

分段式音频检索算法^{*}郑贵滨^{1,2} 韩纪庆¹ 李海峰¹ 郑铁然¹(哈尔滨工业大学 哈尔滨150001)¹ (哈尔滨理工大学 哈尔滨150001)²

摘要 本文提出了一种基于分段的快速音频检索算法。该算法将检索目标划分成多个较小的片段,每个片段可以独立进行检索;检索过程中使用检索窗控制参与检索的片段及数量。该算法的速度不随检索目标的长度变化,检索速度可调,且能获得良好的查全率和查准率,适用于从未知音频数据源中检索任意长度的特定音频数据及实时应用场合。

关键词 音频检索,分段,检索窗

Segmentation-Based Audio Retrieval Algorithm

ZHENG Gui-Bin^{1,2} HAN Ji-Qing¹ LI Hai-Feng¹ ZHENG Tie-Ran¹(Harbin Institute of Technology, Harbin 150001)¹ (Harbin University of Science and Technology, Harbin 150001)²

Abstract In most audio retrieval methods, retrieval target is usually treated as a whole unit, which results in slowing down the retrieval speed linearly due to target length. A novel audio retrieval algorithm based on segmentation is presented in this paper. The retrieve target is divided into a series of segments, each of which can be searched and retrieved independently. In the mean while, a retrieval window is used to decide which segments can be searched simultaneously. This algorithm can achieve a very high retrieval speed that has no relationship with target length and the speed can be controlled by the size of retrieval window. Experimental results show that recall rate and accuracy rate are satisfied.

Keywords Audio retrieval, Segmentation, Retrieval window

1 引言

随着多媒体技术、数据压缩技术和网络技术的不断发展,多媒体信息的数量迅猛增长。为了充分利用已有的多媒体资源,研究多媒体检索技术变得越来越重要。音频检索技术研究如何从大量音频信息中快速准确地检索满足要求的音频数据,是多媒体检索技术的主要方法之一。目前,国内外在音频检索技术方面的研究主要集中在对已知多媒体数据库进行检索,主要研究方法可分四类:(1)综合利用语音识别、话者识别及文本检索等技术实现对语音类数据的检索^[1];(2)针对音乐类数据的检索^[2];(3)基于相似类别查找的多类别音频检索^[3];(4)针对特定音频信息的检索^[4]。

通常,在进行检索时将检索目标作为一个整体直接检索,随着检索目标长度的增加,检索速度呈线性下降。为此,本文提出了一种分段式检索算法,该算法的检索速度稳定而不随检索目标的长度变化,可用于实时同步地从未知音频数据源中快速检索任意长度的音频数据。

2 分段式检索算法

我们将要检索的目标音频数据称为参考模板。分段式检索算法的基本思想如图1所示:将较长的参考模板划分成若干较小的音频片段,从输入数据流中检索片段的出现情况,从而可得到参考模板的检索结果。

2.1 片段划分

将参考模板(用 R 表示)划分成若干片段,片段的长度可

以各自不同,但第一段起始点和最后一段的结束点应和参考模板的头尾对齐。设划分的片段总数为 N ,片段的序号依次记为 $1, 2, \dots, N$ 。用 R_i 表示第 i 个片段,片段长为 $len(i)$,片段 i, j 间的距离记为 $d(i, j)$:

$$d(i, j) = \sum_{k=i}^{j-1} d(k, k+1) \quad i \in [1, N-1], j \in [i+1, N], \\ d(i, i) = 0 \quad (1)$$

在等距等长片段划分中 $len(i) = L, d(j-1, j) = D \quad (i=1, \dots, N, j=2, \dots, N)$ 均为常数,并且 $d(i, j) = |j-i|D$ 。

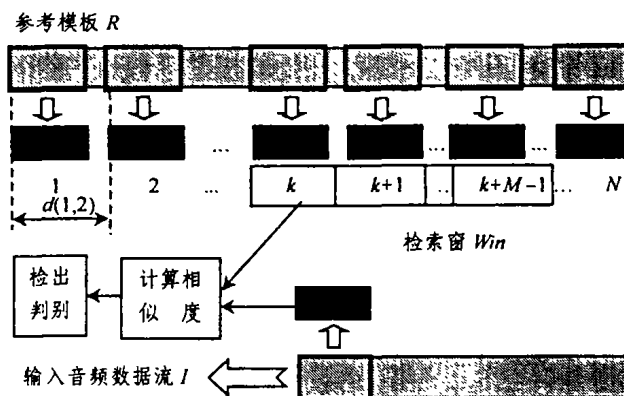


图1 分段式音频检索算法示意图

片段划分时应考虑有利于片段的检出。由于静音帧比例较高的片段在检索时容易误检,因此当参考模板的静音帧比例较小时,可采用等距等长片段划分(如图1所示);否则,可根

^{*}课题获教育部跨世纪优秀人才培养计划基金、哈尔滨市留学回国人员基金(2002AFLXJ005)与哈尔滨工业大学海外留学人员回国工作科研启动基金的资助。郑贵滨 讲师,博士生;韩纪庆 博士,教授,博士生导师。

