文章编号:1000-582X(2007)09-0076-04

基于FPGA的说话人识别系统设计与实现

何 伟,胡又文,张 玲,陈方泉
(重庆大学 通信工程学院,重庆 400030)

摘 要:针对当前基于DSP等硬核处理器的嵌入式说话人识别系统存在训练和辨认时间较长的问题,提出一种基于FPGA平台与矢量量化原理的说话人识别系统实现方案。在采用遗传算法进行矢量量化的说话人识别的系统中,该方案实现的硬件并行运算结构可大大减少求适应度的耗时。经测试,该实现方案在保证识别率前提下,可有效提高训练与识别速度。

关键词:说话人识别;矢量量化;遗传算法;适应度;FPGA
中图分类号:TP391.42 **文献标志码**:A

说话人识别(Speaker Recognition)又称为话者识别,是指根据特定说话人语音波形中反映生理和行为等特征的语音参数来对说话人身份进行识别[1]。说话人识别技术作为一种非接触性识别技术,在司法、军事和信息服务等领域都有广泛的应用前景。

SOPC技术是一种基于FPGA解决方案的SOC,由美国ALTERA公司于2000年提出[2]。基于SOPC平台的开发结合了FPGA灵活可编程与片上NiosII软核处理器的用户可配置等特点。在实现某功能时,可编写C/C++程序运行于NiosII处理器实现,也可设计硬件模块实现,通过硬件实现加速。

目前的嵌入式说话人识别系统通常是基于DSP等硬处理器平台实现,训练与识别耗时长,实时性较差。笔者基于矢量量化(VQ)的说话人识别算法研究的基础上,根据采用遗传算法进行矢量量化的算法特点[3],综合考虑训练与识别时间,资源消耗等因素,在Cyclone II 2C35系列FPGA上实现了嵌入式说话人识别系统。经验证,该说话人识别系统识别率高,实时性优于硬核处理器系统,应用前景良好。

1 基于VQ的说话人识别算法

说话人识别系统包含记录用户语音特征和识别测试者语音2个主要功能,其系统结构框图如图1所示。根据图1可见,系统的任务主要有3项:一是说话人语音采集与语音特征参数提取;二是通过遗传算法计算得到说话人语音码书,称为训练过程;三是在说话人识别时将测试者的语音特征参数与已有码书进行匹配并作出决策,称为识别过程。

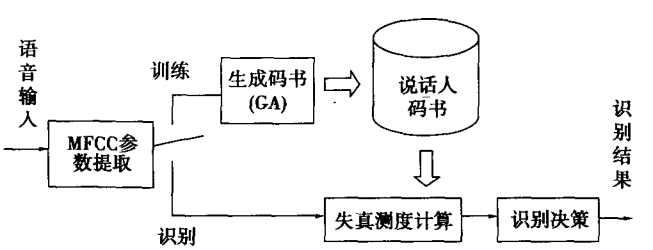

图1 说话人识别系统框图

当系统采集到用户或测试者的语音数据后,首先要提取反映说话人语音特征的参数,系统选用的是MFCC参数。MFCC参数是一种基于人耳对语音频率的非线形感知特征的描述参数[5],在说话人识别中,其性能优于LPC,LPCC等参数。

训练过程就是系统为用户生成码书的阶段,当系统得到用户的一段MFCC参数 $X_n=\{x_{n1},x_{n2},\cdots,x_{nM}\}$ $(n=1,2,\cdots,N)$ 后,要以这段MFCC为依据为该用户

收稿日期:2007-05-30
基金项目:国家自然科学基金资助项目(60472037)
作者简介:何伟(1964-),男,重庆大学副教授,主要从事电子设计自动化、信号与信息处理的研究,(Tel)13908381077;(E-mail)hw@ccee.cqu.edu.cn。

建立一部表征其个人语音特征的码书,X_n 可视为 N 个 M 维矢量点,称为训练序列。码书的生成过程就是将训练序列的 M 个点聚为 K 类,每一类用一个 M 维点表示,用这 K 个点描述训练序列中 M 个点的空间分布情况。聚类得到的序列 $S_n=\{s_{n1},s_{n2},\cdots,s_{nM}\}(n=1,2,\cdots,K)$ 即为一部容量为 K 的码书。

