

双哈希索引的高精度大规模音频样例检索

高晓芳 罗森林 吕 英 罗志军 潘丽敏

(北京理工大学信息系统及安全对抗实验中心 北京 100081)

2014 年 1 月 14 日收到

2014 年 11 月 18 日定稿

摘要 实时音频流中对大规模音频样例进行检索时, 在保证准确率的条件下, 检索速度直接影响音频流实时处理能力。提出一种基于双哈希索引的大规模音频样例检索方法。该方法通过对大规模音频样例的音频特征进行自相似量化后, 分别根据自相似序列的分段向量均值和模值建立线性双哈希索引, 然后在音频流中进行搜索, 最后对搜索结果利用音频的时序和空间信息进行判断得到检索结果。实验结果表明, 本方法实现了大规模音频样例的一次检索, 且当采用 12 维 MFCC 音频特征, 音频样例时长为 16 s、音频样例规模小于 3100 时, 音频样例的检索准确率在 90% 以上, 检索速度大于 12000 倍速, 最高达到 16000 倍速。该方法在有效提高检索精度的基础上, 保证较高的检索速度。

PACS 数: 43.60, 43.72

Retrieval method of large scale audio samples based on Double hashing index

GAO Xiaofang LUO Senlin LÜ Ying LUO Zhijun Pan Limin

(Information System and Security & Countermeasures Experimental Center, Beijing Institute of Technology Beijing 100081)

Received Jan. 14, 2014

Revised Nov. 18, 2014

Abstract The capacity of processing audio stream in real time is affected directly by the detection speed with detection accuracy guaranteed. A method based on double hashing index to test large-scale audio samples is proposed. The method first does weighted self-similarity to the audio feature, secondly establishes double linear hashing indexes to the mean and modulus of self-similarity sequence, then searches in the audio stream and judge the search results by temporal and spatial information to get the detection results. The results of experiments show that the method implements the one detection of large scale audio samples. The real time detection speed is above 12000 xRT, the largest detection speed is 16000 xRT, and the detection accuracy is above 90% when the duration of audio samples is 16 s and the number of audio samples is 3100. The method improves detection speed with higher detection accuracy guaranteed.

引言

基于内容的音频检索技术不断发展, 越来越成为国内外学者关注的焦点, 能够快速有效地在海量音频数据中检索出自己所需要的内容。音频样例检索是在音频流中搜索识别音频样例的过程, 具有重要的应用价值, 可直接应用于网络安全监控和管理、音视频点播和网络电视节目、音频广告统计、网络音频产品的版权保护、音频分类等多个方面。传统的音频样例检索中, 一个音频样例需检索一次, N 个音频样例需检索 N 次, 在上述应用场景中, 需要监控的音频样

例数目巨大, 如能实现对多个音频样例的高精度快速检索, 即 N 个音频样例检索一次得到检索结果, 则能大大提升检索效率, 提高相关应用的实用性能。

目前, 音频检索领域的大多数研究是针对单音频样例进行检索, 相关研究主要包括哈希索引检索、直方图检索、分层蜂窝树 (HCT) 等方法。在哈希索引检索方面, Yu Yi^[1] 分析比较了局部敏感哈希搜索 (Location Sensitive Hash 简称为 LSH) 的 4 种算法, LSH-Dynamic Programming (DP), LSH-Sparse DP (SDP), Exact Euclidian LSH (E2LSH)-DP, E2LSH-SDP, 表明 E2LSH-SDP 方法具有较好的性能; Rynanen^[2] 提出使用 LSH 算法进行哼唱音乐检索; Baluja^[3] 使

用了排列分组的方法来设计局部敏感哈希的映射函数；Zheng Guibin^[4]提出了自相似权重量化方法和一种新的哈希函数的构造方法进行哈希映射，降低了哈希索引检索范围；Yu, Yi^[5]提出基于局部摘要和多层哈希索引的搜索方法，通过索引的粗检索和精检索来快速准确完成检索过程；Cotton C^[6]应用了哈希索引来检索音频中的重复音频事件。唐杰^[7]采用图形学的相关算法，提取查询样例语谱图的特征点，利用 Hash 结构进行特征点匹配，返回相似音频。Zhiyuan G^[8]提出采用分层过滤的方法检索音频片段，依次采用局部哈希索引、线性缩放、递归对齐等方法。Pedraza, C 等^[9]提出基于哈希函数和音频频率特征提取的音频检索指纹识别技术。

在直方图检索方面，McFee^[10]使用了基于梅尔频率倒谱系数 (Mel Frequency Cepstrum Coefficient 简称为 MFCC) 的特征量化，采用学习型相似度算法，对音频语义的代表性和在大规模音频数据库中的组织结构进行研究，分别用来提高检索的准确性和检索速度。齐晓倩^[11]提出一种基于 K-L 距离的两步固定音频检索方法。

