

DATMO (Detection and tracking of moving objects) 相关总结

移动机器人组 尹欢 2017.05.17

Publications:

1. Petrovskaya A, Perrollaz M, Oliveira L, et al. **Awareness of Road Scene Participants for Autonomous Driving**[M]. Springer London, 2012.
2. Sivaraman S, Trivedi M M. **Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis**[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(4):1773-1795.

Thesis:

GARCIA R O C. **Multiple Sensor Fusion for Detection, Classification and Tracking of Moving Objects in Driving Environments**[D]. Université de Technologie de Belfort-Montbéliard, 2014.

一. DATMO 的研究意义与背景

移动机器人在实际运行中要解决三个问题:

- a. 我在哪里 (SLAM / Localization)
- b. 我周围是什么环境 (Localization / DATMO)
- c. 我如何去目的地 (Path Planning / Navigation / Control)

其中, DATMO 的研究是一个重要环节。DATMO 的概念是随无人驾驶以及移动机器人的崛起而提出的。它所解决的是在一个动态的环境, 机器人根据自身携带的传感器(相机、激光雷达等)来检测并跟踪移动的障碍物(车辆、行人等), 并预测它们的速度和位置, 为轨迹规划提供充足的信息。

根据文献 1 的分类, DATMO 的应用场景分为三类:

- a. Pedestrian Zone. 这种场景主要是行人、自行车以及宠物等等, 动态障碍物的速度是 1-2m/s, 检测与跟踪的区域大概是 10-20m 范围内, 从感知-规划-控制的频率大概在 1Hz 左右即可。
- b. Freeway driving. 可以理解为充斥着车辆的单向环境, 如高速公路。该环境的主要问题是车辆的速度很快, 需要更广的探测距离 >100m, 更高频

率的感知-规划-控制>10Hz。

c. Mixed urban driving. 介于前两者的环境，例如校园环境，也更为复杂。

二. DATMO 问题的贝叶斯描述

我们规定系统的状态(DATMO 中为目标的状态)为 S ，观测状态为 Z ，DATMO 的问题是求的后验概率 $bel := p(S | Z)$ ，这个问题和自定位问题类似，只不过所要求的系统状态不是机器人的定位，而是机器人周围障碍物的位置以及速度。

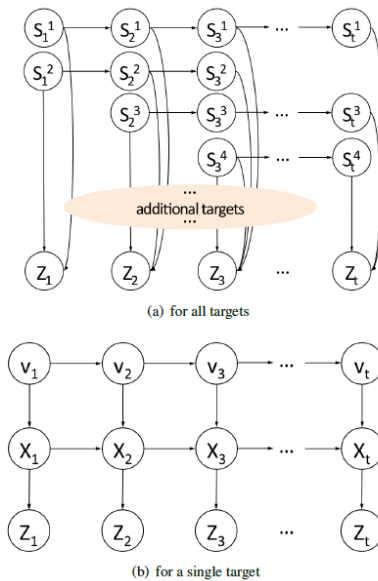
假设系统包含 K_t 个动态障碍物 $T^1, ..., T^{K_t}$ ，其中障碍物的数量 K_t 是随着时间 t 变化的。在二维环境下每个障碍物的位姿 $X_t = (x_t, y_t, \theta_t)$ 和速度 v_t 组成了系统状态 $S_t := (X_t, v_t)$ ，有些文章把物体的形状、类别等等也加入了系统状态描述。整个待估计的贝叶斯量为： $S_t = (S_t^1, ..., S_t^{K_t})$ 。

我们有从零时刻起的所有观测 $bel_t := p(S_t | Z_1, ..., Z_t)$ 以及动态模型 $p(S_t | S_{t-1})$ ，根据马尔科夫过程（实质上是递归）以及贝叶斯定理，可知：

$$bel_t = \eta p(Z_t | S_t) \int p(S_t | S_{t-1}) bel_{t-1} dS_{t-1}$$

对于单个障碍物，一般有三个模型来描述：速度模型 $p(v_t | v_{t-1})$ ，运动模型 $p(X_t | X_{t-1}, v_t)$ 以及观测模型 $p(Z_t | X_t)$ 。

