

第二届中国高校智能机器人创意大赛

The 2nd National University Contest on Intelligent Robotic Innovations

创意设计报告

作品名称： 基于情绪分析的语音陪伴机器人
作品分类： 居家情感交流、陪伴
作品主题： 智能“家”居，情感交互

(2019 年 4 月)

一、作品是什么？

1. 作品简介

以**情绪分析**和**智能语音**为研究对象，作者充分调研了目前家居机器人的研究现状，大部分都是以释放人类体力与脑力为目标，而忽视了“家”的真正含义；在这个基础上作者提出了智能家居的新方向——**情感交互**。整个项目的最终目标是为了营造一个真正的智能“家”居，让科技与人类情感产生共鸣，缓解人们的生活压力，打造一个温馨舒适的情感家居环境。采用全彩 12 阶光立方作为三维动态显示器，基于表情识别和智能语音制作一种**智能陪伴机器人**。

本系统主要由三部分组成：**表情识别摄像头**——树莓派+摄像头搭建完成；**智能音箱**——FPGA 语音识别和麦克风播报；**三维动态显示系统**——包括电源开关电路、控制电路和级联驱动电路，并留下足够空间扩展后续电路。全彩的 3D 显示效果增加了使用者的视觉体验，表情识别和智能语音促进了人机交互，真正实现了人机语音交互，更加智能化、人性化。

2. 实际应用

陪伴机器人功能定位：

家庭娱乐——配合搭载语音识别、麦克风等技术终端的智能音箱让陪伴机器人更加人性化，提高生活的幸福度；

科普教育——全彩光立方动画激发孩子对三维空间的想象力、对智能科技的兴趣，传播三维建模知识；

情感补足——根据情绪分析结果，光立方利用不同旋律的音乐和动画表达自己的感情，让交互更加生动有趣，治愈情感。

二、创意在哪里？

能与人产生情感交互的人机交互才是智能“家”居的真正含义。本作品以情绪分析和智能语音为研究对象，采用全彩光立方三维动画、表情识别、语音识别等综合设计，让科技与人类情感产生共鸣，缓解人们的生活压力、情绪治愈，打造一个温馨舒适的情感家居环境。具体表现在：

1. 改善传统陪伴机器人功能单一、缺乏情感交流的问题；
2. 动态表情识别和裸眼 3D 显示技术的融合，给智能家居领域情感交流带来了新的契机，同时也提出了智能家居的新方向——**情感交互**。
3. 多种情感综合评价能极大可能的获取人的真实情绪状态，更有益于调节情绪、舒缓压力。
4. 三维动画的显示效果更能给人带来真实的**视觉享受**，裸眼 3D 的感观体验更能满足人们对画面的捕捉和丰富。
5. 高阶全彩光立方的**级联驱动电路**设计，解决了电流难以满足万级数量的 LED 驱动问题，可以用此设计方案满足更大数量的三维点阵显示器。

三、做得怎么样？

1. 动态表情识别

简单来说，面部表情是人体（形体）语言的一部分，是一种生理及心理的反应，通常用于传递情感。人类的面部表情有很多种，目前对高兴、吃惊、悲伤、愤怒、厌恶和恐惧 6 种人类基本表情的识别情况比较好，能较好的识别到面部表情，并给出标识、发送指令控制人机交互设备完成相应的动作。由于人情感的复杂性，这些表情还不足以完全确定一个人内心深处的情绪波动，如要提高判断的准确性，还需要通过心率检测、语音处理等综合评价。

