激光前端里程计的点云直接匹配算法实现

金宇强

摘要 点云匹配是利用激光里程计进行导航、定位与地图构建的基础，本文对直接匹配方法 ICP 和 NDT 进行了说明和实现。首先介绍了点对点 ICP 的两种解决方案：SVD 奇异值分解与基于优化的方法。然后介绍了 NDT 方法的基本思想。最后，基于 KITTI 数据集在 ROS 上分别实现了两种方法，并对最终的结果进行了分析，验证了有效性。下一报告中将着力于实现基于特征的匹配方法。

关键词 激光里程计；直接法；位姿估计

引言

定位是智能驾驶领域研究最核心的问题之一，是车辆实现真正自主的前提。一般的，车辆可以基于自身轮速计、GPS 等传感器来进行定位，然而由于车轮打滑漂移以及特殊环境对卫星信号的干扰，单依靠本体传感器并不能提供可靠的定位结果。于是，多传感器融合定位技术逐渐成为现如今最主要的解决方案。目前，最热门的实现方案主要有 VIO（Visual-Inertial Odometry）和 LIO（LiDAR-Inertial Odometry），相比于 VO， LiDar 对光照条件变化不敏感，能够提供更鲁棒的测量数据，而双目 VO 在成本上更有吸引力，且由于计算机视觉的快速发展，其在精度上也有显著提升，更是当前的研究热点。实际上这两种方案本身并不矛盾，甚至有很多共通的地方，如何将两者高效互补组合，也是一项有待研究的课题。而且，不论是定位还是建图，激光里程计都必然是其中重要的一环。

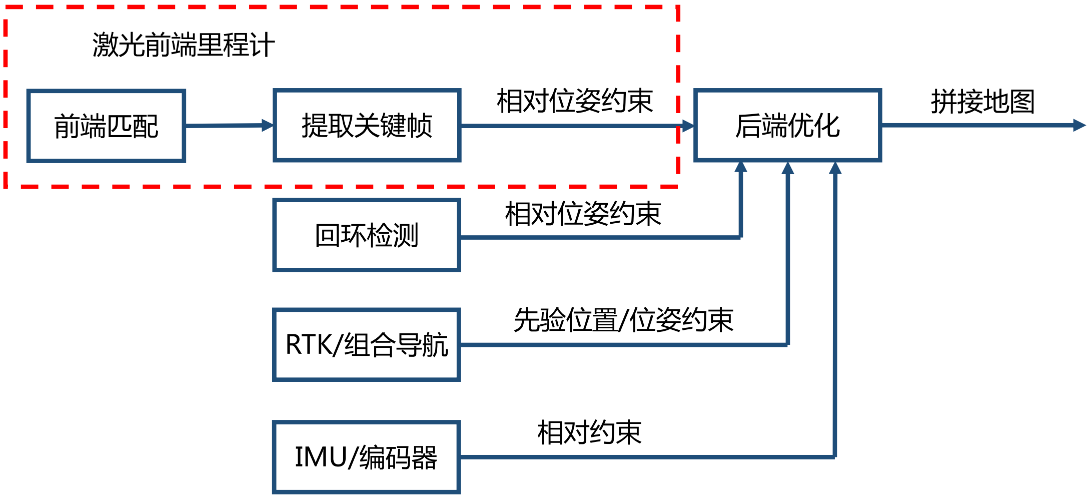


图 1 点云地图构建流程

于是，将以**本文**激光里程计 LO 的点云直接匹配方法为**出发点**，在假期内依次学习并介绍 LO 的特征匹配，建图，IMU 的建模与 LIO 融合，图优化方法、多传感器标定等内容。结合文献资料和前沿框架，以理解为目的，代码实现为方法。希望能学习形成一套系统的理论框架，同时锻炼代码能力。本文简要介绍了直接匹配方法的推导，并整理了一套可复现的代码框架，在整体完善后可用于之后自己的学习与组内传阅交流。

1 点对点的 ICP 算法

迭代最近点算法（Iterative Closest Point, ICP）是点云配准（Point Cloud Registration）中使用非常多的方法，其目的是对于输入的两幅点云的点集X（target）和Y（source），输出一个变换T使得$T(X)$和$T(Y)$的重合程度尽可能的高（本文默认点云数据集经过预处理，即已关联匹配）。一般只考虑变换T是刚性的，即变换只包括旋转R、平移t。具体数学描述为：若令$X = \{x_1, x_2, \dots, x_{N_x}\}$，$Y = \{y_1, y_2, \dots, y_{N_y}\}$，其中$X$和$Y$是原始点云的子集，一般的选取的是两个点集中能够互相关联的点，即$N_x = N_y$，算法的目标可以写为：

$$\min E(R,t) = min \frac{1}{N_y} \sum_{i=1}^{N_y} \|x_i - Ry_i - t\|^2 \tag{1}$$

即，经变换后的$y_i' = Ry_i + t$，该式表示x_i与y_i'的平均距离最小（$i = 1,2 \dots N_y$），直接求解步骤可见图 2，即利用初始R_0、t_0或上一次迭代得到的R_{k-1}、t_{k-1}对点云进行变换，得到k次迭代的变换点云，然后用这个点云和目标点云进行比较，找出源点云中每一个点在目标点云中的最近邻点。这样的方法比较费时，在下文中，将讨论点到点 ICP 算法求解的两个方法，即解析式求解方法 SVD 和基于迭代优化的求解方法。

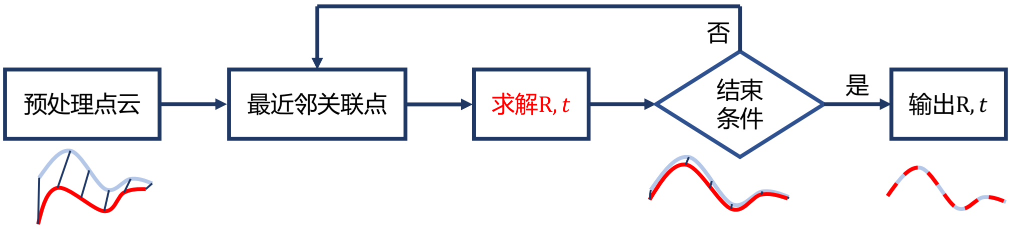


图 2 ICP 算法流程

1.1 基于 SVD 的 ICP 解法

借助 SVD，可以获得最优变换的闭解，具体推导如下。首先定义u_x和u_y为点集X和Y的质心，即：

$$u_x = \frac{1}{N_x} \sum_{i=1}^{N_x} x_i \qquad u_y = \frac{1}{N_y} \sum_{i=1}^{N_y} y_i \tag{2}$$

