研究与开发

基于中心对称局部二值模式的合成伪装语音检测方法

徐嘉[1]，简志华[1]，金宏辉[1]，吴超[1]，游林[2]，吴迎笑[3]
（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;
3. 杭州电子科技大学计算机学院，浙江 杭州 310018）

摘　要：针对基于局部二值模式的伪装语音检测方法的合成语音检测准确度较低的情况，提出了一种基于中心对称局部二值模式的伪装语音检测方法。该方法通过短时傅里叶变换得到语音信号的语谱图，再利用中心对称局部二值模式提取语谱图的纹理特征，并用该纹理特征训练随机森林分类器，从而实现真伪语音的判别。该方法综合考虑语谱图中像素点的数值大小和位置关系，包含了更加全面的纹理信息，并将特征维度降低至16维，有利于减少计算量。实验结果表明，在ASVspoof 2019数据集上，与传统的基于局部二值模式的伪装语音检测方法相比，所提方法将合成伪装语音的串联检测代价函数（t-DCF）降低了16.98%，检测速度提高了89.73%。

关键词：说话人验证；伪装语音检测；中心对称局部二值模式；随机森林
中图分类号：TP391.42
文献标志码：A
doi: 10.11959/j.issn.1000–0801.2023005

Synthetic spoofing speech detection method based on center-symmetric local binary pattern

XU Jia[1], JIAN Zhihua[1], JIN Honghui[1], WU Chao[1], YOU Lin[2], WU Yingxiao[3]
1. School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China
2. School of Cyberspace Security, Hangzhou Dianzi University, Hangzhou 310018, China
3. School of Computer, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: In view of the fact that the local binary pattern (LBP) based speech spoofing detection method has low detection accuracy when detecting synthetic speech, a spoofing speech detection method based on center-symmetric local binary pattern (CSLBP) was proposed. In this method, the spectrogram of the speech signal was obtained through short-time Fourier transform (STFT), and then the texture feature was extracted from the spectrogram using the CSLBP. The random forest classifier was trained by the extracted texture feature to realize the discrimination of genuine and spoofing speech. The CSLBP-based method comprehensively considered the value and position relationship of pixels in the spectrogram so as to contain more texture information, and reduced the feature dimension to 16

收稿日期：2022–05–16；**修回日期：**2022–12–15
通信作者：简志华，jianzh@hdu.edu.cn
基金项目：国家自然科学基金资助项目（No.61201301，No.61772166，No.61901154）
Foundation Items: The National Natural Science Foundation of China (No.61201301, No.61772166, No.61901154)

beneficial to decrease the amount of computation. Experimental results on the ASVspoof 2019 dataset show that, compared with the LBP-based spoofing detection method, the proposed method reduced the tandem detection cost function (t-DCF) of synthetic spoofing speech by 16.98% and increased the detection speed by 89.73%.

Key words: speaker verification, spoofing speech detection, CSLBP, random forest

0　引言

语音是人机交互的一种重要方式，语音信号中包含了说话人特有的身份信息。随着科技的进步，语音合成技术不断发展，生成的语音信号足以欺骗人类听觉系统和计算机[1]。通过语音合成技术生成的高质量语音，若被应用在银行的身份认证上，将对个人的财产安全造成重大影响；若利用合成语音操控智能家居等设备，将造成隐私的泄露；若利用合成语音进行电信诈骗，将对社会治安产生不良影响。使用伪装语音检测技术对语音信号进行分析，从而实现真伪语音的判别，对于提高声纹识别系统的安全性具有重要意义，应用前景广阔[2]。