识别过程中,系统将得到测试者语音 MFCC,称为测试序列,然后与用户码书匹配,如果与某用户码书的失真测度大于阈值,则可认为测试者为该码书表征的用户。

由于码书的生成是高维空间中的点聚类过程,如果用 LBG,K 均值等方法进行聚类易导致结果陷入局部最优点,因而笔者选择具有全局搜索性能的遗传算法进行聚类[5],可得到性能更好的 VQ 码书。针对说话人识别设计的具体步骤如下:

1)群体规模设置为 30,随机初试化群体;

2)采用简单遗传操作,变异与交叉采用无回放随机选择策略,单点交叉,交叉概率 $P_C=90\%$,变异概率 $P_M=10\%$,每一代具有最优适应度的 10% 个体直接保留;

3)计算执行遗传操作后的所有个体的适应度,淘汰 10% 的最差个体;如果遗传代数为 3 的倍数,执行一次 K 均值算法,加快收敛速度;

4)如果遗传代数达到规定阈值或最近三代最优个体适应度比值达一定阈值,停止遗传,否则转步骤 2)。

步骤 3)中个体适应度计算公式为式(1)所示。式(1)中 X 为训练序列,Y 为个体,$d(X_j,Y_i)$ 是训练序列中某点 X_j 与个体中某点 Y_i 之间的欧氏距离,该公式也可计算失真测度,此时,X 为测试序列,Y 用码书序列 S_n 代替;

$$f_i=\frac{1}{\frac{1}{N}\sum_{j=0}^{N-1}\min\left[\underset{Y\in Y_K}{d}(X_j,Y_i)\right]},\qquad(1)$$

遗传结束后,最末代的最优适应度个体即为用户的 VQ 语音码书。

2　硬件系统与算法优化

SOPC 系统中,除含有必要的储存器与语音输入接口外,还外接 PS2 键盘与 LCD,VGA 显示器等人机交互设备,系统整体设计框图如图 2 所示。

2.1　语音采集与语音 MFCC 参数提取

输入的语音在采样芯片中经采样和 A/D 转换,以 8 kHz 的采样频率和 16 bit 采样深度串行传输到 FPGA 芯片端口。FPGA 片内有一个用户制定硬件模块实时接收采样芯片传来的语音数据,同时该模块还计算每

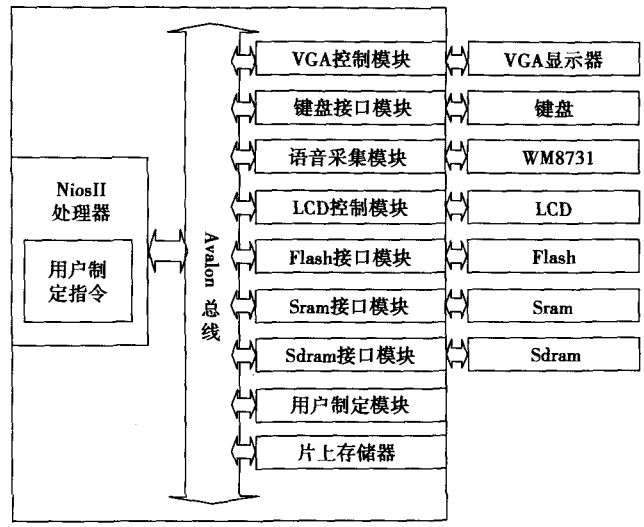

图 2　系统整体设计框图

次接收到的数据的前向差值与平方值,并将 3 个数值传至 SRAM 中的存储区,当存储区满后,模块以中断方式通知 CPU 读取数据。CPU 读出数据后分帧并根据式(2)、(3)求出每一帧的短时能量与过零率,然后通过双门限法检测该段数据是否为有效语音。式(2)、(3)中,N 为每帧采样点数。

由于每个采样点的前向差值与平方值已由数据接受模块算出,CPU 只需提出这些值按帧累加即可。检测为有效语音的数据帧放入 SDRAM 中的循环缓冲区中,当有效语音数据足量后,CPU 停止采集模块工作。

$$\text{过零率}\quad \mathrm{ZCR}(i)=\sum_{n=1}^{N-1}|x_i(n+1)-x_i(n)|,\qquad(2)$$

$$\text{短时能量}\quad e(i)=\sum_{n=1}^{N}x_n^2(n),\qquad(3)$$

由于语音采样频率远低于系统频率,在语音采集与检测的过程中,CPU 可在空闲时间完成 MFCC 提取。MFCC 提取包含大量浮点运算和复数运算,为加快运算,可在 SOPC 中的软核处理器中添加由硬件实现的用户制定指令,这种方式可将提取 MFCC 的速度提高大约 20 倍。

