**第二章 自动驾驶汽车定位**

简介

对于自动驾驶汽车来说，最关键的任务之一就是定位，准确、实时地确定装置的位置。在本章，我们首先研究不同的定位技术，包括GNSS，激光雷达（LiDAR）和高精度地图，视觉里程测量，以及其他航迹推算传感器。我们还研究了应用传感器融合技术来整合多个传感器以提供更精确定位的几个实际应用案例。

2.1 基于GNSS的定位技术

当人类驾驶汽车时，我们通常依靠全球导航卫星系统(GNSS)进行定位。当涉及到自主车辆定位时，我们也从GNSS开始。在本节中，我们将深入研究GNSS技术的细节，并了解应用于自动驾驶时GNSS的优缺点。

2.1.1 GNSS概述

GNSS由几个卫星系统组成:GPS、GLONASS、Galileo和北斗。在这里，我们以GPS为例来提供一个关于GNSS的概述。GPS提供可在GPS接收机中处理的编码卫星信号，允许接收机估计位置、速度和时间[1]。要做到这一点，GPS需要四个卫星信号来计算三维位置和接收机时钟的时间偏移。这些全球定位系统卫星的部署分散在6个轨道平面上，在距地球表面约20 200公里的近乎圆形轨道上，相对赤道倾斜55°，轨道周期约11小时58分钟。

卫星上产生的信号是由基频fo=10.23 MHz[1]产生的。这个信号是用原子钟来计时的，每天误差只有10-13秒。L波段的两个载波信号分别为L1和L2。通过fo的整数乘法生成。载波L1和L2被编码调制成双相位，为接收器提供卫星时钟读数，并传送轨道参数等信息。这些代码由一个状态为+1或-1的序列组成，对应于二进制值0或1。双相位调制是通过载波相位的180°位移来实现的，无论何时代码状态发生变化。卫星信号包含卫星轨道、轨道摄动、GPS时间、卫星时钟、电离层参数、系统状态信息等。导航信息由25个帧组成，每个帧包含1500位，每个帧又被转换成包含300位的5个副帧。

GNSS系统的下一个关键部分是参考坐标系的定义，它对于卫星运动的描述、观测数据的建模和结果的交互都是至关重要的。GNSS的工作需要两个参考系统:(a)用于描述卫星运动的空间固定惯性参考系统; (b)地球固定的地面参考系统，用以确定观测站的位置和说明卫星测地学的结果。利用这两个系统，利用已知的空间固定和地球固定的转换参数，直接在GNSS接收机和后处理软件中计算地球固定系统中接收机的位置。地球参考系按惯例定义为三轴，其中z轴与地球旋转轴重合，这是按惯例国际原点确定的。X轴与格林威治子午线相联系，y轴与Z轴和X轴正交，构成右手坐标系。GPS以WGS84为基准系统，与WGS84相关联的是一个旋转[2]的地心等势椭球体。

近年来，随着近十年来全球卫星定位系统(GNSS)卫星数量的不断增加，支持多个星座的GNSS接收机的出现也保持稳定。随着新GNSS星座的提升，几乎所有的新设备都被期望支持多个星座。支持多个星座的好处包括增加利用能力，特别是在有阴影的地区;提高精度，更多的卫星在视野中提高精度;提高了鲁棒性，因为独立系统更难被欺骗。

2.1.1 GNSS误差分析

理想情况下，使用GNSS，我们可以得到完美的定位结果，而不会有任何误差。然而，在GNSS中有多个地方可以引入错误。在本小节中，我们将回顾这些可能的误差贡献者。

**卫星时钟：**GNSS卫星上原子钟的任何微小误差都会导致接收器计算出的位置出现重大误差。大约10纳秒的时钟误差会导致3米的位置误差。

**轨道误差:** GNSS卫星在非常精确、众所周知的轨道上运行。然而，就像卫星时钟一样，它们的轨道也会有一点变化。当卫星轨道改变时，地面控制系统向卫星发送校正信号，对卫星星历表进行更新。即使有GNSS地面控制系统的校正，在轨道中仍然有小的误差，可能导致高达±2.5 m的位置误差。

