

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2015 年 3 月 5 日 (05.03.2015)



(10) 国际公布号
WO 2015/027751 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2014/081338
- (22) 国际申请日: 2014 年 7 月 1 日 (01.07.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201310378000.0 2013 年 8 月 27 日 (27.08.2013) CN
- (71) 申请人: 复旦大学 (FUDAN UNIVERSITY) [CN/CN];
中国上海市杨浦区邯郸路 220 号, Shanghai 200433 (CN)。
- (72) 发明人: 俞鹏飞 (YU, Pengfei); 中国上海市张衡路 825 号, Shanghai 201203 (CN)。 杨夙 (YANG, Su);
中国上海市张衡路 825 号, Shanghai 201203 (CN)。
- (74) 代理人: 上海专利商标事务所有限公司 (SHANGHAI PATENT & TRADEMARK LAW OFFICE, LLC); 中国上海市桂平路 435 号, Shanghai 200233 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。
- 本国际公布:
— 包括国际检索报告(条约第 21 条(3))。

(54) Title: AUDIO FINGERPRINT FEATURE-BASED MUSIC RETRIEVAL SYSTEM

(54) 发明名称: 基于音频指纹特征的音乐检索系统

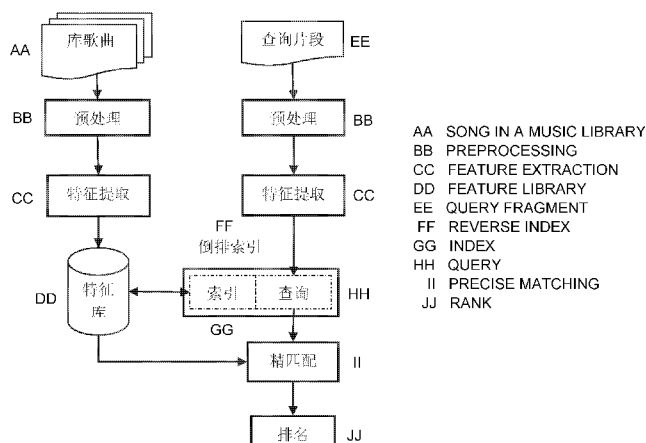


图 1 / FIG. 1

(57) Abstract: An audio fingerprint feature-based music retrieval system, comprising a preprocessing module, a feature extraction module, a reverse index module and a precise matching module. The preprocessing module primarily conducts conversion, resampling, and filtering of audio signals; the feature extraction module is a representation of an audio file employing audio fingerprint features and selecting the most stable points in the frequency spectrum as feature points via double screening based on a dynamic threshold, and using a pair of points to represent a feature; the reverse index module uses the features as keywords, and establishes a reverse index by the features in a music library, and returns an index result according to the number of the same keywords; the precise matching module combines the timing relationships between audio features and employs improved editing distance as the degree of similarity between two feature sequences, so as to optimize the index result.

(57) 摘要:

[见续页]

一种基于音频指纹特征的音乐检索系统。该系统由预处理模块、特征提取模块、倒排索引模块和精匹配模块四部分组成。预处理模块主要完成音频信号的转换、重采样、滤波；特征提取模块是对音频文件的表示，采用音频指纹特征，通过两次基于动态阈值的筛选来选取频谱中最为稳定的点作为特征点，用一个点对表示一个特征；倒排索引模块是将特征作为关键词，由歌曲库的特征建立倒排索引，根据相同关键词多少返回索引结果；精匹配模块是结合音频特征间的时序关系，采用改进的编辑距离作为两个特征序列的相似度，以此优化索引结果。

基于音频指纹特征的音乐检索系统

技术领域

本发明属于信息检索技术领域，具体涉及音频信号处理和多媒体信息检索系统，进一步涉及一种基于音频指纹特征的音乐检索系统。

背景技术

早前，由于音乐信息是非结构化数据，其检索系统一般通过基于文本描述来实现检索。例如在互联网中检索一首歌曲，以歌曲的名字、歌唱者、作歌者、歌曲存取格式等来检索。该方法具有很多的缺点：数据量越来越大，从而人工注释工作量也随之加大；音频感知难以用文字注释表达清楚；信息描述具有一定的主观性。而基于内容的音乐检索系统是根据信息本身的特征参数而非外部属性对内容进行检索，其提取过程由程序自动完成。因此，其不存在对信息描述的主观性，能更好的表示音乐信息，从而使检索更加有效。

音频特征是音频信息的结构化表示，是基于内容的音频检索系统中较为关键的一步，音频特征的好坏直接影响系统性能。音频特征分为时域音频特征和变换域音频特征，时域音频特征较为容易提取，但抗噪能力较差；变换域特征提取过程较复杂，抗噪能力良好，使用较多。在变换域特征中，较为常见的是基于傅里叶和小波变换的特征。早期，Mel 倒谱系数(MFCC)和线性预测倒谱系数(LPCC)特征较为经典，特别是 MFCC，由于其特征是根据人的听觉模型生成的，应用较为广泛。随着音频指纹特征的出现，由于其特征鲁棒性较好，使得大量学者投入在这方面进行研究，发展较快。

