

“纯视觉”与“融合感知”，谁将主导自动驾驶的未来？

李子豪

自动驾驶为人类出行带来了翻天覆地的变化。通过精确的传感器和智能算法，自动驾驶车辆能够大幅提升出行的安全性，减少由人为因素引发的事故，同时还能显著提高道路使用效率，缓解交通拥堵，为城市交通带来更加流畅的体验。此外，自动驾驶技术的广泛应用也将促进环保出行，通过优化行驶路线和减少不必要的行车，有助于降低能源消耗和减少尾气排放。最为重要的是，自动驾驶技术将为人类出行带来前所未有的便捷性和舒适性，乘客可以在旅途中放松身心，享受高品质的出行体验。

据中商产业研究院发布的《2017~2027 全球及中国自动驾驶行业深度研究报告》，我国量产乘用车自动驾驶等级正在由 L2 向 L3+ 过渡。2023 年我国在售新车 L1 渗透率约为 11%，L2 渗透率达 51%，L3 和 L4 渗透率分别为 20% 和 11%。目前，各国政府都在加大对自动驾驶技术的政策支持力度，例如中国发布了《智能网联汽车道路测试管理规定》等相关法规，以推进自动驾驶汽车的研发和推广。

自动驾驶汽车的核心技术是人工智能，其目的在于实现汽车的自主感知、自主决策和自主执行。感知技术作为自动驾驶的第一环，是车辆和环境交互的纽带，自动驾驶系统整体性能很大程度上取决于感知系统的性能。早在 2004 年，斯巴鲁就与日立汽车系统（后合并为日立安斯泰莫）合作研发立体摄像头技术，获取

车辆环境信息。该技术后来发展为斯巴鲁的 EyeSight 系统（如图 1），是业内首个基于感知的驾驶辅助系统。它能检测多种障碍物，实现预碰撞制动、自适应巡航等功能。在日立汽车系统的技术支持下，EyeSight 系统持续升级，加入行人检测、碰撞预警制动等功能，大幅提升驾驶安全。

随着技术的不断进步，如今自动驾驶汽车通过各种传感技术（如摄像头、毫米波雷达和激光雷达等），获取环境数据的能力越来越强（如图 2），使得车辆在各种复杂的路况环境下都能够自主运行。

当前，自动驾驶在感知层面的发展路径主要分为两个方向，分别是“纯视觉方案”和“融合感知方案”。业内对于这两种发展路径也一直争论不休，已经成为当前自动驾驶发展道路上的一大分歧。

1 “纯视觉”感知方案

“纯视觉”顾名思义，仅依靠车载高清摄像头作为感知器，模仿人类视觉系统，通过捕捉并深入分析图像数据，以此来感知车辆周围的环境。为了达到 360° 无死角的环境感知，摄像头通常被精心布局在车辆的四周。

纯视觉方案强调的是大量的数据积累和强大的算力。目前业界最先进的高清摄像头能够捕获高分辨率、高帧率的图像，为自动驾驶系统提供更加细致入微的环境信息，AI 人工智能技术也已在自动驾驶感知领域发挥重要作用，从而助力车辆更精确、更智能地感知四周环境，并据此做出更为合理的驾驶决策。

目前汽车行业走纯视觉方案路线的车企主要是特斯拉，斯巴鲁有部分车型也采取了纯视觉方案。

特斯拉车上的 8 个摄像头和强大的视觉处理能力可实现 360° 视野范围，对周围环境的监测距离最远可达 250m。特斯拉方面还表示，与配备雷达的车相比，采取纯视觉方案的 Model 3 和 Model Y 在美国和欧洲保持或提高了主动安全等级，并且在行人自动紧急制



图 1 斯巴鲁第二代 EyeSight 视觉检测

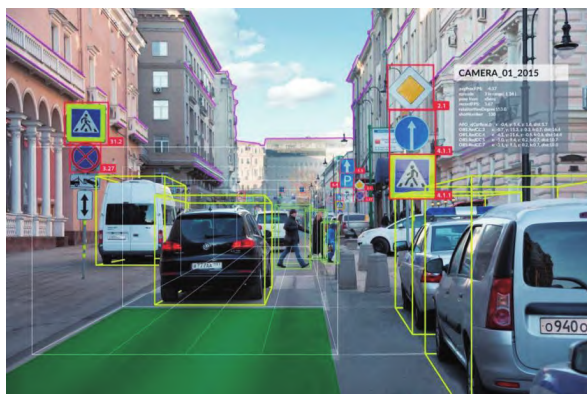


图2 自动驾驶 3D 视觉物体检测

动 (AEB) 干预方面表现更好。多年来,马斯克倡导的“自动驾驶第一性原理”强调:图像数据本质上已经包含驾驶所需的一切信息,人类能做到, AI 司机也应该能做到。他认为多余的传感器其实都是“拐杖”,只会徒增系统的复杂度和资源占用。

