



专题：通信与 AI 融合

采用完整局部二进制模式的伪装语音检测

徐剑¹, 简志华¹, 于佳祺¹, 金易帆², 游林², 汪云路²

(1. 杭州电子科技大学通信工程学院, 浙江 杭州 310018;

2. 杭州电子科技大学网络空间安全学院, 浙江 杭州 310018)

摘要: 针对基于局部二进制模式 (LBP) 的伪装语音检测算法在检测语音转换的欺骗攻击时效果较差的情况, 提出了一种基于完整局部二进制模式 (CLBP) 的伪装语音检测方法。利用变量 Q 变换 (VQT) 生成语音信号的语谱图并应用完整局部二进制模式提取语谱图中的纹理特征向量, 再用该纹理特征向量训练真/伪语音分类器, 实现伪装语音检测。实验结果表明, 所提方法在检测语音转换的欺骗攻击时效果更好, 并且当变量 Q 变换的附加参数 γ 值为 50 时, 由完整局部二进制模式特征向量训练的基于径向基核函数的支持向量机具有最佳的伪装语音检测性能。

关键词: 说话人证实; 伪装检测; 完整局部二进制模式; 变量 Q 变换

中图分类号: TP391.42

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021108

Completed local binary pattern based speech anti-spoofing

XU Jian¹, JIAN Zhihua¹, YU Jiaqi¹, JIN Yifan², YOU Lin², WANG Yunlu²

1. School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

2. School of Cyberspace Security, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: In view of the fact that the disguised speech detection algorithm based on local binary pattern (LBP) is not effective in detecting the spoofing attack from voice conversion, an anti-spoofing method based on completed local binary pattern (CLBP) was proposed. In this method, the spectrogram of speech signals is generated by the variable Q transformation (VQT) and used to train the true/spoofed speech classifier, so as to perform the detection of disguised speech. The experimental results demonstrate that the proposed anti-spoofing method based on the CLBP in the detection of voice conversion deception is better than the LBP-based algorithm, and when the parameter γ in VQT is set to 50, the detection system based on CLBP and SVM-RBF has the best performance for anti-spoofing the disguise speech.

Key words: speaker verification, anti-spoofing, CLBP, VQT

1 引言

随着人工智能技术的深入发展, 说话人证实

得到广泛研究与应用^[1]。但说话人证实系统容易受到伪装语音攻击, 其安全性非常值得关注。伪装语音检测是通过对说话人的声音进行分析, 进

收稿日期: 2021-03-15; 修回日期: 2021-05-01

基金项目: 国家自然科学基金资助项目 (No.61201301, No.61772166)

Foundation Items: The National Natural Science Foundation of China (No. 61201301, No.61772166)



而识别它是真实说话人的语音还是人为恶意伪装的语音^[2]。伪装语音通常由设备回放、语音转换及语音合成等技术生成^[3-5]，通过蓄意的操作能够伪装成特定的说话人声音，从而达到欺骗说话人证实系统的目的。伪装语音识别系统可针对恶意的伪装语音实现伪装检测，提高说话人证实系统安全性能，具有广阔的应用前景。

伪装语音识别通常需要对目标语音信号进行特征提取，再与其对应的真实语音特征对比分析进而判定真/伪。传统的伪装语音识别系统提取的特征参数主要分为两种：一种是语音信号的幅度谱特征，通常有高阶梅尔倒谱系数、梅尔主频率和对数幅度谱等^[6-8]；另一种是语音信号的相位谱特征，通常有修正的群时延和相对相移等^[9-11]。但这些特征在应对未知类型伪装语音的攻击时，检测效果往往不是很理想。

近些年，纹理分析逐渐成为计算机视觉和模式识别领域的研究热点，并在图像的分类处理中得到较大的发展。局部二进制模式（local binary pattern, LBP）是一种流行的纹理分析方法，最初在面部图像识别中有广泛的应用^[12-13]。最近，Alegre 等^[14]提出了一种基于 LBP 的反伪装说话人证实方法，利用了 LBP 对级联排列的声学特征向量进行纹理分析，从而区别真/伪语音。这种基于 LBP 的伪装语音检测策略相较于传统的伪装语音检测策略所需要的先验知识（如伪装语音类型、生成的算法等信息）较少，对于未知伪装语音的检测效果较好，有利于实际的应用。不过，这种方法在面对由语音转换生成的伪装欺骗时检测效果不如传统的伪装语音检测方法，而且它的步骤较为复杂，需预先提取语音信号的声学特征向量并级联处理后才可获取纹理特征向量。

