



带安全验证的声纹识别系统

软件学院创新项目立项答辩

小组成员：胡孝博 方必诚 薛仁捷

指导老师：软件学院 沈莹

2020.4.16



目录

1. 研究背景与研究目的
2. 研究内容
3. 项目进度安排
4. 已有基础与现阶段成果
5. 创新点与项目特色
6. 研究成果的应用前景



1 研究背景 & 研究目的

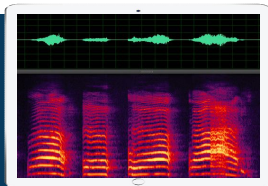
- 随着信息技术的迅猛发展，如何有效地进行身份识别，进而保障个人隐私和信息安全，成为一个愈发重要的问题

声纹识别(Voiceprint-Recognition)

- 优点：
- 非接触性采集(可远程实现)
 - 采集成本低
 - 伪造难度高
 - 应用场景广泛
 - 算法复杂度低

现有技术的不足：

- 对抗环境噪音干扰的能力 (准确性)
- 缺乏辅助安全验证模块 (安全性)



声纹识别技术具有潜在的研究价值以及可行的研究方向！

1 研究背景 & 研究目的

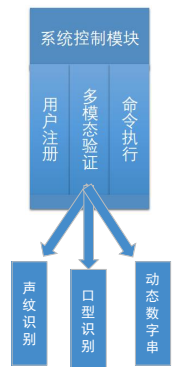
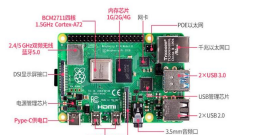
基于嵌入式的离线声纹识别系统结构示意图

Q: 是否可以引入一种辅助的“动态安全验证模块”？

- 将多种身份识别技术相结合 (CV+CA)
- 构成**声纹识别**与**口型识别**相结合的多模态“视听识别系统”

树莓派

功能强大 轻便小巧
可拓展性高 支持Python、Matlab等



2 研究内容

Q: 我们要研究的核心问题有哪些？



实现带文本提示的
声纹识别技术



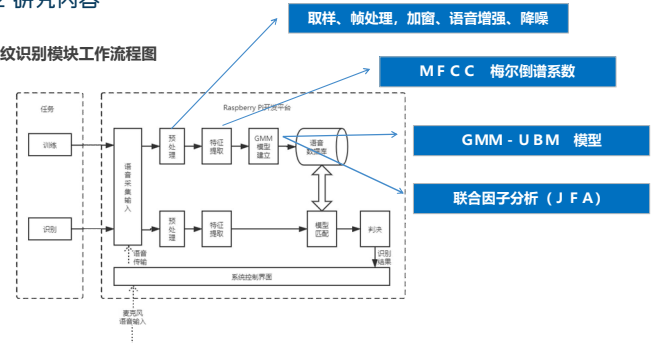
融合时空视听信息
解决音频流与视频流匹配性
(3D-CNNs)



移植至树莓派
实现嵌入式系统

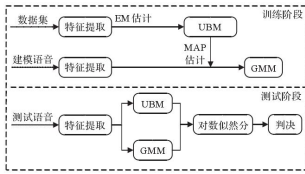
2 研究内容

声纹识别模块工作流程图

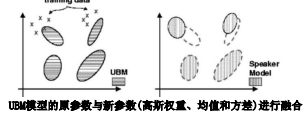


2 研究内容

目前，在声纹识别领域，高斯混合模型（GMM）、联合因子分析模型、i-vector等仍是主流模型构建方式。



基于GMM-UBM的说话人识别算法



UBM模型的原参数与新参数(高斯权重、均值和方差)进行融合

通常在实际中一个说话人的语音数据比较有限而训练高斯混合模型又需要大量的训练数据

基于GMM的通用背景模型(Universal Background Model) GMM-UBM的系统框架

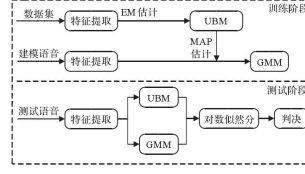
通过背景数据生成不具备表征身份能力的先验模型

极大后验概率(MAP)算法

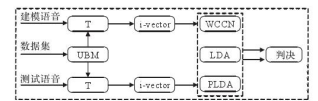
将预先训练过的模型向目标说话人模型进行微调避免过度拟合

2 研究内容

目前，在声纹识别领域，高斯混合模型（GMM）、联合因子分析模型、i-vector等仍是主流模型构建方式。



基于GMM-UBM的说话人识别算法



基于i-vector的说话人识别算法

i-vector是基于JFA的简化版，即用一个全因子空间（Total factor matrix, T）同时描述说话人信息和信道信息，然后把语音映射到了一个固定且低维的向量上

2 研究内容

通过python、matlab完成相关算法的实现、优化与调用

- 声纹识别:
pyaudio (语音预处理)
MFCC (梅尔倒谱系数, 用于声音特征提取)
scikit-learn (机器学习模块 包含GMM模型)
Matlab Audio工具箱
UIS-RNN (用于项目早期数据合理性验证与中后期接口测试)

- 口型识别:
opencv (进行图像处理, 唇唇定位与唇部特征提取)
tensorflow (提供深度学习框架)