摄像头在纯视觉感知方案中扮演着数据采集的核心角色,其重要性不言而喻。目前,市场上主要有单目摄像头、双目立体摄像头两种类型。两者的关键差异在于,单目摄像头仅限于输出二维信息,无法直接获取深度数据,而双目立体摄像头则借鉴了人眼的立体视觉机制,它通过两个摄像头从不同视角捕捉同一目标的影像,并利用视差计算来精确测定目标的距离。这种基于物理测量的方法,相较于单目摄像头依赖逻辑推理和大数据训练来测距的方式,显然具备更高的精确度。

前面我们提到,斯巴鲁与日立汽车系统(日立安斯泰莫)在 2004 年就合作研发了先进的立体摄像头技术。20 年来,日立安斯泰莫在自动驾驶摄像头技术上一直在持续精进。目前,日立安斯泰莫最新的技术是基于左右分离型双目立体视觉摄像头实现 3D Free Space (如图 3)。该技术利用立体视觉识别物体,通过 3D 感知技术对未经学习的物体进行处理,实现更安全的驾驶,即使是未经学习的物体也可以被检测为立体物。结合成像信息和立体检测结果,形成高精度的可行驶区域 Free Space,在危险来临时避开障碍物,确保车辆可以持续行驶至安全场所。此外,该摄像头通过自动校准技术调整左右摄像头画面,实现分离型立体摄像头构造下的立体识别技术,提升安装灵活性、并支持聚焦客户的感兴趣区域。

相比于传统的立体摄像头,3D Free Space 的视角扩大了 3 倍,通过扩大立体摄像头的水平 FOV 来扩

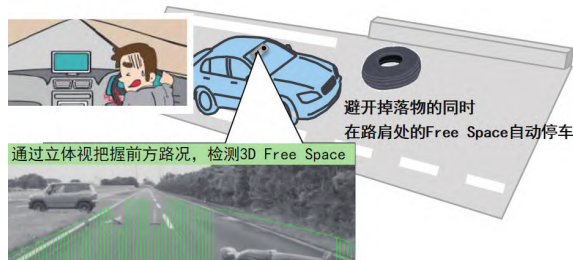


图3 3D Free Space 检测运行示意

大检测范围,在不需要其他传感器的条件下,仅依靠立体摄像头便可在交叉路口实现直行及转弯时自动紧急制动 (AEB),同时实现基于远距离物体检测的 ACC (自适应巡航控制),且无须大幅增加摄像头的像素数。

此基于双目立体视觉技术的自动驾驶系统还可根据 OEM 对自动驾驶功能的需求,进行多组摄像头自由搭配,摄像头灵活布局,可以使用多至 10 余个视角不同的摄像头组合,利用多个摄像头组合区域的立体视觉和多摄像头对物体的识别和追踪,实现单目摄像头无法覆盖的 360° 的 3D 测距,更好地把握驾驶环境,以及通过识别其它车辆的特征(如闪光灯、红灯和制动灯)来预测其行为等。在多组摄像头搭配状态下,可覆盖车辆前方 250 m、后方 180 m、左右分别为 50 m 的范围,可实现纯视觉感知的一般道路自动驾驶和自动泊车等。

摄像头与雷达在感知原理上有所不同,因此摄像头不能被遮挡,这意味着它在低光照或雨雪等恶劣天气条件下的障碍物识别可能受到一定的影响。为了解决这一问题,日立安斯泰莫研发出了一种创新的基于 AI 的能见度诊断技术。该技术旨在提升摄像头在恶劣天气条件下的感知可靠性。若摄像头被遮挡或发生故障,由于 AI 已经深入学习了这些故障模式,因此能够迅速识别每个摄像头中出现的问题,并立即发出警报,提醒驾驶员注意,从而有效预防潜在的危险。此外,

