



报告正文

1. 年度计划要点和调整情况

本项目的研究内容为：

- (1) 基于人类驾驶感知信息处理与融合新模型的车载多模态传感数据的融合；
- (2) 驾驶意图指导下的ACS体系结构与选择性知觉信息处理；
- (3) 基于人-车-路整体决策的认知心理学模型和高层认知模型指导下的决策优化方法；
- (4) 智能无人驾驶车辆集成验证平台。

第一年的年度计划要点为：

- (1) 智能无人驾驶车辆总体设计；
- (2) 车体的改造以及设备的购买、安装与调试；
- (3) 驾驶员感知信息处理与融合的认知机制研究；
- (4) 动态上下文感知理论与模型指导下感知数据融合的计算模型；
- (5) 基于决策需求的车载多模态传感数据的实时融合；
- (6) 驾驶环境下注意力聚焦与选择模型的研究；
- (7) 基于人类认知机制的驾驶意图指导下的ACS体系结构的研究。

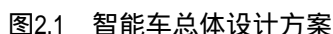
目前研究内容已按以上年度计划要点顺利进行。除此之外，我们还在以下两个方面开展了研究工作：

- (1) 组合导航与全局路径规划研究；
- (2) 运动规划与决策方案设计。

2. 研究工作主要进展和阶段性成果

本年度的研究工作，基于清华大学智能技术与系统国家重点实验室智能汽车课题组原有技术积累与工作基础，进行了更加深入且具创新意义的研究。主要研究工作集中在：1)智能无人驾驶车辆总体设计；2)车体的改造以及设备的购买、安装与调试；3)驾驶员感知信息处理与融合的认知机制；4)环境感知与多模态传感信息融合；5)驾驶环境下注意力聚焦与选择模型；6)组合导航与全局路径规划；7)运动规划与决策方案设计等。目前，研究工作已取得重要进展，并获得了阶段性研究成果。

在本年度的研究工作中，课题组首先完成了智能无人驾驶车辆的总体方案论证，并分别进行了初步设计与详细设计。智能车总体设计方案如图2.1所示。



这五项主要研发任务或功能模块大致可划分为：

- 1) 车辆底层改装；
- 2) 环境感知；
- 3) 组合导航与全局路径规划；
- 4) 运动规划与决策；
- 5) 系统集成。

基于智能车的研发需求以及模块化设计的思想,我们进行了如图2.2所示的具体模块划分。目前,已对各个模块下的构件(功能构件+监控构件)进行了详细的方案论证与设计,并已进入算法实现与单元技术验证阶段。拟采用虚拟现实仿真和半实物仿真等手段,对方案的安全性、可靠性、鲁棒性和实时性进行上车前的预先推演与验证。



图2.2 系统模块划分

2.2 车体的改造以及设备的购买、安装与调试

2.2.1 车体改造

车体改造是智能无人驾驶车辆研发的基础和前提。根据智能车的功能要求，选择SUV汽车较为合适。此类汽车具有车体宽广、高大的特点，适合安装较多设备。具有较好的越野功能，适合于各种道路环境。另外此类车辆具有坚固、稳定、可靠、安全、抗振的特点。

课题组收集了多种SUV汽车的技术资料，并分析了它们各自的特点与智能车的技术要求。参数比较见表2.1所示。

表2.1 适合于智能车改造的SUV车辆的选型

型号	奇骏 X-TRAIL	Jeep 指挥官	宝马 X3	陆地巡洋舰	日本普拉多
指导价	24 万元	48 万元	50 万元	100 万元	50 万元
发动机	2.5L	V8 4.7L	V6 2.5L	V8 4.7L	V6 4.0L
排气量 ml	2488	4701	2497	4664	3956
最大功率 KW/ rpm	135/6000	223/5650	160/6500	202/5400	179/5200
最大扭矩 Nm/ rpm	227/4400	445/3950	218/6500	410/3400	376/378
变速箱	CVT 无级变速	5 速自动变速	6 档自动变速	5 档手自一体	5 档自动变速
最高车速 Km/h	177	202	210	200	180
转向装置	EPS 车速感应式电动助力转向系统	齿轮齿条带助力	电子伺服式助力转向系统	齿轮齿条液压助力	齿轮齿条转向装置
车长 mm	4630	4787	4569	5170	4970
车宽 mm	1785	2261	1853	1970	1875



车高 mm	1685	1919	1674	1945	1905
离地间隙 mm	205	180	201	210	220

最终根据改装的要求，并考虑到改装的方便程度以及车辆的性价比等，我们选择并购买了奇骏 X-TRAIL 2.5L XL 豪华版 CVT 4WD 作为待改造的车辆。

车体改造将主要实现如下目标：

- 1) 方向转角的自动控制，油门的自动控制，刹车的自动控制，档位的自动控制；
- 2) 智能车的自动点火、熄火，自动通断转向灯、喇叭等；
- 3) 多种传感器和测量装置及设备的合理安装，如摄像机、激光雷达、GPS、惯导设备、嵌入式计算机及测控计算机、电机、电磁阀功放及驱动器等；
- 4) 为车载设备提供各类电源。

目前已完成的车辆改造工作包括：

- 1) 汽车发电机安装机构设计、加工、安装与调试，并进行了多次改进；
- 2) 发电机电源（DC28V、DC14V）线（带保护管）的布线；
- 3) 电源管理系统的设计与加工；
- 4) 档位手/自切换机构的设计、加工、安装与调试；
- 5) 方向码盘机构的设计、加工、安装与调试；
- 6) 仪器架、电池盒的设计、加工与安装；
- 7) 驾驶舱摄像机支架的设计、加工与安装；
- 8) 1线激光雷达支架的设计、加工与安装；
- 9) 64线激光雷达台架的设计、加工与安装；
- 10) GPS/IMU组合导航设备台架与天线的设计、加工与安装。

2.2.2 设备购买、安装与调试

目前已完成所有设备（包括车辆、组合导航设备、摄像机、各类激光雷达、车载计算机、千兆网交换机、电源相关模块、无线电台等）的采购，并已初步完成单台设备的测试与安装设计。将在系统集成中进一步完成混成系统的实际调试。

2.3 驾驶员感知信息处理与融合的认知机制研究

该部分研究内容主要针对智能车总体设计、系统架构、传感器选择、感知数据融合与决策模型等，提供认知心理学、人因学尤其是人类感知信息加工的相关理论与知识，以期作为相关研究的设计依据。此外，本研究也以空间信息为切入点研究人类的

感知信息处理机制，可为驾驶员的信息处理研究提供基础。

2.3.1 空间编码和不同线索的关系研究

在动态环境中，新物体和事件的出现伴随着旧物体和事件的消失，例如路边小动物突然跑进车道，前方远处的车突然拐弯消失。这些突现或消失的事件会影响着驾驶员的信息处理。在方位空间信息处理中，当方位词的语义与其所在的位置不一致时，对方位词的位置信息的处理也会受到干扰，这称为空间Stroop效应（如图2.3所示）。本研究通过3个实验分析了线索消失所引起的注意是否同突现线索所引起的注意对方位空间信息处理产生同样的影响。实验1和2比较了突现-消失vs.消失线索的注意效应对空间Stroop效应的影响；实验3比较了突现vs.消失线索的注意效应对空间Stroop效应的影响。结果表明，两种线索均影响空间Stroop效应，但突现-消失线索比消失线索有优势。该论文发表在Psychological Research上。

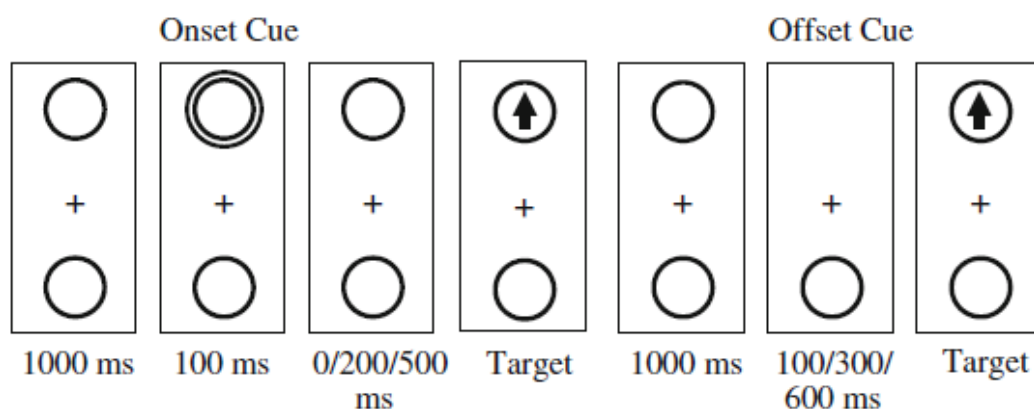


图2.3 空间Stroop任务实验范式

2.3.2 时空信息绑定问题研究

我们采用多人脸追踪的实验范式（如图2.4所示），探讨时空动态变化过程中人脸的身份信息与空间信息绑定的关系。实验1证明了人类在动态过程中的人脸学习以及追踪过程中人脸身份的时空绑定能力；实验2发现正立人脸的追踪绩效显著低于倒立或完全相同的正立人脸的追踪绩效，表明正立人脸激活了某种程度的强制性身份加工，从而与追踪过程竞争注意资源；实验3发现只有作为追踪目标的人脸身份得到加工；实验4表明追踪过程中切换人脸的身份会损害追踪绩效，说明追踪中身份的加工是强制性的；实验5发现注意在身份加工和追踪过程之间的分配权衡关系。这些结果反映了追踪中人脸身份加工的注意偏向，以及时空信息绑定中身份信息的作用。该项工作发表在国际顶级视觉研究期刊Journal of Vision上。