本文提出了一种利用完整局部二进制模式（completed local binary pattern, CLBP）^[15]提取语音信号的纹理特征向量的方法，它可以有效地实现伪装语音检测。该方法的好处在于不用提前获

取、级联语音信号的声学特征向量，而是直接对语音信号的语谱图进行分析获得语音信号的纹理特征向量。CLBP 通过计算语音信号语谱图的完整局部二进制模式的符号差值（CLBP-sign, CLBP_S）、完整局部二进制模式的幅度差值（CLBP-magnitude, CLBP_M）以及完整局部二进制模式的中心灰度（CLBP-center, CLBP_C）而得，所获取的图像纹理特征更加全面，有助于提高伪装语音检测的准确率。

2 基于 LBP 的伪装语音检测

2.1 LBP 算法

LBP 算法是一种对图像局部纹理和空间结构的测量方法，它的中心思想是比较图像像素点间的局部差异，即图像某一像素点跟其周围像素点进行灰度值大小的对比生成一串二进制数，计算式为：

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (1)$$

其中， g_p 表示相邻像素点的灰度值， g_c 表示中心像素点灰度值， P 表示相邻像素点的个数， R 是相邻像素点与中心像素点的半径距离，而 $s(x)$ 表示符号函数，有：

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

若相邻像素点的灰度值 g_p 大于其中心像素点灰度值 g_c ，二进制位设置为 1，否则设置为 0。

2.2 LBP 伪装语音检测方法

基于 LBP 的伪装语音检测方法是应用 LBP 对语音信号的声学特征向量进行纹理分析，从而获取反映语音声学特性的纹理特征向量并用以真/伪语音的检测，具体过程如图 1 所示。包含的步骤有：首先，对一段语音信号进行声学特征提取并形成特征矢量序列，即矩阵。之后用 LBP 对整个矩阵进行纹理分析，同时映射至 $\text{LBP}_{P,R}$ ，由此得到一幅纹理特征模式图；最后将纹理特征模式

图中每一行的 $LBP_{p,R}$ 值用统计直方图表示并垂直级联得到纹理特征向量, 用该纹理特征向量训练分类器, 从而达到伪装语音检测的目的。

这种伪装检测策略有以下方面的缺陷: 第一, 在检测语音转换的欺骗攻击时, 伪装检测效果有欠缺; 第二, 步骤较为复杂, 需预先提取传统的语音信号的声学特征向量, 而且检测的效果依赖该声学特征, 面对部分通过改变语音声学特征生成的伪装欺骗时, 伪装语音检测效果较差。应用 LBP 提取语音信号的纹理特征如图 1 所示。

3 基于 CLBP 的伪装语音检测

3.1 CLBP 算法

CLBP 算法^[15]是一种改进的 LBP 算法, 由 CLBP_S、CLBP_M 和 CLBP_C 3 部分组成, CLBP 的组成框架如图 2 所示。一段语音的声学特征参数矢量序列构成原始图像, 其纹理特征分为两部分: 局部差异和中心灰度。通过局部符号和幅度差异转换分析图像的局部符号差异和局部幅度差

异分别得到 CLBP_S 和 CLBP_M, 再计算中心灰度与整幅图像的灰度平均值的差值得 CLBP_C, 最后将三者组合构成一幅 CLBP 图并统计生成 CLBP 直方图。

3.1.1 CLBP_S 特征

CLBP_S 特征表示图像局部的符号差异, 计算方法与 LBP 相同, 首先读取图像每个像素点灰度值, 设置 3×3 的评估窗口进行二进制编码, 使其相邻像素点与中心像素点进行灰度值比较, 若相邻像素点的灰度值 g_p 大于中心像素点灰度值 g_c , 二进制位设置为 1, 否则设置为 0, 即:

$$CLBP_S = \sum_{p=0}^7 s(g_p - g_c) 2^p \quad (3)$$

该评估窗口可以生成 8 位二进制数, 同时转换为十进制值, 每一个十进制值代表一种纹理模式, 从 00000000(0) 到 11111111(255) 一共有 256 种纹理模式。用该评估窗口分析整幅图像, 可以得到一幅包含 256 种纹理模式的特征图像, 通过直方图统计并生成统计直方图, 将每一种纹理模式