目前，基于音频指纹特征的音乐检索系统以Shazam公司和Philips公司的音乐检索系统较为经典。Shazam公司的音乐检索系统是在频谱上选取局部极值点作为特征点，然后把相邻的两个特征点组成一个点对来表示一个特征；采用hash索引实现检索；查询时，使用直方图统计相同特征点的时间差，一般目标歌曲的时间差是统一的，将会集中出现在某处，从而检索到该歌曲。该系统查询方式并不适用海量音频检索，而且局部极值点非常多，导致特征数据非常多，很多特征抗噪能力差。在Philips公司的音乐检索系统中，特征是在频谱上计

算各频段能量，根据相邻帧的能量大小，将各频段量化成 0 或 1，所有频段组成一个二进制序列，通过哈希(hash)函数，得到最终特征。采用哈希表实现检索，通过统计词频数来进行检索排序。在海量音乐数据下，hash冲突会非常多，也不适用，同时在特征性能上不如前一种指纹特征。

发明内容

本发明的目的在于提供一种基于音频指纹特征的音乐检索系统，该系统能够在海量音乐信息下进行快速准确的检索，且能够对录音查询片段进行有效检索。

本发明提供的基于音频指纹特征的音乐检索系统，包括预处理模块，特征提取模块，倒排索引模块和精匹配模块四个部分。其中：

所述的预处理模块，用于音频文件格式统一，音频重采样和音频滤波。

所述的特征提取模块，用于对音乐文件的结构化表示，采用基于动态阈值的音乐指纹特征。首先对歌曲序列进行分帧，帧之间有较高重叠率，对每帧进行快速傅里叶变换（FFT），处理完所有帧，得到频谱矩阵；接着，对频谱矩阵进行平滑处理；然后，在矩阵中选取极值点，并根据动态阈值对这些点进行两次筛选，取大于阈值的点作为特征点；最后，用一个点对来表示一个特征，并经哈希（Hash）函数变换，一个哈希值即为一个特征。对于每个特征点，在其后续频段的邻近区域内，选取最多 P 个最近邻的特征点与该特征点一一组成特征，所有特征按帧的先后顺序和特征点的筛选顺序组成一个特征序列。

所述的倒排索引模块，用于系统的初次检索，以一个特征作为一个关键词，以歌曲库的所有特征建立倒排索引表；当查询时，通过倒排索引表统计查询片段每个关键词在各歌曲中出现的次数，并将所有关键词在各个歌曲中出现的次数求和，然后对求和的结果进行排序，排序结果所对应的歌曲作为初次检索结果。为了防止查询片段较短或者较为偏僻，在倒排索引中并未加入权重，即各词项在各歌曲中权重是一样的。但这样可能降低目标歌曲与其他歌曲的区分度，对于较长的歌曲，需要对歌曲进行分段，以歌曲片段为单位加入倒排索引表中。

所述的精匹配模块，用于系统的二次检索，先根据倒排索引返回的结果选定候选歌曲，接着读取各候选歌曲的特征序列，并对特征序列按查询特征序列

长度进行分段，对每首歌曲筛选出最为相似的 Q 个特征序列片段（与查询特征序列具有最多的相同特征个数），然后，对这 Q 个片段与查询特征序列进行改进的编辑距离计算（特征值只错一位认为是相同的），取最小的编辑距离作为该歌曲片段与查询片段的相似度，最后，根据相似度对候选歌曲进行排序，得到最终的检索排名，作为系统检索结果。如果某首歌在排名中出现多条记录，只保留第一条记录。

本发明的优点为：系统所用的特征鲁棒性好且数据量小；采用现阶段较为成熟的倒排索引技术作为系统初始检索，可适用于海量音乐检索；精匹配作为系统二次检索，能有效的找出目标歌曲且可以并行计算。

附图说明

图 1 为本发明系统结构示意图。

图 2 为本发明系统音乐指纹特征提取流程图。

图 3 为本发明系统特征表示示意图。

图 4 为本发明系统倒排索引结构示意图。

图 5 为本发明系统精匹配流程图。

具体实施方式

图 1 显示了系统结构，包括预处理模块，特征提取模块，倒排索引模块和精匹配模块四个部分。预处理模块主要完成音频信号的转换、重采样和滤波；特征提取模块是对音频文件的结构化表示，采用基于两次阈值筛选的音频指纹特征；倒排索引模块是根据歌曲库的特征建立倒排索引，当查询时，通过倒排索引统计各歌曲片段与查询片段相同关键词个数多少，并对个数和进行排序，作为初次检索结果；精匹配模块在初次检索的基础上，结合音频特征间的时序关系，采用改进的编辑距离作为两个特征序列的相似度，优化索引结果。对于数据库中每首歌，通过预处理和特征提取，将特征保存于特征库中，用于建倒排索引和精匹配；对于查询片段，做相同的预处理和特征提取，其特征用于查询和精匹配。