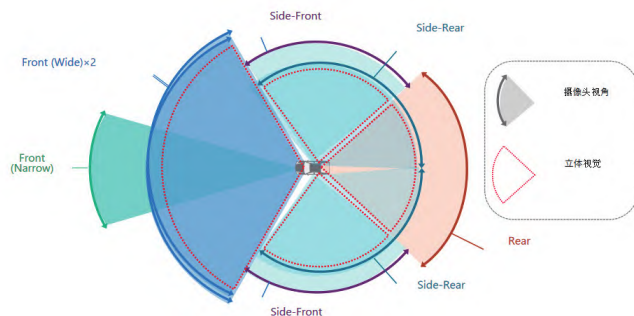


图4 日立安斯泰莫 360° 立体视觉监测范围示意

为了确保系统在投放市场后依然能保持高精度，该系统还配备了自动校准功能，可在行驶过程中进行自适应调整，以确保长期稳定的性能。

同样坚持纯视觉感知方案的斯巴鲁汽车，也算是该领域的先行者了。如图 5 所示基于“预防安全”理念开发的 EyeSight 驾驶辅助系统如今广泛搭载到了斯巴鲁车型上面，EyeSight 就是仅通过双目摄像头便能实现跟车、碰撞预警以及紧急制动等多项功能，在同类技术中显得尤为出色。即使在恶劣的天气条件，如暴雨、夜间或逆光行驶时，EyeSight 也能保持其良好的工作状态，准确识别前车和车道。

具体来说，其防碰撞制动系统（PCB）是在当有碰撞危险情况时，系统首先会通过警示音和警示信号提醒驾驶员注意。如果驾驶员未进行回避操作，系统将主动制动，自动减速或停车。这一功能在车速差小于约 60 km/h 时，可以有效预防碰撞或降低伤害。ACC 功能在 0 ~ 180 km/h 的速度范围内工作，系统可以自动调整车速以保持与前车的安全距离。该功能可以在机动车专用道路上实现跟随行驶，减少对加速和制动踏板的多余操作。除此之外，EyeSight 还包含了起步加速踏板误操作预防功能、车道偏离修正辅助、多目标识别等功能，且性能优异。总的来讲，斯巴鲁 EyeSight 是一个将企业成本管理与实际用户体验完美结合的自动驾驶系统，也是一个让驾驶变得更轻松，为驾驶员提供安全保护和便利的前沿技术。

与昂贵的激光雷达等传感器相比，高清摄像头极具成本优势，这对于降低自动驾驶系统的总体造价、提升其市场竞争力有关键作用。

2 “多传感器融合”感知方案

当前业内很多厂商认为，单一的传感器往往存在



图 5 斯巴鲁 EyeSight 驾驶辅助系统

局限性，无法在各种环境下都提供准确、全面的信息。鉴于安全性和可靠性的重要考量，众多厂商选择激光雷达作为摄像头的辅助和备份，从而构建出“以视觉为主导，激光雷达为辅助”的融合智能驾驶系统。

融合感知方案不仅依赖于摄像头，还结合了能够主动发射的激光雷达与毫米波雷达等设备的回波反射技术，以此来精准地识别外界的静态与动态目标。这种双独立系统相互协作，既能互为备份，又能相互增强，为智能驾驶提供了更为全面和稳定的支持。

在融合感知方案中，还存在“前融合”与“后融合”两种主要策略。它们的区别主要在于数据整合方式的不同。

如图 6 所示，前融合是在原始数据层面进行融合，即所有传感器的原始数据（如激光雷达的点云数据、摄像头的图像数据、毫米波雷达的距离和速度数据等）被同时处理。这些数据通过一套统一的算法进行处理，生成一个综合的感知结果。前融合策略能够充分利用不同传感器之间的信息互补性，提高感知的准确性和鲁棒性，融合后的数据可以视作一个“超级传感器”的输出，具有更全面的环境感知能力。但同时，前融合需要处理大量、多样化的原始数据，所以对算法和算力的要求较高，且融合过程中可能面临时间和空间对齐的问题。

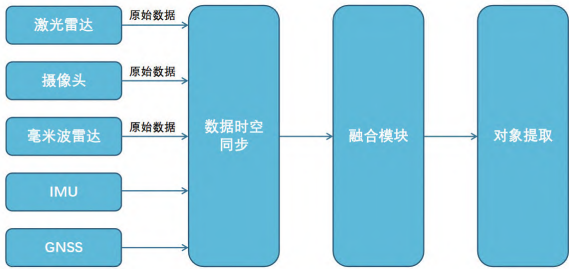


图 6 前融合算法框架

如图 7 所示，后融合是在各个传感器独立完成感知任务后，再进行目标融合，即每个传感器分别生成目标数据（如目标检测、速度预测等），然后由主处理器对这些数据进行校验、比对和合并，形成一个综合的感知输出。因为每个传感器可以独立运行自己的感知算法，对算力和算法的要求相对较低，更容易实现传感器的即插即用和模块化设计，便于系统的扩展和维护。此外，后融合对单个传感器的依赖较低，某个传感器的故障不会导致整个系统失效。出于对系统稳定性的考量，后融合策略也是业内的主流方案。