伪装语音检测技术通过提取语音信号的特征参数并应用分类模型实现真伪语音的判别。常用的特征参数包括以下两大类。一类是短时谱特征，比如梅尔频率倒谱系数（Mel-frequency cepstral coefficient，MFCC）、线性预测倒谱系数（linear prediction cepstral coefficient，LPCC）等。文献[3]比较了 MFCC 等特征的检测性能，在 ASVspoof 2019 数据库的 LA 数据集上，利用 MFCC 进行伪装语音检测的等错误率（equal error rate，EER）为 9.33%。崔兆国[4]比较了 MFCC、LPCC 特征以及 36 阶 MFCC 及其差分倒谱系数的伪装检测效果，实验结果表明，这 3 种特征均可用于伪装语音检测且 36 阶 MFCC 及其差分倒谱系数具有更优的效果。另一类是短时相位特征，比如修正群延迟（modified group delay，MGD）、相对相移（relative phase shift，RPS）等。文献[5]发现群时延特征保留了较多的共振峰结构，证明了其对于语音处理的鲁棒性。文献[6]比较了 MGD 和 RPS 两种特征的检测性能，在使用 MGD 进行伪装语音检测时，EER 为 4.918 7%，利用 RPS 做特征时，EER 为 4.473 0%。上述声学特征都可以实现真伪语音的判别，但对合成语音提取这些特征时，仅保留了原始语音中的幅度或相位信息，保留信息不全面，影响了检测性能。

近年来，纹理分析成为图像处理领域的研究热点，局部二值模式（local binary pattern，LBP）是一种较为常用的纹理描述方法。文献[7-8]利用 LBP 特征实现了颜色纹理分类，文献[9-10]通过提取面部图像的 LBP 特征，实现了人脸识别。在语音伪装检测领域，Alegre 等[11]提出了一种利用 LBP 描述符实现伪装语音检测的方法，该方法对语音分帧后提取特征向量，将每一帧的特征向量级联后运用 LBP 描述符进行纹理信息提取，实现真伪语音的区分，在 NIST 数据集上将 EER 降低至 0.5%。文献[12]提出了对语谱图直接进行纹理分析的方法，在 ASVspoof 2015 数据集上将 EER 从 2.589%降至 0.796%。然而 LBP 特征维数为 256 维，维数较高，检测效率有待提高，且 LBP 特征只利用了像素点之间的大小关系，包含的纹理信息较为单一，检测准确率有待提升。本文提出了一种利用中心对称局部二值模式的合成语音伪装检测方法，通过短时傅里叶变换（short-time Fourier transform，STFT）得到语音信号的语谱图，再利用中心对称局部二值模式（central-symmetric local binary pattern，CSLBP）对语谱图进行纹理分析得到纹理特征图，将特征图映射至统计直方图得到 16 维特征向量，用该向量训练随机森林实现合成语音的伪装检测。该方法直接对语谱图提取 CSLBP 特征，综合考虑了语音信号的幅度和相位信息，并利用语谱图像素点之间的大小和位置关

系，降低了特征维数，提高了系统的检测性能。

1 纹理特征

1.1 LBP 纹理特征

LBP 是一种描述图像局部纹理信息的纹理描述方法，该方法将图像划分为若干个 3×3 的邻域，将每个邻域中心像素点的灰度值作为阈值，将周围 8 个像素点的灰度值与阈值进行比较，若周围像素点的灰度值大于或等于阈值，则该像素点被标记为 1，否则为 0，LBP 纹理示意图如图 1 所示[13]。

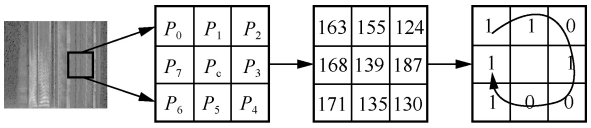


图 1 LBP 纹理示意图

经比较，每个邻域可按顺时针得到一个 8 位二进制数，该 8 位二进制数可转换成十进制，这个十进制数值代表了该邻域中心像素点的 LBP 特征值，且每个十进制数表示一种 LBP 纹理模式，因此 LBP 特征值共有 256 种不同的模式，计算式如下[14]：

$$\mathrm{LBP}(x,y)=\sum_{i=0}^{7}s(p_i-p_c)\times 2^i \tag{1}$$

其中，(x,y) 为中心像素点的坐标，p_i 为该邻域内第 i 个周围像素点的灰度值，p_c 为该邻域内中心像素点的灰度值，$s(x)$ 为符号函数，计算式如下：

