

# Dynamic SLAM:The Need For Speed

- 论文: [Dynamic\\_SLAM\\_The\\_Need\\_For\\_Speed.pdf](#)
- 开源: 目前仍未开源, 实际上可参考 [VDO-SLAM](#)

## Contribution

- 1)一种新颖的位姿变化表示, 用于建模给定刚体相关点集合的运动, 并将该模型集成到SLAM优化框架中, 只需能够跟踪分割的物体, 而不需要三维模型
- 2)第一个能够估计相机姿态、静态和动态结构、场景中所有刚性目标的SE(3)位姿和速度, 并可在真实的户外数据集上演示
- 3)相比与先前DS-SLAM与DynaSLAM单纯的做动态特征点剔除, Dynamic SLAM:TNFS创新性的提出了对动态特征点进行追踪, 并放入非线性优化框架中, 使得动态特征点包含的信息被充分利用
- 4)Dynamic SLAM:TNFS中, 实现了动态物体速度的提取, 这在自动驾驶的碰撞避免之类的任务中可以起到重要帮助

## 内容细节

Dynamic SLAM:TNFS 重点内容在于提出在后端对动态特征点追踪并放入优化, 我们也会着重介绍这部分内容。

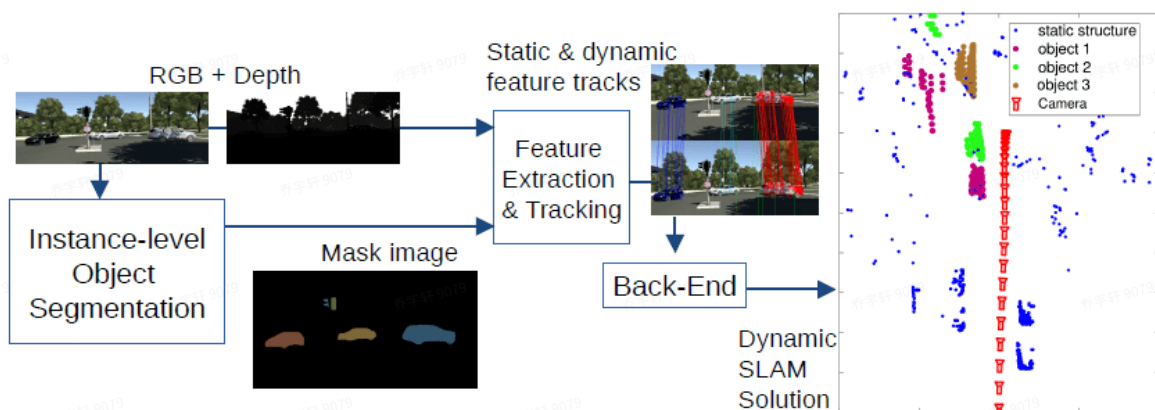


Fig. 4. System overview. Input images are used into an instance level object segmentation algorithm to provide objects masks. The algorithm then detects and tracks features on potentially moving objects. Potentially dynamic features along with tracked static features are used to build the graph, that is then fed into a back-end optimisation.

如上流程, 输入RGBD图像以及感知自身的里程计或IMU, 我们的后端会同时估计相机运动、静态和动态结构以及对象运动情况。而在此之中, 我们要尽量保证动态物体中的特征点被提取。因此我们需要引入(即时的)物体掩膜分割算法来先验地帮助我们追踪动态物体。