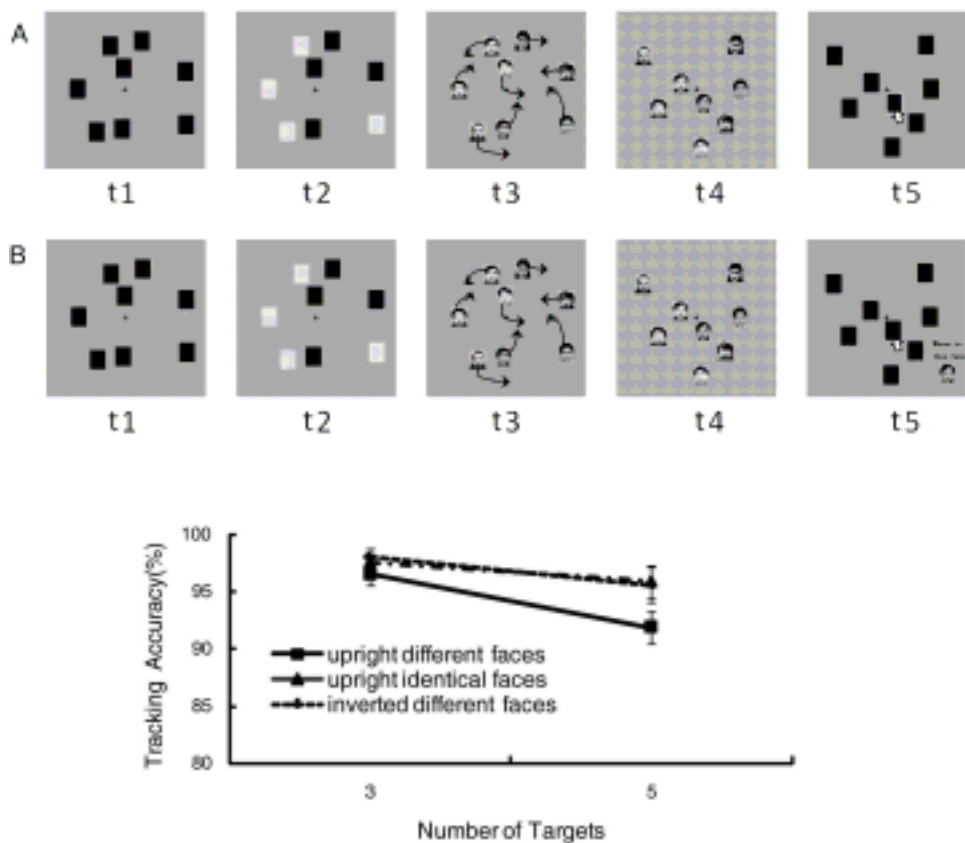


图2.4 MOT实验流程与结果

2.3.3 空间注意与空间记忆研究

采用双任务范式发现，第二任务（多物体追踪，MOT）只影响与空间信息有关的视觉工作记忆，表明视觉记忆与视觉注意只竞争与空间有关的资源。实验1发现物体形状-颜色的记忆不受MOT任务影响，而物体颜色-位置的记忆受MOT任务的影响。实验2发现在空间位置固定条件下，物体的颜色记忆不受MOT任务的干扰，在位置不固定的条件下受MOT任务的影响。而MOT任务的绩效不受记忆负荷的影响。论文即将发表在视觉研究的国际期刊Visual Cognition上。

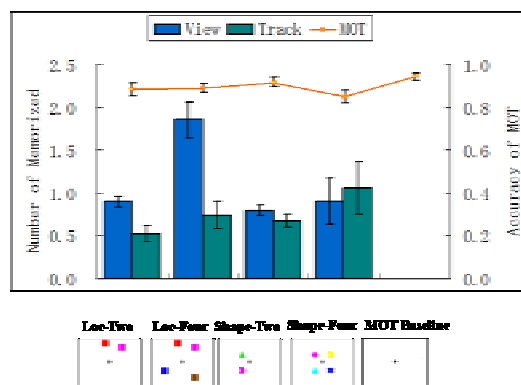


图2.5 不同任务条件下的空间注意和空间记忆绩效



2.4 环境感知与多模态传感信息融合

2.4.1 基于视觉的环境建模与理解

2.4.1.1 研究内容与思路

为了使智能车能在多样的道路、气候和光照环境下可靠地进行自动驾驶，对环境的正确感知起着重要的作用。为了完成特定的任务，智能车需要识别和理解相关的环境与目标。例如自动通过高速公路的收费站、进入高速公路的服务区或在服务区内寻找加油站等。此外，通过对特定地标的识别也可以校准车辆的位置，以消除在没有GPS信号的情况下自动驾驶的误差积累。在对环境的感知过程中，基于视觉信息的处理和理解能够得到关于环境的全局描述，从而可以弥补其他传感器（如激光雷达）的不足。

基于视觉的环境建模与理解是为智能车的导航模块服务的，其主要任务是在智能车行驶的过程中能够识别出当前摄像机视野里的目标环境，同时给出目标环境相对于智能车的位置和方向。基于这些信息，导航模块可以确定出阶段性行驶目标以及给出相应的路径规划。为此需要研究的内容有：1）面向导航的环境建模；2）面向导航的目标环境识别；3）目标环境的定位与跟踪。

目标环境识别的研究通常有以下几种思路：

1) 特定类别物体的检测问题：通过训练特定类别物体的检测器（如收费站的检测器），直接从图像中检测相应的物体。这种思路的好处是可以跨过物体分割问题，直接从原始图像得到语义表述从而给出目标环境的识别结果。但是缺点是检测器只适用于具体的物体类别，不具有通用性。

2) 图像匹配问题：预先采集目标环境的样本图像，然后在实际行驶中通过匹配输入图像和样本图像来完成目标环境的识别。这种思路的好处是匹配算法比较通用，没有假设具体的图像内容。但是缺点是对每个目标环境都要预先采集样本图像。

3) 物体类别的识别问题：先从图像中分割出候选物体，然后识别出相应的类别。这种思路的好处是在得到候选物体的情况下，由于大部分干扰已被过滤掉，可以只考虑某些常见类别的识别。例如在服务器内识别建筑物和加油站的类别，在高速公路上识别ETC提示牌从而可以暗示收费站的出现等。这样，分类问题会相对比较容易，而且具有一定的通用性。但是缺点是需要解决分割问题。

我们先后尝试了前两种思路，但由于不同高速公路收费站的差别较大不易训练检测器（收费站已经是所能考虑的图像结构比较紧凑的物体了），以及图像匹配特征（如SIFT、MSER）的不稳定性（不能总保证输入图像中检测到的特征点与样本图像中特征

