计算机科学
COMPUTER SCIENCE

一个面向短波通信的LHOG话音检测方法

白洁, 田瑞丽, 任一夫, 员建厦

引用本文

白洁, 田瑞丽, 任一夫, 员建厦. 一个面向短波通信的LHOG话音检测方法[J]. 计算机科学, 2024, 51(6A): 230600115-5.

BAI Jie, TIAN Ruili, REN Yifu, YUAN Jianxia. Low-rank HOG Voice Detection Method for Short-wave Communication [J]. Computer Science, 2024, 51(6A): 230600115-5.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230600115

一个面向短波通信的 LHOG 话音检测方法

白 洁[1] 田瑞丽[2] 任一夫[1] 员建厦[1]
1 中国电子科技集团公司第五十四研究所 石家庄 050081
2 中国联合网络通信有限公司河北省分公司 石家庄 050051

摘 要 噪声环境下语音检测准确率偏低是短波通话面临的公开挑战。当前已有方法应用有限,其根源在于难以可靠地在噪音环境下提取准确且高效的语音特征。针对上述问题,提出了一个面向短波通信的低秩方向梯度直方图(Low-rank Histogram of Oriented Gradient,LHOG)话音检测方法。首先,对目标音频源数据进行预处理,实现噪声环境下语音信息的可视化表征;然后,在 HOG 特征提取器中嵌入低秩化结构,缓解特征中的冗余信息,并降低噪声干扰,从而获得准确且高效的特征;最后,通过常用的 SVM 分类模型便可在噪声环境中准确快速地区分话音和噪声。测试结果表明,该方法的准确率达到了 95.12%,误报率仅为 0.96%,漏报率为 13.14%。与现有主流方法的对比实验证明,该方法话音检测准确率高,资源占用少,能够有效提高短波通信侦控效率。

关键词:模式识别;语谱图;方向梯度直方图;低秩结构;支持向量机
中图分类号 TP391.4

Low-rank HOG Voice Detection Method for Short-wave Communication

BAI Jie[1],TIAN Ruili[2],REN Yifu[1] and YUAN Jianxia[1]
1 The 54th Research Institute of CETC,Shijiazhuang 050081,China
2 China United Network Communications Co.,Ltd. Hebei Branch,Shijiazhuang 050051,China

Abstract The low accuracy of voice detection in noisy environment is an open challenge for short wave communication. The application of existing methods is limited,because it is difficult to reliably extract accurate and efficient voice features in the noise environment. To solve the above problem,a Low-rank histogram of oriented gradient(LHOG) voice detection method for short wave communication is proposed in this paper. Firstly,target audio source data is preprocessed to realize visual representation of voice information in noisy environment. Then,a low-rank structure is embedded in the HOG feature extractor to alleviate redundant information and reduce noise interference,so as to obtain accurate and efficient features. Finally,the common SVM classification model can be used to reliably distinguish voice from noise in noisy environment. The test results show that the accuracy of this method is 95.12%,the false positive rate is 0.96%,and false negative rate is 13.14%. Compared with the existing mainstream methods,the experiment shows that the average detection accuracy of this method is higher,and resource occupation is less. Therefore,this method can effectively improve the detection and control efficiency of short-wave communication.

Keywords Pattern recognition,Spectrogram,HOG,Low-rank structure,SVM

在短波侦控工作中,需要对已知频率的短波电台进行全天候控守,以免漏掉突发重要情况。短波信号经过采集、识别、解调[1]后生成音频数据,再由人工进行筛选和复听。为了节省人力,目前多采用自动手段对录音进行检测[2-4],筛选出有效话音后再通过智能语音识别或人工听抄的方式进行处理。然而随着空间频谱环境的恶化,射频频谱变得非常拥挤,信道间相互干扰严重[5],导致被控频率存在强烈的噪声和干扰信号,严重限制了短波通信的发展。因此,亟需发展兼顾准确性与时效性的话音检测方法。