嵌入式系统: 树莓派4B + 摄像头 + 麦克风

基于ARM的微型电脑主板, 其系统基于Linux

3 项目进度安排

【2021.1 - 2021.2】项目需求确认、学习和初步调研

【2021.2 - 2021.3】技术路线初步确定 PC上搭建系统框架 GUI界面

【2021.3 - 2021.4】嘴唇定位 声纹特征的提取 现阶段

【2021.5 - 2021.7】声纹识别的模型训练 口型识别模型建立

【2021.7 - 2021.10】声纹识别模块的测试 口型特征与文本相关性

【2021.10 - 2021.11】移植树莓派平台与测试

【2021.11 - 2022.2】性能升级与算法优化

【2022.2 - 2022.3】完整声纹系统测试 结题准备



4 已有基础与现阶段成果

【2021.1 - 2021.2】项目需求确认、学习和初步调研（寒假前期）

带安全验证的声纹识别系统技术综述

胡名博 郭生院讲学堂 2053410

摘要:

声纹识别是生物识别技术的一大分支, 相关技术的发展日益成熟, 近年来应用广泛, 在金融、安防以及医疗领域, 本文基于带安全验证的声纹识别系统的相关理论, 总结了声纹识别的概念原理, 基本实现框架, 传统模型以及深度学习算法, 并探讨了声纹识别系统的联合识别模式进行了技术综述。

关键词: 声纹识别 深度学习 口型识别

关于带安全验证的声纹识别系统的论述

2052329 方必诚 信息类 (陈勤)

摘要:

随着信息技术的迅猛发展, 如何识别个人的身份, 进而保障个人隐私和信息安全, 逐渐成为一个重要的问题。时至今日, 指纹识别、人脸识别等生物识别方式已较为普遍流行, 而声纹识别的研究热度也在不断提高。一方面, 它具有不易遗忘、不会丢失等优势, 具有较好的快捷性和准确性; 另一方面, 它也在处理环境噪音、判断说话人状态方面需要改进技术。本文针对声纹识别技术的现状, 构建带安全验证的声纹识别系统, 运用阅读文本验证、动态安全验证等方法, 使用 Python 语言在嵌入式系统中完成声纹识别并提供相关指令操作。

关键词:

声纹识别 Python 嵌入式 动态安全验证 MFCC

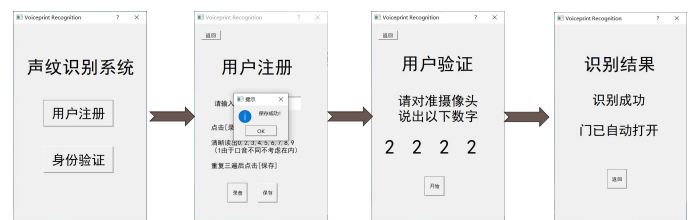
针对声纹识别、口型识别技术的综述

- 了解计算机听觉与计算机视觉领域的技术现状
- 总结了现有算法的优缺点
- 梳理了可行的技术路线

4 已有基础与现阶段成果

【2021.2 - 2021.3】技术路线初步确定 PC上搭建系统框架 GUI界面（寒假后半段——开学前两周）

依据系统框架初步完成的简化版GUI界面（Qt Designer）

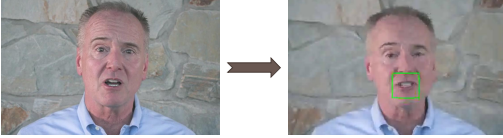


4 已有基础与现阶段成果

【2021.3 - 2021.4】嘴唇定位 声纹特征的提取（当前阶段）

唇部定位与运动特征提取

—3D Convolutional Neural Networks for Cross Audio-Visual Matching Recognition



```

$ ./lib-readimg-delearning $(pwd)/src/lib_readimg.py $(pwd)/src/vizualize.py -p "/phoneme" --dataset-dir="/data/phoneme" --model-dir="/data/phoneme" --data-format="raw" --data-type="float32" --data-channel=1 --data-size=1024 --data-batch-size=16 --data-shape=(720, 544, 3)
frame_shape: (720, 544, 3)


The cropped photo is detected ...
frame number 0 of 28
writing frame 1 with activation 1
frame_shape: (720, 544, 3)

The cropped photo is detected ...
frame number 1 of 28
writing frame 2 with activation 1
frame_shape: (720, 544, 3)

```

- 基于python的opencv库
- 利用ffmpeg进行帧处理、合成与导出

5 创新点与项目特色

- 1. 跨模态的安全验证算法 引入脸型验证模块应用于声纹识别系统
 - 避免播放录音、模仿声音而产生的安全隐患
 - 在噪音等干扰导致音频信息可靠性降低的情况下，提高识别的准确率
 - 2. 离线的嵌入式系统 保证成果的实用性
 - 轻便，可扩展性强
- 



6 研究成果的应用前景

1. 智慧城市建设
- (1) 家居安全: 安全可靠的门禁系统 --> 智能家居的核心一环
 - (2) 公共卫生: 非接触式的身份识别 --> 如新冠疫情下, 口罩阻碍了人脸识别的准确性
2. 人机交互的前沿领域
- 让智能机器人准确识别“主人身份”, 大大提高机器人安全性与隐私性



期待您的指导!

答辩人：
胡孝博(2053410) 工科试验班(信息类)
方必诚(2052329) 工科试验班(信息类)
薛仁捷(2050336) 理科试验班