超星·期刊

第25卷第6期 华东交通大学学报 Vol. 25 No. 6
2008年12月 Journal of East China Jiaotong University Dec. ,2008

文章编号:1005-0523(2008)06-0068-05

一种基于定点DSP的语音识别算法实现

吕 涛,刘百芬,燕贤青

(华东交通大学 电气与电子工程学院,江西 南昌 330013)

摘要:介绍了一种基于定点DSP的特定人的孤立词语音识别系统,着重讨论了算法中参数快速提取的实现. 该算法采用基于动态规划(DP)技术的动态时间弯折算法(DTW),采用VAD两极端点检测方案以提高端点检测的精度,以Mel尺度倒谱系数为语音识别参数. 首先讨论了标准浮点实现方法,然后结合算法特点和定点DSP的结构,采用了一种优化的快速定点实现方法. 测试结果表明,定点算法时间比浮点算法大大降低,使系统的性能得到了提高.

关 键 词:语音识别;动态时间弯折算法;Mel尺度倒谱系数;定点算法

中图分类号:TN912.3 文献标识码:A

语音识别是近年来十分活跃的一个研究领域. 在不远的将来,语音识别技术有可能作为一种重要的人机交互手段,辅助甚至取代传统的键盘、鼠标等输入设备,在个人计算机上进行文字录入和操作控制. 而在手持式PDA、智能家电、工业现场控制等应用场合,语音识别技术有更为广阔的发展前景. DSP技术的出现和发展为语音识别技术的应用提供了强有力的支持. 当前,特定说话人、中等词汇量的单词识别系统已经在DSP上得到实现,但效果并不理想,为了这些成果早日走向实用,走向市场,许多语音实验室及DSP厂商投入了大量的人力物力开发基于DSP的语音识别系统. 我们研制的基于定点DSP的语音识别系统,实现了低成本,训练方便简单的语音识别方案,识别率大大提高.

1 基于定点DSP的语音识别系统

该系统采用TI公司54X系列中的TMS320VC5416作为系统的核心部件,TMS320VC5416是TI公司0.15 μm器件中的第一款定点DSP芯片,有128 k×16位片内RAM,速度为160 MIPS,有3个多通道缓冲串行口(McBSP)[1].

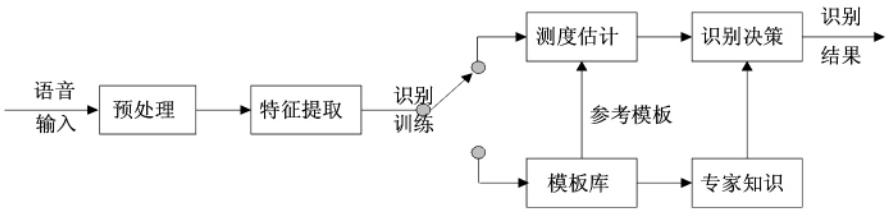

图1 典型的语音识别系统框图

语音识别系统的典型方案如图1所示. 输入的模拟语音信号首先要进行处理,包括预滤波、采样和量化、加窗、端点检测、预加重等. 语音信号经预处理后,接下来很重要的一环就是特征参数提取. 对特征参数的要求是:(1)提取的特征参数能有效地代表语音特征,具有很好的区分性. (2)各阶参数之间有良好的独立性.

收稿日期:2008-09-09

作者简介:吕 涛(1980-),男,河北保定人,硕士研究生,研究方向为DSP技术.

超星·期刊

(3) 特征参数要计算方便,最好有高效的计算方法,以保证语音识别的实时实现[2].

2　语音识别算法

在本系统中我们采用最为简单有效的方法 DTW(Dynamic Time Warping,动态时间弯折) 算法,该算法基于 DP(Dynamic Project,动态规划) 的思想,解决了发音长短不一的模板匹配问题,是语音识别中出现较早、最为经典的一种算法[3].