**电离层延迟:** 电离层是离地80 ～ 600公里的大气层。这一层含有带电粒子，称为离子。这些离子会延迟卫星信号，并可能导致大量的卫星位置误差(通常为±5m)。电离层延迟随太阳活动、一年中的时间、季节、一天中的时间和位置而变化。这使得很难预测电离层延迟对计算位置的影响程度。电离层的延迟也根据通过电离层的信号的无线电频率而变化。

**对流层延迟:** 对流层是离地球表面最近的大气层。对流层延迟的变化是由对流层中湿度、温度和气压的变化引起的。由于对流层环境在一个局域内非常相似，所以基站和漫游接收器所经历的对流层延迟非常相似。这使得RTK GNSS可以补偿对流层延迟，下一小节将对此进行讨论。

**多路径:** 当GNSS信号从物体(如建筑物的墙壁)反射到GNSS天线时，就会发生多路径。由于反射回来的信号传到天线的距离很远，所以反射回来的信号到达接收器的时间稍微有些延迟。这个延迟信号会导致接收器计算错误的位置。

我们在图2.1中总结了这些贡献源的误差范围。要更详细地讨论这些错误，请参考[3,4,5,6]。

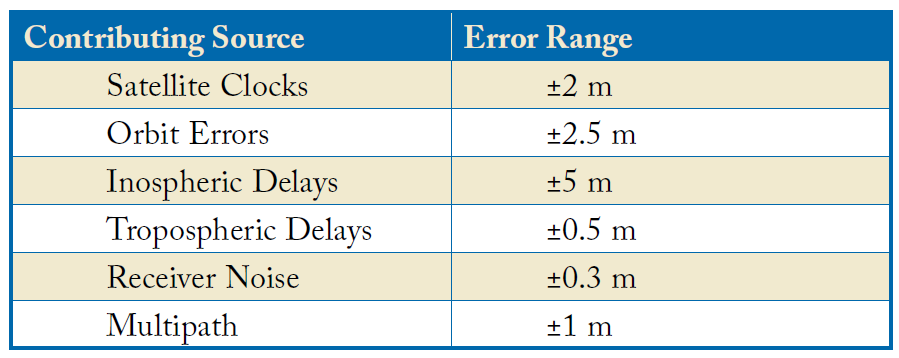


图2.1 GNSS系统误差

2.1.3 基于卫星的增强系统

卫星增强系统(SBAS)补偿现有的GNSS，以减少测量误差。SBAS弥补了GNSS在准确性、完整性、连续性和可用性方面的某些缺陷。SBAS的概念是基于遍布整个大陆的精确定位参考站的GNSS测量。然后将GNSS误差传输到计算中心，以计算差分校正和完整性信息，然后使用地球同步卫星在大陆上广播，作为原始GNSS信息的增强或覆盖。SBAS信息通过能够覆盖广大地区的地球同步卫星广播。

许多国家已经实施了自己的基于卫星的增强系统。欧洲有欧洲地球同步导航覆盖服务(EGNOS)，主要覆盖欧洲。美国有其广域扩增系统(WAAS)。中国已经启动北斗系统(BDS)，提供自己的SBAS实施。日本拥有多功能卫星增强系统(MSAS)。印度推出了自己的SBAS程序，命名为GPS和地理增强导航(GAGAN)，覆盖印度次大陆。所有的系统都符合一个通用的全球标准，因此都是兼容的，不可互操作的。

值得注意的是，大多数商业GNSS接收器都提供SBAS功能。详细地说，WAAS规范要求它为横向和纵向测量提供7.6 m或更少的位置精度，至少95%的时间。在特定位置对该系统的实际性能测量表明，该系统的横向好于1.0米，纵向好于1.5米。

2.1.4 实时运动和差分GPS

根据我们的经验，大多数商用的多星座GNSS系统提供的定位精度不超过2米半径。虽然这对人类司机来说已经足够了，但为了让自动驾驶汽车沿着道路行驶，它需要知道道路的位置。为了保持在特定的车道上，它需要知道车道的位置。自动驾驶汽车在车道上行驶时，其定位要求按分米计算。幸运的是，实时运动学(RTK)和差分GNSS确实提供分米级的定位精度。在本小节中，我们将研究RTK和差分GNSS如何工作。