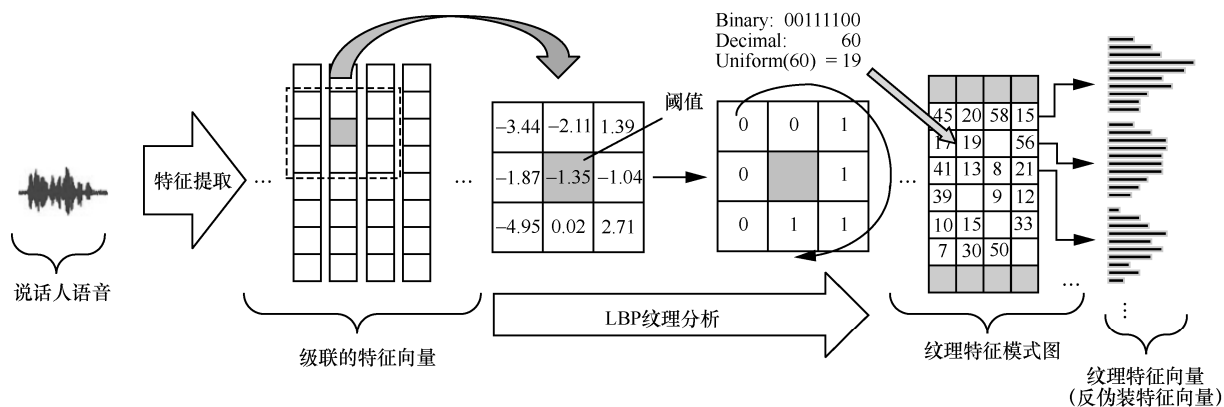


图1 应用 LBP 提取语音信号的纹理特征

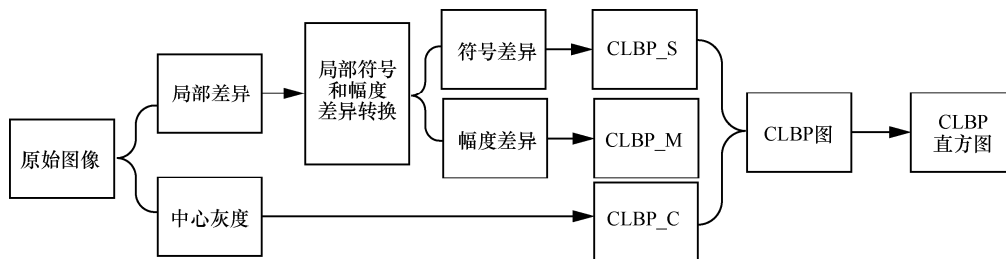


图2 CLBP 的组成框架



作为特征向量的一个维度，再将该纹理模式的数量作为特征向量在该维度下的值，所以一段语音可以得到 1×256 维的 CLBP_S 特征向量。

3.1.2 CLBP_M 特征

CLBP_M 特征表示的是图像局部的幅度差异。先对相邻像素点灰度值与中心像素点灰度值的差值取绝对值，记为相邻像素点与中心像素点的幅度差值 m_p ；再取整幅图像所有幅度差值的平均，记为幅度差值的阈值 c ，即：

$$m_p = |g_p - g_c| \quad (4)$$

$$c = \frac{1}{N} \sum_{k=1}^N m_p^{(k)} \quad (5)$$

其中， g_p 为相邻像素点的灰度值， g_c 为中心像素点灰度值； k 为幅度差值 m_p 的序号， N 为 m_p 的总数。将每个相邻像素点与中心像素点的幅度差值的绝对值 m_p 作为新的相邻像素点灰度值，幅度差值的阈值 c 作为新中心像素点灰度值。设置 3×3 的评估窗口进行二进制编码，使其新相邻像素点与新中心像素点进行灰度值比较，若新相邻像素点的灰度值大于其新中心像素点灰度值，二进制位设置为 1，否则设置为 0。该评估窗口同样可以生成 8 位二进制数并转换为十进制值，每一个十进制值代表一种纹理模式，共有 256 种纹理模式。这样一段语音的特征矢量序列构成的图就可以得到包含 256 种纹理模式的纹理特征，再将所有的纹理特征采用直方图统计就可以得到 1×256 维的完整局部二进制模式的幅度差值 CLBP_M 特征向量。CLBP_S 特征和 CLBP_M 特征提取过程示意图如图 3 所示。