所述的特征提取模块，其特征提取过程如图 2 所示，采用基于两次阈值筛选的音频指纹特征。首先，对音频序列 $X=\{x_1, x_2, \dots, x_L\}$ 进行分帧，L 为音

频序列长度，帧之间有较高的重叠率，共分成 M 帧；接着，对每一帧进行 N 点快速傅里叶变换，即取 N 个频段点，处理完所有帧后，得到 N*M 维的频谱矩阵 S，并对频谱矩阵 $S=[S_{i,j} \mid i=1,2,\dots,N; j=1,2,\dots,M]$ 进行平滑处理，平滑计算公式如下：

$$S_{i,j} = \log_{10}(\max(\text{abs}(S_{i,j}), e^{-5})) \quad i = 1, 2, \dots, N, j = 1, 2, \dots, M \quad (1)$$

其中 $\text{abs}()$ 为取模运算，M 由音频长度决定，N 可取 129（傅里叶变换取 256 个点，由于对称性，取一半），然后，在 S 中选取极大值点，即 $S_{i,j} > S_{i,j-1}$ 且 $S_{i,j} > S_{i,j+1}$ ，作为特征点，并根据阈值对特征点进行两次筛选；用 N 维向量 thresh 表示频谱中各频段的阈值，在 S 矩阵中，取前 R 帧各频段的最大值来初始化对应频段的阈值，一般 R 取 10；初次筛选：顺序扫描所有特征点，若该点值大于对应维度的阈值，则保留该特征点，否则删除该特征点，同时按以下公式更新阈值向量 thresh ：

$$\text{thresh}_k = \begin{cases} \text{thresh}_k * 0.97 & s_{i,j} \leq \text{thresh}_i \\ \max(\text{thresh}_k, \text{thresh}_i * e^{\frac{-(k-i)^2}{1000}}) & \text{其它} \end{cases}, k = 1, 2, \dots, N \quad (2)$$

第二次筛选：从最后一个特征点开始，逆序扫描所有特征点，按相同规则筛选特征点和更新阈值；最后，用一个点对来表示一个特征，对于每一个特征点，用它与其邻近区域的每个特征点组成一个特征；当邻近区域内特征点较多时，选取与它最相邻的 P 个点与该特征点一一组成特征。一般，P 取 3~5 的整数，P 越大，包含信息越多，但特征数量明显增多。用这些点对的时间，频率来表示特征，并进行了哈希（Hash）函数计算，具体见图 3。按第一次筛选顺序逐个表示这些特征点，处理完所有帧得到一维特征序列。

图 3 显示了特征的表示，点 A(t, f) 为要表示的特征点，矩形为它的邻近区域 $[t+1: t+T, f-F/2: f+F/2]$ ，区域中最相邻的 P 个特征点与点 A 组成 P 个特征，在程序中，P 取 3，T 取 32，F 取 64，如图中三个箭头。由于使用一个查询片段来进行检索，我们用第一个点的频率 F1，第二点与第一个点的频率差 ΔF 及它们的时间差 Δt 来表示一个特征。为了便于后续检索，我们对特征进行了哈希函数运算，公式如下。

$$\text{Feature} = F1 \ll (fbits + tbits) + \Delta F \ll tbits + \Delta t \quad (3)$$

其中， $\ll$ 为向左移位运算，用 fbits 位表示频率差 ΔF ，tbits 位表示时间差 Δt ，Feature 为特征值。当 fbits 或者 tbits 较小时，还需要进行取模运算，

在程序中，tbits 取 6，fbits 取 8。一个哈希值便是一个音乐指纹特征，一般用 2-3 个字节来表示。

所述的倒排索引模块由两部分组成，如图 4 所示，左边部分叫做字典，即由词项组成，是一系列字符串的集合，字典在索引中通常是以字典序列存储，系统中，所有哈希值相同的特征组成一个词项；右边部分是包含某个字符串的文档编号的集合，称之为“倒排链表”，每一个词项都对应一个属于自己的“倒排链表”，该表记录了包含该词项的歌曲编号或者歌曲片段编号。当查询时，通过倒排索引表统计各歌曲片段与查询片段相同关键词个数多少，然后计算个数和（对于查询片段中出现的重复关键词进行累加计算），并按个数和进行排序，作为倒排索引的结果。考虑到查询片段的特征很可能是目标歌曲的偏僻特征，在链表中并未加入权重，也就是各词项在各歌曲中具有相同的权重。然而这样大大降低了目标歌曲与其他歌曲的区分度，通常需要对长歌曲进行分段，以歌曲片段为单位建立索引，能有效的提高它们之间的区分度。