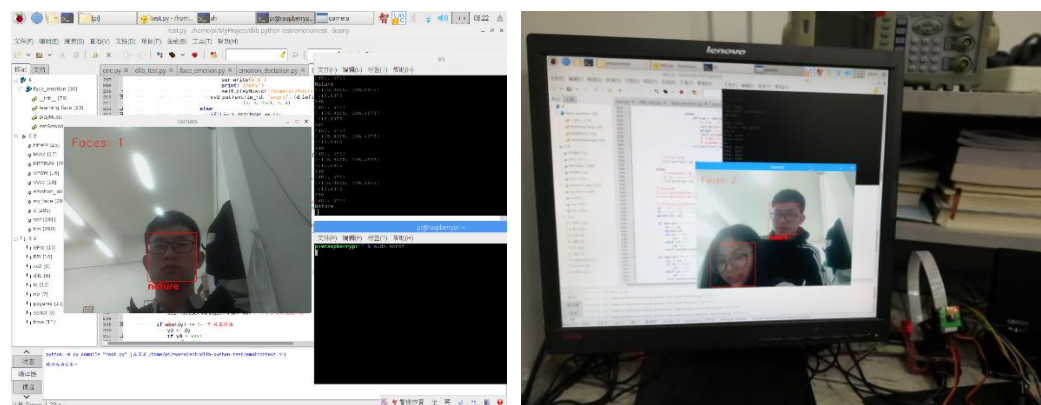


图 1 表情识别效果图

2. 智能语音实现情况

语音识别采用 LDV5 语音模块和 FPGA 的语音采样、滤波、识别实现。经过长时间的调试和参数设定，针对常用的语句识别率较高，通过本地语义分析检索，完成语音聊天功能。智能语音对答、轻音乐播放、讲故事、语音控制光立方显示等功能现都比较完善，但语音识别率还有待提高，响应时间也需要再降低。

3. 全彩光立方显示

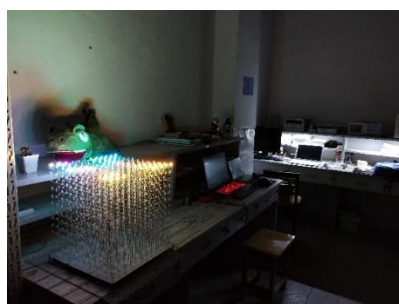


图 a “风车”

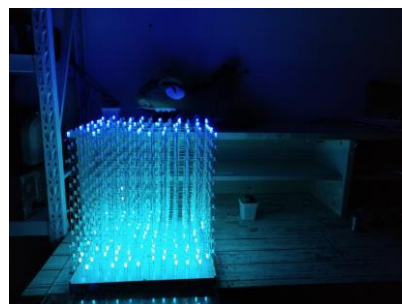


图 b “雨滴”



图 c 语音控制

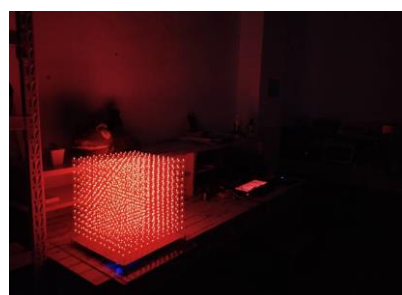


图 d 夜灯模式

图 2 光立方三维动态显示效果图

4. 整体实现情况

树莓派上电后打开摄像头，进行动态表情识别，当识别到面部表情后，播放调节情绪的背景音乐。同时，通过蓝牙通信发送指令给 STM32，控制光立方显示不同的三维动画，配合播放的背景音乐营造出轻松舒适而温馨的家居环境，能有效缓解人的精神压力、舒缓情绪。