传统语音信号处理方法常用梅尔频率倒谱系数(Mel Frequency Cepstral Coefficents,MFCC)提取音频数据特征,再用随机森林(Random Forest,RF)、支持向量机(Support Vector Machine,SVM)、高斯混合模型(Gaussian Mixture Model,GMM)等传统分类器进行分类[6]。近年来,随着人工智能技术的发展和计算能力的提升,深度学习算法在音频识别领域展现出了优异的性能,如卷积神经网络(Convolutional Neural Networks,CNN)和长短时记忆网络(Long Short Term Memory Network,LSTM)现已被广泛使用。越来越多的科研人员将深度学习算法和传统机器学习算法相结合来优化识别和分类效果[7]。例如,Chen 等[4]提取音频信号的梅尔频率倒谱系数作为特征向量,再以支持向量机作为分类器识别汽车喇叭声,取得了较高的识别准确率;Sailor 等[8]利用 CNN 和卷积限制玻尔兹曼机进行环境音频分类,得到的结果比传统方法有所提高;Chen 等[9]用一维深度卷积神经网络与

基金项目:河北省智能化信息感知与处理重点实验室发展基金项目(SXX22138X002)
This work was supported by the Development Fund Project of Hebei Key Laboratory of Intelligent Information Perception and Processing (SXX22138X002).
通信作者:白洁(13086194@qq.com)

梅尔滤波器组特征集成对音频场景进行分类，取得了不错的效果；Choi 等[10]用 CNN 融合梯度向量机算法生成新的模型对音频场景进行分类，取得了较高的准确率；Qiu 等[11]把从语音中提取的 FBANK 特征作为输入，训练基于双向长短时记忆网络模型对民航陆空通话进行识别，取得了较好的效果。但是上述各种方法也存在一些不足。例如，用 MFCC 提取特征要经过短时傅里叶变换（Short-Time Fourier Transform，STFT）、梅尔滤波器组过滤、对数运算、离散余弦变换等多个步骤，对算力要求较高，资源占用较多，且信号经过多次转换后丢失了很多细节，因此 MFCC 对噪音的适应性不好，当噪音出现时，基于 MFCC 特征的分类效果就明显下降[12]；CNN 和 LSTM 等深度神经网络模型则调参复杂，对训练样本需求量大，模型训练难度较高，且深度神经网络模型运算量大，导致对硬件要求高，使得模型应用成本较高。

为解决上述问题，本文提出一个新颖的 LHOG 短波通话话音检测方法。选择 HOG 特征，是基于其在图像处理领域的广泛应用和良好的性能。HOG 特征在目标检测和图像分类任务中被广泛采用，其优势在于对局部形状和边缘信息的敏感性。与采用其他图像特征相比，在语谱图特征提取中，HOG 特征可以捕捉到音频频谱中的边缘和纹理信息，可有效提高语音识别的准确性和鲁棒性。本文的创新性在于：在 LHOG 特征提取器中设计了低秩化结构，能够有效地缓解特征中的冗余信息，并降低噪声干扰，从而获得准确且高效的话音特征。此外，由于通过设计的 LHOG 可获得准确高效的话音特征，使用常规的 SVM 分类模型就能够可靠地区分话音和噪声，因而与 CNN 和 LSTM 等深度神经网络模型相比，本文检测方法具有运算量少、执行速度快等优点。

1 LHOG 短波通信话音检测方法

LHOG 分为训练和检测两个阶段。在训练阶段，先要对音频数据进行标注，区分话音和噪音，再做短时傅里叶变换生成语谱图，从语谱图中提取 LHOG 特征后作为输入去训练 SVM 分类器。在检测阶段，直接用待检测音频数据生成语谱图，再从语谱图中提取 LHOG 特征输入训练好的 SVM 分类器进行分类，检测并提取出音频数据中的话音。LHOG 短波通信话音检测方法的实现步骤如图 1 所示。

