

# 产生音频数据索引的有效方法

李 应, 侯义斌

(西安交通大学电脑与信息技术研究所, 陕西西安 710049)

**摘 要:** 支持基于例子检索的音频多媒体数据库, 应当能够自动分析音频数据而产生音频数据的索引。本文把基于离散小波变换和基于时频分析的索引方法相结合, 提出一种利用小波变换、小波子带的过零率、小波子带幅度的均值、一定比例的小波子带的高频系数和全部近似系数等来组成索引向量的方法。提出的方法采用双重索引搜索, 即在用过零率向量和平均幅度向量作为索引进行预先扫描的基础上, 再用小波变换的关键系数向量作为索引进行详细扫描, 实验表明, 这种方法具有较高的检索精度和搜索效率, 在基于例子索引和搜索的音频多媒体数据库中具有重要意义。

**关键词:** 小波变换; 时频分析; 关键系数; 过零率; 平均幅度

**中图分类号:** TP311.134.3; TP912.3 **文献标识码:** A **文章编号:** 0372-2112 (2002) 11-1613-04

## The Efficient Method for Producing Audio Data Index

LI Ying, HOU Yi-bing

(Computer and Information Technology Institute, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

**Abstract:** To support the retrieval by example, audio multimedia database ought to be capable of automatically analyzing audio data for producing the index of the audio data. We combined the index method based on discrete wavelet transform with the method based on time-frequency analysis in this paper. And a method was proposed that utilize the wavelet subband, a certain proportional high-frequency coefficient of the wavelet subband and all the approximately coefficient etc. to form the index vector. The proposed method introduced double index search, that is using zero-cross vector and average magnitude vector at first pre-search and then using key coefficient of the wavelet transform as index to make more detailed search. The experiment shows that this method has a high retrieval precision and search efficiency, and has importance sense in the audio multimedia database based on example index and search.

**Key words:** wavelet transform; time-frequency analysis; key coefficient; zero cross rate; average magnitude

### 1 引言

与图象和视频等其他多媒体数据一样, 音频数据需要基于例子的索引和近似搜索。支持基于内容和例子检索的音频多媒体数据库, 应当能够自动分析音频数据来产生音频数据的索引。文献[1]描述了基于内容的音频的分类、搜索和回取系统。它主要用统计技术来分类、分析声音, 并得出由均值、变量、自相关和延续等组成的特征向量, 并用特征向量作为基于内容分类和声音检索的唯一信息。在文献[2]中, 描述了通过例子查询的搜索和检索系统, 但它只支持音乐数据和查询一些较规则的震动声音。对于可以对任何类型的音频数据进行分析并产生索引的方法, 文献[3, 4]提出了通过短时 Fourier 变换和小波变换产生索引的方法。用小波变换的关键系数作为音频数据搜索的索引, 这种方法在文献[3]中取得了较好的检索精度。作者在文献[4]中提出了利用短时快速离散余弦变换(FDCT)产生音频数据索引的方法, 也取得了一定的效果。由

于这两种方法都需要抽取大量的系数作为索引, 因此, 它对于大型的音频多媒体数据库并不实用。在文献[5]中提出了通过小波变换, 计算各子带的标准误差、过零率和均值来产生索引向量。而在文献[6]中采用线性预测编码对连续通用音频数据的进行分类, 在文献[7, 8]中主要通过抽取音频数据的常用特征: 短时能量函数、短时平均过零率、短时基音周期和频谱峰轨迹来实现对音频数据的分类。虽然, 文献[5~7]中的方法对于音频数据的有效分类具有一定的实用价值, 但它们不能对各种音频数据进行有效的搜索。

由于上述各种方法的局限性, 一定程度地影响了音频多媒体数据库的应用。本文把文献[3, 4]中的方法与文献[5, 7]中用到的基于音频信号时频分析方法相结合, 提出一种利用小波变换、小波子带的过零率、小波子带的均值、一定比例的小波子带的高频系数和全部近似系数的综合作为索引的方法。这种方法使得基于例子的音频数据查询的搜索效率和检索精度得以较大提高。

### 2 离散小波变换

#### 2.1 小波分析

在连续的情况,小波序列为<sup>[9]</sup>

$$a_{j,k}(t) = |a|^{-1/2} \left( \frac{t-b}{a} \right),$$

$$b \in R, a \in R_+, \text{且 } a > 0.$$

其中  $a$  为伸缩因子,  $b$  为平移因子. 这里,  $a$  的允许条件为

$$C = \int_0^1 \frac{|\hat{\psi}(\xi)|}{|\xi|} d\xi < \infty$$

通常,把连续小波变换中尺度参数  $a$  和平移参数  $b$  进行离散化,并分别取作  $a = a_0^j, b = ka_0^j b_0, j \in Z$ . 如果采用二进制的动态网格,即  $a_0 = 2, b_0 = 1$ ,则对于离散的情况,小波序列为  $\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), j, k \in Z$