点的一致性)而最终放弃。在经过一段时间的论证和探索后,最终选择了第三种识别思路。

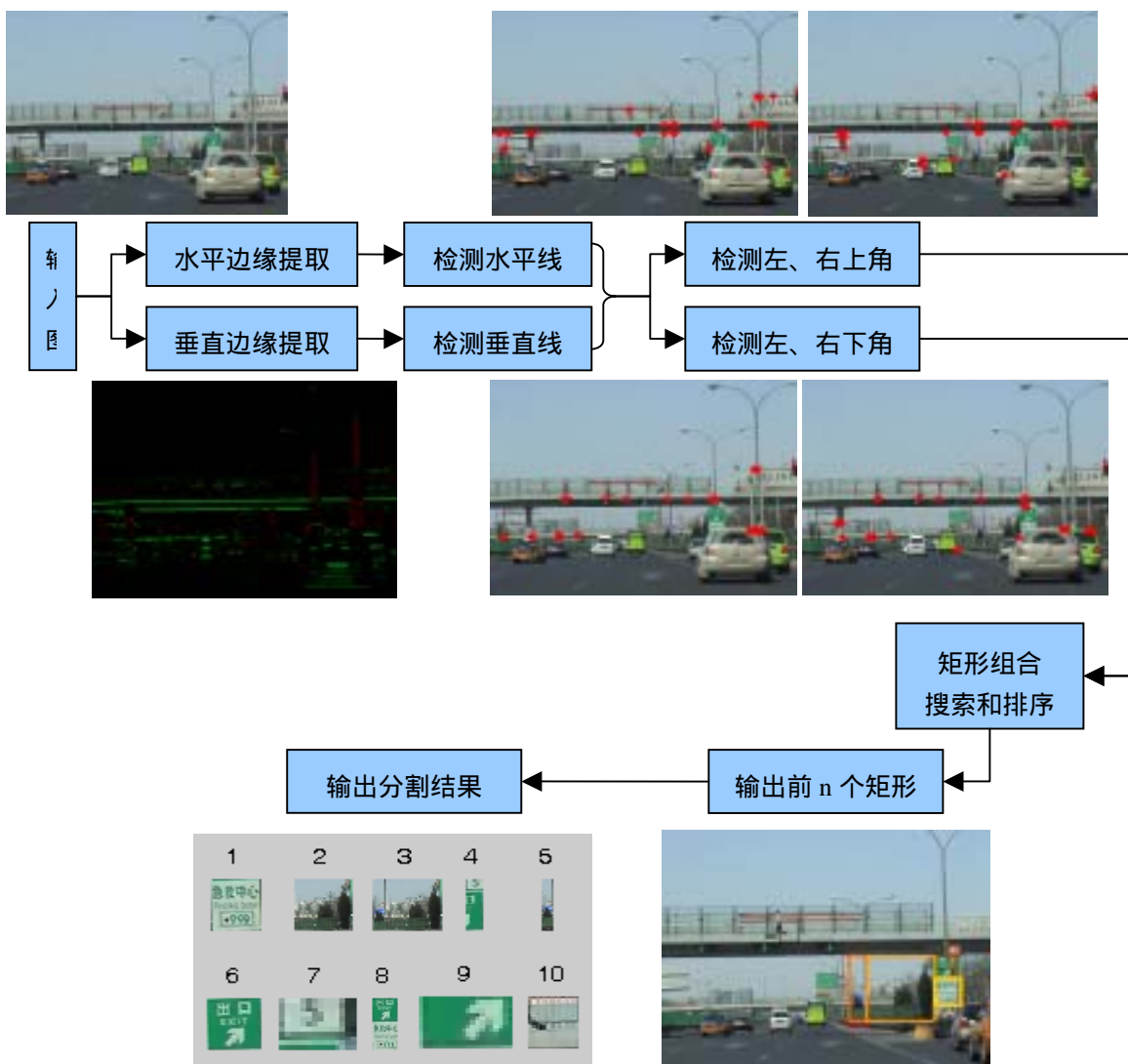


图 2.6 矩形检测的算法流程

在这种识别思路下,首先要解决的关键问题是物体的分割问题,其中包括确定要分割什么样的物体以及怎么实现分割。由于矩形的普遍性(在建筑物、加油站、收费站以及交通标志牌等中都存在),我们以检测矩形框作为研究的切入点。后续还可以考虑检测三角框、箭头线、立交桥以及人行横道线等其他常见的几何机构,以丰富用于目标环境识别的线索。给定每帧中的候选物体的列表,目标环境的识别可以分为2个阶段:1)从候选列表中识别出所期待的类别作为识别的线索;2)从每帧得到的线索时序中推理目标环境是否出现(如通过判断先经过绿色指示牌再经过黄色ETC牌来确定是否到达某收费站)。

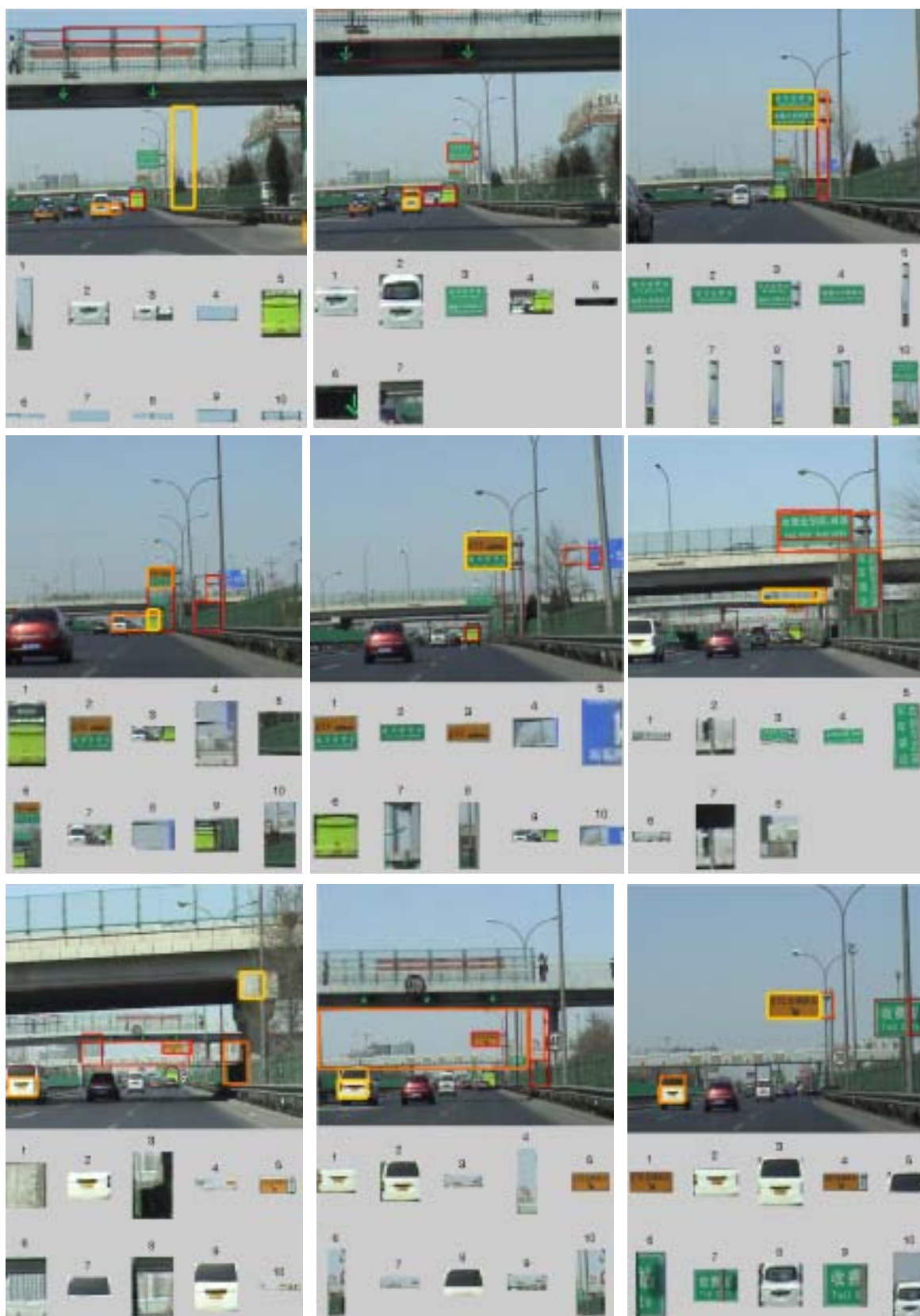


图 2.7 矩形检测的实验结果

在以上的识别思路下，目标环境的建模包括：1) 学习物体的分类器；2) 确定特定目标环境的识别规则。其中分类器可以利用颜色信息和区域内边缘结构特征等来实

现。而识别规则可以基于一阶谓词逻辑或随机文法。规则的设计是在语义层次的，从而不用对每个目标环境采集样本图像，使建模过程变得方便。在识别目标环境后，可以持续跟踪目标环境，然后利用得到的运动信息恢复结构和实现定位。

2.4.1.2 研究进展

课题的进展包括两个方面：1) 通过尝试不同研究思路和方法，确定了最终的具体技术路线；2) 实现了矩形检测算法，给出了物体分割的结果，可为后续识别进行必要的准备。矩形检测算法的流程如图2.6所示，实验结果如图2.7所示。

2.4.2 基于 Velodyne HDL-64 激光雷达的数据可视化与障碍物检测

- 1) 实现了Velodyne数据的实时可视化，并有效地去除了噪声；
 - 2) 对每帧数据量为200,000左右的点云进行地面模型的建模求解，并进一步完成障碍物检测（每秒10~15帧）；
 - 3) 设计并实现了多种障碍物检测算法；
 - 4) 对可通行区域的计算方法进行了优化。
- 部分实验结果如图2.8~图2.10所示。