首先利用掩膜分割得到静态和(潜在的)动态物体,前端进行特征提取关联和跟踪,这是算法成功的关键。后端输入里程计测量与3D测量,进行估计。

## 1.目标运动模型

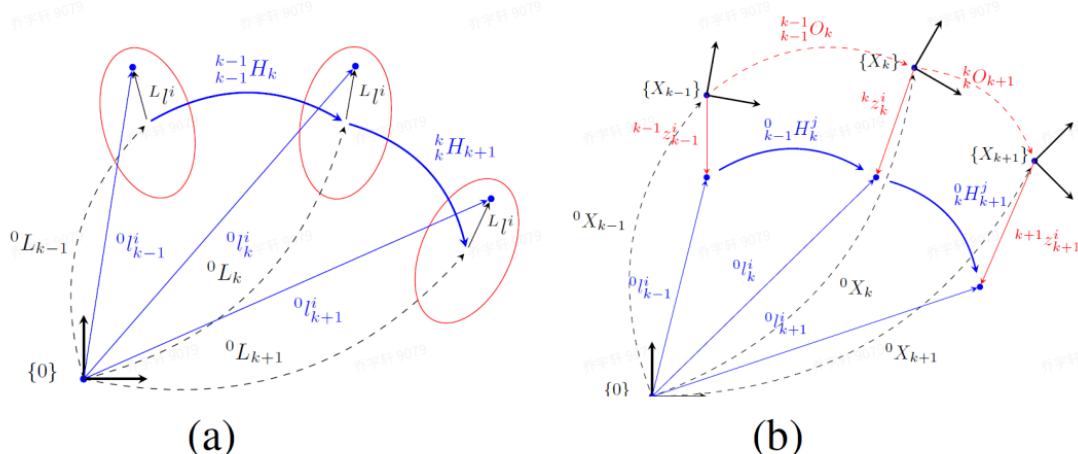


Fig. 2. (a): Coordinates of the rigid body in motion. The points  $L^i$  are represented relative to the rigid body pose  $\{L\}$  at each step. (b): Robot poses, moving landmark positions (black) and the measurements (red) at three different time steps.

当我们想对刚体的位姿变换进行估计时,问题之一在于坐标架的选取。记 $\{0\}$ 为刚体外的参考坐标架(可以认为是固定的世界坐标系), $\{L\}$ 为建立在移动刚体本身的坐标架(随时间在变动)。

目标特征点 $i$ 相对于 $\{0\}$ 在连续时间 $k$ 和 $k-1$ 下确定的刚体的相对位姿变换为:

$${}^0\bar{l}_k = {}_{k-1}^0H_k \cdot {}^0\bar{l}_{k-1} \quad (5)$$

此为本方法的关键公式,因为它消除了估计目标姿态的需要,并允许我们在 $\{0\}$ 中直接处理点坐标。

我们来对这个公式进行导出:

记 ${}^0L_k \in SE(3)$ 为移动刚体在时间 $k$ 相对于 $\{0\}$ 的位姿变换,对刚体上的某个特征点,记 $L^i \in \mathbb{R}^3$ 为 $\{L\}$ 下的坐标,注意此处 $L^i$ 不随刚体的运动发生改变。记 ${}^0l_k^i$ 为该点在时间 $k$ 相对于 $\{0\}$ 的坐标。

位姿 ${}^0L_k$ 是 $\{L_k\}$ ( $k$ 时刻的 $\{L\}$ )到 $\{0\}$ 的坐标变换,从而我们有 $L^i = {}^0L_k^{-1} \cdot {}^0l_k^i$  (1)

记 ${}_{k-1}^0H_k \in SE(3)$ 表示刚体 $L$ 从 $k-1$ 时刻到 $k$ 时刻的刚体变换,注意左上标表示该变换矩阵是在 $\{L_{k-1}\}$ 坐标架下得到的。依据该定义可以得到 ${}_{k-1}^0H_k = {}^0L_{k-1}^{-1} \cdot {}^0L_k$  (2)与 ${}^0L_k = {}_{k-1}^0H_k \cdot {}^0L_{k-1}$  (3)

结合(1),(2)两式,对于 $k$ 时刻与 $k-1$ 时刻分别带入可得 ${}^0\bar{l}_k = {}^0L_{k-1} \cdot {}_{k-1}^0H_k \cdot {}^0L_{k-1}^{-1} \cdot {}^0\bar{l}_{k-1}$  (4)

结合(2),(3)两式,可知 ${}^0\bar{l}_k = {}_{k-1}^0H_k \cdot {}^0\bar{l}_{k-1}$ ,即我们所需的关键式。(3)(4)中 $\bar{l}$ 为 $l$ 的齐次坐标。

关于式(2),(3),(4)的细节可参考下图

(2): 坐标系:  $\{L_{k-1}\} \xrightarrow{{}^0L_k} \{0\} \xrightarrow{{}^0L_{k-1}^{-1}} \{L_{k-1}\}$   
 坐标:  $L^i$   ${}^0L_k \cdot L^i$   ${}^0L_{k-1}^{-1} \cdot {}^0L_k \cdot L^i$   
 求  ${}_{k-1}^0H_k$ , 即考虑  $L^i$  在变换后在  $\{L_{k-1}\}$  下的坐标  
 而  $\{L_{k-1}\}$  中  $L^i$  在变换后的对应点正是  $\{L_k\}$  中的  $L^i$ . 该点在  $\{0\}$  下的坐标  ${}^0L_k \cdot L^i$

