

# 基于 PCA 和非线性映射改进的 MFCC 特征提取方法<sup>\*</sup>

符恬恬<sup>1,2</sup> 郑斌琪<sup>1</sup> 李成娟<sup>1</sup> 夏利杰<sup>1</sup>

(1. 中国科学院上海微系统与信息技术研究所微系统技术重点实验室 上海 201800; 2. 中国科学院大学 北京 100049)

**摘要:** 使用传统的梅尔倒谱系数(MFCC)作为特征进行野外目标识别时,由于 MFCC 对环境干扰较敏感,所以会导致识别率的下降。针对这个问题,提出了使用主成分分析法(PCA)代替 MFCC 提取过程中使用的离散余弦变换,并且使用非线性函数对梅尔滤波后所获得的对数能量进行映射。改进后的 MFCC 更贴合实际数据、可以增强目标信号所在频段的权重、有着良好的可分性和鲁棒性。经过实验证,使用 PCA 和非线性映射改进后的 MFCC 作为分类特征时,准确率为 93.36%,优于传统的 MFCC。

**关键词:** 目标识别;梅尔倒谱系数;主成分分析;非线性映射

中图分类号: TN911.7 文献标识码: A 国家标准学科分类代码: 510.4040

## Improved MFCC feature extraction method based on PCA and nonlinear mapping

Fu Tiantian<sup>1,2</sup> Zheng Binqi<sup>1</sup> Li Chengjuan<sup>1</sup> Xia Lijie<sup>1</sup>

(1. Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201800, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** When using traditional Mel-frequency cepstral coefficients (MFCC) as features for target recognition in wild filed environments, their sensitivity to environmental interference often leads to a decline in recognition accuracy. To address this issue, this study proposes replacing the discrete cosine transform used in the MFCC extraction process with principal component analysis (PCA) and applying a nonlinear function to map the logarithmic energy obtained after Mel filtering. The improved MFCC are more aligned with actual data, can enhance the weighting of frequency bands containing the target's signal, and have better separability and robustness. Experimental results show that using the improved MFCC based on PCA and nonlinear mapping as classification features achieves an accuracy of 93.36%, surpassing the performance of traditional MFCC.

**Keywords:** target recognition; Mel-frequency cepstral coefficient; principal component analysis; nonlinear mapping

## 0 引言

科学技术的快速发展改变了社会各个领域的运作方式,在军事领域,以信息战为主的作战模式已经成为主流。信息战的关键在于如何获取准确的战场情报,对战场的实时监控和侦查是获取战场情报的一种重要方式。无线传感网络可以实现对战场环境的感知和信息的处理<sup>[1]</sup>,无线传感器网络由多种传感器组成<sup>[2]</sup>,声音传感器是无线传感器网络的重要组成部分,它可以实现野外目标的检测和识别,有着被动式检测、能耗低、受地形影响小、隐蔽性好等多种优点,在无线传感器网络中发挥了巨大的作用<sup>[3]</sup>。

基于声音传感器进行目标识别主要分为数据采集和数

据处理两个部分。

数据处理的主要步骤为预处理、信号增强、特征提取和分类等。为了提升目标识别的准确率,需要提取有着良好可分性、贴合实际数据的声学特征。比较常见的声学特征有以下几种:基于语音信号功率谱提取的语谱图(spectrogram)特征、基于梅尔滤波器的滤波器组(filter-bank, FBank)特征和梅尔倒谱系数(mel-frequency cepstral coefficient, MFCC)特征<sup>[4]</sup>、基于 Gammatone 滤波器组的 GFCC(gammatone cepstral coefficient)特征<sup>[5]</sup>、滑动差分倒谱(shifted delta cepstrum, SDC)<sup>[6]</sup>、基于经验模态分解的特征<sup>[7]</sup>等。同时还有一些基础特征的改进算法,如周大春等<sup>[8]</sup>提出了了 fisher 加权的 GFCC 特征,Masumura 等<sup>[9]</sup>提

出了 12 维的 MFCC 特征和其一、二阶差分特征组成的混合特征, 张楠等<sup>[10]</sup> 提取了一种基于 LMD 改进的 Gammatone 滤波器组图谱特征。