对于$E(R,t)$，有：

$$
\begin{aligned}
E(R,t) &= \frac{1}{N_y}\sum_{i=1}^{N_y}\|x_i - Ry_i - t\|^2 \\
&= \frac{1}{N_y}\sum_{i=1}^{N_y}\left\|x_i - Ry_i - t - u_x + Ru_y + u_x - Ru_y\right\|^2 \\
&= \frac{1}{N_y}\sum_{i=1}^{N_y}\left\|x_i - u_x - R\left(y_i - u_y\right) + u_x - Ru_y - t\right\|^2 \\
&= \frac{1}{N_y}\sum_{i=1}^{N_y}\left(\left\|x_i - u_x - R\left(y_i - u_y\right)\right\|^2 + \left\|u_x - Ru_y - t\right\|^2\right) \qquad (3)
\end{aligned}
$$

其中将第三个等号的平方项展开，因为有$\frac{1}{N_y}\sum_{i=1}^{N_y}(x_i - u_x) = \frac{1}{N_y}\sum_{i=1}^{N_y}\left(y_i - u_y\right) = 0$，所以$2\left(x_i - u_x - R\left(y_i - u_y\right)\right)^T\left(u_x - Ru_y - t\right) = 0$恒成立，即第四个等式成立。那么，原目标函数可以写为式（3）的形式，容易发现，在式中，第二项与i无关，因此可以放在求和符号外。令：

$$E_1(R,t) = \sum_{i=1}^{N_y}\left\|x_i - u_x - R\left(y_i - u_y\right)\right\|^2 \qquad (4)$$

$$E_2(R,t) = \left\|u_x - Ru_y - t\right\|^2 \qquad (5)$$

则$E(R,t) = E_1(R,t) + E_2(R,t)$。其中，式（4）只与旋转有关，式（5）只用于求平移部分，并且对于任意的R，均可以找到一个t，使得式(5)最小，即$E_2(R,t) = \left\|u_x - Ru_y - t\right\|^2 = 0$。因此，可以将原先一步优化的步骤拆开成两步，即先根据$E_1(R,t)$求得旋转矩阵，然后再根据$E_2(R,t)$求平移向量。在此基础上，有：

$$
\begin{aligned}
E_1(R,t) &= \frac{1}{N_y}\sum_{i=1}^{N_y}\left\|x_i - u_x - R\left(y_i - u_y\right)\right\|^2 \\
&= \frac{1}{N_y}\sum_{i=1}^{N_y}\|x_i' - Ry_i'\|^2 \\
&= \frac{1}{N_y}\sum_{i=1}^{N_y}(x_i'^T x_i' + y_i'^T R^T R y_i' - 2x_i'^T R y_i') \qquad (6)
\end{aligned}
$$

其中$x_i' = x_i - u_x$，$y_i' = y_i - u_y$，基于旋转矩阵R的性质（单位正交阵），$R^T R = I$，容易见第一、二项均与R无关，因此令$E_1'(R,t) = \sum_{i=1}^{N_y} x_i'^T R y_i'$，则：

$$\arg\min_R E_1(R,t) = \arg\max_R E_1'(R,t) = \arg\max_R \sum_{i=1}^{N_y} x_i'^T R y_i' \qquad (7)$$

为了求解式(7)，需要将E_1'转化为矩阵形式，然后根据矩阵性质进一步化简：

$$E_1'(R,t) = \sum_{i=1}^{N_y} x_i'^T R y_i' = \sum_{i=1}^{N_y} \mathrm{tr}(x_i'^T R y_i') = \sum_{i=1}^{N_y} \mathrm{tr}(R y_i' x_i'^T)$$
$$= \mathrm{tr}\left(\sum_{i=1}^{N_y} R y_i' x_i'^T\right) = \mathrm{tr}(RH) \qquad (8)$$

其中$H = \sum_{i=1}^{N_y} y_i' x_i'^T$，由于常数项的迹等于它本身，第二个等号成立；且有推论$\mathrm{tr(AB)} = \mathrm{tr(BA)}$，第三个等式成立。至此，式（7）的优化问题转化为：找到合适的R，使得$\mathrm{tr}(RH)$达到最大值。进一步的，引入下述定理：

定理 1：若有正定矩阵AA^T，则对于任意正交矩阵B，有$\mathrm{tr}(AA^T) \geq \mathrm{tr}(BAA^T)$

证明：
同样根据$\mathrm{tr(AB)} = \mathrm{tr(BA)}$，有：

$$\mathrm{tr}(BAA^T) = \mathrm{tr}(A^T BA) = \sum_i a_i^T (Ba_i)$$

其中a_i为A的列向量。根据柯西-施瓦兹不等式，有:

$$a_i^T (Ba_i) \leq \sqrt{(a_i^T a_i)(a_i^T B^T B a_i)} = a_i^T a_i$$

其中$B^T B = I$，因此，

$$\mathrm{tr}(BAA^T) = \sum_i a_i^T (Ba_i) \leq a_i^T a_i = \mathrm{tr}(AA^T)$$

该定理的意义在于说明了：R是$\max\limits_R \mathrm{tr}(RH)$的解当且仅当$R$可以将$\mathrm{tr}(RH)$转化为$\mathrm{tr}(AA^T)$的形式。基于该定理的结果，就可以使用 SVD 来获得最终的解析解。首先，对H进行 SVD，有：

$$H = U\Sigma V^T \qquad (9)$$

取$R = VU^T$，则：有问题

$$RH = VU^T U\Sigma V^T = V\Sigma V^T = V\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}V^T = V\Sigma^{\frac{1}{2}}\left(V\Sigma^{\frac{1}{2}}\right)^T \qquad (10)$$