其他检索方法方面，郑贵滨^[12]提出基于分段的实时声频检索方法；Knight^[13]提出了一种基于语音识别后用文本进行搜索相关音频的方法；Sturm^[14]提出基于自适应稀疏相似性的音频样例搜索概念；Cont A^[15]提出了基于音频流几何信息的相似度计算概念，通过统计音频流中信息几何来替代音频特征进行分析研究，适用于音频流中相似度计算；丁向荣^[16]提出了一种基于 HCT 的音频检索方法，通过最小距离控制 HCT 的增长和检索，能够精确的检出目标。

上述音频样例检索的研究预设场景均是针对给定的单个音频样例，在音频库中进行匹配，并不能实现多个待检音频样例的高精度检索。如基于模糊聚类的搜索方法准确性和搜索速度上较差；直方图搜索方法具有检索速度快和检索准确率高的特点，但是每次仅能使用单个样例进行快速直方图搜索。基于高维索引方法的难点在于如何建立有效的音频检

索结构，并利用音频的最近邻搜索方法有效地搜索该音频检索结构。本文提出一种双哈希索引的大规模音频样例检索方法，实现对大规模音频样例的一次检索，即仅需对音频流文件进行一次检索动作，即可完成对 N 个音频样例的检索定位，同时保证检索的准确性。

1 哈希索引音频检索算法

1.1 算法原理

为实现大规模音频样例检索，本文提出双哈希索引的音频样例检索方法，使用自相似权重量化方法生成音频特征段向量，建立段向量均值和模值的双哈希表索引库，再对待检音频流采用哈希索引方法分别在段向量均值和模值哈希表中进行检索，结果求交集得到最终检索结果。原理如图 1 所示。所提方法分为离线处理和在线检索两个阶段。

在离线处理阶段，提取大规模音频样例的音频特征，对音频特征利用平滑处理生成特征矢量，利用自相似量化方法^[4]生成段向量，对段向量的模值和段向量的均值进行哈希映射，建立两个哈希索引表的哈希索引库，完成离线处理过程。

在在线检索阶段，对待检索音频流文件，提取音频特征然后利用自相似进行量化得到段向量，对段向量的模值和均值进行哈希映射，利用索引值在大规模音频样例的哈希索引库中进行搜索并得到两个哈希表交集对应音频段的特征数据，然后与音频流特征数据进行相似度比较获得检索结果。

1.2 段向量模值、均值生成

段向量的模值、均值是生成双哈希索引的依据，因此首先对提取的大规模音频样例的音频特征采用自相似量化法^[4]进行降维处理。设查询的向量为 $\overline{Q}(q_0, q_1, \dots, q_{m-1})$ ，待检音频流给定的数据集为 $\phi = \{\overline{P}_i | i = 0, 1, \dots, N\}$ ， N 是音频文件数量，其中的段向量 $\overline{P}_i(p_i, p_{i+1}, \dots, P_{i+m-1})$ 开始于第 i 帧，段长为 m

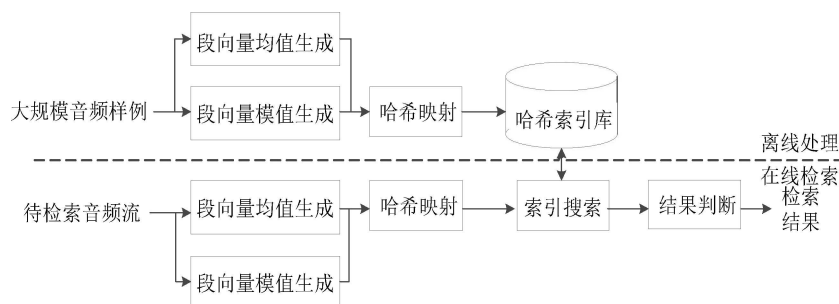


图 1 基于段向量模值和均值的双哈希索引检索算法原理图

时,如图 2 所示,对于音频帧的特征向量序列,设第 i 帧特征向量 \mathbf{X}_i ,生成的量化值即自相似序列表示为:

$$p_i = \sum_{j=-\text{Range}}^{\text{Range}} c_j \cdot d(\mathbf{X}_i, \mathbf{X}_{i-j}), \quad (1)$$

$$\left(\begin{array}{l} \text{Range} > 0, \\ \sum_{j=-\text{Range}}^{\text{Range}} c_j = 1 \end{array} \right)$$