为了更好地适应环境的干扰和变化,提高端点检测的精度. 使用 VAD 两极端点检测算法,使所有说话人的识别率都有不同程度的提高. 传统的 LPC(线性预测系数) 模型是基于发音模型建立的,LPCC(LPC 倒谱系数) 系数也是一种基于合成的参数. 这种参数没有充分利用人耳的听觉特性. 实际上,人的听觉系统是一个特殊的非线性化系统,它响应不同频率信号的灵敏度是不同的,基本上是一个对数的关系. 所以我们采用了一种能够比较充分利用人耳这种特殊感知特性的参数,即 Mel 尺度倒谱参数(Mel_scaled Cepstru Coeffi-cients) ,简称为 MFCC,MFCC 参数能够比 LPCC 参数更好地提高系统的识别性能[4]. MFCC 的计算首先用 FFT 将时域信号转化成频域,之后对其对数能量谱用依照 Mel 刻度分布的三角滤波器组进行卷积,最后对各个滤波器的输出构成的向量进行离散余弦变换 DCT,取前 N 个系数.

2.1　未改进的 MFCC 算法

MFCC 参数计算的要点是将线性功率谱 $S(n)$ 转换成为 Mel 频率下的功率谱,在计算之前需要在语音的频谱范围内设置若干个带通滤波器 $Hm(n)$,$m=0,\cdots,M-1$,$n=0,\cdots,N/2-1$. M 为滤波器个数,N 为一帧语音信号的点数. 每个滤波器具有三角形特性,其中心频率为 f_m,它们在 Mel 频率轴上是均匀分布的. 在线性频率上,当 m 较小时相邻的 f_m 间隔很小,随着 m 的增加,相邻的 f_m 间隔逐渐拉开. Mel 频率和线性频率的转换关系如下:

$$mel=\ln\left(1+\frac{f}{700}\right)\left(\frac{1000}{\ln\left(1+1000/700\right)}\right) \tag{2-1}$$

这些带通滤波器的参数是事先计算好的. 让 M 选择为 26,FFT 点数 N 为 256. 语音信号的采样频率为 8 000 Hz.

下面是 MFCC 参数的计算方法[5]:

(1) 首先确定每一帧语音采样序列的点数,本系统取 $N=256$ 点. 对每帧序列 $S(n)$ 进行预加重处理后再经过离散 FFT 变换,取模的平方得到离散功率谱 $S(n)$.

(2) 计算 $S(n)$ 通过各 $H_m(n)$ 后所得的功率值,即计算 $S(n)$ 和 $H_m(n)$ 在各离散频率点上乘积之和,得到 M 个参数 P_m,$m=0,\cdots,M-1$.

(3) 计算 P_m 的自然对数,得到 L_m,$m=0,\cdots,M-1$.

(4) 对 $L_0,\cdots,L_{m-1}$ 计算其离散余弦变换,得到 D_m,$m=0,\cdots,M-1$.

(5) 舍去代表直流成分的 k,取 $D_1\cdots D_k$ 作为 MFCC 参数. 此处 $k=12$.

但是,考虑到系统的适时性能要求,单纯的浮点计算需要大量的计算量,无法满足.

2.2　MFCC 算法的定点化

以上算法如果用浮点 C 程序在 DSP 上实现,计算一次需耗时 12.20 ms 左右,耗时太长,另外还要考虑语音编码程序所需的时间,因此计算量太大,无法满足实时操作. 为此根据 MFCC 算法的特点,进行了定点化处理. TMS320VC5416 是 16 位定点处理器,定点化处理应该注意以下几个问题:

(1) TMS320VC5416 支持小数和整数两种运算方式,在计算时应选择小数方式,使计算结果的绝对值都小于 1;

(2) 采用双字定点运算库代替 C 语言的浮点库,提高运算速度;

(3) 注意在每次乘加运算之后进行饱和操作,防止结果上溢或下溢;

(4) 由循环处理后的一组数据可能有不同的指数,要进行指数归一化,以便后续定点操作对指数和尾数部分分别处理.

超星·期刊

2.3 不含任何优化的 MFCC 参数定点实现方法[6].

(1) 计算输入信号的离散 FFT 变换: 在 FFT 变换的程序中采用 16 位算法,每层蝶形计算完毕后提取最大的指数作为归一化参数 e,所有数据据此进行一次归一化,保证结果在 $[-0.5.TIF,+0.5]$ 范围之内. 最后得到的离散 FFT 结果为:

$$Y = DFFT(X) = Y_a + jY_b = 2^e Y_a + j2^e Y_b \tag{2-2}$$

其中 2^e 为归一化参数,而 $Y_a(n)$、$Y_b(n)$ 都在 $[-0.5.TIF,+0.5]$ 之内.

