基于双哈希索引的高效语音生物哈希安全检索算法

黄羿博[1], 陈德怀[1], 张秋余[2]

[1]西北师范大学 物理与电子工程学院 兰州 中国 730070
[2]兰州理工大学 计算机与通信学院 兰州 中国 730050

摘要 针对语音数据在信道传输与云端存储时的安全性问题, 以及由于语音数据数目大、维数高、空间复杂度高带来的检索效率问题, 提出了一种基于双哈希索引的高效语音生物哈希安全检索算法。首先, 在服务端分别提取语音信号的频谱通量与峭度因子特征并将两种特征融合, 利用 Bagging 分类对语音信号的差分哈希分类, 并基于分类结果构建密钥分配索引表; 然后, 根据密钥分配索引表建立具有单一映射密钥的生物特征模板, 并将其量化构造生物哈希, 得到哈希索引; 同时, 采用混合域置乱加密算法对原始语音加密, 构建密文语音库; 最后, 将哈希索引与密文语音库上传至云端并构建云端生物哈希索引表。在移动端, 采用归一化汉明距离进行匹配检索。实验结果表明: 本文算法的匹配阈值区间为(0.2694, 0.4173), 说明该检索算法能够灵活选取匹配阈值, 具有较好的鲁棒性和区分性; 检索过程中单条语音平均检索时间仅为 9.4957×10^{-4}s, 并且经过 15 种内容保持操作后的查全率与查准率均为 100%, 说明该算法具有较好的检索性能, 可以满足各种环境下的语音检索需求; 同时提出的加密算法密钥空间大小为 10^{60}, 说明能够抵御穷举密钥攻击、保证语音数据的安全; 此外, 构建的生物特征模板具有良好的多样性、安全性和可撤销性。

关键词 安全语音检索; 双哈希索引; 生物特征模板; 生物哈希; 密文语音
中图法分类号 TP391.3; TN912.3 DOI 号 10.19363/J.cnki.cn10-1380/tn.2024.03.06

Efficient Speech Biological Hashing Secure Retrieval Algorithm Based on Double Hash Index

HUANG Yibo[1], CHEN Dehuai[1], ZHANG Qiuyu[2]

[1] College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China
[2] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

Abstract Aiming at the security of speech data in channel transmission and cloud storage, as well as the problems of retrieval efficiency caused by the large number, high dimension and high spatial complexity of speech data, an efficient speech biological hashing secure retrieval algorithm based on double hash index is proposed. Firstly, the spectral flux and kurtosis factor features of speech signal are extracted in the server terminal, and then the two features are fused, Bagging classification is used to classify speech signals by differential hashing, and the key distribution index table is constructed based on the classification results; then, according to the key distribution index table, the biometric template with a single mapping key is established, and its biometric hash is quantized to obtain the hash index; at the same time, the mixed domain scrambling encryption is used to encrypt the original speech and construct the encrypted speech database; finally, the hash index and encrypted speech database are uploaded to the cloud and the biological hash index table is constructed. In the mobile terminal, using normalized hamming distance for matching retrieval. The experimental results show that the matching threshold interval obtained by the algorithm is (0.2694, 0.4173), which shows that the retrieval system can flexibly select the matching threshold and has good robustness and discrimination; the average retrieval time of a single speech in the retrieval process is only 9.4957×10^{-4}s, and the recall and precision after 15 kinds of content preservation operations are 100%, it shows that the algorithm has good retrieval performance and can meet the needs of speech retrieval in various environments; at the same time, the size of the encryption algorithm key space is 10^{60}, which shows that it can resist exhaustive key attack and ensure the security of speech data; in addition, the constructed biometric templates have good diversity, security and revocability.

Key words secure speech retrieval; double hash index; biometric template; biological hashing; encrypted speech

通讯作者: 黄羿博, 博士, 副教授, Email: huang_yibo@nwnu.edu.cn。
本课题得到甘肃省科技计划项目资助, 甘肃省自然科学基金(No. 21JR7RA120)和国家自然科学基金(No. 61862041)资助。
收稿日期: 2022-04-05; 修改日期: 2022-12-21; 定稿日期: 2023-11-02

1 引言

1.1 研究背景

云计算的发展推动了语音等多媒体数据的制造、传播和存储。要从海量多媒体数据中准确、高效地查询到精确信息, 仅仅依靠简要文本标签难以实现。基于内容的语音检索技术研究利用音频的物理特征、听觉特征以及语义特征实现音频信息检索[1], 成为现阶段音频检索领域的一项重要研究课题。此外, 云存储技术作为云计算领域的一项重要技术, 已经成为用户将数据外包存储的关键方式, 极大的缓解了日益严重的“数据丰富”问题。近年来, 由于云环境半开放的特点, 数据安全问题频发, 用户隐私信息遭受泄露。因此, 当云服务器中存储敏感数据时, 如何在保护用户隐私的前提下实现高效、安全的语音检索, 成为目前音频检索领域亟须解决的关键问题之一。

针对语音内容检索的需求, 基于哈希算法[2]的语音检索技术应运而生。哈希算法具有良好鲁棒性和区分性的同时耗费较低的计算和存储空间, 因而被广泛用于语音内容检索以实现近似最近邻 (Approximate nearest neighbor, ANN) 搜索[3]。用户将原始语音数据量化构造为特定长度的二进值哈希码, 并通过计算哈希码之间的汉明距离获得对应语音数据, 在实现精确检索的同时避免了原始数据直接暴露于云环境而带来的安全隐患。但是, 在语音数据海量化、高维化的发展趋势下, 传统基于哈希的近邻检索算法的检索效率降低, 难以达到在有效时间内寻找到目标的要求。

索引表[4]的出现既能保证数据的灵活访问与共享又可以减少计算开销, 实现快速查询。索引表是根据关键码值而直接访问内存存储位置的数据结构, 可以有效提升查询速度。由于哈希算法与索引表的种种优势与特征, 基于索引的语音哈希检索技术[5]被相应提出, 将二进制哈希码作为索引, 构建出具有安全性、高效性的哈希索引表。在实际生活中, 不同语音数据的特征分布存在差异, 可以通过机器学习算法来学习这些语音数据的特征分布, 构建一种“可学习”索引结构, 以达到优化索引结构的目的。因此, 双哈希索引的高效语音检索方案的研究具有重大的实际意义和应用价值。

1.2 研究现状

语音检索的研究由来已久, 早期的语音检索主要通过语音识别技术[6]实现。但是识别系统的性能受多方面的影响, 检索结果差强人意。20 世纪 70 年代, Bridle 首先提出关键词检出技术[7], 并定义为“给定词识别”。随后, Christiansen 等人[8]提出“关键词”的概念, 并沿用至今。其提出的关键词检出技术是通过信号的线性预测编码技术(Linear Predictive Coding, LPC)实现对连续语音中关键词的检索和定位, 对 10 个数字以内的识别具有较高准确率。在 1985 年, Higgins 等人[9]提出了填充模型(Filler)的概念, 用来对关键词以外的发音现象建模, 并利用模型连接的方法构建了关键词检索系统。在 2008 年, Felipe 等人[10]将 R-Tree 与叠加的文本签名相结合, 引入了一种称为 IR2-Tree(信息检索 R-Tree)的索引结构, 并使用该索引结构实现 top-k 空间关键字查询。在 2015 年, Singh 等人[11]提出了一种基于投影和多尺度哈希(ProMiSH)的关键词搜索方案, 使用随机投影和基于多尺度哈希的索引结构, 显示出优越的可扩展性, 实验结果表明, 该方案的检索效率远远高于基于树的检索系统。

虽然关键词检出技术已经发展得非常成熟, 但是容易遭受字典和统计攻击, 无法保证搜索隐私性。因此, 基于内容的语音检索技术逐渐成为研究重点。其思想是对语音内容特征进行分析和提取, 然后将这些内容特征作为索引并采用一种近似匹配的思想进行检索, 根据匹配距离与阈值的大小关系获得检索结果。其中, 结合感知哈希[12]是基于内容的语音检索的一项重要技术手段, 研究人员已经在此方面取得了众多研究成果。如: He 等人[13]提出了一种基于音节级感知哈希的加密语音检索算法, 该算法利用声学模型提取语音特征并构造感知哈希, 对各种信号环境下的语音数据都具有良好的查全率、查准率和感知鲁棒性。Zhang 等人[14]通过提取语音信号的短期互相关特征作为特征摘要构造感知哈希, 并通过度量哈希码之间的汉明距离实现语音检索。

尽管感知哈希技术因其单向转换的特点而具有良好的鲁棒性与单向性, 但是没有更好的权衡安全性与可撤销性。因此, 基于生物特征模板[15-16]保护的语音检索方案引起了广泛关注。生物哈希[17]法又叫加盐法, 是一种有效的生物特征模板保护方法, 其在提取生物特征的基础上, 使用特定密钥定义了一个正交随机变换函数, 并利用该函数对生物特征进行变换, 进一步量化获得生物哈希序列。Teoh 等人[18]提出了一种基于随机多空间量化(RMQ)的生物哈希, 该方法首先将原始特征转换为低维特征向量, 然后映射到指定的随机子空间序列上, 最后将特征向量重新映射量化后得到 RMQ 生物哈希序列。Wang 等人[19]提出了一种基于能零比和改进 LP-MMSE 参数

