

文章编号:1001-9081(2010)01-0230-03

一种基于随机段的固定音频检索方法

杨继臣, 王伟凝

(华南理工大学 电子与信息学院, 广州 510640)

(NisonYoung@yahoo.cn)

摘要:在固定音频检索的整体检索方法中,当检索目标较长时,检索时间会变得很长。为了减小检索时间,提出了一种基于随机段的音频检索方法。把整个检索过程分成随机段检索和整体匹配两个阶段:随机段检索是从参考模板中随机选择一段(随机段)作为检索目标进行检索;整体匹配是在随机段检索出的基础上,判断潜在目标信号是否为参考模板。把这种随机检索的方法应用到计算特征距离和直方图交集方法中,结果证明该检索方法的准确率可以达到90%以上,而且平均检索时间可以降低到随机段与参考模板的比值和整体检索时间的积。

关键词:直方图交集法;特征距离;过零率

中图分类号: TN912.3 **文献标志码:** A

Method of specific audio retrieval based on randomly segment

YANG Ji-chen, WANG Wei-ning

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou Guangdong 510640, China)

Abstract: This paper proposed a specific audio retrieval method based on random segment in order to decrease the retrieval time for relatively long object in the total retrieval of audio retrieval. The whole retrieval process was composed of random segment retrieval and total matching: the first was to select a segment from template model as object to retrieve in stored signal and the second was to judge whether the potential object signal was the template model based on random segment. Then using this method in computing feature distance directly and histogram intersection retrieval, the experimental results show retrieval accuracy over 90% and average retrieval time declined to the ratio of random segment to template model multiplying total retrieval time.

Key words: histogram intersection; feature distance; Zero Crossing Rate (ZCR)

0 引言

随着现代信息技术、多媒体技术和网络技术的发展,多媒体信息的数据量急剧地增多,如何在海量的多媒体库中找到感兴趣或有用的信息,为人们在娱乐、商业、教育等方面更好地服务是一个研究热点。在多媒体检索中,音频检索是一个受人们关注的富有挑战性的研究课题^[1],相对于文本检索和图像检索,音频检索发展比较缓慢。

目前音频检索可以分为两大类:一类是基于内容的,它主要是利用高层信息对音频进行分类和识别,例如音频分类、音频索引、关键词检索等^[2-3];另一类是基于特征相似度的(或称为基于模板的),又称为固定音频检索。它是指给定一个查询音频段(模板),在待检音频库(或音频流)中检索与其同源的片段^[4-5]。这种检索思想最早由文献[6]提出的,后来文献[4-5, 7-10]对这种检索思想进行了进一步的发展。

固定音频检索方法主要有基于距离的方法、基于直方图的方法^[6-7]和二者相结合的算法^[9]。这类检索大多将检索目标作为一个整体进行直接检索,所以本文就把此类检索方法称为整体检索。整体检索存在以下问题:计算代价随着检索目标长度的增加呈线性增长,检索时间随着检索目标长度的增加而增加;当检索目标比较长时,计算代价比较大和需要的检索时间比较长。文献[8]虽然提出了将整个检索目标分成若干较小的片段,但是每个片段都要作为一个小目标独立检

索出来,总的检索时间上仍然和整体检索差不多,需要的检索时间并没有明显减少。

针对整体检索存在的上述问题,文中提出了一种解决方法:基于随机段的音频检索方法。当检索目标比较长时,首先在参考模板(检索目标)中随机选择一段(随机段)作为检索目标进行检索;当随机段检出后,按照随机段在参考模板中对应的位置关系,再判断参考模板与潜在目标信号是否匹配;如果匹配,即潜在目标信号与参考模板相似度达到设定的阈值,就认为找到了目标信号;否则,就可以认为这个潜在目标信号不是目标信号。实验结果表明,该检索方法速度快,准确率可以达到90%以上。该方法不仅可以在计算特征距离的检索中使用,还可以应用在直方图交集的检索中,可以快速地从未知数据中检索出多个任意长度的音频数据。

1 基于随机检索的检索方法

1.1 随机检索的检索思想

本文将检索目标称为参考模板,将检索源信号称为存储信号,目标是在存储信号中找到检索目标,文中的检索方法由两阶段构成:首先在参考模板中随机选择一段(随机段)作为检索目标,可以选在参考模板的任何地方,在存储信号中也选取与随机段信号同样长的滑动窗信号并以一定的步长滑动,通过计算特征距离或直方图交集值来计算二者的相似度,直到找到随机段在存储信号中相对应的信号为止;找到随机段