其中, $d(\mathbf{X}_i, \mathbf{X}_{i-j})$ 表示两个特征向量间的距离, Range 为加权范围,决定了段向量中的元素值大小, c_j 为加权系数,其值可采用如下方式确定:

$$c_j = \frac{\text{Range} - |j|}{\sum_{i=-\text{Range}}^{\text{Range}} (\text{Range} - |i|)}, \quad (2)$$

$$(-\text{Range} \leq j \leq \text{Range})$$

之后对 P_i 按照给定的段长和段移进行划分获得相应的段向量。

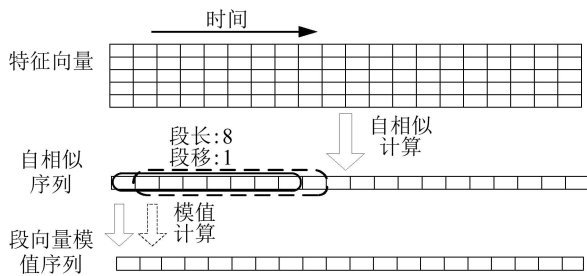


图 2 段向量模值生成

查询段向量为 $\overline{Q}(q_0, q_1, \dots, q_{m-1})$, $q_j (0 \leq j \leq m-1)$ 按照式 (1) 计算,则段向量模值为:

$$\|\overline{Q}\| = \sqrt{\sum_{j=0}^{m-1} q_j^2}, \quad (3)$$

段向量均值为:

$$M_{\overline{Q}} = \frac{\sum_{i=0}^{m-1} q_i}{m}. \quad (4)$$

实验对段向量模值、均值的最大值、最小值及其分布进行研究。哈希索引是根据段向量的模值和均值建立的,在音频样例检索中常采用 3 s 长的音频片段作为基本检索片段,因此段向量长度取值对应 3 s 时长的音频片段,段向量中的元素值大小取决于自相似权重量化范围 Range,本文对不同长度的量化取值范围进行实验(见图 3)。当自相似权重量化计算范围越大时,计算复杂度呈线性增加,当自相似权重计算范围为 5 时,如图 4 和图 5 中曲线所示,可以获

得较好的自相似权重计算速度以及数据分布;另一方面,由段向量模值、均值分布如图 4、图 5 所示,当自相似权重量化计算范围增大时,数据分布更加平坦,有利于建立索引进行检索时加快检索速度。反之,如果段向量模值、均值都集中于一个区域,则哈希映射的哈希索引值会较为集中,容易造成误检,重新检索会减缓检索速度。综上所述,此时段向量的模值和均值取值范围的上界和下界分别取与最大值和最小值相差 0.5% 处的值,使其包含大部分音频样例同时去除个别音频样例的影响。