据合适的片段划分位置采用不等长和不等距片段划分。

2.2 检索窗的更新及长度设定

由于参考模板中的片段具有时序性,在检索时,序号小的片段应先于序号大的片段被检出,因此在检索过程中可设定检索窗,用于限定每一时刻能够参与检索的片段及数量。只有在检索窗中登记的片段才可以参与检索,并且片段在检索窗中从左至右按照片段序号从小到大的顺序登记。设检索窗长度为 M ,即同时可有 M 个片段参与检索,当检索开始尚无片段被检出时,检索窗中记录的片段序号依次是 $1, 2, \dots, M$ 。检索过程中,当有片段被检出后,则根据与输入数据流时间同步的原则调整检索窗内登记的片段。

当有片段被检出后,更新检索窗时,用参考片段序号确定检索窗应登记的片段。设最后一个被检出的片段是 R , 检出时间是 $t_d(i)$, 当前时间是 t_{cur} 。参考片段序号 k 由下式计算:

$$\begin{aligned} k &= \max(k', l) \\ k' &= \max\{j | d(i+1, j) \leq t_{cur} - t_d(i), \\ &\quad j = i+1, \dots, N\} - \lfloor \lambda \cdot M \rfloor \end{aligned} \quad (2)$$

其中: λ 是引入的修正系数,考虑在输入音频流中的检索目标数据有可能缺失片段,以及片段误识造成检索窗错误滑动,一般取 $0 < \lambda \leq 0.5$ 。 $\lfloor \dots \rfloor$ 表示取整。从片段 R_k 开始将尚未检出的 M 个片段的序号依次登记到检索窗中,如图1所示(假设 R_k 后的所有片段均未检出)。

由于本算法只有在一个片段被检出后才能随时间移动检索窗,使后继片段参与检索,完成整个检索过程。因此在使用检索窗的检索过程中,必须保证窗内登记的片段至少有一个能被正确检出。否则,将造成后继片段无法参与检索而导致参考模板漏检。因此检索窗的长度应满足一定的条件。在分段式检索算法中,并不限定片段本身的检索算法,不同的片段检索算法具有不同的片段查全率(设为 HR_{ug})。假设片段的检出是概率事件,每个片段的检出概率均为 HR_{ug} 。检索窗中登记的片段至少有一个被正确检出的概率:

$$HR_{win} = 1 - (1 - HR_{ug})^M$$

如限定 HR_{win} 的最小数值为 HR_{win}^{min} , 根据上式有:

$$M \geq \lceil S_f \cdot \ln(1 - HR_{win}^{min}) / (\ln 1 - HR_{ug}) \rceil \quad (3)$$

其中, S_f 是引入的可靠系数, $S_f > 1 + \lambda / (1 - \lambda)$ 。

检索窗越大越有利于目标数据的检出,但会增加计算负担。在实用中可根据计算机能同时检索的最大片段数和需要同时检索的参考模板个数灵活设定检索窗的长度。

2.3 一次匹配结束及参考模板的检出判别

参考模板有一个片段被检出后,我们称参考模板的一次匹配开始。一次匹配开始后,则每当一个片段(设为片段 R_i) 被检出,根据该片段的检出时间 $t_d(i)$, 以及该片段和最后一个片段的距离 $D(i, N)$, 可按下式计算本次匹配应该结束的时间:

$$T_{finish} = \max(OldT_{finish}, t_d(i) + d(i, N)) \quad (4)$$

其中: $OldT_{finish}$ 表示 R_i 被检出前的匹配结束时间,如 R_i 是第一个被检出的片段,则 $OldT_{finish} = 0$ 。如果参考模板的最后一个片段被检出或当前时间大于 T_{finish} , 则本次匹配结束。设置参考模板的片段检出数阈值如下:

$$\eta = \delta \cdot N + C \quad (5)$$

其中, δ 是片段检出比率,其值反映对检索要求的严格程度,越大则检出样本与检索目标的长度越接近; C 是常数,用以避免在参考模板较短时片段数阈值过低而容易造成的参考模板误检。当检出片段数量不小于 η 时,即认为参考模板被检出。