融合的多格式语音生物哈希方案, 构建了具有可撤销性的生物特征模板。Huang 等人[20]提出了一种基于特征融合的生物哈希语音检索算法, 首先通过 FFT 与 IFFT 提取语音特征, 并使用改进的 Marotto 混沌矩阵生成密钥, 最后将该密钥与语音特征迭代内积获得生物哈希序列。该算法构建的生物特征模板具有良好的安全性、多样性以及可撤销性。黄羿博等人[21]提出了一种基于混沌测量矩阵的生物哈希密文语音检索算法, 该算法将混沌测量矩阵用于生物哈希构造, 提高了生物特征模板的安全性、多样性和隐私性, 缺点在于索引方式不够完善, 处理高维数据时效率不佳。

在传统基于内容的语音检索方案中, 建立有效的索引结构可以极大提高检索系统的效率, 实现快速查询。现有的索引结构大多依据语音数据的特征分布差异, 主要包括: 树型索引[22]、哈希表索引[23]、语义视觉索引[24]和基于音节 Lattice 索引[25]等。Zhang 等人[26]构建了一种基于 B+Tree 的安全聚类索引结构, 实现了全文检索, 有效提高了检索效率。Cao 等人[27]提出了深度柯西哈希(Deep Cauchy Hashing, DCH), 主要思想是设计一个基于柯西分布的成对交叉熵损失, 通过对汉明距离大于给定阈值的相似特征对进行惩罚, 生成紧凑且集中的二进制哈希码, 从而实现高效的汉明空间检索。Agrawal 等人[28]提出了一种新的索引结构 HashFile, 该算法选择随机投影作为哈希函数生成哈希码, 然后利用线性扫描递归地划分密集的桶并组织成树形结构。给定查询点 q, 检索算法以自顶向下的方式查询该点附近的桶, 将每个节点中的候选存储桶按照哈希值的升序顺序存储, 最后加载到内存中进行线性扫描获得查询结果。由于随机投影有助于过滤掉远处的数据点, 而线性扫描可以有效地处理剩余的候选数据点, 因此检索性能得到了提高。

此外, 针对语音信号在云存储环境下的内容泄露问题, 语音加密技术引起了国内外学者广泛关注, 并将其与语音检索算法结合以实现安全语音检索, 解决语音数据传输和存储安全问题。常见的语音加密方法有: 时频域置乱加密[29-30]、混沌加密[31-32]和音频隐写术[33-34]等。Khaleel 等人[35]提出了一种语音加密算法, 并依赖于量子混沌映射和 k-means 聚类设计了两个置乱阶段对语音加密, 使语音信号能够抵御不同类型的攻击, 提高了算法的安全性。Abdullah 等人[36]提出了一种基于量子计算机制的音频隐写术新方法, 该算法根据加密量子音频信号的状态修改宿主量子音频信号, 选定最低有效量子位的状态, 确保了宿主量子音频与其隐写版本之间的高度不可感知性, 解决了原始音频的安全问题。

1.3 本文贡献

本文针对现有检索方案中存在的索引结构复杂、生物特征模板可撤销性差以及明文数据泄露问题, 以基于内容保护的语音检索为背景, 提出了一种基于双哈希索引的高效语音生物哈希安全检索方案, 具体贡献如下:

(1) 针对生物特征模板可撤销性差的问题, 利用 Bagging 分类对差分哈希分类, 根据分类结果构建密钥分配索引表并建立具有单一映射密钥的生物特征模板。当受到攻击时, 可以查询索引表快速更新密钥, 生成新的生物特征模板, 实现了生物特征模板的安全性与可撤销性。

(2) 针对云环境下的用户隐私安全问题, 提出混合域置乱加密算法对原始语音数据加密。实验结果表明, 该加密方案的密钥空间大小为 10^{60}, 可以抵御穷举密钥攻击, 防止明文数据泄露。

(3) 针对现有索引结构复杂的问题, 利用分级检索的思想, 设计了一种双哈希索引的索引结构。用户先通过一级密钥分配索引表查询所属类别, 获得对应密钥后量化构造生物哈希码。检索时只需与二级云端生物哈希索引表中同类的哈希索引进行匹配, 实现了类内检索, 从而构建高效的语音哈希检索算法。

2 提出的安全语音检索算法

图 1 是本文提出的安全语音检索算法模型, 主要包括密文语音库构建、生物特征模板构建、云端生物哈希索引表构建、以及移动端语音检索四个部分。

在密文语音库构建方面, 采用混合域置乱加密算法对原始语音数据加密, 并将加密语音上传至云端构建密文语音库。

在生物特征模板构建方面, 首先, 分别提取原始语音的频谱通量特征和峭度因子特征, 进行特征融合得到原始语音的融合特征向量; 然后, 对差分哈希分类并基于分类结果构建密钥分配索引表; 最后, 查询密钥分配索引表建立具有单一映射密钥的生物特征模板, 生成原始语音的生物哈希序列并上传至云端。

在云端生物哈希索引表构建方面, 根据上传的密文语音库及生物哈希索引在云端构建生物哈希索引表。

在移动端语音检索方面, 用户输入待检索语音, 利用上述方法生成待检索语音的生物哈希码并发送至云端。检索时分别将该条生物哈希码与云端生物

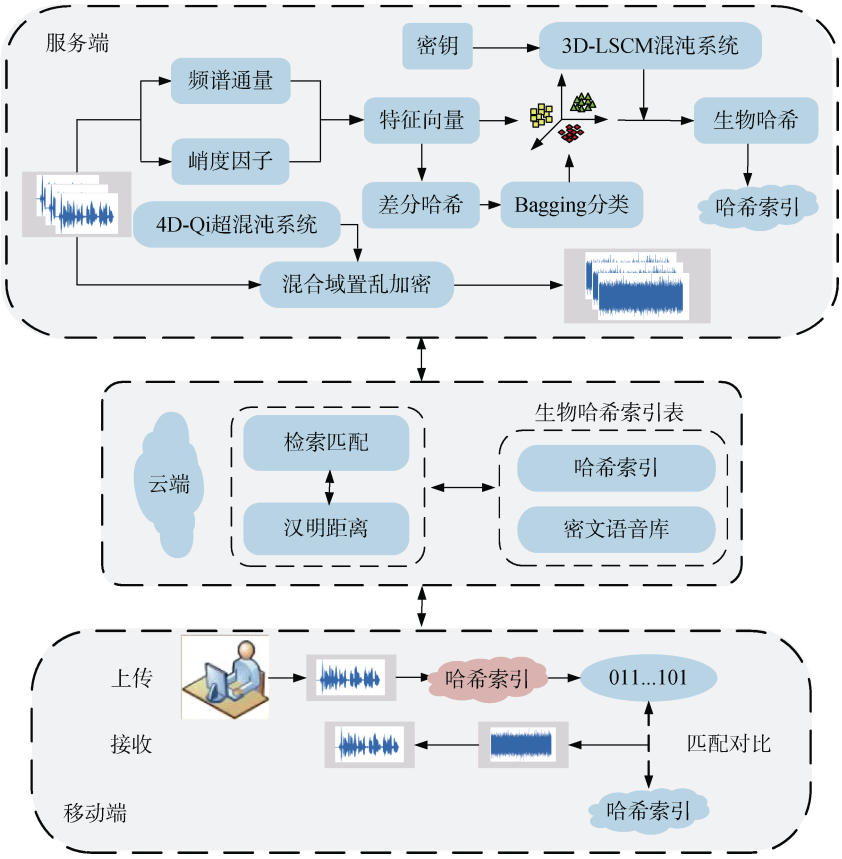

图 1　语音检索算法模型

Figure 1　Speech retrieval algorithm model

哈希索引表中对应类属的哈希索引进行检索匹配, 并将密文语音库中符合阈值条件的密文语音发送给用户, 用户在移动端可以根据对应密钥对加密语音进行解密。

2.1　密文语音库的构建

传统的置乱加密广泛应用于语音安全领域并取得了良好的加密效果, 但它仍然存在一定的局限性, 比如使用单一的一维伪随机序列作为密钥不能抵御已知明文攻击, 且可能存在较多无效密钥。针对上述问题, 本文采用了混合域置乱加密算法, 并利用 4D-Qi 超混沌系统[37]生成四个一维混沌序列作为该加密算法在不同域内置乱加密的初始密钥。

其中 4D-Qi 超混沌系统的动力学方程如式(1)所示:

$$\begin{cases} x_{i+1} = \alpha(y_i - x_i) + y_i z_i w_i \\ y_{i+1} = \beta(x_i + y_i) - x_i z_i w_i \\ z_{i+1} = -\delta z_i + \phi x_i y_i w_i \\ w_{i+1} = -\eta w_i + x_i y_i z_i \end{cases} \tag{1}$$

式中, $[x_0, y_0, z_0, w_0]$为系统初始值, α, β, δ, ϕ, η 为控制参量, 且当 α=50, β=4, δ=13, ϕ=4, η=20 时, 系统有两个正李雅普诺夫(Lyapunov)指数, 呈现超混沌状态。该超混沌系统相图如图 2 所示。