收稿日期:2009-06-23;修回日期:2009-08-14。 **基金项目:**国家自然科学基金资助项目(60972132;60602014)。

作者简介:杨继臣(1980-),男,安徽界首人,博士研究生,主要研究方向:语音信号处理;王伟凝(1975-),女,江西南昌人,副教授,博士,主要研究方向:信号处理、模式识别。

的匹配对象后,以此为定位点,在存储信号中确定与参考模板相同长度的信号段,然后计算模板信号与潜在目标信号的相似度,判断二者是否匹配,从而确定在存储信号中是否找到检索目标。由于本方法的关键是首先在存储信号中找到与随机段对应的音频信号段,所以本文就把这种检索方法称为随机检索法,以区别于整体检索法。在存储信号中找到随机段的匹配信号可以通过计算特征距离或直方图交集值两种方法得到。

1.2 基于计算特征距离的随机段的检索方法

如图1所示,首先在参考模板中随机选择一段(随机段 R)作为检索目标,然后在存储信号中找一个同样长度的滑动窗(W)。对于参考模板而言,随机信号段的位置有三种:开头、中间和结尾。因为在存储信号中要检索的整个检索目标和模板信号是整体对应的,为了避免漏检,所以要保证随机段的信号与滑动窗在时间上的对应。若随机段在开头,那么滑动窗也要在存储信号的开头;若随机段在中间或结尾,那么滑动窗与存储信号开始位置的距离也要与随机选段与模板信号开始位置的距离保证一样,然后再以一定的步长滑动此滑动窗。

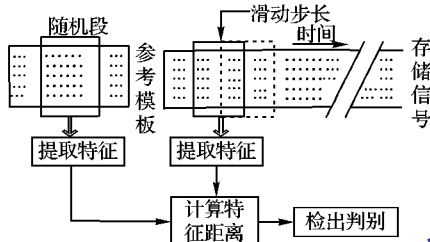


图1 直接计算特征距离的检索方法

在这一步中,提取的语音信号特征是13阶的MECC(Mel Frequency Cepstrum Coefficient)。使用的相似度测量可以选择使用余弦距离或AHS(Arithmetic Harmonic Sphericity)^[11]距离。

$$D(R, W)_{\cos} = \frac{\sum_{i=1}^q R_i \cdot w_i}{\sqrt{\sum_{i=1}^q R_i^2 \times \sum_{i=1}^q w_i^2}} \quad (1)$$

式(1)中: R_i 与 w_i 分别是 R 和 W 中的特征值; q 是特征值总数。

$$D(R, W)_{\text{AHS}} = \log[\text{tr}(\Sigma_R \Sigma_W^{-1}) \times \text{tr}(\Sigma_W \Sigma_R^{-1})] - 2\log(d) \quad (2)$$

式(2)中, tr 表示矩阵的迹; Σ_R 和 Σ_W 分别是 R 和 W 中的特征值的协方差; Σ_R^{-1} 和 Σ_W^{-1} 分别是它们的逆; d 是特征值的维数。当二者的距离达到设定的阈值时,就被认为找到了随机段所对应的检索目标。

1.3 基于直方图的随机段的检索方法

直方图在音频检索中获得了广泛的应用^[6-7],如图2所示。和图1不同的是,提取的音频特征是过零率(Zero Crossing Rate, ZCR)。

分别对随机段和滑动窗提取特征后,将其归一化,再分别生成相应的特征直方图。

随机段:

$$h^R = (h_1^R, h_1^R, \dots, h_q^R)$$

滑动窗:

$$h^W = (h_1^W, h_1^W, \dots, h_q^W)$$

其中 q 表示桶或直方条数。

随机段和滑动窗直方图的相似度可以通过计算二者的直方图交集值得到,直方图交集值通过直方图交集法^[6-7]计算,其公式为:

$$S(h^R, h^W) = \sum_{i=1}^q \min(h_i^R, h_i^W) \quad (3)$$

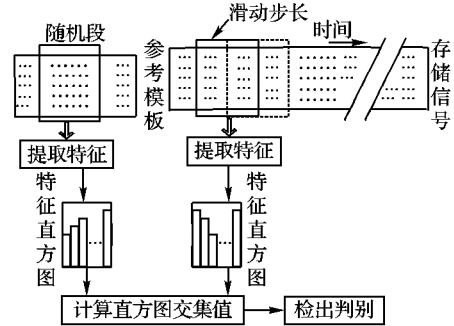


图2 直方图检索法

1.4 整体匹配

无论使用上面的哪一种方法,当在存储信号中找到随机段信号所对应的滑动窗信号后,就要判断整体目标是否匹配,如图3所示。因为音频信号具有时序性,所以选择的目标信号就要保证和参考模板在时间上是对应的,这就要保证滑动窗到潜在目标信号左右端点的距离与随机段到参考模板左右端点的距离一样,然后再使用上面三种方法中的任何一种方法来计算二者的相似度,当相似度达到设定的阈值时,就可以认为整体匹配。

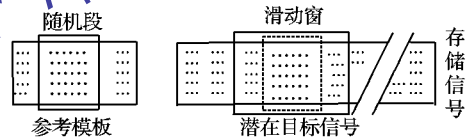


图3 参考模板与潜在目标信号的匹配

2 实验结果及分析

本实验数据文件有三个:第一个选自汉语普通自然口语对话语料库(CADCC)里中的长度为498 s的对话;第二个取自电视连续剧《Friends》中的长度为608 s的伴音;第三个是取自北京奥运会开幕式(OCBO)中的长度为888 s的伴音。以上三段数据格式都是单声道,采样率为16 kHz,量化精度为16 bit。首先把每个文件都去除静音,然后分成帧长是32 ms,帧移是16 ms的帧序列。滑动窗滑动时,滑动步长取0.1 s,参考模板取自实验数据文件,本文中,无论是随机段检索还是整体匹配阶段,使用直方图交集值法和COS距离法的阈值是1,而使用AHS距离法阈值是0。由于音频文件短时包含的信息比较少,在实验中发现,随机段长取大于等于1 s的效果比较好。本文用查全率(Recall Rate, RR)、查准率(Precision Rate, PR)和时间(T_r 代表采用整体检索法所用时间; T_R 代表采用随机段法检索所用时间)三个指标来评价本文方法的检索性能,它们的定义如下:

$$RR = \frac{\text{Number}_{\text{正确检出的目标数}}}{\text{Number}_{\text{检索源中目标总数}}} \times 100\%$$

$$PR = \frac{\text{Number}_{\text{正确检出的目标数}}}{\text{Number}_{\text{检索出的目标总数}}} \times 100\%$$

另外实验采用的机器配置为Pentium 4、2.80 GHz CPU、512 MB内存。表1是对上述三个音频文件进行检索得到的结果。

表1 三个实验文件检索结果

实验文件	采用方法	参考模板/s	T_T /s	随机段长/s	T_R /s	RR/%	PR/%
CADCC	AHS 距离法	10	5 800	1	486	90.91	95.24
				2	1 013	95.24	100.00
				3	2 186	100.00	100.00
Friends	AHS 距离法	20	13 295	1	488	93.75	93.75
				2	1 015	96.77	100.00
				3	2 188	100.00	100.00
OCBO	直方交集法	20	3 066	1	153	93.02	95.24
				2	307	95.24	97.56
				3	602	100.00	100.00

由实验结果可以看出:

1) 使用随机检索法比使用整体检索法节省了检索时间,基本上 T_R 等于参考模板与随机段的比值。

2) 在参考模板长度固定的情况下,随机段越长,检索出目标的时间就越长;随机段固定的情况下,参考模板越长,检索出目标的时间就越长。

3) 由于这种方法只是通过计算相似性来进行检索,不关心详细的语义内容,因此它可以用来检索任意类型的音频数据。

在实验中发现,滑动步长的大小影响实验结果,当滑动步长比较小时(本实验取 0.1 s),检索目标一般都可以检索到,但计算量比较大,需要的时间比较长;当滑动步长较大时,计算量比较小,需要的时间比较短,但因检索是在相同的时间位置上找相同的数据,若滑动步长比较大时,则内容有可能会错位导致相似度降低,若低于设定的阈值时检索目标就会漏检。比如本来随机段与滑动窗信号的内容都是“本届奥运会中国金牌第一”,其后紧接内容是“美国第二”,但若因滑动步长取得比较大的原故,“本届奥”这三个字没能进入滑动窗,进入滑动窗的内容变成“运会中国金牌第一美国第”。