RTK GNSS通过减少卫星时钟、不完全轨道、非球面延迟和三次方波延迟中的误差来达到高精度。图2.2展示了RTK GNSS的基本概念，纠正这些GNSS误差的一个好方法是在一个位置已知的基站上设置一个GNSS接收机。基站接收器根据卫星数据计算其位置，并将其与实际已知位置进行比较，然后识别出两者之间的差异。由此产生的误差修正可以从基地传达到车辆。

具体来说，RTK使用基于载波的测距，并提供比基于代码的定位精确数量级的范围(以及位置)。基于代码的定位是一种利用卫星伪随机码中包含的信息，通过粗采集码进行数据采集的处理技术。经过微分校正，这种处理技术的结果是5米的精度。载波测距是另一种通过载波相位接收器采集数据的处理技术，利用载波信号计算位置。载波信号具有比伪随机码高得多的频率，比单纯使用伪随机码具有更高的精度。伪随机码缩小参考量，载波码在差分校正后进一步缩小参考量，从而提高亚米精度。在基于carrier的测距中，通过确定卫星和飞行器之间的载波周期数，然后用这个数乘以载波波长来计算距离。计算范围仍然包括来自卫星时钟和星历表、电离层和对流层延迟等来源的误差。为了消除这些误差并利用基于carri er的测量精度，RTK性能要求测量数据从基站传输到车辆上。

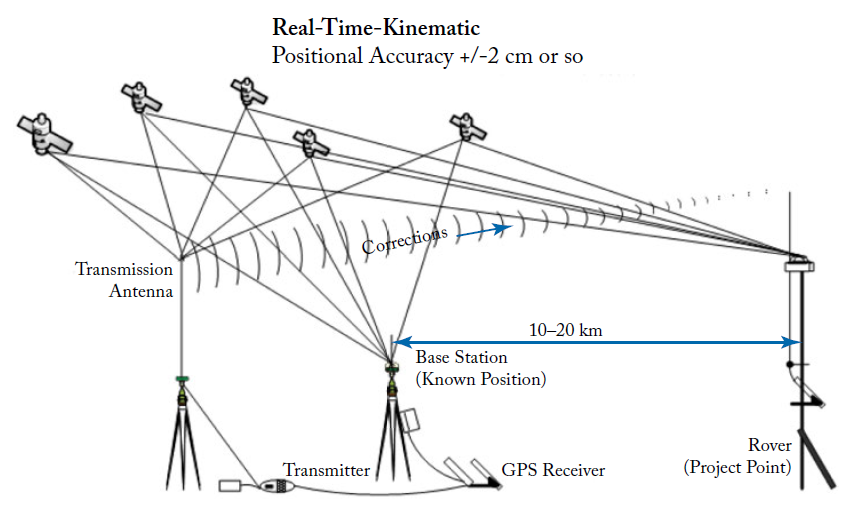


图2.2 RTK GNSS

在RTK GNSS系统中，车辆通过融合模糊度分辨和差分校正的算法来确定自己的位置。车辆所能达到的位置精度取决于它到基站的距离和差分校正的精度。修正的精度与已知的基站位置和基站卫星观测的质量相同。因此，选址对于尽量减少干扰和多路径径等环境影响至关重要，基站和车载接收器和天线的质量也是如此。

2.1.5 精确的单点定位

虽然RTK GNSS系统提供了满足自动驾驶要求所需的亚分米级精度，但这种解决方案通常需要用户部署自己的基站，维护成本较高。在本小节中，我们将研究精确点定位(PPP) GNSS系统如何帮助缓解这一问题[7,8]。

