

# 基于匹配层融合的身份识别算法研究与实现

李傲梅, 胡正豪, 周川川

(陆军炮兵防空兵学院 信息工程系, 安徽 合肥 230031)

**摘要:** 提出一种基于音视频匹配层自适应加权融合的身份识别方法。在不同程度的噪声情况下, 图像与声音的识别率会随噪声的增强而降低, 凭借单个生物模态的识别, 难以达到很好的预测结果; 而且两种模态融合时的权值不同, 融合系统的稳定性效果也不同。采用双模态的自适应加权融合不仅可以有效地弥补不同生物模态识别之间的优缺点, 而且可以自适应选择最优的权值进行决策。实验表明, 该方法的理论推测成立, 比单模态的身份识别具有更高的识别率与鲁棒性。

**关键词:** 音视频融合; 匹配层; 自适应加权; 身份识别; 鲁棒性

中图分类号: TN06; TP391.4

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.191426

中文引用格式: 李傲梅, 胡正豪, 周川川. 基于匹配层融合的身份识别算法研究与实现[J]. 电子技术应用, 2020, 46(7): 57-59.

英文引用格式: Li Aomei, Hu Zhenghao, Zhou Chuanchuan. Research and implementation of recognition algorithm based on matching layer fusion[J]. Application of Electronic Technique, 2020, 46(7): 57-59.

## Research and implementation of recognition algorithm based on matching layer fusion

Li Aomei, Hu Zhenghao, Zhou Chuanchuan

(Department of Information and Communication Engineering, Army and Artillery Air Defense Force College, Hefei 230031, China)

**Abstract:** This paper proposes an adaptive weighted fusion based on audio-video matching layer. In the case of different degree of noise, the recognition degree of image and sound will decrease with the increase of noise. And the weight of the two modes is different, the stability effect of the fusion system is also different. The adaptive weighted fusion of two modes can not only make up the advantages and disadvantages of different biological modes, but also choose the optimal weight to make the decision. Experiments show that the proposed method is feasible and has higher recognition rate and robustness than single mode identification.

**Key words:** audio and video fusion; matching layer; adaptive weighting; identification; robustness

### 0 引言

目前, 人工智能迅速崛起, 身份识别技术更是应用在生活各个领域之中, 如检票、付款、登录系统等。但是日常应用的环境都是在条件较好的情况, 并不满足全天候嘈杂的复杂作战环境, 如在阴雨天或者夜晚光线不充足时, 人脸图像的采集就会受到影响; 在战车舱内进行身份识别驾驶时, 会因为发动机或炮火声的影响, 导致语音信号受噪声干扰。所以, 依靠任何一种单模态的身份识别都极易受到外部环境的影响。采用多模态的身份识别技术(如人脸、语音、指纹、虹膜等<sup>[1-3]</sup>), 可以通过对来自多种传感器信号的处理、分析、决策, 弥补不同模态信息的不足, 降低单一模态受干扰产生的误差, 得到更加可靠、准确的结果。

近年来, 多模态的融合识别技术备受关注, 一般情况下主要分为数据层、特征层、匹配层和决策层融合<sup>[4-5]</sup>。由于图像属于二维信号, 语音属于一维信号, 数据层和特征层融合相容性不强, 且运算复杂; 决策层融合依赖信息匮乏单一, 抗噪能力弱。因此, 本文主要提出一种针对声纹和人脸特征的匹配层融合识别算法, 首先对不

同说话人的无噪声语音信号建立高斯混合模型, 然后对加入噪声的语音信号进行模型匹配获得匹配分数, 再对含有噪声的人脸图像依次进行小波分解、主成分分析(Principal Component Analysis, PCA)降维、支持向量机(Support Vector Machine, SVM)分类获取匹配分数, 最后利用本文提出的自适应加权融合算法进行加权匹配分析, 获得最优决策结果。本文创新点在于规范一个衡量系统稳定性的标准 $\xi$ , 并利用这个标准选择最佳的权值进行融合, 判断结果。实验结果验证表明, 该方法比采用单一模态的识别算法更加有效, 具有较高且稳定的识别率。