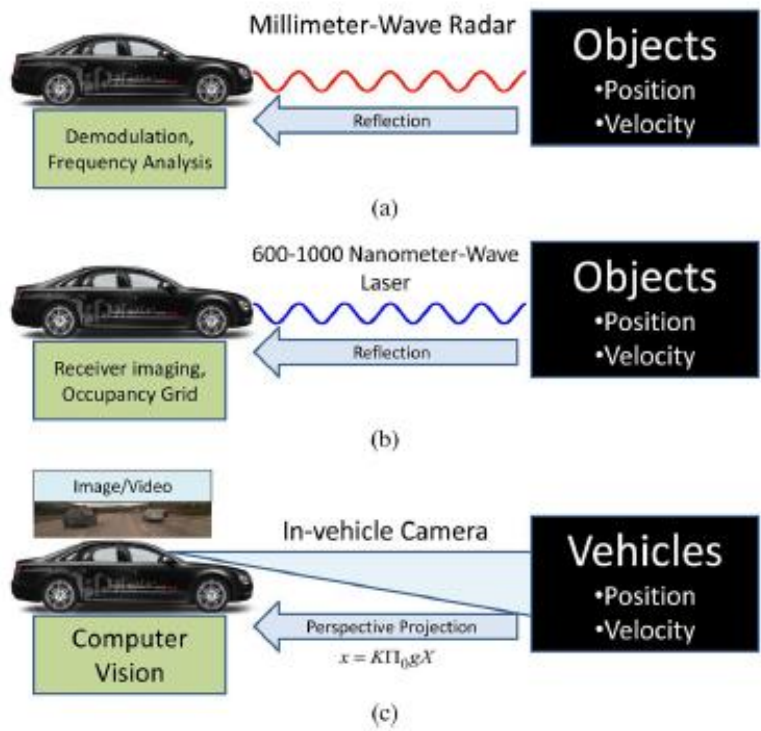
下面的两张图分别描述了整个 DATMO 模型以及单个障碍物的模型：



三. 传感器

障碍物的检测依赖于移动机器人搭载的传感器，一般有三类传感器可供选择：相机、激光雷达、（毫米波）雷达。

如下图所示：



各个传感器之间的简单比较如下表所示：

模式	观测数据	描述单位	价格（相对）	鲁棒性	探测角度
雷达	距离	米	差异较大	鲁棒	10-70°
激光	距离， 反射率	米	单线： 较便宜 多线： 昂贵	极端天气不 鲁棒	90- 360°
视觉	RGB	像素	便宜 但全景相机 略贵	夜间不鲁棒	~30°

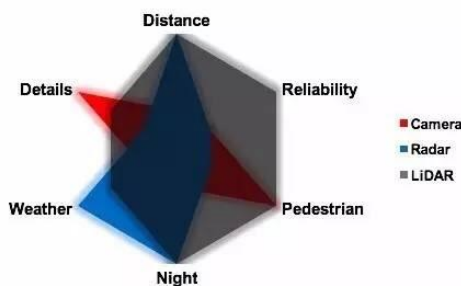
其他的说明如下：

- a. 相机。图像作为视觉的载体提供了丰富的 RGB 信息，因此强大的计算机

视觉技术可以检测出车辆和行人等等，但是缺乏深度信息；少部分人尝试用双目恢复深度来估计障碍物状态，但不是主流。此外，视觉信息处理会占用大量的计算资源，因此一般会分配 GPU 来处理。

- b. 激光。对于检测障碍物，以前是通常用平行于地面的 2D 激光；08 年之后多线激光逐渐崛起，Velodyne-64 线激光一秒钟可以在 360 度范围内获取百万多个点。最近几年也出现了基于 learning 的方法在激光点云上检测车辆与行人。
- c. 毫米波雷达。探测的范围较为狭窄，一般用于机器人的前部，探测距离与激光传感器差不多，但是在距离深度上误差较大。目前常用的品牌是 delphi ESR，国内售价 3.3 万左右。

下面这张图对比了三种传感器：



四. 动态障碍物的建模

1. 动态障碍物的运动模型

a. 布朗（Brownian）运动模型

一般用于行人的运动建模，因为人的运动没有约束，是预测不准的。在此模型下，位姿加入了零均值的高斯噪声，方差也随时间增大而增大。

b. 线性运动模型

一般用于车辆的运动模型，因为汽车的运动是不完整的。在一段时间内，速度的值是不变的，而朝向则根据最大转向 $d\theta_{\max}$ 来确定。

c. 自行车运动模型

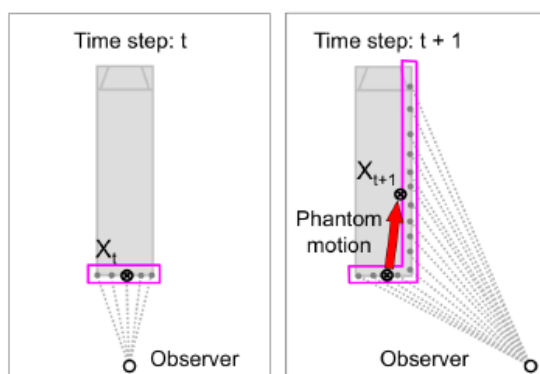
另一种车辆运动模型，在线性速度上附加了角速度，并且假设车辆是一直延弧线运动。相关公式在《概率机器人》上有解释。