(2) 计算功率谱

$$S = |Y|^2 = |2^e Y_a + j2^e Y_b|^2 \tag{2-3}$$

或

$$S(n) = 2^{2n}(Y_a(n)^2 + Y_b(n)^2) = 2^{2e}P(n), n = 0,\cdots,N/2-1 \tag{2-4}$$

(3) 计算滤波器输出

$$H_m = \sum_{n=0}^{N/2-1} W_m(n) \cdot S(n) = 2^{2e}\sum_{n=0}^{N/2-1} W_m(n) \cdot P(n) \tag{2-5}$$

其中 W_m 是第 m 个滤波器. 注意在此乘加循环中,累加的结果很可能大于 1,因此必须再提取一个最大的指数 2^b,将结果归一化到 [0,1],即:

$$H_m = 2^{2e} \cdot 2^b \cdot F_m, m = 0,\cdots,M-1 \tag{2-6}$$

其中 M 为滤波器的个数,在这里 $m = 26$.

(4) 计算对数

$$L_m = \ln(H_m) = \ln(2^{2e} \cdot 2^b \cdot F_m) = \ln(2^{2e} \cdot 2^b) + \ln(F_m), m = 0,\cdots,M-1 \tag{2-7}$$

(5) 计算离散余弦变换: $D = dct(L)$ 或 $D_k = \frac{2c(k)}{M}\sum_{m=0}^{M-1} L(m)\cos\left(\frac{(2m+1)k\pi}{2M}\right), k = 0,\cdots,M-1$ (2-8)

其中:

$$c(k) = \begin{cases} \sqrt{2}/2, k = 0 \\ 1, k = 1,2,\cdots,M-1 \end{cases}$$

最后取 $D(1),\cdots,D(12)$ 作为 MFCC 参数即可.

2.4 优化的定点化 MFCC 算法

由以上步骤可以看到,在计算 DCT 参数时,由第一步和第三步得到的归一化参数 2^{2e}、2^b 是以常数的形式存在的. 由 DCT 的定义可知,该常数将作为直流分量的一部分体现在 D_0 中,而 D_0 又恰恰是要被舍去的. 因此实际计算时不必考虑 e、b 的影响. 这样就得到了 MFCC 参数的快速定点算法:

(1) 计算离散 FFT 并归一化,去除指数部分,只保留小数部分

$$Y = DFFT(X) = Y_a + jY_b + 2^e Y_a + j2^e Y_b$$

$$\hat{Y} = Y_a + jY_b \tag{2-9}$$

(2) 计算功率谱

$$\hat{S}(n) = Y_a(n)^2 + Y_b(n)^2 = P(n) \tag{2-10}$$

(3) 采用定点算法计算滤波器输出并归一化,去除指数部分,只保留小数部分

$$\hat{H}_m = \sum_{n=0}^{N/2-1} W_m(n) \cdot \hat{S}(n) = \sum_{n=0}^{N/2-1} W_m(n) \cdot P(n) \tag{2-11}$$

$$\hat{H}_m = 2^b \cdot F_m$$

(4) 计算对数

$$\hat{L}_m = \ln(F_m) \tag{2-12}$$

(5) 直接计算所需的 12 点 DCT

$$D_k = 1/M\sum_{m=0}^{M-1} \hat{L}(m)\cos((2m+1)k\pi/(2M)), k = 1,\cdots,12 \tag{2-13}$$

其中,上式中的系数事先计算好. 经过以上优化,计算一帧 MFCC 参数所需的时间减少到 1.35 ms. 再除去 ADPCM 编码所需的时间,系统有几乎 85% 的时间是空闲的. 在空闲时令 DSP 处于 IDLE(空闲)状态,可使 DSP 的工作电流降到 12 mA 以下,从而有效地降低了系统的功耗.

