.$$

所以，IMU误差状态的线性连续时变模型为：

$$\dot{\tilde{\mathbf{X}}}_{\text{IMU}} = \mathbf{F}\tilde{\mathbf{X}}_{\text{IMU}} + \mathbf{G}\mathbf{n}_{\text{IMU}}$$

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：

$$\mathbf{F} = \begin{bmatrix} -\lfloor\hat{\boldsymbol{\omega}}\times\rfloor & -\mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ -\mathbf{C}_{\hat{q}}^T\lfloor\hat{\mathbf{a}}\times\rfloor & \mathbf{0}_{3\times3} & -2\lfloor\boldsymbol{\omega}_G\times\rfloor & -\mathbf{C}_{\hat{q}}^T & -\lfloor\boldsymbol{\omega}_G\times\rfloor^2 \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix}$$

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*ω_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×6N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：

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

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：

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

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：

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

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：

$$ \mathbf{J} = \begin{bmatrix} \mathbf{C}\left({}^{C}_{I}\bar{q}\right) & \mathbf{0}_{3\times 9} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 6N} \\ \lfloor \mathbf{C}_{\hat{q}}^{T}\, {}^{I}\mathbf{p}_{C} \times \rfloor & \mathbf{0}_{3\times 9} & \mathbf{I}_{3} & \mathbf{0}_{3\times 6N} \end{bmatrix} $$

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**r**，该残差与状态误差$\tilde{X}$成正比，一般形式为：

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

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：

$$ \mathbf{z}_i^{(j)} = \frac{1}{{}^{C_i}Z_j} \begin{bmatrix} {}^{C_i}X_j \\ {}^{C_i}Y_j \end{bmatrix} + \mathbf{n}_i^{(j)}, \qquad i \in \mathcal{S}_j $$

上式中，$n_i^{(j)}$是2×1的图像噪声向量，其协方差为$R_i^{(j)} = \sigma_{im}^2 I_2$。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：

$$ \hat{\mathbf{z}}_i^{(j)} = \frac{1}{{}^{C_i}\hat{Z}_j} \begin{bmatrix} {}^{C_i}\hat{X}_j \\ {}^{C_i}\hat{Y}_j \end{bmatrix}, \quad \begin{bmatrix} {}^{C_i}\hat{X}_j \\ {}^{C_i}\hat{Y}_j \\ {}^{C_i}\hat{Z}_j \end{bmatrix} = \mathbf{C}({}^{C_i}_{G}\hat{\bar{q}})({}^{G}\hat{\mathbf{p}}_{f_j} - {}^{G}\hat{\mathbf{p}}_{C_i}) $$

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：

$$\mathbf{H}_{\mathbf{X}_i}^{(j)} = \begin{bmatrix} \mathbf{0}_{2\times 15} & \mathbf{0}_{2\times 6} & \dots & \underbrace{\mathbf{J}_i^{(j)} \lfloor {}^{C_i}\hat{\mathbf{X}}_{f_j} \times \rfloor \quad -\mathbf{J}_i^{(j)} \mathbf{C}({}^{C_i}_{G}\hat{q})}_{\text{Jacobian wrt pose i}} & \dots \end{bmatrix}$$

$$\mathbf{H}_{f_i}^{(j)} = \mathbf{J}_i^{(j)} \mathbf{C}({}^{C_i}_{G}\hat{q})$$

$$\mathbf{J}_i^{(j)} = \nabla_{{}^{C_i}\hat{\mathbf{p}}_{f_j}} \mathbf{z}_i^{(j)} = \frac{1}{{}^{C_i}\hat{Z}_j} \begin{bmatrix} 1 & 0 & -\frac{{}^{C_i}\hat{X}_j}{{}^{C_i}\hat{Z}_j} \\ 0 & 1 & -\frac{{}^{C_i}\hat{Y}_j}{{}^{C_i}\hat{Z}_j} \end{bmatrix}$$

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$R^{(j)} = \sigma^2 I_{2M_j}$。

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$2M_j \times 3$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$(2M_j - 3) \times 1$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;

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

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

下面细节性的描述EKF使用多个特征建立的约束进行更新的过程，EKF更新可以通过以下两个之一的事件进行触发：

- 当一个特征已经被一定数量的图像跟踪到，且不再被检测到时，该特征相关的相机位姿按照1.4中所述的进行处理。此事件一般当特征出了相机的视角时便会触发。
- 每当一张图像被记录时，当前相机位姿估计量便会被加到状态向量里（见1.3）。如果记录的图像数量达到了设定的最大阈值N_{max}，会移除掉至少一张过去时刻的图像（此操作应该是为了减小计算量，主要会用在缓慢移动和静止运动中）。在取消状态之前，为了使用他们的定位信息，所有发生在对应时刻的特征观测都会被用掉。在算法里使用时，会选择$N_{max}/3$个从倒数第二个过去时刻的位姿pose起在时间上均匀分布的poses。这些位姿在进行一次基于相同特征的多位姿约束的EKF更新之后会被丢掉。作者选择总是保持最老的状态位姿pose保留在状态向量里，因为将来新进的位姿pose和最旧的位姿产生的地理约束会有一个比较大的基线长度，会存在更有价值的定位信息。该方法在实际中表现很不错。