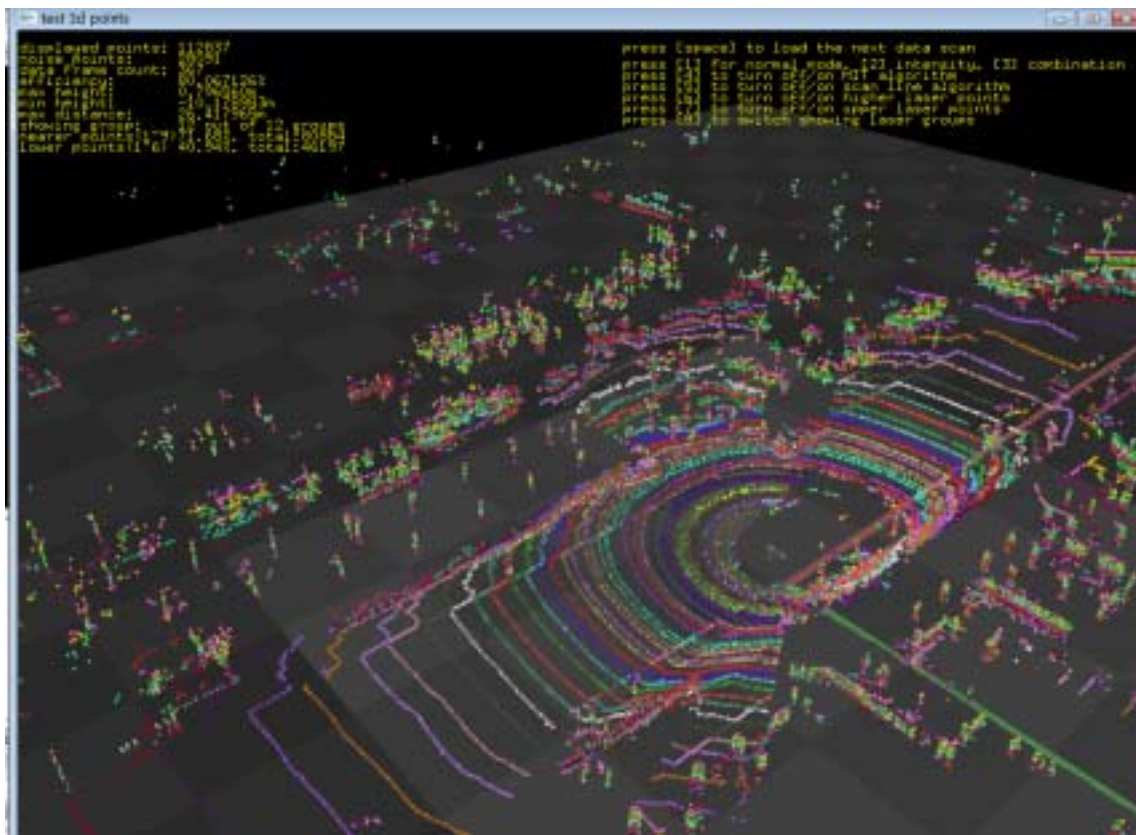


图2.8 3D数据可视化

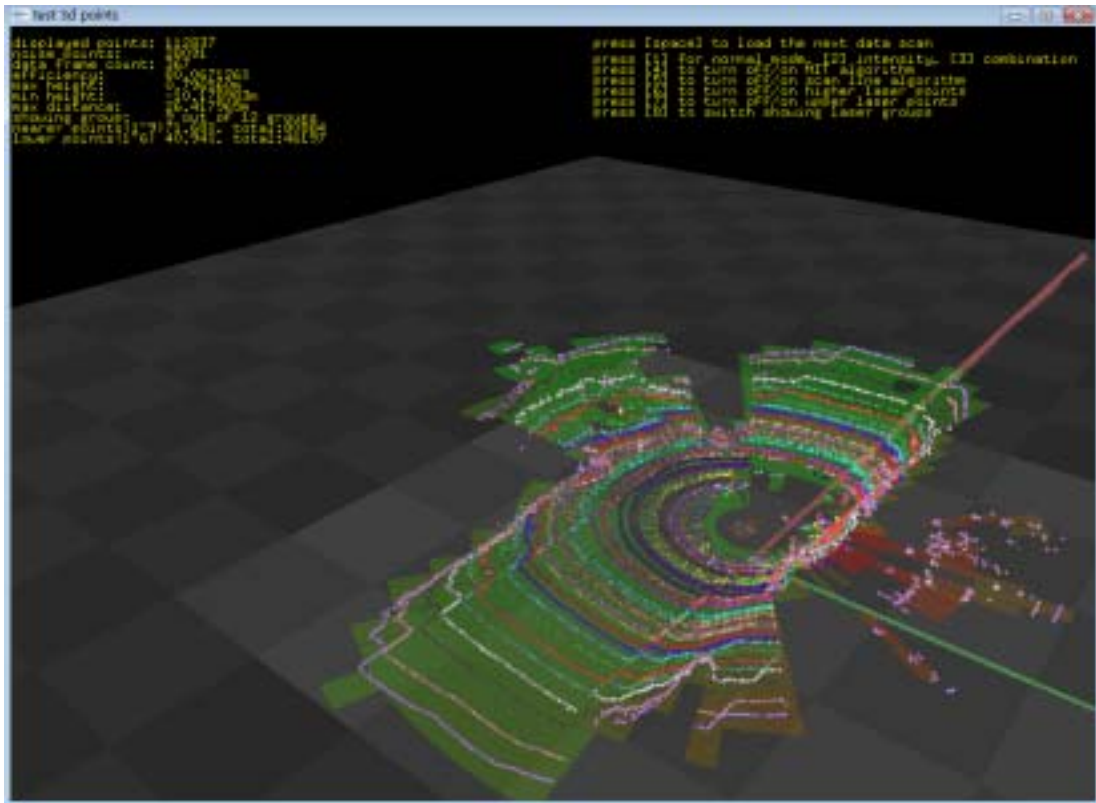


图2.9 道路模型建立

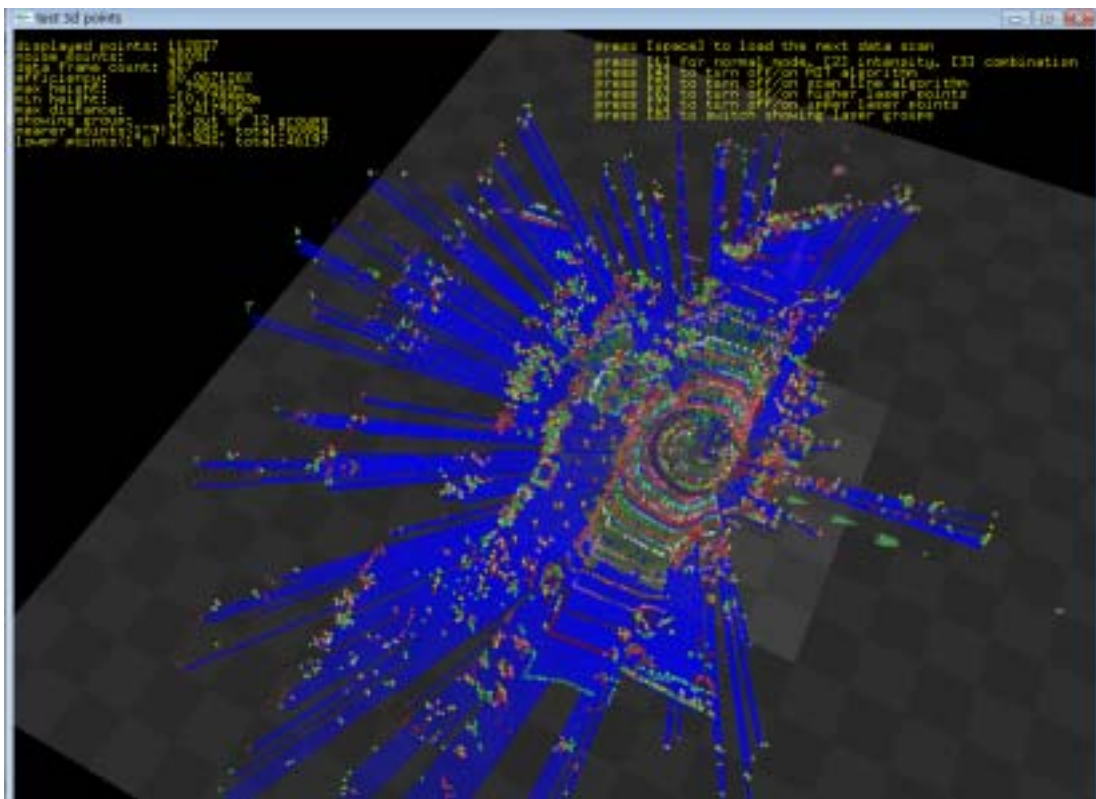


图2.10 可通行区域的算法实现



2.4.3 基于多传感器信息融合的车道线与路面识别

在车道线、路面以及车辆前方障碍物检测的过程中，我们采用了两种不同类型的传感器：摄像机与激光雷达，其中按照用途的不同，激光雷达又包括平面线扫描激光雷达与多线扫描激光雷达。摄像机观察到的物体信息十分丰富且直观，但缺点也是显而易见的。例如，检测结果容易受环境（天气、光照等）条件的影响，而且信息量巨大。如何从海量的图像信息中提取出对检测有用的信息，无疑是一个十分艰巨的任务。与摄像机相比，激光雷达是主动式传感器，其数据基本不受环境的影响，在各种环境下均表现良好，且数据量较小，易于实时处理。但缺点是数据是一维距离数据（对1线激光雷达），不太直观，即使人眼也很难快速地从激光雷达数据中，得到目标物体的真实形状与位置。为了能够有效地利用两种异构传感器的优点，避免使用单一传感器所带来的限制，在本项目中，我们采用摄像机与激光雷达数据融合的方法，对车道线与路面进行识别。

2.4.3.1 摄像机模型及其标定算法

为了建立摄像机所获得的二维图像与三维真实世界之间的对应关系，我们首先需要对摄像机进行标定。对于摄像机标定过程中使用的模型，我们采用最简单直观的针孔模型，即对于一个物理世界中的点 Q ，其坐标为 (X, Y, Z) ，以某些偏移的方式投影到图像平面上的点 (x_{screen}, y_{screen}) ，如式(2.1)所示：

$$x_{screen} = f_x \left(\frac{X}{Z} \right) + c_x, y_{screen} = f_y \left(\frac{Y}{Z} \right) + c_y \quad (2.1)$$

这里，引入 f_x 和 f_y 是因为在摄像机的成像平面上，像素并不是正方形而是矩形， f_x 和 f_y 分别表示摄像机沿 x 和 y 方向的焦距， c_x 和 c_y 分别表示光轴的中心在成像平面上的位置。通常，式(2.1)可以表示为齐次坐标的形式，即

$$q = MQ \quad (2.2)$$

$$\text{其中, } q = \begin{bmatrix} x \\ y \\ w \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

对于径向畸变，我们用式(2.3)对其进行描述，有

$$\begin{aligned} x_{corrected} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{corrected} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned} \quad (2.3)$$

对于切向畸变，我们用式(2.4)对其进行描述，这时有

$$\begin{aligned} x_{corrected} &= x + [2p_1y + p_2(r^2 + 2x^2)] \\ y_{corrected} &= y + [p_1(r^2 + 2y^2) + 2p_2x] \end{aligned} \quad (2.4)$$