(3): 坐标系:  $\{0\} \xrightarrow{{}^0L_{k-1}} \{L_{k-1}\} \xrightarrow{{}^0L_k} \{0\}$   
 坐标:  ${}^0\bar{l}_{k-1}$   ${}^0L_{k-1} \cdot {}^0\bar{l}_{k-1}$   ${}^0L_k \cdot {}^0L_{k-1} \cdot {}^0\bar{l}_{k-1}$

$$\begin{aligned} \text{坐标} &= {}^0\bar{l}_{k-1}^i \rightarrow {}^0L_{k-1} {}^0\bar{l}_{k-1}^i \rightarrow {}^0L_k {}^0L_{k-1}^{-1} {}^0\bar{l}_{k-1}^i \\ {}^0\bar{l}_{k-1}^i &\text{是}\{L_{k-1}\}\text{中向量在}\{0\}\text{下的表示, } {}^0H_k = {}^0L_k \cdot {}^0L_{k-1}^{-1} \\ (4): {}^0\bar{l}_k^i &= {}^0L_k \cdot {}^L\bar{l}^i = {}^0L_{k-1} \cdot {}^L_{k-1} H_k \cdot {}^L\bar{l}^i = {}^0L_{k-1} \cdot {}^L_{k-1} H_k \cdot {}^L_{k-1}^{-1} \cdot {}^0\bar{l}_{k-1}^i \end{aligned}$$

我们可以看到式(5)的优越之处在于我们想要估计  ${}^0_{k-1}H_k$  时，并不要求刚体本身在k-1与k时刻的位姿  ${}^0L_{k-1}$  与  ${}^0L_k$ ，只需依赖其中点的坐标即可。而关于点的信息，我们已由特征点提取得到。

还有一优越性在于我们可以根据  ${}^0_{k-1}H$  导出刚体运动的速度： $v = {}^0_{k-1}t_k - (I_3 - {}^0_{k-1}R_k)c_{k-1}$

其中  ${}^0_{k-1}t_k$  与  ${}^0_{k-1}R_k$  分别代表  ${}^0_{k-1}H_k \in SE(3)$  的平移量与旋转矩阵， $I_3$  是单位矩阵， $c_{k-1}$  为物体的重心位置(相对于  $\{0\}$  的重心坐标)

推导过程：已知  ${}^0_{k-1}H_k = {}^0L_k \cdot {}^0L_{k-1}^{-1}$ ，记  ${}^0L_k$  的平移量与旋转矩阵分别为  ${}^0t_{L_k}$  与  ${}^0R_{L_k}$ ，代入可得  ${}^0_{k-1}H_k$  的平移量与旋转矩阵分别为  ${}^0t_{L_k} - {}^0R_{L_k} \cdot {}^0R_{L_k}^T \cdot {}^0t_{L_{k-1}} - (I_3 - {}^0R_{L_k} \cdot {}^0R_{L_k}^T) {}^0t_{L_{k-1}}$  与  ${}^0R_{L_k} \cdot {}^0R_{L_k}^T$ 。把这两个结果代入，速度计算公式右端变为  ${}^0t_{L_k} - {}^0R_{L_k} \cdot {}^0R_{L_k}^T \cdot {}^0t_{L_{k-1}} - (I_3 - {}^0R_{L_k} \cdot {}^0R_{L_k}^T) {}^0t_{L_{k-1}}$ ，(在坐标系  $\{0\}$  下， $c_{k-1}$  即可认为是坐标系  $\{L_{k-1}\}$  的原点坐标)，进一步变为  ${}^0t_{L_k} - {}^0t_{L_{k-1}}$ ，实际上就为  $k-1$  时刻到  $k$  时刻的平均速度。

注意我们的算法是稀疏的(基于提取的特征点)，所以对于重心的位置坐标采用的是通过特征点的3D坐标重心来估计的办法。实际上整个过程中我们主要规避了  ${}^0L_k$  的旋转部分的求解。

## 2.Dynamic SLAM中的移动因子(motion factor)

本方法同时估计相机姿态、静态和动态结构以及目标移动状态，故需要联合优化相关项：

$$\begin{aligned} \theta^* = \operatorname{argmin}_{\theta} \Big\{ & \sum_{k=1}^{m_k} \rho_h((h(x_k, l_k^i) - z_k^i)^\top \Sigma_{w_k}^{-1} (h(x_k, l_k^i) - z_k^i)) + \\ & \sum_{i=1}^{m_i} \rho_h((f(x_{k-1}, x_k) - o_k)^\top \Sigma_{v_k}^{-1} (f(x_{k-1}, x_k) - o_k)) + \\ & \sum_{i,j}^{m_s} \rho_h((g(l_{k-1}^i, l_{k,k-1}^j, {}^0H_k^j)^\top \Sigma_q^{-1} (g(l_{k-1}^i, l_{k,k-1}^j, {}^0H_k^j)) \Big\} \end{aligned} \quad (7)$$