在上述特征中, MFCC 应用广泛, 因为其计算较简单, 能模拟人耳的听觉特性, 在识别纯洁声音时效果较好, 被广泛应用于声音识别中。但在实际场景中, MFCC 特征对噪声敏感度较高, 受环境影响较大<sup>[11]</sup>, 从而影响了后续的识别效果。研究者们提出了大量 MFCC 的改进算法, 以提高识别率。陶瀚宇等<sup>[12]</sup> 将梅尔滤波器组翻转后获得翻转梅尔倒谱滤波器, 并且用翻转梅尔倒谱滤波器获得 IMFCC (inverted mel-frequency cepstral coefficients), 再将 MFCC-IMFCC 联合起来获得了混合倒谱特征, MFCC 和 IMFCC 特征呈现互补关系, 两者结合可以更完整地捕捉信号在各个频段的特征; 张伟等<sup>[13]</sup> 使用多正弦窗估计代替 MFCC 中的 FFT 对信号进行谱估计, 多正弦窗可以减少频谱泄露, 并用幂函数压缩代替 MFCC 中的对数压缩, 幂函数可以减少低频干扰, 获得改进后的 MFCC 特征; 岳倩倩等<sup>[14]</sup> 依据人耳听觉的特征, 使用听觉神经发放率-声音强度曲线来对梅尔能量进行映射, 获取了改进的 WMFCC 特征。大多改进方案的思路从 MFCC 的提取过程入手, 或其关键步骤的提取方法进行替换, 或改变滤波器, 或进行特征融合, 或通过人耳听觉曲线使改进特征更加符合人耳听觉特征, 未考虑到目标信号的声学特性。

为了增强 MFCC 特征的稳健性, 使其更贴合目标信号的声学特征。本文提出了使用 PCA 和非线性映射改进的 MFCC 特征提取方法, 具体工作如下: 使用了 PCA 来代替 MFCC 提取过程中的离散余弦变换 (discrete cosine transform, DCT)。通过训练集获得的 PCA 变换矩阵会比纯粹是数学计算的 DCT 变换矩阵更贴近实际数据。同时考虑到目标信号集中于低频段, 所以使用非负、递减的非线性映射函数对低频部分进行加强, 非线性映射函数的常数参数可以使用粒子群优化算法 (particle swarm optimization, PSO) 获得, 将令识别率达到最高的常数参数作为最终参数, 再将改进后的特征使用 XGBoost 分类, 获得最终的分类结果。

经过实验验证, 本文提出的基于 PCA 和非线性映射的改进的 MFCC 特征提取方法能够有效提升野外目标分类识别的准确率。

## 1 基本理论

### 1.1 传统的 MFCC 特征

MFCC 特征是基于人耳的听觉特性的声音频谱特征。人耳所能听到声音的高低与声音的频率并非是线性关系, 随着声音频率的增强, 人耳对声音信号的接受能力是越来越弱的<sup>[15]</sup>。所以使用 Mel 频率尺度更加贴近人耳的感知特性。Mel 频率尺度与实际频率的映射关系可用公式表示:

$$F_{\text{Mel}}(f) = 1125 \ln(1 + f/700) \quad (1)$$

式中:  $F_{\text{Mel}}$  为 Mel 频率,  $f$  为实际频率。其在 1000 Hz 以下, 基本保持线性分布, 在 1000 Hz 以上呈对数关系。

根据 Mel 频率可以将声音频率分割成一系列的三角滤波器组, 这些三角滤波器在 Mel 频率上是等带宽的。若三角滤波器的个数为  $M$  个, 第  $m$  个滤波器的中心频率为  $f_m$ , 则第  $m$  个滤波器的传递函数为: 第  $m$  个滤波器的中心频率为  $f_m$ , 则第  $m$  个滤波器的传递函数为:

$$H_m(k) = \begin{cases} 0, & k < f_{m-1} \text{ or } k > f_{m+1} \\ \frac{2(k - f_{m-1})}{(f_{m+1} - f_{m-1})(f_m - f_{m-1})}, & f_{m-1} \leq k \leq f_m \\ \frac{2(f_{m+1} - k)}{(f_{m+1} - f_{m-1})(f_m - f_{m-1})}, & f_m \leq k \leq f_{m+1} \end{cases} \quad (2)$$

根据 Mel 频率尺度构建的三角 Mel 滤波器组如图 1 所示, 可以看出在低频部分 Mel 滤波器组分布较为密集, 高频分布较为稀疏, 符合人耳的听觉感知特性。

