# [毫米波雷达与视频前融合算法基本框架](http://10.102.1.36:8091/pages/viewpage.action?pageId=1346837)

**摘要**

传感器融合在机器人感知领域中有着不可替代的作用，这是由于人造传感器对于环境的感知从来都无法做到对环境的完美还原。一方面原因是电子电路对于环境的采样噪声和离散采样，另一方面则是由于物理原理性的采样载体缺陷。对于利用电磁波进行的环境感知方法来说，电磁波的波段决定了感知的可探测性和探测精细程度。例如x光，能够起到穿透人体表皮组织对内脏和骨骼进行成像的作用，但对于希望忽略真实目标前面噪声的采样需求，这种手段由于足够精细，而无法避开噪声影响。更长的波段能够绕射微粒直径小于自身波长的噪声，直接在被测目标上成像，例如空地扫描雷达。采用了更长波长的电磁波，能够不受传播路径上的大气等吸收和散射，直接成像在地面上，从而在卫星环绕地表时，扫描出地形地貌。

考虑到各类传感器设备的感知波段差异和主被动测量调制手段的差异，综合了多频段电磁波的优势，在信号层级上将多元输入进行融合，可以使组合设备对环境的测量感知具有更强的鲁棒性以及不落下风的感知精度，在质的层面上提升传感器套件的感知鲁棒性和感知精度，助力自动驾驶感知设备的发展。

**绪论**

现有技术存在的问题：

1. 单个波段传感器感知的局限性问题

采用单传感器的感知方案是完全无法做到感知道路的瞬息万变情况。在当前的自动驾驶感知方案中，精细的三维感知大多数采用的是激光雷达的方案。诚然，激光雷达起到了对环境的精细感知，通过激光扫描的方式，在很大的距离下还原出周围环境的精细三维信息，并且由于主动探测方案，从而在夜间也有很好的表现。但是，当环境中出现肉眼可见的噪声时，激光雷达的表现将会急剧下降。例如在出现团雾、暴雨、大雪、沙尘等天气时，由于空气中的粒子直径与激光雷达的波长相当并且量足够大，激光在入射到这些粒子聚合体中时，发生衰减和后向散射。一方面产生了对于空气中粒子聚合体的误检测回波，另一方面降低了激光的传输功率，缩小了激光雷达的探测范围。

当然，此时采用更长波长的电磁波进行感知探测，就不再受到噪声的影响，可以实现测到真实目标的作用。但是，改用更长的波长也就意味着更粗略的探测精度。一般毫米波雷达能做到1°的角分辨率，就已经是相当高的测角分辨率了，而激光雷达动辄0.2°角分辨率。因此，增加测量波长虽然起到了抗干扰的作用，但同时也意味着测量精细度的降低。

对于可见光相机而言，则更是如此，由于其被动的采样方式，在光线条件不好的条件下甚至比主动发光的激光雷达的探测效果更差，因此寻求多传感器融合是唯一的解决方法。

2. 目标级融合导致的感知虚检测漏检测问题

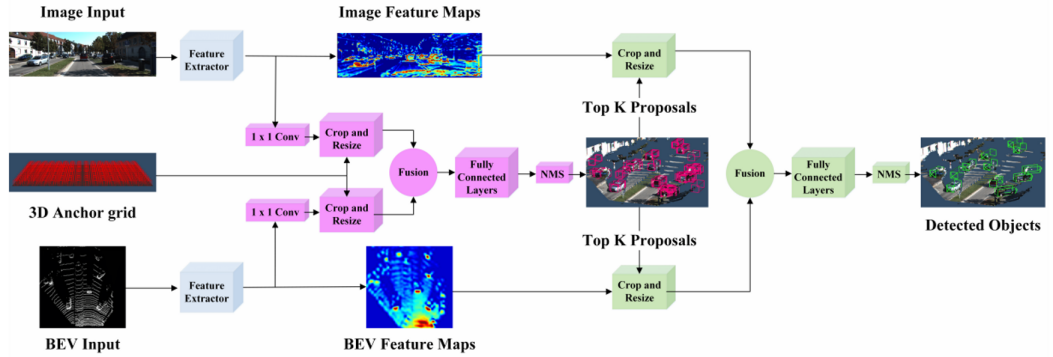
当然，大家很早就意识到可以采用多传感器融合来提升感知的鲁棒性，但由于手段的限制，现在大多数方案是集中于目标级融合之上。这是由于，一方面传感器造价成本高，封装性好，多数方案只能拿到传感器输出的后处理结果进行进一步的操作。这就限制了融合算法的表现。当单个传感器遇到挑战性信号情况时，无法将微弱但是真实的信号作为高置信度测量结果输出，就造成了漏检。若强行降低传感器的输出阈值，则容易造成虚警。