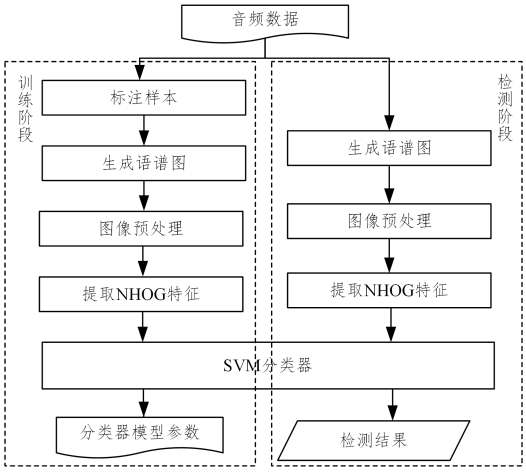


图 1 方法的实现步骤

Fig. 1 Implementation steps of the proposed method

1.1 音频数据预处理

语谱图作为一种常用的语音特征表示方法，已经被广泛应用于语音信号处理和语音识别任务中。它将语音信号在时间和频率上进行了分析，提供了关于声音能量、频谱轮廓以及语音中的共振峰等信息。已有研究表明，与传统的语音特征相比，语谱图在需要更多频谱信息的情况下，能够提供更多细节和上下文信息，有助于语音信号的分析和识别。

受此启发，本文采用短时傅里叶变换（STFT）处理音频数据来生成语谱图。其过程是先通过分帧获得一段长音频的短时平稳信号，再将每一帧乘以 Hamming 窗，以增加帧两端的连续性，再通过快速傅里叶变换（Fast Fourier Transformation，FFT）获得时频信号，如式（1）所示：

$$X_m(k,i)=\sum_{n=0}^{L-1}X_m(n)\mathrm{e}^{-\frac{i2\pi k}{n}} \tag{1}$$

其中，m 表示帧数，L 表示帧长，i 表示第几帧[13]。

最后将每一帧频域信号相叠加得到语谱图。噪音和话音经过短时傅里叶变换后生成的语谱图如图 2 所示，该图横坐标代表时间（单位为 s），纵坐标代表频率（单位为 Hz）。

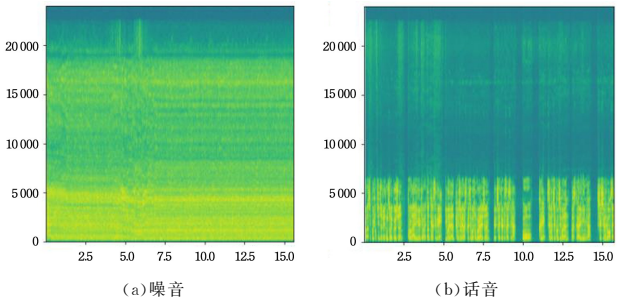


图 2 噪音与话音的语谱图对比

Fig. 2 Spectrogram comparison of noise and vioce

1.2 LHOG 特征提取

1.2.1 传统 HOG 特征提取方法

本文提取语谱图的 LHOG 特征作为分类器的输入。HOG 特征是一种在计算机视觉和图像处理中常用的特征描述子[14]，该特征通过计算和统计图像局部区域的水平和竖直的梯度幅值及像素点的梯度方向来构成特征[15]，在图像处理领域得到非常广泛的应用[16]，如人脸检测[17]、行人检测[18]等。首先，算法将图像分成若干个小的连通区域，称为胞元（cell），将多个胞元（cell）组成像素块（block），以 block 为单位计算其中 cell 的所有像素处的梯度幅值和梯度方向；然后，统计每个 cell 的梯度方向特征向量；最后，将所有 cell 的梯度方向特征向量按照空间顺序组合起来，以保存空间信息。详细的 HOG 特征提取步骤如下。