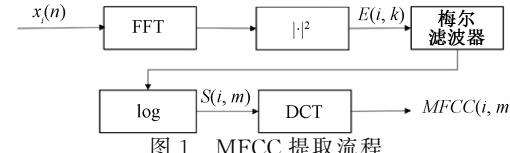


图 1 MFCC 提取流程

Fig. 1 MFCC extraction flowchart

MFCC 的提取流程如图 1 所示。

步骤 1) 预处理, 预处理包含了预加重、分帧、加窗等步骤。声音在传播过程中, 高频部分损失比低频部分更严重, 使用高通滤波器对信号进行预加重来弥补高频信号的损失, 因为声音信号是短时平稳的信号, 所以需要对信号进行分帧处理, 使每一帧是一段有着平稳特征的声音信号。为了减少每一帧信号的频谱泄露, 还需要对分帧后的信号进行加窗。语音信号  $x(n)$  经过预处理后第  $i$  帧语音信号表示为  $x_i(n)$ 。

步骤 2) 计算每一帧的谱线能量: 对每一帧信号进行 FFT 变换, 根据 FFT 变换后的数据计算出信号的谱线能量, 表示为  $E(i, k)$ 。

步骤 3) 计算通过 Mel 滤波器的能量: 将步骤 2) 中获得的谱线能量通过 Mel 滤波组, 获得 Mel 滤波器能量后取对数, 通过第  $m$  个 Mel 滤波器所获得对数能量  $S(i, m)$ :

$$S(i, m) = \log \left[ \sum_{k=0}^{N-1} E(i, k) H_m(k) \right] \quad (3)$$

步骤 4) 计算倒谱: 最后对获得的对数能量做离散余弦变换获得 MFCC 特征<sup>[16]</sup>:

$$MFCC(i, m) = \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} S(i, m) \cos \left[ \frac{\pi n (2m-1)}{2M} \right] \quad (4)$$

### 1.2 PCA

PCA 是广泛应用的降维方法, 同时也可以去除数据之

间的相关性。PCA 可以将  $n$  维特征映射至  $k$  维, 通过计算数据的协方差矩阵的特征值和特征向量, 选择前  $k$  大的特征值所对应的特征向量组成转换矩阵, 这样就可以将  $n$  维数据矩阵转换为  $k$  维数据矩阵。其具体步骤如下:

步骤 1) 对数据进行中心化, 这一步可以消除平移效应, 让数据集的中心保持在坐标原点。

步骤 2) 计算协方差矩阵。协方差矩阵的对角线元素表示了数据矩阵各维度的方差, 剩下的元素表示了数据矩阵不同维度的协方差。

步骤 3) 计算特征值和特征向量。对协方差矩阵进行数学计算获得特征值与特征向量。特征值代表了该成分的重要性, 特征向量则代表了该成分的方向。

步骤 4) 选取主成分并映射到新的坐标空间。将特征值按从大到小的顺序排列, 选取前  $k$  个特征值对应的特征矩阵组成投影矩阵, 将原始数据矩阵映射到新空间中, 完成降维<sup>[17]</sup>。

### 1.3 粒子群优化算法

粒子群优化算法是一种基于群体的搜索算法, 它是通过简化社会模型得来的, 使用一种叫粒子的随机群体进行初始化的, 每一个粒子都有一个速度, 代表一种潜在解, 粒子可以根据其历史行为对速度大小和方向进行调整, 使用调整后的速度飞过搜索空间, 在搜索的过程中, 粒子会朝着越来越好的搜索空间飞去。它的设计初衷是模拟鸟类在麦田中寻找食物, 鸟类会通过与周围(在其附近)的其他鸟类通过合作来寻找食物, 然后再将这种行为扩展到多维搜索 PSO 可以算法被描述为如下公式:

$$\mathbf{v}_{id} = w\mathbf{v}_{id} + c_1 \text{rand}(0)(\mathbf{p}_{id} - \mathbf{x}_{id}) + c_2 \text{Rand}()(\mathbf{p}_{gd} - \mathbf{x}_{id}) \quad (5)$$