3.1.3 CLBP_C 特征

CLBP_C 特征表示整幅图像的中心灰度水平，通过计算整幅图像中所有中心像素点灰度值的平均 c_l ，记为新的阈值，即：

$$c_l = \frac{1}{M} \sum_{k=1}^M g_c^{(k)} \quad (6)$$

其中， $g_c^{(k)}$ 表示第 k 个中心像素点 g_c 的灰度值， M

9	12	34
10	25	28
99	64	56

(a) 3×3 样本区域

-16	-13	9
-15		3
74	39	31

(b) 局部差异

0	0	1
0		1
1	1	1

(c) 符号差异

16	13	9
15		3
74	39	31

(d) 幅度差异

图 3 CLBP_S 特征和 CLBP_M 特征提取过程示意图

为 g_c 的总数。将每一局部区域的中心像素点的灰度值 g_c 与 c_l 比较，若中心像素点的灰度值大于 c_l 记为 1，否则记为 0。这样通过二进制编码就可以得到 CLBP_C 特征的值。

在分别计算 CLBP_S 特征、CLBP_M 特征以及 CLBP_C 特征之后，将这些特征进行串联，得到用作伪装语音检测的 CLBP 特征向量。按照上述步骤，从所有真实语音和伪装语音的特征参数矢量构成的图中求得 CLBP 特征向量，用作支持向量机的训练。

3.2 训练支持向量机

本文采用支持向量机^[16]作为真/伪语音的分类器。给训练集设置标签，记为 (x_i, y_i) , $i = 1, 2, 3, \dots, N$, $x_i \in R_M$, $y_i \in [-1, 1]$ 。其中， N 是训练样本的总数， M 是样本空间的维数， y_i 是样本的分类类别， $y_i = 1$ 表示伪装语音， $y_i = -1$ 表示真实语音。选用径向基核函数 (radial basic function, RBF)，引入参数 γ 和错误惩罚因子 C 同时进行优化，训练最优支持向量确定一个最优超平面。使目标函数最大化，即：

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (7)$$

受限条件为：

$$\sum_{i=1}^N \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \quad (8)$$

采用的 RBF 核函数为:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

其中, α_i 为每个样本对应的拉格朗日乘子, 错误惩罚因子 C 在确定 RBF 核函数后, 控制错误分类样本的惩罚程度。解中只有一部分 α_i 不为零, 所对应的样本就是训练所得的支持向量。

分类器支持向量机示意图如图 4 所示, 支持向量机共有 3 层, 从下至上分别为 SVM 输入层、SVM 隐藏层以及输出层。通过上述步骤, 可得分类函数:

$$y = f(x) = \text{sgn}\left[\sum_{i=1}^l \alpha_i y_i K(x_i, x_j) + b\right] \quad (10)$$

其中, b 为分类阈值, 可由支持向量求得。应用此分类决策函数可对待识别的语音信号的纹理特征分类, 达到识别真/伪语音的目的。

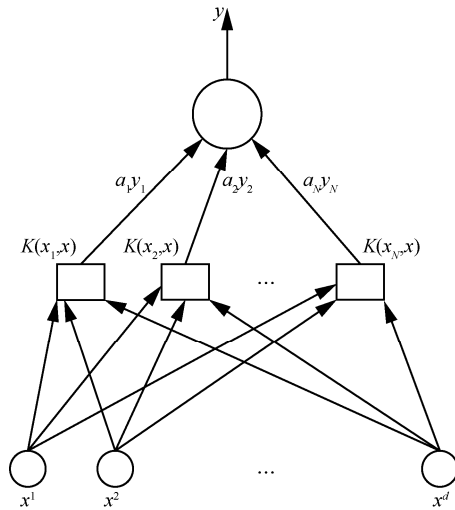


图 4 分类器支持向量机示意图

3.3 CLBP 伪装语音检测

本文提出的基于 CLBP 的伪装语音检测算法, 通过 CLBP 分析语音信号的特征参数矢量序列图提取用以伪装检测的语音纹理特征向量。在提取语音信号的语谱图时, 首先对说话人语音信号进行变量 Q 变换 (VQT) 得到清晰纹理的语谱图,

表示为:

$$X^{\text{VQT}}(k) = \frac{1}{N_k} \sum_{n=0}^{N_k-1} x(n) w_{N_k}(n) e^{-j \frac{2\pi Q}{N_k} n} \quad (11)$$