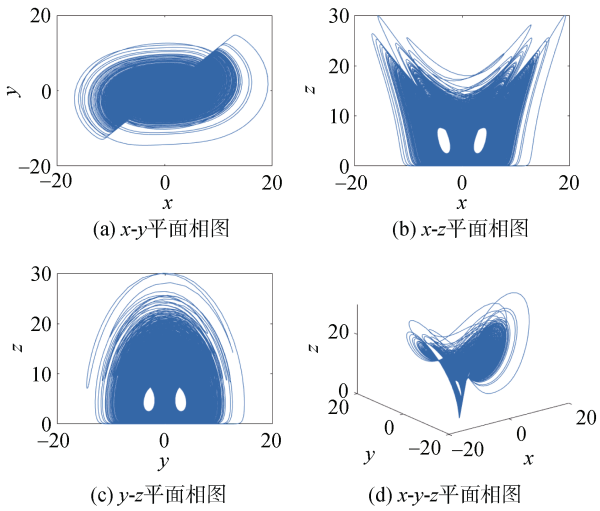

图 2　4D-Qi 超混沌系统相图

Figure 2　The phase portraits of 4D-Qi hyperchaotic system

由图 2 可以看出, 该超混沌系统的动力学行为复杂, 具有显著的混沌特性。因此将该混沌系统应用于语音加密, 能够显著提高加密系统的安全性, 扩大密钥空间。

采用的混合域置乱加密算法模型如图 3 所示。

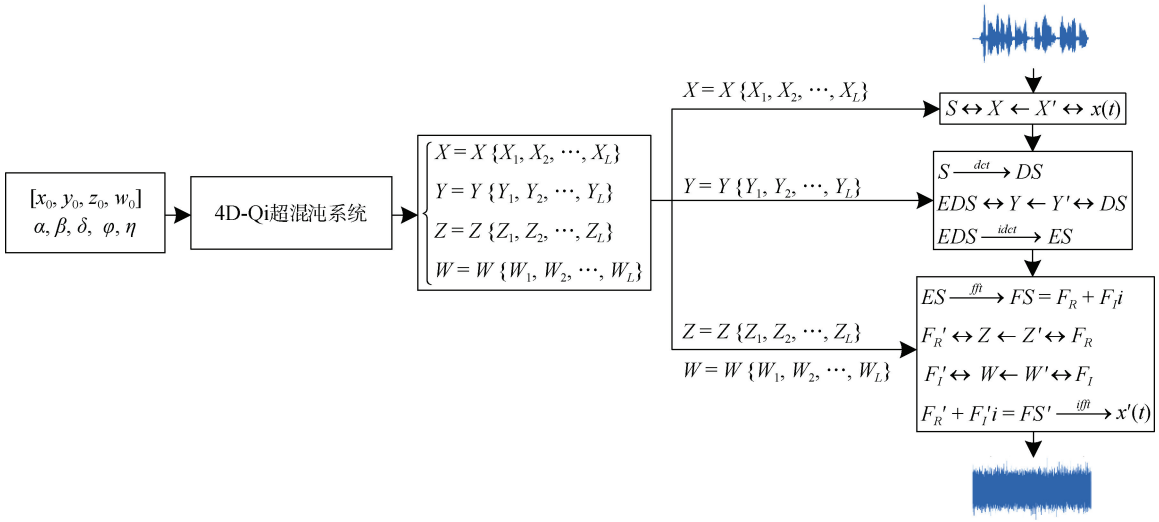

图 3　混合域置乱加密算法模型

Figure 3　The mixture domain scrambling encryption algorithm model

具体加密过程如下:

Step 1: 通过加密密钥 *Key′* 激发 4D-Qi 超混沌系统生成大小为 4×*L* 的随机序列(2)。利用序列 *X, Y, Z, W* 对原始语音进行混合域置乱加密。

$$\begin{cases} X = X\{X_1, X_2, \cdots, X_L\} \\ Y = Y\{Y_1, Y_2, \cdots, Y_L\} \\ Z = Z\{Z_1, Z_2, \cdots, Z_L\} \\ W = W\{W_1, W_2, \cdots, W_L\} \end{cases} \tag{2}$$

Step 2: 通过序列 *X* 对原始语音 *x*(*t*)的时域方面进行置乱加密得到密文序列 *S*。

$$S \leftrightarrow X \leftarrow X' \leftrightarrow x(t) \tag{3}$$

Step 3: 将密文序列 *S* 进行离散余弦变换(DCT), 得到 DCT 域频谱序列 *DS*。通过序列 *Y* 对序列 *DS* 进行置乱加密得到加密 DCT 域频谱序列 *EDS*, 进行 DCT 逆变换得到时域加密序列 *ES*。

$$EDS \leftrightarrow Y \leftarrow Y' \leftrightarrow DS \tag{4}$$

Step 4: 将序列 *ES* 进行频域方面快速傅里叶变换(FFT), 得到复数值形式的采样序列 *FS*, 然后分别提取其实部序列 F_R 与虚部序列 F_I。通过序列 *Z* 和序列 *W* 分别对实部序列 F_R 与虚部序列 F_I 进行置乱加密, 得到加密后的实部序列 F_R 和虚部序列 F_I'。根据序列 F_R 和序列 F_I 得到新的复数值形式的频域序列 *FS′*, 最后进行 FFT 逆变换得到加密语音 *x′*(*t*)。

$$\begin{cases} FS = F_R + F_I i \\ F_R' \leftrightarrow Z \leftarrow Z' \leftrightarrow F_R \\ F_I' \leftrightarrow W \leftarrow W' \leftrightarrow F_I \\ F_R' + F_I' i = FS' \end{cases} \tag{5}$$

Step 5: 将原始语音库中的语音片段按照上述加密算法进行加密操作, 构建密文语音库。

2.2　特征融合

频谱通量(Spectrum Flux, *SF*)表征所有相邻两个音频帧在频谱内差异的平均值, 可用来描述某段语音信号的频谱能量变化并能表现语音信号能量的突变程度。其定义如式(6), 式(7)所示:

$$Z(m,k) = \left| \sum_{n=-\infty}^{\infty} x(n) w(mM - n) e^{j\frac{2\pi}{M}kn} \right| \tag{6}$$

$$SF = \frac{1}{(M-1)(P-1)} \times \sum_{m=1}^{M-1} \sum_{p=1}^{P-1} \left[\log Z(m,p) - \log Z(m-1,p) \right] \tag{7}$$

式中, *x*(*n*)是经过预处理后的语音信号, *w*(*n*)是窗函数, *M* 是窗的长度, *P* 是离散傅里叶变换的系数。

峭度因子(Kurtosis Factor, *Ku*)是描述语音波形尖峰度的一个评价指标, 用于表达语音信号在某个频率范围内峭度值的大小, 其定义如式(8)所示:

$$Ku(f) = \frac{E(|X(t,f)|^4)}{(E(|X(t,f)|^2))^2} - 2 \tag{8}$$

式中, 符号| •|表示取模运算, *E*(•)表示数学期望, *X*(*t*, *f*)为语音信号 *x*(*t*)在频率 *f* 处的复包络。

本文提出的检索框架是基于内容的语音检索方法, 故提取语音信号的频谱通量特征和峭度因子特征作为原始输入特征。由于语音信号由清音与浊音组成, 相邻帧间能量的突变程度较大, 导致单一的语音特征不能完整的表达语音信号。因此本文算法将语音信号的频谱通量特征与峭度因子特征迭代内积得到融合特征向量, 如式(9)所示:

$$SF \bullet Ku \to V \tag{9}$$

融合后的特征向量能够同时体现语音帧内以及帧与帧之间的特征, 更加完整地表达语音信号特征。

2.3 密钥分配索引表的构建

采用索引的方式可以对系统中的数据进行快速的定位和操作, 因此建立索引表可以有效组织存储系统中的数据, 提高存储系统插入和查询性能。本文基于特征向量和 Bagging 分类算法构建密钥分配索引表, 具体构建过程如下:

Step 1: 设原始语音为 $x(t)$, 将 $x(t)$进行预处理得到帧长为 L, 帧移为 M, 总帧数为 N 的语音信号 $x_i(n)$。其中, $x_i(n)$表示第 i 帧的第 n 个采样值。然后, 利用上一节提出的特征提取方法得到一维融合特征向量 $V=\{V(i)|i=1,2,\ldots,N\}$。最后, 将融合特征向量 V 进行二值化处理, 得到原始语音的差分哈希序列 $h=\{h(i)|i=1,2,\ldots,N\}$。

二值化处理过程: 设差分哈希序列 $h(1)$为 1。如果特征向量 $V(i)$的第 i 个采样值大于第 i–1 个采样值, 那么差分哈希序列 $h(i)$的值为 1, 否则为 0。如式(10)所示:

$$h(i)=\begin{cases}1, V(i)>V(i-1)\\ 0, else\end{cases} \tag{10}$$

式中, i =1,2,…,N。

Step 2: 将原始语音信号的差分哈希序列 h 作为训练集, 并定义分类标签为 K, 使用 Bagging 分类算法进行训练学习得到其分类模型。然后, 根据分类结果定义各类属语音数据的单一映射密钥 $Key=\{Key(i)|i=1,2,\ldots,K\}$。

其中, Bagging 算法是一种将多个独立的基分类器集成为一个分类器的集成学习方法, 适用于对训练数据微小变化敏感的模型。其可重复取样的采样方式使得各训练子集间相互独立, 从而得到具有较高泛化性能及较大差异度的基分类器。Bagging 算法的训练过程如算法 1 所示。

Step 3: 根据原始语音的逻辑地址 S、融合特征向量 V、差分哈希 h, 类别 K 以及对应密钥 Key 之间的单一映射关系构建密钥分配索引表。构建的索引表模型如图 4 所示。

算法 1: Bagging 算法

输入: 样本集 $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_m,y_m),\}$, 其中 x 为样本, y 为样本标签;

训练轮数 P;