$$\mathbf{x}_{id} = \mathbf{x}_{id} + \mathbf{v}_{id} \quad (6)$$

其中,  $w, c_1$  和  $c_2$  为常数,  $w$  为惯性权重, 用于平衡全局和局部的搜索能力,  $c_1$  和  $c_2$  为学习因子,  $\text{rand}()$  和  $\text{Rand}()$  是用于生成  $[0, 1]$  范围内的数的两个随机函数;  $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{id})$  表示第  $i$  个粒子;  $\mathbf{P}_i = (\mathbf{p}_{i1}, \mathbf{p}_{i2}, \dots, \mathbf{p}_{id})$  表示第  $i$  个粒子的之前的最佳位置;  $g$  表示种群中所有粒子中最好的粒子的索引;  $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{id})$  表示第  $i$  个粒子的速度。式(5)展示了粒子群中所有粒子的速度变化, 它由 3 个部分组成, 第 1 个部分是动量部分, 因为速度不能突然改变, 第 2 个部分是“认知部分”, 是从该粒子自身的飞行经验中学习到的部分; 第 3 部分是“社会”部分, 是该粒子从群体飞行经验中学习到的部分。式(6)是该飞行粒子的位置更新过程。组合在一起就是群体中某个粒子的飞行轨迹方程。

PSO 的实施流程如下:

步骤 1) 在  $D$  维的问题空间上初始化粒子群, 使粒子群具有随机的位置和速度。

步骤 2) 对于每个粒子, 计算适应度函数的值。

步骤 3) 将每个粒子的适应度函数值与其之前的最佳

位置的值  $\mathbf{P}_{best}$  进行比较, 如果当前值优于  $\mathbf{P}_{best}$  则更新  $\mathbf{P}_{best}$  和  $\mathbf{P}_i$  的值。

步骤 4) 识别至今为止最好的邻域粒子, 更新最佳索引  $g$  的值。

步骤 5) 改变粒子的速度和位置, 循环步骤 2), 直到满足一个标准, 如达到最好的适应度函数或者达到最大迭代次数时退出循环<sup>[18]</sup>。

## 2 改进的 MFCC 特征

在识别无噪声的声音信号时, 传统的 MFCC 特征的分类效果较好, 但 MFCC 特征对噪音感染特别敏感, 为了提高 MFCC 特征的抗干扰性, 提升训练集和测试集的特征匹配度, 扩宽 MFCC 特征使用的场景限制, 本文对 MFCC 特征进行了改进。

### 2.1 基于 PCA 改进的 MFCC 特征

由于 Mel 滤波器在频域上有所重叠, 所以经过 Mel 滤波后的对数能量存在一定程度的相关性, 在传统的 MFCC 方法中, 使用 DCT 对对数能量进行去相关来获得最终的倒谱特征。DCT 是一种纯粹的数学矩阵运算, 与实际的数据没有任何关系, 未贴合实际数据的特征<sup>[19]</sup>, 就会导致传统的 MFCC 特征对环境适应性的下降。如果基于实际的数据的训练集来计算变换矩阵, 再将此矩阵应用于实际数据获得经过改进的倒谱特征, 新获得的特征是由大量实际数据驱动得来, 能充分表征实际数据的特点, 有着更强的可分性和抗干扰能力。由于 PCA 变换不仅可以对数据进行降维, 也可以对数据去相关, 所以使用 PCA 变换代替 MFCC 特征提取中的 DCT, 获得全新的 PCA-MFCC 特征。PCA-MFCC 特征的提取的流程如图 2 所示。

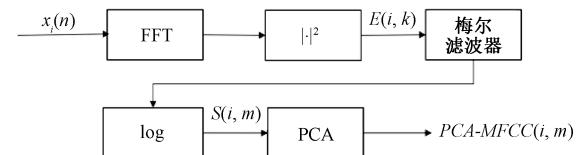


图 2 PCA 改进的 MFCC 提取流程

Fig. 2 PCA-improved MFCC extraction flowchart

### 2.2 基于非线性映射改进的 MFCC 特征

野外目标的声音信号都集中于低频部分, 可以通过强调低频部分的特征来增强 MFCC 的抗干扰能力。本文使用两种非线性映射函数<sup>[20]</sup>来增强声学特征的低频部分。非线性映射函数应该满足非负、递减、非线性这 3 个条件, 经过声音经过梅尔滤波器后所获得的对数能量是非负的所以映射函数应该是非负的, 为了强调低频部分所以映射函数应该是随着频率减小的, 而非线性的映射函数能够强调声学特征的非线性。