2. 观测模型建立

a. 基于聚类方法

基于聚类的方法来估计物体的质心比较容易实现，但是会有一个问题。如果对障碍物的观测不是完整的，那么估计的质心会产生偏移，从而对障碍物的速度估计产生影响。一个静止的车辆会被估计出一个速度。

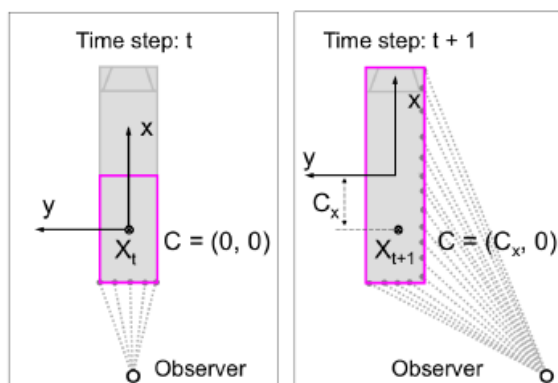
如下图所示：



b. 基于几何的模型

基于几何模型的方法是将利用一个 **Model** 来描述物体的几何形状，对于车辆一般是矩形或者长方体，对于行人是圆柱体或多变形。

为了避免出现聚类方法时质心偏移的问题，会在障碍物的状态描述中引入“锚点”的概念。在第一次观测到障碍物的时候，锚点即为质心，在障碍物坐标系下为 $C = (0,0)$ ，随着对同一个障碍物的进一步观测，质心发生了偏移，锚点的局部坐标也随之变化，如下图的 $C = (C_x, 0)$ 。而障碍物的状态变量中含有锚点，因此对速度的估计不会有影响。



c. 基于栅格的模型

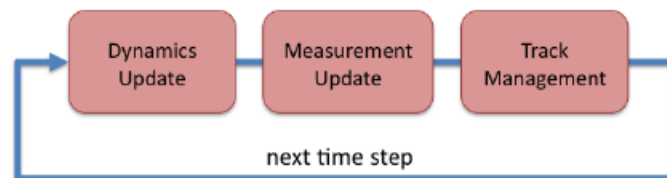
导航和 SLAM 中常用栅格地图，因此用栅格占用来描述障碍物信息是一

种更高效更简洁的方法。

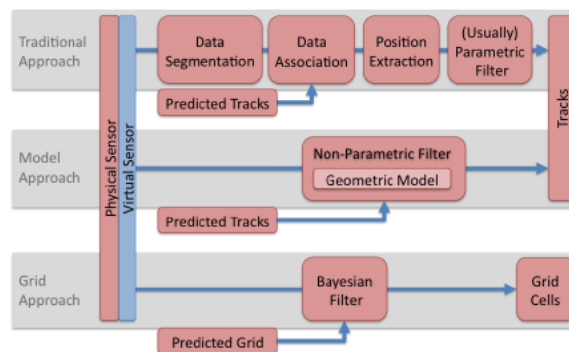
在传感器信息融合上，栅格的表示方法也更为有效；作为低级的描述方式，这种模型中没有单个物体的概念，也没有 tracking 的概念。

五. 三种基本的 DATMO 形式

经典的 DATMO 流程如下：



和机器人自定位流程类似，在相关综述中又将 DATMO 分为以下三类，我个人认为这种分类不是很合理，如图所示：



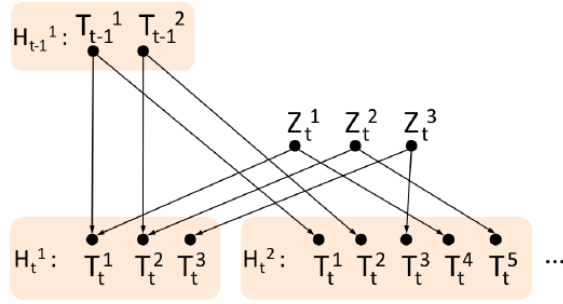
前两种实际上是用参数化滤波（KF/EKF/UKF）以及非参数滤波（PF/HF）解决问题。参数化滤波计算量相对较小，（pseudo）???

1. 经典 DATMO

无论是基于数据分割（聚类）还是其他的模式识别方式，原始数据都被分为若干个抽象的物体点。在相邻两帧之间要判断障碍物的 Data Association。

最基本的 association 方法是基于 NN(Nearest neighbor)，比如 GNN；或者基于 Track-tree。在多目标跟踪任务时更高级的两种常用算法是 MHT(multiple hypothesis tracking)和 JPDAF(joint probabilistic data association filter)。

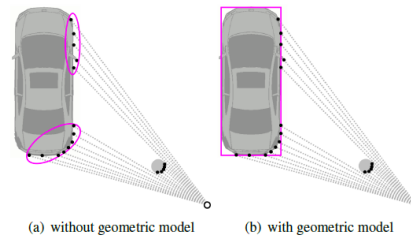
MHT 算法的思想是分析两帧之间的各种可能性，再对各种可能性打分。如下图所示， H_{t-1}^1 到下一时刻得到了三个观测值，因此有很多可能性，图中显示了 H_t^1 和 H_t^2 两种可能性。对这些可能性计算 SPRT(sequential probability ratio test)，取值最大的为结果。