因此，我们总共需要五个畸变系数(k_1, k_2, p_1, p_2, k_3)。一般我们可以用一个 5×1 的畸变向量来表示它们。

对每一幅摄像机得到的特定物体的图像，我们可以在摄像机坐标系上用旋转和平移来描述物体的相对位置，如图2.11所示。

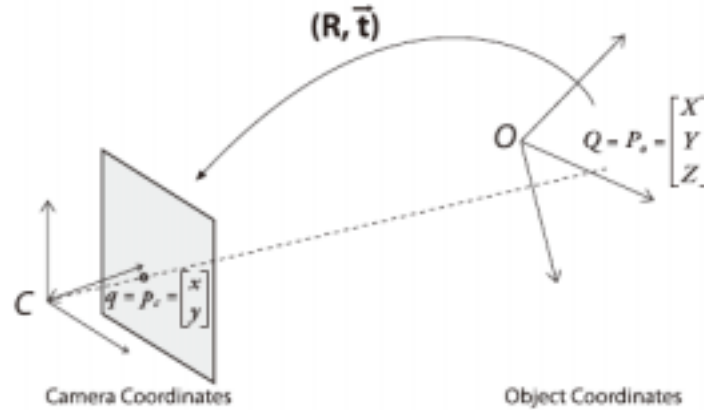


图2.11 旋转向量与平移矩阵

通常，任意维的旋转可以表述为坐标向量与合适尺寸的方阵的乘积。最终一个旋转等价于在另一个不同坐标系下对点位置的重新表述。在三维空间中，旋转可以分解为绕各自坐标轴的二维旋转，其中旋转轴线的度量保持不变。如果依次绕 x ， y 和 z 轴旋转角度 ψ ， ϕ 和 θ ，那么总的旋转矩阵是三个矩阵 $R_x(\psi)$ ， $R_y(\phi)$ 和 $R_z(\theta)$ 的乘积，即 $R = R_z(\psi)R_y(\phi)R_x(\theta)$ ，旋转矩阵 R 的特性是它的逆矩阵就是它的转置阵。因此有 $R^T R = R R^T = I$ 。

平移向量是用来表示怎样将一个坐标系的原点移动到另一个坐标系的原点，或者说，平移向量是第一个坐标系原点与第二个坐标系原点的偏移量。因此，从以目标中心为原点的坐标系移动到以摄像机中心为原点的另一个坐标系，相应的平移向量为 $T = origin_{object} - origin_{camera}$ 。点在世界坐标系中的坐标 P_o 到在摄像机坐标系中的坐标 P_c 的变换方程为：

$$P_c = R(P_o - T) \quad (2.5)$$

我们可以用三个角度来表示三维旋转，用三个参数(x, y, z)来表示三维平移。因此我们总共有6个参数。由于摄像机的内参数矩阵有4个参数，因此对每个视场的求解需



要10个参数。

我们采用张正友的方法对摄像机的内参数、外参数和畸变参数进行求解。该方法的一个显著优点是只需要使用一个平面的棋盘格图像。在每个视角中，我们改变其相对摄像机的位置，对于每个视角的图像，采用角点检测的方法检测棋盘格线的交点，通过求解单应性方程来获得摄像机的内参数以及每个视角下的外参数。对于每个视角，我们把物体坐标系的原点固定在棋盘格的第一个交点上， x 和 y 轴分别沿棋盘格平面上的角点的横向与纵向伸展， z 轴垂直于棋盘格平面。这样，对于棋盘格上所检测到的每个角点，都有 $z=0$ ，因此，三维空间中的点到摄像机二维平面上的映射就转化为二维空间平面到摄像机二维平面上的映射。在采集了多个视场的图像后，棋盘格的角点所能提供的约束将大于未知变量的个数，因此可以用最小二乘的方法来求得摄像机的内参数和每个视角下的外参数。

以上的标定过程没有考虑到摄像机的畸变参数，即在以上的标定过程中，我们假设所有的畸变参数都为0。为了求解畸变参数，我们以未考虑畸变参数时得到的摄像机内参数作为初始值，来求解以下方程，如式(2.6)和(2.7)所示，即

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} f_x X^w / Z^w + c_x \\ f_y X^w / Z^w + c_y \end{bmatrix} \quad (2.6)$$

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \begin{bmatrix} x_d \\ y_d \end{bmatrix} + \begin{bmatrix} 2p_1 x_d y_d + p_2 (r^2 + 2x_d^2) \\ p_1 (r^2 + 2y_d^2) + 2p_2 x_d y_d \end{bmatrix} \quad (2.7)$$

其中， (x_p, y_p) 为点的真实位置， (x_d, y_d) 为畸变后点的位置。

2.4.3.2 摄像机与激光雷达的联合标定

为了能将激光雷达的数据可视化，并与摄像机数据进行融合，我们需要将两者的数据统一到同一坐标系下。由于摄像机得到的是二维平面，而激光雷达得到的只是一条线上的距离值，因此我们将激光雷达的数据投影到摄像机坐标系下，这就必须知道摄像机与激光雷达两者之间的相对位置，即旋转矩阵与平移向量。为了避免直接测量的不便和由此可能产生的较大误差，我们采用将两者进行联合标定的方法来自动获取它们之间的旋转矩阵和平移向量。

首先进行坐标系的定义，如图2.12所示。

假设从摄像机坐标系到激光雷达坐标系的旋转矩阵为 Φ ，平移向量为 Δ ， P^f 为激光雷达坐标系中，空间某点的坐标， P 为摄像机坐标系下同一点的坐标，我们可以得

到如式(2.8)所示的方程。

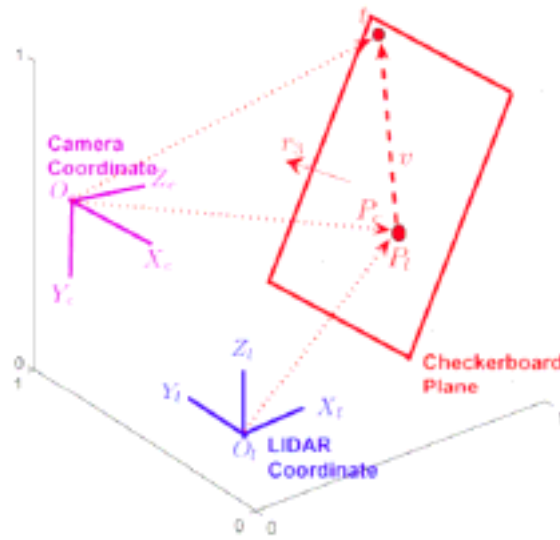


图2.12 激光雷达与摄像机坐标定义

$$P^f = \Phi P + \Delta \quad (2.8)$$

同样地，从激光雷达坐标系变换到摄像机坐标系的坐标变换可以表示成

$$P = \Phi^{-1}(P^f - \Delta) \quad (2.9)$$

我们同样用一个规则的平面棋盘格来进行摄像机与激光雷达的联合标定，我们将标定用的棋盘格平面称为标定平面，并将物体坐标系（世界坐标系）定义在标定平面上，且使标定平面的 $Z=0$ 。对于一个物体坐标系上的点 $P(X,Y,Z)$ ，其在摄像机成像平面上的坐标为 $p(u,v)$ ，物体坐标系到摄像机坐标系的旋转矩阵与平移向量分别为 R 和 t ，因此有

$$p = sM(RP+t) \quad (2.10)$$

其中， s 为尺度常量，用以表示物体坐标系的尺度与摄像机坐标系的尺度关系。

在摄像机坐标系下，我们用一个向量 N 来表示标定平面，向量 N 的方向定义为标定平面的法向， N 的模定义为摄像机坐标系原点到标定平面的距离，不难从式(2.10)中得到 N 的表达式，即

$$N = -(R_3^T t) R_3 \quad (2.11)$$

在标定过程中，我们固定好摄像机和激光雷达，保持他们的相对位置不变，然后采集多个视角下的图像。在每个视角下，保证标定平面上所有的角点在摄像机拍摄的

图像中可见，并以此求出该视角下标定平面的外参数，通过式(2.11)，求出该视角下的标定平面 N 。同时，保证激光雷达的扫描线通过标定平面，并手工拾取激光雷达在标定平面上的投影点，并求出这些点在激光雷达坐标系下的坐标。假设其中某点在摄像机坐标系下的坐标为 P ，由于其落在标定平面上，因此 $N \cdot P = \|N\|^2$ ，结合式(2.9)，可以得到

$$N \cdot \Phi^{-1}(P^f - \Delta) = \|N\|^2 \quad (2.12)$$

在得到标定平面 N 与一个激光雷达坐标系下的点 P^f 后，式(2.12)提供了关于 Φ 和 Δ 的一个约束。在多个视角、每个视角下有若干个 P^f 的情况下，我们很容易用线性方法求出 Φ 和 Δ 。

线性方法优化的是代数上的距离，并不是我们直接测量到的距离。因此，我们用非线性的方法来求激光雷达的测量点与校正平面之间的距离的最小值，如式(2.13)所示。

$$\sum_i \sum_j \left(\frac{N_i}{\|N_i\|} (\Phi^{-1}(P_{ij}^f - \Delta)) - \|N_i\| \right)^2 \quad (2.13)$$

关于求该函数的最小值，有多种方法，如经典的Levenberg-Marquardt算法(LMA)等。但这些算法的一个显著缺点是，能否找到全局最优值依赖于所给定的初始值，如果初始值离真实值偏差较大，则很可能无法收敛到全局最优值。一般以线性优化的结果作为非线性优化的初始值。为了避免复杂的代数计算，并且能在一个较大的初始值范围内都能收敛到最优值，我们采用粒子群优化(Particle Swarm Optimization, PSO)方法来优化式(2.13)。

在进行PSO优化之前，由于激光雷达落在校正平面上的点必然在一直线上，我们首先采用直线拟合的方法拟合这些点，以去除部分噪声的干扰。

在获得了旋转矩阵 Φ 和平移向量 Δ 之后，我们以它们为初始值，再对包括摄像机内参数矩阵 M 在内的所有参数做一次联合优化，以提高联合标定的精度，优化函数如式(2.14)所示：

$$\sum_i \sum_j d^2(P(\Phi, \Delta, P_{ij}^f) - N(R_i, t_i)) + \alpha \sum_i \sum_j \|p_{ij} - p(M, R_i, t_i, P_j)\|^2 \quad (2.14)$$