3 结语

文中提出了一种快速有效的基于随机段的音频检索方法,当参考模板取自实验数据文件时,实验结果表明,该检索方法减少了计算量,提高了检索速度;解决了整体检索方法中当检索目标比较长时,检索时间过长的问题。最后又分析了滑动步长的选择问题。

参考文献:

- [1] FOOTE J. An overview of audio information retrieval[J]. *Multimedia Systems*, 1999, 7(1): 2-10.
- [2] HANSEN J H L, HUANG RONGQING. SpeechFind: Advances in

spoken document retrieval for a national gallery of the spoken word [J]. *IEEE Transactions on Speech and Audio processing*, 2005, 13(5): 712-730.

- [3] CHECHIL G, LE E, REHN M, *et al.* Large-scale content-based audio retrieval from text queries[C]// *Proceedings of 1st ACM International Conference on Multimedia Information Retrieval*. New York: ACM, 2008: 105-112.
- [4] 张卫强, 刘加. 网络音频数据库检索技术[J]. *通信学报*, 2007, 28(12): 152-155.
- [5] 张卫强, 刘加. 一种基于仿生模式识别思想的固定音频检索方法[J]. *自然科学进展*, 2008, 18(7): 808-813.
- [6] SMITH G, MURASE H, KASHINO K. Quick audio retrieval using feature search[C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*. New York: IEEE, 1998, 6: 3777-3780.
- [7] KASHINO K, KUROZUMI T, MURASE H. A quick search method for audio and video signals based on histogram pruning[J]. *IEEE Transactions on Multimedia*, 2003, 5(3): 384-357.
- [8] 郑贵滨, 韩纪庆. 基于分段的实时声频检索方法[J]. *声学学报*, 2006, 31(2): 101-108.
- [9] ZHANG W Q, LIY J. two-stage method for specific audio retrieval [C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*. New York: IEEE, 2007, 4: 85-88.
- [10] YAO J C, WAN W W, YU X Q, *et al.* A quick specific audio retrieval algorithm based on general prediction[C]// *IEEE 2008 International Conference on Audio, Language and Image Processing*. New York: IEEE, 2008, 1180-1184.
- [11] JOHNSON S E, WOODLAND P C. A method for direct audio search with application to indexing and retrieval[C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*. New York: IEEE, 2000, 3: 1427-1430.

(上接第229页)

向量机作为分类器,建立了红虫及淡水浮游生物图像的分类识别模型,该模型基本能正确识别红虫,具有比较好的分类精度。本文仅对淡水中的红虫、剑水蚤和猛水蚤进行了分类识别实验,待今后条件成熟了,还可以对水中的浮游藻类与浮游生物之间的识别进行有效地研究。同时本文实验均在培养皿中进行实验,如果条件成熟可在水源地捕获数据源进行训练和分类,这样能更好地验证本文方法的有效性。进一步可以通过统计红虫数量比例等手段,为水厂红虫爆发预警,水厂水质监测,投药量设置等方面提供支持,同时为水厂浮游生物自动检测奠定了基础。

参考文献:

- [1] 左金龙, 崔福义, 孙兴滨. 饮用水处理工艺中摇蚊幼虫污染防治

技术的研究进展[J]. *水处理技术*, 2006, 32(9): 1-5.

- [2] 陈孝敬, 吴迪, 何勇, 等. 基于小波包和偏最小二乘支持向量机的多光谱纹理图像的大米分类研究[J]. *光谱学与光谱分析*, 2009, 29(1): 222-225.
- [3] 唐远炎, 王玲. 小波分析与文本文字识别[M]. 北京: 科学出版社, 2004.
- [4] MUEZZINOGLU M K, ZURADA J M. RBF-based neurodynamic nearest neighbor classification in real pattern space[J]. *Pattern Recognition*, 2006, 39(5): 747-760.
- [5] 荣海娜, 张葛祥, 金伟东. 系统辨识中支持向量机核函数及其参数的研究[J]. *系统仿真学报*, 2006, 18(11): 3204-3226.
- [6] 赵晶莹, 孙兴滨, 吕伟民, 等. 基于小波分析的红虫识别[J]. *东北林业大学学报*, 2009, 37(4): 112-114.