1）采用非线性 Gamma 校正方法，调节图像亮度的不均匀性，降低图像阴影和光照变化带来的影响，如式（2）所示：

$$g(I)=I^{\gamma} \tag{2}$$

其中，I 是图像中某点的像素值，γ 为 Gamma 值。当 $\gamma<1$ 时，$g(I)$ 运算可提高低灰度值区间的像素对比度；当 $\gamma>1$ 时，$g(I)$ 运算可增强高灰度值区间的对比度。

2）将语谱图分解成固定大小的 cell，再将相邻的 cell 组合形成 block。进行梯度计算时，从图像的一角开始，以 cell 的边长作为前进的步长，遍历整幅图像。

3）选择水平梯度算子[－1 0 1]，竖直梯度算子[1 0 －1]$^{\mathrm{T}}$。首先用水平梯度算子对原图像做卷积运算，得到 x 方向的梯度分量$G_x(x,y)$；然后用竖直梯度算子对原图像做卷积运算，得到 y 方向的梯度分量$G_y(x,y)$[16]。每个像素的

梯度幅值 $G(x,y)$和方向 $\theta(x,y)$的计算公式如下：

$$G(x,y)=\sqrt{G_x(x,y)^2+G_y(x,y)^2} \tag{3}$$

$$\theta(x,y)=\tan^{-1}(G_y(x,y)/G_x(x,y)) \tag{4}$$

4)将360°方向平均分成 9 个条带(bin)，角度区间分别为[0,40)，[40,80)，[80,120)，[120,160)，[160,200)，[200,240)，[240,280)，[280,320)，[320,360)。以 cell 为单位，遍历 cell 内的所有像素点，统计落入每个 bin 角度区间内的梯度幅值，形成单个 cell 的特征，即一个 9 维的向量[19]。

5)将 block 中所有 cell 的特征串联起来，构成该 block 的特征向量，然后使用式(5)进行归一化，即把每个特征除以该特征向量的范数，以消除光照对图像的影响。

$$v^* \leftarrow \frac{v}{\sqrt{\| v \|_2^2+\omega^2}} \tag{5}$$

其中，v 表示特征向量；$\| v \|_k$表示 v 的 k 阶范数；ω 为一个取值很小的系数，用于防止分母为 0。

6)最后，将所有 block 的特征向量串联起来就是图像的最终特征向量，即 HOG 特征。

1.2.2　LHOG 特征提取方法

区别于传统 HOG 特征提取方法，LHOG 中的低秩化机制主要体现在 3 个方面：取消 block 运算，裁剪归一化运算，基于 PCA 方法的特征降维。上述操作在保证话音特征信息准确性的同时，能够有效地降低特征维度，提高话音检测效率。

具体而言，采用传统 HOG 特征提取方法提取特征向量，在步骤 2)中会划分 cell 和 block。设语谱图分辨率为 240×600，cell 大小取 12×12，block 大小取 24×24，扫描步长取 12，可将语谱图分解为 19×49=931 个 block，每个 block 包含 4 个 cell，则特征向量的维数为：931×4×9=33 516。这是由于以 block 为单位扫描语谱图时会有 cell 在 block 中重复被计算的情况，而特征总数过多将会影响特征提取效率，延长算法运行时间。步骤 5)中以 block 为单位进行归一化运算的目的是调节图像亮度，降低光照变化对图像带来的影响。而语谱图是对音频文件进行短时傅里叶变换得到的时频映射图形，不存在光照变化对其特征提取的影响。如图 3 所示，经过对 HOG 特征提取方法进行优化，将语谱图直接分解为 cell，不再组合形成 block，而是直接以 cell 为单位进行梯度计算，这样分辨率为 240×600 的语谱图将分解为 20×50=1 000 个 cell，每个 cell 有 9 个特征，特征向量的维数降为 1 000×9=9 000 个，可有效减轻运算压力。另外，裁剪步骤 5)归一化运算，也可减轻预算压力。