对于任意的函数  $f(t) \in L^2(R)$  ( $L^2(R)$  表示平方可积的实数空间,即能量有限的信号空间)的连续小波变换为

$$W_f(a, b) = \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt$$

而离散化小波变换系数则可表示为

$$C_{j,k} = \int_{-\infty}^{\infty} f(t) \overline{\psi_{j,k}(t)} dt$$

在小波变换中,它们的时频窗口为两个矩形

$$[b-a, b+a] \times [(\pm 0 - \frac{1}{2})/a, (\pm 0 + \frac{1}{2})/a]$$

这两个窗口中心为  $(b, \pm 0/a)$ , 时窗和频窗宽分别为  $a$

和  $1/a$ . 其中  $b$  仅仅影响窗口在相平面时间轴上的位置,而  $a$  不仅影响窗口在频率轴上的位置,也影响窗口的形状. 这样小波变换对不同的频率在时域上的取样步长是调节性的,即在低频时小波的时间分辨率较差,而频率分辨率较高;在高频时小波变换的时间分辨率较高,而频率分辨率较低,这正符合低频信号变化缓慢而高频信号变化迅速的特点. 本文正是通过对这些特点的综合来产生音频数据的索引,并取得较好的效果.

#### 2.2 音频信号的多尺度小波分解

在基于小波索引的方法中,把音频数据进行多尺度小波分解. 其中输出的分解结构包含小波分解系数向量  $C$  和系数长度向量  $L$ . 分解向量  $C$  是由近似和详细的系数组成的系数集,它的结构按如下组织:

$$C = [\text{app. coef}(N), \text{det. coef}(N), \dots, \text{det. coef}(1)]$$

系数长度向量  $L$  的各分量表示如下:

$$L(1) = \text{lengthof}(\text{app. coef}(N)),$$

$$L(i) = \text{lengthof}(\text{det. coef}(N - j + 2)), j = 2, \dots, N + 1$$

其中,  $\text{app. coef}(N)$  表示近似系数集,  $\text{det. coef}(i)$  表示详细系数集且  $1 \leq i \leq N$ ,  $\text{lengthof}(\text{app. coef}(N))$  近似系数集的长度,  $\text{lengthof}(\text{det. coef}(N - j + 2))$  表示详细系数集的长度且  $2 \leq j \leq N + 1$ ,  $N$  表示分解的最大尺度. 这里取  $N = 6$ , 对某段音频数据进行小波分解,并得到如图 1 所示的分解系数.