### 1 说话人识别

由于语音信号采集方便且不易失真的优点, 说话人识别技术受到广泛关注, 提取特征和识别方法也很多, 但是比较通用的还是采用基于梅尔倒谱系数(Mel-scale Frequency Cepstral Coefficients, MFCC)及其一阶差分特征参数构建高斯混合模型(Gaussian Mixed Mode, GMM)的方法<sup>[6-7]</sup>。本文也是通过该方法, 获得声音匹配分数。具体流程如图 1 所示。

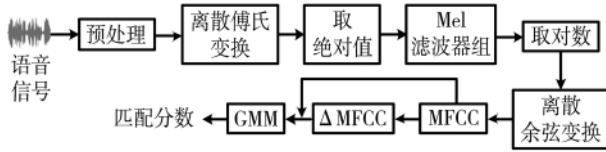


图1 语音匹配分数计算流程图

## 2 人脸识别

由于人脸特征的直观、易辨认性,人脸识别成为了人们日常生活当中确认身份常用的方法,并随着人们对人脸识别的不断研究逐步改进。在本文中,采用了PCA与径向基函数(Radial Basis Function, RBF)核的SVM相结合的方法<sup>[8-9]</sup>,具体的处理流程如图2所示。



图2 人脸匹配分数计算流程图

## 3 匹配层融合算法

通过以上两种方法分别可以得到待识别样本的语音匹配分数 $\{a_i\}_{i=1,2,\dots,n}$ 和人脸匹配分数 $\{b_i\}_{i=1,2,\dots,n}$ ,其中 $i$ 代表样本标签号, $n$ 表示总的标签数。然后对匹配分数计算如下:

(1)归一化处理。由于两种模式下得到的匹配值量纲是不相同的,不能直接作为计算的依据,因此在融合前必须进行归一化处理。本文中采用的是Min-Max归一化方法,转换公式如下:

$$a'_i = \frac{a_i - a_{\min}}{a_{\max} - a_{\min}} \quad i=1, 2, \dots, n \quad (1)$$

$$b'_i = \frac{b_i - b_{\min}}{b_{\max} - b_{\min}} \quad i=1, 2, \dots, n \quad (2)$$

(2)计算每种模式不同类别的匹配分数,匹配分数高的对系统的信任度就越高,通过匹配分数可以更直观地反映不同类别判断的相差程度。

$$\text{Score}_i(a) = \frac{a'_i}{\sum_{i=1}^n a'_i} \quad i=1, 2, \dots, n \quad (3)$$

$$\text{Score}_i(b) = \frac{b'_i}{\sum_{i=1}^n b'_i} \quad i=1, 2, \dots, n \quad (4)$$

(3)将权重 $\alpha$ 以0.02为步长,从0递加到1,带入式(5)进行计算可得整体匹配分数矩阵 $S^{51 \times n}$ (其中 $S^{51 \times n}$ 表示51行 $n$ 列数组, $S_{ki}$ 代表第 $k$ 行 $i$ 列元素)。

$$S_{ki} = \alpha_k \times \text{Score}_i(a) + (1 - \alpha_k) \times \text{Score}_i(b) \quad i=1, 2, \dots, n; k=1, 2, \dots, 51 \quad (5)$$

(4)求取矩阵 $S^{51 \times n}$ 中每行的最大值与第二大值,得数组 $\text{Max}^{51}$ 和 $\text{Sec}^{51}$

$$\text{Max}_k = \max(S_k) \quad k=1, 2, \dots, 51 \quad (6)$$

$$\text{Sec}_k = \text{sec}(S_k) \quad k=1, 2, \dots, 51 \quad (7)$$

(5)计算每个权值下最大值与第二大值的差得数组 $\xi^{51}$ ,取 $\xi^{51}$ 最大值所在的权值,即为最佳权值,该权值下匹配分数最大的类别即为最终判断结果。 $\xi$ 越大,说明该类别与其他类别的差别越大,分类越清晰,系统就越稳定。具体计算公式如下:

