

2022 年英特尔杯大学生电子设计竞赛嵌入式系统专题邀请赛

2022 Intel Cup Undergraduate Electronic Design Contest

- Embedded System Design Invitational Contest

# 作品设计报告

## Final Report



Intel Cup Embedded System Design Contest

报告题目： 私人定制--科技馆智能互动型机器人

学生姓名： 梁嘉辉 见迎超 覃维敬

指导教师： 刘恒

参赛学校： 成都理工大学

## 2022 年英特尔杯大学生电子设计竞赛嵌入式系统专题邀请赛

# 参赛作品原创性声明

本人郑重声明：所提交的参赛作品报告，是本人和队友独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果，不侵犯任何第三方的知识产权或其他权利。本人完全意识到本声明的法律结果由本人承担。

参赛队员签名：

梁嘉辉 见迎超 覃维敬

日期： 2022 年 07 月 27 日

# 私人定制--科技馆智能互动型讲解机器人

## 摘要

传统的科技馆参观存在预约讲解难、讲解员数量不够的问题，存在服务人数有限、参观者体验感弱的弊端。对比现有的博物馆参观模式，拥有自主动态路径和定制化参观讲解服务授权的智能互动讲解机器人提供了高效、方便的参观方式，提供智能人机交互的一站式讲解服务。以 GNS-V40 为边缘计算核心,搭建边缘计算网络,实现算力分摊,保证数据实时性。

**关键词：**定制化参观服务，动态路径规划，智能交互，边缘计算

# **Personal Tailor--INTELLIGENT INTERACTIVE EXPLANATION ROBOT FOR SCIENCE AND TECHNOLOGY MUSEUM**

## **ABSTRACT**

The traditional science and technology museum visit has the problems of difficult to make an appointment and explain, insufficient number of interpreters, and the disadvantages of limited number of service members and weak experience of visitors. Compared with the existing museum visit mode, the intelligent interactive explanation robot authorized with independent dynamic path and customized tour and explanation service provides an efficient and convenient way to visit, and provides a one-stop explanation service of intelligent human-computer interaction. With Seavo GNS-V40 as the edge computing core, the edge computing network is built to realize the computing power sharing and ensure the real-time data.

**Key words:** Custom tour services, Dynamic path planning, Intelligent interaction, Edge computing

# 目 录

第一部分 项目背景	1
1.1 立项背景及意义	1
1.2 国内外研究现状	3
1.2.1 国外研究现状	4
1.2.2 国内研究现状	4
1.2.3 国内外研究对比	4
1.3 项目应用前景	4
第二部分 项目设计方案	6
2.1 研究开发内容及系统架构	6
2.1.1 功能描述	6
2.1.1.1 获取预约信息	6
2.1.1.2 人脸识别解锁	6
2.1.1.3 智能语音交互	6
2.1.1.4 动态规划路线	6
2.1.1.5 自主导航跟随	6
2.1.1.6 实时定位讲解	6
2.1.2 系统总设计	7
2.1.2.1 服务器	7
2.1.2.2 下位机	7
2.1.3 系统硬件设计	7
2.1.4 系统软件架构	8
2.1.4.1 语音唤醒	8
2.1.4.2 人脸解锁	8
2.1.4.3 用户跟随	8
2.1.4.4 人机交互	9
2.1.4.5 动态路径规划	9
2.1.4.6 信息传输	10
2.2 技术关键	10
2.2.1 二维码定位	10
2.2.2 自主导航运行	10
2.2.2.1 红外巡线导航	11
2.2.2.2 激光雷达 SLAM 建图定位与导航	13
2.2.3 自主避障	14
2.2.4 智能语音问答	15
2.2.4.1 语音识别	15
2.2.4.2 模糊匹配	15
2.2.5 智能人脸解锁	16
2.2.5.1 预处理	16
2.2.5.1.1 dlib 模块简介	16
2.2.5.1.2 Resnet 网络架构介绍	17
2.2.5.1.3 Resnet 残差结构的公式推导	19
2.2.6 动态路径规划	19

2.2.6.1 权重动态赋予 .....	19
2.2.6.2Dijkstra 算法 .....	19
第三部分 测试与验证 .....	21
3.1 语音模块 .....	21
3.1.1 字符串分割识别 .....	21
3.1.2 文字转拼音 .....	21
3.2 电机调整 .....	21
3.3 局域网通信 .....	22
3.3.1 模拟键盘输入 .....	22
3.3.2 人脸识别解锁 .....	22
3.3.3 收发指令 .....	22
3.4 动态路径规划 .....	22
第四部分 预期目标 .....	23
第五章 参考文献 .....	24
第六部分 致谢 .....	25

## 第一部分 项目背景

在进行选题过程中，小组注意到，在疫情影响人们生活的现在，人们更注重无接触式服务，其中主要包括快递和外卖的无人配送，酒店自助入住，无人化测温，和景区无人化咨询等。其次，随着疫情防控政策的调整，各地的文化交流活动陆续恢复，科技馆展览正成为人们出行的主要选择。因此，小组决定针对科技馆讲解过程中所遇到的问题，设计智能讲解机器人，来最大化的代替真人讲解员，并为参观者提供更为舒适的参观流程。

### 1.1 立项背景及意义

当今社会科技发展，机器人作为最有代表的高科技产物，其融合了电子、机械、人工智能等交叉学科。目前机器人正在逐渐代替人类来完成那些重复性高且繁重的工作，进而提高了人们的生活质量。

当代科技馆的主要功能是教育，因此科技馆的参观过程中的教育价值与功能的实践程度非常重要，这将会受到社会各界的广泛关注。图 1-1 是 2022 年春节期间各地科技展馆的参观人数统计，可以看到科技馆展览正成为人们出行的主要选择。在疫情受影响的现在，无接触化参观更是人们所追求的。在这个前提下，讲解机器人如果能够提供不亚于普通真人讲解员的功能，这将会大大提高该类机器人的社会效益，给消费者带来极致的体验。因此平衡参观内容的丰富和不同用户的需求，最大的发挥展览教育与科普作用尤为重要。此外一个好的参观体验，对于科技馆的运营也是极为重要的，好的参观效果将会给展馆带来良好的收益。观众的参观体验主要是由他们的参观需求所决定的，根据观众自身以及参观形式的不同可以大致区分他们的参观需求。

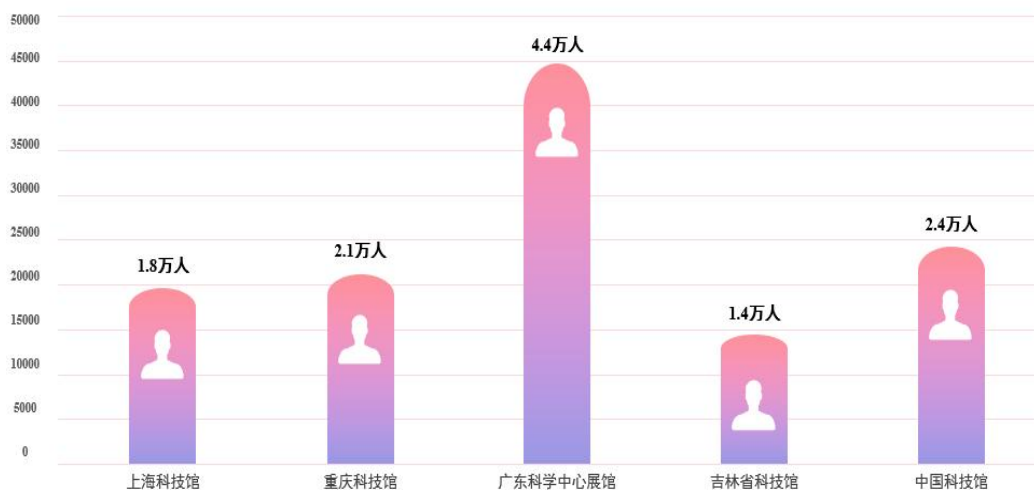


图 1-1 2022 年春节期间各地科技展馆的参观人数

#### 1.1.1 按参观者划分

对于观众自身来说，参观需求的不同主要取决于他们的年龄阶段，主要可以划分为以下四个不同的需求阶段：儿童观众（3 岁至 12 岁），青少年观众（13 岁至 17 岁），成年观众（18 岁至 59 岁），老年人（60 岁以上）。

### 1.1.1.1 儿童观众

儿童观众的需求。儿童体验科学的过程是与玩乐结合在一起的。儿童具有强烈的好奇心，希望看到新鲜的东西。此外，“好动”也是儿童的一大特性，这两大特性决定了为儿童制作展品的特点。对于此类观众，科技馆展品的内容和形式都要新鲜。“形式新鲜”容易理解，而“内容新鲜”则是指设计展品时，对于科学原理的切入点要新鲜，要符合儿童的心理。展品要能够动手参与，要让儿童玩得高兴，并且通过玩能悟出一些道理；在考虑儿童需求的时候，也要关注家长的意向，家长对孩子的参观活动经常有一定的指导作用。家长希望孩子通过参观科技馆拓宽知识面，增长见识，并且能在孩子内心形成对科学的感性认识，培养其对科学的学习兴趣。

### 1.1.1.2 青少年观众

青少年观众的需求。青少年正处于增长知识，积累经验的阶段，参观科技馆，希望通过实践来验证他们在学校里学到的知识，这一过程对他们来说是非常重要的。学习是为了应用，了解知识在生活中的应用，了解科学原理如何存在于生活中，是青少年好奇心的表现。这一特点使得青少年群体同样希望看到新鲜的展示内容，以及更高领域或层面的科学原理的展示。因为年龄的特点，他们善于发现问题，挑战问题，因此对展示内容、展品和教育活动的设计提出更高的要求。

### 1.1.1.3 成年观众

成年观众的需求。成年人的要求会因学历、经历等不同差异较大。总的来说他们更注重宏观的事物，希望了解科技发展的现状、科技与社会发展的关系、现阶段科学的热点问题以及科技在社会生活中的应用。他们的视野更广阔，希望了解一些与自己工作和生活相关度较小，平时不太了解的科技层面的问题，尤其是一些新领域的发现，以增长自己的见识，开阔眼界。因此，成年观众需要的是知识容量大、反映科技前沿的展览。