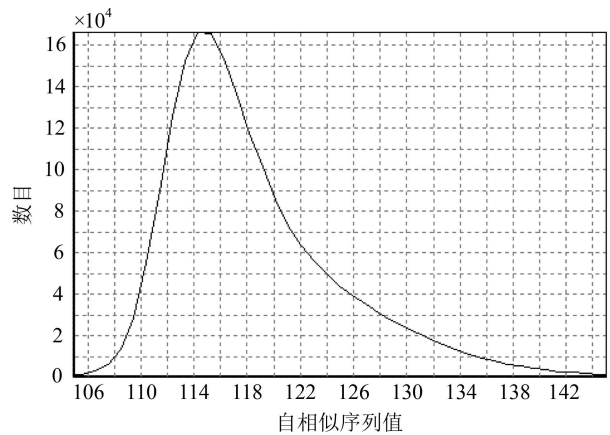


图 3 自相似序列值分布图

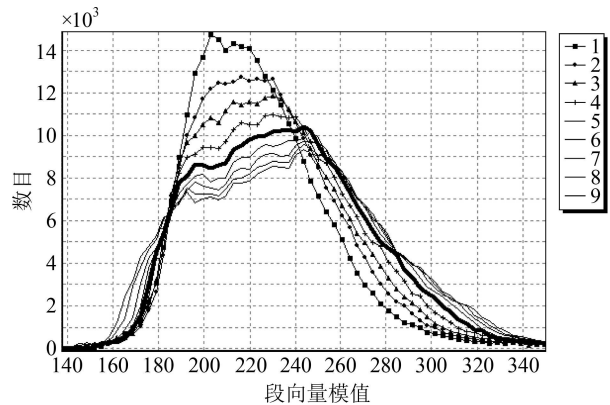


图 4 段向量模值分布图

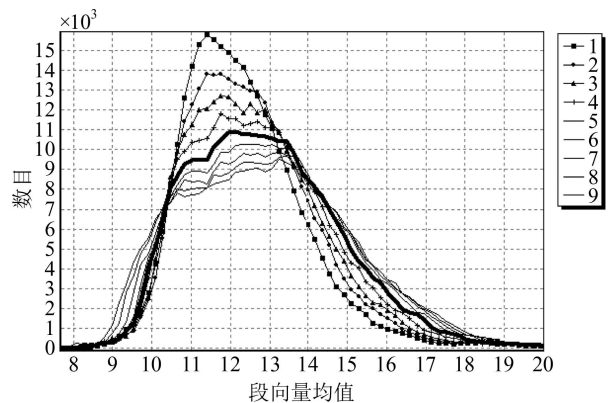


图 5 段向量均值分布图

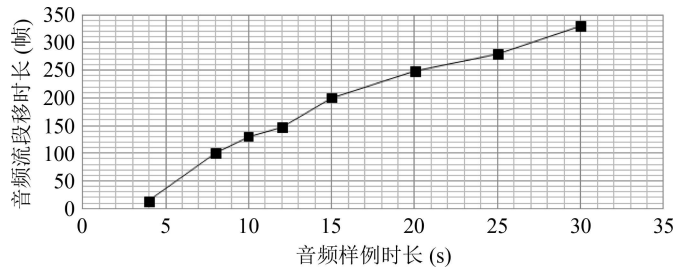


图 6 不同时长音频样例检索时段移长度 (帧)

段向量在音频样例和音频流中的段移具有不同的表现性能。采用不同的段向量段移可以减少数据查询量从而加快搜索速度，对于时长为 T 的音频样例，设 t_0 为段向量的时长， S_1 为音频样例中段向量的段移时长， S_2 为音频流中段向量的段移时长，要利用音频样例的起止时间和文件名来判断检索到的音频的准确性，必须在 $T - t_0$ 时间内完成多次检索，即

$$\begin{cases} \frac{(T - t_0)}{S_1} > 4, \\ \frac{(T - t_0)}{S_2} > 4. \end{cases} \quad (5)$$

由上式可见，对于不同时长音频样例，其 S_1 和 S_2 取值有较大差异，根据实验当 S_1 取值为 1 和 2 时，对于不同时长 S_2 均有较好的检索性能，从速度方面考虑，故取 $S_1 = 2$ 。不同时长音频样例对应的音频流中段向量段移 S_2 在保证良好的检索准确率情况下 (F 值大于 95%)，取值如图 6 所示。

1.3 哈希映射

所提方法通过对段向量模值、均值建立哈希索引表，完成双哈希索引库。

对段向量模值，使用如下哈希函数获得哈希索引值，

$$k = h(x) = (N - 1)g(x), \quad (6)$$

其中， x 为段向量模值， N 为哈希表桶尺寸大小， $g(x)$ 为模值标准化函数。

$$g(x) = \begin{cases} 0, & x \leq x_{\text{Lower}}, \\ \frac{x - x_{\text{Lower}}}{x_{\text{Upper}} - x_{\text{Lower}}}, & x_{\text{Lower}} < x < x_{\text{Upper}}, \\ 1, & x \geq x_{\text{Upper}}, \end{cases} \quad (7)$$

式中：

$$x_{\text{Lower}} = x_{\text{min}} + \frac{(x_{\text{max}} - x_{\text{min}})}{N},$$

$$x_{\text{Upper}} = x_{\text{max}} - \frac{(x_{\text{max}} - x_{\text{min}})}{N},$$

N 的取值根据实验需要存储的数据量的多少和算法

相似度搜索的精度来确定，文中取值为 30000。

同时，为段向量均值单独建立一个哈希表，通过获得的哈希表中桶编号，把音频样例的起止时间和文件名存入哈希表中，得到双哈希索引的索引表，建立大规模音频样例的哈希索引库。

1.4 索引搜索

使用段向量均值作为搜索索引时，如图 7 所示，如果 L 为中心线，则向量的均值算法就是把向量投影到该中心线上，当 $\sqrt{\theta/k} \geq \text{abs}(M_{\bar{Q}} - M_{\bar{P}_i}) \geq \sqrt{d(\bar{Q}, \bar{P}_i)/k}$ 时说明此时向量 \bar{P}_i 处于阈值为 θ 的向量 \bar{Q} 的邻域，即 $M_{\bar{Q}} - \sqrt{\theta/k} \leq M_{\bar{P}_i} \leq M_{\bar{Q}} + \sqrt{\theta/k}$ 时向量 \bar{P}_i 为邻域向量。

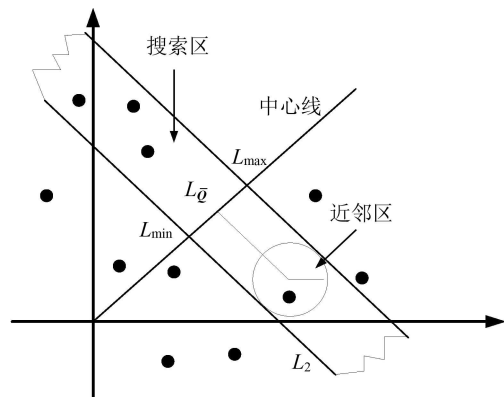


图 7 等均值算法搜索范围图

此时，对等均值邻域的哈希键值搜索范围为：

$$\begin{cases} K_i^{\text{Lower}} = h\left(x - \sqrt{\frac{\theta}{k}}\right), \\ K_i^{\text{Upper}} = h\left(x + \sqrt{\frac{\theta}{k}}\right). \end{cases} \quad (8)$$