图2.3显示了PPP GNSS方案如何工作。全世界部署了许多参考站，这些参考站实时接收精确的参考卫星轨道和参考GNSS卫星时钟。这些参考站然后计算应用于卫星定位结果的校正。一旦修正被计算出来，它们将通过卫星或互联网传递给最终用户。精确的卫星位置和时钟使卫星时钟误差和轨道误差最小化。然后，我们可以应用一个双频GNSS接收器来消除电离层的一阶效应，该效应与载波频率成正比。因此，采用双频GNSS测量组合可以完全消除一阶导数球延迟。此外，利用UNB模型对对流层延迟进行了修正[9]:为进一步提高精度，在估算位置和其他未知数[10]时，对剩余的对流层延迟进行了估计。通过结合这些技术，PPP能够提供分米到厘米级别的位置解决方案。

具体来说，PPP算法使用双频段接收机的输入码和相位观测，以及精确的卫星轨道和时钟，以计算精确的接收机坐标和时钟。来自所有卫星的观测数据都在一个过滤器中一起处理，比如扩展卡尔曼滤波器(EKF)。估计了位置、接收机时钟误差、对流层延迟和载波相位不确定性。EKF最大限度地降低了系统中的噪声，并能够利用厘米级精度来估计位置。通过连续的GNSS测量来改进EKF状态的估计，直到它们收敛到稳定而准确的值。

PPP与RTK定位的不同之处在于它不需要访问一个或多个临近基站，而且PPP提供的是绝对定位，而不是RTK提供的相对于基站的位置。PPP只需要精确的轨道和时钟数据，这些数据由地面处理中心通过相对稀疏的台站网络的参考站测量数据计算出来。请注意，PPP只涉及一个GPS接收器，因此，在用户附近不需要参考站。因此，PPP可以被视为一种全局位置方法，因为它的位置解引用了一个全局参考框架。因此，PPP提供了比RTK方法更大的定位一致性，在RTK方法中，位置解决方案相对于本地基站或多个基站。此外，PPP在结构上类似于SBAS系统。与SBAS相比，PPP提供的关键优势在于，它要求实时提供精确的参考GNSS轨道和时钟，从而达到厘米级精度，而SBAS只提供米级精度。此外，PPP系统允许在全球范围内使用单一校正流，而SBAS系统则是区域性的。