所述的精匹配模块，采用多个步骤实现精匹配，其过程如图 5 所示，首先，根据初次检索返回结果，寻找一“拐点”，假定倒排索引表返回的第 i 首歌曲中具有相同特征个数之和为 num_i ，如果存在一点 K ，使得：

$$\sum_{i=1}^K num_i > \sum_{i=K+1}^{3*K} num_i, \quad 3 * K \leq N \quad (4)$$

则认为该点为“拐点”，目标歌曲就在这前 K 个候选歌曲片段中；接着，读取前 K 个候选歌曲片段的特征序列，对这些序列进行分段，找出最为相似的 Q 个片段，它们与查询序列具有最多的相同特征个数，一般， Q 取 3~6， Q 越小，计算改进的编辑距离次数越少，速度越快，但有可能无法包含目标片段，对于一般查询片段， Q 取 3；然后，将这 Q 个片段与查询特征序列进行改进的编辑距离计算，把最小距离的片段作为与查询序列最相似的片段，并取最小距离作为与该候选歌曲片段的相似度。设查询特征序列 $A=\{A[1], A[2], \dots, A[m]\}$ ，比较的特征序列 $B=\{B[1], B[2], \dots, B[n]\}$ ，长度分别为 m 和 n ，距离矩阵 $d=\{d[i, j]=0 \mid i=1, 2, \dots, m; j=1, 2, \dots, n\}$ ， $d[i, j]$ 为子序列 $A[1 \dots i]$ 和 $B[1 \dots j]$ 的距离，改进的编辑距离算法步骤如下：

- (1) 初始化距离矩阵 d ，读入特征序列 A 和 B ；

(2) 循环遍历特征序列 A, 逐次取数 $A[i]$, 依次执行操作步骤 (3)、(4)、(5) ;

(3) 循环遍历特征序列 B, 逐次取数 $B[j]$, 依次执行操作步骤 (4)、(5) ;

(4) 计算代价 cost, 如果数 $A[i]$ 与数 $B[j]$ 相等或只有 1 位 (bit) 不同, cost 为 0, 否则为 1, 如公式:

$$\text{cost} = \min((A[i] \wedge B[j]) \& ((A[i] \wedge B[j]) - 1), 1) \quad (5)$$

其中, $\wedge$ 为位异或运算, $\&$ 为位与运算;

(5) 调整距离矩阵, 计算出当前最小距离 $d[i, j]$, 公式如下:

$$d[i, j] = \min(d[i-1, j]+1, d[i, j-1]+1, d[i-1, j-1]+\text{cost}) \quad (6)$$

(6) $d[m, n]$ 即为改进的编辑距离。

最后, 我们根据相似度进行排序, 得最终的检索排名, 如果某首歌在排名中出现多条记录 (较长歌曲建立倒排索引时分段), 只保留第一条记录。

权 利 要 求

1. 一种基于音频指纹特征的音乐检索系统，其特征在于包括预处理模块，特征提取模块，倒排索引模块和精匹配模块四个部分；其中：

所述的预处理模块，用于音频文件格式统一，音频重采样和音频滤波；

所述的特征提取模块，用于对音乐文件的结构化表示，采用基于动态阈值的音乐指纹特征；首先对歌曲序列进行分帧，对每帧进行快速傅里叶变换，处理完所有帧，得到频谱矩阵；接着，对频谱矩阵进行平滑处理；然后，在矩阵中选取极值点，并根据动态阈值对这些点进行两次筛选，取大于阈值的点作为特征点；最后，用一个点对表示一个特征，并经哈希函数变换，输出一个哈希值作为一个特征；对于每个特征点，在其后续频段的邻近区域内，选取最多 P 个最近邻的特征点与该特征点一一组成特征；所有特征按帧的先后顺序和首次特征点筛选顺序组成一维特征序列；

所述的倒排索引模块，用于系统的初次检索，以一个特征作为一个关键词，对数据库中的每首歌曲的特征建立倒排索引表；当查询时，通过倒排索引表统计查询片段每个关键词在各歌曲中出现的次数，并将所有关键词在各个歌曲中出现的次数求和，然后对求和的结果进行排序，排序结果所对应的歌曲作为初次检索结果；

所述的精匹配模块，用于系统的二次检索，先根据初次检索返回的结果选定候选歌曲，接着读取各候选歌曲的特征序列，并对特征序列按查询特征序列长度进行分段，对每首歌曲筛选出最为相似的 Q 个特征序列片段，即其与查询特征序列具有最多的相同特征个数；然后，对这 Q 个片段与查询特征序列进行改进的编辑距离计算，取最小的编辑距离作为该歌曲片段与查询片段的相似度；最后，根据相似度对候选歌曲进行排序，得到最终的检索排名。

2. 根据权利要求 1 所述的音乐检索系统，其特征在于：在特征提取模块中，所述的采用基于动态阈值的音乐指纹特征，具体实现过程为：首先，对音频序列 $X=\{x_1, x_2, \dots, x_L\}$ 进行分帧，L 为音频序列长度，帧之间有较高的重叠率，共分成 M 帧；接着，对每一帧进行 N 点快速傅里叶变换，处理完所有帧后，得到 $N \times M$ 维的频谱矩阵 S，并对频谱矩阵 $S=[S_{i,j} \mid i=1, 2, \dots, N; j=1, 2, \dots, M]$