提出的两种非线性映射函数如下:

$$MAP_1(x) = 1 - \frac{a}{1 + e^{-b(x-c)}} \quad (7)$$

$$MAP_2(x) = a + \frac{1}{(1+bx)^c} \quad (8)$$

非线性映射函数的常数参数值可以通过粒子群优化算法来寻找,适应度函数为野外目标识别的准确率。

MAP1-MFCC 和 MAP2-MFCC 的特征提取流程如图 3 所示。

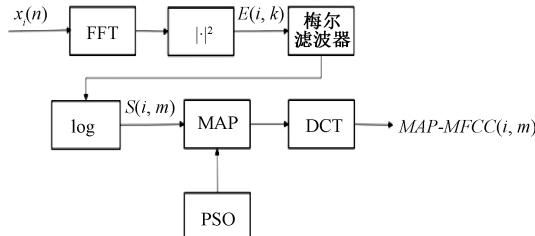


图 3 非线性映射改进的 MFCC 提取流程

Fig. 3 Nonlinear mapping-improved MFCC extraction flowchart

与传统的 MFCC 的不同在于将式(3)的提取过程改进为式(9)。

$$S(i, m) = \text{map}(m) \log \left[ \sum_{k=0}^{N-1} E(i, k) H_m(k) \right] \quad (9)$$

### 2.3 基于 PCA 和非线性映射改进的 MFCC 特征

将 PCA 和非线性映射都应用于 MFCC 特征提取过程中, 可以获得 MAP1-PCA-MFCC 和 MAP2-PCA 特征, 其特征提取的流程如图 4 所示。

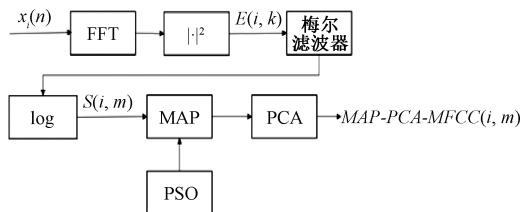


图 4 PCA 与非线性映射改进的 MFCC 提取流程

Fig. 4 PCA and nonlinear mapping-improved MFCC extraction flowchart

### 3 仿真实验

为了验证本文提出的特征提取算法的有效性,使用在外场采集的数据集,并对其进行降噪处理,对数据集分别提取 MFCC、PCA-MFCC、MAP-MFCC、PCA-MAP-MFCC 特征,并使用 XGBoost 对其进行分类<sup>[21]</sup>,对比实验结果。实验流程包括数据采集、特征提取和分类 3 个步骤。

### 3.1 数据集

数据集是来自于在外场采集的数据集,采集场景如下图所示,使用的采集设备是 ICS40720 传声器,传感器布设位置到道路中心的位置小于 50 m,道路总长为 1 000 m 以内,采样率为 8 192 Hz,一共采集了三种车辆的声音信号,自制数据集包括了轻型轮式车(LW)、装甲车(AW)和履带车(T)。信号原本的采样频率为 8 192 Hz,本方法先将其降采样至 1 024 Hz,设置其帧长为 1 s,帧移为 0.5 s。3 种运

动目标的帧数如表 1~2 所示。

表 1 训练集

**Table 1** Training dataset

训练集	条数	帧数
LW	15	17 321
AW	15	14 575
T	15	13 175
合计	45	45 071

表 2 测试集

**Table 2** Testing dataset

训练集	条数	帧数
LW	6	4 890
AW	6	4 796
T	6	4 812
合计	18	14 498

### 3.2 参数设置和评价指标

在声音预处理部分首先将声音信号从 8 192 Hz 降采样至 1 024 Hz, 分帧过程中帧长设置为 1 s, 帧移设置为 0.5 s。

在特征提取部分, Mel 滤波器组中的三角滤波器设置为 24 个, 提取 24 维的 MFCC 特征, MFCC 特征的第一维为直流分量, 可以去除, 最终获得 23 维的 MFCC 和 MAP-MFCC 特征, 由于 PCA 在计算时, 第一步就是去中心化, 去除了直流分量, 为了保持维度的一致, 可以直接获得 23 位的 PCA-MFCC 和 MAP-PCA-MFCC 特征。