这里， $d^2(P, N)$ 为点 P 到 N 平面的距离。

2.4.4 道路标志的识别

2.4.4.1 交通标志的检测与识别

交通标志是一种图形与文字相结合的公共标识，有很显著的颜色和形状特征，起到指示、提示和警示等作用。世界各国使用的交通标志大多类似，但不完全相同。1986年我国第一次颁布了交通标志设置标准GB5768 - 86，并于1999年对该标准进行了修订，升级为GB5768 - 1999。我国现行的交通标志标准就是1999年6月1日起实行的GB5768 - 1999。

交通标志可以分为主标志和辅助标志两大类。主标志分为警告标志、禁令标志、指示标志、指路标志、旅游区标志和道路施工安全标志。辅助标志是附设在主标志下，起辅助说明作用的标志。主要交通标志的作用分别如下：

- 警告标志：警告车辆、行人注意危险地点的标志。
- 禁令标志：禁止或限制车辆、行人交通行为的标志，如图2.13所示。
- 指示标志：指示车辆、行人行进的标志，如图2.14所示。
- 指路标志：传递道路方向、地点、距离信息的标志。
- 旅游区标志：提供旅游景点方向、距离的标志。
- 道路施工安全标志：通告道路施工区通行的标志。



图2.13 禁令标志示例



图2.14 指示标志示例

交通标志的检测与识别是智能驾驶的重要组成部分，对于提高行车的安全性有着重要意义。交通标志的检测与识别是计算机视觉中通用物体检测与识别的一个特殊情况，一方面可以借鉴计算机视觉中很多经典的算法，另一方面又对实时性、准确性和鲁棒性都有着更高的要求。

我们对交通标志和交通标志的检测与识别进行了广泛调研，进而提出了一整套解决方案。系统分为检测和识别两个子系统，参见图2.15。并且进一步对于标志牌的检

测、文字的提取、数字的识别等方法进行了深入探索。特别是设计并实现了一套完整的限速标志的检测和识别算法，并且通过实验验证了有效性。

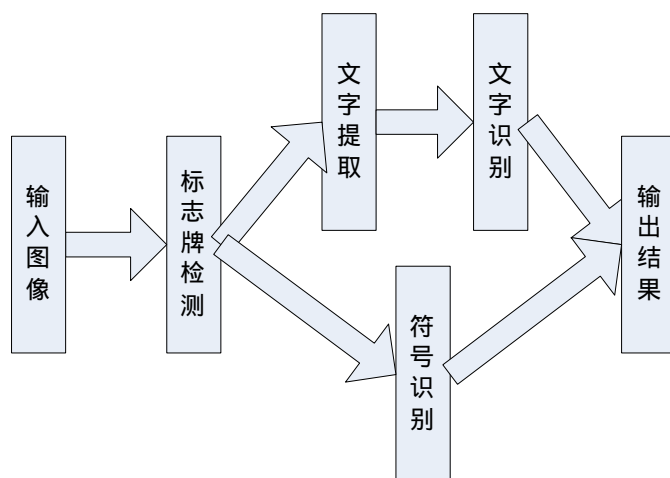


图2.15 交通标志的检测与识别框图

2.4.4.2 限速标志的检测与识别

(1) 限速标志识别的特点

公路上限速标志的种类通常是固定的，而且作为交通标志，其尺寸和规格也有明确的规定。通常为白底，红圈，黑字；形状为圆形。常见的如：40，60，80，100，120等。标准的限速40Km/h的标识示例如图2.16所示。



图2.16 标准的限速40Km/h的标志

但是，实际场景下采集的视频中的图片，通常会受到多种客观因素的影响而使得限速标志图像的效果比较差，难于识别。最常见的情况，由于天气、光照等条件的影响，不同时间拍摄的同一个限速标志图像，其亮度差别可能会比较大。也可能由于拍摄的图像受晚霞的影响，而背景由白色变成粉红色。或者限速标志由于使用时间过长而褪色，造成它与相邻路段的限速标志颜色不同。拍摄过程中，也可能会发生抖动，而造成图像的模糊。拍摄角度的不同，可能会使得原来的圆形变成椭圆。如图2.17所

示的图片，就由于标志本身放置角度的问题，而在图片上呈现出椭圆。

另外，虽然GB5768 - 1999中对于限速标志中数字的字体有比较明确的规定和示例，但是实际拍摄的数据和通过网络下载得到的限速标志的数据，在字体上通常会出现和标准中规定字体不同的情况。此外，复杂多变的背景，也通常会给检测和识别带来比较大的困难。



图2.17 实际场景下形状发生变化的限速标志

(2) 系统原理

根据限速标志检测和识别问题的特点，我们采取了图2.18所示的方法。首先对于一幅输入图片，借鉴人脸检测领域的方法，较为快速、准确、鲁棒地找出含有限速标志的目标区域。然后使用Canny算子在目标区域中提取边缘信息，使用基于RANSAC的椭圆检测方法将其中的椭圆拟合出来，去掉复杂的背景和边框，最后使用一种简单自适应阈值和基于深度优先的分割算法，将其中的数字提取出来。由于边缘信息的鲁棒性，使得该算法对于质量比较差的图像仍能够很好的处理。在得到数字图像之后，进行数字识别，通过对识别出的数字进行组合，得到限速标志的类别。在识别阶段引入了Transfer Learning思想，较好地解决了训练数据集难以收集的问题，显著地改进了识别的效果，使得识别结果较为准确而鲁棒。

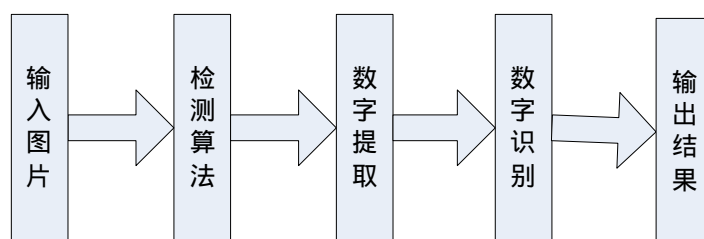


图2.18 限速标志识别原理

(3) 实验结果

使用了家用DV对北京八达岭高速公路德胜门到城角立交桥段的限速标志进行了拍摄。由于设备和车辆条件所限，得到的数据基本上全部都受到了抖动的影响，视频中大部分的帧拖影现象比较严重，如图2.19所示。因此，目前从导出的视频的帧中挑选了部分数据集，标注了560张图片，其中280张图片做训练集，280张图片做测试集。在得到的限速标志中，有30，40，60，80，100，120几种，图2.20给出了一些限速标志的示例。从视频中导出的彩色图片的规格是 720×576 大小，位深度为24。由于拍摄条件的限制，整个数据集的建立是相当困难的。



图2.19 受到拖影影响的图片

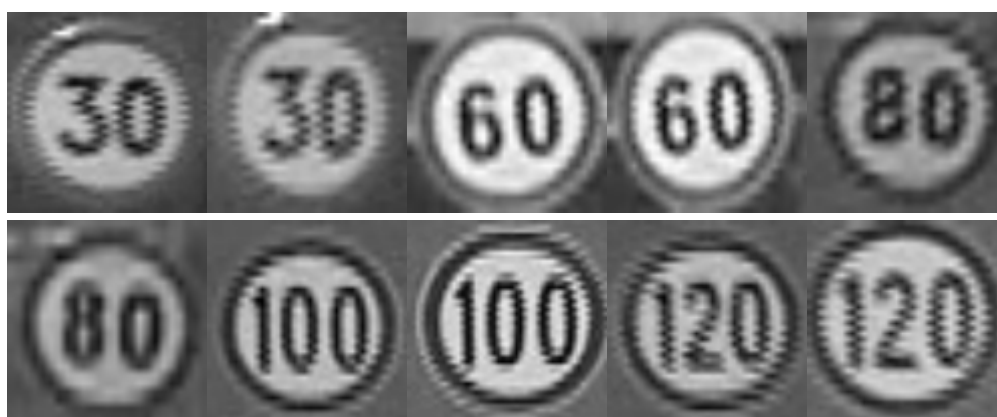


图2.20 限速标志实例

实验平台配置了Intel(R) Core(TM)2 Duo 2.66GHz的CPU和2G内存。为了平衡速度和精度，把原来的图像缩放到 540×432 大小。

检测子系统：

对于280张 540×432 大小的图片进行了测试。图片中共有311个限速标志，限速标



志的大小在 17×17 至 75×75 之间。结果检测出了其中的294个，误检的负例窗口为60个，负例出现最多的一张图片有4个误检的窗口。因此检测率为94.53%，平均每个图片的错误个数为0.2143。检测一副图片平均耗时52.3毫秒。

识别子系统（含数字提取）：

对217张限速标志进行测试，结果只有2张图片识别错误。正确率为99.08%。识别一个标志平均耗时13.0毫秒。

整体性能：

在280张 540×432 大小的图片上测试系统。这280张图片共有311个限速标志，结果检测出了其中的279个，还有32个没有被正确检测出。这32个标志中，24个为漏检，即没有对图像做出标注。原因可能是没有被检测到，也可能是在识别的阶段被排除掉了。8个为误判，即对图像进行了错误的标注。检测结果中，未出现非限速标志被判为限速标志的情况。因此检测率为89.71%，每张图片平均误检个数为0。对一副图片进行限速标志的检测和识别，整个过程平均耗时74.5毫秒。

2.5 驾驶环境下注意力聚焦与选择模型的研究

首先，以认知心理学对视觉注意、变化知觉、工作符合的理论为基础，根据智能车合理配置感知系统的要求，结合当今驾驶人因学研究的最新发现，搜集、整理了大量文献资料，整合为专题综述“驾驶中视觉注意的聚焦和转移”，对智能车感知设备的配置和感知信息的处理起到一定的参照与指导作用。