进一步讨论更新过程的细节。给定L个经过以上策略检测通过的特征点约束。接着根据之前描述的步骤，针对每个特征计算残差向量$r_o^{(j)}$，$j=1,2,3...L$，以及对应的测量矩阵$H_o^{(j)}$，$j=1,2...L$。通过堆叠所有的残差进一个向量里，则可以得到：

$$r_o = H_x \tilde{X} + n_o \tag{3}$$

由于这些特征测量在统计意义上是相互独立的，噪声向量$n_o^{(j)}$是互不相关的。因此，噪声的协方差矩阵是$R_o = \sigma_{im}^2 I_d$，$d=\sum_{j=1}^{L}(2Mj-3)$是残差r_o的维度。实际中可能出现的一个问题是，d会非常大。比如，如果在10个相机位姿处都看到10个特征，残差的维度是170。为了减小EKF更新的计算复杂度，对Hx进行QR分解。具体为：

$$\mathbf{H}_{\mathbf{X}} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{T}_H \\ \mathbf{0} \end{bmatrix}$$

上式中$Q1$和$Q2$是单位矩阵，其列分别构成H_x的基础范围和零空间，而且T_H是上三角矩阵。基于此，可得：

$$\mathbf{r}_o = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{T}_H \\ \mathbf{0} \end{bmatrix} \widetilde{\mathbf{X}} + \mathbf{n}_o \Rightarrow$$
$$\begin{bmatrix} \mathbf{Q}_1^T \mathbf{r}_o \\ \mathbf{Q}_2^T \mathbf{r}_o \end{bmatrix} = \begin{bmatrix} \mathbf{T}_H \\ \mathbf{0} \end{bmatrix} \widetilde{\mathbf{X}} + \begin{bmatrix} \mathbf{Q}_1^T \mathbf{n}_o \\ \mathbf{Q}_2^T \mathbf{n}_o \end{bmatrix}$$

由上一个等式可见，将残差r_0投影到H_x的基向量空间，并保持利用所有的有用测量信息。残差$Q_2^T r_o$只是噪声，可以被直接完全去掉。基于此，采用下式进行EKF更新：

$$\mathbf{r}_n = \mathbf{Q}_1^T \mathbf{r}_o = \mathbf{T}_H \widetilde{\mathbf{X}} + \mathbf{n}_n$$

上述表达中，$n_n = Q_1^T n_o$是一个噪声向量，其协方差等于$R_n = Q_1^T R_o Q_1 = \sigma_{im}^2 I_r$，r是$Q_1$的列数。EKF的更新过程中的卡尔曼增益计算为：

$$\mathbf{K} = \mathbf{P}\mathbf{T}_H^T \left(\mathbf{T}_H \mathbf{P} \mathbf{T}_H^T + \mathbf{R}_n\right)^{-1}$$

最后，这个状态协方差可基于下式更新：

$$\mathbf{P}_{k+1|k+1} = \left(\mathbf{I}_\xi - \mathbf{K}\mathbf{T}_H\right) \mathbf{P}_{k+1|k} \left(\mathbf{I}_\xi - \mathbf{K}\mathbf{T}_H\right)^T + \mathbf{K}\mathbf{R}_n\mathbf{K}^T$$

上式中$\xi = 6N + 15$是协方差矩阵的维度。

1.6 讨论

如上所述，滤波器的最终计算的复杂度是与观测到的特征数量成线性关系的，而且至多还会有状态数三次方个状态被包含在状态向量里。因此，被包含在状态里的位姿pose数量是影响算法时间复杂的主要因素。由于该数量是一个可配置的参数，可以根据计算资源以及应用需求场合修改。如果要求适应变化的的可用计算资源，滤波器状态可以被自适应的在滤波过程中调整。

使用相机观测在递推状态估计中困难的一个原因是测量模型的非线性。基于视觉的运动估计对噪声非常敏感，尤其是当观测到的特征距离较远时，错误的局部极小值会引起不一致的协方差。非线性导致的该问题，在已有的文献中可以使用 **Sigma-point**卡尔曼滤波，粒子滤波，以及对特征进行逆深度转换处理。所述算法的下面两个特点是的对线性不确定性具有鲁棒性：1）在测量模型中使用逆特征点深度参数，见附录；2）观测量的滞后线性化【17】。基于此算法的结构，每个特征的多个观测被采集，优先使用它们进行EKF更新，便可以实现测量雅格比更精确的评估。