对段向量模值的哈希索引表以及段向量均值的哈希索引表搜索的结果求交集，如图 8 所示阴影部分，得到哈希索引搜索的段向量集合。可见，求取段向量均值和模值哈希索引的交集，就可以进一步压缩 ϵ - 近似最近邻查询相似度搜索范围，加快检索速度。

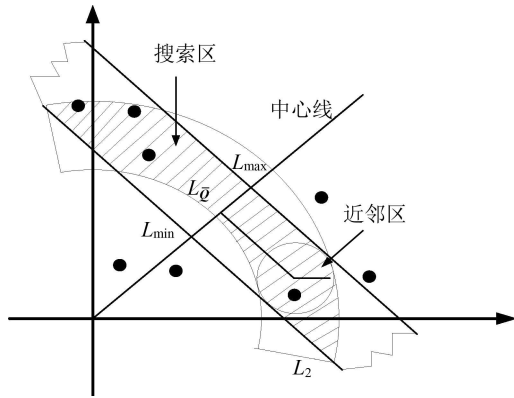


图 8 综合搜索范围图

1.5 结果判断

结合音频的起止时间和文件名判别检索结果，如图 9 所示。获得的交集检索结果可能为多个检索值，使用检索结果中的信息，如对应音频样例起止时间和文件名，对检索结果进行判别，得到准确的检索结果。

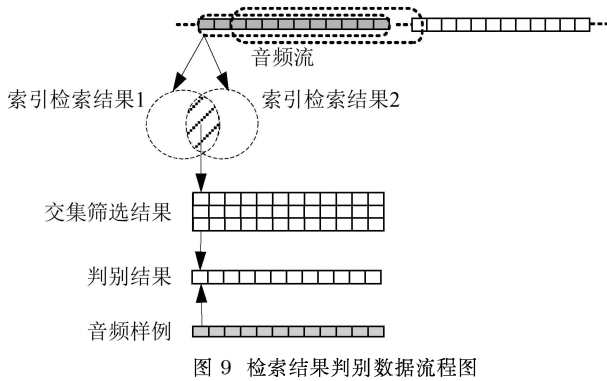


图 9 检索结果判别数据流程图

为加快算法的检索速度，采用不同的段向量段移进行搜索。检索开始后，对待检音频流每帧的段向量进行检索，当检索得到一定的音频样例信息后，就可以使用不同的段向量段移进行搜索达到快速检索的目的，按照以下方法进行判断：

(1) 逐帧检索：当前没有检索结果时，对音频流进行逐帧搜索，这可以避免段移过大造成漏检。当音频流中的音频帧与哈希表中匹配成功时，随即通过音频流后续的音频帧在哈希表里匹配到的音频样例的音频帧的起止时间连续性和文件名一致性来判决是否开始更加快速的逐段检索。判决的标准为：如果音频流的连续 4 个音频帧在哈希表中匹配到的音频样例的音频帧同属一个音频样例文件(通过文件名是否一致来判断)且起止时间连续(即 4 个匹配到的音频样例的音频帧起止时间有交叉)，则对音频流开始逐段检索，并将首帧的起始时间作为音频流检索结果的起始时间；否则继续逐帧检索。

(2) 逐段检索：对音频流的搜索按照段向量段移

S_2 进行逐段搜索，搜索后筛选的结果再按照音频样例时序进行判断时，时间计算按照段向量的段移为 S_2 进行计算。相对于按帧进行哈希索引检索的算法，其搜索的速度理论上提高到 $S_1 * S_2$ 倍。

(3) 检索结束判决：通过音频流中连续音频段在哈希表里匹配到的音频样例段起止时间的连续性和文件名的一致性来判断是否结束逐段检索：如果音频流的连续 4 个音频段在哈希表中匹配到的音频样例段不能满足以下任一条件，则结束检索，将最后一个音频段的段尾时间记为音频流检索结果的终止时间，输出检索结果即音频流中对应的音频样例和音频流中的起止时间：

