# 摘要：

本文着重于使用惯性测量和自然发生点特征的观察来进行运动估计。迄今为止，这项任务主要是使用过滤方法来解决的，该方法从已知的初始条件开始跟踪系统状态。然而，当没有初始系统状态的先验知识可用时，（例如，在系统开始运行时），现有方法不适用。为了解决这个问题，在这项工作中，我们提出了直接从传感器测量中计算系统可观察量（平台姿态和速度、特征位置和 IMU 相机校准）的算法，无需任何先验知识。这项工作的一个关键贡献是一种基于凸优化的算法，用于计算相机和 IMU 之间的旋转矩阵。我们表明，一旦计算了这个旋转矩阵，所有剩余的量都可以通过求解二次约束最小二乘问题来确定。为了提高他们的准确性，初始估计由迭代最大似然估计器细化。

# I.INTRODUCTION

近年来，人们对使用视觉和惯性测量的运动估计方法越来越感兴趣，这项任务通常被称为视觉辅助惯性导航（参见例如 [1]-[7] 和其中的参考文献）。这有很多原因。首先，相机和 MEMS 惯性测量单元 (IMU) 都很紧凑、价格低廉、并且具有低功率要求。其次，这些传感器几乎可以在任何环境中运行，并允许进行全 3D 姿态估计，从而为导航提供了一个非常通用的解决方案。第三，近年来我们看到了设备的激增（例如，手机），其中包含相机和惯性传感器作为其标准传感器有效载荷的一部分。这些设备正逐渐变得无处不在，因此需要新技术来实现在 GPS 拒绝环境中的高精度导航。

绝大多数使用相机和 IMU 测量进行导航的现有技术要么采用递归贝叶斯估计方法 [1]-[4]，要么采用平滑公式 [5]。在这两种情况下，对状态的准确初始猜测（先前估计）对于可靠估计是必要的。这是因为这两种方法都依赖于测量模型的线性化，因此在没有准确的初始估计的情况下，大的线性化误差会导致发散。在当前实践中，为了初始化上面讨论的任何状态估计方法，通常使用域具体知识根据具体情况。例如，在某些应用中，额外的传感器（例如，倾角仪和/或 GPS）可能是可用的，或者可能知道平台最初是静止的。然而，这些方法并不普遍适用。

为了解决这个限制，在最近的工作 [8]-[10] 中，提出了仅基于对自然发生特征的观察来初始化系统状态的算法。这些方法提供了对移动平台的姿态（滚动和俯仰）、其速度、相机观察到的特征位置的初始估计，并且，在 [9] 的情况下，对于加速度计偏差。为了计算这些估计，这些方法假设相机到 IMU 的转换（旋转和平移）是先验已知的。虽然在已知用于导航的平台的情况下这可能是一个有效的要求，但并不总是能满足。例如，考虑移动电话用户想要使用该设备进行室内导航的情况。理想情况下，用户应该能够简单地将应用程序下载到他/她的设备，并立即能够使用它进行导航，而无需任何冗长的初始设置过程。实现这一目标的最简单方法是让导航算法能够估计所有必要的参数，包括相机到 IMU 的转换，而无需事先进行初始猜测。这是本文要解决的问题。

我们在这里提出算法，利用自然发生的点特征的观察，结合惯性测量，估计 (i) 相机姿态，(ii) 相机速度，(iii) 特征的位置，和 (iv)相机到 IMU 的转换。解决此问题的一种直接方法是将其表述为非线性最小二乘问题，我们尝试将特征相对于未知参数的重投影误差最小化。这当然是一种有效的方法，但它的局限性在于它要收敛到有意义的估计，它需要对未知数进行良好的初始猜测

本文的主要贡献是提供这种初始估计的方法。如先前的工作 [8]、[9] 所示，可以通过求解适当公式化的线性系统来解析计算相机速度和姿态以及特征位置。我们在这里展示了相机-IMU 平移也可以通过类似的线性系统来计算，而对于相机到 IMU 的旋转，我们提出了两种不同的方法：