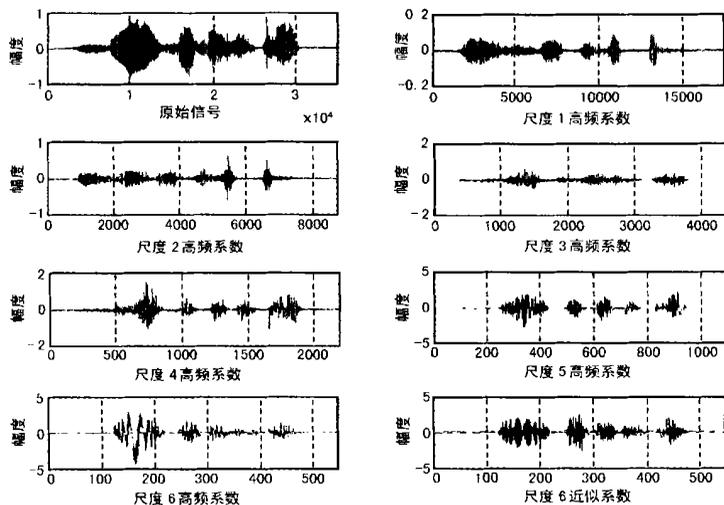


图 1 原始信号和尺度 6 近似系数及尺度 6、5、4、3、2、1 的高频系数

### 3 产生音频数据索引

#### 3.1 基于小波的索引

在图 2 中,显示了从离散小波变换(DWT)系数中得出索引的原理. 在作为索引的关键系数向量中,选择尺度  $N$  的所有近似系数. 在

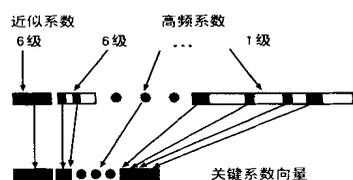


图 2 从离散小波变换中产生索引

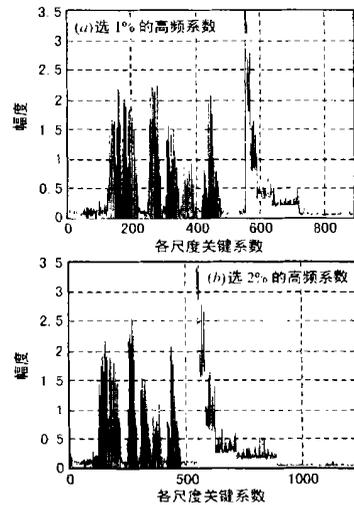


图 3 离散小波变换系数中产生的关键系数

不同尺度的高频系数中,分别选择其中一定比例幅值最大的系数作为关键系数. 从 2.2 所述的音频信号的多尺度小波分解系数  $C$  结构中,通过系数的选择得出  $C'$ ,即关键系数向量. 从上述的反映选择系数的数量和位置的  $L$  中,并得出  $L'$ ,它反映  $C'$  中系数的数量和位置. 它们组成近似和详细系数的数量(和位置),其中,  $C'$  被用于索引,而  $L'$  作为辅助的信息. 如果用  $n'_a$  和  $n'_i$  分别表示近似系数和选出的  $i$  级高频系数的数量,用  $l_1, \dots, l_{n'_i}$  表示在  $L$  中  $i$  级高频系数的位置. 那么,

$$L' = [ n'_a, l_1, \dots, l_{n'_a}, n'_6, l_1, \dots, l_{n'_6}, n'_5, l_1, \dots, l_{n'_5}, \dots, n'_1, l_1, \dots, l_{n'_1} ]$$

对于  $N=6$  (6级分解) 的情况. 图 3(a) 和 (b) 分别表示当选择所有近似系数、在不同尺度的高频系数中取 1% 和 2% 的最高幅值系数的情况 (幅值取绝对值, 为了突出关键问题, 这里忽略去噪声等环节).

### 3.2 基于音频信号的时域处理的索引

利用时域方法来分析的音频信号特征参数包括音频的短时平均幅度和能量、短时过零率、短时自相关函数和短时平均幅度差函数等<sup>[10]</sup>. 音频信号经过多尺度小波分解后产生的各尺度高频及近似系数也提供了这些指标. 这些指标既可以作为单纯的特征, 也可以作为进一步复合分析的基础. 这里, 把其中的小波子带过零率和小波子带幅度均值作为音频信号的时域处理的索引.

小波子带过零率  $Z$ : 对于音频这种前后关系的离散信号, 如果连续采样值有不同的算术符号, 则发生过零. 这意味着在过零和信号的频率之间存在一定的关系. 在每个子带中, 音频信号的小波系数以震动波形的形式出现. 在时域处理中, 过零表示子带信号频率在一个周期中变换的频度. 我们选择小波子带中小波系数过零率作为一个特征. 在小波子带中, 我们定义过零率如下:

$$z_i = \frac{1}{2 * s_i} \sum_{k=0}^{s_i} | \text{sgn}[x(k)] - \text{sgn}[x(k-1)] |$$

式中  $s_i$  为小波子带中系数的数量,  $\text{sgn}[x(n)]$  满足关系

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases}$$

小波子带幅度均值  $M$ : 音频的时频特性中的平均幅度与小波系数的平均幅度相对应. 带有相同频率和响度的音调, 在不同的包迹中, 产生不同的声音. 系数序列的平均幅度, 是一种信号包迹的指标. 因此, 使用的另一个特征向量是小波子带系数  $x(j)$  均值, 定义为

$$m_i = \frac{1}{s_i} \sum_{j=1}^{s_i} | x(j) |$$

对于上述的两个特征: 小波子带过零率子向量  $Z \{ z_1, z_2, \dots, z_n \}$  和小波子带幅度均值子向量  $M \{ m_1, m_2, \dots, m_n \}$  可以通过每个小波子带中的系数来计算.