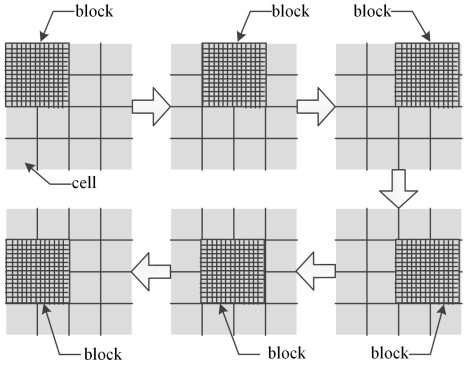


图 3　扫描图像进行梯度计算示意

Fig. 3　Schematic of scanning image for gradient calculation

上述操作能够有效减少特征向量的维数，但特征中仍然存在冗余和干扰。可采用主成分分析方法(Principal Components Analysis，PCA)对特征信息进行处理、压缩和抽提，将高维特征映射到低维上。方法步骤如下：

1)将分辨率为 240×600 的语谱图的每一列相互连接构成一个 144 000 维的向量，则数量为 n 的样本集可以表示向量 $X=\{x_i \in R^{144000}, i=1,2,\cdots,n\}$，其中 x_i 表示第 i 个样本构成的向量。

2)将样本向量 X 的每一行进行零均值化运算，得到向量 A。令 η 为均值向量：

$$\eta=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{6}$$

则有 $A=\{x_1-\eta_1, x_2-\eta_2, \cdots, x_n-\eta_n\}$

3)用式(7)求出协方差矩阵 C。

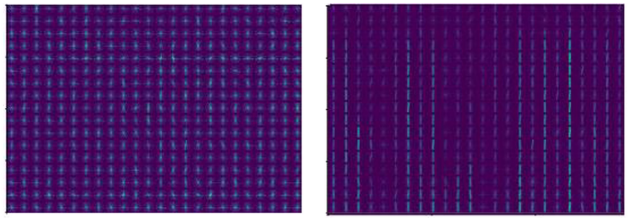
$$C=\frac{1}{n}AA^{\mathrm{T}} \tag{7}$$

4)求出协方差矩阵 C 的特征向量 μ。

5)将特征向量按对应特征值大小从上到下按行排列成矩阵，取前 k 行组成矩阵 P。

6)$Y=PX$ 即为降维到 k 维后的数据。

由此，通过 LHOG 便可得到准确且高效的特征信息，具体可视化评估表现如图 4 所示。



(a)噪音　　(b)话音

图 4　噪音和话音的 HOG 特征对比

Fig. 4　HOG features comparison of noise and voice

1.3　话音与噪声分类模型

通过上述设计的 LHOG 可获得准确高效的话音特征，因此使用常规的 SVM 分类模型就能够可靠地区分话音和噪声。本文中的 SVM 分类模型选用线性核函数从语谱图中提取 LHOG 特征为训练样本数据，通过训练计算求得分离超平面函数后，构建分类决策函数，使用时用分类决策函数对输入进行分类。

1)训练分类器

训练过程的输入为训练数据集 T，输出为分离超平面的函数。

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$$

$$x_i \in R^n, y_i \in \{+1,-1\}, i=1,2,\cdots,N$$

选择惩罚参数 $C>0$，构造并求解以下二次规划问题。其中，α_i 为拉格朗日乘子，y_i 为数据分类标签，且 $y \in \{-1,1\}$；N 是样本的个数。

$$\begin{cases}\min\limits_{\alpha} \dfrac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j)-\sum\limits_{i=1}^{N}\alpha_i \\ \text{s.t.} \sum\limits_{i=1}^{N}\alpha_i y_i=0, 0\leqslant\alpha_i\leqslant C, i=1,2,\cdots,N\end{cases} \tag{8}$$