其次，为满足智能车对环境信息，尤其是环境变化快速检测、实时融合的要求，从人类驾驶者的视觉模式出发，开展在引入真实驾驶中视觉注意分布模式基础上的、对驾驶场景中变化探测绩效与机制的系统研究。该研究不仅在驾驶心理学领域具有一定的理论意义，更可望为智能车海量感知数据的实时融合与快速决策提供有益的参照。目前已形成较为完整的研究方案并经反复论证，进入了实验材料制备、预实验实施的阶段。

2.5.1 用于目标检测的层级注意力模型

复杂环境中的目标检测和识别是计算机视觉所面临的一个很困难的问题。传统算法大多借助于滑动窗口的方式在输入图像中遍历所有可能的区域，以检测任务需要的物体。另一方面，人可以通过注意力的机制将视觉快速聚焦于感兴趣的区域，并把所要识别的物体与周围环境的干扰物隔离开。智能车在实际行驶中，需要通过分析和识



别大量具有复杂背景的视觉输入，以避免障碍物，从而选择合适的行驶轨迹。我们从人的视觉系统的特性出发，建立了一个注意力模型以解决复杂背景中的物体检测问题。

注意力按其形成机制来说可以分为两种：一种是自底向上（bottom-up），基于特征本身的注意力，多称为saliency机制。一般来说，saliency机制通过编码底层特征的神经单元间相互作用来实现，使得这些单元的反应程度对应着某一空间位置的显著程度，这样整个神经单元群体的反应就表示了一张saliency map。另外一种是自顶向下（top-down），基于任务信息，利用先验知识进行反馈，加强那些与任务相关的区域的显著程度。我们的模型结合了这两种类型的注意力，通过自底向上的注意力机制来尽可能的去除无意义的背景信息，再通过自顶向下的反馈对剩余信息的显著程度进行调整，最终得到输入图像中各个位置的显著性信息，结合阈值分割，便可以得到所要探测物体所在的区域。

本模型从视觉通路的层级结构特性出发，将自底向上和自顶向下两种机制进行结合，并以一种生理学上可行的方式加以实现。模型中，编码单元不仅编码了视觉输入信息，同时也编码了注意力信息，即显著性信息。基于已有理论和模型，我们对初级视皮层编码单元之间的局部调节和来自高层视觉皮层的反馈进行了建模和实现，并将这两种作用机制通过视觉通路的信息传递而整合到一起。

我们将模型与其他注意力模型进行了对比。在复杂环境中的物体检测任务中，我们的模型取得了最好的效果。同时我们发现，经过自底向上的注意力处理，输入视觉信息具有更好的区分性。为了说明这一点，我们利用互信息的工具对自顶向下中所用到得视觉特征进行分析，即经过自底向上的处理，这些特征与类别之间的互信息变得更大。这进一步体现了自底向上注意力机制在整个注意力系统中所起到的重要作用。

本工作的研究意义在于，首次基于视觉通路的层级特性将两种类型的注意力机制以一种可行、简便的方式结合在一起，模型对比的实验结果说明了这种仿生模型的有效性。通过对不同类型注意力所起作用的分析，我们的结果也支持了自底向上注意力的重要性。

该工作文章已整理投稿于国际会议：The 23th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)。

2.6 组合导航与全局路径规划

2.6.1 设备采购



依据智能车系统的综合性能要求，在详细了解国内外组合导航设备性价比的基础上，经过充分的论证，最终确定了采用Novatel Span系统作为本智能车的组合导航系统。为了进一步提升系统的性能，我们还完成了与其他设备或系统进一步组合的方案设计。已先后经过了系统论证、试验测试、招标采购等环节。目前设备已经到货并进行了初步的测试，如表2.2所示。

表2.2 高精度Novatel Span GPS/INS组合导航设备

品 目 号	硬件配置名称		数 量	产品特征
1	基准站设备组成	NovAtel FlexPak-V2-L1L2S 接收机	1	双频 RTK：1cm +1ppm
2		NovAtel GPS-702GG GPS 双频天线	1	接收 GPS L1L2 和 GLONASS 信号
3		接收机电源电缆	1	汽车点烟器接头的电源电缆
4		接收机数据电缆	2	nPPS 输出，两个 RS-232 串行接口
5		接收机天线电缆	1	防水 TNC 母头、2 米
6		接收机天线安装磁基座	1	防水铜柱
7		12V 40Ah 铅酸电池	1	工业级、超稳电流、6 小时
8		电池充电器	1	德国工艺、耐用
14	移动站设备组成	NovAtel ProPak-V3-424-RT2j 接收机	1	双 频 RTK：1cm +1ppm；支持 200Hz IMU-FSAS
15		iMAR IMU FSAS-EI-O 惯性测量单元（含 IMAR-IMWS-V2 车轮传感器）	1	闭环技术、陀螺偏差小于 0.75 度/小时、加速度计偏差小于 1mg
16		IMU 原厂标配数据电源电缆	1	MIL-C-38999-，22 针，2 米
17		NovAtel GPS-702GG GPS 双频天线	1	接收 GPS L1L2 和 GLONASS 信号
18		接收机电源电缆	1	汽车点烟器接头的电源电缆
19		接收机数据电缆	2	nPPS 输出，两个 RS-232 串行接口
20		接收机天线电缆	1	防水 TNC 母头，2 米
21		接收机天线安装磁基座	1	防水铜柱
22		12V 40Ah 铅酸电池	1	工业级、超稳电流、6 小时
23		电池充电器	1	德国工艺、耐用
24		移动站电台 South W800(2W ,500 ~ 800 MHz)	1	

2.6.2 系统与算法设计

2.6.2.1 高精度组合导航系统

针对智能自主车辆对导航系统的要求，并结合实际驾驶环境和驾驶经验，我们设计了详细的系统组合方案和相关算法，主要包括：

1) DGPS (差分GPS) + INS (惯性导航系统) + Wheel Speed (轮速) 系统分析(Span



系统)与改进；

2) Compass辅助对准算法(主要提供航向信息)；

3) SINS(捷联惯性导航系统)+车辆模型限制约束(车辆本身约束)；

4) 环境(结构化道路)的环境约束(车道宽度、连续性等约束、车道线跟踪结果的融合)；

5) 双轮速编码器+方向编码器+SINS组合导航；

6) 信标(Beacon)匹配、路标匹配算法(Landmark)——绝对位置纠偏；

7) 地图匹配算法(包括地图的修正和特征点设置、提取)——绝对位置纠偏；

8) 上下文场景约束融合(上下文感知计算——Context)。

目前已经完成系统的整体设计和多传感器组合导航算法的设计，针对系统的工作特点，在上述算法基础上确定了技术实现途径。

(1) 系统初始化

系统初始化主要包括：

- 给定在载体的初始位置和初始速度等信息；
- 数学平台的初始对准，即确定姿态矩阵的初始值；
- 惯性传感器的校准，对陀螺的标度系数、漂移即加速度的标度系数进行标定。

(2) 惯性传感器的误差补偿

静态误差补偿和动态误差补偿。

(3) 姿态矩阵的计算

(4) 导航参数的计算

导航参数的计算就是将加速度计的输出变换到导航坐标，再进一步计算载体的速度、位置等导航参数。

(5) 导航和控制信息的提取

载体的姿态信息、角速度和线加速度等信息可以从姿态矩阵的元素和陀螺仪、加速度计的输出中提取。

2.6.2.2 地理信息和全局路径规划算法

目前已经完成基本的地理信息系统(GIS)方案设计和全局路径规划算法的设计，基本实现了全局规划算法等，具有一定的地理信息系统开发条件。

2.7 运动规划与决策方案



2.7.1 设计原则

运动规划与决策模块是智能车的核心模块之一，主要任务是充分、高效地利用环境感知模块获取的局部动态环境信息和导航模块提供的全局导航信息，进行实时局部运动规划，提供智能车行驶的参考轨迹，同时根据实际路况与车辆状态，通过推理系统，给出智能车的行为决策结果与控制输入，指导车辆安全、可靠、实时地完成自主驾驶任务。

为了使智能车具有类人的驾驶行为表现，在运动规划与决策模块的设计过程中，需充分结合人类认知、推理、决策的行为习惯和方式，充分考虑人类在驾驶车辆过程中的注意力转移方式、决策方式、驾驶意图的体现等方面。并最终设计出简单、可行、完备和稳定的运动规划与决策系统。

2.7.2 整体架构

运动规划与决策模块中总体上包括三大系统。

- 1) 高层行为决策系统：负责对导航信息和环境感知信息的高层理解，对获取的所有信息进行分类与整理，构成规则库的规则前件要素，从而触发有限状态机的状态转移，进而得到不同的环境感知意图、导航意图以及行车模式；
- 2) 运动规划系统：在不同行车模式下，实现地理坐标系下的期望运动轨迹的生成；
- 3) 底层智能车控系统：根据车辆当前的实际运动位姿与期望运动位姿，通过模糊推理系统，实现对汽车方向盘、油门、刹车的行为决策。

2.7.3 输入输出数据定义

方案中首先定义了输入数据的内容，并统一表达格式，供运动规划与决策模块使用。输入数据主要分为两类：1) 导航模块数据；2) 环境感知模块数据。

导航模块数据主要包括：全局路径（Route）数据、自定位（Self-positioning）数据；环境感知模块数据主要包括：车道（Lane）数据、路面（Road surface）数据、障碍物（Obstacles）数据（包括行人、车辆和其他物体）、交通标志（Signs）与信号灯（Traffic Lights）数据、典型路口（Intersection）数据、汽车车灯（Taillights）数据等。

其次，输出的数据主要分为三类：1) 导航意图（ Intention_N ）；2) 环境感知意图（ Intention_s ）；3) 智能车控向量（ \mathbf{D}_C ）。三类输出数据的具体功能描述如下：