1. 第一种方法需要在图像中观察到五个或更多特征，并计算表示相机到相机的单位四元数。 IMU 旋转作为线性最小二乘问题的解决方案。
2. 第二种方法只需要两个（或更多）特征，并采用一系列凸问题来获得解决方案。

如上所述获得的初始估计具有直接使用传感器数据计算的优点，但在统计上不是最优的。因此，它们随后使用最大似然估计器 (MLE) 进行改进，其公式导致非线性最小二乘问题，使用 L-M算法。除了为未知参数提供统计上的最优估计外，MLE 还为我们提供了估计误差的协方差矩阵。这使得在不同设置下测试估计的准确性成为可能，并得出关于人们可以预期的估计质量的实用结论。这些结果以及证明直接求解方法和迭代 MLE 的性能的测试在第 VII 节中介绍

# II. RELATED WORK

视觉辅助惯性导航系统的可观察性特性已在 [1]、[2] 中进行了检验。这些工作表明，在没有已知全局坐标的参考点的情况下，IMU 的全局位置以及围绕重力轴的旋转（即偏航）是不可观测的。另一方面，通常可观察到以下量：

1. IMU 相对于水平面的姿态（即横滚和俯仰）
2. IMU 轨迹（位置、速度和方向）相对于初始 IMU 坐标系
3. 相对于初始 IMU 帧的特征位置。
4. IMU 和相机帧之间的转换（即相机到 IMU 校准）
5. IMU 陀螺仪和加速度计偏差。

正如第一节所讨论的，我们在本文中的重点是估计上面的量 1-4。我们假设 IMU 偏差的估计已经可用（例如，可以假设这些偏差最初接近于零，然后可以在第 VI 节中描述的 MLE 细化中计算它们的高精度估计）。[1]、[2] 的可观测性结果证明（除了奇异轨迹，例如恒速运动）我们能够估计感兴趣的量，但没有显示如何直接从传感器数据计算这些量，而无需先前的估计。

如果已知相机到 IMU 的转换，则可以应用 [8]、[9] 的方法来估计 (O1)-(O3) 的量。为了计算这种转换，[11] 和 [12] 提出了依赖于观察已知校准模式的方法（以及在 [11] 的情况下使用专门的转盘）。虽然这些方法会很好适用于实验室环境，对专用设备的需求使其不太适合广泛使用。在 [1]、[2]、[13]、[14] 中，基于卡尔曼滤波器的估计器用于估计相机到 IMU 的转换，但由于这些需要良好的初始猜测，它们不适用于不存在先验知识的相关情况。与上述所有方法相比，我们在此提出了直接从传感器测量值估计相机到 IMU 转换的方法，无需先验知识，并且仅使用对自然发生特征的观察。在下文中，我们将介绍我们的工作细节。

# III. PROBLEM FORMULATION

在本节中，我们将介绍相机和 IMU 的问题公式和测量方程。我们表明，除了相机和 IMU 之间的旋转之外，我们试图估计的所有量（第 II 节中定义的（O1）-（O4））都可以线性估计。在第四节中，我们展示了如何从测量中恢复这种旋转，最后在第六节中，我们提出了一个最大似然估计器，它改进了这些初始估计，以提供系统中所有状态的高精度估计

考虑在时刻 t0、t1、...记录 N 个图像的情况。 . . , tN-1。通过采用合适的图像处理算法（例如，KLT 跟踪 [15] 或 SIFT 关键点提取和匹配 [16]），我们跟踪图像中的 M 个特征点。除了特征观察，时间间隔 [t0, tN−1] 的 IMU（陀螺仪和加速度计）测量可用。 IMU 陀螺仪和加速度计的测量值由以下等式 [17] 给出：

其中表示IMU坐标系中表示的3D旋转速度向量，是全局坐标系中的IMU加速度，是全局坐标系中表示的重力加速度向量，而和表示噪声分别在陀螺仪和加速度计测量中。我们在这里假设偏差至少是大致已知的（例如，从先前的传感器校准中）并且已经从 IMU 测量中移除。在本节的其余部分，我们将忽略测量中的噪声，并制定一个线性方程组，使我们能够估计除相机和 IMU 之间的旋转矩阵之外的所有量。