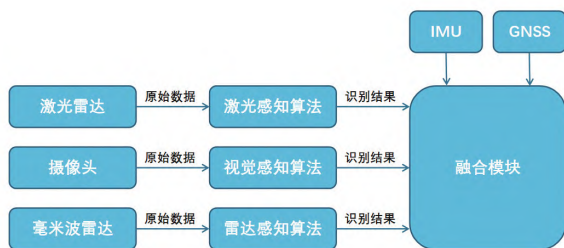


图7 后融合算法框架

总的来看，前融合与后融合两者各有优势，前融合原始信息精度高，后融合直接给出判定，胜在高效率。

目前最新的“全融合”技术，是一种更高级的多传感器融合方法，能够将两者优势结合，取长补短，并通过进一步的数据处理和算法优化，实现了对环境更全面、准确的感知。在数据整合方面，全融合技术建立在前融合和后融合的数据基础之上。它既可以处理原始传感器数据（类似前融合），也可以处理经过初步处理的传感器数据（类似后融合）。而在算法上就是先用前融合的思路得到明确信息，再用后融合的算法得出相应的判断，最后系统再进行融合比对、校验，得出一个感知精度和计算效率双赢的结果。

目前，上汽飞凡汽车的旗下车型搭载的智驾系统RISING PILOT就采用了行业首创的Full Fusion全融合算法（如图8），依托33个顶级高阶感知硬件，包括全球首发量产的LUMINAR 1550 nm高规激光雷达和采埃孚Premium 4D成像雷达等，将前融合与后融合两种算法同步进行，实现了全要素、全观测周期的融合。详细来看，该算法将前融合组合、多任务、多特征网络、深度神经网络输出的探测结果，与后融合多个传感器独立输出的探测结果进行综合比对。依靠高带宽、超算力芯片，能够实现在毫秒级内完成感知、融合、预测、决策和执行等一整套应急响应，确保行车安全，并满足L3~L4级别自动驾驶的算力需求。

3 结束语

关于纯视觉路线与激光雷达路线何者更为优越的问题，业界内依旧众说纷纭，一直未有定论。从实用角度出发，我们不难发现，纯视觉与感知融合视觉在性能上各有千秋，难以一概而论。然而，从成本控制的角度来看，纯视觉方案因其更为简洁的结构，在成本控制上显得更为得心应手，而多传感器融合方案虽在短时间内解决了感知距离、物体形状等难题，而多传感器融合确实是比较复杂，摄像头和雷达独立去做



图8 上汽飞凡的全融合算法

感知数据采集，决策过程中还要做取舍和判断，如果规则设定或者系统架构不合理，也有可能出现相对危险的情况。

人类驾驶主要依赖的感知系统是我们的眼睛，所以纯视觉路线更加类似人类驾驶，考虑到人工智能技术目前正处于突飞猛进的发展阶段，纯视觉技术在此基础上能不断地自我学习进化，而例如激光雷达在硬件原因上无法很好地应对恶劣天气等复杂驾驶状况，这样来看，纯视觉方案潜力似乎更为巨大，甚至超越人类。当前，纯视觉方案的核心难点在于要读懂拍到的东西到底是什么，基于摄像头收集到的仅是2D数据，需要进行图像分割、物体分类、目标跟踪、在线标定、视觉SLAM和ISP等一系列步骤进行匹配与深度学习，分析图像中的每个像素，将其还原成真实的3D场景。

自动驾驶技术的崛起，正在为人类智慧出行赋予全新的可能。除了“纯视觉”和“融合感知”两种方案，也许未来还将出现更多的感知方案。因为有像特斯拉、上汽飞凡和斯巴鲁等在内的众多前沿科技企业在不断研发新产品与新技术，它们的目的无一例外都是为将人们的出行变得更加智能、便捷和舒适，同时也为城市交通规划和道路安全带来新的解决方案。自动驾驶正逐步改变我们的出行方式，引领人类迈向一个更加智慧的交通未来。