得到最优解 $\alpha^*=(\alpha_1^*,\alpha_2^*,\cdots,\alpha_N^*)^{\mathrm{T}}$ 后，通过计算得到分类超平面的参数 ω^* 和 b^*。

$$\omega^*=\sum_{i=1}^{N}\alpha_i^* y_i (x_i \cdot x_j) \tag{9}$$

选择α^*的一个分量α_j^*满足条件$0<\alpha_j^*<C$，计算b^*。

$$b^*=y_j-\sum_{i=1}^{N}\alpha_i^* y_i(x_i\cdot x_j) \tag{10}$$

再用参数ω^*和b^*构建如式(11)所示的分离超平面的函数。

$$\omega^*\cdot x+b^*=0 \tag{11}$$

2)分类判断

根据分离超平面的函数构建如式(12)所示的分类决策函数，以待检测音频数据为输入，做出分类判断。当$0<y\leqslant 1$时，音频数据属于话音；当$-1\leqslant y\leqslant 0$时，音频数据属于噪音。

$$y=sign(\omega^*\cdot x+b^*) \tag{12}$$

2 实验结果与分析

2.1 实验环境

本文实验的硬件环境：CPU为Intel i5-11400F，内存16 GB。软件环境：操作系统为Windows 10，编程平台为pyCharm2019.2.1，模型库为OpenCV v4.5.4。全部实验程序均基于python语言构建，基于OpenCV组件从语谱图中提取LHOG特征，并实现SVM分类器的训练与应用。

2.2 实验数据集

本文实验使用从工程实际运行过程中采集并积累的话音和噪音数据构建样本集，用于训练和测试，采样率为48 kHz，存储格式为wav。话音数据来自约20个不同的说话人，从语种上涵盖了汉语、英语、俄语、日语、韩语、蒙语，从性别上涵盖男性和女性，从对话形式上涵盖了双男性对话、双女性对话、男女对话、单男性自话、单女性自话，每个音频文件时长5 s，共7451个，总时长约10.3 h；噪音数据来自不同音量、不同频段下的短波通信背景噪音，每个音频文件时长5 s，共15 680个，总时长约21.8 h。

2.3 对比实验方法

本文实验的目的是验证LHOG方法的准确率和资源占用率。为了验证本文方法的有效性，选用常规的HOG+SVM方法、基于多优化机制的CNN方法、MFCC+SVM方法作为对比方法。未优化的HOG+SVM方法直接使用传统HOG特征提取方法提取语谱图特征，再通过SVM分类器进行分类；基于多优化机制CNN方法采用CNN和双向门控循环单元网络构建基本并行框架，再融合批标准化机制和分层注意力机制网生成模型用于音频分类识别[10]；MFCC+SVM方法提取音频数据的梅尔倒谱系数特征作为输入，训练SVM分类器用于分类判断[4]。

2.4 实验参数设置

1)HOG+SVM方法参数设置

对音频数据做短时傅里叶变换时，窗口长度为2048，步长为512，帧重叠率为窗口大小的1/4，生成的语谱图分辨率为240×600。

HOG特征提取窗口大小为240×600，cell尺寸为12×12，bin数量为9，每个窗口的特征总数为9000个，经过PCA方法特征降维后特征总数降为8647个。

SVM分类器选择线性核函数，训练误差上限Nu为0.5，损失函数系数P为0.1，惩罚系数C为0.01。

2)MFCC+SVM方法参数设置

MFCC预加重系数为0.97，加窗类型为Hamming窗，帧长50ms，帧位移25ms，梅尔过滤器个数为26。

SVM分类器参数与HOG+SVM方法相同，选择线性核函数，训练误差上限Nu为0.5，损失函数系数P为0.1，惩罚系数C为0.01。

3)基于多优化机制的CNN方法参数设置

输入为经过处理生成的梅尔声谱图，尺寸为129×517。深度可分离卷积层为2D(64,(5,5))，激活函数为ReLU；最大池化层为2D(2,2)；全连接层为Dense(2)。

2.5 实验结果与分析

为了保证对比实验的公平性，4种方法使用相同的样本数据进行模型训练和测试。将话音、噪音文件各取90%作为训练样本集，用于模型训练；其余10%的话音和噪音文件作为测试样本集，用于测试训练好的方法的检测效果。