为了估计区间 内的 IMU 方向变化，我们整合了以下微分方程 [18]：

在 。这产生了旋转矩阵 ，它描述了和 之间的 IMU 旋转。 IMU 在时间 的全局位置由下式给出：

其中 是初始位置，是初始速度， = - 。使用（2）和上面的等式，我们得到：

其中：

接下来我们注意到 IMU 在时间 相对于 的位置由 给出。使用这个表达式，我们可以重新排列（4）得到：

其中 是相对于帧 表示的时间 的 IMU 速度，并且 是相对于同一帧表示的重力矢量。方程。 (6) 将在接下来的内容中有用，因为它只包含可观察的量，而没有全局（不可观察的）量。接下来我们将展示如何将相机测量值表示为 (6) 中出现的数量的函数。

假设一个经过内在校准的相机，在时间 对第 j 个特征的观察由透视相机模型描述：

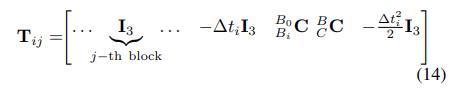
其中 是第 j 个特征在时间 相对于相机帧的位置，是测量噪声。我们使用集合 来描述描述可用测量的所有索引对

利用帧变换的基本性质，我们可以将向量 表示如下：

其中 是特征相对于 的位置，而 表示 IMU 和相机之间的恒定变换（旋转和平移）。使用 (6)，我们可以将 (8) 重写为：

上式右边的未知量是向量 , , 和 IMU 相机外在校准 而旋转矩阵 和向量 使用 IMU 测量值计算。我们现在使用 (7) 来获得（忽略测量噪声）：

使用 (9) 并重新排列项，上式可以写为 ，其中 是以下 (3M+9)×1 向量：



方程 源自对第 i 个图像中第 j 个特征的观察。通过收集所有特征和 IMU 测量结果的方程，我们得到线性系统:

其中 A 是具有块行 的矩阵，b 是具有块元素 的块向量，对于所有

现在让我们检查 (15) 中线性系统的性质。首先，我们注意到向量 x 包含恢复可观察项 (O1)-(O4) 所需的所有未知数，除了相机到 IMU 的旋转。具体来说，知道局部坐标系中的重力方向相当于知道 IMU 相对于水平面的姿态（量 (O1)）。此外，如果 和 已知，那么我们可以使用（6）估计局部帧中的 IMU 轨迹（数量（O2））。特征位置（O3）和相机到-IMU的杆臂 （(O4) 的一部分）明确包含在 x 中。因此，求解这个线性系统将使我们能够确定我们寻求的所有参数，除了 。从 (12)-(14) 我们看到矩阵 A 和向量 b 可以使用特征测量、IMU 测量和旋转矩阵 来计算。因此，我们看到，如果我们能够确定相机到 IMU 的旋转，我们可以轻松恢复所有剩余数量

## 所需的测量次数

一个重要的考虑因素是确定所需的图像和特征的数量，以便能够估计系统中的所有未知参数。我们检查的第一个问题是图像的最小数量。在[19]中，我们表明如果图像数量为 N ≤ 3，则线性系统 (15) 中的矩阵 A 是秩至少差 3。换句话说，即使 已知，我们仍然无法唯一确定系统中的未知参数。因此，所需的最小图像数量为 N = 4。

为了确定所需的特征数量，我们使用了一个计数参数：对于每个图像中的每个特征测量，我们从 Aijx = bij 获得 2 个标量方程。因此，对于 N 个图像和 M 个特征，测量约束的数量为 2NM。另一方面，可观察到的未知数为 3M +11（特征位置为 3M，6 用于相机到 IMU 的转换，3 用于初始速度，2 用于滚动和俯仰）。为了能够唯一确定所有未知数，我们必须有 2MN ≥ 3M+11，从中我们看到所需的最小特征数是 M = 3，如果有四张图像可用，则 M = 2，如果有五张图像可用, 和 M = 1,如果有七个或更多图像可用。在所有这些最小情况下，测量的数量高于未知数的数量，因此问题受到过度约束，并且可以计算出唯一的解决方案。