其中, w_{N_k} 是长度为 N_k 的窗函数, Q 是 VQT 中的变量因子, k 表示 VQT 谱的频率序号, N_k 值和 k 值有关。变量 Q 因子通过引入一个附加参数 γ 使 Q 因子向低频平滑减小, 以提高低频处的分辨率, 即:

$$B_k = Q f_k + \gamma \quad (12)$$

将得到的语谱图转换为灰度图, 并提取 CLBP 特征, 将其用作 SVM 训练及分类的语音纹理特征向量。采用上述方法获取真实语音库及伪装语音库中所有语音信号的 CLBP 纹理特征向量用作训练集, 将训练得到的 SVM 作为识别伪装语音的分类器。在识别时, 同样需要提取待识别的说话人语音纹理特征向量, 进而采用训练得到的 SVM 区分真/伪语音。

4 实验设计及结果分析

4.1 数据集

实验选用 ASVspoof 2015 语音库^[17]评估本文所提算法的性能, 该语音库包含了真实语音和伪装语音, 真实语音是在无噪环境下由 106 位说话人录制, 其中包含 45 名男性和 61 名女性, 而伪装语音是采用多种不同算法修改真实语音而得。语音信号的采样频率为 16 kHz, 以 16 位比特量化, 音频文件格式为 RIFF/WAVE。将语音库中的数据分为训练集、开发集和评估集 3 个子集, 各个子集之间没有重复, 具体情况见表 1。

表 1 训练集、开发集和评估集的说话人和语音数量

子集	说话人		语音	
	男生	女生	真实	伪装
训练集	10	15	3 750	12 625
开发集	15	20	3 497	49 875
评估集	20	26	9 404	184 000



实验语音库中的伪装语音由语音转换和语音合成两种伪装方式生成见表 2，共分为 10 个子集，分别表示为 S1~S10，具体情况介绍可以参考文献[18]。

表 2 语音库中各子集语音的数量及伪装方式

子集	语音			声码器	伪装方式
	训练集	开发集	评估集		
真实	3 750	3 497	9 404	无	无
S1	2 525	9 975	18 400	STRAIGHT	语音转换
S2	2 525	9 975	18 400	STRAIGHT	语音转换
S3	2 525	9 975	18 400	STRAIGHT	语音合成
S4	2 525	9 975	18 400	STRAIGHT	语音合成
S5	2 525	9 975	18 400	MLSA	语音转换
S6	0	0	18 400	STRAIGHT	语音转换
S7	0	0	18 400	STRAIGHT	语音转换
S8	0	0	18 400	STRAIGHT	语音转换
S9	0	0	18 400	STRAIGHT	语音转换
S10	0	0	18 400	STRAIGHT	语音合成

4.2 伪装语音检测系统性能评估方法

实验采用等错误率 (equal error rate, EER) 的指标来评价伪装语音检测系统的性能。在伪装语音检测系统中，为判别语音样本是否是真实的说话人声音，会为该语音样本计算打分。分数越高，是真实语音的概率越大；反之越小，更可能是伪装语音。检测系统预先设定阈值 θ ，若得分高于阈值 θ 判定为真实语音，反之判定为伪装语音。由此，可以得到两种常见的指标。

(1) 错误拒绝率 (false rejection rate, FRR)

真/伪语音分类问题中，若真实语音样本的得分小于阈值 θ 被系统误认为伪装语音，则为错误拒绝。错误拒绝数量在所有真实语音案例的比例即错误拒绝率 $P_{FR}(\theta)$ ，计算式为：

$$P_{FR}(\theta) = \frac{\{\text{真实语音得分} \leq \theta \text{ 的数量}\}}{\{\text{所有真实语音数量}\}} \quad (13)$$

(2) 错误接受率 (false acceptance rate, FAR)

若伪装语音样本的得分大于阈值 θ 被系统误认为真实语音，则为错误接受。错误接受数量在所有伪装语音数量中的比例即错误接受率 $P_{FA}(\theta)$ ，计算式为：

$$P_{FA}(\theta) = \frac{\{\text{伪装语音得分} \geq \theta \text{ 的数量}\}}{\{\text{所有伪装语音数量}\}} \quad (14)$$