显然，式（10）的结果符合定理 1。因此当$R = VU^T$时，$\mathrm{tr}(RH)$最大。最后，平移矩阵可以根据式（5）计算所得，即：

$$t = u_x - Ru_y \qquad (11)$$

1.2 基于优化的 ICP 解法

SVD 的思路是一步求解，基于点云数据先验知识（关联关系），通过一步运算求得R和t。而当缺少先验知识时，即匹配关系未知时，无法获得解析解，只能使用非线性优化的方法，其思想是先构造损失函数，然后通过迭代的方式（梯度下降）找到最优值。基于泰勒展开，迭代的策略有诸如最速下降法、牛顿法、高斯牛顿法和 LM 方法等，实际过程中可以根据问题特性选择合适的方法[2]。

对于 ICP 求解问题，损失函数可以表示为：

$$\min_{T} = \frac{1}{2}\sum_{i=1}^{n}\|x_i - Ty_i\|_2^2 \tag{12}$$

其中$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$。为了求式（12）的雅克比矩阵，需要将其转化为李代数模式：

$$\min_{T} = \frac{1}{2}\sum_{i=1}^{n}\|x_i - \exp(\xi^\wedge)\, y_i\|_2^2 \tag{13}$$

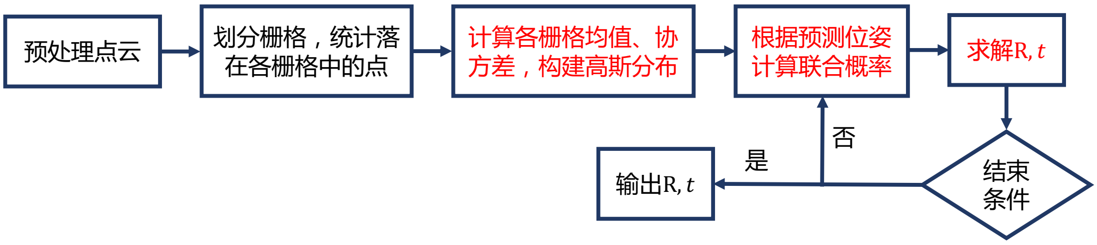
其中ξ为T对应的李代数，对应的雅克比为：

$$J = \frac{\partial e}{\partial \delta\xi} = -(\exp(\xi^\wedge)\, y_i)^{\odot} \tag{14}$$

然后就可以根据优化的固定步骤求解位姿。其中的李群李代数的相关知识在 SLAM 中的应用已非常广泛，在本文中只基于文献[2]给出相应的结论，不做展开，在后续报告中或将给出李群李代数的相关推导。

2 正态分布变换算法 NDT

基于直接匹配的另一种方法是 NDT（Normal Distributions Transform）算法，其是一种基于概率的匹配方法。同样地，对于点集 target：$X = \{x_1, x_2, \dots, x_{N_x}\}$，source：$Y = \{y_1, y_2, \dots, y_{N_y}\}$，首先在$Y$中划分栅格，在每个栅格中计算均值、协方差并构建高斯分布，然后基于X中落在Y栅格上的点与栅格上的均值协方差计算联合概率，将联合概率作为位姿估计的评价指标，具体步骤可见图 3。



预处理点云
划分栅格，统计落在各栅格中的点
计算各栅格均值、协方差，构建高斯分布
根据预测位姿计算联合概率
求解R, t
结束条件
否
是
输出R, t

图 3 NDT 算法流程

算法的目标函数可表示为[3]：

$$max\,\Psi = max\prod_{i=1}^{N_y} f\big(X, T(p, y_i)\big) \tag{15}$$

其中对于 2D 模型：$p = p_3 = [t_x \quad t_y \quad \phi_z]^T$，对于 3D 模型：$p = p_6 = [t_x \quad t_y \quad t_z \quad \phi_x \quad \phi_y \quad \phi_z]^T$。根据预测的位姿，对点进行平移和旋转，令$y_i' = Ry_i + t$，旋转和平移后的点$y_i'$与目标点集$X$中的点在同一坐标系下，此时可计算各点的联合概率：

$$f(X, y_i') = \frac{1}{\sqrt{2\pi}\sqrt{|\Sigma|}} exp\left(-\frac{(y_i' - \mu)^T \Sigma^{-1}(y_i' - \mu)}{2}\right) \tag{16}$$

因此可得所有点的联合概率，为了简化问题，取其对数：

$$ln\,\Psi = \sum_{i=1}^{N_y}\left(-\frac{(y_i' - \mu)^T \Sigma^{-1}(y_i' - \mu)}{2} + ln\left(\frac{1}{\sqrt{2\pi}\sqrt{|\Sigma|}}\right)\right) \tag{17}$$

并去除第二项（常数项），式（15）可写作：

$$max\,\Psi = max\,ln\,\Psi = min\sum_{i=1}^{N_y}(y_i' - \mu)^T\Sigma^{-1}(y_i' - \mu) \tag{18}$$

又因为$y_i' = Ry_i + t$，待求参数是R和t，所以可以定义损失函数$f_i(p) = y_i' - \mu$，基于高斯牛顿法的流程，只需计算损失函数关于待求参数的雅克比$J_i = \frac{df_i(p)}{dp}$，便可进行迭代优化。

3 畸变补偿

通常我们将雷达内部旋转一周扫描得到的点集作为一帧点云来处理，用这种方法分帧的优势在于一周扫描正好是对周围环境的一次完整采集，并且足够的激光点点数量是匹配的前提。但是由于每个激光点的坐标是相对于雷达的，而雷达运动时，不同激光点的坐标原点均不相同，这也就导致了点云的运动畸变，可见图 4，其可分为平移畸变和旋转畸变。