2.2　适应度计算硬件结构及遗传算法实现

CPU 提取出 MFCC 参数完成后得到 N 帧 M 维 MFCC,然后通过遗传算法聚类生成用户码书。根据实验情况并综合考虑系统资源与识别性能,设定系统的典型参数为:总帧数 $N=512$,码书容量 $K=64$,$M=12$,每一帧 MFCC 与其一阶差分参数组合,扩为 24 维,遗传个体 $T=30$。

遗传聚类算法中,交叉和变异等遗传操作主要是对存储器的读写与位操作,采用硬件加速效果提升不大,因此这部分功能由软件在处理器上实现。完成一代遗传操作后,需要评估群体中每一个体的适应度,根

据式(1)估算,执行一代群体适应度计算至少需作$(2 \cdot M) \cdot N \cdot T \cdot K = 23\ 592\ 960 \approx 2.4 \times 10^7$次乘法和$4.8 \times 10^7$次加减法。实验表明,遗传算法实现收敛大约需要40~150代,显然,如果直接用软件程序在运算速度仅67 MIPS的Nios II嵌入式系统上实现,必然导致耗时过长。

根据式(1)可知,适应度求解过程主要是计算欧氏距离$d(X_j, Y_i)$,而高维空间两点间距离的求取有良好的并行特性,因此在笔者提出的适应度计算硬件结构中,用K路运算器并行运算来加快适应度的求解。该硬件结构框图如图3所示,CPU将训练序列与单个个体通过地址分配单元按维写入K路数据存储与运算单元,由选择与控制单元启动运算,得到的K路输出并行进入K输入加法器,再由距离运算单元作开方处理即可得到两点距离,选择与控制对输出结果进行比较,搜索,并累加该值,N次处理后,得到该个体适应度的倒数,并由控制与选择单元以中断方式将该值返回给CPU。然后CPU将下一个个体写入存储单元,重复计算过程。

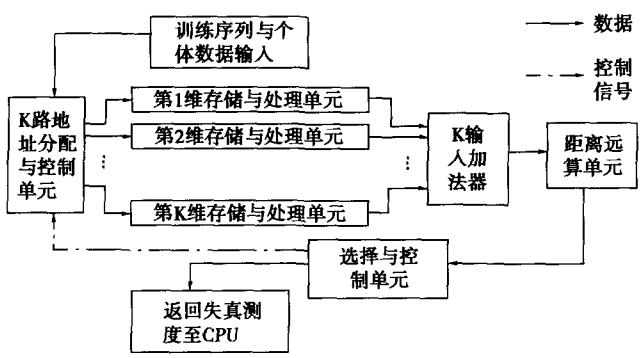

图3 适应度计算硬件结构框图

根据此思路,用硬件逻辑单元流水设计出该适应度计算模块,该模块作为用户制定模块添加到系统总线上,由CPU调用。该模块执行一代群体适应度计算需要时钟周期数仅为:$(F+1) \cdot N \cdot T + (2 \cdot M) \cdot T \cdot F = 1\ 044\ 480 \approx 1 \times 10^6$,该模块峰值运算速度可达6 400 MIPS,远优于软件实现。

在嵌入式系统设计中,通常为了降低求适应度的运算量而对算法作各种简化,如稳态方式[6],通过限制每一代发生变化的个体数量来减少运算,但是这些改进一定程度上限制了算法的随机性,聚类结果容易陷入局部最优解。

2.3 实现说话人识别

说话人识别过程是对测试者的MFCC参数和用户码书进行匹配的过程。识别过程,先提取一段测试者语音的MFCC,由CPU调用适应度计算硬件结构求出测试序列相对每一部码书的失真测度,由于匹配度计算公式与适应度计算公式相同,仅是训练序列和测试序列长度不同,故可以用同样的硬件模块来实现。匹配得到的失真测度与经验阈值比对,大于阈值则认为测试者是码书对应的用户,小于则测试者被拒绝。