# IV. DETERMINING THE CAMERA-TO-IMU ROTATION

在本节中，我们提出了两种计算相机到 IMU 旋转矩阵的方法。当至少观察到 M = 5 个特征时，第一个适用。在这种情况下，我们可以估计不同图像之间的相对相机方向 ，使用 image-based motion estimation算法，例如 [20]（如果观察到 M = 4 个特征，同样可以完成，但所涉及的算法要复杂得多 [21]）。当 M < 4 时，不能单独使用特征测量来估计相机旋转。对于这种情况，我们在第四节B中描述 一种只需要 M ≥ 2 个特征来恢复 camera-to-IMU 旋转矩阵的方法

## Solution for the case M ≥ 5

我们首先考虑观察到足够特征的情况，以便我们可以仅使用特征观察来恢复不同时刻之间的相对相机方向 。通过使用 IMU 测量，我们可以通过 (3) 估计相同时间间隔内的 IMU 方向变化 。然后，我们可以对未知的 使用以下等式：

为了恢复矩阵 ，我们可以将上述方程转换为它们的等效单位四元数表示 [22]、[23]。具体来说，使用这种表示，我们有 [24]：

其中，对于 4 × 1 单位四元数 ，我们表示：

方程(18) 是 形式的线性系统，其中未知数是描述相机到 IMU 旋转的四元数 。通过使用所有可用的图像对，我们可以构建一个过约束的线性系统：

其中 B 是具有块行 的矩阵。最小二乘解 是对应于 B 的最小奇异值的右单位奇异向量，我们可以从中直接恢复旋转矩阵 [24]。为了使解决方案是唯一的，至少有两对图像，其中系统围绕不同的轴旋转，是必需的[23]。因此，我们看到至少需要三个图像，其中至少跟踪五个特征，这种方法才能确定相机到 IMU 的旋转。

## Solution for M ≥ 2 points

我们现在提出了一种替代方法，它可以对任意数量的点 M ≥ 2 进行操作，即，即使单独的特征测量不能用于确定图像之间的相对相机旋转，它也可以恢复矩阵 。

由叉乘的性质，我们知道对于任意向量 和 ，以下成立： 。因此，对于相机观察到的任意两个特征 m 和 n的时刻 ，我们可以写

接下来，我们使用 (7) 来写 ，因此：

在最后一个等式中， 是使用特征测量计算的已知向量，而 是使用 IMU 测量计算的已知矩阵。矩阵 和向量 是未知的。我们可以在向量 y 中收集所有未知数：

其中 是 3 × 3 旋转矩阵 的元素，并且 是 Δp 的元素。我们现在看到 (25) 是 y 的元素中的二次方程。从每幅图像中我们可以提取一个这样的方程，因此从 N 幅图像中我们获得了 y 元素中的 N 个二次方程。此外，旋转矩阵必须满足正交性约束，由 c 中的六个二次方程表示。最后，由于 (25) 在 ∆p 中是齐次的，我们必须对这个向量强制执行范数约束：

这是 Δp 中的另一个二次方程。因此，我们看到，总共有 N 幅图像，我们在 y 中获得了 N +7 个二次约束。由于 y 是一个 12 × 1 的向量，当至少有 5 张图像可用时，我们有许多方程等于未知数的数量。