进行平滑处理，平滑计算公式如下：

$$S_{i,j} = \log_{10}(\max(abs(S_{i,j}), e^{-5})) \quad i = 1, 2, \dots, N, j = 1, 2, \dots, M \quad (1)$$

其中 $abs()$ 为取模运算，然后，在 S 中选取极大值点，即 $S_{i,j} > S_{i,j-1}$ 且 $S_{i,j} > S_{i,j+1}$ ，作为特征点，并根据阈值对特征点进行两次筛选；用 N 维向量 $thresh$ 表示频谱中各频段的阈值，在 S 矩阵中，取开始 R 帧各频段的最大值来初始化对应频段的阈值；初次筛选：顺序扫描所有特征点，若该点值大于对应维度的阈值，则保留该特征点，否则删除该特征点，同时按以下公式更新阈值向量 $thresh$ ：

$$thresh_k = \begin{cases} thresh_k * 0.97 & s_{i,j} \leq thresh_i \\ \max(thresh_k, thresh_i * e^{\frac{-(k-i)^2}{1000}}) & \text{其它} \end{cases}, k = 1, 2, \dots, N \quad (2)$$

第二次筛选：从最后一个特征点开始，逆序扫描所有保留的特征点，按与上述相同规则筛选特征点和更新阈值；最后，用一个点对来表示一个特征，对于每一个特征点，用它分别与其后续频段的邻近区域的每个特征点组成一个特征；当邻近区域内特征点较多时，选取与它最相邻的 P 个点与该特征点一一组成特征，并按第一次筛选顺序逐个表示这些特征点，处理完所有帧得到一维特征序列。

3. 根据权利要求 1 所述的音乐检索系统，其特征在于：所述的倒排索引模块由两部分组成，一部分为字典，字典由词项组成，所有哈希值相同的特征组成一个词项；另一部分是倒排索引表，其中，每一个词项都对应一个属于自己的“倒排链表”，该表记录了包含该词项的歌曲编号或者歌曲片段编号。

4. 根据权利要求 1 所述的音乐检索系统，其特征在于：所述的精匹配模块，采用多个步骤实现精匹配，首先，根据初次检索返回的结果，寻找一“拐点”，假定倒排索引表返回的第 i 首歌曲中与查询片段具有的相同特征个数之和为 num_i ，如果存在一点 K ，使得：

$$\sum_{i=1}^K num_i > \sum_{i=K+1}^{3*K} num_i, \quad 3 * K \leq N \quad (4)$$

则认为该点为“拐点”，目标歌曲就在这前 K 个候选歌曲片段中；接着，读取前 K 个候选歌曲片段的特征序列，对这些序列进行分段，找出最为相似的

Q 个片段，它们与查询序列具有最多的相同特征个数；然后，将这 Q 个片段与查询特征序列进行改进的编辑距离计算，把最小距离的片段作为与查询序列最相似的片段，并取最小距离作为与该候选歌曲片段的相似度；设查询特征序列 $A=\{A[1], A[2], \dots, A[m]\}$ ，比较的特征序列 $B=\{B[1], B[2], \dots, B[n]\}$ ，长度分别为 m 和 n，距离矩阵 $d=\{d[i, j]=0 \mid i=1, 2, \dots, m; j=1, 2, \dots, n\}$ ， $d[i, j]$ 为子序列 $A[1 \dots i]$ 和 $B[1 \dots j]$ 的距离，改进的编辑距离算法步骤如下：

(1) 初始化距离矩阵 d，读入特征序列 A 和 B；

(2) 循环遍历特征序列 A，逐次取数 $A[i]$ ，依次执行操作步骤 (3)、(4)、(5)；

(3) 循环遍历特征序列 B，逐次取数 $B[j]$ ，依次执行操作步骤 (4)、(5)；

(4) 计算代价 cost，如果数 $A[i]$ 与数 $B[j]$ 相等或只有 1 位 (bit) 不同，cost 为 0，否则为 1，如公式：

$$\text{cost} = \min((A[i] \wedge B[j]) \& ((A[i] \wedge B[j]) - 1), 1) \quad (5)$$