基于时域的音频数据索引是通过分别计算尺度小波系数的幅度平均值和过零率得到的两个子向量  $Z$  和  $M$ . 对于 6 尺度离散小波分解, 这两个子向量的维数较低, 便于作为音频数

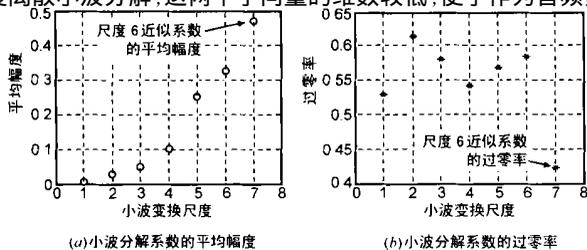


图 4 基于时域的两个索引

据的索引对音频数据库进行高效的搜索. 对于图 1 所示的各尺度小波变换系数, 它们的小波子带幅度均值子向量和小波子带过零率子向量如图 4 中的 (a) 和 (b) 所示.

### 4 音频数据的索引与搜索方法

当音频数据存入音频数据库时, 首先按上述方法产生一组索引向量  $\{ Z, M, C \}$ .

把这一组信息插入到索引结构中作为原始数据的索引, 并保留一个指向原始音频块的指针. 从音频数据中产生索引的大致过程如图 5 所示.

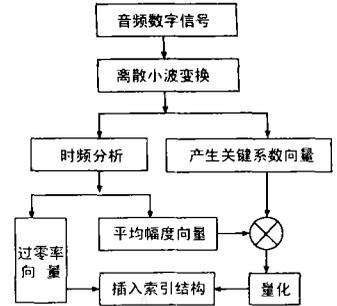


图 5 从音频数据中产生索引

当需要进行基于例子的查询时, 同样把待查询的例子音频进行上述变换后, 产生一组索引向量  $\{ Z, M, C \}$ .

当对数据库中的索引进行搜索时, 先对小波子带幅度均值子向量  $Z$  和小波子带过零率子向量  $M$  进行搜索. 如果查询的子向量用  $q$  表示, 而在数据库音频夹中的子向量用  $x$  表示, 那么, 在查询和音频夹的子向量之间的 Euclidean 距离  $D_t$  定义为

$$D_t = \left( w_z \sum_{i=1}^n (q_z^i - x_z^i)^2 + w_m \sum_{i=1}^n (q_m^i - x_m^i)^2 \right)^{1/2} \quad (1)$$

这里,  $w_z$  和  $w_m$  分别表示每个特征组件的权值. 对于满足这两个子向量的索引, 再用小波变换的关键系数子向量进行搜索. 查询小波变换关键系数子向量  $q$  和数据库音频夹中的小波变换关键系数子向量  $x$  之间的 Euclidean 距离  $D_w$  定义为

$$D_w = \left( w_{a6} \sum_{i=1}^{s_{a6}} (q_{a6}^i - x_{a6}^i)^2 + w_{c6} \sum_{i=1}^{s_{c6}} (q_{c6}^i - x_{c6}^i)^2 + w_{c5} \sum_{i=1}^{s_{c5}} (q_{c5}^i - x_{c5}^i)^2 + \dots + w_{c1} \sum_{i=1}^{s_{c1}} (q_{c1}^i - x_{c1}^i)^2 \right)^{1/2} \quad (2)$$

这里,  $w_{a6}$  和  $w_{c6}, w_{c5}, \dots, w_{c1}$  分别表示不同尺度时的关键系数权值,  $a6$  表示尺度 6 的近似系数,  $c_i$  表示尺度  $i$  的高频系数, 其中  $i=1, 2, \dots, 6$ . 如果两串间的相互距离足够小, 它们被认为是匹配的, 返回搜索的原音频数据的指针, 最后通过指针返回搜索结果.

通过例子查询的搜索方法的大致过程如图 6.

### 5 实验与结果分析

数据文件由语音、音乐和多个其他种类的数据如动物和鸟声、汽笛声等特殊

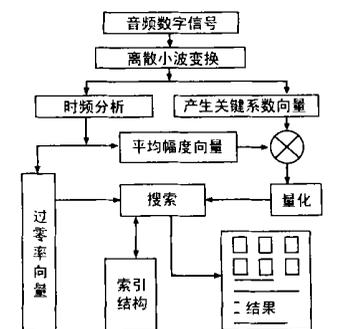


图 6 通过例子查询的搜索

的声音组成. 文件和查询大小范围为 (1 - 5 秒), 总共 200 个音频数据文件. 采用上述基于离散小波变换的关键系数向量、过零率向量和平均幅度向量为索引, 来对音频数据文件进行搜