- 导航意图（ Intention_N ）：在道路发生变化或完全堵塞的情况下，发出指令，要求导航模块进行全局路径的重新规划和导航；



- 环境感知意图(Intention_s):指导各类传感器在获取环境信息时的注意力方向；
- 智能车控向量(\mathbf{D}_c):给底层车控模块发送控制指令或者直接参与控制。

2.7.4 高层行为决策系统

行为决策系统主要包括智能车行为模式有限状态机与规则库(状态转移条件)两大部分,主要实现对智能车行为模式的粗分类。同时通过信息预处理,给出导航意图和环境感知意图,进一步指导导航模块的路径规划,提供环境感知模块的注意力方向。

首先,定义了智能车的原子动作,所谓原子动作,即该行为是具有单一属性的不可再分的最小动作单元。汽车的所有驾驶行为模式均可以由原子动作复合而成。定义如下:

- 点火;
- 熄火;
- 换挡;
- 变速(刹车与油门);
- 转弯;
- 其他(灯光,喇叭等)。

然后,对智能车行为模式进行了定义与划分。智能车行为模式是指能完成某项固定行驶任务的车辆状态的概括与描述,由原子动作复合而成。是高层行为决策的基础。智能车在行驶中的驾驶行为模式与原子动作关系如表 2.3 所示。

表2.3 智能车行为模式与原子动作关系

序号	类别	点火	熄火	换挡	变速	转弯	其他
1.	启动						
2.	停车						
3.	紧急停车						
4.	车道跟踪						
5.	车辆跟踪						
6.	车道变换						
7.	超车						
8.	U型转弯						
9.	通过路口						
10.	倒车						
11.	泊车						
12.	其他						

行为模式有限状态机是智能车行为模式随着路况条件的变化而自动转移的状态

转移集合，它描述了智能车在行驶过程中所有行为模式之间的转移关系。该状态机包含两个关键问题的设计与描述：1) 状态之间的转移条件（或称为状态转移触发条件）；2) 各个状态中智能车行为决策的设计。

有限状态机转移条件决定了智能车各个行为模式之间的转移关系。根据行为模式转移的方式不同，一共设计了 33 个不同的转移条件，并将转移条件分为主动转移与被动转移条件两大类。目前，已对各转移条件进行了详细定义。

2.7.5 运动规划系统

总体规划方法：首先确定不同行车模式下所需的多个必经路点，然后采用二次或高次曲线（如：抛物线）正切曲线等进行拟合，获得光滑的局部动态期望运动轨迹，如图 2.21 所示。

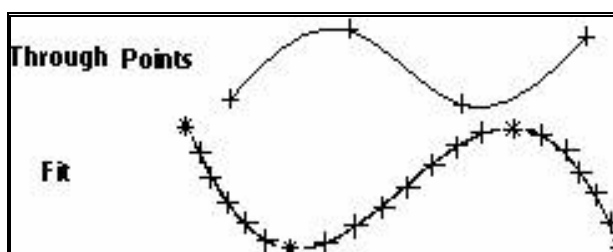


图2.21 必经点曲线拟合

2.7.6 底层智能车控系统

智能车行为决策系统主要模拟人类驾驶汽车的操作行为进行决策，包括对方向盘、油门、刹车等主要执行机构的操作以及对档位、转向灯、喇叭等辅助装置的操作。其主要过程采用模糊推理系统加以实现。首先，根据信息预处理系统提供的智能车实际运动位姿和运动规划系统提供的智能车期望运动位姿获得位姿误差向量，然后，根据事先建立的模糊规则库，通过模糊推理运算，在智能车运动位姿误差与车控向量之间建立映射关系，从而给底层提供可靠合理的执行指令。理论上，可通过图2.22所示原理框图实现。

然而在实际设计过程中，直接采用上述方法有两大难点：1) 很难直接将人类驾驶汽车的经验表达成模糊推理规则；2) 模糊变量个数多且存在耦合，使模糊规则数非常庞大。

人类在驾驶汽车时，一般是通过控制车的航向和车速来达到驾驭汽车的目的，正确的航向和合理的速度能够使汽车按既定轨迹行驶。因此，我们将智能车控系统分成两大子系统：1) 侧向（转向——方向盘）控制系统；2) 纵向（车速——油门、刹车）控制系统。

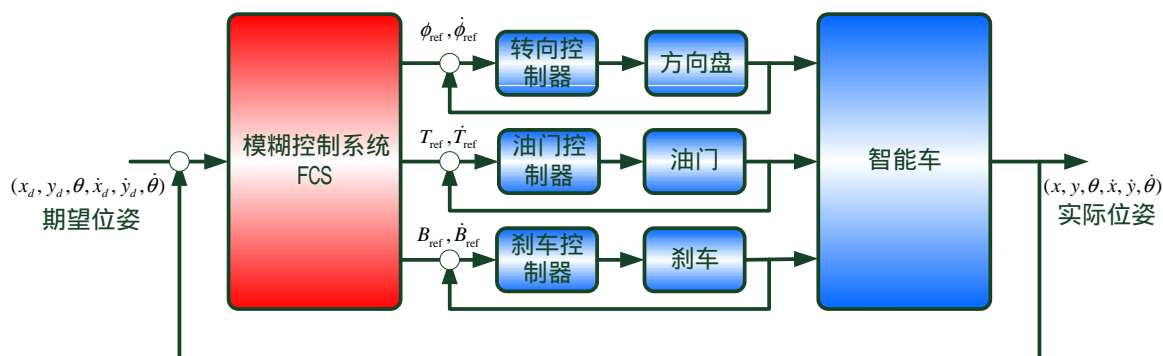


图2.22 智能车控系统框图

对于前述的侧向和纵向控制系统，我们已经完成了输入输出模糊变量的分割、输入输出隶属函数的先验定义和模糊规则库的设计（分别含75条和56条模糊规则数），并进行了初步的仿真研究，结果较为满意。

下一步将利用半实物仿真，对智能车控系统的安全性、可靠性等进行进一步的单元技术验证。



3. 下一年度工作计划

下一年度我们将按照计划书中的年度计划,深入开展研究工作。目前对计划书中的原研究内容和主要成员无调整,课题组除毕业离校的学生外,将保持稳定。

4. 当经费使用情况与下一年度经费预算

2009年度拨入经费180万元,累计使用经费59.84万元,包括科研业务费(主要用于国内外学术交流、西安等地差旅费、论文版面费、复印资料费等),实验材料费(主要包括机械设计与加工费、电子元器件、汽车发电机、进口电液制动装置、步进电机驱动器、转速表、转向控制用光电编码器等购置费),仪器设备费(购置调试用计算机),以及其他费用(包括项目管理费等)。

本项目获得了清华信息科学与技术国家实验室(筹)、清华大学智能技术与系统国家重点实验室和清华大学计算机系的配套支持。2009年度,已为本项目购买了总计达228.13万的设备。

下一年度经费预算将按计划执行。

5. 存在的问题、建议及其他需要说明的情况

无。

6. 附件:论文目录

(以下论文均已标注获本基金资助)

A. 已发表和已有录用通知的论文

- [1] Ren, D., Chen, W., Liu, C. H., and Fu, X. (2009). Identity processing in multiple-face tracking. *Journal of Vision*, 9(5), 18: 1-15.
- [2] Sun, Y. H., Wang, Z., Liu, Y., and Fu, X. L. (2010). Naturally-Formed Objects Categorized As Artifacts: Effect of Objects' Functional Depictions. *Chinese Science Bulletin*, 55(3): 1-6. (in press).
- [3] Liu, Y. Fu, Q. F., & Fu, X. L. (2009). The interaction between cognition and emotion. *Chinese Science Bulletin*. (in press).
- [4] Zhang, H., Xuan, Y., Fu, X., and Pylyshyn, Z. W. (2009). Do objects in working memory compete with objects in perception? *Visual Cognition*. (in press).
- [5] Luo, C., Lupiáñez, J., Fu, X., and Weng, X. (2009). Spatial Stroop and spatial orienting: The role of onset versus offset cues. *Psychological Research*. DOI: 10.1007/s00426-009-0253-z. (in press).
- [6] Wang, B. and Fu, X. (2009). Gender differences in the effect of daytime sleep on declarative memory for pictures. *Journal of Zhejiang University-SCIENCE B*, 10(7), 536-546.
- [7] He, Y., Xuan, Y., Chen, W., and Fu, X. (2009). Subjective image quality assessment: A method based on signal detection theory. *Proceedings of 2009 IEEE International Conference on Systems, Man, and Cybernetics (SMC2009)*, San Antonio, Texas, USA, October 11-14, 2009.



- [8] Shang, J., Fu, X., Liu, Y., and Luo, C. (2009). The role of trait anxiety in the interaction between eye gaze and facial expressions. *Proceedings of the 5th International Conference on Natural Computation (ICNC' 09)*, pp.342-344.
- [9] Ren, D., Zhou, H., and Fu, X. (2009). A deeper look at gender difference in multitasking: gender-specific mechanism of cognitive control. *Proceedings of the 5th International Conference on Natural Computation (ICNC '09)*, pp.13-17.
- [10] 刘焯, 付秋芳, 傅小兰 (2009). 认知与情绪的交互作用. 科学通报, 54(18): 2783-2796.
- [11] 孙宇浩, 王哲, 刘焯, 傅小兰 (2010). 物体的功能会抑制自然成因对人造物分类的影响, 科学通报, 55(1): 1-8. (in press).

B. 已投稿论文

- [1] 赵楠、傅小兰. 驾驶中视觉注意的聚焦和转移. 人类工效学, 2009.