超星·期刊

3　人机交互界面

系统采用 VC ++ 来进行人机交互界面的设计,通过调用 MATLAB 来进行矩阵运算; 在 PC 机上通过 CCS 对程序调试成功后,把程序下载到 DSP 上,通过 VC ++ 来实现对 DSP 的控制,如图 2 所示:

图 2　语音识别交互界面

从采样率的下拉列表框中可以选取语音信号的采集频率,声道数列表框中选取单声道还是双声道. 可以先点击与 Matlab 通信按钮,实现程序的仿真,然后点击与 DSP 通信,实现程序在 DSP 上的运行. 对着麦克风录音,分别在下面显示信号的输入和识别结果的输出,如一致,则说明识别成功. 在此处采样率选择 8 000 Hz,声道数选择单声道,通过麦克风输入“电气”,发现识别正确.

4　实验结果

第一组: 先采集 10 个学生(5 男 5 女,普通话均较标准) 的共 100 个样本(内容为 10 个系名如计算机、自动化、中文、新闻等) ,并对所有的样本进行训练,做成 10 个模板,采用传统的 MFCC 算法对该系统进行测试.

第二组: 利用第一组采集的语音数据,采用改进的 MFCC 算法对该系统进行测试. 所得测试结果如表 1 所示.

表 1　语音系统的测试结果

内　容	第一组	第二组
样本数	100	100
错误识别数	2	3
识别率(%)	98	97

从表中得知,两种的识别率很接近,但是后者识别需要的时间只有前者的 1 /3,大大节省了时间,而且降低了系统的功耗,提高了性能.

5　结论

随着 DSP 的大量普及,用 DSP 芯片构成的语音识别系统有着广泛的应用背景. 在定点 DSP 上实现浮点运算,操作数必须变成定点数,然后再返回浮点数,极大地降低了运算速度. 因此,采用对算法的定点化处理,

超星·期刊

充分利用了定点 DSP 的定点计算能力,提高了代码效率,提升了系统性能. 本文的定点 DSP 算法可以应用在对实时性要求较高、效率较高的场合,例如在飞机等要求较高的语音控制. 在编写语音识别算法时,对其进行定点化处理以及一些原则和方法对其它类似的算法也有着现实指导意义. 在实际的应用中,应注意根据 DSP 芯片的特点,对算法进行优化,使得 DSP 芯片的性能得到提升.

参考文献:

[1] 苏明武. 基于 DSP 的语音识别技术研究及实现 [D]. 哈尔滨: 哈尔滨工程大学,2005.
[2] 杨行峻,迟惠生. 语音信号数字处理 [M]. 北京: 电子工业出版社,1998.
[3] 吴晓平,崔光照,路 康. 基于 DTW 算法的语音识别系统实现 [J]. 电子工程师,2004,12(7):17 - 19.
[4] 俞一彪,许允喜,芮贤义. 一种语音特征参数子分量分析与有效性评价的新方法 [J]. 信号处理,2007,23(2):23 - 25.
[5] 但志平,王以治,等. 基于倒谱参数和支持向量机技术的说话人识别系统 [J]. 电声技术,2004,6(3):39 - 43.
[6] 余小青,万旺根,等. 用 DSP 技术开发的一种新的语音识别系统 [J]. 上海大学学报,1999,5(2):41 - 43.

The Realization of A Fixed_point Speech Recognition Algorithm Based on DSP

LV Tao,LIU Bai_fen,YAN Xian_qing

(School of Electrical and Electronic Engineering,East China Jiaotong University,Nanchang 330013,China)

Abstract: The paper introduces a specific isolated word speech recognition systems based on fixed_point DSP,focusing on achieving the rapid extraction algorithm parameters. The algorithm is based on dynamic programming (DP) technology Dynamic Time Warping algorithm (DTW),by using VAD bipolar endpoint detection programmers to improve the accuracy of endpoint detection,and MFCC for speech recognition parameters. The paper firstly discusses realizing methods of standard floating_point. Then combining with its fixed_point DSP algorithm characteristics and the structure,it adopts an optimized fast fixed_point method. The results show that time of the fixed_point algorithms is substantially reduced compared to floating_point arithmetic,therefore,the performance of the system has been enhanced.

Key words: speech recognition; dynamic time warping algorithm; MFCC; fixed_point

(责任编辑: 周尚超)