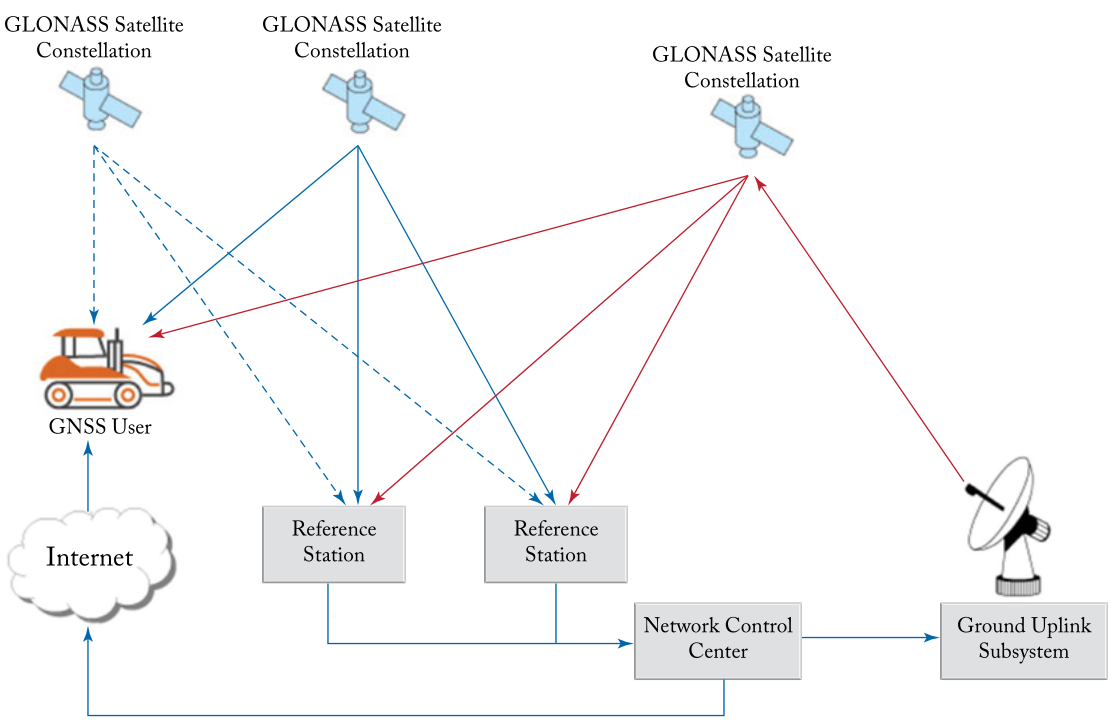


图2.3:PPP GNSS。由NovAtel公司提供，经允许使用。

PPP解决方案面临的一个主要问题是，为了解决任何局部偏差，如大气条件、多路径环境和卫星几何形状，通常需要很长一段时间(30分钟)才能收敛到分米精度。目前，有几个合并的后处理PPP服务。相反，实时PPP系统正处于发展初期[11,12,13]。

2.1.6 GNSS与惯性导航的集成

在前面的小节中，我们研究了不同年代的GNSS技术，在本小节中，我们将探索如何利用惯性数据来改进GNSS定位方法[14]。惯性导航系统(INS)使用来自惯性测量单元(IMU)的旋转和加速度信息来计算相对位置随时间的变化。一个典型的六轴IMU是由六个互补的传感器排列在三个正交轴上。在每一个轴上上都有一个加速度计和一个陀螺仪。加速度计测量线性加速度，陀螺仪测量旋转加速度。有了这些传感器，IMU就可以测量其在3D空间的精确相对运动。惯性导航系统使用这些测量数据来计算位置和速度。此外，IMU测量提供了关于三个轴的角速度，这些角速度可用于推断局部姿态(翻滚、俯仰和方位角)。

通常，INS系统以1kHz的频率运行，提供非常频繁的位置更新。然而，INS系统也有几个缺点:首先，INS只从初始点提供一个相对的解决方案。这个初始起点必须提供给INS。其次，更重要的是，使用IMU在三维空间中导航实际上是每秒数百/数千个样本的总和(或积分)，在此期间错误也在累积。这意味着，如果没有外部参考来校正，未校正的INS系统将很快偏离真实位置。因此，在使用INS执行定位任务时，必须为INS提供准确的外部参考，允许INS使用数学滤波器(如卡尔曼滤波器)最小化定位误差。

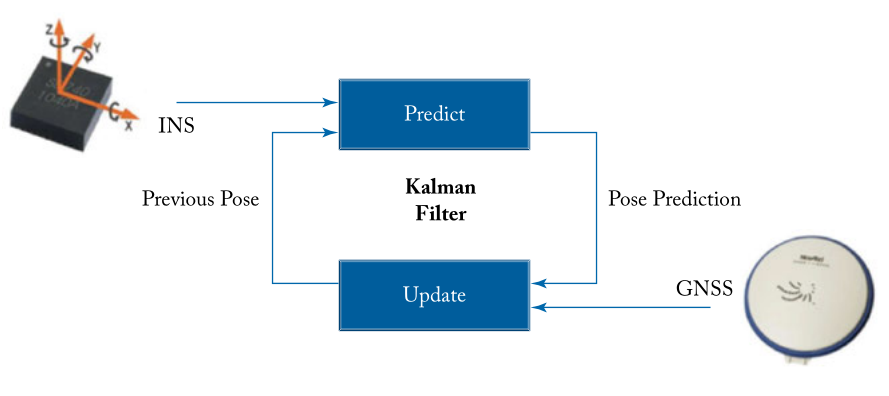


图2.4 使用卡尔曼滤波进行GNSS与INS融合

如图2.4所示，该外部参考可以由GNSS提供。GNSS提供了一组可以用作初始起点的绝对坐标。此外，GNSS提供了连续的位置和速度，以更新INS滤波估计。当GNSS信号因信号障碍而受损时(如在隧道中行驶时)，INS系统可在短时间内定位车辆。