### 1.1.1.4 老年观众

老年观众的需求。随着老龄化社会的到来，老年观众已成为科技馆的重要组成部分。老年人有较多的空闲时间，通常把休闲和学习科学知识结合在一起。因此，他们对展厅的环境布置，休息场所的安排非常关注。此外，老年群体对生活科技的发展相对关注，他们常常会与自己的经历作对比，对与生活密切相关的科学知识更感兴趣。由于老年人对于现代科技知识的掌握速度要慢于年轻人，因此在展览设计时可以设计一些适合老年人操作的展品。同时，还可以举办一些以图片、实物为主的短期展览，以满足老年观众对健康的关心，对科技进步的关注。

## 1.1.2 按参观形式划分

根据参观者参观形式的不同可以定位他们不同的需求，因此太为他们制作专属于他们的定制化路线，科技馆的参观形式主要有以下几种：家庭式的参观形式和旅游团的参观形式。

### 1.1.2.1 家庭参观形式

家庭式的观众群体。此类形式通常以家庭为单位，包括家长和年龄较小的少年儿童（初中的学生通常不再用家长陪同），时间多在周末或节假日（多由于平时家长要上班或学生要上学）。

### 1.1.2.2 旅游团参观形式

旅游团式的观众群体。此类形式通常为异地观众，成人多数，也有少年儿童。他们可能有家庭式观众群体的想法和需求，也可能是服从旅行社的安排。总之，他们通常是长途跋涉到科技馆的，参观时间有严格限定。他们需要在限定的时间内，尽可能多地看到最能代表科技馆特点的内容。

### 1.1.2.3 其他观形式

此外还有就是单人式的观众群体。此类形式通常为对科技馆具有非常浓厚的兴趣的观众，



他们只身来到科技馆，其参观一般是带着目的性的，可能对于某个特别的版块感兴趣，并且一般参观时间较长，与讲解员的互动较多。其他形式的观众群体，实际上大多数以结伴而来的为多数。此类观众除了没有严格的时间和纪律要求外，其他自然要求与前两种无太大区别。因此在设置问题时主要从以上这几种参观形式入手，以这几种不同的参观形式定位不同的顾客需求，并为他们提供科学规划的，符合他们需求的参观路线。

在实际参观过程中，参观时长也是很重要的一个因素，大部分的真人讲解其模式是提前确定好的，也就是说在参观过程中他们的参观时长基本是呈现确定性的，因此对于不同的参观者可能并不受用。在上实际参观过程中，如果能够提前知晓顾客的预计参观时间，并且针对其预约时间和其他上述关键信息，一对一的定制路线，这样将会极大的提高用户的体验，并且能够大大缩减讲解过程中的冗余时长来增加机器人的使用效率，提高展馆的客流量。

此外根据国际机器人联合会的统计数据显示，2000 年以前，全球服务机器人数量为 6600 台左右。进入 21 世纪后，服务机器人行业蓬勃发展，经过 20 年的发展，2020 年服务机器人总数已达到 2700 万台以上，为 2000 年服务机器人数量的 4060 倍。由图 1-2 可以看出，2005-2010 年，专用服务机器人的销售量较低，在 2015 年之前整体增速缓慢，之后增长幅度迅速提升。可以看到在 2017 年至今，专用服务机器人一直处于爆炸式增长的过程中，因此科技馆机器人应当抓住本次浪潮，发展优秀产品来迎合市场需求。

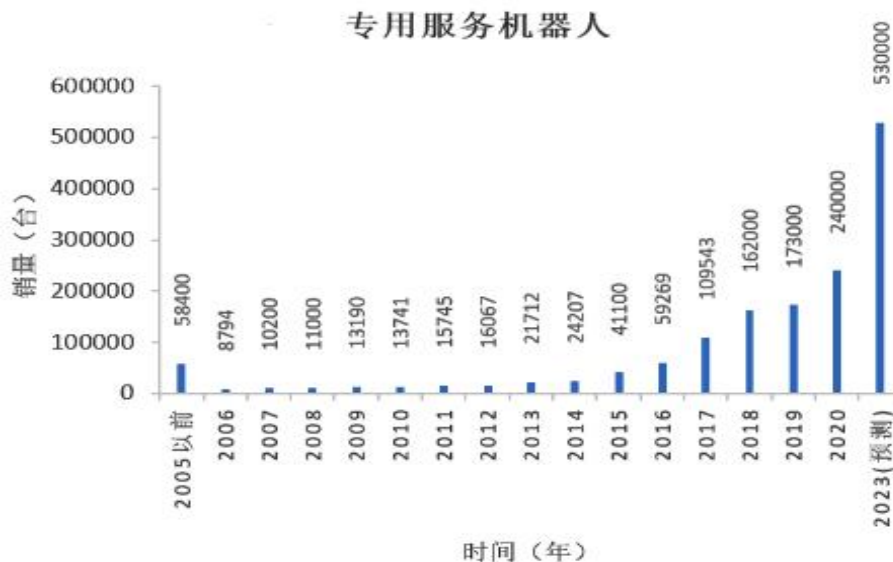


图 1-2 专用服务机器人销量统计

## 1.2 国内外研究现状

2010 年以来，计算机市场飞速发展，2012 年，英国剑桥大学发售了世界上最小的台式机 Raspberry Pi 电脑板，中文译名“树莓派”，该电脑仅有一张信用卡大小。树莓派的出现，代表低成本微型计算机开始进入市场。随着计算机的微型化和性能的提升，微型的计算机可以装入小型机器人并为机器人带来大量的算力，使得机器视觉等所需要的复杂的计算量变得可以接受，同时使得更加智能的服务机器人成本降低，逐渐进入市场。

导游型机器人 Rhino 就是这类机器人的代表，它是由美国 Carnegie Mellon 大学、德国波昂大学及德国 Aachen 科技大学共同研究发展，它的控制软件整合低阶的概率推论与高阶问题的解决方法共同嵌入在一阶逻辑层内，它的硬件由超过 20 个模块所组成，经由 3 台计算机加以平行处理执行命令，同时还有 2 台经由无线以太网网络连结的工作站来做监督，处理相关的工

作事宜,在传感器部分有激光测距仪、超声波传感器、活动式红外线传感器和触碰平板及摄影头等。

### 1.2.1 国外研究现状

欧美国家在服务机器人产品研制开发方面起步较早。欧洲在以康复机器人为代表的服务机器人方面的研究起源于 20 世纪 70 年代中期。当时的机器人多以手操机器人项目为基础,把机械手安装在轮椅或者其他代步工具上,辅助病人康复,增强病人的运动能力。美国的 DeVAR 机器人、加拿大的 Regencies 机器人、法国的 Spartacus 和联邦德国研发技术部支持的 Heidelberg 都属于此类项目。另一类服务机器人则是在机械手上装有力传感器和距离传感器以保证工作安全可靠,可以穿行到各个房间。这类机器人一般工作在未知环境中。由于在未知环境下控制的复杂性以及控制成本等原因,目前这类服务机器人大多处于研究开发阶段。

### 1.2.2 国内研究现状

我国在服务机器人领域的研发起步相对较晚。但在各研究机构的不懈努力下,已经取得了显著的进步。近年来,在国家高技术研究发展计划的支持下,我国在服务机器人基础技术研究和产品研发方面已开展了大量工作,并取得了一定的成绩。目前已初步形成了水下自主机器人、消防机器人、搜救/排爆机器人、仿人机器人、医疗机器人、机器人护理床和智能轮椅、烹饪机器人等系列产品,展示出一定的市场前景。多家研发机构在“迎宾机器人”、“导游机器人”、“清扫机器人”的成果也令人耳目一新。

据国际机器人联合会数据统计,在 2019 年,全球范围内有 890 家企业正在研制服务机器人。在专业服务机器人领域,约 90% 的物流机器人和医疗机器人在欧洲和北美生产,而仅有 10% 在亚洲生产,可见亚洲整体的服务型机器人仍然处于赶超阶段。

### 1.2.3 国内外研究对比

美国目前处于行业前列。2013 年扫地机器人全球累计销量超过 1000 万台。中国在 2018 年扫地机器人累计销量为 560 万台左右。虽然美国服务机器人行业发展较早,技术积累较多,但美国服务机器人市场较小,应用场景也较少。就扫地机器人而言,2019 年中国扫地机器人年销售量已经超越美国两倍以上,这表明中国相对于美国具有更广阔的服务机器人市场。日本在 20 世纪 60 年代末从美国引进机器人技术后,已经成为机器人第一大生产国,且日本的机器人文化氛围较好,日本企业热衷于研发服务机器人。与中国不同的是,日本机器人研发流程主要针对市场,根据需求设计产品。然而 2008 年全球经济危机后,日本经济持续低迷,由于市场环境以及机器人零部件的价格问题等因素,日本服务机器人行业发展也屡遇瓶颈。

## 1.3 项目应用前景

美国斯坦福大学 Kaplan 教授曾评价人工智能、机器人技术的不断发展是为了将人类从事务性工作中解放出来。在 2018 年世界机器人大会上,中国科学院丁汉院士指出,未来机器人技术具有广阔的发展空间,特别是随着新材料技术、人工智能技术等应用科技的快速发展,机器人的明天会更加美好。科技馆作为以展览教育为主的公益性科普教育机构,它在当代的主要功能就是负责对群众的科普,丰富国民的科学知识,它的产生是为满足社会和国家发展的需求,其目的是让公众了解科学,激发科学兴趣,提高公众的科学素养,进而为国家培养创新型人才,因此科技馆的发展在当代尤为重要。随着国家经济飞速发展,近年来国内各类展馆数量快速增长。而增长速度最快的,就要数科技馆了,因此机器人代替传统人工讲解员进行参观讲解市场巨大,此外机器人作为讲解员出现在科技馆中更能体现科技带来的变化,凸显科技馆的科技氛围。

智慧旅游概念的提出使得旅游业找到新的发展点,旅游业作为服务业的龙头产业,必然会因为信息技术的变革以及游客自主旅游的更多需求发生革命性的变化。此外,随着生产生活的不断发展,新兴旅游方式正在快速发展,旅游业如何去满足这部分新兴需求,同样也离不开