(1) 次匹配到的音频样例段同属一个文件(通过文件名是否一致来判断)；

(2) 4 次匹配到的音频样例段起止时间连续(即音频样例段的起止时间不交叉、有间断)。

2 实验及分析

2.1 实验目的和数据源

利用段向量模值和均值的哈希索引检索算法进行检索，分别对单个哈希表搜索与双哈希表搜索的结果、算法有效性进行实验分析，实验目的是：

(1) 测试单、双哈希表对搜索命中数目的影响。

(2) 系统性能测试，测试不同时长音频样例对检索性能的影响。

(3) 系统性能测试，测试音频样例数目对检索性能的影响。

(4) 系统性能测试，比较所提方法和动态直方图搜索方法的性能。

实验数据资源：检索公开数据库 Content-Based Retrieval of Audio、The Musical Instrument Samples (MIS) dataset、The General Series 6000 Sound Effects Library、Cmu_arctic database 等音频文件数量和时长不能满足本文大规模音频样例的需求，因此本文采用自行构建的音频库，使用包含电影和电视音频文件约 120 小时原始音频库作为待检音频，并抽取时长为 30 s、静音帧比率小于 0.2、最大静音帧长度小于 300 ms 的音频段作为基本音频段，其中每个基本音频段在原始音频库中仅出现一次。如表 1 所示，基本音频段的总数为 11000 个，其中语音 3452 个、音乐 2070 个、其它音频 5478 个。实验时从每个基本音频段中截取一定时长的音频段作为音频样例，用来代表基本音频段来进行检索，截取的音频样例时长分别为 4 s, 8 s, 10 s, 12 s, 15 s, 20 s, 25 s, 30 s。

所有音频文件采用单声道、采样速率 8 kHz、解析度 16 位的 Windows PCM 格式。

表 1 音频样例分布

时长 (s)	音频样例数量 (个)			
	总数量	语音	音乐	其它音频
30	11000	3452	2070	5478

2.2 实验环境和条件

实验环境: Intel(R) Xeon(R) E5620 2.4 GHz 处理器、12 G 内存的服务器 (Windows Server 2003), 使用 C++ 进行编程实验, 采用单核单线程进行测试。

2.3 评价方法说明

评价方法: 采用基于音频样例数的评价参数描述系统检索的准确度。采用 xRT 描述大规模音频样例在实时音频流中的检索速度, 区别于用来描述单一音频样例在音频库中的实时搜索倍速^[5]。

算法的检索对象是音频流, 音频流经过检索后被切分为若干片段, 若检索出的音频样例片段与真实的音频样例片段在时序上有重叠, 则认为该音频片段被正确检索; 若没有重叠, 则一定存在误检或漏检的现象, 具体见图 10 所示。

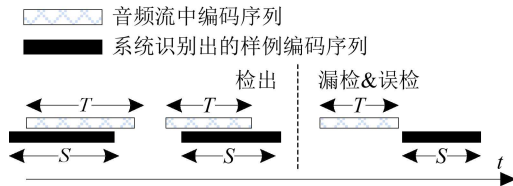


图 10 检出准则示意图

评价方法 I: 音频样例数准确率 P_f 和召回率 R_f

设 X = 系统正确检出的音频样例数, Y = 系统检出的音频样例总数, Z = 真实的音频样例总数。则有:

$$P_f = \frac{X}{Y}, \tag{9}$$

$$R_f = \frac{X}{Z}, \tag{10}$$

$$F_f = \frac{2P_f R_f}{P_f + R_f}. \tag{11}$$

评价方法 II: 检索倍速 xRT

$$xRT = \frac{\text{被检索的音频长度}}{\text{每个音频样例的平均检索时间}}. \tag{12}$$

2.4 实验结果及分析

根据检索算法和参数实验, 实验中自相似权重的计算长度取 5, 段向量模值取值范围为 [160,320], 段

向量均值的取值范围 [8.5,190], 段向量均值的相似度阈值 ϵ_1 取值为 0.001。

对大规模音频样例进行检索时, 音频样例的时长和规模不同会对检索结果有很大影响, 因此, 本文分别对不同的音频样例时长和音频样例规模进行实验, 具体如下: 音频样例时长取值范围为 4 s 到 30 s, 步长为 4 s; 音频样例的规模取值范围为 100 到 11000, 步长为 500。

(1) 搜索命中数目测试

实验首先测试了采用单哈希索引和双哈希索引对不同时长的音频样例进行搜索时, 每个音频样例通过哈希表搜索到的结果命中数目。对比分析如图 11 所示, 横轴为音频样例时长, 纵轴为每个音频样例检索的统计平均数。可见, 通过双哈希表搜索的方法, 其搜索范围得到了极大的减小, 对 4 s 到 30 s 时长的音频样例, 检索到的目标数据的数量减少 6 到 8 倍, 并且随着音频样例时长呈缓慢增长状态, 表明该方法可以有效的减少搜索范围, 提高检索速度。