到目前为止，只考虑了两个特征。为了将我们的公式扩展到观察到 M ≥ 2 个特征的情况，我们注意到（25）可以重写为：

类似地，我们根据以下未知向量得到一个二次方程组：

其中 c 受六个正交约束，而 y 的其余部分受尺度约束，使其范数等于 1。向量 y 包含 n = 3M + 6 个未知数

### Solving the system of equations using convex iteration: 1) 使用凸迭代求解方程组：

上面导出的方程形成了一个多元二次方程系统。已知以代数方式求解这样的系统是 NP 完全的 [25]，因此在这里我们将采用基于迭代凸近似的求解方法。具体来说，y 中的方程组可以写为，其中 p 是可用约束的数量（在 M = 2 的最小情况下为 12），是实对称矩阵， 是等于 0 或 1 的常数。因此，可以通过解决可行性问题来找到 y：

使用属性 ，我们可以将上述问题等效地写为：

其中 表示 n × n 半正定矩阵的圆锥。一旦找到该问题的解决方案，y 由 Y = 给出。上述可行性问题可以通过将秩约束放松为不等式来精确地重新表述：

由于唯一的秩为零的矩阵是零矩阵（这不是一个解），因此上述问题的解与（31）相同.为了获得上述问题的解决方案，我们将使用[26]的结果。具体来说，[26]表明，可以通过迭代解决以下两个凸优化问题来找到（32）的解决方案：

并且

其中 ¹ 表示半正定意义上的矩阵不等式。过程如下：我们为所谓的方向矩阵W选择一个初始值（在我们的实现中设置为零），并解决问题（33）得到Y？ .然后我们在问题（34）中使用这个值找到一个新的值W，并重复这个过程直到收敛。

在 [26] 中表明，当该迭代收敛时，获得的解决方案是原始问题 (32) 的精确解决方案（因此，也是我们寻找 y 的原始问题）。我们注意到问题 (33) 和 (34) 都是半定程序 (SDP)，并且可以使用现成的算法有效地解决。

1. Addressing the presence of noise: 解决噪声的存在：

方程。 (25) 仅在测量完美时才成立。当噪音出现时，我们将有

其中 是一个（小）误差。为了能够使用上面介绍的凸迭代公式解决问题，我们可以计算误差的上限，使得：

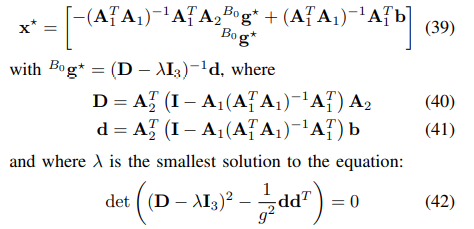
因此，我们现在可以将估计未知向量 y 的问题写成可行性问题：

其中不等式约束来自测量，而等式约束是由于对 y 的元素的正交性和单位范数约束。重要的是要注意，如在（31）中，这些约束为 Y 定义了一个凸集，因此仍然可以应用 [26] 的方法。所以，为了找到问题 (37) 的解决方案，我们使用类似于 (33) 和 (34) 的凸迭代，唯一的区别是 (33) 中的等式约束现在由等式和不等式的组合代替约束如上图所示。

向量 y（以及我们感兴趣的旋转矩阵 C BC ）的估计精度取决于边界 e b ij 的选择。如果选择过松的界限，那么得到的解决方案将是不准确的。另一方面，如果选择的边界太小，问题将变得不可行。为了解决这个问题，在我们的实现中，我们使用传感器噪声的已知统计数据来计算边界的初始低估计。我们开始使用这些值运行凸迭代，如果问题不可行（导致迭代停止 [26]），我们逐渐增加边界，直到收敛成功。

# V. SOLUTION OF THE LINEAR SYSTEM Ax = b

一旦通过上一节中描述的两种方法之一确定了旋转矩阵 C BC ，我们就可以继续求解线性系统 (15) 以恢复剩余的未知参数。在存在噪声的情况下，该系统将没有精确解，因此我们可以改为计算最小二乘解，即我们可以最小化函数。众所周知，这个问题的 x 的最优值由 给出。但是请注意，该解决方案没有利用重力加速度矢量的范数可能事先已知的事实。为了利用这些附加信息，我们可以制定一个有约束的最小二乘问题：