基础学习算法 δ。

输出: 分类结果 $Z(x)$。

过程:

for t =1,2, …,P; do

D_t=Bootstrapping(D); // 对数据集 D 采样

$z_t=\delta(D_t)$; // 在数据集 D_t 上训练模型

end

$Z(x)=\arg\max_y \sum_{t=1}^{T} I(z_t(x)=y)$ // I 为指示函数

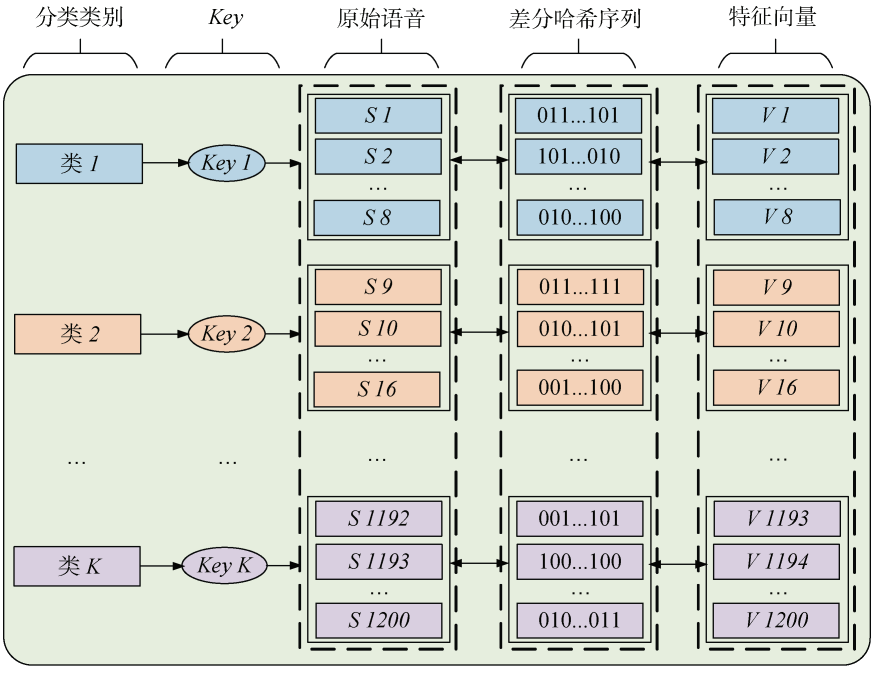

图 4 密钥分配索引表

Figure 4 Key distribution index table

2.4 生物特征模板与生物哈希索引表的构建

利用上一节构建的密钥分配索引表可查询不同类别的特征向量 V 及其对应密钥 Key, 建立具有单一映射密钥的安全生物特征模板。生物特征模板构建模型如图 5 所示。

具体构建过程如下:

Step 1: 查询密钥分配索引表, 获得各类语音的单一映射密钥 Key。然后, 根据不同的密钥 Key 激发 3D-LSCM 混沌系统[38]产生 K 个长度为 N 的一维随机序列 $q(i)$。最后, 将序列 $q(i)$进行施密特正交化得到相互正交的随机序列 $Q(i)$。

其中, 3D-LSCM 混沌系统由 Logistic 映射与 Sine 映射相互嵌套耦合, 具有更大的混沌映射范围和分布不均匀性。其数学表达式如式(11)所示:

$$\begin{cases} x_{i+1} = \sin(\pi\eta(z_i+3)x_i(1-x_i)) \\ y_{i+1} = \sin(\pi\eta(x_{i+1}+3)y_i(1-y_i)) \\ z_{i+1} = \sin(\pi\eta(y_{i+1}+3)z_i(1-z_i)) \end{cases} \tag{11}$$

式中, η 为控制变量且 $\eta\in[0, 5]$时, 系统处于混沌状态。

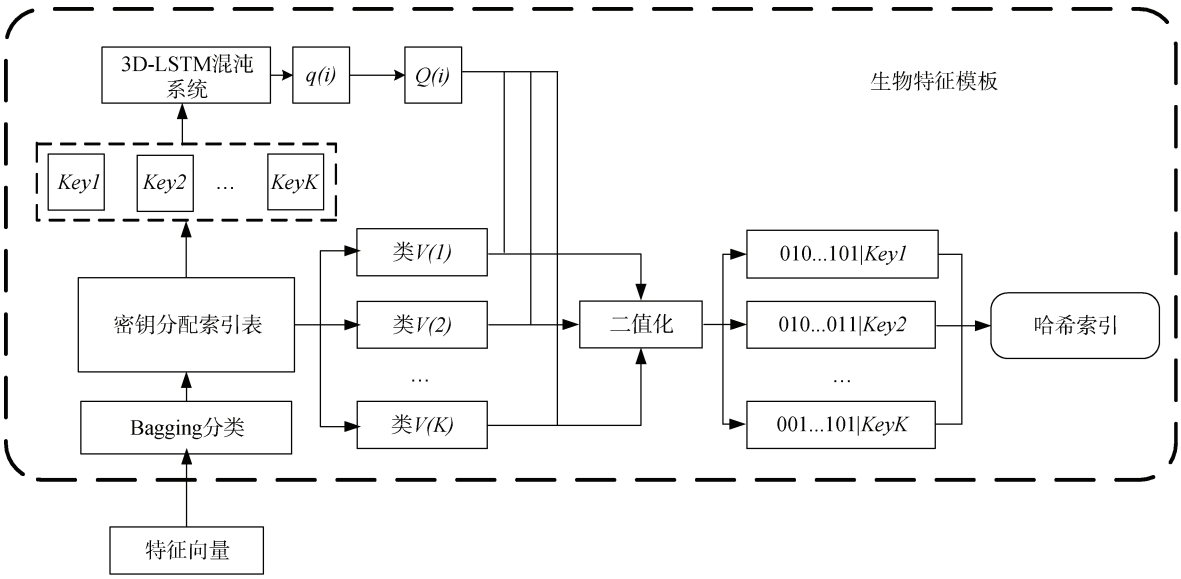

图 5 生物特征模板模型

Figure 5 Biometric template model

Step 2: 将不同类别语音数据的特征向量 V 分别与其对应的随机序列 $Q(i)$进行迭代内积, 可以得到相应的各类别生物特征向量 $D=\{D(i)|i=1,2, \ldots, N\}$。然后, 将生物特征向量 $D(i)$ 二值化生成生物哈希序列 $H=\{H(i)|i=1,2, \ldots, N\}$, 即: 哈希索引。哈希序列计算如式(12)所示:

$$H(i)=\begin{cases} 1, D(i)>D(i-1) \\ 0, else \end{cases} \tag{12}$$

Step 3: 将哈希索引上传至云端储存。

该算法建立在可撤销生物特征模板的思想上, 将差分哈希序列和 Bagging 分类算法结合, 构建可供准确查询与定位的密钥分配索引表, 并基于该索引表建立了 K 个生物特征模板, 其中每个生物特征模板将根据其单一映射密钥构建。对于受到攻击后的生物特征模板, 可以更新密钥, 快速生成新的生物特征模板。因此, 该算法构建的生物特征模板具有多样性、安全性和可撤销性。

在云端生物哈希索引表构建方面, 云端服务器接收到上传的密文语音库以及哈希索引后, 根据密文语音库中加密语音的逻辑地址 AD, 生物哈希序列 H 及其单一映射密钥 Key 之间的一一对应关系, 构建用于语音检索的云端生物哈希索引表。构建的生物哈希索引表模型如图 6 所示。

2.5 移动端语音检索

移动端发出查询请求时, 云端服务器获得待检索语音的哈希序列 H' 并查询云端生物哈希索引表, 与所有符合权限的哈希序列进行匹配检索, 将满足阈值条件的哈希序列所对应的密文语音作为检索结果反馈给用户。具体检索过程如下:

Step1: 将待检索语音 $x'(t)$ 按照上述相同方法进行特征提取并得到差分哈希序列 h'。

Step2: 将序列 h' 作为测试数据并根据 Bagging 分类判断该条语音所属类别, 查询密钥分配索引表获得对应密钥 Key。然后, 通过 Key 激发对应的生物特征模板, 得到待检索语音的生物哈希序列 H' 并上传至云端。

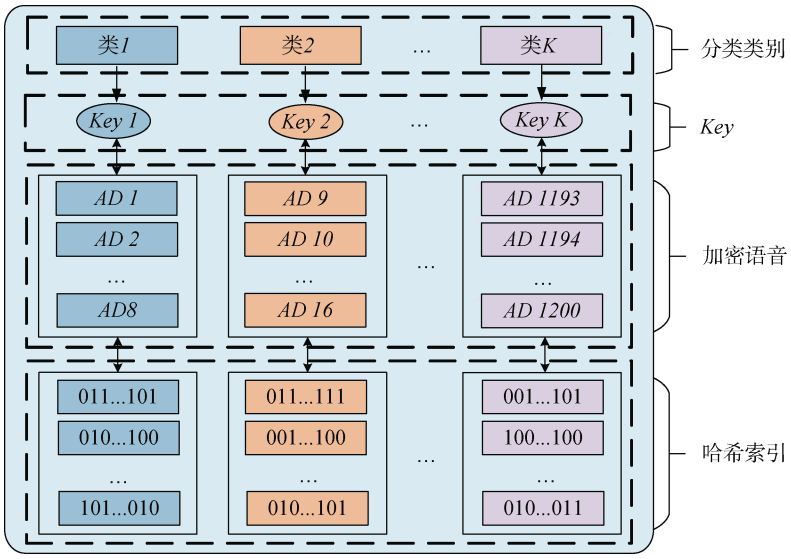

图 6　云端生物哈希索引表

Figure 6　Cloud biological hash index table

Step 3: 将序列 H' 与生物哈希索引表中对应类别的哈希索引进行检索匹配。在检索过程中, 根据归一化汉明距离 $D(:,:)$进行匹配查询, 即比特误码率(BER)。其计算公式(13)为:

$$\mathrm{BER}=D(H,H')=\frac{1}{N}\sum_{i=1}^{N}\left|H(i)\oplus H'(i)\right| \tag{13}$$