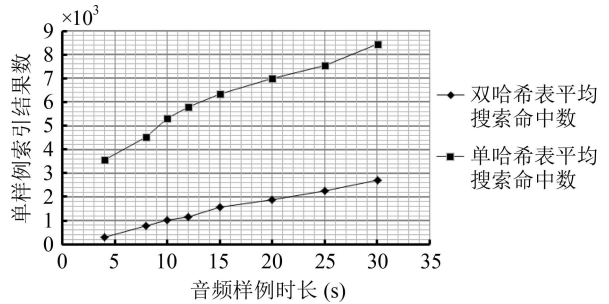


图 11 哈希索引搜索结果命中数对比图

(2) 音频样例数和时长对检索准确性的影响测试

为了测试相同的音频样例规模下, 音频样例的时长对检索准确性和检索速度的影响, 截取不同长度的音频样例进行检索, 检索方法中的段向量的参数对应参数实验中图 4 以及图 6 的取值。使用 F_f 和 xRT 来评价系统检索准确性和检索速度。

当音频样例中段向量的段移取值一定时, 音频样例的检索速度与待检索目标数目、音频流中音频段向量的段移有关, 当音频样例时长增加一倍时, 其待检索目标数目也会增加一倍, 从而增加了检索范围, 降低检索速度; 另一方面, 当音频样例时长增加时, 根据前述的参数选择实验, 增加音频流中段向量的段移可以加快检索速度, 由图 12 所示, 在音频样例时长较短时, 音频流中段向量的段移起到了决定作用, 使检索的速度快速上升; 随着音频样例时长增加, 同一近邻搜索结果越多, 将会导致检索速度越来越慢。

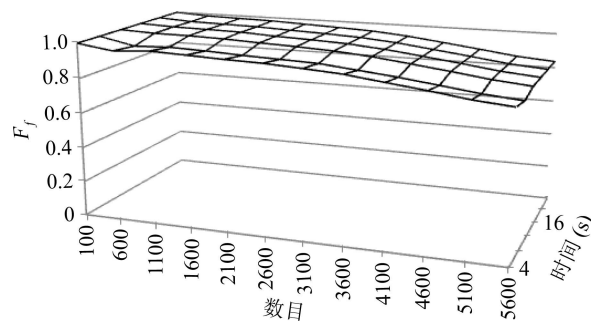


图 12 音频样例数目和时长对检索准确性影响

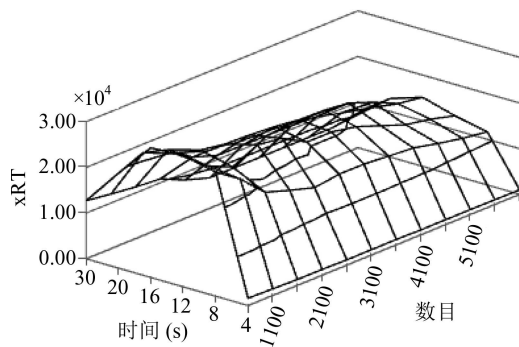


图 13 大规模音频样例时长和数目对检索速度的影响

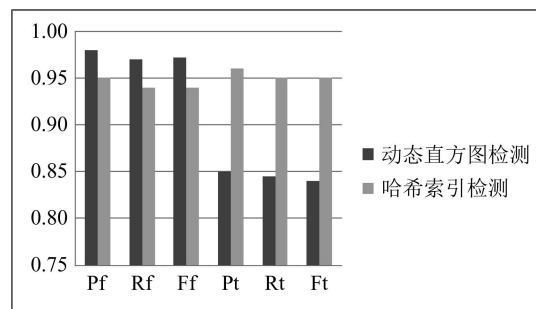


图 14 大规模音频样例时长和数目对检索速度的影响

(3) 音频样例数目和时长对检索速度的影响测试

为了测试不同的音频样例规模和不同时长时系统的检索性能,分别截取不同时长、不同样例规模的音频样例进行检索实验。

实验结果如图 13 所示,在音频样例规模较小时速度较快,当音频样例的数目为 100 时,检索速度 xRT 达 16000,随着音频样例数目增加,检索速度逐渐减小。另一方面,在保证检索的准确性的基础上,音频样例的时长为 16 s 时,检索的速度较高。这是由于当音频样例的时长较短时,检索时设置的音频流段移较短,对音频流的检索次数多,检索速度根据检索时音频流的段移将会呈倍数下降;当音频的时长较长时,为了保证检索的准确性,对音频流检索时的段移并没有随着时长的增加而相应增加,从而造成检索速度的下降。

(4) 测试本文方法与动态直方图检测方法的性能

本文检测方法中的阈值参数、音频样例的时长、规模分别使用实验 1、实验 2、实验 3 得到的优化取值;动态直方图检测方法是利用动态直方图进行相似性搜索的方法对比测试,哈希索引搜索和动态直方图检测方法的检测准确性如图 14 所示,哈希索引的检测倍速是 16000,动态直方图检测的检测倍速是 79。

3 结束语

提出一种基于哈希索引的大规模音频样例检索

方法,实现了大规模音频样例快速精确检索,不需重复检索。并且该方法在保证较高的检索精度的同时,提高了音频样例检索速度。本文通过实验对阈值参数取值、样例时长、样例规模对检测准确性及检测速度的影响进行了分析,实验表明在保证一定准确率的基础上,音频样例的时长、音频样例规模是影响检索速度的关键因素。下一步工作可以针对音频特征筛选、参数优化等方向展开。