对比实验以准确率、误报率、漏报率和资源占用率作为衡量方法有效性的指标。准确率是指话音样本和噪音样本总的分类正确率；误报率是指将噪音误分类为话音的比率；漏报率是指将话音误分类为噪音的比率；资源占用率是指在方法运行期间对计算机设备的CPU和内存的占用率。

4种方法的检测结果如表1所列。LHOG方法经过样本集的训练后，对测试样本集的检测准确率达到了95.12%，误报率低至0.96%，漏报率为13.14%，且CPU、内存等资源占用率在4种方法中最低；常规HOG+SVM方法的准确率、误报率、漏报率与LHOG方法表现持平，但资源使用率要高出约30%；基于多优化机制的CNN方法的准确率为89.5%，误报率最高，达到了6.38%，漏报率为16.49%，资源消耗在4种方法中最高；MFCC+SVM方法虽然误报率也不高，仅有3.38%，但漏报率最高，达到了21.85%，远高于LHOG方法，资源占用率也高于LHOG方法。通过表1的检测结果对比可以看出，LHOG方法的准确率除了比常规的HOG+SVM方法略低0.2%外，明显优于其他两种方法，资源占用率则明显低于所有其他方法，在短波通信话音检测应用上具有明显优势。

表1 4种对比方法的检测结果

Table 1 Detection results of four comparison methods

(%)

检测方法	准确率	误报率	漏报率	CPU占用率	内存占用率
LHOG	95.12	0.96	13.14	42.51	47.31
常规的HOG+SVM	95.33	0.83	12.73	72.41	62.3
基于多优化机制的CNN	89.5	6.38	16.49	78.24	68.13
MFCC+SVM	90.67	3.38	21.85	68.58	54.47

2.6 消融实验

由于LHOG中的低秩化机制包含取消block运算、裁剪归一化运算和基于PCA方法的特征降维3个操作，为了分析3个操作对LHOG特征提取方法的影响，采用消融实验逐一进行了测试。实验结果表明：取消block运算会略微降低音频检测准确率，幅度小于0.3%，但能够显著降低资源占用率，幅度大于17%；裁剪归一化运算对准确率几乎无影响，但能够小幅降低资源占用率，幅度为5%左右；基于PCA方法的特征降维对准确率几乎无影响，但也能够适当降低资源占用率，幅度为2%～3%。

结束语 针对短波通信侦控业务中由于噪音干扰导致话音检测准确率不高的问题，提出了一个LHOG短波通信话音检测方法。首先，对音频信号进行短时傅里叶变换形成语谱

图；其次，用 LHOG 特征提取方法提取语谱图中准确且高效的 LHOG 特征；最后，用 LHOG 特征作为输入训练 SVM 分类器进行分类。本文提出的 LHOG 方法有效提高了短波通信话音的检测准确率和效率。

本文研究内容从实际工程出发，以工程实际运行过程中采集的话音和噪音数据为样本进行测试，并与常规的 HOG＋SVM、基于多优化机制的 CNN、MFCC＋SVM 等传统音频分类算法进行对比分析。测试结果表明，LHOG 的准确率达到了 95.12%，误报率低至 0.96%，漏报率为 13.14%；常规的 HOG＋SVM 方法准确率为 95.33%，误报率为 0.83%，漏报率为 12.73%，但资源占用率远高于 LHOG 方法；基于多优化机制的 CNN 方法准确率为 89.5%，在 4 种方法中最低，资源占用率与常规的 HOG＋SVM 方法持平；MFCC＋SVM 方法的准确率相对较低，为 90.67%，虽然误报率不高，但漏报率最高。从实验结果可以看出，LHOG 方法从准确率、误报率、漏报率、执行效率 4 个方面综合效果更好，因此该方法具有优越性和较强的可用性。未来工作中，将尝试在更多实际工程数据集以及通用数据集进行验证推广。

参 考 文 献

[1] WANG J R，LI Y B. Design on all-digital demodulation algorithm for HF multitone parallel signal[J]. Radio Engineering，2016，46(1)：76-79.