2.2 使用激光雷达和高精度地图进行定位

包括Waymo、百度、BMW等在内的大多数商用无人驾驶汽车原型，都依赖于激光雷达和高清地图进行定位。在本节中，我们将研究激光雷达和高清地图如何工作，以及如何将两者结合起来为自动驾驶车辆提供准确的定位。

2.2.1 激光雷达概述

在本小节中，我们提供了激光雷达技术的概述。激光雷达是光探测和测距的缩写，它通过用脉冲激光照射目标，然后用传感器[15]测量反射的脉冲来测量到目标的距离。激光返回时间和波长的差异可以用来制作目标的数字3d表示。激光雷达的基本工作原理如下:首先，激光雷达仪器在一个表面发射快速的激光脉冲，有些脉冲高达每秒15万次。然后，仪器上的传感器测量每个脉冲反弹所需的时间。当光以恒定和已知的速度移动时，激光雷达仪器可以高精度地计算出自身与目标之间的距离。通过快速连续地重复这一过程，该仪器构建了一个复杂的它所测量表面的“地图”。

通常有两种激光雷达检测方法:非相干检测(也称为直接能量检测)和相干检测[16]。相干系统最适合多普勒或相位敏感测量，通常使用光学外差检测，一种提取编码为电磁辐射相位或频率调制的信息的方法。这使它们能够以低得多的功率运行，但代价是需要更复杂的收发器。具体来说，当非相干光发射时，它会向各个方向传播。相反，相干光使用高度专门化的二极管，这种二极管在电磁波谱的光学部分或附近产生能量，这意味着所有的单个能量波都在同一个方向上运动，从而大大降低了功耗。

在相干型和非相干型激光雷达中，存在两种主要的脉冲模型:高能脉冲系统和微脉冲系统。高能源系统发射大功率光,可以对人眼有害,这些系统常用到大气研究,他们通常用于测量各种大气参数如高度、分层和密度的云,云粒子特性,温度、压力、风、湿度、和示踪气体浓度。相反，微脉冲系统的功率较低，属于眼睛安全级别，因此在使用时几乎没有什么安全防范措施。在其原始设计[17]中，微脉冲激光雷达发射机是一个二极管泵浦的微焦耳脉冲能量，高重复率激光器。眼睛的安全是通过光束扩展来实现的。接收器使用光子计数固态盖革模式雪崩光电二极管探测器。在自动驾驶中使用的激光雷达设备大多是一致的微脉冲系统，这样它们就满足了所有类别中最安全的一类激光安全要求。

激光雷达中使用的激光可按其波长分类。600-1000纳米激光器是最常用的，通常其最大功率限制在满足一类要求。波长为1550纳米的激光器也被广泛使用，因为它们可以用于更大的范围和更低的精度。此外，1550nm波长的激光在夜视护目镜下不显示，因此非常适合军事应用。机载激光雷达系统使用1064纳米二极管泵浦YAG激光，而测深系统使用532纳米双二极管泵浦YAG激光，这种激光穿透水中的衰减比机载1064纳米激光要小得多。如果接收检波器和电子器件有足够的带宽来处理不断增加的数据流量，那么较短的脉冲可以获得更好的分辨率。

一个典型的激光雷达系统由两个主要部件组成，激光扫描仪和激光接收器。图像生成的速度受到扫描到系统中的速度的影响。各种各样的扫描方法可用于不同的目的，如方位角和仰角，双振荡平面镜，双轴扫描仪，和多面镜。它们的光学类型决定了系统可以探测到的分辨率和范围[18,19]。激光接收器读取并记录返回到系统的信号。有两种主要类型的激光接收技术，硅雪崩光电二极管和光电倍增管[20]。