自身的现代化，从技术层面说，这里面最重要的就是实现智慧旅游。

2011年7月15日，国家旅游局局长邵琪伟正式提出，旅游业要落实国务院关于加快发展旅游业的战略部署，走在我国现代服务业信息化进程的前沿，争取用10年时间，在我国初步实现智慧旅游。根据《2019年中国机器人市场分析及政策汇编》显示，我国的机器人行业发展势头大好，我国目前已经成为全球最大的机器人市场，并且年均增长达到了28%，其覆盖领域将不断扩大，预计将会达到全行业类别的覆盖。展览馆作为一个普遍的社会公共服务平台，其服务模式将会迎来翻天覆地的变化。移动机器人的应用将会极大提升用户体验，提高服务效率。针对大型展览馆，移动型机器人因较为广阔的服务范围，可以很大程度的减少购入数量，为展览馆减少相应开支。

## 第二部分 项目设计方案

### 2.1 研究开发内容及系统架构

根据对行业痛点的分析以及对市场需求的调研，团队提出了开发一款可提供私人定制服务的智能讲解机器人，实现参观路线定制化以及参观过程动态化。

#### 2.1.1 功能描述

本产品的主要功能为“定制”+“讲解”。“定制”需要做到对参观者身份的核验，并且根据参观者信息提供服务。“讲解”需要做到实时定位，移动导航以及正确行驶规定路线。团队将功能拆分为以下六个部分，包括：获取预约信息、人脸识别解锁、智能语音交互、动态规划路线、自主导航跟随、实时定位讲解。

##### 2.1.1.1 获取预约信息

###### 1.定制化服务

为准确的提供定制化服务，产品需要提前进行预约，预约形式为网上预约。需要用户填写包含参观时间和参观形式在内的参观信息，除此之外，还需要用户提供一张人脸照片。

###### 2.主题化研学服务

如果用户在参观形式一栏选择学生团体，将开启为配合国家“双减”政策的“主题类研学”服务。截至目前为止，我国科技馆数量已达 1000 个，正成为青少年的“第二课堂”。产品将科技馆的内容以学科分类，用互动性更强的参观路线提升知识黏性，有效的发挥科技馆教育与科普的作用。

##### 2.1.1.2 人脸识别解锁

为确保准确匹配预约人，为其提供“私人”参观服务，产品设置唤醒解锁功能，在获取预约信息之后，产品将提取用户提交的照片，做预处理后训练得到人脸数据存储到机器中，在现场识别到预约人脸的时唤醒为其提供服务。

##### 2.1.1.3 智能语音交互

为避免触控式交互有可能带来的疫情传播风险，本产品选择智能语音问答作为交互方式，扩大交互范围和使用人群。

##### 2.1.1.4 动态规划路线

用户到达现场并唤醒机器，机器会再次向用户提问获取具体信息，结合预约时提供的信息，确定最终的参观路线，并为用户提供导航讲解服务。在参观的过程中，可以根据用户需求和场馆内人员密度等信息动态修改路线。

##### 2.1.1.5 自主导航跟随

为保证产品可以正确带领用户到达指定地点并进行讲解，产品需要根据科技馆内部构造进行地图建模。根据路线信息做出正确引导，并在特定位置暂停或讲解。在导航过程中，产品需要实时检测用户是否跟跟随，若用户不再跟随，则停留提醒用户并等待，等用户回到检测范围内时，继续导航。

##### 2.1.1.6 实时定位讲解

为了给用户提供正确的讲解内容，产品需要实时定位，了解自己的位置信息。因为作为讲解机器人，讲解内容的播放是必不可少的，所以我们将定位与讲解内容储存相结合，利用计算机视觉识别二维码，二维码内包含当前位置信息以及讲解内容，实现实时定位的同时，

降低成本。

## 2.1.2 系统总设计

在方案设计上，团队选择服务器+多下位机机器人的形式：

### 2.1.2.1 服务器

#### (1) 服务器选择

考虑到科技馆面积较大，且来访者的密度会根据社会热点而变化，比如 2022 年 2 月 4 日，天问一号从火星向北京发来冬奥会祝贺之后，包含航空航天主题的路线参观人数增多。所以，如果选择单个服务器，将会面临体积较大，成本较高，更新较快且还有可能造成某个节点数据流过大而导致数据延迟。所以，我们依托 Intel GNS-V40 的边缘计算功能，在每个展馆内放置一个服务器，搭建边缘计算服务器网络，每个服务器控制三个私人定制讲解机器人，当某个节点的计算压力较大时，分担给其余的服务器，保证数据处理的实时性，降低成本。

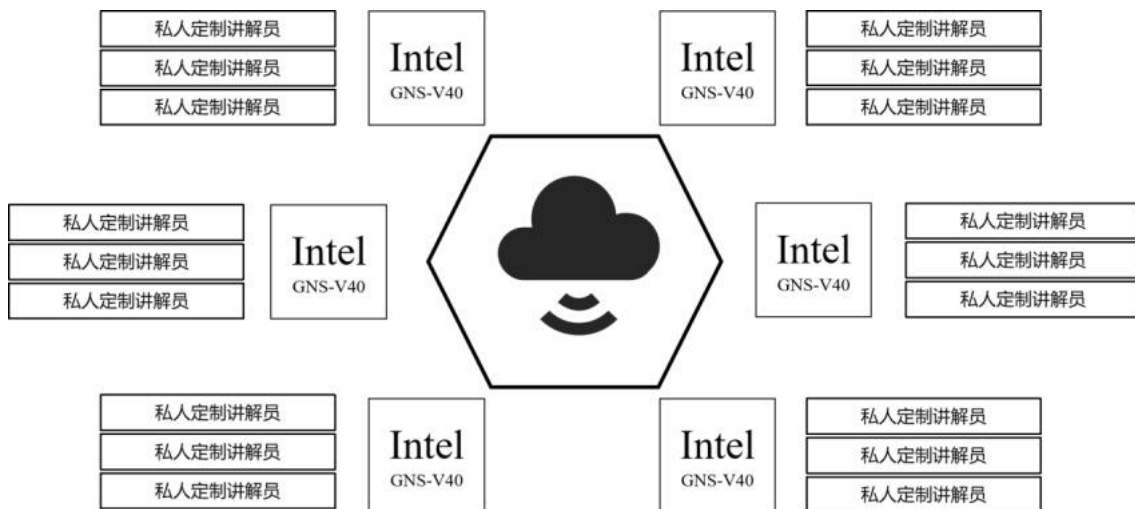


图 2-1 系统总设计图

#### (2) 服务器功能

Intel GNS-V40 作为服务器有两个功能：

- 接受下位机传回的图像信息进行数据处理，解锁启动机器。
- 接受下位机传回的实时视频信息进行数据处理，实现机器用户跟随。
- 接受语音信息并识别与用户进行智能交互，获取动态参观信息。
- 根据动态参观信息寻找最优路径，并返回路径指令给下位机进行导航。

### 2.1.2.2 下位机

#### (1) 下位机选择

考虑到下位机不仅需要接受传感器信息，还需及时将信息传输给服务器并接受服务器做出判断，团队选择树莓派 4B 作为下位机。

#### (2) 下位机功能

树莓派作为下位机具有以下几个功能：

- 接收图像信息、传感器数据以及实时视频信息并及时传输至服务器。
- 接收服务器传回的路线指令并控制相应电机实现移动。

### 2.1.3 系统硬件设计

对于单个服务器来说，需要 3 个模块，分别是：语音输入模块、视觉处理模块以及 AI 语音处理模块。

对于单个下位机机器人来说，需要包含 4 各模块，分别是：巡线模块、驱动模块、图像

输入模块以及红外避障模块。

单个产品硬件框图如下：

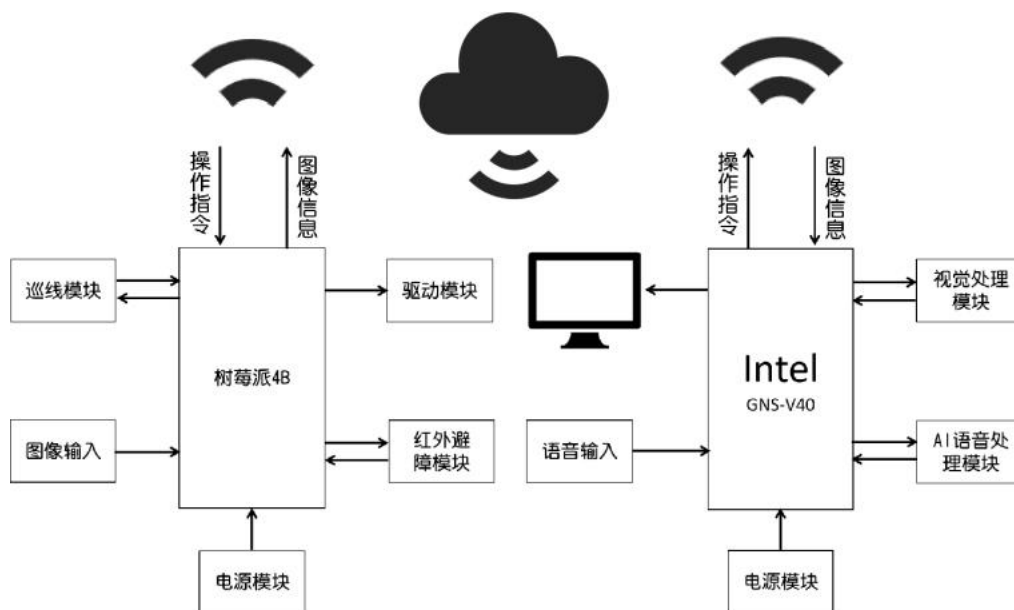


图 2-2 硬件框图

## 2.1.4 系统软件架构

### 2.1.4.1 语音唤醒

在每次调用程序时需要使用唤醒词“小佳慧”对系统进行启动，一方面为了避免一开启系统就进入后面的功能而显得突兀，另一方面也增加了系统的可交互性和人性化体验。若识别不到用户说话则提示“小佳慧没听懂您说什么，请再说一次”；若用户的唤醒词错误则提示“唤醒词识别失败，请重新输入”；直到用户唤醒成功，系统会提示“唤醒词识别成功”。