式中, $\oplus$表示异或逻辑运算, N 表示生物哈希序列的长度。

采用 BER 的假设检验对哈希序列的匹配结果进行描述。

U_0: 如果两条语音片段内容相同, 则有: $D \leqslant \tau$。

U_1: 如果两条语音片段内容不同, 则有: $D>\tau$。

其中, τ 为匹配阈值。当两条序列的汉明距离 D 小于等于阈值 τ 时, 表示成功检索到相关语音片段; 反之, 当两条序列的汉明距离 D 大于阈值 τ 时, 表示未检索到相关语音片段。

Step 4; 将匹配成功的哈希序列对应于密文语音库中相应的密文语音片段, 并将相应密文语音反馈给用户。用户在移动端可以根据相应解密方法对加密语音进行解密操作, 其中解密为加密的逆过程, 解密后的语音片段即为查询语音。

3　实验结果及分析

3.1　实验环境

实验所用语音均来自 TIMIT (texas instruments and massachusetts institute of technology) 语音数据库和 TTS (text to speech) 语音数据库。在原始语音数据库中有 1200 条不同的语音片段, 其中, 语音片段的格式为 WAV, 长度为 4s。

实验硬件平台为 Intel(R) Core(TM) i5-4200H CPU, 2.80GHz, 计算机内存为 8G。操作软件环境为 Windows 10 系统的 MATLAB R2020b。

本文算法所用参数如下: 帧长 L=200, 帧移 M=80, 总帧数 N=798, 分类类别 K=150。

3.2　区分性分析

区分性用于表征不同内容语音或者相同语音的可靠性。不同语音信号生物哈希码的 BER 值基本服从正态分布。本实验中, 1200 条原始语音片段经过计算一共可以得到 719400 个 BER 值, 它们基本服从正态分布, 如图 7 所示。

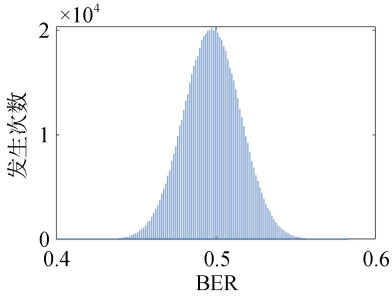

图 7　BER 正态分布图

Figure 7　Normal distribution of BER

根据棣莫弗-拉普拉斯中心极限定理, 哈希序列的归一化汉明距离近似服从 $\mu=p$, $\sigma=\sqrt{p(1-p)/N}$ 的正态分布, 其中, μ 为均值, σ 为标准差, p 为生物哈希序列中 0 或 1 发生的概率, N 为总帧数。表 1 给出了不同生物哈希序列长度时正态分布参数的理论值

与实验值。

表 1 不同哈希序列长度的正态分布参数

Table 1 Normal distribution parameters of different biological hash sequence lengths

N	理论值		实验值	
	μ_t	σ_t	μ_e	σ_e
798bits	0.50	0.0177	0.4976	0.0179
639bits	0.50	0.0198	0.4971	0.0199
532bits	0.50	0.0217	0.4966	0.0219
441bits	0.50	0.0238	0.4957	0.0239

由表 1 可以看出, 当哈希序列长度改变时, 本文算法经过实验得到的正态分布参数均近似等于理论值, 说明该算法有较好的随机性与抗碰撞性。

为了进一步衡量本文算法的区分性, 引入误识率(FAR)作为评价指标。FAR 表示将不同语音片段错误判断为相同语音片段的概率, 其公式(14)如下:

$$\mathrm{FAR}=\int_{-\infty}^{\tau} f(x\mid\mu,\sigma)\mathrm{d}x=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\tau} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \tag{14}$$

式中, FAR 代表误识率, τ 代表匹配阈值, μ 代表 BER 均值, σ 代表 BER 标准差。其中, FAR 的值越小, 表示误判语音的概率越小, 算法的区分性越好。

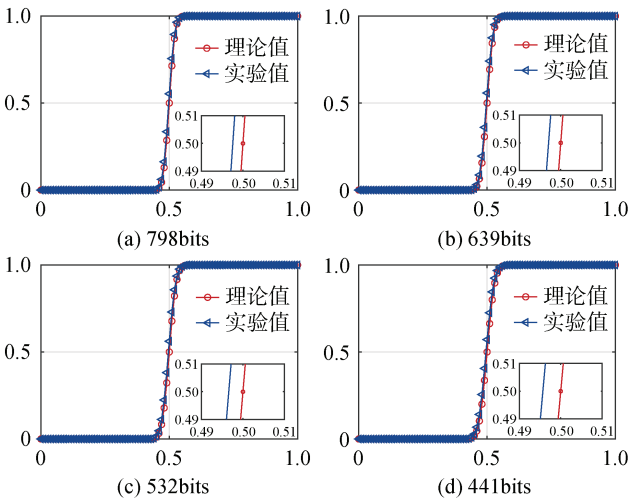

图 8 不同生物哈希序列长度的 FAR 曲线

Figure 8 FAR of different biological hash sequence lengths

由图 8 可以看出, 选择不同生物哈希序列长度时实验得到的 FAR 曲线与理论 FAR 曲线基本重合。因此, 本文算法具有较低的误识率。

表 2, 表 3 分别对比了本文算法中选取不同生物哈希序列长度时的 FAR 值以及本文算法与其他算法的 FAR 值。从表 2 可以看出, 随着哈希序列长度增加, FAR 值逐渐变小, 当生物哈希序列长度为 798 时, FAR 值最小, 区分性达到最佳。

表 2 不同哈希序列长度的 FAR 值

Table 2 FAR of different biological hash sequence lengths

τ	798bits	639nits	532bits	441bits
0.10	4.764e-110	1.627e-88	7.264e-74	1.126e-61
0.20	1.283e-62	1.719e-50	3.069e-42	2.355e-35
0.25	5.455e-44	1.477e-35	8.100e-30	5.102e-25
0.30	9.568e-29	2.463e-23	1.190e-19	1.468e-16
0.35	7.082e-17	8.147e-14	9.934e-12	5.733e-10
0.40	2.321e-08	5.605e-07	4.941e-06	3.179e-05

表 3 不同算法的 FAR 值

Table 3 FAR of different algorithms

τ	本文算法(798bits)	文献[39]	文献[40]	文献[41]	文献[42]
0.10	4.763e-110	8.006e-40	9.913e-47	3.654e-42	6.773e-47
0.20	1.283e-62	2.948e-23	5.271e-27	1.405e-24	2.992e-27
0.25	5.455e-44	1.014e-16	3.009e-19	1.215e-17	1.668e-19
0.30	9.568e-29	2.422e-11	6.870e-13	6.166e-12	3.893e-13
0.35	7.082e-17	4.092e-07	6.403e-08	1.874e-07	3.877e-08
0.40	2.321e-08	5,078e-04	2.540e-04	3.540e-04	1.714e-04

由于 FAR 值易受哈希序列长度影响, 为了更好地反映本文算法在不同哈希序列长度时的区分性,可以引入熵率(ER)作为另一衡量标准。ER 主要比较哈希算法的综合性能, 且不受哈希序列长度的影响, 是算法区分性和摘要性的联合评价指标, 其取值范围是(0,1)。ER 值越接近 1, 表示算法的区分性越好。ER 的计算公式如式(15), 式(16)所示:

$$\mathrm{ER}=-P\log_2 P-(1-P)\log_2(1-P) \tag{15}$$

$$P=\frac{1}{2}\left(\sqrt{\frac{|\sigma_t^2-\sigma_e^2|}{\sigma_t^2+\sigma_e^2}}+1\right) \tag{16}$$

式中, σ_t 和 σ_e 分别表示 BER 值的理论与实验标准差; ER 是熵率, 且 ER 值越大, 区分性越好。

由表 4 可以看出, 本文算法在不同生物哈希序列长度时的 ER 值均接近于 1, 可以说明该算法有较好的区分性。

由表 3, 表 5 可以看出, 相比于其他算法, 本文算法具有更理想的 FAR 值和 ER 值。当匹配阈值为 0.35 时, 本文算法的 FAR 值为 7.082×10^{-17}, 表示判断 10^{17} 条语音片段时大约有 7.082 条被误识, 在相同条件的匹配阈值下, 是文献[39]的 5.778×10^{9} 倍, 文献[40]的 9.041×10^{8} 倍, 文献[41]的 2.646×10^{9} 倍, 文献[42]的 5.474×10^{8} 倍。

表 4 不同哈希序列长度的 ER 值

Table 4 ER of different biological hash sequence lengths

N	ER
798bits	0.9919
639bits	0.9964
532bits	0.9934
441bits	0.9970

表 5 不同算法的 ER 值

Table 5 The ER of different algorithms

	σ_t	σ_e	ER
本文算法	0.0177	0.0179	0.9919
文献[39]	0.0264	0.0304	0.8964
文献[40]	0.0264	0.0277	0.9651
文献[41]	0.0264	0.0295	0.9187
文献[42]	0.0250	0.0279	0.9196