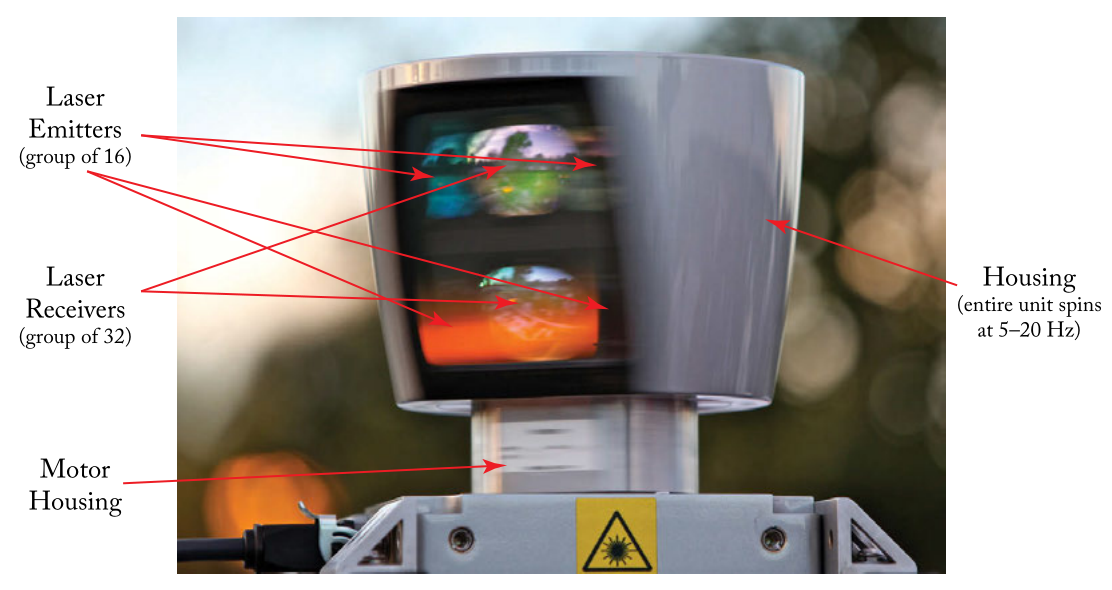


图2.5 Velodyne 64线激光雷达

图2.5显示了一个Velodyne HDL-64激光雷达，它被广泛应用于自动车辆。它利用64个从+2.0°到-24.9°排列的激光雷达通道，提供26.9°的垂直视场，并通过其旋转头的设计提供实时360°水平视场。旋转速率可由用户选择，范围为5-20 Hz，以便用户确定由激光雷达传感器产生的数据点的密度。HDL-64设备产生的激光波长为905 nm，脉冲为5 ns，可捕获高达220万点/秒的点云，范围可达120米，典型精度为±2厘米。装置的上半部分由激光发射器(4组，每组16个)组成，下半部分由激光接收器(2组，每组32个)组成。

在实际应用中，激光雷达制造商和用户面临的一个主要问题是校准[21]。激光雷达器件的性能在很大程度上取决于其校准。在良好校准下，可以很容易地从环境中提取出精确的三维数据来提取线性或平面特征。相反，如果传感器校准不当，这些特征的提取可能是困难的、不可靠的或不可能的。

一个多束激光雷达系统被建模为一组射线，即直线。这些光线定义了在一个传感器固定的坐标系中激光束的位置和方向。这类系统的固有校准是定义每一束激光的位置和方向的参数的估计。校准技术的基本原理是对激光雷达校准参数进行优化估计，使激光雷达获得的三维数据与地面真实值相匹配。标定过程是一个涉及多个参数的优化过程，可分为以下几个步骤。

**参数化选择**:在三维坐标系中定义一束激光至少需要5个参数，其中2个角度定义相关线的方向，3个参数定义光束的原点。如果需要一个距离校正因子来校正激光束所做的测量，则每束激光束的校正参数数可达6或7个。

**目标函数的选择**:目标/成本函数C是优化过程的基础，用于对获取的三维点云数据与实际环境进行定量比较。当获取的三维数据与地面真实环境的差异较大时，C应提供更高的成本，当获取的三维数据与真实环境的匹配程度提高时，C应提供更低的成本。

**数据分割**:这一步包括从获得的数据中提取实际对应于校准对象的数据，该校准对象的地面真实值已知。为校准过程选择的环境应该被设计和制造以允许适当的数据分割。

随着光束数量的增加，校准参数的数量也随之增加。因此，对于光束数较多的器件，标定过程较为困难。这是为什么拥有更多光束的设备比设备更贵的一个主要原因。

2.2.2 高精度地图概述