JPDAF 算法不需要多种假设,要跟踪的目标数量是固定的;但也不像 GNN 那样多帧之间目标的连接是强连接,只是用概率的软连接方式。最后以联合概率的方式表示 data association 结果。JPDAF 的变种方法有很多,就不一一阐述了。

2. 基于模型的 DATMO

基于模型的 DATMO 不太关注分割的障碍物之间的 association,因为几何模型可以把零碎的障碍物连接起来,如下图所示:



基于模型的 DATMO 假设 data association 是正确的,更关心的是用非参数滤波的方法来估计位姿,比如粒子滤波系列 RB-PF,以及 MCMC 等等。

3. 基于栅格的 DATMO

如果一个物体检测错误,或者 tracking 错误,那么对 DATMO 的影响很大。因此基于栅格的障碍物表示方法可以降低此类错误的影响。

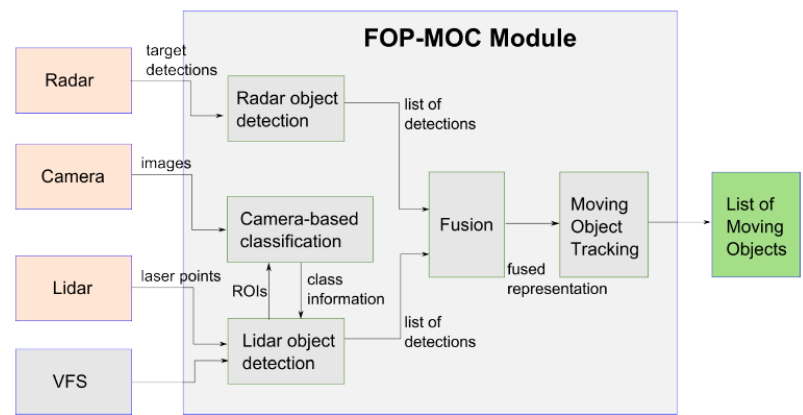
第一个在 2D 栅格上做 DATMO 的方法是 BOF(Bayesian Occupancy Filter)。该方法中每一个栅格都有两个概率分布来描述,分别是障碍物占有率以及速度分布。当障碍物很多时,这种方法计算量较大。另一种普遍的算法是 FCTA(Fast Clustering and Tracking Algorithm),也是基于 2D 栅格,但是是基于 Object Level。

六. 传感器融合技术

最近几年,在 DATMO 领域,传感器融合的技术开始兴起。在无人驾驶技术中 ADAS(Advanced Driver Assistant Systems)以及 ACC(Adaptive Cruise

Control) 采用了视觉-激光-雷达或者视觉-雷达等方案。

一种多传感器融合的例子如下所示：

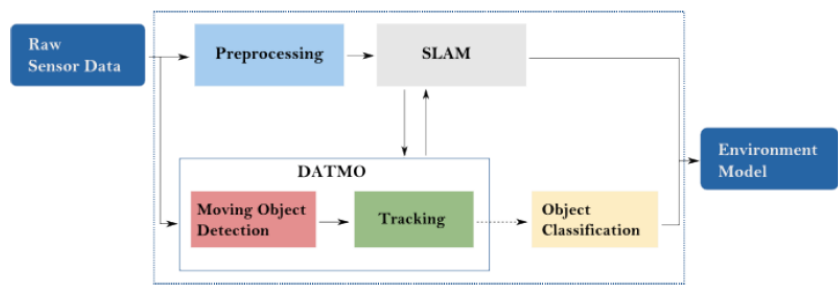


激光和视觉都有自己的单独的检测方案。激光的优势是在 **range** 上精准，但是在类别检测上不足；视觉方案则相反。而雷达的技术更成熟，可以提供障碍物相对于车辆的相对速度，有些毫米波雷达可以提供物体的类别。

在融合方面，一般是将检测结果进行融合（滤波），融合的层面可以在 **2D** 栅格上，或者在障碍物的状态空间上。方法一般是基于贝叶斯滤波。

七. **DATMO 与 SLAM**

作为移动机器人的感知端的两个模块，**DATMO** 和 **SLAM** 承担的作用不同，二者的结果相互影响，一种移动机器人的感知结构如下图所示：



SLAM 的定位精度以及地图构建的效果很容易受移动障碍物的影响，如果通过 **DATMO** 去除这些影响，那么会提高定位精度；而 **DATMO** 对障碍物在世界坐标系下的位姿估计又依赖于机器人自身的位姿估计。二者共享同样的传感器数据，相辅相成。