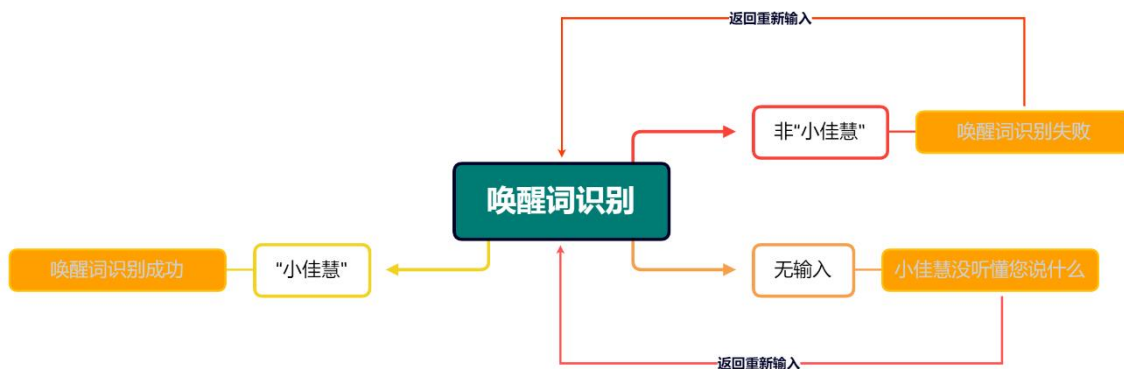


图 2-3 语音唤醒流程图

### 2.1.4.2 人脸解锁

在获取预约信息中的照片后，将照片做光学预处理和几何预处理，以扩大数据集。之后在边缘终端 GNS-V40 上训练人脸的框架，后使用 Resnet 残差网络进行识别，在保障识别率的基础上让网络尽快收敛。在识别时设置 0.2 的欧氏距离阈值，若是识别结果大于 0.2 则为识别失败，语音播报“人脸认证失败，请重新验证”；若是识别结果小于 0.2 则为识别成功，释放摄像头，语音播报“人脸验证成功，小佳慧竭诚为您服务”。

### 2.1.4.3 用户跟随

产品使用深度摄像头来获取周围的视觉信息，根据人像景深距离数据判断用户是否跟随

机器，当距离数值大于 0.5m 时，机器暂停运行并给出语音提示：“请跟随小佳慧不要走丢哦”，当人像回归检测范围并距离数值小于 0.5 时，继续运行导航。

#### 2.1.4.4 人机交互

在人脸识别时，我们识别到预先有预约过的用户人脸就把预约的展馆信息填入，减少不必要的交互内容和冗余选择，节省用户的体验时长，把更多的交互放在参观上，提高系统的反应速度。语音播报内容：“根据您的预约信息，将为您推荐定制化服务/主题类研学参观服务”。在确认好预约信息之后开始进入服务阶段，在提供的所有服务当中若用户选择“参观展馆”，将进入到另一个交互模块—参观路线选择，根据提前调问题来询问分析用户属性，进而为其生成专属于他的定制化路线。

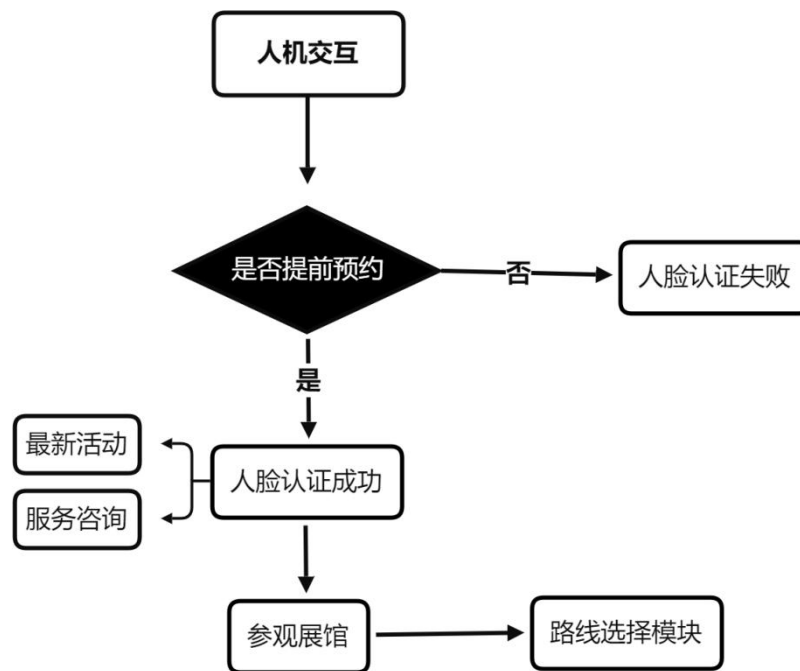


图 2-4 人机交互流程图

#### 2.3.4.5 动态路径规划

使用动态规划数学模型，根据科技馆的路线信息，搭建动态规划路线模型，将权重点分为三层，第一层为预约信息层，根据用户的预约信息填写对应权重；第二层为语音交互层，根据用户参观现场过程中的信息填写对应权重；第三层为环境信息层，获取展馆内部和两个参观之间路线上的监控信息，根据 yolov4 算法分割人群计算得到人群密度，根据人群密度填写对应权重，避免在参观过程中发生拥堵。

其中，语音交互层和环境信息层权重可动态更改，通过 Dijkstra 算法计算最优路径，实现动态路径规划。

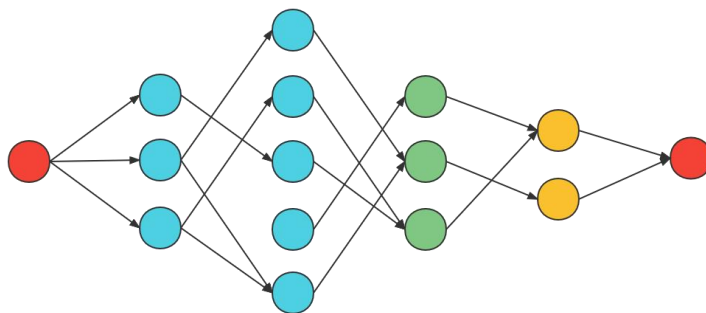


图 2-5 动态规划路线模型

#### 2.1.4.6 信息传输

人脸识别解锁、智能语音交互、动态规划路线、自主导航跟随、实时定位讲解功能都在服务器中完成计算，当用户选择“参观展馆--主题类研学/定制化服务--路线”后，语音播报“xx 路线涉及 xx 展馆的内容”，并传输运行指令至下位机，由下位机电机转动指令至驱动底板，机器开始运行。在机器运行的过程中，上下位机进程保持挂起，下位机摄像头不断扫描前方二维码传送回来的扫描信息并将其发送回上位机，上位机若是收到下位机信息则做出对应的操作。上位机收到数字 x（1~10）则播报“即将进入 x 馆”，收到数字 xx（1~10）则播报“正在讲解的是 x 馆的内容”，若是收到字母则播报“x 馆的内容已经全部讲解完了，是否更换路线”。此时进程挂起，等待用户的语音输入，若是不更换则上位机向下位机发送原指令，更换则从用户的语音中解析出指令并向下位机发送。

## 2.2 技术关键

### 2.2.1 二维码定位

项目当中二维码来确定位置信息，使用摄像头来获取周围的视觉信息。在项目过程中，前置摄像头与树莓派直接相连，通过调用机器摄像头功能，识别到位于地面上的立面二维码后识别出相应结果，以此来区分不同讲解内容和实现机器人的定位功能，二维码信息将在树莓派上识别，省去了图像信息的传输过程，在识别到信息之后与位置库进行比对获知准确位置，Intel 板通过无线局域网读取位置信息，进行播报讲解内容。对于不同讲解内容，设备会通过二维码来调用不同的语料库来实现针对不同展馆的不同内容讲解。在二维码识别过程中，采用了 ZBar 算法。ZBar 的图像处理采用了二阶微分边缘检测，其包含了：

(1) 对运动均值后图像做二阶差分、一阶差分。边缘判定规则为二阶导数为零的位置是一阶时的最大值或最小值，因此认为是边缘点；对二阶导数符号发生变化的地方一定存在边缘点，由于进行的是差分运算，需要进行插值找到导数为零的近似位置。另外还需要满足阈值。一阶差分主要用来判定边界两侧像素大小变化。

(2) 计算边缘阈值。对一阶差分计算阈值利于抗噪。取得当前阈值，如果当前阈值小于最小阈值，或者边缘宽度为 0，返回最小阈值。求相对阈值，上一次的阈值乘以当前边缘和上一次边缘之间的距离与再上一个距离的比值。如果上一次的阈值大于相对阈值，则用上一次的阈值减去相对阈值，结果如果大于最小阈值，则返回这个结果，否则返回最小阈值。

(3) 边缘处理。对满足边界判定规则的点进行边缘处理。

(4) 更新新的边缘。通过线性插值算法更新新的边缘。更新阈值，对运动均值后的图像一阶差分，然后乘以一个常数。

通过以上算法获知位置储存在二维码当中的信息，进而实现二维码定位的功能。

### 2.2.2 自主导航运行



### 2.2.2.1 红外巡线导航

#### (1) 红外线传感器结构

红外线传感器包括光学系统、检测元件和转换电路三大部分。光学系统按结构不同可分为透射式和反射式。检测元件按工作原理可分为热敏检测元件和光电检测元件。热敏检测元件中最常见的就是热敏电阻，热敏电阻受到红外线辐射时温度升高，电阻发生变化，通过转换电路变成电信号输出。

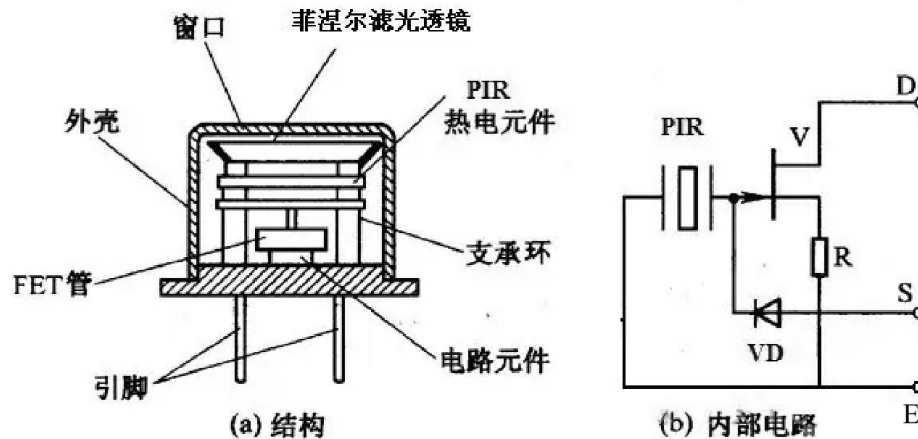


图 2-6 红外线传感器结构

#### (2) 红外传感器分类

##### • 热敏电阻型传感器

热敏电阻是由锰、镍、钴的氧化物混合后烧解而成的，热敏电阻一般制成薄片状，当红外辐射照射在热敏电阻上，其温度升高，电阻值减少。测量热敏电阻值变化的大小，即可得知入射的红外辐射的强弱，从而可以判断产生红外辐射物体的温度。