另一个比较感兴趣的观测是，在典型的图像序列里或者当重新在某些地方定位时，大部分特征只能在很少的图像帧（随机特征）上被可靠跟踪到，而且几乎没有特征可以被长期跟踪到（持久特征）。这是因为相机视角的限制，以及障碍，图像噪声，还有可视点的变化，这些导致特征跟踪算法的失败。如之前所讨论的，如果已经被看到的一个特征的所有位姿pose已经被包含在状态向量里，此时提出的测量模型除了线性化导致的精度损失，是最优的。因此，对于实际图像序列，文章提出的算法可以最优的处理随机特征的位置信息。而且，状态向量X_k不要求只包含IMU和相机位姿pose。如果可以，持久特征也可以被包含在滤波器状态里，用来进行定位和建图。这将会在一片区域里使用回环提高定位精度。

附录

为了计算一个跟踪到的特征f_i的位置估计，作者采用$intersection$算法【27】。为了避免局部极小，而且为了更好的数值稳定性，在该过程中使用特征点的逆深度参数。特别是，如果$\{C_n\}$是第一次观测到特征的相机图像帧，接着这个特征坐标系相对相机在第i时刻可以表达为：

$$^{C_i}\mathbf{p}_{f_j} = \mathbf{C}(^{C_i}_{C_n}\bar{q})^{C_n}\mathbf{p}_{f_j} + {}^{C_i}\mathbf{p}_{C_n}, \quad i \in \mathcal{S}_j$$

上式中，$C(^{C_i}_{C_n}\bar{q})^{C_n}$和$^{C_i}P_{C_n}$分别是时刻$n$和时刻$i$间相机帧的旋转和平移。上式可进一步表达为：

$$\begin{aligned}
^{C_i}\mathbf{p}_{f_j} &= {}^{C_n}Z_j\left(\mathbf{C}(^{C_i}_{C_n}\bar{q})\begin{bmatrix}\frac{^{C_n}X_j}{^{C_n}Z_j}\\ \frac{^{C_n}Y_j}{^{C_n}Z_j}\\ 1\end{bmatrix} + \frac{1}{^{C_n}Z_j}{}^{C_i}\mathbf{p}_{C_n}\right)\\
&= {}^{C_n}Z_j\left(\mathbf{C}(^{C_i}_{C_n}\bar{q})\begin{bmatrix}\alpha_j\\ \beta_j\\ 1\end{bmatrix} + \rho_j\ {}^{C_i}\mathbf{p}_{C_n}\right)\\
&= {}^{C_n}Z_j\begin{bmatrix}h_{i1}(\alpha_j,\beta_j,\rho_j)\\ h_{i2}(\alpha_j,\beta_j,\rho_j)\\ h_{i3}(\alpha_j,\beta_j,\rho_j)\end{bmatrix}
\end{aligned}$$

在上述表达中，h_{i1}，h_{i2}，h_{i3}分别是变量α_j,β_j,ρ_j的尺度函数，其分别定义为：

$$\alpha_j = \frac{^{C_n}X_j}{^{C_n}Z_j}, \quad \beta_j = \frac{^{C_n}Y_j}{^{C_n}Z_j}, \quad \rho_j = \frac{1}{^{C_n}Z_j},$$

因此，综合前述算法中的公式，可以将测量等式表示为只包含α_j,β_j,ρ_j的函数：

$$\mathbf{z}_i^{(j)} = \frac{1}{h_{i3}(\alpha_j,\beta_j,\rho_j)}\begin{bmatrix}h_{i1}(\alpha_j,\beta_j,\rho_j)\\ h_{i2}(\alpha_j,\beta_j,\rho_j)\end{bmatrix} + \mathbf{n}_i^{(j)}$$

基于上述测量$z_i^{(j)}, i \in S_j$，以及状态向量里相机位姿的估计量，使用高斯牛顿最小二乘可以得到估计值$\hat{\alpha}_j, \hat{\beta}_j, \hat{\rho}_j$。接着，特征的全局坐标可以由下式计算：

$$
{}^G\hat{\mathbf{p}}_{f_j} = \frac{1}{\hat{\rho}_j}\mathbf{C}^T({}^{C_n}_G\hat{\bar{q}}) \begin{bmatrix} \hat{\alpha}_j \\ \hat{\beta}_j \\ 1 \end{bmatrix} + {}^G\hat{\mathbf{p}}_{C_n}
$$

需要注意的是，在最小二乘过程中，相机位姿估计值被认为是常量，它们的协方差是被忽略掉的。因此，在特征位置估计的优化消耗上，最小化的效率可以很高。回想，尽管这是一阶近似，但是这些估计误差不会影响测量残差。因此，不会对系统性能产生明显的负面影响。