综上所述, 本文算法具有较好的区分性, 可以满足基于内容的语音检索需求。

3.3 鲁棒性分析

为了测试本文算法的鲁棒性, 首先对语音库中的 1200 条原始语音进行了表 6 所示的 15 种内容保持操作(Content preservation operations, CPOs), 然后分别计算内容保持操作后的 BER 均值和 BER 最大值。计算结果如图 9 所示。

表 6 内容保持操作

Table 6 Content preservation operations

操作手段	操作方法	简称
音量调节 1	音量增加 50%	V.1
音量调节 2	音量降低 50%	V.2
低通滤波 1	6 阶 FIR 滤波, 截止频率 3.4kHz	F
低通滤波 2	6 阶巴特沃斯滤波, 截止频率 3.4kHz	B
重采样 1	采样率下降至 8 kHz, 再上升至 16 kHz	R.1
重采样 2	采样率上升至 32 kHz, 再下降至 16 kHz	R.2
添加回声	延时 300ms, 叠加衰减 25%	E
窄带噪声 1	30dB, 中心频率 0～4 kHz	G.1
窄带噪声 2	40dB, 中心频率 0～4 kHz	G.2
窄带噪声 3	50dB, 中心频率 0～4 kHz	G.3
MP3 压缩 1	32kb/s	M.1
MP3 压缩 2	64kb/s	M.2
MP3 压缩 3	96kb/s	M.3
MP3 压缩 4	128kb/s	M.4
MP3 压缩 5	192kb/s	M.5

由图 9 可知, 原始语音经过上述 15 种内容持操作后 BER 最大值的分布区间为(0.0163, 0.2694), BER 均值的分布区间为(0.0016, 0.1265), 说明本文算法具有较好的鲁棒性。在以上 15 种内容保持操作中, 由于音量调节与重采样操作仅对语音信号的幅值产生影响, 而对信号频域影响较小, 因而鲁棒性较好; 而 FIR 滤波和巴特沃斯滤波操作会滤除语音信号大于 3.4kHz 的数据, 在时域内降低部分语音强度, 频域内滤除部分语音频谱, 因而鲁棒性较差; 在回声操作方面, 由于添加窄带噪声会将原始语音与添加的高斯白噪声相叠加, 从而在时域内改变语音强度, 频域内影响语音的声纹, 因而鲁棒性较差。

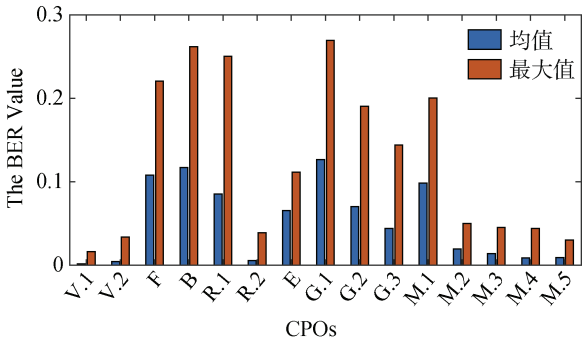

图 9 BER 均值和最大值

Figure 9 The BER mean and max value

表 7 不同算法的 BER 均值

Figure 7 The BER mean value of different algorithms

CPOs	本文算法	文献[39]	文献[40]	文献[41]	文献[42]
V.1	0.0016	0.0052	0.0925	0.0264	0.0231
V.2	0.0044	0.0014	0.0089	4.60e-05	7.91e-04
F	0.1081	-	-	-	-
B	0.1169	-	-	-	-
R.1	0.0855	0.0283	0.0304	0.0010	9.45e-04
R.2	0.0055	-	-	0.0104	0.0062
E	0.0653	0.1505	0.2375	0.1427	-
G.1	0.1265	-	-	0.0779	0.0055
G.2	0.0701	-	-	-	-
G.3	0.0441	0.0267	0.0416	-	-
M.1	0.0986	-	-	-	0.0214
M.2	0.0194	-	-	-	-
M.3	0.0137	-	-	-	-
M.4	0.0086	0.1928	-	0.2189	0.0019
M.5	0.0092	-	-	-	0.0019

为了进一步证明本文算法的鲁棒性, 将本文算法的BER均值与文献[39-42]中算法的BER均值进行比较, 结果如表 7 所示。

由表 7 可知, 相较于其他算法, 本文算法经过 15 种内容保持操作后 BER 均值的最大值为 0.1265, 远远小于文献[39]中算法的 0.1928, 文献[40]中算法的 0.2375 和文献[41]中算法的 0.2189, 说明该算法有更

好的鲁棒性。

误拒率(FRR)表示将正确语音错误判断为错误语音的概率, 可以更加直观的衡量算法的鲁棒性和区分性, 其算法如式(17)所示:

$$\mathrm{FRR}=1-\int_{-\infty}^{\tau} f(x\mid\mu,\sigma)\mathrm{d}x=1-\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\tau} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \tag{17}$$

式中, FRR 代表误拒率, FRR 的值越小, 说明算法的鲁棒性越好。

根据式(17), 得出 15 种内容保持操作后语音库中不同语音片段 BER 的 FRR 值, 并结合原始语音库中不同语音片段 BER 的 FAR 值, 绘制 FRR-FAR 曲线图。如图 10 所示。

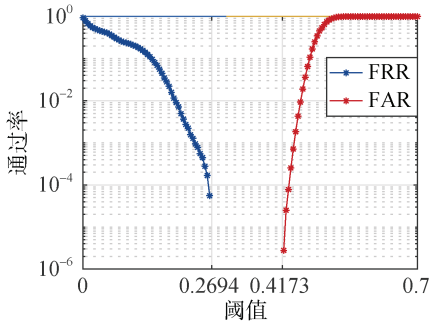

图 10 FRR-FAR 曲线图

Figure 10 FRR-FAR curve

由图 10 可知, 本文算法的 FRR-FAR 曲线图没有发生交叉, 并且匹配阈值的选择区间为(0.2694, 0.4173), 说明该算法可以灵活选择阈值, 准确识别内容保持操作和不同语音内容。

3.4 安全性分析

为提高云存储检索系统在语音数据传输与存储过程的安全性, 并保证用户从服务器上获取数据的完整性, 本文采用了混合域置乱加密算法。

3.4.1 密钥空间

理论上, 如果加密系统的密钥空间大于 2^{100}, 则认为在当前计算速度下, 加密系统可以有效抵御穷举攻击。提出的混合域置乱加密算法的密钥包括 4D-Qi 超混沌系统的初始值[x_0, y_0, z_0, w_0], 其步长均为 10^{-15}, 则密钥空间大小为 10^{60}, 密钥长度为 199bit。由于密钥空间远远大于 2^{100}, 说明该加密算法密钥空间足够大, 能够有效的抵御穷举密钥攻击。

3.4.2 安全性与完整性分析

为验证该加密算法的安全性, 从原始语音库中随机选取一条语音并分别对其进行加密与解密操作, 结果如图 11 所示。

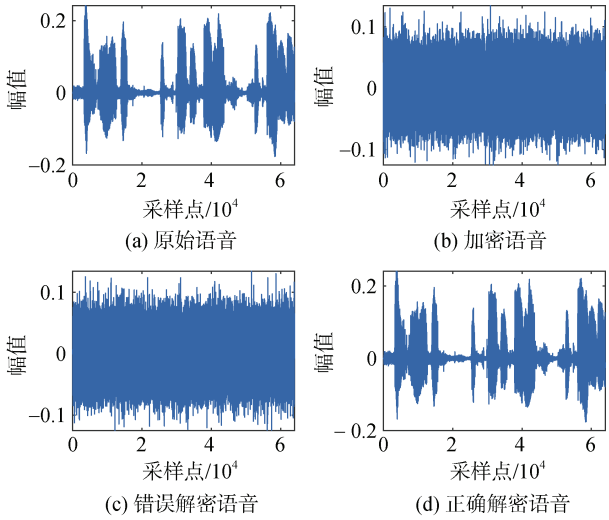

图 11 语音加解密对比图

Figure 11 Comparison of speech encryption and decryption

由图 11 可以看出, 加密后的语音波形与原始的语音波形完全不同, 加密效果十分明显。并且正确解密后的语音波形与原始语音波形相同, 说明该加密系统有较好的安全性, 能够保证用户从服务器上获取完整的语音数据。

为了进一步说明该加密算法的安全性, 验证加密前后语音信号的弱相关性, 随机选取该条原始语音的 32000 个样本点作为采样点, 以 $x(i)$为横坐标, $x(i+1)$为纵坐标, 将其加密前后的散点图进行对比, 结果如图 12 所示。

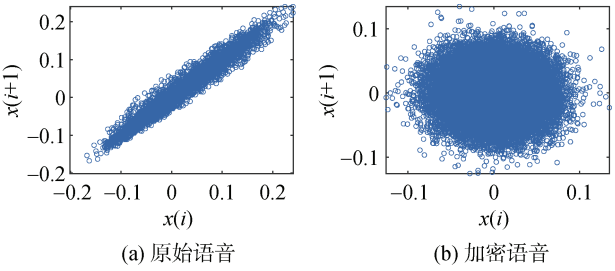

图 12 加密前后散点图

Figure 12 Scatter plot before and after encryption

由图 12 可知, 加密前后该条语音的采样点位置产生了明显的差异, 说明该加密算法极大的降低了语音信号的相关性。可以通过斯皮尔曼相关系数计算语音加密前后的相关性, 其定义如式(18)所示:

$$\rho=\frac{\sum_i (x_{(i)}-\bar{x})(y_{(i)}-\bar{y})}{\sqrt{\sum_i (x_{(i)}-\bar{x})^2 (y_{(i)}-\bar{y})^2}} \tag{18}$$