##### • 热电偶型传感器

热电偶是由热电功率差别较大的两种材料构成。当红外辐射到这两种金属材料构成的闭合回路的接点上时，该接点温度升高。而另一个没有被红外辐射辐照的接点处于较低的温度，此时，在闭合回路中将产生温差电流。同时回路中产生温差电势，温差电势的大小，反映了接点吸收红外辐射的强弱。利用温差电势现象制成的红外传感器称为热电偶型红外传感器，因其时间常数较大，相应时间较长，动态特性较差，调制频率应限制在 10HZ 以下。

##### • 莱气动型传感器

高莱气动型传感器是利用气体吸收红外辐射后，温度升高，体积增大的特性，来反映红外辐射的强弱。它有一个气室，以一个小管道与一块柔性薄片相连。薄片的背向管道一面是反射镜。气室的前面附有吸收模，它是低热容量的薄膜。红外辐射通过窗口入射到吸收模上，吸收模将吸收的热能传给气体，使气体温度升高，气压增大，从而使柔镜移动。在室的另一边，一束可见光通过栅状光栏聚焦在柔镜上，经柔镜反射回来的栅状图像又经过栅状光栏投射到光电管上。当柔镜因压力变化而移动时，栅状图像与栅状光栏发生相对位移，使落到光电管上的光量发生改变，光电管的输出信号也发生变化，这个变化量就反映出入射红外辐射的强弱。这种传感器的特点是灵敏度高，性能稳定。但响应时间性长，结构复杂，强度较差，只适合于实验室内使用。

##### • 热释电型传感器

热释电型传感器是一种具有极化现象的热晶体或称“铁电体”。铁电体的极化强度(单位面积上的电荷)与温度有关。当红外线辐射照射到已经极化的铁电体薄片表面上时，引起薄片温度升高，使其极化强度降低，表面电荷减少，这相当于释放一部分电荷，所以叫做热释电

型传感器。如果将负载电阻与铁电体薄片相连，则负载电阻上便产生一个电信号输出。输出信号的大小，取决于薄片温度变化的快慢，从而反映入射的红外辐射的强弱。由此可见，热释电型红外传感器的电压响应率正比于入射辐射变化的速率。当恒定的红外辐射照射在热释电传感器上时，传感器没有电信号输出。只有铁电体温度处于变化过程中，才有电信号输出。所以，必须对红外辐射进行调制(或称斩光)，使恒定的辐射变成交变辐射，不断的引起传感器的温度变化，才能导致热释电产生，并输出交变的信号。

- 外光电传感器

当光辐射在某些材料的表面上时，若入射光的光子能量足够大时，就能使材料的电子逸出表面，这种现象叫外光电效应或光电子发射效应。光电二极管、光电倍增管等便属于这种类型的电子传感器。它的响应速度比较快，一般只需几个毫秒。但电子逸出需要较大的光子能量，只适宜于近红外辐射或可见光范围内使用。

- 光电导传感器

当红外辐射照射在某些半导体材料表面上时，半导体材料中有些电子和空穴可以从原来不导电的束缚状态变为能导电的自由状态，使半导体的导电率增加，这种现象叫光电导现象。利用光电导现象制成的传感器称为光电导传感器，如硫化铅、硒化铅、碲化铟、碲隔汞等材料都可制光电导传感器。使用光电导传感器时，需要制冷和加一定的偏压，否则会使响应率降低，噪声大，响应波段窄，以致使红外线传感器损坏。

- 光生伏特传感器

当红外辐射照射在某些半导体材料的PN结上时，在结内电场的作用下，自由电子移向N区，如果PN结开路，则在PN结两端便产生一个附加电势，称为光生电动势。利用这个效应制成的传感器或PN结传感器。常用的材料为砷化铟、碲化铟、碲化汞、碲锡铅等几种。

- 光磁电传感器

当红外辐射照射在某些半导体材料表面上时，半导体材料中有些电子和空穴将向内部扩散，在扩散中若受强磁场的作用，电子与空穴则各偏向一方，因而产生开路电压，这种现象称为光磁电效应。利用此效应制成的红外传感器，叫做光磁电传感器。

光磁电传感器不需致冷，响应波段可达 $7\mu\text{M}$ 左右，时间常数小，响应速度快，不用加偏压，内阻极低，噪声小，有良好的稳定性和可靠性。但其灵敏度低，低噪声前置放大器制作困难，因而影响了使用。

### (3) 红外四路巡线传感器

本产品选择红外四路巡线传感器，当检测到黑色时，传感器输入电平为0，反之则为一，通过四路电频共16种组合，产品可以实现完整的巡线功能。传感器电路图如下：

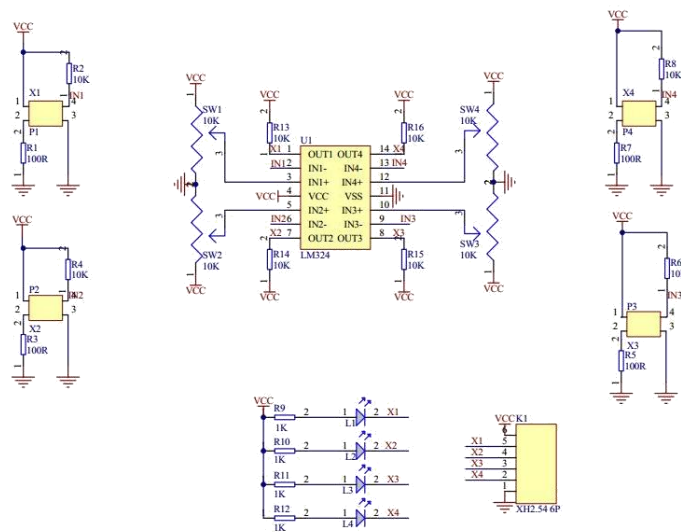


图 2-7 红外四路巡线传感器电路图

(4) 巡线导航算法

将传感器的 16 种分类情况与机器运动情况相匹配,当遇到路口即四个输入电平全为 0 时,机器停止运动并开启摄像头获取定位信息。若当前位置在规划路线中,则前进进入展馆,若不在当前规划路线中,则转弯前往下一个展馆,下图为算法流程图:

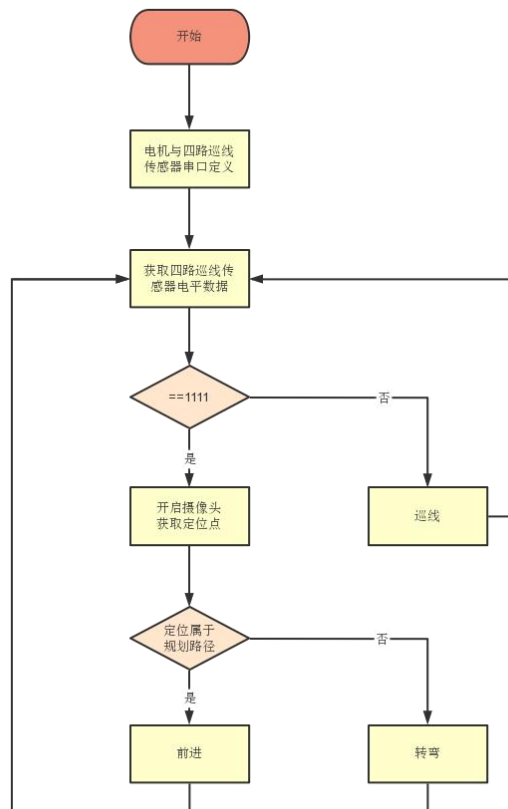


图 2-8 巡线导航算法流程图

2.2.2.2 激光雷达 SLAM 建图定位与导航

(1) SLAM

即时定位与地图构建 (Simultaneous Localization and Mapping, SLAM) 技术可以很精确地实现环境的地图构建、定位以及多点导航。目前 SLAM 技术可以分为激光 SLAM(Lidar SLAM) 和视觉 SLAM (Visual SLAM)，激光 SLAM 采用的传感器为激光雷达，而视觉 SLAM 则采用深度摄像头。激光 SLAM 技术较为成熟、误差少，且足以满足当前环境的使用。

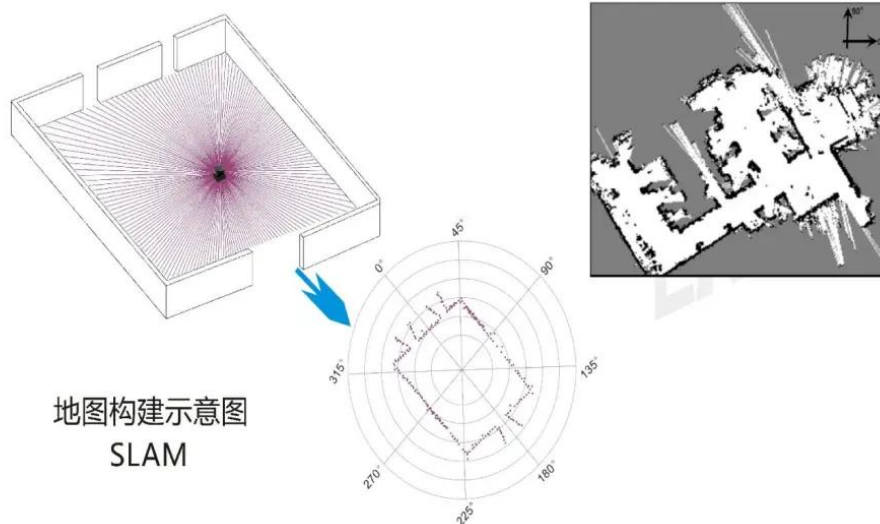


图 2-9 SLAM 地图构建示意图

### (2) SLAM 系统

SLAM 系统一般分为五个模块，包括传感器数据、视觉里程计、后端、建图及回环检测。

- 传感器数据：主要用于采集实际环境中的各类型原始数据。包括激光扫描数据、视频图像数据、点云数据等。
- 视觉里程计：主要用于不同时刻间移动目标相对位置的估算。包括特征匹配、直接配准等算法的应用。
- 后端：主要用于优化视觉里程计带来的累计误差。包括滤波器、图优化等算法应用。
- 建图：用于三维地图构建。
- 回环检测：主要用于空间累积误差消除。