$$s(x)=\begin{cases}1, & x\geqslant 0\\ 0, & \text{其他}\end{cases} \tag{2}$$

1.2 CSLBP 纹理特征

由于 LBP 纹理特征维数较高，在计算时具有较高的时间复杂度，学术界提出了 CSLBP 纹理特征。CSLBP 依旧将图像划分为若干个 3×3 邻域，但不是将周围像素点与中心像素点进行比较，而是比较关于中心像素点对称的几对像素点的灰度值[15]。CSLBP 纹理示意图如图 2 所示，将图像划分为若干个 3×3 的邻域，每个邻域的中心像素点为 p_c，p_0 和 p_4、p_1 和 p_5、p_2 和 p_6、p_3 和 p_7 为关于 p_c 对称的 4 对像素点，分别比较这 4 对像素点的大小关系。在 CSLBP 中，引入了一个预先设定的全局阈值 T，描述了纹理区域的平坦性。若 p_0 与 p_4 的差值大于或等于全局阈值 T，则给 p_0 点赋值 1，否则为 0，p_1、p_2、p_3 同理。经过比较后，每个邻域可以得到一个 4 位二进制数，将该二进制数转换为十进制即可得到该点的 CSLBP 特征值。

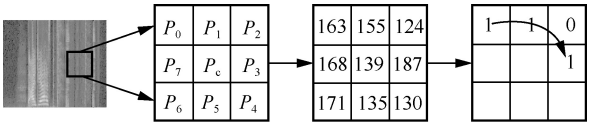


图 2 CSLBP 纹理示意图

CSLBP 的计算式如下：

$$\mathrm{CSLBP}(x,y)=\sum_{i=0}^{3}s(p_i-p_{i+4})\times 2^i \tag{3}$$

$$s(x)=\begin{cases}1, & x\geqslant T\\ 0, & \text{其他}\end{cases} \tag{4}$$

用 CSLBP 进行纹理描述时，每个像素点仅由 4 位二进制数表示，因此 CSLBP 共有 16 种不同的模式，大大降低了纹理特征的维数，且包含了梯度方向上的信息[16]。

2 利用 CSLBP 的伪装语音检测

基于 CSLBP 的伪装语音检测方法，首先将语音信号转换为语谱图，再提取语谱图的 CSLBP 纹理特征，然后将该特征输入随机森林网络进行训练和分类，从而实现伪装语音检测。CSLBP 特征提取流程如图 3 所示。

首先通过 STFT 得到语音信号的语谱图，然后将语谱图转换为灰度图，再将灰度图划分为若干个 3×3 邻域，对每一邻域内的像素点提取十进制的 CSLBP 值，即可得到整幅图像的 CSLBP 矩阵，最后对整幅图像的 CSLBP 值进行直方图统计，统计每种模式下像素点的数目，最终得到 16 维的 CSLBP 特征向量。

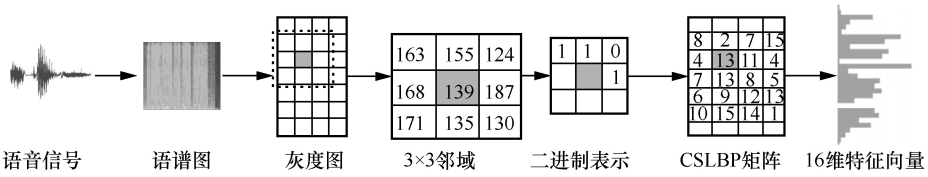


图 3　CSLBP 特征提取流程

本文提出的利用 CSLBP 的伪装语音检测整体流程如图 4 所示，首先用图 3 的方法获取训练集中语音信号的 CSLBP 特征向量，并输入随机森林网络进行训练得到分类器，检测时提取待测语音的 CSLBP 特征向量，然后利用随机森林分类器实现真伪语音的判别。