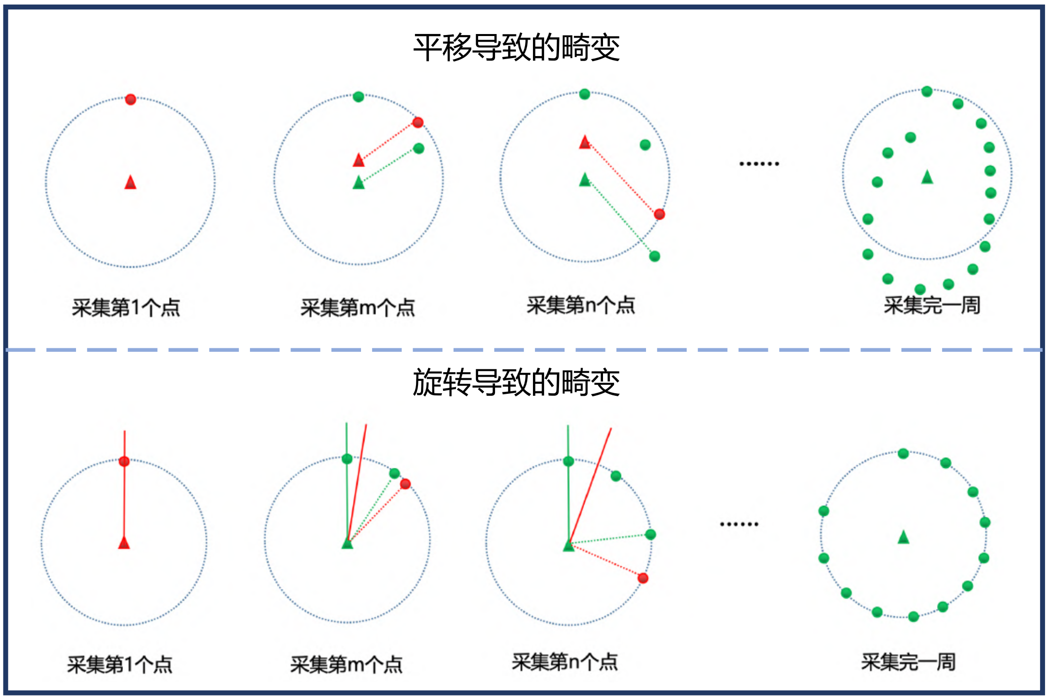


图 4 点云的运动畸变

图 4 中虚线圆圈表示真实物体，红色点为激光束打到的位置，红色三角为当前采集时刻雷达的位置，绿色三角为一帧的坐标原点，绿色圆圈为一帧点云激光点的坐标，下图中的红色实线表示雷达 0 度的坐标轴，绿色实现为一帧点云 0 度坐标轴。畸变的原因理解起来比较简单，因此不做过多介绍。

解决点云畸变基本思路为求解每一个激光点时刻相对于一帧点云激光点的坐标，这需要知道每一时刻雷达的相对位移，然后对每一个激光点坐标做补偿，补偿量为激光点原点（即当前雷达坐标）相对于该帧起始时刻的变化。假设一帧点云中，起始时刻的雷达位姿为$T_0 = \begin{bmatrix} R_0 & t_0 \\ 0 & 1 \end{bmatrix}$，第$i$个激光点采集时，雷达的位姿为$T_i = \begin{bmatrix} R_i & t_i \\ 0 & 1 \end{bmatrix}$，第$i$个激光点坐标$P_i$为$[p_{ix} \quad p_{iy} \quad p_{iz}]^T$,则第$i$个激光点补偿畸变

后的坐标应为：

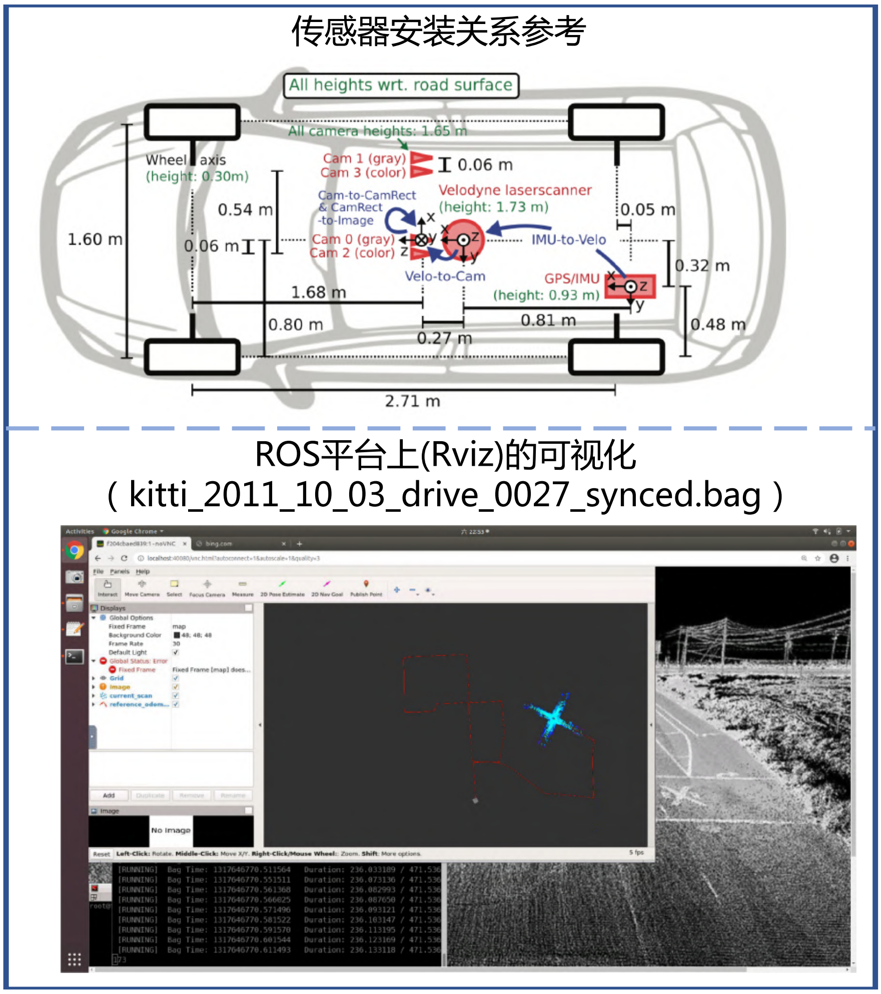
$$\bar{P}_i = T_0^{-1} T_i P_i \tag{19}$$