3 实验分析与结论

VQ说话人识别中,参数的选择对系统性能有一定影响。主要可选参数有训练序列长度与MFCC维数;被影响的性能参数有误识率,FPGA资源消耗与训练与识别时间。

实验测试环境为普通实验室,参与实验者共24人,男15人,女9人,测试语音时长不低于5 s。实验中,随机选不同人员语音生成用户码本,然后全体人员参与测试。

表1 不同参数设置下系统性能与资源耗用

码本聚类维数	训练语音帧数	误识率/%	FPGA资源耗用		遗传算法平均训练时间/s	10模板下说话人平均识别时间/s
			逻辑单元(门)	存储位单元/位		
12	256	2.43	4 106	61 496	1.827	0.331
16	256	1.97	5 479	81 920	1.807	0.355
12+一阶差分	256	1.88	7 825	122 936	2.154	0.393
16+一阶差分	256	1.82	10 774	163 896	2.405	0.422
12+一阶差分+二阶差分	256	1.69	14 275	184 376	2.561	0.437
12	512	1.74	4 142	110 664	2.858	0.335
16	512	1.55	5 604	147 528	3.132	0.357
12+一阶差分	512	0.85	8 250	196 662	3.413	0.395

根据实验结果表1可知:在相同的训练语音时长(即训练序列帧数)的基础上,使用MFCC+差分参数的系统识别率优于单纯使用MFCC,但带来的数据处理量、存储单元和逻辑单元的消耗也相应增大;同时,训练序列帧数对识别率的影响相对提高维数更加重要。这是因为在训练语音帧数有限的情况下,训练语音时长对用户码书的修正效果更加明显,使码书更能反映用户的语音特征,但是这样也带来大量存储单元的消耗与训练时间的增加。

同时,还进行了不同平台上相同算法的耗时比较实验,结果如图4所示。

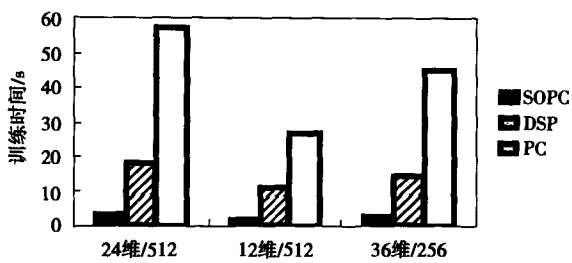

图4　不同硬件平台上的训练用时

图4中DSP平台是采用C5502,PC平台是主频1.6G的AMD处理器,纵轴表示完成训练过程的用时。可见,采用适应度计算模块的SOPC系统速度性能远远优于硬处理器系统。

参考文献:

[1] O'SHAUGHNESSY D. Speaker recognition [J]. IEEE Acoustic, Speech and Signal Processing Magazine, 1986, 3 (4): 4-7.

[2] 任爱锋,初秀琴,常存,等.基于FPGA的嵌入式系统设计[M].西安:西安电子科技大学出版社,2004.

[3] 赵力.语音信号处理[M].北京:机械工业出版社,2003.

[4] 张军英.说话人识别的现代方法与技术[M].西安:西北大学出版社,1994.

[5] 陆金桂,李谦.遗传算法原理及其工程应用[M].北京:中国矿业大学出版社,1997.

[6] BORNHOLDT S, GRAUDENZ D. General asymmetric neural networks and structure design by genetic algorithms [J]. IEEE Transactions on Neural Networks, 1992, 5 (2): 327-334.

Disign and Implementation of Speaker Recognition System Based on FPGA

HE Wei, HU You-wen, ZHANG Ling, CHEN Fang-quan

(College of Communications Engineering, Chongqing University, Chongqing 400030, China)

Abstract: Aiming at the shortcoming of low recognition and training speed in embedded speaker - recognition system based on DS Phard-core processor, a new scheme of system based on FPGA and vector quantization principle is presented. In the speaker-recognition system based on the vector quantization and generation algorithm, a fitness parallel process hardware structure is presented by the scheme which consumes much less time than software processing while getting the fitness. The test shows that the system uses this method obtains both high recognition rate and higher speed in training and recognition.

Key words: speaker recognition; vector quantization; generation algorithm; fitness; FPGA

(编辑　陈移峰)