### (3) 工作流程

传感器读取数据后，视觉里程计估计两个时刻的相对运动 (Ego-motion)，后端处理视觉里程计估计结果的累积误差，建图则根据前端与后端得到的运动轨迹来建立地图，回环检测考虑了同一场景不同时刻的图像，提供了空间上约束来消除累积误差。

## 2.2.3 自主避障

自主避障是通过红外线传感器探头利用被检测物对发射红外光的反射或遮挡，光敏接收管接收前方物体反射光，从而判断有障碍物。红外线避障探头电路图如下图所示，在实际进行测试当中，当前方有障碍物时，红外线避障指示灯引脚电压会变为高电压，从而使其变亮，若无障碍物则一直处于熄灭状态。在系统逻辑上面，当指示灯处于熄灭状态时，树莓派对应引脚为高电压，逻辑数字为 1，当指示灯处于明亮状态时，树莓派的逻辑数字为 0。在进行探头测试当中通过调节探头上光敏电阻的大小来控制实际测量的灵敏度，其测量结果如下。

表 2-1 左侧红外线探头测量结果

距离	2cm	5cm	8cm	16cm	32cm

第一次测量	亮	亮	亮	暗	暗
第二次测量	亮	亮	暗	暗	暗
第三次测量	亮	亮	暗	暗	暗

表 2-2 右侧红外线探头测量结果

距离	2cm	5cm	8cm	16cm	32cm
第一次测量	亮	暗	暗	暗	暗
第二次测量	亮	亮	暗	暗	暗
第三次测量	亮	亮	暗	暗	暗

经过调试，把避障范围精确到 5cm-8cm 之间，这样在确保能够正常通行的前提下可以实现机器的自主避障功能。

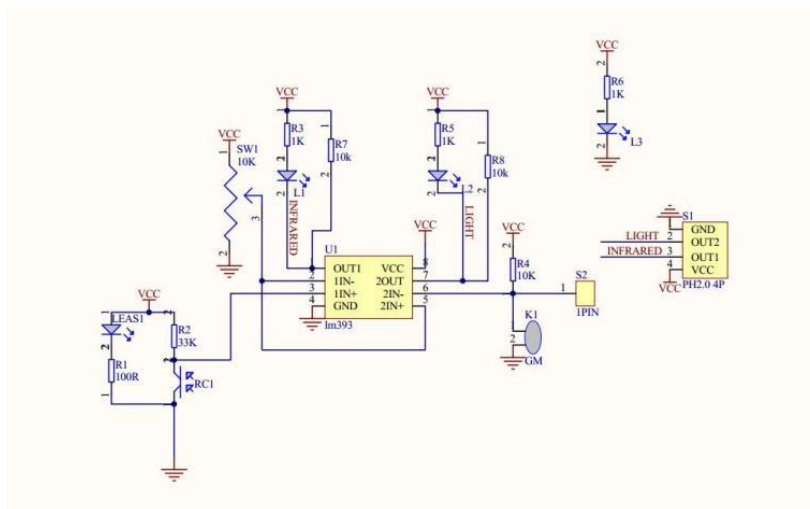


图 2-10 红外线避障探头电路图

## 2.2.4 智能语音问答

### 2.2.4.1 语音识别

在语音识别上调用目前市面上成熟的百度语音 aip 进行识别，在保证识别准确率的基础上使用算法对识别出的字符串进行分割、转拼音操作。字符串分割操作有效避免语音频段中部分文字识别出错而导致整个语音频段无效的情况出现，提高对音频段和关键词识别成功率，从而大大提升系统的可交互性。转拼音操作主要是针对同音字和谐音字的情况，调用 Python 的 xpinyin 库，防止用户说出关键词后因为语音库中的同音词和谐音词干扰到系统进程，识别过程中的有效提升系统的认证成功率。

### 2.2.4.2 模糊匹配

在保留语音识别的声音模型后，团队还考虑到我国的部分方言存在前后鼻音的差异，所以在匹配的时候，使用模糊匹配，使前后鼻音都可以识别。



图 2-11 语音模糊匹配图

## 2.2.5 智能人脸解锁

### 2.2.5.1 预处理

#### (1) 光学预处理

考虑到不同场景下环境因素的不同，尤其是光强对人脸识别准确率的影响，团队将预约照片做光学预处理，将输入照片通过傅里叶变化降噪后，为其赋予 4 种曝光条件，得到不同曝光条件下的照片，补充至数据集中，以提供人脸在不同光强下的识别率。

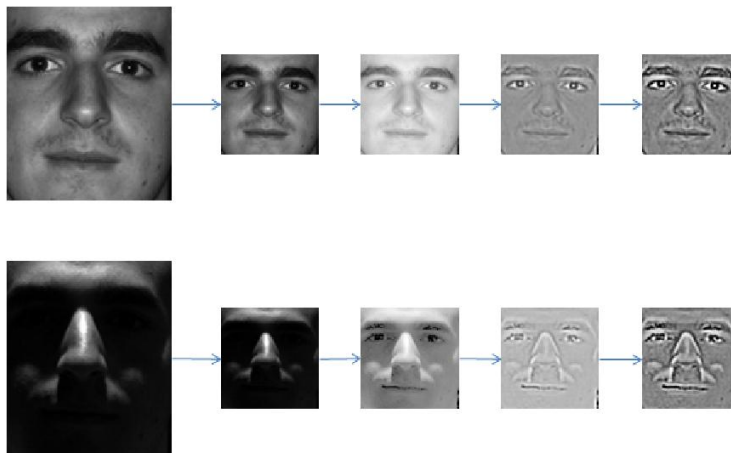


图 2-12 光学预处理图

#### (2) 几何预处理

将输入照片进行左旋、右旋、前旋、后旋、放大、缩小操作，设置 3 档不同的角度和缩放比例梯度，得到不同几何变化下的照片，补充到数据集中，以提供人脸在不同角度下的识别率。

### 2.2.5.1 dlib 模块简介

在人脸识别上项目调用现有主流的深度学习开发工具 dlib，以及其下的 Resnet 网络模型 dlib\_face\_recognition\_resnet\_model\_v1。Dlib 是一个包含机器学习算法的 C++ 开源工具包，已被广泛机器人，嵌入式设备，移动电话和大型高性能计算环境。等领域。Dlib 的主要特点有：

#### (1) 文档齐全

不像很多其他的开源库一样，Dlib 为每一个类和函数提供了完整的文档说明。同时，还提供了 debug 模式；打开 debug 模式后，用户可以调试代码，查看变量和对象的值，快速定位错误点。另外，Dlib 还提供了大量的实例，本项目就借用其网络实例模型 dlib\_face\_recognition\_resnet\_model\_v1。

#### (2) 高质量的可移植代码

Dlib 不依赖第三方库。可移植性好，可用在 Window、Mac OS、Linux 系统上。

(3) 提供大量的机器学习和图像处理算法

包括但不限于基于 SVM 的分类和递归算法、针对大规模分类和递归的降维方法、聚类、多层感知机、相关向量机、与支持向量机相同的函数形式稀疏概率模型，对未知函数进行预测或分类。稀疏概率模型的训练是在贝叶斯框架下进行的，与 SVM 相比，不需要估计正则化参数，其核函数也不需要满足 Mercer 条件，只需要更少的相关向量。

2.2.5.2 Resnet 网络架构介绍

项目选用 dlib 这个模块是因为其包含 Resnet 网络架构模型，Resnet 模型提出早、开发成熟且易于上手，相比于之前的网络，不论是架构上还是收敛速度上都有很大改进。下图是 Resnet34 的具体网络架构和 Resnet18/34/50/101/152 的各项参数图：

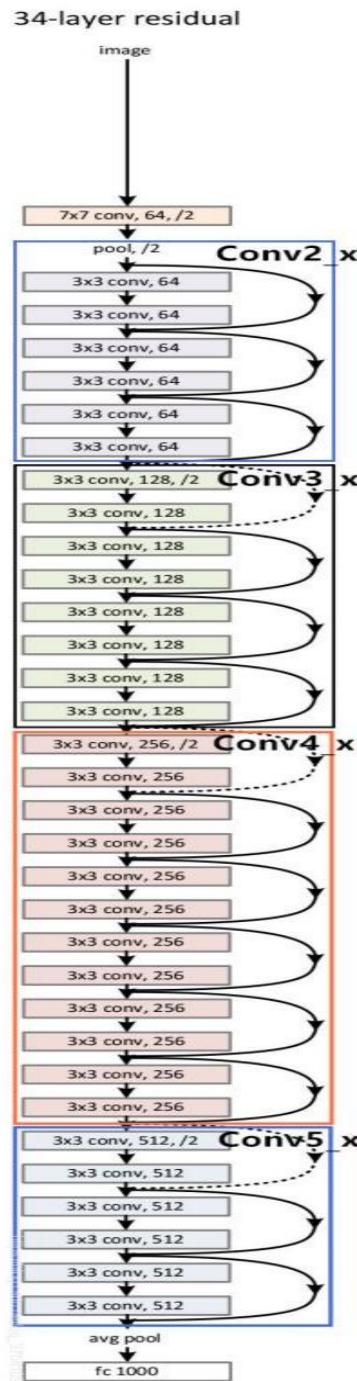


图 2-13 Resnet34 网络结构图

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2.x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3.x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4.x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5.x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

图 2-14 Resnet18/34/50/101/152 各项参数图

Resnet 最先提出时是为了解决神经网络中梯度消失的(又称“梯度退化”)问题。梯度消失反映到实际当中就是随着神经网络训练和测试层数的增加，最终识别效果反而变差的情况。下图显示的传统 CNN 网络在实际测试时出现的梯度退化问题，表现为随着自变量网络层数的增加，图像识别的错误率并没有得到有效改善。