实际上，雷达点云是局部坐标系下的表示，当以 0 时刻雷达的位姿为基准坐标系时，此时T_0为单位矩阵，T_i为 0 到i时刻的相对旋转和平移，因此只需要知道 0 到i时刻的平均角速度和平均速度即可（可由 IMU 或轮速计获得）。

畸变补偿在实际问题中存在很多变动（如 KITTI 数据），像雷达内部旋转的方向不同，点云时间戳定位不同等，需要应对各种情况对补偿算法作出适应性改动，因此，重点在于理解雷达运动造成点云畸变的核心机理。

4 基于 KITTI 数据集的实现

本文选用 2012 年开源的 KITTI[4]作为算法实现的数据集，受限于当时的硬件水平，该数据集有一些数据缺漏或不完整性等，但不影响算法的验证。但是经过阅读源码发现 KITTI 中的畸变补偿算法与第三章介绍不同，因为 KITTI 提供的数据打乱了点云排列的格式，但据文[4]以及诸多网络博客的说明，这样的问题在实际工程中不容易出现，因此在本文不做额外考虑。采集数据的传感器安装关系以及 KITTI 在 ROS 平台上的 rosbag 可视化可见图 6。



传感器安装关系参考

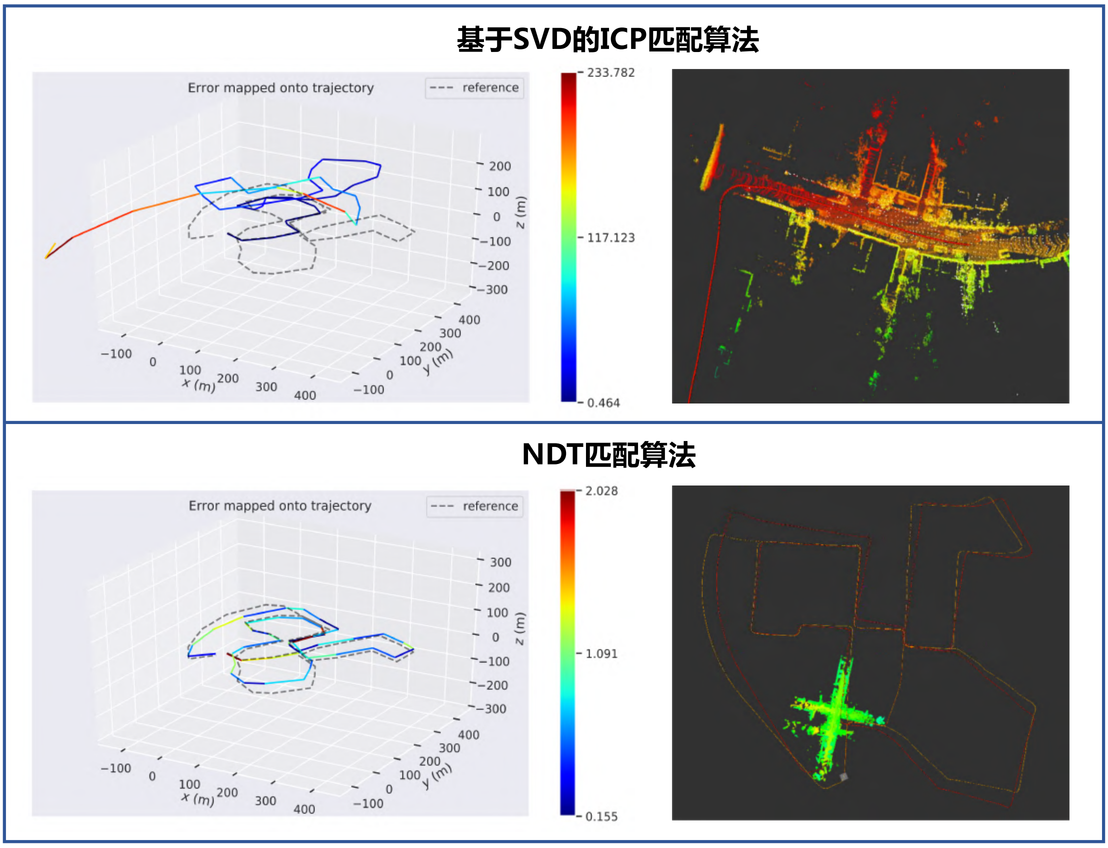
ROS平台上(Rviz)的可视化
(kitti_2011_10_03_drive_0027_synced.bag)

```
[RUNNING]  Bag Time: 1317646770.511564   Duration: 236.033189 / 471.536
[RUNNING]  Bag Time: 1317646770.551511   Duration: 236.073136 / 471.536
[RUNNING]  Bag Time: 1317646770.561368   Duration: 236.082993 / 471.536
[RUNNING]  Bag Time: 1317646770.566025   Duration: 236.087650 / 471.536
[RUNNING]  Bag Time: 1317646770.571496   Duration: 236.093121 / 471.536
[RUNNING]  Bag Time: 1317646770.581522   Duration: 236.103147 / 471.536
[RUNNING]  Bag Time: 1317646770.591570   Duration: 236.113195 / 471.536
[RUNNING]  Bag Time: 1317646770.601544   Duration: 236.123169 / 471.536
[RUNNING]  Bag Time: 1317646770.611493   Duration: 236.133118 / 471.536
173
```

图 5 传感器安装关系[4]以及 rosbag play kitti

本文以及接下来在 KITTI 上实现的代码都是在[5]的框架结构进行修改和功能实现，NDT 的代码实现基于 PCL（Point Cloud Library）[6]，其是一个用于图像和点云处理的大型开源库，封装了大量的先进算法（包括滤波、特征估计、表面重构、配准、模型拼合和分割）。ICP 的代码实现参考了[7]的方法，经比较，其与 PCL 库实现的 ICP 方法结果相近（误差均较大）。

本文对两种基于直接匹配的位姿估计算法在 KITTI_2011_10_03_drive_0027_synced 数据包上进行实验。将 KITTI 中组合导航的结果（bag 中已给出）作为真值，来对激光里程计精度结果进行评价。最终的定位轨迹可视化和点云视图可见图 6，左边图表示最终的定位结果，灰色虚线为组合定位真值，实线为直接点匹配法位姿估计结果，其中，线条颜色越趋近于红色，则相对误差越大。右图为实验中的点云视图，可见基于 ICP 匹配算法在 KITTI 数据及上最终完全偏离了真值轨迹（有待检验 debug），而 NDT 方法能够给出完整且较准确的定位。