其中， $\wedge$ 为位异或运算， $\&$ 为位与运算；

(5) 调整距离矩阵，计算出当前最小距离 $d[i, j]$ ，公式如下：

$$d[i, j] = \min(d[i-1, j]+1, d[i, j-1]+1, d[i-1, j-1]+\text{cost}) \quad (6)$$

(6) $d[m, n]$ 即为改进的编辑距离。

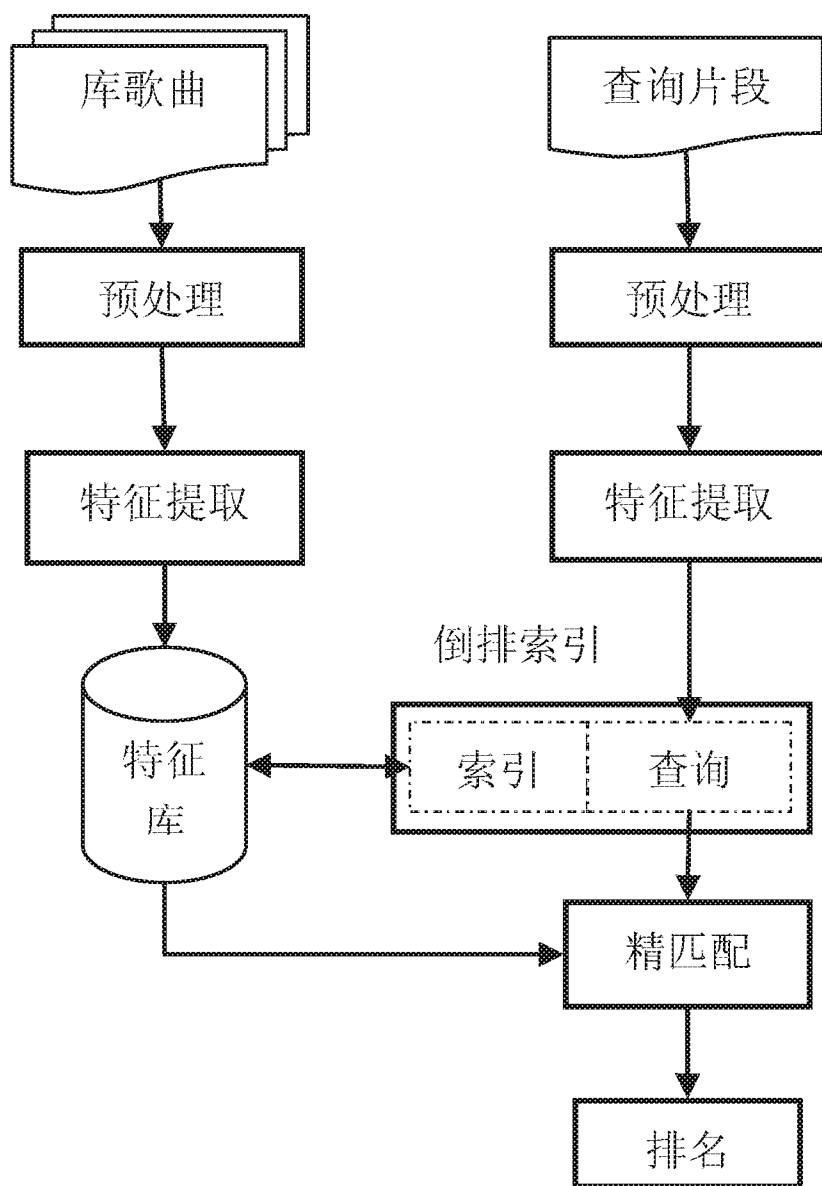


图 1

2/4

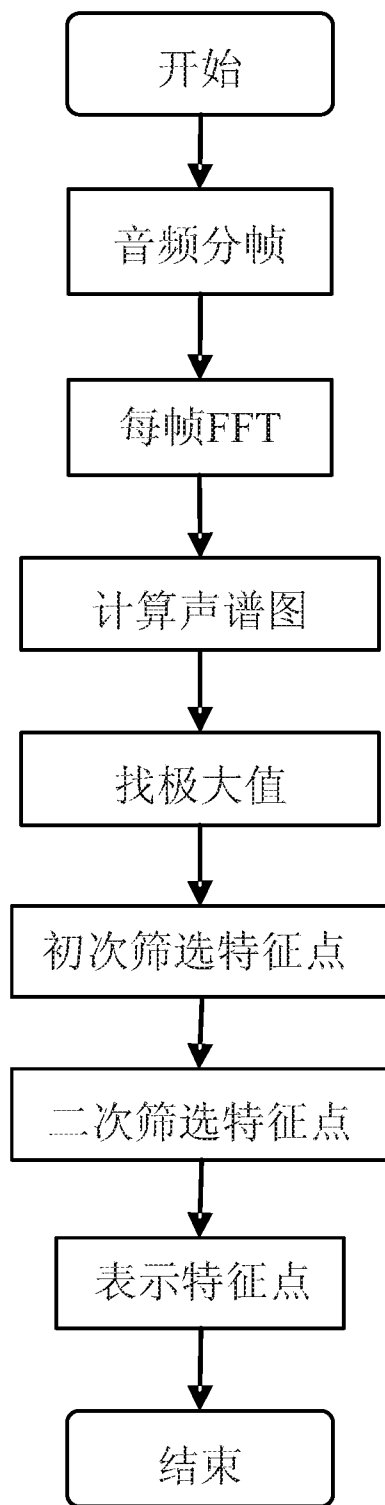


图 2

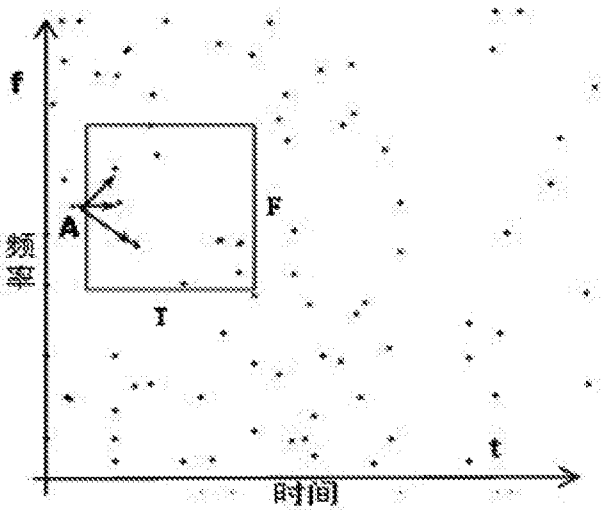


图 3

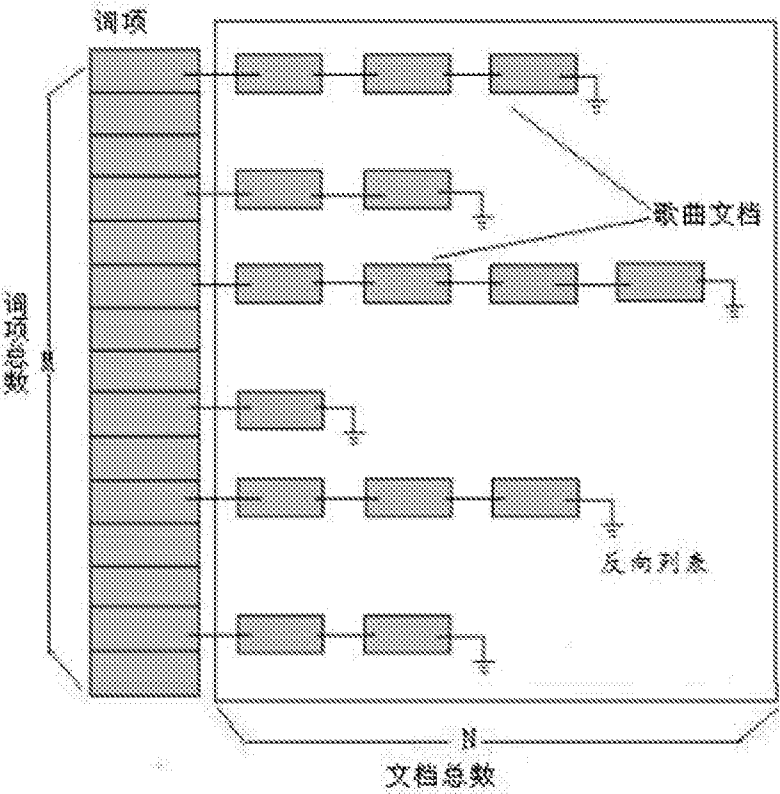


图 4

4/4

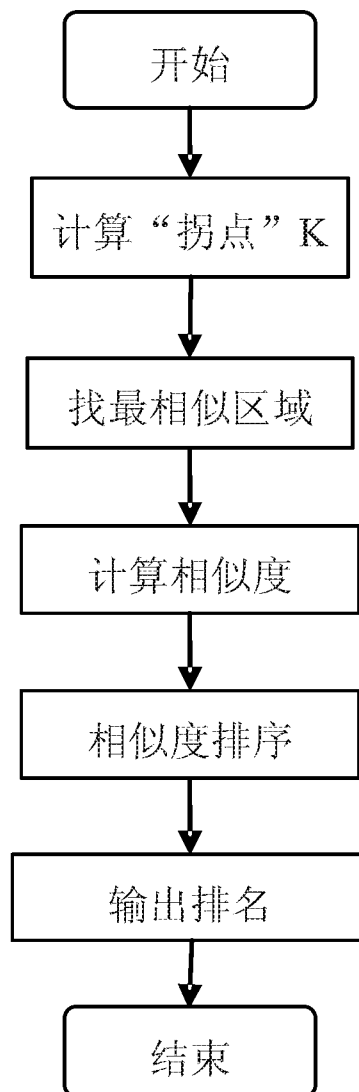


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/081338

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: music+, audio+, voice?, sound?, speech??. multimedia, media+, video, fingerprint+, retriev+, search+, revers+, invert+, invers+, index+, hash+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 103440313 A (FUDAN UNIVERSITY) 11 December 2013 (11.12.2013) claims 1 to 4	1-4
Y	Tang Jie, et al. "Improved Algorithms of Music Information Retrieval based on Audio Fingerprint." 2009 Third International Symposium on Intelligent Information Technology Application Workshops. 22 November 2009 (22.11.2009), ISSN: pages 367 to 371	1-3
Y	Pedro Cano, et al. "A Review of Algorithms for Audio Fingerprinting." 2002 IEEE Workshop on Multimedia Signal Processing., 11 December 2002 (11.12.2002), ISSN: pages 169 to 173	1-3