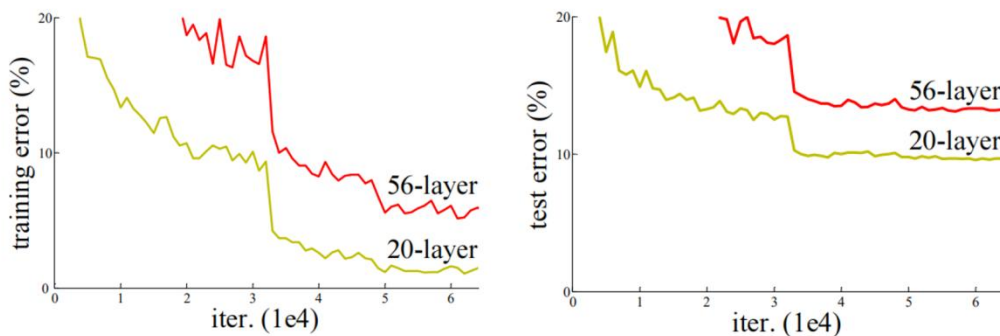


图 2-15 传统 CNN 网络在测试时梯度消失

ResNet 网络结构在 2015 年由何凯明提出。它源自于这样一个想法。从直觉上来说，如果一个神经网络变深了以后，它的学习效果变差，那问题应该出在加深的部分，换句话说，是新加入的层次让网络变差了。ResNet 开创性地提出了 residual 结构(又称残差结构，如图 4 图 5 所示)来减轻网络退化的问题，具体实现为在层数和识别率达到一定。下图是使用 residual 结构的卷积网络，可以看到随着网络的不断加深，效果并没有变差，反而变的更好了。

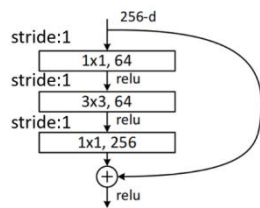


图 2-16 残差结构示意图 1

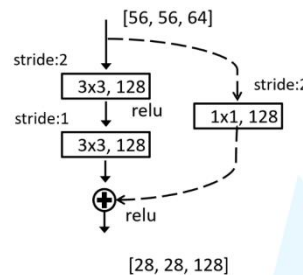


图 2-17 残差结构示意图 2

在 ResNet 之前，也有一些针对此问题的研究成果，比如批标准化(Batch normalization)，



在一定程度上减小了深层次神经网络的误差，改善了精度，但依然没有能比较彻底地解决这个问题。所以那个时候的“深层次”神经网络，其实都不是很深。

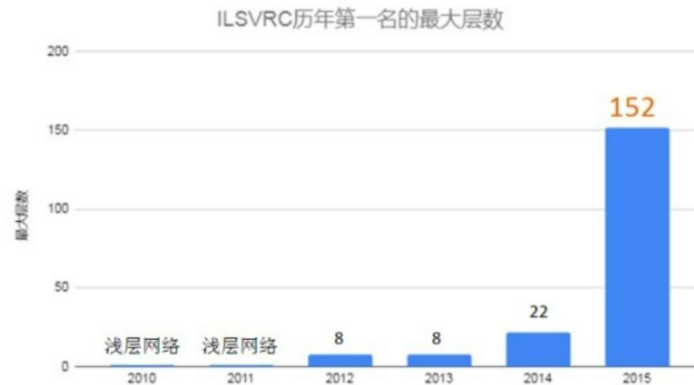


图 2-18 ILSVRC 竞赛 2010 - 2015 年第一名的网络数量

如图 6 所示，在 ILSVRC (ImageNet Large Scale Visual Recognition Challenge, ImageNet) 2010-2014 年历年第一名的网络模型当中，最深的是 2014 年的 22 层 (GoogLeNet, GoogleNet 在赛后又陆续推出 v2v3v4 版本，其性能均有所提升，这里仅讨论比赛时的 v1 版本)。2015 年 ResNet 出现之后，模型深度上升到 152 层，真正在网络层次上飞跃。

### 2.2.5.3 Resnet 残差结构的公式推导

为什么残差学习相对更容易，从直观上看残差学习需要学习的内容少，因为残差一般会比较小，学习难度小点。不过我们可以从数学的角度来分析这个问题，首先残差单元可以表示为：

$$y_l = h(x_l) + F(x_l, W_l)$$

$$x_{l+1} = f(y_l)$$

其中  $x_l$  和  $x_{l+1}$  分别表示的是第  $l$  个残差单元的输入和输出，注意每个残差单元一般包含多层结构。F 是残差函数，表示学习到的残差，而  $H(x_l) = x_l$  表示恒等映射，相较于学习训练后得到输出，恒等映射中残差  $F = 0$ ，这样可以在层数很深时保障识别成功率的基础上很快输出本轮的训练结果。F 是 ReLU 激活函数。基于上式，我们求得从浅层 1 到深层 L 的学习特征为：

$$\frac{\partial loss}{\partial x_l} = \frac{\partial loss}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial loss}{\partial x_L} \cdot \left( 1 + \frac{\partial}{\partial x_L} \sum_{i=l}^{L-1} F(x_i, W_i) \right)$$

### 2.2.6 动态路径规划

#### 2.2.6.1 权重动态赋予

建立对应表，将传感器数据、预约信息数据以及语音交互数据与数值对应，并将其存放在 sensor.txt 文件中，新数据传入时覆盖元数据。动态规划路线模型搭建为矩阵形式，矩阵持续读取文件获取最新数据，实现权重动态赋予。

#### 2.2.6.2 Dijkstra 算法

##### (1) 描述

是由荷兰计算机科学家 Edsger Wybe Dijkstra 在 1956 年发现的算法，戴克斯特拉算法使用类似广度优先搜索的方法解决赋权图的单源最短路径问题。Dijkstra 算法原始版本仅适用于找到两个顶点之间的最短路径，后来更常见的变体固定了一个顶点作为源结点然后找到该顶点到图中所有其它结点的最短路径，产生一个最短路径树。本算法每次取出未访问结点中距离最小的，用该结点更新其他结点的距离。

(2) 算法伪码

**Input :** Directed graph  $G = (V, E, W)$  with weight

**Output :** All the shortest paths from the source vertex  $s$  to every vertex  $v_i$   
 $\in V \setminus \{s\}$

```
1 :  $S \leftarrow \{s\}$ 
2 :  $\text{dist}[s, s] \leftarrow 0$ 
3 : for  $v_i \in V - \{s\}$  do
4 :    $\text{dist}[s, v_i] \leftarrow w(s, v_i)$ 
   (when  $v_i$  not found,  $\text{dist}[s, v_i] \leftarrow \infty$ )
5 : while  $V - S \neq \emptyset$  do
6 :   find  $\min_{v_j \in V} \text{dist}[s, v_j]$  from the set  $V - S$ 
7 :    $S \leftarrow S \cup \{v_j\}$ 
8 :   for  $v_i \in V - S$  do
9 :     if  $\text{dist}[s, v_j] + w_{j,i} < \text{dist}[s, v_i]$  then
10 :        $\text{dist}[s, v_i] \leftarrow \text{dist}[s, v_j] + w_{j,i}$ 
```

## 第三部分 测试与验证

### 3.1 语音模块

#### 3.1.1 字符串分割识别

在测试与验证的过程中，由于项目的交互系统没有采用触屏式方案，在交互上语音模块的性能需求突出。考虑到百度提供的 aip 识别精确度和高效性，以及其在市面上推行的时间长、接口成熟、反应速度快，因此起初项目的重心放在智能机器巡线和动态路径规划上，对语音部分的功能直接调用其 aip 取回识别结果即可。

随着项目推进，语音交互的需求越发明显。如果只是针对同一音频段的只识别一次则识别结果中的某个汉字识别出错就会导致整个音频段报错；此外，在实际测试时总是无可避免地会遇到各种各样的环境音效和背景音效，若是对 aip 的识别结果不加修改则会对系统的运行进程产生较大影响。考虑到交互过程中对安全性的需求不大，小组对此的解决方案是，对目标关键词进行拆分，只要检测到关键词中的一个即可识别成功。经过修改调整之后的语音识别代码运行顺畅，在获取语音信息给下位机发送指令时效果比未经修改的代码语音识别效果好出很多；目前就算在说话过程中遇到环境杂音或是背景音效也能够取得较好的识别效果。例如小组在给游客的主题研学类方案当中有参观路线名为“电磁基础”“虚拟技术”“航空航天”的主题。在处理字符串时可讲预设答案分到一个 theme 列表当中：

```
theme = ["电", "磁", "基", "础",  
"虚", "拟", "技", "术",  
"航", "空", "航", "天"]
```

在识别时为按顺序读取列表中每个字符是否在语音片段当中，若是出现则取该次语音片段为这个关键字对应的主题作为识别结果。

#### 3.1.2 文字转拼音

在遇到多项选择和长字符串时使用字符串分割识别效果较为明显。可是当遇到单项识别结果、较短字符串时，字符串分割识别法就比较难以解决此类语音互动的需求。为此，小组提出的方案是文字转拼音。只需调用 Python 的 xpinyin 库文件，对 aip 识别回来的结果进行转拼音处理。同样对目标关键词的拼音进行拆分处理，在识别时逐拼音进行读取，判断目标拼音中是否有出现在语音段中。经过测试，文字转拼音可以很好地解决诸如谐音字和同音字的问题。例如小组在开启系统时的唤醒词“小佳慧”，经常被识别为“肖佳慧”“小家伙”等。有时关键词“小佳慧”中只有 1~2 个字能够识别成功，这种时候把小佳慧的文字转为拼音“xiao”“jia”“hui”，再逐词去识别认证，往往“xiao”“jia”的音识别成功率为最高，“hui”的成功率较低。

### 3.2 电机调整

在机器实际的运行过程当中，除软件之外，硬件方面也会遇到各种各样的问题。相较于软件可以在 IDE 和环境当中调整，若是硬件出现问题则较为难以调整和适配。鉴于小组在演示时使用的是机器作为机器人雏形，载重量上较为吃紧，经常会出现行驶尤其是转弯时电压不