基于SVD的ICP匹配算法

NDT匹配算法

图 6 轨迹可视化结果

对轨迹的误差分析图可见图 7，使用相对位姿误差(RPE：relative pose error)来计算误差，本实验设置间隔$\Delta = 100$，对于估计位姿$P_1, \dots, P_n \in SE(3)$，真实位姿$Q_1, \dots, Q_n \in SE(3)$，其中下标表示帧数，对于第$i$帧的 RPE，有：

$$RPE_i = (P_i^{-1} P_{i+\Delta})(Q_i^{-1} Q_{i+\Delta})^{-1} \qquad (20)$$

得到所有帧的 RPE 后，计算均方根误差 RMSE，其中$m = n/\Delta$：

$$RMSE(RPE_{1:n},\mathit{\Delta}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}||RPE_i||^2} \tag{21}$$

该标准用于估计系统的漂移，最终结果可见图 7。

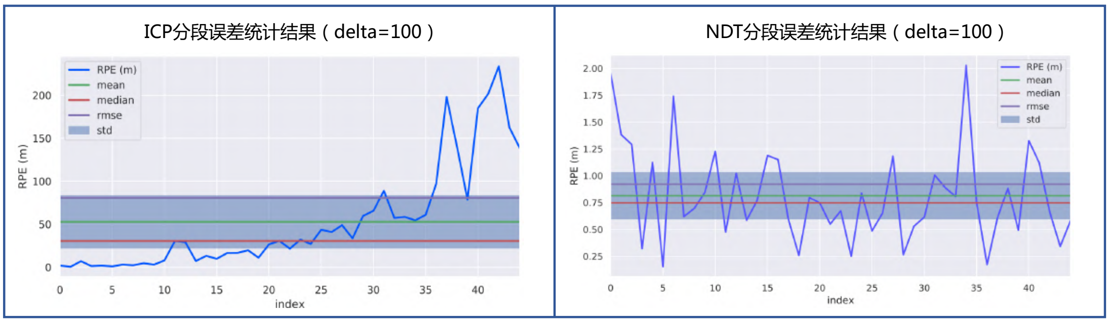


图 7 分段 RPE 统计结果

可以发现，ICP 算法匹配的误差非常大，尤其是对于 KITTI 这种长路段，最终会出现误差发散的问题，经过仔细的代码调试以及结果分析，认为原因可能有四个（可能与 KITTI 数据集的特性有关，还需进一步讨论）：(1)缺少足够的先验：点间关联正确是使用 SVD 进行位姿估计的的前提，但这样的条件较为苛刻；(2)收敛问题：ICP 算法总是收敛到局部极值，导致最终结果偏离真值，这样的误差往往难以建模；(3)欠约束导致的误差：一些环境下没有足够的信息来估计完整的位姿信息，如长直线路段或圆弧转弯路段；(4)传感器噪声引起的误差：从理论上理解，直接点-点匹配的方法对于传感器和环境噪声非常敏感，其会带来较大的误差。 相较之下，同样是基于点的扫描匹配方法，NDT 利用数学性质（概率密度）来刻画点云数据，和 ICP 不同，其不需要建立点间关联，对于存在错误关联的情况也更加鲁棒，因而最终结果也表现的较好，但其在求解过程中使用了优化算法，仍存在可能陷入局部最优等优化常见的问题。

5 小结

经过理论学习和实验分析，介绍了应用成熟的点云匹配解决方案 NDT 和 ICP，实验中 NDT 表现出更好的效果。但经调研，实际中为了保证算法效率，基于 SVD 的 ICP 仍有较多的应用，且综合反响较好，因此本文中的 ICP 实验结果还有待考证，误差发散的结果仍需进一步讨论。

参考文献

[1] Sorkine-Hornung O, Rabinovich M. Least-squares rigid motion using svd. 2017. https://igl.ethz.ch/projects/ARAP/svd_rot.pdf.

[2] 高翔，张涛，刘毅，颜沁睿．视觉 SLAM 十四讲:从理论到实践．北京：电子工业出版社，2019．第六章

[3] A. Das and S. L. Waslander, “Scan registration using segmented region growing NDT,” Int. J. Rob. Res., vol. 33, no. 13, pp. 1645–1663, 2014.

[4] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3354–

3361, 2012.

[5] Github：https://github.com/Little-Potato-1990?tab=repositories

[6] Point Cloud Library：https://pointclouds.org/

[7] Github: https://github.com/tttamaki/SICP-test

[8] EVO package：https://pypi.org/project/evo/

文献阅读 1

Some interesting facts about Paper:

ClusterVO: Clustering Moving Instances and Estimating Visual Odometry for Self and Surroundings

of **CVPR2020**：

1. Comparison of some dynamic SLAM results:

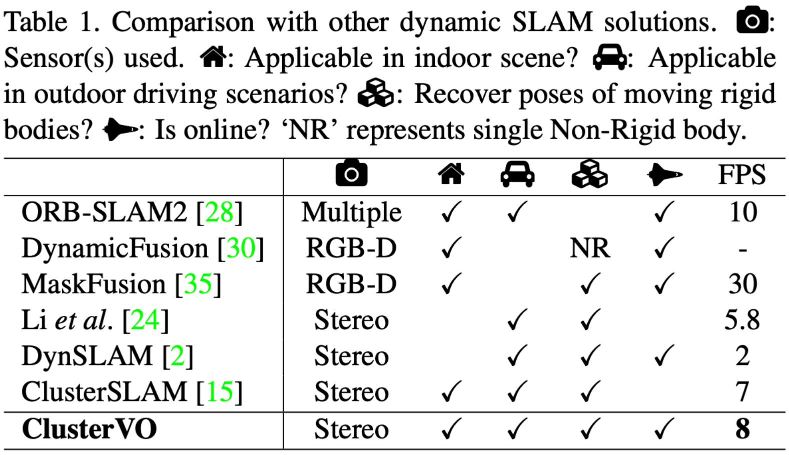


Table 1. Comparison with other dynamic SLAM solutions. 📷: Sensor(s) used. 🏠: Applicable in indoor scene? 🚗: Applicable in outdoor driving scenarios? 📦: Recover poses of moving rigid bodies? ✈: Is online? ‘NR’ represents single Non-Rigid body.