其中 g 是重力加速度范数的已知值，是包含地标位置、IMU 速度和 IMU 相机平移的向量（参见（11）），A 的划分与X。上述问题是一个二次约束的最小二乘问题。其最优解可以使用拉格朗日乘子法[19]推导出来，由下式给出

我们在 (42) 中计算其行列式的矩阵是一个 3 × 3 矩阵，其元素是 λ 中的二次多项式。因此，(42) 是 λ 中的六阶多项式方程。要找到最小的根，可以简单地计算所有根（这可以用非常低的计算成本以数值方式完成），然后选择最小的实数。我们的测试表明，当存在噪声时，使用已知的重力信息（即，使用受约束的最小二乘解（39）而不是无约束的解）可以显着提高估计精度。

# VI.MAXIMUM-LIKELIHOOD ESTIMATOR

至此提出的直接解决方案提供了无需任何事先初始猜测即可提供结果的优势。然而，它们在统计上并不是最优的，因为噪声的存在没有正确建模。为了正确考虑测量中的噪声，

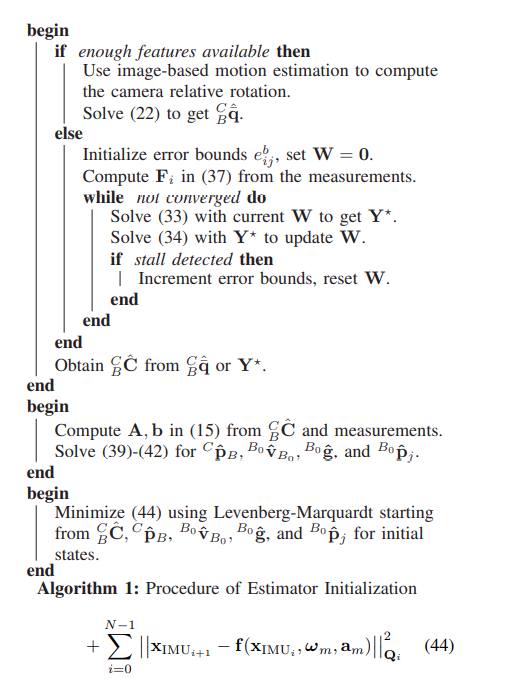
我们制定了一个用于估计参数向量 θ 的 MLE，包括

1. 在记录图像的每个时刻的 IMU 状态（位置、速度、方向），
2. 所有特征的位置 pj，j = 1。 . . M
3. 相机到 IMU 的转换。数量 (i)-(ii) 相对于原点与初始 IMU 框架的原点重合的框架表示，而其 z 轴与重力方向对齐。

按照标准做法，我们将图像测量噪声向量 建模为具有协方差矩阵 的高斯零均值随机变量。具体来说，我们使用时间间隔 [ti , ti+1] 中的 IMU 测量值来计算 IMU 状态的变化：

\

其中 是噪声向量，建模为具有协方差矩阵 Qi 的零均值高斯随机变量。函数 f 和协方差矩阵 Qi 使用连续时间运动方程的数值积分来计算 [18]。

最大化测量的可能性等同于最大化对数似然，这反过来又等价于最小化成本函数：2022-06-09 23-35-36 的屏幕截图\

其中 h(·) 是描述透视测量模型的函数（参见（7）），并且

成本函数 c(θ) 是非线性的，并且通过应用 Levenberg Marquardt 方法 [27] 迭代地执行其最小化。在我们的测试中，我们观察到如果使用第 IV 和 V 节中描述的直接解决方案来提供迭代的初始猜测，收敛速度很快，只需要几次（通常少于 10 次）迭代。我们还注意到，即使在推导直接解时假设 IMU 偏差是已知的，但如果需要，这些偏差可以作为未知数包含在 MLE 中，并与所有其他参数一起估计。

算法 1 总结了用噪声测量确定可观察量的完整过程。三个处理块对应于三个估计阶段，即 IMU 相机旋转估计（第 IV 节）、约束最小二乘解（第 V 节）和 MLE （第六节）。