在粒子群优化部分,适应度函数为分类准确率,最大迭代次数设置为 120 次,  $w$  为 0.8,  $c_1$  和  $c_2$  的值为 2。为了保持映射函数的非负性和防止溢出,MAP1 的参数  $a$  搜索范围为  $[0,1]$ , 参数  $b$  的搜索范围为  $[0,10]$ , 参数  $c$  的搜索范围为  $[-24,24]$ , MAP2 的参数  $a, b, c$  的搜索范围均为  $[0,5]$ 。在 XGBoost 部分,迭代次数为 100 次。在 XGBoost 部分,迭代次数为 100 次。

实验结果采用准确率,漏警率和虚警率进行衡量。

### 3.3 实验过程与结果

为了本文提出的特征提取技术的有效性,对声音数据集分别提取 MFCC, PCA-MFCC, MAP-MFCC 和 MAP-PCA-MFCC 特征,对于使用了映射函数的 MAP 类特征,每个数据集会有其最佳分类率所对应的映射函数常数参数的值,如将原始数据集 MAP-MFCC 所对应的常数参数带入 MAP-PCA-MFCC 中,比较其分类结果,可以进一步验证 PCA 改进的 MFCC 参数的有效性。同时提取文献[12]提出的混合倒谱特征,文献[13]提出的改进 MFCC 特征和文献[14]提出的 WMFCC 特征,将这些改进算法与本文的改进算法进行对比。

对数据集提取 MFCC、PCA-MFCC、MAP-MFCC、MAP-PCA-MFCC 特征,并计算出其对应的识别准确率,虚警率和漏警率。对于 MAP 类特征,其最佳识别率对应的常数参数如表 3 所示,对应的最佳参数组合记为 MAP-BEST 和 MAP-PCA-BEST。

表 3 不同特征对应的最佳参数组合

Table 3 Optimal parameter combinations for different features

特征类型	a	b	c
MAP1-MFCC	0.52	8.97	20.08
MAP1-PCA-MFCC	0.91	10.00	24.00
MAP2-MFCC	0.46	0.63	4.23
MAP2-PCA-MFCC	5.00	2.64	0.53

### 3.4 实验结果分析

从表 4 中可以看出,相比于 MFCC 特征,使用 PCA-MFCC 特征进行分类的准确率上升了 6.88%,虚警率下降了 6.35%,漏警率降低了 6.88%,再观察表 5,在使用同一种映射函数且常数参数组合相同的情况下,使用 MAP-PCA-MFCC 特征进行分类的准确率均比使用 MAP-MFCC 特征的准确率高,漏警率和虚警率均比使用 MAP-MFCC 时的低,说明使用 PCA 来代替 MFCC 提取过程中的 DCT 是有效的。

表 4 不同特征的分类结果

Table 4 Classification results for different features %

特征类型	准确率	虚警率	漏警率
MFCC	86.24	12.96	13.76
文献[12]	86.70	12.45	13.3
文献[13]	86.94	12.03	13.06
文献[14]	87.25	12.4	12.75
PCA-MFCC	93.12	6.61	6.88
MAP1-MFCC	88.50	10.75	11.50
MAP2-MFCC	86.49	13.01	13.51
MAP1-PCA-MFCC	93.36	6.40	6.64
MAP2-PCA-MFCC	93.32	6.42	6.68

从表 4 可以看出,相比于 MFCC 特征,使用 MAP1-MFCC 特征进行分类的准确率相比 MFCC 特征上升了 2.26%,虚警率下降了 2.21%,漏警率降低了 2.26%;相比于 MFCC 特征使用 MAP2-MFCC 特征进行分类时,准确率上升了 0.25%,虚警率上升了 0.05%,漏警率下降了 0.25%,说明使用非线性映射函数改进后的 MFCC 可以提升目标识别的准确率,并且 MAP1 映射函数的效果比 MAP2 好。

从表 4 可以看出,相比于 MFCC 特征,PCA-MFCC 特征和 MAP-MFCC 特征,MAP-PCA-MFCC 的准确率是最

表 5 使用相同参数组合的 MAP 类特征分类结果

Table 5 Classification results of MAP features using the same parameter combination %