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 08 August 2014	Date of mailing of the international search report 19 September 2014
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer KANG, Dandan Telephone No. (86-10) 62413593

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/081338

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GUAN, Yaohua et al. "Improved Algorithm of Data Preprocessing in Audio Fingerprint Search." Computer Engineering and Applications, vol. 46, no. 21, 31 December 2010 (31.12.2010), ISSN: pages 145 to 147 and 170	1-3
A	CN 103093761 A (TENCENT TECH SHENZHEN CO., LTD) 08 May 2013 (08.05.2013) the whole document	1-4
A	US 2006149552 A1 (AEC ONE STOP GROUP INC.) 06 July 2006 (06.07.2006) the whole document	1-4
A	US 8492633 B2 (THE ECHO NEST CORPORATION) 23 July 2013 (23.07.2013) the whole document	1-4
A	US 2011173208 A1 (ROVI TECHNOLOGIES CORPORATION) 14 July 2011 (14.07.2011) the whole document	1-4

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/081338

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103440313 A	11 December 2013	None	
CN 103093761 A	08 May 2013	None	
US 2006149552 A1	06 July 2006	US 8352259 B2	08 January 2013
		WO 2006073802 A2	13 July 2006
		US 7567899 B2	28 July 2009
		US 2009259690 A1	15 October 2009
US 8492633 B2	23 July 2013	US 2013139674 A1	06 June 2013
		US 2013139673 A1	06 June 2013
		US 8586847 B2	19 November 2013
US 2011173208 A1	14 July 2011	WO 2011087757 A1	21 July 2011

A. 主题的分类 G06F 17/30(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06F 17/- 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNPAT, WPI, EPODOC, CNKI:音乐, 音频, 语音, 声音, 话音, 多媒体, 媒体, 视频, 指纹, 检索, 搜索, 倒排, 反向, 反转, 反排, 逆排, 索引, 哈希, music+, audio+, speech?+, voice?, sound?, media+, video, fingerprint+, search+, retriev+, revers+, invert+, invers+, index+, hash+		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 103440313 A (复旦大学) 2013年 12月 11日 (2013 - 12 - 11) 权利要求1-4	1-4
Y	Tang Jie, et al. "Improved Algorithms of Music Information Retrieval based on Audio Fingerprint." 2009 Third International Symposium on Intelligent Information Technology Application Workshops., 第卷, 第期, 2009年 11月 22日 (2009 - 11 - 22), ISSN: , 第367-371页	1-3
Y	Pedro Cano, et al. "A Review of Algorithms for Audio Fingerprinting." 2002 IEEE Workshop on Multimedia Signal Processing., 第卷, 第期, 2002年 12月 11日 (2002 - 12 - 11), ISSN: , 第169-173页	1-3
☒ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件		
国际检索实际完成的日期 2014年 8月 08日		国际检索报告邮寄日期 2014年 9月 19日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国 传真号 (86-10)62019451		授权官员 康丹丹 电话号码 (86-10)62413593

C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	关耀铎等. "音频指纹搜索中数据预处理的改进算法." Computer Engineering and Applications计算机工程与应用., 第第46卷, 卷, 第第21期, 期, 2010年 12月 31日 (2010 - 12 - 31), ISSN: , 第145-147, 170页	1-3
A	CN 103093761 A (腾讯科技深圳有限公司) 2013年 5月 08日 (2013 - 05 - 08) 全文	1-4
A	US 2006149552 A1 (AEC ONE STOP GROUP INC.) 2006年 7月 06日 (2006 - 07 - 06) 全文	1-4
A	US 8492633 B2 (THE ECHO NEST CORPORATION) 2013年 7月 23日 (2013 - 07 - 23) 全文	1-4
A	US 2011173208 A1 (ROVI TECHNOLOGIES CORPORATION) 2011年 7月 14日 (2011 - 07 - 14) 全文	1-4

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/081338

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103440313	A	2013年 12月 11日	无			
CN	103093761	A	2013年 5月 08日	无			
US	2006149552	A1	2006年 7月 06日	US	8352259	B2	2013年 1月 08日
				WO	2006073802	A2	2006年 7月 13日
				US	7567899	B2	2009年 7月 28日
				US	2009259690	A1	2009年 10月 15日
US	8492633	B2	2013年 7月 23日	US	2013139674	A1	2013年 6月 06日
				US	2013139673	A1	2013年 6月 06日
				US	8586847	B2	2013年 11月 19日
US	2011173208	A1	2011年 7月 14日	WO	2011087757	A1	2011年 7月 21日

表 PCT/ISA/210 (同族专利附件) (2009年7月)