$$\xi_k = \text{Max}_k - \text{Sec}_k \quad k=1, 2, \dots, 51 \quad (8)$$

该方法在每次融合决策前进行,就可以通过自适应调整获得最佳融合权值,得到最稳定的融合系统和最优匹配结果。

## 4 实验结果与分析

本实验在Spyder下使用Python3编程,利用sklearn模块相应库函数,采用ORL数据库作为人脸实验数据库,库中总共有40个人的400幅图像信息;利用麦克风采集40个人的400段语音信号作为语音数据库,与人脸数据库相对应。之后,采用十折交叉验证法随机分为10组,分别选每1组作为测试样本,另外9组作为训练样本进行测试,共执行10次求得识别率。

### 4.1 实验数据处理

对人脸图像与语音信号分别进行加噪处理,模拟在恶劣环境下采集的数据信息。如图3所示,图3(b)是椒盐噪声比例为0.1的人脸图像,图3(d)是SNR(信噪比)为10的声音信号。

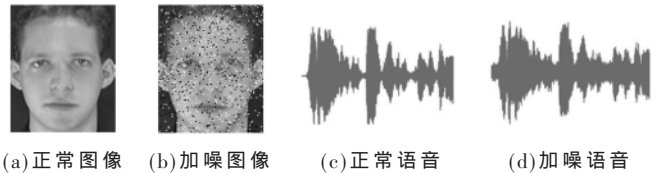


图3 不同噪声处理得到的数据

### 4.2 实验对比分析

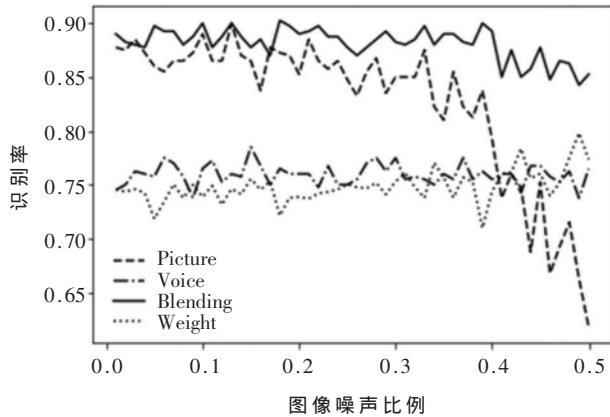
在不同噪声条件下进行了多次实验,得到如图4所示实验结果:图4(a)为声音信噪比为8时图像椒盐噪声在0.01~0.5范围内的变化图,图4(b)为图像椒盐噪声为0.1时声音噪声在0.01~0.4(信噪比为40~8)范围内的变化图。其中,Picture曲线为人脸识别率,Voice曲线为语音识别率,Blending曲线为融合识别率,Weight为声音权重变化。声音信噪比公式为:

$$\text{SNR} = 20 \log_{10} \frac{P_s}{P_n} \quad (9)$$

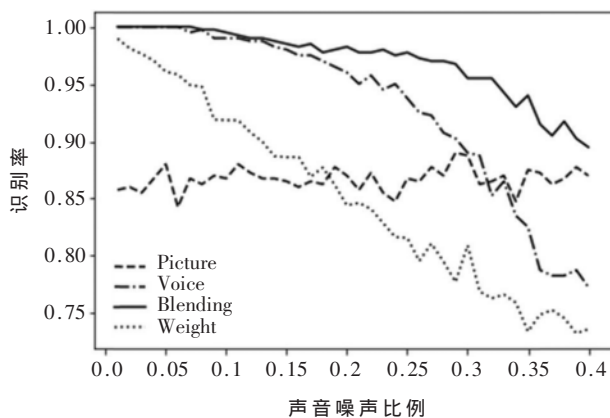
其中, $P_s$ 为信号有效功率, $P_n$ 为噪声有效功率。

由图4(a)可知,随图像噪声的增强,人脸识别率不断降低,融合后的识别率虽有影响但仍然保持在0.88以上;声音权重有升高趋势,但变化不明显,说明虽然图像识别率降低了,但是系统的稳定性变化并不大。