其中， $P_{FR}(\theta)$ 会随着阈值 θ 的增大单调增大，而 $P_{FA}(\theta)$ 会随着阈值 θ 的增大单调减小。当阈值 $\theta = \theta_{EER}$ 时，这两种错误检测率相等，便是等错误概率 EER，计算式为：

$$EER = P_{FR}(\theta_{EER}) = P_{FA}(\theta_{EER}) \quad (15)$$

错误拒绝率和错误接受率相等时，EER 大小可以用来评估伪装语音检测系统性能，EER 越小，伪装语音检测效果越好，反之伪装语音检测效果越差。

4.3 伪装语音检测系统性能测试

实验采用 LBP 算法与本文所提 CLBP 算法进行性能对比，采用的特征参数如下。

(1) 分别提取线性倒谱系数 (linear frequency cepstral coefficient, LFCC) 和梅尔倒谱系数 (mel-frequency cepstral coefficient, MFCC) 这两种声学特征参数。应用 SPro⁵ 工具提取训练集中语音信号的 38 维 LFCC 特征 (12 LFCC、12 Δ LFCC、12 $\Delta\Delta$ LFCC、 Δ Energy、 $\Delta\Delta$ Energy) 以及 38 维 MFCC 特征 (12 MFCC、12 Δ MFCC、12 $\Delta\Delta$ MFCC、 Δ Energy、 $\Delta\Delta$ Energy)，并用 LBP 和 CLBP 的纹理分析方法提取上述两种声学特征中各自的 LBP 特征向量和 CLBP 特征向量。

(2) 利用 VQT 将训练集中每一个语音转换为语谱图，并用 LBP 和 CLBP 的纹理分析方法分别提取每一张语谱图的 LBP 特征向量和 CLBP 特征向量。

将上述各类参数在开发集和评估集中进行测试，比较各种情况下的伪装语音检测性能。在进

行 VQT 时,为了选择合适的附加参数 γ 的值,实验在 3 个不同的分类器中进行了性能测试。这 3 个分类器包括 GMM-UBM、具有因子分析 (factor analysis, FA) 信道补偿的 GMM (GMM-FA) 以及 SVM-RBF 系统,实验结果见表 3。从表 3 的实验结果来看,在各类分类器,当 $\gamma=50$ 时,检测系统的等错误率最小,性能最好。另外,不论附加参数 γ 取何值,本文采用的方案 CLBP_S/M/C+SVM-RBF 具有最佳的性能。

采用不同特征参数在开发集数据中进行性能对比,实验结果见表 4。由表 4 可知,当提取特征的对象相同时,采用 CLBP_S/M/C 的系统检测效果普遍要好于采用 LBP 的系统;当提取特征的方法相同时,对语谱图提取纹理特征的检测性能优于 LPCC 和 MFCC。因此,从实验结果来看,采用 CLBP_S/M/C 方法对语谱图提取纹理特征向量

具有最好的检测效果。同时,实验也发现,不管采用哪个方法,在检测 S2 这种伪装语音类型时,EER 都明显增大,因为 S2 是通过改变声学特征而生成的伪装语音,单纯依靠声学特征参数进行检测会比较困难。

从 MFCC 特征可以看出,通过它提取的纹理特征向量训练的系统检测 S2 欺骗的效果最差,这是由于 S2 为通过调整第一阶梅尔倒谱系数生成的语音转换欺骗。在检测 S3、S4 这两种语音合成伪装欺骗时,得到的等错误率普遍小于检测 S1、S2 和 S5 这 3 种由语音转换伪装欺骗求得的等错误率,可见基于纹理特征的检测方法对语音合成的欺骗检测效果较好。在检测 S1、S2 和 S5 这 3 种语音转换伪装欺骗时,采用 CLBP_S/M/C 方法对语谱图提取纹理特征向量进行伪装检测相比于传统基于 LBP 的方法具有更小的 EER 值,可见

表 3 评估集中各个伪装语音检测系统的等错误率比较

检测系统	$\gamma=0$			$\gamma=25$			$\gamma=50$		
	已知伪装欺骗	未知伪装欺骗	平均值	已知伪装欺骗	未知伪装欺骗	平均值	已知伪装欺骗	未知伪装欺骗	平均值
LBP+GMM	5.381%	9.332%	7.357%	0.894%	4.350%	2.622%	0.477%	2.306%	1.392%
LBP+GMM-FA	5.053%	7.905%	6.479%	0.669%	3.873%	2.271%	0.451%	2.152%	1.302%
LBP+SVM-RBF	3.566%	9.830%	6.698%	0.482%	4.866%	2.674%	0.360%	2.270%	1.315%
CLBP_S/M/C+GMM	4.549%	8.025%	6.287%	0.369%	3.441%	1.905%	0.090%	1.502%	0.796%
CLBP_S/M/C+GMM-FA	3.934%	7.326%	5.630%	0.215%	2.508%	1.362%	0.065%	1.389%	0.727%
CLBP_S/M/C+SVM-RBF	3.206%	7.800%	5.503%	0.097%	2.630%	1.364%	0.029%	1.394%	0.712%

