

山东科技大学

本科毕业设计（论文）开题报告

题 目 基于 MFCC-GMM 声纹识别树莓派声控小车

学 院 名 称 计算机科学与工程学院

专 业 班 级 物联网工程 15-1

学 生 姓 名 郑天田

学 号 201501061332

指 导 教 师 房胜

填表时间：二〇一九年三月十五日

填表说明

1. 开题报告作为毕业设计（论文）答辩委员会对学生答辩资格审查的依据材料之一。
2. 此报告应在指导教师指导下，由学生在毕业设计（论文）工作前期完成，经指导教师签署意见、相关系主任审查后生效。
3. 学生应按照学校统一设计的电子文档标准格式，用 A4 纸打印。装订在左侧。
4. 参考文献不少于 8 篇，其中应有适当的外文资料（一般不少于 2 篇）。
5. 开题报告作为毕业设计（论文）资料，与毕业设计（论文）一同存档。

设计（论文） 题目	基于 MFCC-GMM 声纹识别的树莓派声控小车				
设计（论文） 类型（划“√”）	工程设计	应用研究	开发研究	基础研究	其它
		√			

一、 本课题的研究目的和意义

声纹识别是通过分析声音的特征来识别说话人身份的技术，用来解决“谁在说话”的问题。对声音的识别包括语音识别和声纹识别，也就是通常所说的说话人识别。声纹识别与指纹识别和虹膜识别类似，通过分析人体的生物特征和行为来自动识别人身份。该项技术是由贝尔实验室在上个世纪四十年代末开发，主要用于军事情报领域，六十年代后期美国的法医鉴定，法庭证据等领域都使用了这项技术。如今声纹识别已经应用到各行各业，尤其在法庭审判和案件侦破中提供强有力的证据。

以指纹特征，人类面部特征，人类语音特征为代表的生物识别技术，在今后数年内将成为 IT 行业重要的技术革命。随着智能设备进入人们的日常生活，安全和隐私成为人们越来越关注的问题，由于面相，指纹等生物特征的独一无二，可以作为身份识别最好的 ID。所以近些年来，指纹识别，人脸识别大范围的应用到包括智能手机在内的智能设备当中，作为一种有安全保障的识别方式为用户提供身份验证。由于每个人的声腔尺寸和发生器被操纵的方式不同，声音也可以作为确定一个人身份的工具。联邦调查局曾对 2000 多起与声纹识别相关的案件进行的统计，其中利用声纹识别作为证据的案件只有 0.31% 的错误率。所以声纹是一种与指纹，面相特征一样的，可靠的身份特征。而且相比较指纹认证需要用户与设备进行直接接触，声音可以在空气中传输，因此可以在很远的距离进行身份的验证。目前，声纹识别已经应用到证券交易、银行交易、个人设备声控锁、汽车声控锁、公安取证等各行各业。

二、 本课题的主要研究内容（提纲）

学习了解说话人识别的相关算法技术。将传统机器学习模型与深度学习理论和说话人识别相结合。学习了解基于 MFCC 的声音特征提取、基于 GMM 和 HMM 的说话人识别、基于深度学习的说话人识别以及基于 PCA 和谱减法的语音降噪。

对各个模型和算法进行对比，进而分析各个模型和算法的优缺点。并针对已有的硬件设施设计出合理的系统。提出合理的功能和性能需求分析。

针对系统设计和需求分析，实现一个完整的基于说话人识别的树莓派小车系统，并对系统进行测试，检验系统是否达到设计要求。

三、 文献综述（国内外研究情况及其发展）

声纹识别的发展和研究现状。声纹识别技术分为两个大类：即说话人辨别和说话人确认技术。说话人辨别技术是从多个人中判断出当人说话的是其中的哪一个人，说话人确认是用来判断未知说话人是否为某一个已知的人。声纹识别的常用方法有模板匹配法，VQ 聚类法，最近邻近法。伴随着深度学习概念的问世，深度神经网络（DNN），循环神经网络（RNN）等方法的出现，对声纹识别技术的研究提供了越来越多的思路和方法。

这些方法虽然处理手段各不相同，但是基本原理是类似的。都是利用语谱图，通过提取说话人的基音频谱和包络，共振峰及其出现的频率和轨迹等参数特征，然后再与其他方法结合进行声纹识别。但是这些方法还存在着很大的局限和弊端。在对声音处理之前，需要对其进行很多处理，例如语音检测（VAD）、去噪，语音增强等。去除环境噪音和混响不仅对声纹的检测很重要，同时对语音含义的识别也很重要。

VAD 通常有两种方式，Long-Term Spectral Divergence (LTSD) 和能量检测，此外声音的特征主要通过梅尔频率倒谱系数 (MFCC) 提取，传统的机器学习模型主要有隐马尔可夫模型 (HMM) 和混合高斯模型 (GMM)。虽然声纹特征有着很好的唯一性，但目前声纹识别技术还不能提供极其精确的分辨，尤其是在混合说话人的环境下和噪音很大的情况下。

随着近几年深度学习的发展带来了模式识别的提升，但对于声纹识别的研究进展仍不大。目前声纹识别技术存在的主要问题有：

声纹的采集和特征的建立

噪音问题

混合人说话

说话人身体状况

说话人情绪

声纹识别还面临着诸多的问题，关于发展趋势，声纹识别和其他生物特征识别一样，向着深度学习的方向发展，但是由于深度学习是基于数据驱动模型，要有庞大的真实场景的数据集，以及对数据的精确标注，目前还比较难实现。所以传统机器学习的算法和模型还会在目前的声纹识别研究中占有很大比重。

四、 拟解决的关键问题

要解决的核心问题有：

1. 声纹特征的提取
2. 声学模型的训练
3. 语音的降噪。

五、 研究思路和方法

小车基于树莓派开发板，所以考虑硬件条件以及性能需求，开发板并不支持深度神经网络的大量数据采集及训练，所以采用传统的机器学习模型来搭建声纹识别系统，对于声纹特征的采集，采用目前流行的 MFCC 算法。树莓派对 python 有着很好的支持，并且通过 python 程序可以控制树莓派上的 GPIO，进而可以控制大量的传感器，同时 python 有着大量机器学习的库支持，所以选用 python 作为本系统的开发语言。针对功能需求的分析，系统将会分为四大模块，树莓派小车基本功能模块，语音唤醒模块，语音识别模块，声纹识别模块和数据采集模块，并且各个模块之间直接相互独立。

六、 本课题的进度安排

第 3 周到第四周，对所需的技术进行调研，调研该技术是否适合应用于本系统。

第 5 周到第 6 周，对系统提出需求分析，包括性能需求和功能需求。

第 7 周到第 12 周，编写代码，实现各个模块的功能。

第 13 周到第 16 周，测试系统，弥补不足以及完善论文。

七、 参考文献

- (1) 蔡宇, 采用子带谱减法的语音增强, 计算机应用, 2015
- (2) 刘光辉, 基于神经网络的全基因组 DNA 甲基化预测研究, 2017
- (3) Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4
- (4) 王凌之, 语义成分分析法解析, 2014
- (5) 于海涛, 网络上的信息传播模型与算法研究, 青岛大学报 2013
- (6) Pattern Recognition and Machine Learning, 2006 ,Christopher M. Bishop.
- (7) Artificial Intelligence: A modern approach, 3rd edition, 2010, Stuart Russell and Peter Norvig.
- (8) Sahidullah, Md.; Saha, Goutam. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Communication. May 2012, 54 (4): 543 – 565.

指导教师意见

指导教师（签名）:

年 月 日

所在系（所）意见

负责人（签章）：

年 月 日