由图4(b)可知,无噪声情况下,声音识别率特别高,随声音噪声的增强,声音识别率逐渐降低,融合后的识别率也随之降低,但仍然在0.92以上;声音权重持续变小,但仍然较高,说明随声音噪声的增强,声音识别的稳



(a)不同图像噪声



(b)不同声音噪声

图4 不同噪声识别率变化图

定性也降低,但仍强于人脸识别系统。

综合图4(a)、图4(b)可知,融合后的识别效果明显要强于任何一种单模态的识别效果,并且具有较高的识别率。

## 5 结论

本文针对在单一模态受噪声影响下识别率下降问

题,提出对语音、图像信息分别进行提取特征值和计算匹配分数,然后采用自适应加权融合算法选取最优匹配权值,得出最优决策的方法,并进行实验验证。经实验证明,高斯混合模型的声音识别稳定性远远强于SVM分类的人脸识别;而且在噪声背景下,采用该匹配层融合的方法远远优于任何一种单模态的识别效果,同时也证明了采用多模态的融合识别将是身份识别的重要发展方向。

## 参考文献

- [1] 孙贵华,陈淑荣.一种改进的Refine多尺度人脸检测方法[J].电子技术应用,2019,45(8):34-39.
- [2] 范叶平,李玉.基于深度集成学习的人脸智能反馈认知方法[J].电子技术应用,2019,45(5):5-8,13.
- [3] 余成波,李彩虹,曾亮.K-means指纹定位的优化算法[J].电子技术应用,2018,44(2):70-74.
- [4] 柳欣,耿佳佳.多生物特征融合发展现状及其展望[J].小型微型计算机系统,2017,38(8):1792-1799.
- [5] 陈倩.多生物特征融合身份识别研究[D].杭州:浙江大学,2007.
- [6] 柯晶晶,周萍.差分和加权Mel倒谱混合参数应用于说话人识别[J].微电子学与计算机,2014,31(9):88-91.
- [7] 周萍,沈昊.基于MFCC与GFCC混合特征参数的说话人识别[J].应用科学学报,2019,37(1):24-32.
- [8] 张庶,李子月,刘玉超.基于PCA与SVM的人脸识别技术[J].指挥与控制学报,2019,5(3):249-253.
- [9] 刘艳丽,王铁建.SVM算法在人脸识别中的应用研究[J].电脑知识与技术,2017,13(18):176-177.

(收稿日期:2019-12-30)

## 作者简介:

李傲梅(1972-),女,博士,教授,主要研究方向:信息与通信工程、电子工程。

胡正豪(1992-),通信作者,男,硕士研究生,主要研究方向:智能控制、模式识别,E-mail:289449169@qq.com。

周川川(1990-),男,硕士研究生,主要研究方向:电子工程、智能控制。

(上接第56页)

on Pattern Recognition.IEEE,1994.

- [11] KRIZHEVSKY A,SUTSKEVER I,HINTON G E.Imagenet classification with deep convolutional neural networks[J].Advances in Neural Information Processing Systems,2012,25(2):1106-1114.
- [12] ZEILER M D,FERGUS R.Visualizing and understanding convolutional networks[J].ECCV,2014,8689(53):818-833.
- [13] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
- [14] SZEGEDY C,Liu Wei,Jia Yangqing,et al.Going deeper with convolutions[C].Conference on Computer Vision and Pattern Recognition(CVPR).IEEE,2015.

- [15] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[C].LCLR2015,2015.

- [16] He Kaiming,Zhang Xiangyu,Ren Shaoqing,et al.Spatial pyramid pooling in deep convolutional networks for visual recognition[C].ECCV 2014,European Conference on Computer Vision,2014:346-361.

(收稿日期:2019-12-25)

## 作者简介:

李太文(1994-),男,硕士,主要研究方向:深度学习、模式识别。

范昕炜(1978-),男,博士,高级工程师,主要研究方向:机器人、模式识别与智能控制、数据挖掘。

## 版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所