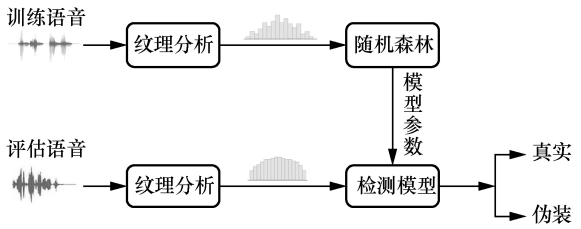


图 4　利用 CSLBP 的伪装语音检测整体流程

3　实验与结果

3.1　数据集

实验是在 ASVspoof 2019 语音数据库中的 LA 数据集上进行的，其中包含真实、合成和转换 3 种语音，本实验仅采用其中的真实语音和合成语音进行训练和测试。LA 数据集分为训练集、开发集和评估集，3 个子集之间无重复语音，LA 数据集见表 1。该数据集中的语音由神经波形模型、声码器、波形拼接等 17 种算法生成[17]。

表 1　LA 数据集

子集	语音数目/个		
	真实	伪装	合成
训练集	2 580	22 800	15 200
开发集	2 548	22 296	14 864
评估集	7 355	63 882	49 140

3.2　性能评价方法

本实验采用串联检测代价函数（tandem detection cost function，t-DCF）作为语音伪装检测性能的评价方法[18]。在实际应用中，伪装语音检测系统往往需要与自动说话人验证（automatic speaker verification，ASV）系统结合使用，且检测结果同时受到伪装语音检测系统和 ASV 系统的影响，若仅使用 EER 作为评价标准，无法反映检测模型的整体性能。t-DCF 综合考虑了错误拒绝和错误接受发生的代价、错误率以及真伪语音的先验概率，用来评估伪装语音检测系统较为合理，且 t-DCF 值越小，系统检测效果越好。

t-DCF 的计算式为：

$$\text{t-DCF} = C_{\text{FRR}}\pi_{\text{tar}}P_{\text{FRR}}^{\text{CM}} + C_{\text{FAR}}\left(1-\pi_{\text{tar}}\right)P_{\text{FAR}}^{\text{CM}} \tag{5}$$

其中，C_{FRR} 和 C_{FAR} 分别表示错误拒绝一个真实语音的代价和错误接受一个伪装语音的代价，实际应用中这两个参数应根据具体场景来确定。若对安全性的需求更高，则将 C_{FAR} 设置得更高；若用户体验更加重要，则将 C_{FRR} 设置得更高。由于本实验中伪装语音检测系统既需要对伪装语音拒绝访问，也需要对真实语音允许访问，因此在实验中将 C_{FRR} 和 C_{FAR} 均设置为 1。$P_{\text{FRR}}^{\text{CM}}$ 和 $P_{\text{FAR}}^{\text{CM}}$ 分别表示错误拒绝率和错误接受率，计算式如下：

$$P_{\text{FRR}}^{\text{CM}} = \frac{N_{\text{FR}}}{N_{\text{bonafide}}} \tag{6}$$

$$P_{\text{FAR}}^{\text{CM}} = \frac{N_{\text{FA}}}{N_{\text{spoof}}} \tag{7}$$

其中，N_{FR} 表示真实语音中被错误拒绝的个数，N_{bonafide} 表示真实语音的总数量，N_{FA} 表示伪装语音中被错误接受的个数，N_{spoof} 表示伪装语音的总数量。π_{tar} 表示目标说话人出现的先验概率，在本

实验中即真实语音随机出现的先验概率，计算式如下：

$$\pi_{\text{tar}} = \frac{N_{\text{bonafide}}}{N_{\text{bonafide}} + N_{\text{spoof}}} \tag{8}$$