	📷	🏠	🚗	📦	✈	FPS
ORB-SLAM2 [28]	Multiple	✓	✓		✓	10
DynamicFusion [30]	RGB-D	✓		NR	✓	-
MaskFusion [35]	RGB-D	✓		✓	✓	30
Li *et al.* [24]	Stereo		✓	✓		5.8
DynSLAM [2]	Stereo		✓	✓	✓	2
ClusterSLAM [15]	Stereo	✓	✓	✓		7
ClusterVO	Stereo	✓	✓	✓	✓	**8**

2. Why do this?

① The existing dynamic SLAM/VO is based on a priori assumptions of the object or environment, which leads to its limited generalization ability.

② Is it possible to estimate and optimize the motion of both ego and surrounding rigid clusters/objects at the same time.

③ Data association problem between 2D geometric road signs, 2D semantic road signs, 3D point clouds and motion states.

3. How to do it?

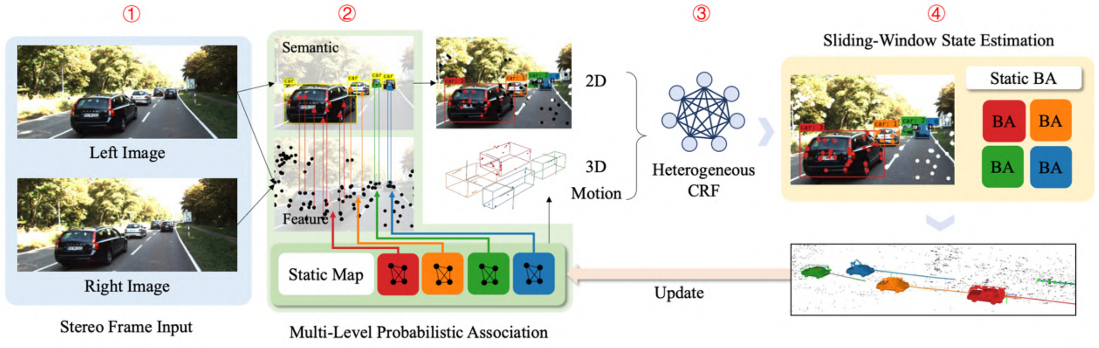
① For dynamic instances, a multi-level probabilistic data association technique is proposed to track both low-level features and high-level detections over time in the 3D space.

② Use heterogeneous CRF(Conditional Random Field) inference that combines semantics, space, and motion to refine existing clusters or create new clusters.

③ During optimization, marginalization error terms and motion prior error terms were added to the energy functions of static scenes and dynamic scenes.

④ A keyframe management method with dual channels in time and space is designed to ensure speed and efficiency, as well as motion prediction when occluded.

4. Pipeline of ClusterVO



ClusterVO takes synchronized and calibrated stereo frame as input, and outputs the camera and object pose for each frame. For each incoming frame, use YOLO to detect the semantic bounding box, and extract ORB features to match on the entire stereo image.

① Firstly, the detected bounding box and the extracted features are respectively associated with the previously discovered clusters and landmarks through a multi-level probability association formula.

② Then, we perform heterogeneous conditional random field (CRF) on all features with relevant map landmarks to determine the cluster segmentation of the current frame.

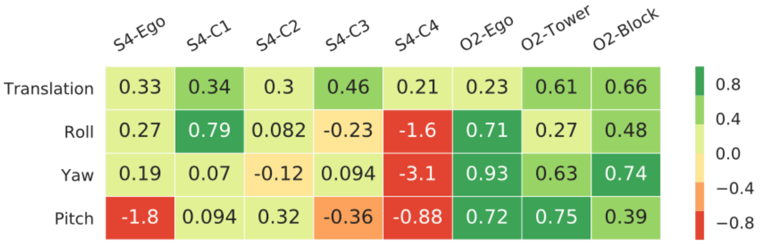
③ Finally, the state-estimation step optimizes all states on the sliding window through marginalization and smooth motion prior. The flow is shown in the figure below.

5. Experiments and results

① Indoor scene evaluations:

Dateset: Oxford Multimotion;

Heatmap of the ratio of decrease in error using ClusterVO for different trajectories and measurements.(可参考这样的误差可视化图)



	S4-Ego	S4-C1	S4-C2	S4-C3	S4-C4	O2-Ego	O2-Tower	O2-Block
Translation	0.33	0.34	0.3	0.46	0.21	0.23	0.61	0.66
Roll	0.27	0.79	0.082	-0.23	-1.6	0.71	0.27	0.48
Yaw	0.19	0.07	-0.12	0.094	-3.1	0.93	0.63	0.74
Pitch	-1.8	0.094	0.32	-0.36	-0.88	0.72	0.75	0.39

Figure 4. Performance comparison with MVO on S4 and O2 sequence in Oxford Multimotion [17] dataset. The numbers in the heatmap show the ratio of decrease in error using ClusterVO for different trajectories and measurements.

② Outdoor driving evaluations:

Dateset: KITTI

The effect is similar, the main advantage is the improvement in efficiency.

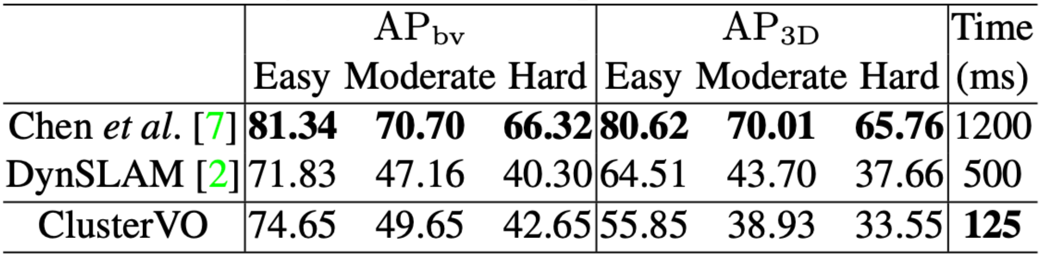
③ Ablation study: CRF sensitivity analysis

6. Some problems

① : In this paper,the 3D term can eliminate the divorced points of the 2D association, but it based on the assumption that the landmark points obey the Gaussian distribution.Is this assumption reasonable?