$\rho_h$  为鲁棒Huber核函数，其作用在于减小较大单边误差在二范数下的影响，使其在某阈值外变为一次增长，对误差的梯度给出上界。h为3D点测量模型，z为点测量坐标；f为里程计模型，o为里程计测

量； $g$ 为移动目标点运动模型， $g(l_{k-1}^i, l_{k-1}^0, H_k^j) = ({}^0l_{k-1}^i) - {}_{k-1}^0 R_k^j \cdot l_{k-1}^0 - {}_{k-1}^0 t_k^j + q_{s_j}$ ， $q_{s_j} \sim N(0, \Sigma_q)$ 为高斯噪声，上式是原文中给出的公式，个人认为()中 ${}^0l_{k-1}^i$ 应该改为 ${}^0l_k^i$ 。

$H$ 为相对变换测量因子。总体实现的是三个加权最小二乘误差求和，协方差矩阵代表对各分量的权重，当协方差矩阵为单位矩阵时即为经典的最小二乘。所有优化变量放在集合  $\theta$  中，其中包括所有相机坐标 $x$ ，所有刚体坐标 $l$ 以及所有相对变换矩阵 $H$ 。

可以看到TNFS与静态SLAM后端的非线性优化的区别在于它加入了对目标点移动的捆绑调整。通过优化该最小二乘误差，实际上我们就已经把动态特征点的信息纳入考量了。但本文仅仅只给出了优化的目标函数，至于如何优化该函数，如何减少运算成本都是本文还未提到的。

### 3.因子图

Dynamic SLAM:TNFS给出了两张因子图来描述不同情景下的优化想法

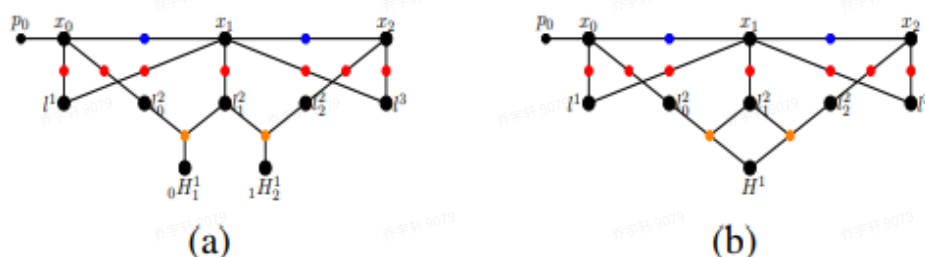


Fig. 3. Back-end (a): Factor graph representation of a problem with multiple pose change vertices for the same object. (b): Factor graph representation of a problem with a unique pose change vertex for the same object.

蓝色节点表示里程模型，红色节点表示测量模型，橘色节点表示刚体运动的位姿变换模型


- 城市情景下，物体运动变化较为频繁，随时都可能有加速减速。图3(a)表示了该情景之下的因子图，它为同一个物体的运动提供了两个时间变化下的两个顶点  ${}_0H_1^1$  与  ${}_1H_2^1$ ，原文中提出一个可能的约束：最小化两者的变化，可以参考VDO-SLAM中的平滑运动因素。
- 高速公路情景下，车辆局部时间上可以认为是匀速(速度变化较为缓慢)，对一个物体我们只给出一个移动估计。并且提出：刚体坐标系 $\{L\}$ 下恒定的变换对应参考坐标系 $\{0\}$ 下的恒定变换，这个推导相对来说没那么重要，在此处省略，有兴趣可参考原文。

在文章此部分给出的因子图优化中，与先前2中给出的优化目标貌似有所不同。在2中的观测误差部分，每个 $x_k$ 对应的只有一个 $l_k$ 进行观测误差的计算。在因子图中3个 $x$ 变量对应的红色节点远多于3个。猜测应该是对于位置 $x$ 附近所有(可能选取3个)静态或动态的目标都进行观测并计算误差。

### 后续

其实可以看出，Dynamic SLAM:TNFS仅仅只是提出了后端优化的一个想法：放入动态特征点追踪统一优化。这个想法很有意义。事实上，在自动驾驶的复杂道路环境下(如堵车情况),我们提取的大量特征点都来自动态物体，仅仅通过先前语义+几何直接剔除的办法显然不够优越。如此多的信息可根据本篇

文章的内容得到一个被利用的思路。本篇文章相对来说没那么完整，但在该文章发表的一年以后，相同的作者在此基础上提出了VDO-SLAM，一个更加完整成熟的Dynamic SLAM框架。

 VDO-SLAM