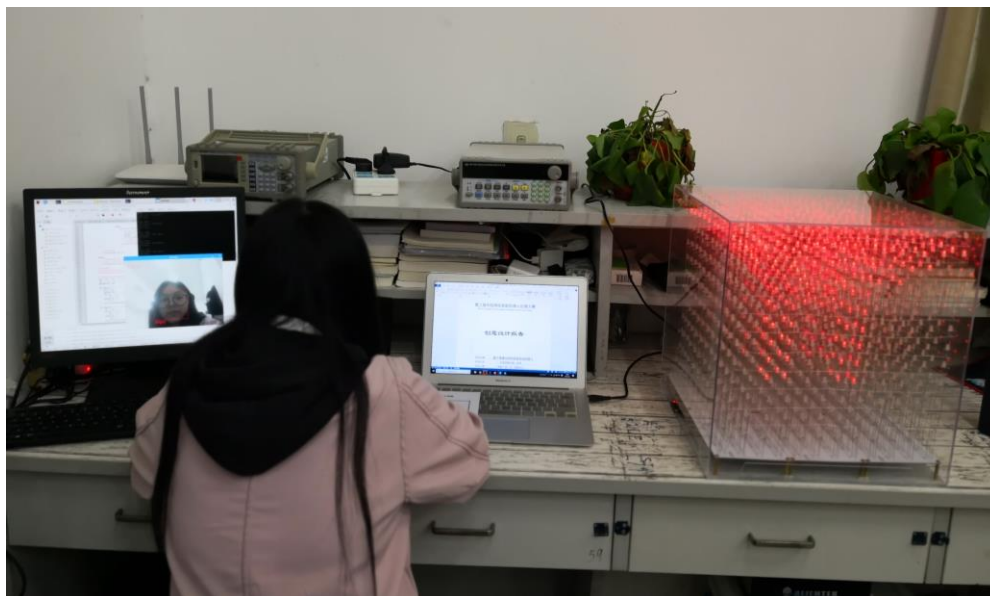


图 3 整体实现效果图

四、怎么做的？

1. 系统总体设计

如图 4 所示，该机器人系统由树莓派和 STM32 为主控制器，将表情识别结果发送至 STM32 控制光立方显示 3D 全彩动画。树莓派完成对云台舵机的控制，使得摄像头能跟踪人的移动；采用 Dlib 和 Numpy 进行人脸识别、面部特征点的提取和算法归一化处理，完成表情识别。STM32F4 控制 12 阶全彩光立方的三维位点的 LED 灯显示全彩动画，呈现 3D 画面，增强视觉效果。（具体实现流程可参考图 5）