检出长度 L_{Detect} 由下式计算:

$$L_{Detect} = (t_d(e) - t_d(b) + len(b)) \quad (6)$$

其中: $b, len(b)$ 是第一个被检出的片段序号及其长度; e 是最后一个被检出的片段序号。在检索时,参考模板首尾片段漏检会造成检出长度与输入流中的实际长度之间存在误差。适当减小片段划分的间距有利于减小这种误差。

2.4 分段式检索算法

设要在输入流(用 I 表示)中检索参考模板的片段 R_j , 从最新的输入帧(第 i 号帧)开始向左取长度为 $len(j)$ 的数据段作为输入模板片段,用 I_i 表示。计算输入模板 I_i 与参考模板的片段 R_j 间的相似度 $S(I_i, R_j)$, 判断片段 R_j 是否在输入流中出现。

分段式检索算法主要步骤如下:

- (1) 参考模板片段划分, 设片段总数为 N , 计算并保存片段的特征数据;
- (2) 检索窗初始化: 设定检索窗长度 M , 将参考模板的起始 M 个片段序号放入检索窗中, 即 $Win(h) = h, h = 1, \dots, M$; Win 是检索窗数组, 存放片段序号。
- (3) 输入数据流结束, 转(5), 否则:
 - a) 如果参考模板已有片段被检出, 更新检索窗;
 - b) 读取新的输入片段 I_i , 并令 $h = 1$;
 - c) 计算输入片段 I_i 和检索窗中第 h 个片段 R_j 的相似度 $S(I_i, R_j)$, 并判别片段 R_j 是否出现;
 - d) 如 $h < M$, 则 $h = h + 1$, 转 c); 否则, 转步骤(4);
- (4) 如一次匹配开始, 且满足结束条件:
 - a) 若检出片段数不小于 η , 则参考模板被检出, 输出本次检索结果;
 - b) 结束本次匹配转(2)。
 否则转(3)。
- (5) 检索结束

由于分段检索算法的固有特点, 即每个片段可独立检出, 且检出的片段达到一定的数量后便可认为参考模板被检出。因此, 即使检索源(输入数据流)中输入模板有不规则残缺(缺少部分片段), 该算法也能获得很好的性能, 具有一定的稳健性。

3 实验

实验采用中央电视台一台的节目录音(长13个小时, 16KHz、单声道、16位采样)。计算机配置: Pentium IV 1.4GHz CPU、256M RAM, 编程工具是 Visual C++ 6.0。片段检索算法和整体检索算法均采用 MFCC 特征和欧氏距离。采用等长等距片段划分(L, D 均为1秒)。

图2列出的是: 从录音数据库中检索单个模板时, 整体直接检索算法和分段式检索算法所用时间与参考模板长度的关系。其中, 分段式检索算法所用的检索窗长度分别是: 5、15、30。整体直接检索算法的检索时间与参考模板的长度成正比, 随着参考模板长度的增加, 检索时间将难以控制。而分段检索算法的检索时间仅和检索窗的长度有关(成正比), 参考模板的长度越大, 算法的速度优越性越明显。

在实验中, 整体直接检索算法的查全率和查准率均可达100%。分段式检索算法的检索性能如图3所示。当检索窗的大小增大时查全率也随之上升, 检索窗大小达到6以上时, 查全率可达100%, 同时查准率略有下降, 但仍在99.2%以上。可见分段式检索算法在提高检索速度的同时, 也能确保获得很高

的查全率和查准率。

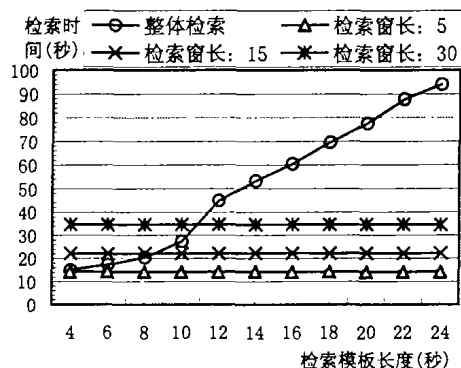


图2 分段式检索与整体直接检索的速度比较

结论 分段式检索算法的速度快,并且可通过检索窗的长度来调节检索速度,并能获得较高的查全率和查准率(可达或接近100%);片段的检索可以采用各种合适的算法,如果片段检索算法采用动态规划技术等会进一步提高检索速度。综合考虑检索速度、查全率和查准率,分段检索算法优于整体直接检索算法,适用于从未知音频数据源中检索任意长度的指定音频数据以及实时应用场合。我们将进一步研究分段检索方法在实时数据检索中的应用问题。

参考文献

1 Makhoul J, Kubala F, et al. Speech and Language Technologies for

(上接第61页)