[2] WAN L，WANG Q，LI J. End-to-End Speech Recognition with Recurrent Neural Networks for Mandarin Chinese[J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2017，25(10)：1974-1983.

[3] LI B. Speech Activity Detection Based on Deep Neural Networks Trained with Noise-Robust Features[J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2017，25(11)：2193-2203.

[4] ALDARMAKI H，ULLAH A，RAM S，et al. Unsupervised automatic speech recognition：A review[J]. Speech Communication，2022，139：76-91

[5] DONG B H，LI S Q. Current status and developing tendency for high frequency communications[J]. Information and Electronic Engineering，2007，5(1)：1-5.

[6] YIN F M，WANG S J，ZHAO L. Environmental sound classification using DeepESC convolutional neural networks[J]. Technical Acoustics. 2019，38(5)：590-593.

[7] CHEN D，HUANG Z P. Car honking recognition based on mel frequency cepstrum coefficient and support vector machine[J]. Science Technology and Engineering，2021，21(11)：4486-4491.

[8] SAILOR H B，AGRAWAL D M，PATIL H A. Unsupervised filterbank learning using convolutional restricted boltzmann machine for environmental sound classification[C]// Proceedings of Conference on the International Voice Communication Association，2017：3107-3111.

[9] CHEN H T，LIU Z Z，LIU Z M，et al. Integrating the data augmentation scheme with various classifiers for acoustic scene modeling[J]. arXiv：1907.006639，2019.

[10] CHOI Y，ATIF O，LEE J，et al. Noise-robust sound-event classification system with texture analysis[J]. Symmetry，2018，10(9)：402.

[11] QIU Y，JIA G M，YANG J F，et al. Voice recognition model of civil aviation radiotelephony communication based on BiLSTM[J]. Journal of Signal Processing，2019，35(2)：293-300.

[12] YU Q Q，LI Y，LI Y. Eco-environmental sounds classification under noise conditions[J]. Journal of Chinese Computer Systems，2011，32(8)：1689-1693.

[13] YANG L D，HU J T. Audio scene recognition of deep neural network under multiple optimization mechanisms[J]. Journal of Signal Processing，2021，37(10)：1969-1976.

[14] DALAL N，TRIGGS B. Histograms of briented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR 2005). IEEE，2005：886-893.

[15] GENG Y N，LIU S S，LIU T T，et al. Survey of pedestrian detection technology based on computer vision[J]. Journal of Computer Applications，2021，41(S1)：43-50.

[16] LE V，ZHU Y，NGUYEN A. Research on depth image gesture segmentation and HOG-SVM gesture recognition method[J]. Computer Applictions and Software，2016，33(12)：122-126.

[17] ALBIOL A，MONZO D，MARTIN A，et al. Face recognition using HOG-EB-GM[J]. Pattern Recognition Letters，2008，29(10)：1537-1543.

[18] BAO X M，REN W J，LV W T. A novet algorithm for Pedestrian recognition based on gabor wavelet and HOG feature[J]. Radio Engineering，2017，47(10)：25-29，48.

[19] ZHANG L，ZHANG Y，CHEN L L. A method of low illumination image target recognition[J]. Radio Engineering，2020，50(8)：656-660.

[20] CORTES C，VAPNIK V. Support vector networks[J]. Machine Learning，1995，20：273-297.

[21] XU X Y，YAO P. Palm vein recognition algorithm based on HOG and improved SVM[J]. Computer Engineering and Applications，2016，52(11)：175-180.

[22] SRIVASTAVA R K，PANDEY D. Speech recognition using HMM and Soft Computing[J]. Materials Today：Proceedings，2022，51：1878-1883.

BAI Jie，born in 1981，postgraduate. His main research interests include big data and artificial intelligence technology applications.