3.3 性能测试

在计算 LBP 特征时，若中心像素点的灰度值为 30，相邻像素点 p_1 和 p_2 的灰度值分别为 29 和 30，则 p_1 像素点会被赋值为 0，p_2 像素点被赋值为 1，而 p_1 和 p_2 像素点灰度值的差异很小，可以忽略不计，因此 LBP 对这种可以忽略的变化不具有鲁棒性。为了改善这种情况，在 CSLBP 的计算中，引入全局阈值 T，该值可以降低像素点之间的灰度差异。但是若该值过大，会导致所有像素点均被赋值为 0，进而导致 CSLBP 特征失去其有效性。因此，为保证 CSLBP 的鉴别能力，T 值应为一个较小的值。文献[16]发现将该值设置为 0 时，纹理特征的鲁棒性最强。在文献[19]中，该值设置为 3。在本实验中，为了找到最优的阈值，比较了 T 值取为 0～10 时，利用 CSLBP 特征做特征向量，支持向量机（support vector machine，SVM）和随机森林作为后端分类器进行伪装语音检测的 t-DCF 值计算，在这里用训练集中的真实语音和 A01～A04 4 种类型的合成语音进行训练，用评估集中的真实语音和 A07～A16 类型的合成语音一起进行评估。不同阈值 T 下 CSLBP 特征的 t-DCF 见表 2。实验结果表明，无论是 SVM 还是随机森林，均在 T 值为 5 时取得了最佳检测结果，因此后续实验均在 T 值为 5 的基础上开展。

表 2 不同阈值 T 下 CSLBP 特征的 t-DCF

T	0	1	2	3	4	5	6	7	8	9	10
SVM	0.218 4	0.202 9	0.164 8	0.134 4	0.131 4	0.123 1	0.124 7	0.128 5	0.142 3	0.154 6	0.167 2
随机森林	0.189 2	0.190 6	0.179 0	0.144 0	0.116 4	0.101 7	0.102 1	0.106 3	0.116 3	0.130 6	0.147 9

为了验证 CSLBP 特征的有效性，比较了 36 维 MFCC 特征、LPCC 特征、对语谱图提取的 LBP 特征和对语谱图提取的 CSLBP 特征在 SVM 和随机森林做后端分类器时的 t-DCF，在这里用训练集中的真实语音和 A01～A04 4 种类型的合成语音进行训练，用评估集中的真实语音和 A07～A16 类型的合成语音一起进行评估。4 种特征的 t-DCF 见表 3。同时，对 4 种不同特征的检测时间进行了比较，4 种特征的检测耗时见表 4。

表 3 几种特征的 t-DCF

模型	MFCC	LPCC	LBP	CSLBP
SVM	0.275 4	0.646 8	0.156 3	0.123 1
随机森林	0.391 8	0.541 5	0.122 5	0.101 7

表 4 几种特征的检测耗时

模型	检测耗时/s			
	MFCC	LPCC	LBP	CSLBP
SVM	16	31	247	17
随机森林	65	42	516	53

从表 3 和表 4 来看，利用 CSLBP 特征训练随机森林的方法在合成语音检测时取得了最佳检测结果，与 LBP 相比，CSLBP 特征的 t-DCF 降低了 16.98%，且检测速度提高了 89.73%，这是因为 CSLBP 特征不仅利用了像素点之间的大小关系，还利用了像素点之间的空间位置关系，纹理信息更为全面，极大地提高了合成语音的检测性能。评估集中不同类型合成语音检测的 t-DCF 见表 5。

表 5 对评估集中不同类型的合成语音检测性能进行了比较，在这个比较实验中，训练所用样本是训练集中的真实语音和 A01～A04 4 种类型的合成语音，然后对评估集中 A07～A16 10 种类型的合成语音分别进行伪装检测。在检测 A07、A08、A14 类型的合成语音时，CSLBP 的 t-DCF 值比 LBP 略高，这是因为对于这几种类型的合成语音来说，其语谱图中邻域中心像素点的灰度值