表 4 开发集不同特征参数情况下等错误率对比

提取纹理特征的对象	提取纹理特征的方法	EER					
		S1	S2	S3	S4	S5	平均值
LFCC	LBP	1.318%	3.257%	0.860%	1.216%	2.442%	1.819%
	CLBP_S/M/C	1.059%	3.163%	0.507%	0.753%	1.962%	1.489%
MFCC	LBP	1.422%	5.869%	0.882%	1.304%	2.096%	2.315%
	CLBP_S/M/C	1.139%	5.291%	0.694%	0.873%	1.640%	1.927%
语谱图	LBP	0.690%	0.982%	0.150%	0.289%	0.747%	0.572%
	CLBP_S/M/C	0.248%	0.331%	0.000%	0.000%	0.318%	0.179%



表 5 评估集中几种伪装语音检测方法的 EER 对比

提取纹理特征的对象	提取纹理特征的方法	已知的伪装欺骗						未知的伪装欺骗						所有伪装欺骗
		S1	S2	S3	S4	S5	平均值	S6	S7	S8	S9	S10	平均值	平均值
LFCC	LBP	1.140%	2.951%	0.545%	0.940%	2.060%	1.527%	2.553%	1.364%	1.615%	1.253%	11.467%	3.650%	2.589%
	CLBP_S/M/C	0.896%	2.463%	0.315%	0.507%	1.431%	1.122%	1.758%	0.959%	0.914%	0.819%	10.339%	2.958%	2.040%
MFCC	LBP	1.029%	5.385%	0.529%	0.899%	1.968%	1.962%	4.857%	1.194%	1.231%	1.520%	17.720%	5.304%	3.633%
	CLBP_S/M/C	0.779%	4.890%	0.403%	0.492%	1.404%	1.594%	4.504%	0.785%	0.857%	0.900%	15.879%	4.585%	3.090%
语谱图	LBP	0.661%	0.796%	0.084%	0.145%	0.698%	0.477%	0.803%	0.752%	0.670%	0.596%	8.709%	2.306%	1.392%
	CLBP_S/M/C	0.125%	0.174%	0	0	0.153%	0.090%	0.190%	0.145%	0.103%	0.080%	6.991%	1.502%	0.796%

本文所提出的基于 CLBP 对语谱图提取纹理特征向量的伪装语音检测方法比传统的基于 LBP 的方法有更好的性能，提升了对语音转换伪装欺骗的检测效果。

表 5 是系统在评估集的性能测试结果。从表 5 可知，由语谱图提取纹理特征向量在 S1~S9 这 9 种伪装欺骗中进行伪装检测的等错误率较小，检测效果较好。同时，不管哪种情况，在检测未知欺骗 S10 时，EER 值均明显增大。这说明与其他攻击相比，使用 MaryTTS 的 S10 欺骗语音是最难检测的攻击类型，原因是 S10 欺骗语音不使用声码器，而是使用了单元选择合成算法选择语音段并级联生成欺骗语音，保留了较多的纹理特征。在检测 S2 和 S6 伪装欺骗时，由 LPCC 和 MFCC 这些声学参数提取纹理特征向量的性能表现较差，因为这依赖于语音信号提取的声学特征，导致在检测由语音信号的声学特征生成的语音转换欺骗时系统的伪装检测性能明显降低。在检测未知的语音转换欺骗攻击（S6~S9）时，传统基于 LBP 对声学特征提取纹理特征向量的伪装语音检测性能较差，而采用 CLBP_S/M/C 方法对语谱图提取纹理特征向量的检测性能较好。整体来看，基于 CLBP_S/M/C 纹理提取方法相比基于 LBP 的纹理提取方法在检测效果上有了较大的提升。

5 结束语

本文提出了一种基于 CLBP 的伪装语音检测

算法，它有效地改善了传统的基于 LBP 的伪装语音检测方法的性能，特别是在面对由语音转换生成的伪装语音时效果更加明显。实验结果表明，当 VQT 的附加参数 $\gamma=50$ 时，由 CLBP 特征参数和 SVM-RBF 分类器构建的伪装语音检测系统具有最佳的性能。