(1)动态服务质量保证机制 动态 QoS 保证机制是一种基于资源预留的服从综合服务模型的方法。资源预留请求规定了一个预约请求范围,网络实体通过对此请求范围进行检测,从而可以灵活地提供服务。预约请求范围从能接受的最小服务级别到最大服务质量等级,各种网络实体在此范围内根据网络资源状况进行动态的自适应调整,从而提供一种在动态环境下灵活的 QoS 保障机制。

(2)具有 QoS 能力的中间知配机制 即采用带有中间适配件的网路框架来动态地适应网络性能的变化,这种方法考虑了网络的性能和端到端的资源状况,可以向应用提供有用的信息用于重构,从而使整个系统获得最优的 QoS 保障等级。

(3)支持 QoS 的 MAC 协议 其目标是在移动 Ad hoc 网络中使共享介质的各个节点能在尽可能不影响其他节点的前提下实现自身的 QoS 要求。其中,组分配多址接入协议 GAMA 是一种用于提供 QoS 保证的 MAC 协议,协议中,一个竞争阶段通过使用 RTS 和 CTS 对话,为随后的无竞争阶段预留带宽;多址接入/分组预留协议 MACA/PA 可以在无冲突阶段发送一个确认消息来通知相邻的节点是否需要预留带宽,每个节点可以基于分组中携带的预约请求来获得信道的状态。

(4)QoS 路由 该机制首先通过带有 QoS 信息的路由,实施有效的允许控制来防止网络过载,然后寻找满足 QoS 要求的路由在无线网络中实施负载均衡。这里可以基于各种现有的路由算法来构造 QoS 路由协议,每个节点在路由表中增加相应的 QoS 信息,计算最短路径的同时,计算各种 QoS 信

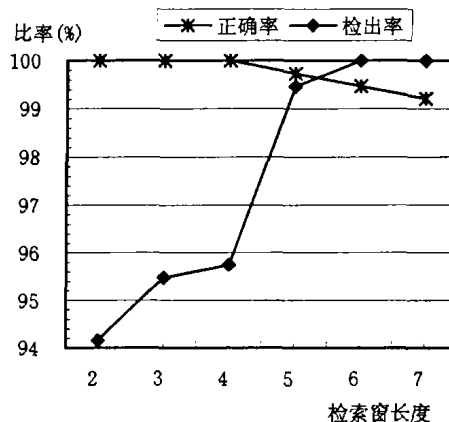


图3 分段式检索算法的性能曲线

Audio Indexing and Retrieval. In: Proc. of the IEEE, 2000, 88(8): 1338~1353

- 2 Tseng H Y. Content-Based Retrieval for Music Collections. In: Proc. of 4th ACM Conf. on Digital Libraries, 1999. 176~182
- 3 Liu Mingchun, Wan Chunru. A Study on Content-Based Classification and Retrieval of Audio Database. In: Intl. Symposium on Database Engineering & Applications, 2001. 339~345
- 4 Smith G, Murase H, Kashino K. Quick audio retrieval using active search. In: Proc. Int. Conf. Acoustics, Speech, Signal Processing, 1998. 3777~3780

息,每个节点根据 QoS 信息来决定是否接受新的连接请求。然而,计算满足多个约束条件的路由是一个 NP 完全问题,因此需要借助启发式算法来计算一个满足多目标的次优化路由。

结束语 在高等院校发展无线自组织网络的意义是不言而喻的,伴随着“移动办公”的需求逐渐强烈,有线接入心有余而力不足,由于投资巨大,学校不可能将每个教室布设有固定电缆及通信设施,随着笔记本电脑及无线网卡等无线通信设施成本的下降,及笔记本电脑的普及,拥有笔记本电脑的教师和学生可以经常组成自组织网络,例如:临时办公会议,教师上课,集体备课,辅导答疑等都要求经常组成自组织网络。综上,我们提出了基于无线自组织网络的 CAES 应用系统的研究。这是具有高新技术及其产品的开发研究工作。市场占有率大,具有高科技含量和高技术附加值的技术或网络产品。

参考文献

- 1 肖永康,山秀明,任勇.无线 Ad hoc 网络及其研究难点.电信科学,2002(6)
- 2 王海涛,郑少仕.自组网的路由协议及 QoS 保障.现代电信科技,2001(2)
- 3 英春,史美林.自组网体系结构研究.通信学报,1999(9)
- 4 倪鹏云.计算机网络系统结构分析.国防大学出版社,1999
- 5 Chunhung R L, Jainshing L. QoS Routing in Ad Hoc Wireless Networks. IEEE Journal on selected areas in communication. 1999, 17(8): 1426~1438
- 6 胡凯,宋京民,等.网络教育新技术.科学出版社,2001
- 7 冯永新.基于移动 Ad hoc 网络重构技术的研究:[东北大学博士论文].2003