表 5　评估集中不同类型合成语音检测的 t-DCF

特征	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	平均
MFCC	0.363 5	0.232 1	0.444 0	0.413 0	0.518 8	0.586 5	0.318 4	0.354 0	0.332 3	0.398 4	0.396 1
LPCC	0.481 9	0.748 4	0.604 2	0.454 9	0.529 9	0.597 6	0.184 3	0.758 0	0.703 6	0.384 7	0.544 8
LBP	0.020 8	0.014 2	0.067 3	0.133 8	0.238 7	0.100 0	0.050 5	0.104 6	0.396 4	0.093 6	0.122 0
CSLBP	0.050 9	0.032 3	0.055 7	0.086 1	0.070 8	0.026 1	0.050 4	0.238 0	0.336 5	0.062 2	0.100 9

包含了丰富的纹理信息，而 CSLBP 特征没有将中心像素点的灰度值利用起来，因此效果相对较差。另外，从表 5 可以看出，这几种检测方法在检测 A14、A15 这两种类型的合成语音时，相比于其他几种类型，t-DCF 值较高。原因在于 A14、A15 这两种类型的语音在生成时不仅利用了语音合成算法，还利用了语音转换技术，生成的语音更加贴近真实语音，保留了较多的纹理信息。虽然 A13 的语音也采用了类似于 A14、A15 的生成方法，但检测效果优于 A14、A15，这是因为 A14、A15 采用基于长短期记忆（long short-term memory，LSTM）网络的声学模型生成语音[20]。而生成 A13 语音时，采用传统的语音转换方法以及基于矩匹配的损失函数进行训练[21]，用直接波形修改的方式生成输出波形，纹理信息也随之改变，因此易于检测。整体来看，CSLBP 特征降低了合成语音的 t-DCF，减少了检测所需时间，提高了合成语音检测系统的整体性能。

4　结束语

本文提出了一种利用 CSLBP 特征的合成语音检测方法。该方法首先通过 STFT 得到语音信号的语谱图，然后利用语谱图中像素灰度值和空间位置的差异，提取 CSLBP 特征进行纹理分析，再进行直方图统计得到特征向量，然后利用随机森林对特征向量进行训练和分类，实现伪装语音的检测。该方法不仅比较了像素点之间的灰度值，还利用了像素点之间的位置关系，提取了更多的纹理信息，降低了特征维度，有效地改善了传统 LBP 特征的检测性能，提高了检测速度，降低了 t-DCF。实验结果表明，在全局阈值为 5 时，由 CSLBP 特征和随机森林构建的合成语音检测系统在 ASVspoof 2019 数据集上取得了最佳的检测性能。利用 CSLBP 的合成语音检测方法虽较传统方法有所改进，但仍存在问题需要解决，如全局阈值自适应化等，在未来的工作中，将继续进行优化。

参考文献：

[1] KANERVISTO A, HAUTAMÄKI V, KINNUNEN T, et al. Optimizing tandem speaker verification and anti-spoofing systems[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 477-488.

[2] LEI Z C, YAN H, LIU C H, et al. Two-path GMM-ResNet and GMM-SENet for ASV spoofing detection[C]//Proceedings of ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2022: 6377-6381.

[3] ALZANTOT M, WANG Z Q, SRIVASTAVA M B. Deep residual neural networks for audio spoofing detection[C]//Proceedings of Interspeech 2019. Cary: ISCA, 2019: 1078-1082.

[4] 崔兆国. 基于 SVM 的反蓄意模仿说话人识别研究[D]. 桂林: 桂林电子科技大学, 2013.
CUI Z G. Research on speaker recognition of anti-deliberate imitation based on SVM[D]. Guilin: Guilin University of Electronic Technology, 2013.

[5] PADMANABHAN R, PARTHASARATHI S H K, MURTHY H A. Robustness of phase based features for speaker recognition[C]//Proceedings of Interspeech 2009. Cary: ISCA, 2009: 2299-2302.

[6] SARATXAGA I, SANCHEZ J, WU Z, et al. Synthetic speech detection using phase information[J]. Speech Communication, 2016(81): 30-41.