式中, ρ 为斯皮尔曼相关系数, $x_{(i)}$和 $y_{(i)}$分别为加密前与加密后的第 i 个采样点, $\bar{x}$ 与 $\bar{y}$ 分别为加密前后采

样点的均值。

通过公式计算出原始语音的相关性是 1, 为极强相关; 而加密语音的相关性是 9.3685×10^{-4}, 为极弱相关或无相关。可知该加密算法完全打破了明文的统计特性, 可以有效的抵御穷举统计分析的攻击。

为了更加充分的验证该加密算法的安全性, 从原始语音库中随机选取 100 条语音片段并计算斯皮尔曼相关系数, 计算结果如图 13 所示。

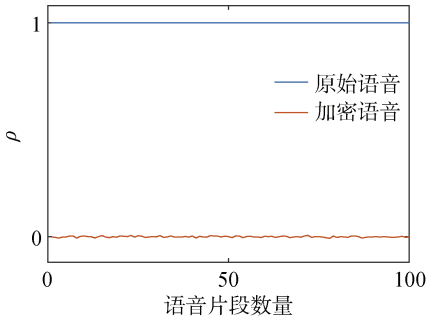

图 13 加密前后斯皮尔曼系数对比

Figure 13 Comparison of Spearman correlation coefficient

由图 13 可以看出, 原始语音的相关性为 1, 而加密语音的相关系数绝对值的最大值为 6.6×10^{-3}, 远远小于 10^{-2}。说明该加密算法能够极大降低语音的相关性, 解决云存储环境下语音检索系统的安全性问题。

3.5 分类效率

为了验证本文算法对差分哈希分类时的分类效率, 可引入分类正确率(Accuracy)作为评价指标。其计算公式如式(19)所示:

$$R_{\text{Accuracy}}=\frac{C_T}{C_T+C_F}\times100\% \tag{19}$$

式中, C_T 表示分类正确的语音个数, C_F 表示分类错误的语音个数。R_{Accuracy} 表示分类正确率, 且其值越接近于 1, 表示分类正确率越高。

为了衡量本文分类方法的性能, 验证分类正确率与分类数目 K 的关系, 计算得到 15 种内容保持操作在不同 K 时的分类正确率, 结果如表 8 所示。

此外, 分类时间也是衡量分类效率的重要指标, 因此计算不同 K 时分类一条语音所需的平均时间 T, 如表 9 所示。

综合表 8, 表 9 可以得出, 当 K 变大时, 分类正确率会更高, 所需的平均分类时间也会越长。这是因为随着 K 的增加, Bagging 算法对样本集的属性分析更加精确, 从而提高了分类正确率。同时, 由于需要分析的样本集属性增多, 算法运行时间也会相应变长, 导致算法的分类效率有所下降。

表 8 不同 K 时的分类正确率

Table 8 Classification accuracy of different K

(%)

CPOs	K=50	K=100	K=150	K=200
V.1	100	100	100	100
V.2	100	100	100	100
F	100	100	100	100
B	96	98	100	100
R.1	99	100	100	100
R.2	100	100	100	100
E	98	98	100	100
G.1	96	99	100	100
G.2	100	100	100	100
G.3	100	100	100	100
M.1	100	100	100	100
M.2	100	100	100	100
M.3	100	100	100	100
M.4	100	100	100	100
M.5	100	100	100	100

表 9 不同 K 时的平均分类时间

Table 9 Average classification time of different K

K	T(s)
50	0.8062e-2
100	0.8660e-2
150	1.0574e-2
200	1.3323e-2

结合不同分类数目 K 时的分类正确率和平均分类时间 T, 选择本文算法的分类数目 K=150。本文算法中所用的分类方法对上述 15 种内容保持操作后语音的差分哈希进行分类时的分类正确率均为 100%, 并且在保证极高分类正确率的同时耗费较短的分类时间。说明 Bagging 分类算法对本文算法获得的差分哈希序列有较好的分类效果, 可以满足对不同内容保持操作语音的分类, 从而实现语音类内检索, 提高系统的检索效率。

3.6 检索性能分析

3.6.1 检索精度

在语音检索系统中, 查全率(Recall, R)与查准率(Precision, P)是衡量检索性能的重要指标, 其计算公式如式(20), 式(21)所示:

$$R=\frac{S_T}{S_T+S_N}\times100\% \tag{20}$$

$$P=\frac{S_T}{S_T+S_F}\times100\% \tag{21}$$

式中, S_T 表示检索结果中正确语音且被检索到的数量, S_N 表示正确语音且未被检索到的数量, S_F 表示错误

语音且被检索到的数量。

根据图 10 可以得出, 匹配阈值的取值范围是(0.2694, 0.4173)。为了保证本文算法在不同内容保持操作后仍然具有较好的查全率和查准率, 本文采用 0.35 作为门限值。

由表 10, 表 11 可以看出, 该算法经过 15 种内容保持操作后, 仍能保证查全率和查准率均为 100%, 可知该算法具有极高的检索精度。同时相较于文献[39-40,42], 该算法在保证检索精度的同时能够满足更多种类内容保持操作后语音的检索需求。

表 10 不同算法查准率对比

Table10 Comparison of precision of different algorithms (%)

CPOs	本文算法	文献[39]	文献[40]	文献[42]
V.1	100	100	100	100
V.2	100	100	100	100
F	100	-	-	100
B	100	-	-	-
R.1	100	100	100	100
R.2	100	-	-	100
E	100	99	100	100
G.1	100	-	-	-
G.2	100	-	-	-
G.3	100	100	100	100
M.1	100	-	-	-
M.2	100	-	-	-
M.3	100	-	-	-
M.4	100	98	-	-
M.5	100	-	-	-

表 11 不同算法查全率对比

Table 11 Comparison of recall of different algorithms (%)

CPOs	本文算法	文献[39]	文献[40]	文献[42]
V.1	100	100	100	100
V.2	100	100	100	100
F	100	-	-	100
B	100	-	-	-
R.1	100	100	100	100
R.2	100	-	-	100
E	100	100	100	100
G.1	100	-	-	-
G.2	100	-	-	-
G.3	100	100	100	100
M.1	100	-	-	-
M.2	100	-	-	-
M.3	100	-	-	-
M.4	100	100	-	-
M.5	100	-	-	-

3.6.2 检索效率

为了衡量该算法的检索效率, 随机选取 1200 条语音进行检索匹配并计算检索时间, 结果如表 12 所示:

表 12 不同算法的检索效率对比

Table 12 Retrieval efficiency of different algorithms

	主频(GHz)	语音片段长度(s)	平均检索时间(s)
本文算法	2.8	4	9.4957e-4
文献[39]	2.5	4	0.1000
文献[40]	2.5	4	0.1467
文献[42]	2.5	4	0.0667

由表 12 可知, 本文算法平均检索时间为 9.4957 $\times 10^{-4}$s, 是文献[39]中算法的 105 倍, 是文献[40]中算法的 154 倍, 是文献[42]中算法的 70 倍, 相较于其他算法, 检索效率极大提升。这是因为本文算法在移动端对待检索语音的差分哈希序列进行分类, 使得在云端检索时仅需进行类内检索, 无需与整个语音库进行匹配, 从而极大提高了检索效率。

4 结论

本文将超混沌系统与 Bagging 分类应用于语音生物哈希检索算法, 提出了一种基于双哈希索引的高效语音生物哈希安全检索算法。在该算法中, 提出混合域置乱加密对原始语音加密并构建密文语音库, 有效防止了语音数据的泄露, 保证了语音数据传输与存储过程的安全性; 通过采用 Bagging 分类对语音数据的差分哈希分类, 构建密钥分配索引表, 建立了具有多样性、安全性的可撤销生物特征模板, 确保受到攻击后的模板能够快速更新生成新的生物特征模板; 根据密文语音库和哈希索引构建云端生物哈希索引表, 减少了云端匹配时的哈希索引数量, 显著提高了检索效率与准确性, 平衡了不同内容保持操作语音的检索性能。

由于本文算法经受低通滤波器操作和窄带高斯噪声后的鲁棒性不够理想, 且采用的特征融合技术会在一定程度上影响算法的整体效率。因此, 未来的研究目标是进一步优化算法的特征提取与融合算法, 提升算法的鲁棒性, 获得理想的检索效率。

参考文献

[1] Foote J. An Overview of Audio Information Retrieval[J]. *Multimedia Systems*, 1999, 7(1): 2-10.

[2] Wang J B, Xu S, Zheng F, et al. Learning Efficient Hash Codes for Fast Graph-Based Data Similarity Retrieval[J]. *IEEE Transactions on Image Processing*, 2021, 30: 6321-6334.

[3] Pan Z B, Wang L Z, Wang Y, et al. Product Quantization with Dual Codebooks for Approximate Nearest Neighbor Search[J]. *Neurocomputing*, 2020, 401: 59-68.

[4] Mishra S, Khetarpaul S. UR-Tree: A Spatial Index Structure for Handling Multiple Point Selection Queries[J]. *Multimedia Tools and Applications*, 2023, 82(6): 8093-8111.

[5] Norouzi M, Punjani A, Fleet D J. Fast Search in Hamming Space with Multi-Index Hashing[C]. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 3108-3115.