**相关工作**

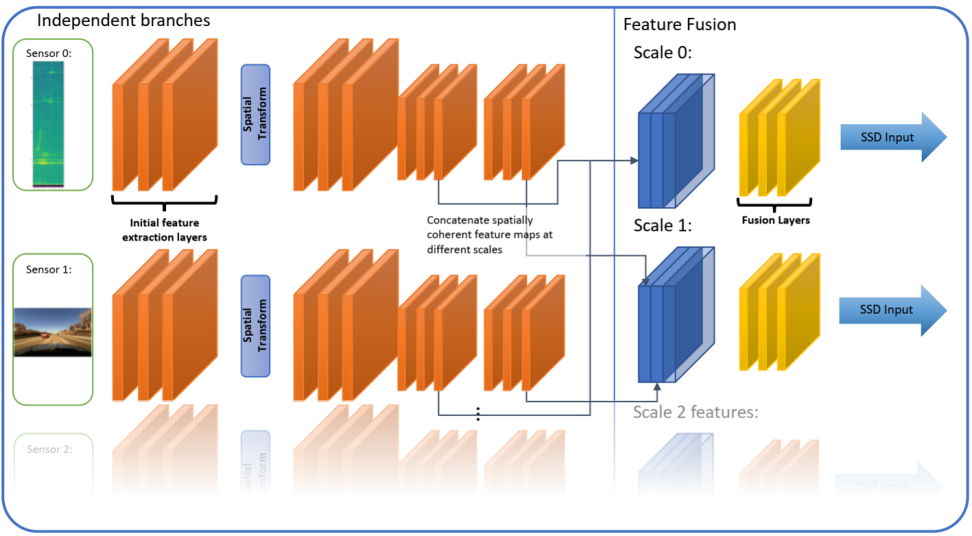
简单来说，毫米波雷达与视频融合的基础出发点分为以下几个方面，在数据已经完成了时间同步和空间同步后，研究多模态信号数据的融合方法成为关键问题。以深度学习方法为架构基础，将数据对齐后输入模型进行特征提取和语义空间转换，数据驱动模型完成拟合。

在融合进行目标检测任务下，第一个方法是以毫米波雷达点云为roi，为图像提供目标检测锚点的融合方式。论文[1]提出的，使用毫米波雷达点提供位置信息，图像提供检测的分类信息。这个方法属于强行组合了各个不同传感器特点，为融合目标检测提供了新信息而已。并没有整整起到融合的作用。论文[2]通过标题表示出来使用了毫米波雷达作为区域锚框，来进行融合。这是一个耦合性较强的方法，很好地利用了anchor based目标检测框架里的特点，很好地利用了毫米波雷达数据点稀疏的特点。

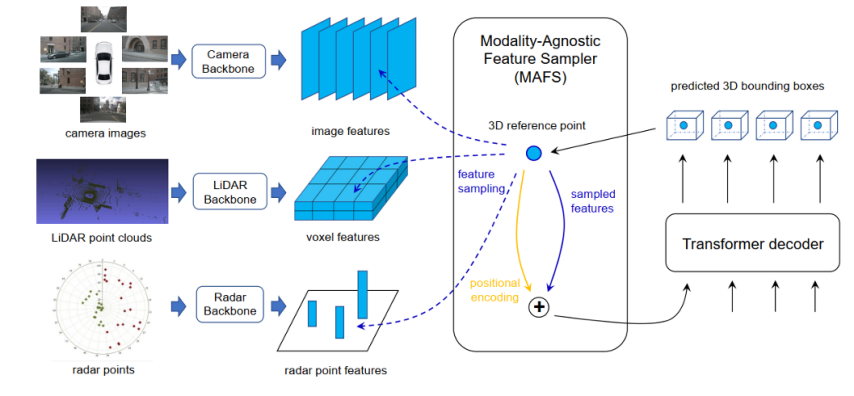
第二类方法则是继承于AVOD[3]激光雷达点云与视频融合的方法，利用二阶段anchor based目标检测框架，制定锚框分别投影到两个通道数据上提取特征后，进行特征图的拼接，最后从拼接后的特征图重回归目标检测框。AVOD的网络框架如下图所示。这类方法比较完整地考虑了两个通道融合特征对应的问题。



第三类方法是将数据对齐在不同视角下后，送入网络进行学习，特征提取和拼接后，回归目标检测结果。例如论文[4]中提到的将毫米波雷达信号和图像数据都放到俯视图视角下进行对齐，后进行特征提取拼接。其网络结构如下图所示，将数据分别进行简单的特征提取后，加入自相关模块计算特征图的自相关权重，分别生成不同尺度的多个特征图后，在不同尺度下拼接这些特征图，之后送入单阶段SSD目标检测网络进行目标检测。值得关注的是，这类方法在数据各通道进行特征提取时采用了注意力模块，这是transformer架构结合CNN架构在融合领域的典型应用方式。



最后一类方法则是直接利用transformer架构来进行传感器融合的例子。如下图所示，是理想汽车研究团队提出的一种利用transformer架构进行传感器融合的"标准"架构[5]的网络结构。众所周知，使用transformer架构的关键在于tokenization和positional embedding，在这个模型里，论文作者加入了参考点选择器，利用这个参考点来作为tokenization的anchor point，从多通道数据特征图上提取特征向量，再拼接position信息，送入transformer模块作为query, key, value。另外这个预测position还会回到MAFS模块来作为备选anchor点。以提高目标检测的精度。具体的reference point 提取方法可以去论文里具体看。



总结上述方法，我们必须感谢深度学习方法，它帮助我们对于多模态传感器特征对齐提供了新的手段。按照传统方法想要进行数据的特征空间对齐，必须是数据存在某种人可理解的维度上的关系，例如惯导的卡尔曼滤波器融合应用，但是很多传感器事实上没有类似的你是观测量，它是控制量之类的联系，因此很难将此类数据融合起来。而深度学习模型帮我们解决了这类问题，我们并不需要关注数据的具体联系到底是什么，而只是相信他们能够在进行一系列特征提取之后，在某个高维空间达成了统一的对齐，可以实现数据融合。

**参考文献**

[1]An Object Detection and Classification Method using Radar and Camera Data Fusion

[2]RRPN-Radar Region Proposal Network for Object Detection in Autonomous Vehicles

[3]AVOD: an Aggregate View Object Detection network

[4]Radar and Camera Early Fusion for Vehicle Detection in Advanced Driver Assistance Systems

[5]FUTR3D: A Unified Sensor Fusion Framework for 3D Detection