足、机器无法启动等问题。对此小组的调整方案是将无用的传感器拆除，调慢机器行驶速度以确保其能够顺利行驶。

### 3.3 局域网通信

#### 3.3.1 模拟键盘输入

在上位机给下位机发送指令时，为实现全流程自动化交互，小组的方案是在同一片局域网下利用上位机的 Xshell 链接上下位机的终端，然后运行到规定地点时在下位机终端内使用 Python 的 `pyautogui` 和 `pynput` 库模拟电脑键盘鼠标输入对应指令，从而使下位机收到指令后驱动电机运行。

#### 3.3.2 人脸识别解锁

在预先查看到下位机的终端 IP 地址后，将其地址记为 URL。只要在同一片局域网内即可在 Python 代码中设置 `cap = cv2.VideoCapture(url)` 即可获取到树莓派摄像头中的实时画面。之后可有效识别人脸并解锁机器。

#### 3.3.3 收发指令

在上位机的 Python 代码内调取 `paramiko` 库文件，生成 Transport 和 SFTP 实例，输入下位机的 IP 地址和账户密码，在上位机里用一个 TXT 文件收取下位机发来的指令，后上位机做出相应响应。下位机摄像头拍摄到二维码后读取二维码内容，通过 SSH 协议向上位机发送信息。

### 3.4 动态路径规划

产品可及时将信息对应数值存贮到 `sensor.txt` 文件中并通过生成 Transport 和 SFTP 实例，输入下位机的 IP 地址和账户密码，实时读取文件填写至模型矩阵中。

## 第四部分 预期目标

项目预计通过封闭式路径的博物馆模型来模拟最终效果，其中用颜色条带固定在地面后作为讲解过程中的行动路径。在人机交互方面，用户可以进行参观展馆、讲解互动。二维码在机器人运行中来进行位置定位和展板信息获取。通过人脸识别来进行服务授权以及在讲解过程中针对个体的互动。在进行参观之前通过数据收集提前获知用户基本信息已经基本需求来为用户提前规划路径，根据提前调研得知影响参观者观看体验的因素来设置问题，根据用户不同需求分层次提前制定符合用户需求的参观路线，使整个参观规划更为科学。在运行过程中实现动态路径规划，为用户规划更科学的参观路线，这主要是通过询问用户的基本信息为用户规划基础路线，在用户初步参观之后提供符合参观主题的问题，根据用户回答答案来为用户重新规划新路径。基于以上方法最终实现定制化的讲解服务。其次在面对“双减”政策的实施，研学活动成为素质教育领域的重要赛道之一，科技馆作为整个研学体系当中比较突出的一个活动点，我们在本次的项目当中也设计了专属于学生群体的预约渠道，分期提供不同主题的研学参观路线，最终可以实现在既定科学规划路线之下的定制化服务，为学生群体提供寓教于乐的知识渠道。

## 第五章 参考文献

- 【1】刘玉花,龙金晶,王恒. 科技馆观众的类型、特点与需求分析[C]//. 科技传播创新与科学文化发展——中国科普理论与实践探索——第十九届全国科普理论研讨会暨 2012 亚太地区科技传播国际论坛论文集.
- 【2】黄体茂. 关于科技馆观众需求的思考[C]//. 科技馆研究文选（2006-2015）
- 【3】李军. 使用与满足理论视阈下的博物馆观众研究[J]. 中国博物馆, 2016(02):16-23.
- 【4】王田苗,陶永,陈阳. 服务机器人技术研究现状与发展趋势[J]. 中国科学:信息科学, 2012, 42(09)
- 【5】刘逸伦,何倩鸿,王科盛. 服务机器人行业发展现状分析[J]. 信息记录材料
- 【6】刘妮丽. “双减”倒逼研学旅行市场走向何方[N]. 中国文化报

## 第六部分 致谢

首先非常感谢 Intel 官方给我们这次参加比赛的机会，在本次比赛过程中，我们不但接触到并使用到了最新的处理器，这极大地拓展了我们对于计算机领域的认知。也非常感谢庄杰老师、刘恒老师跟所有负责此次比赛的老师在此次比赛当中不遗余力地为我们提供帮助。此外我们也非常感谢自己的付出，感谢每一个队员的辛勤付出才会有此次项目的圆满完成！

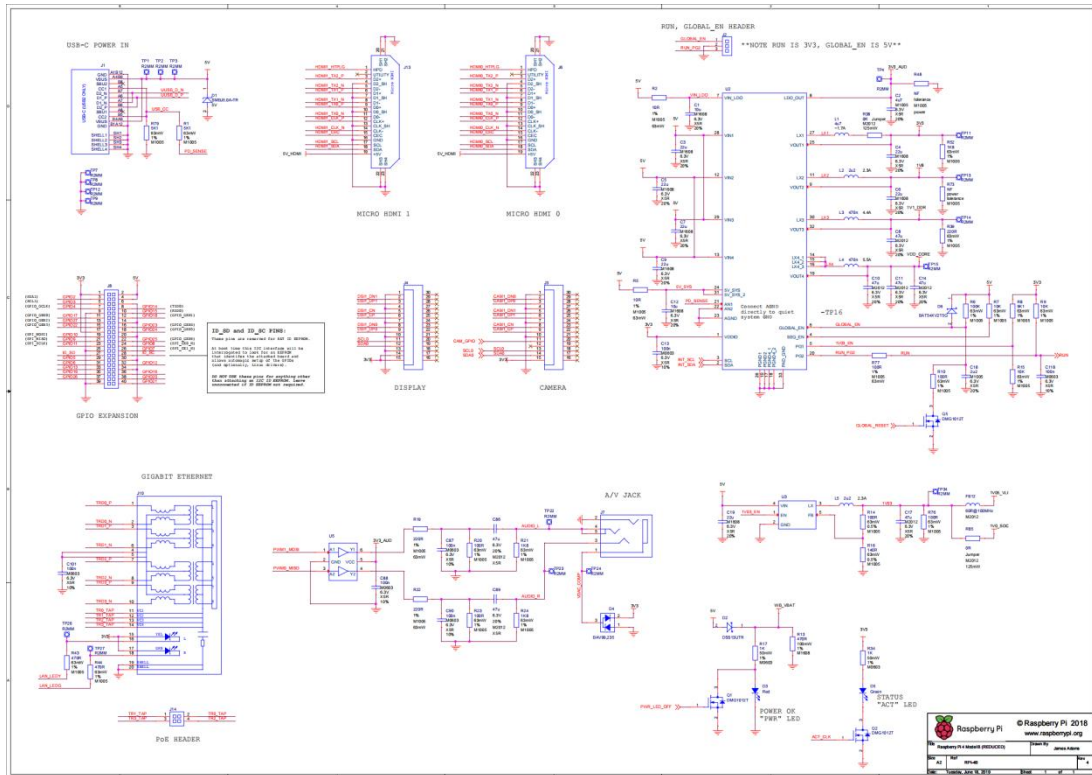
## 附录 1:

程序目录	类型	功能
Shoot.py	Python	从 python 文件中调取摄像头进行拍摄，保存图片并设置亮度和对比度为随机数，提高样本多样性。
Pic	Folder	存储 shoot.py 拍照保存下来的照片
To_csv.py	Python	将人脸照片转换为 csv 文件，后用于人脸识别调取
features2_all_QWJ.csv	Csv	储存组员人脸数据的 csv 文件
Net_camera.py	Python	上位机调取下位机的视频内容时用到网络摄像头模块
Fundamental.py	Python	Python 模拟键盘和鼠标输入，用于上位机和下位机通信
receive_txt.py	Python	上位机用于接收下位机发来的指令，用于上位机和下位机通信
audio2.py	Python	语音交互处理模块，获取用户发来的语音指令
Tracking.py	Python	下位机用于控制底板电机转动和导航功能
dlib_face_recognition_resnet_model_v1.dat	Dat	Python - Dlib 库中附有的人脸识别模块，基于 Resnet 的残差网络
shape_predictor_68_face_landmarks.dat	Dat	Python - Dlib 库中附有人脸 68 个特征点的预处理器

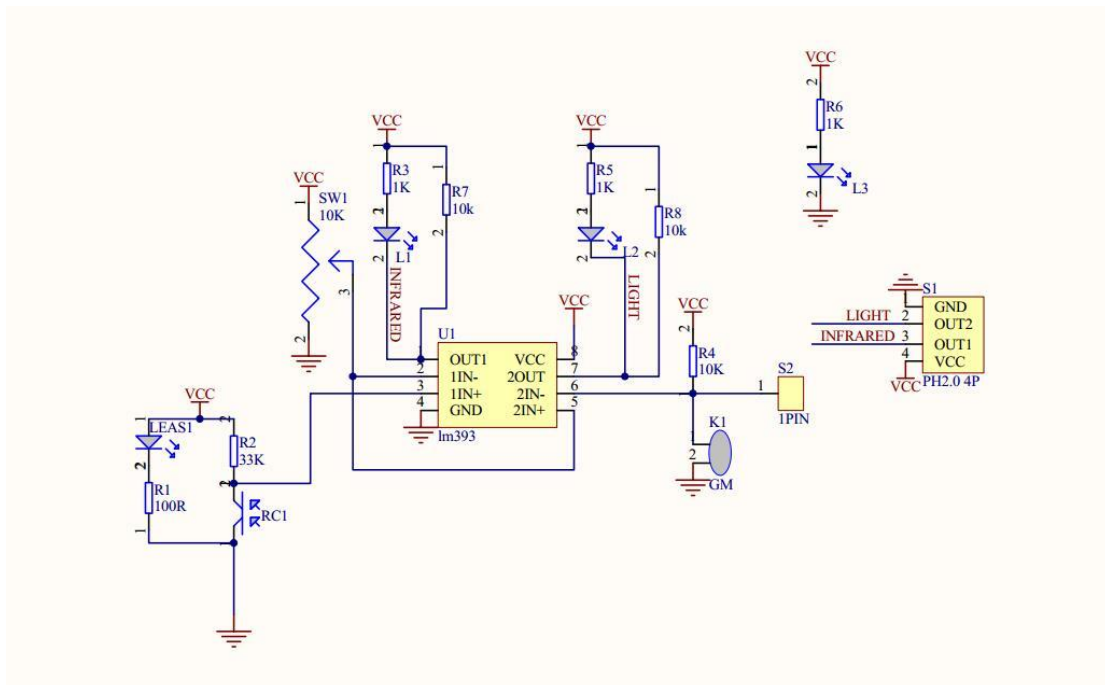


## 附录 2:

电路图:



树莓派电路原理图



红外避障传感器