索. 实验分成四组, 每组采用的小波变换系数中高频系数比例分别为 1%、2%、3% 和 4% (系数中的幅值绝对值最高部分) 和所有的近似系数. 每组重复 10 次, 每次随机抽取 100 个数据文件. 每次实验结束后, 把这些数据放回, 重新随机抽取作为下次实验的 100 个数据文件. 这里我们使用 Daubechies - 4 小波. 匹配时, 采用上述的 Euclidean 距离式 (1) 和 (2) 来作为距离度量指标. 基于离散小波变换的关键系数子向量、过零率和平均幅度子向量的索引方法和实验结果如表 1. 我们使用检索精度作为指标, 其定义如下:

$$\text{检索精度} = \frac{\text{返回的相关对象数}}{\text{返回的对象总数}} \times 100\%$$

从表 1 可以看出, 采用基于离散小波变换和信号时频分析的索引具有较高的检索精度, 且对于采用不同比例的各尺度变换系数, 检索精度相对稳定. 当各尺度的高频成分的比例增加时, 检索精度明显提高.

## 6 结论

本文提出的关于基于离散小波变换和信号时频分析的索引与搜索方法, 使得基于例子的音频数据文件的搜索精度与效率具有较大的提高. 与常用的基于信号统计的音频数据搜索回取方法、基于短时快速离散余弦变换 (FDCT)<sup>[4]</sup> 方法、单纯的离散小波变换<sup>[3]</sup> 和单纯的小波变换或短时 Fourier 变换的时频分析特征<sup>[5,7]</sup> 相比, 由于提出的方法采用双重索引搜索 (首先, 用小波变换系数的过零率子向量和平均幅度子向量作为索引进行搜索, 然后在用小波变换的关键系数子向量作为索引进行搜索), 因此这种方法可以在不影响搜索精度的前提下, 提高了搜索的效率, 便于在音频多媒体数据库中实现基于例子的索引搜索. 对于本文提出的方法, 尚未涉及音频数据进行去噪声处理、对音频数据的有效分段以及索引结构的搜索方法的高效组织. 如果考虑对音频数据进行去噪声处理, 并对音频数据进行有效的分段, 那么这种方法可能更加稳定和可靠, 检索精度也可以进一步提高. 对于精度更高和更加有效的搜索方法, 我们将进一步研究.

## 参考文献:

- [ 1 ] Wold E, Blum T, Keislar D, Wheaton J. Content-based classification, search and retrieval of audio [J]. IEEE Multimedia FAL 1996, 3: (3) 27 - 36.
- [ 2 ] Chias A, Logan J, Chamberlin D, et al. Query by humming: musical information retrieval in an audio database [A]. Scan F. Proceedings of MULTIMEDIA '95 [C]. New York: ACM, 1995. 231 - 236.
- [ 3 ] Subramanya S R. Indexing and searching schemes for audio data in audio/ multimedia databases (multimedia database) [D]. Washington: George Washington Univ., 1999.
- [ 4 ] 李应, 侯义斌. 抽取音频数据特征的快速离散余弦变换方法 [J]. 西安交通大学学报, 2001, 35(8): 817 - 820.
- [ 5 ] Li Guohui, Khokhar A A. Content-Based Indexing and retrieval of audio data using wavelets [A]. Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on [C]. 2000. 885 - 888.
- [ 6 ] Li Dongge, Ishwar K, Sethi, Nevenka D, Tom M. Classification of general audio data for content-based retrieval [J]. Pattern Recognition Letters, 2001, 22: 533 - 544.
- [ 7 ] Zhang Tong, Jay Kuo C C. Audio content analysis for online audiovisual data segmentation and classification [J]. IEEE Trans. on Speech and Audio Processing, 2001, 9(4): 440 - 457.
- [ 8 ] Rao Wang, Zhu Liu, Jincheng Huang. Multimedia content analysis using both audio and visual clues [J]. IEEE Signal Processing Magazine, 2000, 11: 12 - 25.
- [ 9 ] 陈逢时. 子波变换理论及其在信号处理中的应用 [M]. 北京: 国防工业出版社, 1998.
- [ 10 ] 胡光锐. 语音处理与识别 [M]. 上海: 上海科学技术文献出版社, 1994.

## 作者简介:

李 应 男, 1964 年 8 月生于福建闽清, 1993 年在西安交通大学获系统工程专业硕士学位, 现为福州大学计算机与信息科学讲师, 并在西安交通大学攻读博士学位研究生, 主要研究方向为计算机仿真、多媒体数据库及音频数据处理.

侯义斌 男, 1952 年 4 月生于陕西武功, 1982 年在西安交通大学获计算机工程与科学专业硕士学位, 1986 年在 Eindhoven 技术大学获电子工程专业博士学位, 现为西安交通大学计算机和信息技术教授, 并为 Eindhoven 技术大学客座教授, 主要研究方向包括中文信息处理、多媒体处理、用户-系统交互和设计方法学.