参考文献：

- [1] 吴震东, 潘树诚, 章坚武. 基于 CNN 的连续语音说话人声纹识别[J]. 电信科学, 2017, 33(3): 59-66.
WU Z D, PAN S C, ZHANG J W. Continuous speech speaker recognition based on CNN[J]. Telecommunications Science, 2017, 33(3): 59-66.
- [2] WU Z Z, DE LEON P L, DEMIROGLU C, et al. Anti-spoofing for text-independent speaker verification: an initial database, comparison of countermeasures, and human performance[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2016, 24(4): 768-783.
- [3] PHAPATANABURI K, WANG L B, NAKAGAWA S, et al. Replay attack detection using linear prediction analysis-based relative phase features[J]. IEEE Access, 2019(7): 183614-183625.
- [4] 林朗, 王让定, 严迪群, 等. 基于逆梅尔对数频谱系数的回放语音检测算法[J]. 电信科学, 2018, 34(5): 90-98.
LIN L, WANG R D, YAN D Q, et al. A playback speech detection algorithm based on log inverse Mel-frequency spectral coefficient[J]. Telecommunications Science, 2018, 34(5): 90-98.
- [5] KHODABAKHSH A, MOHAMMADI A, DEMIROGLU C. Spoofing voice verification systems with statistical speech synthesis using limited adaptation data[J]. Computer Speech & Language, 2017(42): 20-37.
- [6] WU Z Z, EVANS N, KINNUNEN T, et al. Spoofing and countermeasures for speaker verification: a survey[J]. Speech Communication, 2015(66): 130-153.

- [7] PAUL D, PAL M, SAHA G. Spectral features for synthetic speech detection[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(4): 605-617.
- [8] MUCKENHIRN H, KORSHUNOV P, MAGIMAI-DOSS M, et al. Long-term spectral statistics for voice presentation attack detection[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2017, 25(11): 2098-2111.
- [9] MOWLAEE P, SAEIDI R, STYLIANOU Y. Advances in phase-aware signal processing in speech communication[J]. Speech Communication, 2016(81): 1-29.
- [10] WANG L B, NAKAGAWA S, ZHANG Z F, et al. Spoofing speech detection using modified relative phase information[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(4): 660-670.
- [11] HANILCI C. Speaker verification anti-spoofing using linear prediction residual phase features[C]//Proceedings of 2017 25th European Signal Processing Conference (EUSIPCO). [S.l.: s.n.], 2017.
- [12] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
- [13] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: application to face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 2037-2041.
- [14] ALEGRE F, VIPPERLA R, AMEHRAEY A, et al. A new speaker verification spoofing countermeasure based on Local Binary Patterns [C]// Proceedings of 14th Annual Conference of the International Speech Communication Association (INTERSPEECH). Piscataway: IEEE Press, 2013.
- [15] GUO Z H, ZHANG L, ZHANG D. A completed modeling of local binary pattern operator for texture classification[J]. IEEE Transactions on Image Processing, 2010, 19(6): 1657-1663.
- [16] CHANG C C, LIN C J. Libsvm[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27.
- [17] WU Z Z, YAMAGISHI J, KINNUNEN T, et al. ASVspoof: the

automatic speaker verification spoofing and countermeasures challenge[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(4): 588-604.

- [18] 徐剑. 基于 CLBP 和 HOG 的伪装语音检测方法[D]. 杭州: 杭州电子科技大学, 2020.

XU J. A method for detection of spoofing speech based on CLBP and HOG [D]. Hangzhou: Hangzhou Dianzi University, 2020.

[作者简介]



徐剑 (1995-), 男, 杭州电子科技大学通信工程学院硕士生, 主要研究方向为语音处理和语音伪装检测。

简志华 (1978-), 男, 博士, 杭州电子科技大学通信工程学院副教授、硕士生导师, 主要研究方向为语音转换、伪装语音检测、声纹识别等。

于佳祺 (1997-), 男, 杭州电子科技大学通信工程学院硕士生, 主要研究方向为语音伪装检测、特征提取与分析。

金易帆 (1996-), 男, 杭州电子科技大学网络空间安全学院硕士生, 主要研究方向为语音伪装检测、声纹识别。

游林 (1966-), 男, 博士, 杭州电子科技大学网络空间安全学院教授、硕士生导师, 主要研究方向为生物信息处理、信息安全、密码学等。

汪云路 (1980-), 女, 博士, 杭州电子科技大学网络空间安全学院讲师, 主要研究方向为音频信息处理、信息隐藏。