参 考 文 献

- 1 Yu Y, Joe K, Downie J. S. Efficient query-by-content audio retrieval by locality sensitive hashing and partial sequence comparison. *Transactions on Information and Systems*, 2008; **E91D**(6): 1730—1739
- 2 Ryyanen M, Klapuri A. Query by humming of midi and audio using locality sensitive hashing. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Las Vegas: IEEE*, 2008: 2249—2252
- 3 Baluja S, Covell M, Ioffe S. Permutation grouping: Intelligent hash function design for audio image retrieval. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Las Vegas: IEEE*, 2008: 2137—2140
- 4 Zheng G, Li M. A fast audio retrieval method based on negativity judgment. *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Piscataway: IH-MSP*, 2009: 1156—1159
- 5 Yu Y, Crucianu M. Local summarization and multi-level LSH for retrieving multi-variant audio tracks. *17th ACM International Conference on Multimedia, MM'09, with Co-located Workshops and Symposia, Beijing: ACM*, 2009: 341—350

- 6 Cotton C, Ellis D P W. Finding similar acoustic events using matching pursuit and locality-sensitive hashing. 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz: IEEE, 2009: 125—128
- 7 唐杰. 基于内容的音频检索技术研究. 北京邮电大学, 2010
- 8 Guo Zhiyuan, Wang Qiang, Yin Liang *et al.* Query by humming via hierarchical filters. Tsukuba, 2012: 3021—3024
- 9 Pedraza C, Vitola J, Sepulveda J *et al.* Fast content-based audio retrieval algorithm. Bogota, 2013: 1—5
- 10 McFee B, Barrington L, Lanckriet G. Learning Content Similarity for Music Recommendation. *Audio, Speech, and Language Processing*, 2012; **8**(20): 2207—2218
- 11 齐晓倩, 陈鸿昶, 黄海. 基于 K-L 距离的两步固定音频检索方法. 计算机工程, 2011; **37**(19): 160—162
- 12 郑贵滨. 基于分段的实时声频检索方法. 声学学报, 2006; **31**(2): 101—108
- 13 Knight A, Almeroth K. Fast caption alignment for automatic indexing of audio. *International Journal of Multimedia. A Data Engineering & Amp.*, 2010; **1**(2): 1—17
- 14 Sturm B L, Daudet L. On similarity search in audio signals using adaptive sparse approximations. Spain: Madrid, 2011: 59—71
- 15 Cont A, Dubnov S, Assayag G. On the information geometry of audio streams with applications to similarity computing. *IEEE Transactions on Audio, Speech and Language Processing*, 2011; **19**(4): 837—846
- 16 Ding X, Yang J. Movie audio retrieval based on HCT. *Mechatronics and Industrial Informatics, PTS 1-4: Applied Mechanics and Materials*, 2013: 321—324, 1129—1132

声学消息

李倩倩等同志荣获《声学学报》2014 年度优秀论文奖

根据声学学报优秀论文评选办法, 经过声学学报编委会两轮评选, 声学学报编委会一致同意, 决定授予李倩倩, 李整林, 张仁和等同志的“不确定海洋环境下的贝叶斯声源定位方法”(声学学报, 2014 年第 39 卷第 5 期 535—543 页) 一文为声学学报 2014 年度优秀论文奖。

编委会认为: 李倩倩、李整林、张仁和等同志的“不确定海洋环境下的贝叶斯声源定位方法”(声学学报, 2014 年第 39 卷第 5 期 535—543 页) 一文及其相关研究工作, 针对水下声源匹配场定位对海洋环境依赖性问题的研究, 研究并提出了海洋环境聚焦的声源定位方法, 把海洋环境参数与声源位置同步反演, 得到声源位置的最优估计值, 并利用海底声学参数经验公式降低了反演维数, 改善了匹配场定位的环境宽容性, 有效提高了声源定位精度。实验结果证明本方法对水下声源匹配场定位的环境适应性改进具有重要应用价值。

声学学报优秀论文的评选于 1998 年 12 月 18 日召开的声学学报(京区)编委会上提出, 并于 1999 年 2 月召开的主编会议上确立。1999 年进行了第一次评选, 自今已进行了 16 届。旨在促进声学领域基础研究和应用研究的发展, 鼓励作者开展创新、系统、深入的研究工作, 写出高质量论文。

《声学学报》2014 年度优秀论文奖于 2015 年 10 月 17 日在西安召开的中国声学学会第十一届青年学术会议上由声学学报(中、英文版)主编王小民先生颁奖, 山东科技大学青年教师李倩倩博士代表所有获奖作者领奖。

(中国声学学会声学学报编辑委员会)