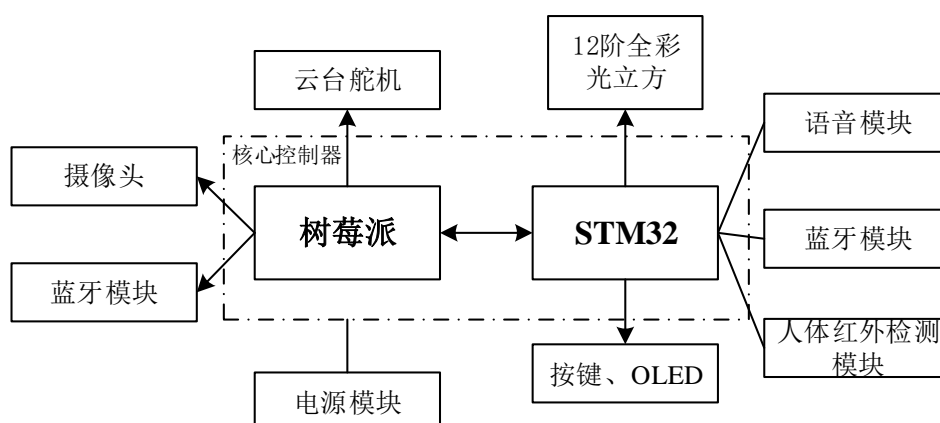


图 4 系统结构框图

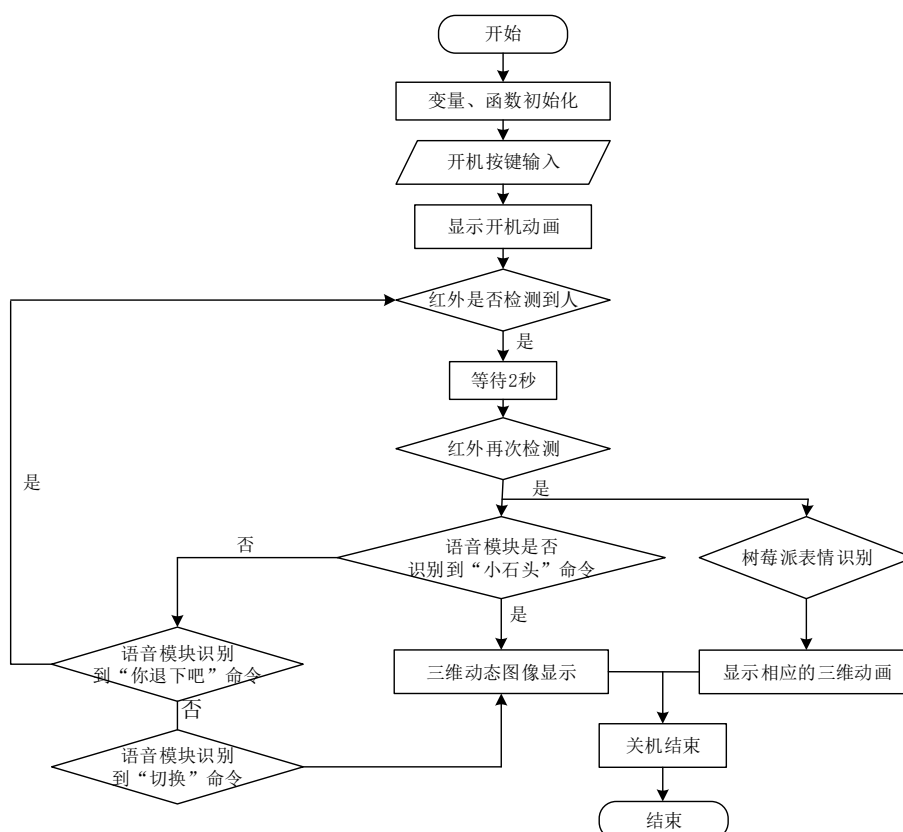


图 5 系统实现流程图

2. 动态表情识别

首先使用 dlib 完成人脸识别，并提取人脸信息 68 个特征点。然后实例化一个 shape_predictor 对象，使用 dlib 训练人脸特征检测，并进行人脸的特征点标定。标定的时候使用 opencv 的 circle 方法，在特征点的坐标上面添加水印，内容就是特征点的序号和位置。接下来需要根据这个 68 个特征点的坐标信息，进行综合计算，作为每个表情的判断指标。

根据我们提出的判断指标，先计算嘴巴的张开比例。在选择指标的标准数值之前，先对多个开心的人脸照片进行分析，计算开心时嘴巴张开比例的平均。用拟合直线的斜率近似代表眉毛的倾斜程度，同理计算出每个表情的眉高、嘴宽的数据，然后分类进行讨论，给出判别阈值。通过对多个不同表情数据的分析，得出每个指标的参考值，可以写出简单的表情分类标准，完成表情的识别。

3. 智能语音实现

对于音频信号的处理，我们采用 LDV5 模块来进行处理。FPGA 模拟 SPI 与模块通信读取其寄存器的值，并对其做 FFT 快速傅里叶变换进行语音信号处理，通过语音信号来发送相应的指令。FPGA 语音采样、滤波、识别，经过长时间的调试和参数设定，针对常用的语句识别率较高，通过本地语义分析检索，完成语音聊天功能。

4. 全彩高阶光立方

采用 12 片 SM16126 串转并强驱动芯片级联形成 144 个输出口控制，通过 DateIN 数据端口输入时序控制，进行 PWM 调制，产生 RGB 颜色调节。通过我们设计的光立方三维建模的上位机，设计不同的动画，标出每个 LED 等的 6 位数据，最终在 DMA 模式下读取 SD 卡的中的 12 阶光立方的每一帧动画的全部数据，实现光立方的全彩动画显示。