[6] Gaikwad S K, Gawali B W, Yannawar P. A Review on Speech Recognition Technique[J]. *International Journal of Computer Applications*, 2010, 10(3): 16-24.

[7] López-Espejo I, Tan Z H, Hansen J H L, et al. Deep Spoken Keyword Spotting: An Overview[J]. *IEEE Access*, 2021, 10: 4169-4199.

[8] Christiansen R, Rushforth C. Detecting and Locating Key Words in Continuous Speech Using Linear Predictive Coding[J]. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, 1977, 25(5): 361-367.

[9] Higgins A, Wohlford R. Keyword Recognition Using Template Concatenation[C]. *ICASSP '85. IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing*, 2003: 1233-1236.

[10] De Felipe I, Hristidis V, Rishe N. Keyword Search on Spatial Databases[C]. *2008 IEEE 24th International Conference on Data Engineering*, 2008: 656-665.

[11] Singh V, Zong B, Singh A K. Nearest Keyword Set Search in Multi-Dimensional Datasets[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(3): 741-755.

[12] Niu X M, Jiao Y H. An Overview of Perceptual Hashing[J]. *Acta Electronica Sinica*, 2008, 36(7): 1405-1411.
(牛夏牧, 焦玉华. 感知哈希综述[J]. *电子学报*, 2008, 36(7): 1405-1411.)

[13] He S F, Zhao H A. A Retrieval Algorithm of Encrypted Speech Based on Syllable-Level Perceptual Hashing[J]. *Computer Science and Information Systems*, 2017, 14(3): 703-718.

[14] Zhang Q Y, Zhou L, Zhang T, et al. A Retrieval Algorithm of Encrypted Speech Based on Short-Term Cross-Correlation and Perceptual Hashing[J]. *Multimedia Tools and Applications*, 2019, 78(13): 17825-17846.

[15] Gomez-Barrero M, Galbally J, Rathgeb C, et al. General Framework to Evaluate Unlinkability in Biometric Template Protection Systems[J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(6): 1406-1420.

[16] Wu L F, Ma Y K, Zhou P, et al. Review of Biometric Template Protection[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(11): 2407-2420.
(毋立芳, 马玉琨, 周鹏, 等. 生物特征模板保护综述[J]. *仪器仪表学报*, 2016, 37(11): 2407-2420.)

[17] Shahreza H O, Marcel S. Towards Protecting and Enhancing Vascular Biometric Recognition Methods via Biohashing and Deep Neural Networks[J]. *IEEE Transactions on Biometrics*, *Behavior*, *and Identity Science*, 2021, 3(3): 394-404.

[18] Teoh A B J, Goh A, Ngo D C L. Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 1892-1901.

[19] Wang Y, Huang Y B, Zhang R, et al. Multi-Format Speech Bio-Hashing Based on Energy to Zero Ratio and Improved LP-MMSE Parameter Fusion[J]. *Multimedia Tools and Applications*, 2021, 80(7): 10013-10036.

[20] Huang Y B, Li H, Wang Y, et al. A High Security BioHashing Encrypted Speech Retrieval Algorithm Based on Feature Fusion[J]. *Multimedia Tools and Applications*, 2021, 80(25): 33615-33640.

[21] Huang Y B, Wang Y, Zhang Q Y, et al. Biohashing Encrypted Speech Retrieval Based on Chaotic Measurement Matrix[J]. *Journal of Huazhong University of Science and Technology* (*Natural Science Edition*), 2020, 48(12): 32-37.
(黄羿博, 王勇, 张秋余, 等. 基于混沌测量矩阵的生物哈希密文语音检索[J]. *华中科技大学学报(自然科学版)*, 2020, 48(12): 32-37.)

[22] Sudjianto A, Yang Z B, Zhang A J. Single-Index Model Tree[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(3): 3101-3114.

[23] Ng W W Y, Li J Y, Tian X, et al. Bit-Wise Attention Deep Complementary Supervised Hashing for Image Retrieval[J]. *Multimedia Tools and Applications*, 2022, 81(1): 927-951.

[24] Krishnaraj N, Elhoseny M, Lydia E L, et al. An Efficient Radix Trie-Based Semantic Visual Indexing Model for Large-Scale Image Retrieval in Cloud Environment[J]. *Software*: *Practice and Experience*, 2021, 51(3): 489-502.

[25] Zheng T R, Han J Q. Syllable Lattice Based Chinese Speech Retrieval Techniques and Removing Redundancy Method from Indices[J]. *Acta Acustica*, 2008, 33(6): 526-533.
(郑铁然, 韩纪庆. 基于音节 Lattice 的汉语语音检索技术及其索引去冗余方法[J]. *声学学报(中文版)*, 2008, 33(6): 526-533.)

[26] Zhang K J, Zhang G L, Jiang C, et al. Research and Implementation of Security Cipher-Text Clustered Index Based on B Tree[C]. *2016 International Conference on Network and Information Systems for Computers*, 2017: 274-278.

[27] Cao Y, Long M S, Liu B, et al. Deep Cauchy Hashing for Hamming Space Retrieval[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1229-1237.

[28] Zhang D X, Agrawal D, Chen G, et al. HashFile: An Efficient Index Structure for Multimedia Data[C]. *2011 IEEE 27th International Conference on Data Engineering*, 2011: 1103-1114.

[29] Zhou Y, Li C L, Li W, et al. Image Encryption Algorithm with Circle Index Table Scrambling and Partition Diffusion[J]. *Nonlinear Dynamics*, 2021, 103(2): 2043-2061.

[30] Sahasrabuddhe A, Laiphrakpam D S. Multiple Images Encryption Based on 3D Scrambling and Hyper-Chaotic System[J]. *Information Sciences*, 2021, 550: 252-267.

[31] Hashemi S, Ali Pourmina M, Mobayen S, et al. Multiuser Wireless Speech Encryption Using Synchronized Chaotic Systems[J]. *International Journal of Speech Technology*, 2021, 24(3): 651-663.

[32] Yasser I, Mohamed M A, Samra A S, et al. A Chaotic-Based Encryption/Decryption Framework for Secure Multimedia Communications[J]. *Entropy*, 2020, 22(11): 1253.

[33] Hameed A S. A High Secure Speech Transmission Using Audio Steganography and Duffing Oscillator[J]. *Wireless Personal*

Communications, 2021, 120(1): 499-513.

[34] Wu J Q, Chen B L, Luo W Q, et al. Audio Steganography Based on Iterative Adversarial Attacks Against Convolutional Neural Networks[J]. *IEEE Transactions on Information Forensics and Security*, 2020, 15: 2282-2294.

[35] Khaleel A H, Abduljaleel I Q. A Novel Technique for Speech Encryption Based on K-Means Clustering and Quantum Chaotic Map[J]. *Bulletin of Electrical Engineering and Informatics*, 2021, 10(1): 160-170.

[36] Abdullah A A, Abbas Y K. Quantum audio steganography system[J]. *Journal of Engineering Science and Technology*, 2020, 15(3): 1562-1588.

[37] Qi G Y, Du S Z, Chen G R, et al. On a Four-Dimensional Chaotic System[J]. *Chaos*, *Solitons & Fractals*, 2005, 23(5): 1671-1682.

[38] Zeng K, Yu S M, Hu Y C, et al. Image Encryption Using 3D Logistic-Sine Cascade Map[J]. *Application of Electronic Technique*, 2020, 46(1): 86-91.

(曾珂, 禹思敏, 胡迎春, 等. 基于 3D-LSCM 的图像混沌加密算法[J]. *电子技术应用*, 2020, 46(1): 86-91.)

[39] Zhang Q Y, Ge Z X, Hu Y J, et al. An Encrypted Speech Retrieval Algorithm Based on Chirp-Z Transform and Perceptual Hashing Second Feature Extraction[J]. *Multimedia Tools and Applications*, 2020, 79(9/10): 6337-6361.

[40] Zhang Q Y, Ge Z X, Qiao S B. An Efficient Retrieval Method of Encrypted Speech Based on Frequency Band Variance[J]. *J Inf Hiding Multim Signal Process*, 2018, 9: 1452-1463.

[41] Zhang Q Y, Qiao S B, Huang Y B, et al. A High-Performance Speech Perceptual Hashing Authentication Algorithm Based on Discrete Wavelet Transform and Measurement Matrix[J]. *Multimedia Tools and Applications*, 2018, 77(16): 21653-21669.

[42] Zhang Q Y, Li G L, Huang Y B. An Efficient Retrieval Approach for Encrypted Speech Based on Biological Hashing and Spectral Subtraction[J]. *Multimedia Tools and Applications*, 2020, 79(39/40): 29775-29798.

黄羿博 于 2015 年在兰州理工大学获得博士学位。现任西北师范大学物理与电子工程学院副教授, CCF 会员。研究领域为多媒体信息安全。研究兴趣包括: 多媒体信息处理、信息安全。Email: huang_yibo@nwnu.edu.cn

陈德怀 于 2019 年在天津科技大学电子信息工程专业获得学士学位。现在西北师范大学电子信息专业攻读硕士学位。研究领域为多媒体信息处理。研究兴趣包括: 语音信号处理、语音识别。Email: 2020222192@nwnu.edu.cn

张秋余 于 1986 年在甘肃工业大学计算机应用专业毕业。现任兰州理工大学计算机与通信学院研究员, 博士生导师。研究领域为网络与信息安全。研究兴趣包括: 信息隐藏和隐写分析、多媒体通信技术。Email: zhangqylz@163.com