常数参数组合	特征类型	准确率	虚警率	漏警率
MAP1-BEST	MAP1-MFCC	88.50	10.75	11.50
	MAP1-PCA-MFCC	92.78	6.87	7.22
MAP1-PCA-BEST	MAP1-MFCC	87.82	11.11	12.17
	MAP1-PCA-MFCC	93.36	6.40	6.64
MAP2-BEST	MAP2-MFCC	86.49	13.01	13.51
	MAP2-PCA-MFCC	92.65	6.99	7.35
MAP2-PCA-BEST	MAP2-MFCC	85.54	13.53	14.46
	MAP2-PCA-MFCC	93.32	6.42	6.68

高的,MAP1-PCA-MFCC 的准确率比 MFCC 特征高 7.12%,比 PCA-MFCC 特征高 0.24%,比 MAP1-MFCC 特征高 4.86%,MAP2-PCA-MFCC 的准确率比 MFCC 特征高 7.08%,比 PCA-MFCC 特征高 0.2%,比 MAP2-MFCC 特征高 6.83%,可以证明经过 PCA 和非线性映射改进的 MFCC 是最有效的,并且 MAP1 的映射效果更好。

从表 4 可以看出,使用文献[12]、文献[13]、文献[14]改进后的 MFCC 特征的识别准确率均高于传统的 MFCC 特征,虚警率和误警率均低于传统的 MFCC 特征。且文献文献[12]、文献[13]、文献[14]改进后的 MFCC 特征的识别准确率低于本文提出的除 MAP2-MFCC 以外所有改进方案的准确率,进一步验证了经过 PCA 和非线性映射改进后的 MFCC 的有效性和 MAP1 映射函数的更优性。

### 4 结论

针对 MFCC 特征易受环境干扰导致野外目标识别准确率低等问题,从 MFCC 的提取流程入手,基于 PCA 和非线性映射函数对 MFCC 进行改进,PCA 在去相关的同时使 MFCC 特征更加贴近实际数据的特征,增强了特征对实际样本的适应性,非线性映射函数能够强调目标声音信号所在的低频部分的特征,增强了声学特征的非线性,突出了目标信号所在的频段,在自制数据集上设计实验,使用 MFCC 特征和改进的 MFCC 特征对目标进行分类,实验结果说明了这两种改进思路分开使用和联合使用均有效,且使用 MAP1-PCA-MFCC 特征时准确率达到了 93.36%。在后续的研究中,可以对声音信号进行增强,进一步减少风的干扰,提升野外目标识别的准确率。

### 参考文献

- [1] QIAN Y L, TANG H Y, RAN Y, et al. Target classification in unattended ground sensors with a two-stream convolutional network [J]. IEEE Sensors Journal, 2022, 23(4): 3747-3755.
- [2] PANDEY M, MISHRA G. Types of sensor and their