② : I cannot understand the comparison of the results in Table 3.

Table 3. 3D object detection comparison on KITTI dataset.



	AP_{bv}			AP_{3D}			Time
	Easy	Moderate	Hard	Easy	Moderate	Hard	(ms)
Chen *et al.* [7]	**81.34**	**70.70**	**66.32**	**80.62**	**70.01**	**65.76**	1200
DynSLAM [2]	71.83	47.16	40.30	64.51	43.70	37.66	500
ClusterVO	74.65	49.65	42.65	55.85	38.93	33.55	**125**

According to the paper, it does not seem to calculate the time for semantic detection?

③ : In the experiment in this article, only the data association of one type of object is involved. Is this still valid in the actual environment of multiple types of objects?

文献阅读 2

Some interesting facts about Paper:

Robust Method for Removing Dynamic Objects from Point Clouds

of **ICRA2020**：

1. Why do this?

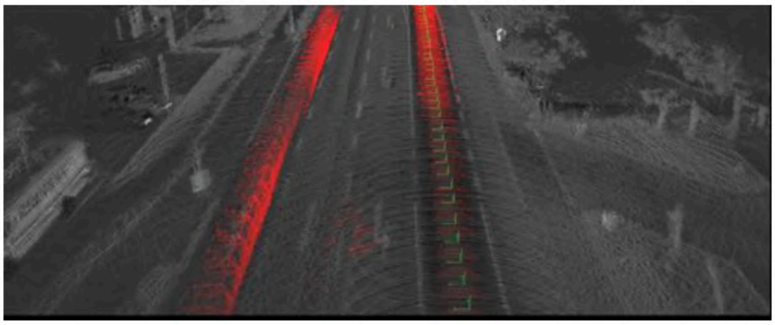


Fig. 1. Dynamic objects, highlighted in red, causing ghost trail effect in a point cloud map

① : Laser scans represent a snapshot of the surrounding at the time of capture, hence they often contain moving objects which may not be observed at all times. Dynamic objects in point cloud maps decrease the quality of maps and affect localization accuracy.

② How to quickly build a high-precision map?

2. How to do it?

① : Build a long-term static map under the Octomap framework, that is, a map after excluding dynamic objects.

② : Use neural networks to identify dynamic objects in the point cloud, and use this information to accelerate the process of building static maps.

③ : Update the map by generating artificial endpoints.

3. Related Work:

① : Model-free method: it is based on comparing the current laser beam with the historical laser beam (or the laser beam generated in the future) to detect dynamic obstacles. The main problem faced by this type of method is that the dynamic obstacles that can be detected must move in the voxel, and there can be no pauses (or excessively long pauses?). For example, when you encounter a red light, cars that have been parked together for a while will be regarded as static objects.

② : Model-based method: Use neural network to recognize dynamic obstacles in the point cloud, generate bounding-box, and find these laser endpoints belonging to the dynamic category.

③ : Map filter method: Generate a static map and filter the point cloud. Anything that does not fit the static map is a dynamic obstacle. The popular 3D map representation is Octomap based on octree, but Octomap is very sensitive to updates and is not suitable for building static maps.

4. Pipeline

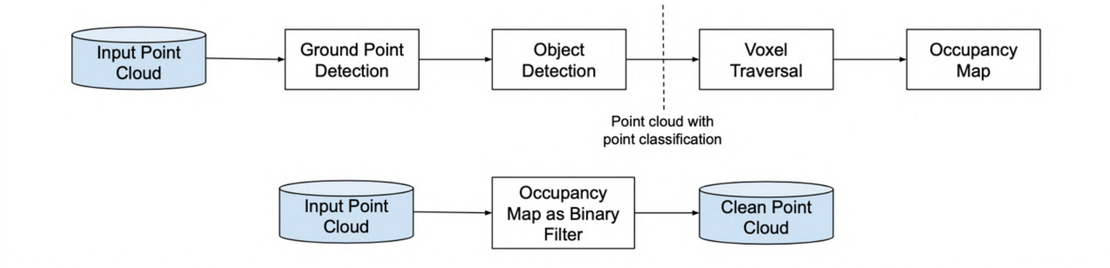


Fig. 2. The overall pipeline of our system. (top) Input Point Cloud goes through ground point detection and object detection. Processed point cloud contains points classified as ground, object and free. Voxel traversal is done and an occupancy map is built. (bottom) Octree occupancy filter is applied to input point cloud map to get clean point cloud map.

① : Use the AVOD network to perceive the point cloud and generate the bounding box of the dynamic object, and then you can divide which points belong to the dynamic object and which points belong to the static background.

② : After dividing the two types of points, use Voxel travelsal to traverse all Voxel on the path. For the end point of the dynamic point, it is updated to occupy, for the end point of the static point, it is updated to the static point, and the point on the path is considered free.

③ : Map update method: The main body is based on Octomap's log-odds update method, with clamping policy (that is, the threshold at both ends), but a free counter is set on each voxel. When voxel is regarded as free, it increments the counter by 1, and when it is not regarded as free, it is reduced by 1. The reason for this is to reduce the sensitivity of static maps to dynamic obstacles. If a certain voxel is free for a long period of time, then this n is very large, and it will be difficult to update it later.

5. Experiments

   Dateset:KITTI

   Experiments were conducted on wide roads, narrow city roads and highways respectively. It assesses the performance of approach statistically by using precision and recall. But from the results, this method does not seem to be very competitive.

6. Some problems

   ① : There are too many hyperparameters in the algorithm that need to be manually set, which may weaken the generalization ability of the method.

   ② : Unable to understand the update of free counter.

   ③ :The result of the experiment is not obvious, and there are few comparative experiments.

   ④ : Octomap may be a method to learn from, but I prefer model-based learning methods.