[7] HOANG V T. Unsupervised LBP histogram selection for color texture classification via sparse representation[C]//Proceedings of 2018 IEEE International Conference on Information Communication and Signal Processing. Piscataway: IEEE Press, 2018: 79-84.

[8] SHU X, SONG Z, SHI J, et al. Multiple channels local binary pattern for color texture representation and classification[J]. Signal Processing: Image Communication, 2021(98): 116392.

[9] KARANWAL S. A comparative study of 14 state of art descriptors for face recognition[J]. Multimedia Tools and Applications, 2021, 80(8): 12195-12234.

[10] SHI L, WANG X, SHEN Y. Research on 3D face recognition method based on LBP and SVM[J]. Optik: International Journal for Light and Electron Optics, 2020(220):165157.

[11] ALEGRE F, VIPPERLA R, AMEHRAYE A, et al. A new speaker verification spoofing countermeasure based on local binary patterns[C]//Proceedings of Interspeech 2013. Cary: ISCA, 2013: 940-944.

[12] 徐剑, 简志华, 于佳祺, 等. 采用完整局部二进制模式的伪装语音检测[J]. 电信科学, 2021, 37(5): 91-99.
XU J, JIAN Z H, YU J Q, et al. Completed local binary pattern based speech anti-spoofing[J]. Telecommunications Science, 2021, 37(5): 91-99.

[13] XIA Z H, YUAN C S, LYU R, et al. A novel weber local binary descriptor for fingerprint liveness detection[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 50(4): 1526-1536.

[14] TOFFA O K, MIGNOTTE M. Environmental sound classification using local binary pattern and audio features collaboration[J]. IEEE Transactions on Multimedia, 2021(23): 3978-3985.

[15] SHAH A, EL-ALFY E. Comparative analysis of feature extraction and fusion for blind authentication of digital images using chroma channels[J]. Signal Processing: Image Communication, 2021(95): 116271.

[16] 王科俊, 曹逸, 邢向磊. 基于 MB-CSLBP 的手指静脉加密算法研究[J]. 智能系统学报, 2018, 13(4): 543-549.
WANG K J, CAO Y, XING X L. Finger-vein encryption algorithm based on MB-CSLBP[J]. CAAI Transactions on Intelligent Systems, 2018, 13(4): 543-549.

[17] WANG X, YAMAGISHI J, TODISCO M, et al. ASVspoof 2019: a large-scale public database of synthesized, converted and replayed speech[J]. Computer Speech & Language, 2020(64): 101114.

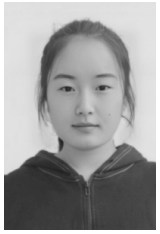
[18] KINNUNEN T , DELGADO H , EVANS N , et al. Tandem assessment of spoofing countermeasures and automatic speaker verification: fundamentals[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020(28): 2195-2210.

[19] HEIKKILA M, PIETIKAINEN M. A texture-based method for modeling the background and detecting moving objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 657-662.

[20] LIU L J, LING Z H, JIANG Y, et al. WaveNet vocoder with limited training data for voice conversion[C]//Proceedings of Annual Conference of the International Speech Communication Association (Interspeech). Cary: ISCA, 2018: 1983-1987.

[21] LI Y J, SWERSKY K, ZEMEL R. Generative moment matching networks[C]//Proceedings of International Conference on Machine Learning (ICML). [S.l.: s.n.], 2015: 1718-1727.

[作者简介]



徐嘉（1998– ），女，杭州电子科技大学通信工程学院硕士生，主要研究方向为伪装语音检测。

简志华（1978– ），男，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。

金宏辉（1999– ），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转换和伪装语音检测。

吴超（1988– ），男，杭州电子科技大学通信工程学院讲师，主要研究方向为导航信号处理及欺骗干扰检测。

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

吴迎笑（1980– ），女，杭州电子科技大学计算机学院特聘教授，主要研究方向为毫米波感知用于声纹识别与认证、射频信息处理和工业互联网。