- applications, advantages, and disadvantages [C]. Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 3. Springer Singapore, 2019: 791-804.
- [3] SUNU J, PERCUS A G. Dimensionality reduction for acoustic vehicle classification with spectral embedding[C]. 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). IEEE, 2018: 1-5.
- [4] 韩令敏, 陈仙红, 熊文梦. 基于语音和文本的双模态情感识别综述[J]. 计算机应用, 2025, 45(4): 1025-1034.
- HAN L M, CHEN X H, XIONG W M. A survey of bimodal emotion recognition based on speech and text[J]. Computer Applications, 2025, 45(4): 1025-1034.
- [5] KUMARAN U, RADHA R S, NAGARAJAN S M, et al. Fusion of mel and gammatone frequency cepstral coefficients for speech emotion recognition using deep C-RNN [J]. International Journal of Speech Technology, 2021, 24(2): 303-314.
- [6] 王延凯, 龙华, 邵玉斌, 等. 基于联合端点检测和动态范围控制的语种识别[J]. 激光与光电子学进展, 2022, 59(13): 205-212.
- WANG Y K, LONG H, SHAO Y B, et al. Language recognition based on joint endpoint detection and dynamic range control [J]. Laser & Optoelectronics Progress, 2022, 59(13): 205-212.
- [7] SHARMA R, VIGNOLO L, SCHLOTTHAUER G, et al. Empirical mode decomposition for adaptive AM-FM analysis of speech: A review [J]. Speech Communication, 2017, 88: 39-64.
- [8] 周大春, 邵玉斌, 张昊阁, 等. 应用于噪声环境下语种识别的 GFCC 改进算法[J]. 云南大学学报(自然科学版), 2024, 46(2): 246-254.
- ZHOU D CH, SHAO Y B, ZHANG H G, et al. An improved GFCC algorithm for language recognition in noisy environments[J]. Journal of Yunnan University (Natural Sciences), 2024, 46(2): 246-254.
- [9] MASUMURA R, ASAMI T, MASATAKI H, et al. Parallel phonetically aware DNNs and LSTM-RNNs for frame-by-frame discriminative modeling of spoken language identification [C]. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017: 5260-5264.
- [10] 张楠, 陈媛媛, 陈鑫钰, 等. 基于 LMD 改进特征提取的三路病理语音识别[J]. 电子测量技术, 2024, 47(12): 140-147.
- ZHANG N, CHEN Y Y, CHEN X Y, et al. Pathological speech recognition based on LMD improved feature extraction [J]. Electronic Measurement Technology, 2024, 47(12): 140-147.
- [11] ZHAO Q, GUO F, ZU X SH, et al. An acoustic signal enhancement method based on independent vector analysis for moving target classification in the wild[J]. Sensors, 2017, 17(10): 2224.
- [12] 陶瀚宇, 陈换过, 彭程程, 等. 基于 MFCC-IMFCC 混合倒谱的托辊轴承故障诊断[J]. 机电工程, 2024, 41(7): 1215-1222.
- TAO H Y, CHEN H G, PENG CH CH, et al. Fault diagnosis of idler bearings based on MFCC-IMFCC hybrid cepstrum [J]. Mechanical & Electrical Engineering, 2024, 41(7): 1215-1222.
- [13] 张伟, 杨宏波, 潘家华, 等. 基于改进 MFCC 与 IMFCC 的心音分类研究[J]. 计算机仿真, 2023, 40(10): 346-352.
- ZHANG W, YANG H B, PAN J H, et al. Research on heart sound classification based on improved MFCC and IMFCC[J]. Computer Simulation, 2023, 40(10): 346-352.
- [14] 岳倩倩, 周萍, 景新幸. 基于非线性幂函数的听觉特征提取算法研究[J]. 微电子学与计算机, 2015, 32(6): 163-166.
- YUE Q Q, ZHOU P, JING X X. Research on auditory feature extraction algorithm based on nonlinear power function [J]. Microelectronics & Computers, 2015, 32(6): 163-166.
- [15] PALO H K, CHANDRA M, MOHANTY M N. Recognition of human speech emotion using variants of mel-frequency cepstral coefficients [J]. Advances in Systems, Control and Automation, 2018: 491-498.
- [16] HIZLISOY S, ARSLAN R S. Text independent speaker recognition based on MFCC and machine learning[J]. Selcuk University Journal of Engineering Sciences, 2021, 20(3): 73-78.
- [17] GREENACRE M, GROENEN P J F, HASTIE T, et al. Principal component analysis[J]. Nature Reviews Methods Primers, 2022, 2(1): 100.
- [18] MARINI F, WALCZAK B. Particle swarm optimization(PSO). A tutorial[J]. Chemometrics and Intelligent Laboratory Systems, 2015, 149: 153-165.
- [19] 张玉欣. 声信号的分解方法及其在车型分类中的应用研究[D]. 重庆: 重庆交通大学, 2020.
- ZHANG Y X. Decomposition methods of acoustic signals and their application in vehicle type classification [D]. Chongqing: Chongqing Jiaotong University, 2020.
- [20] DING L, XIAO L, ZHOU K Q, et al. A new RNN

- model with a modified nonlinear activation function applied to complex-valued linear equations[J]. IEEE Access, 2018, 6: 62954-62962.
- [21] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system [C]. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.

## 作者简介

符恬恬(通信作者),硕士研究生,主要研究方向为声音信

号处理与目标识别。

E-mail: futiantian22@mails.ucas.ac.cn

郑斌琪,正高级工程师,硕士生导师,主要研究方向为特种无线传感网研究。

E-mail: bqzheng@mail.sim.ac.cn

李成娟,中级工程师,主要研究方向为数字信号处理。

E-mail: lichengjuan@mail.sim.ac.cn

夏利杰,正高级工程师,主要研究方向为边缘计算。

E-mail: xialj@mail.sim.ac.cn