第 35 卷 第 5 期 计 算 机 工 程 2009 年 3 月
Vol.35 No.5 Computer Engineering March 2009
·开发研究与设计技术· 文章编号：1000—3428(2009)05—0265—03 文献标识码：A 中图分类号：TP311.12

MFCC 中 DCT 结构的设计与实现

孔维功，张国杰，张效军
(解放军信息工程大学信息工程学院，郑州 450002)

摘 要：根据 MFCC 中 DCT 的特点，设计一种基于 DA 算法的实现结构，采用先分解 ROM 再偏移二进制编码的方法对 DA 算法进行优化，将 ROM 表的大小由 2^N 减小到 $(N/K)2^{K-1}$。通过仿真与 FPGA 测试，验证了该设计的正确性，能够满足说话人识别中 MFCC 参数提取的实时性要求和精度要求。

关键词：说话人识别；美尔频率倒谱系数；离散余弦变换；分布式算法

Design and Implementation of DCT Structure in MFCC

KONG Wei-gong, ZHANG Guo-jie, ZHANG Xiao-jun
(School of Information Engineering, PLA Information Engineering University, Zhengzhou 450002)

【Abstract】This paper presents an implementation structure based on Distributed Arithmetic(DA) according to DCT character in MFCC, which optimizes DA by using ROM reduction and offset binary coder, and reduces the size of ROM table from 2^N to $(N/K)2^{K-1}$. The results of simulation and FPGA test show this kind of design is correct, which meets the requirement of real-time and precision in MFCC computation for speaker recognition.

【Key words】speaker recognition; Mel-Frequency Cepstral Coefficients(MFCC); discrete cosine transform; distributed arithmetic

1 概述

说话人识别[1]是项重要的生物特征识别技术，它通过分析语音信号，提取不同说话人的特征参数，并加以利用，以完成对说话人身份的确定。当前主流的说话人特征参数为美尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)[1]，它是基于人耳听觉特性提出的特征参数，是对人耳听觉特性的工程化模拟，其中，DCT(Discrete Cosine Transform)是计算 MFCC 的主要步骤[1]。

目前，说话人识别技术应用主要采用软件或者 DSP 方法来实现，实时性差，若采用集成电路实现核心算法，可以大幅度提高处理速度，改善其实时性。本文针对说话人识别芯片设计中 MFCC 计算时用到的 DCT 特点，采用 DA 算法加以实现，并运用先 ROM 分解再偏移二进制编码的方法改进 DA 算法，将 ROM 表的大小由 2^N 减小到 $(N/K)2^{K-1}$，同时设计其硬件结构，并将该设计在 FPGA 上进行验证。

2 算法的基本原理

2.1 美尔频率倒谱系数(MFCC)

美尔频率倒谱系数是采用滤波器组的方法计算出来的，这组滤波器是对人耳听觉特性的工程化模拟。

美尔频率倒谱 $C_{mel}[n]$ 在美尔刻度谱上可以采用修改的离散余弦变换(DCT)[1]求得：

$$C_{mel}[n]=\sum_{k=0}^{N-1}\log_x(S[k])\cos[n(k+0.5)\frac{\pi}{M}] \tag{1}$$

其中，n=1,2,…,L，L 为 MFCC 参数的阶数；$S[k]$表示第 k 个滤波器的输出能量；n=1,2,…,N-1。

MFCC 的计算过程如图 1 所示。

MFCC 中 DCT 的特点有：(1)为非 2^N 点的 DCT 变换；(2)在 M 点的变换结果中只取了其中序号为 1~L 的 L 个，即部分输出；(3)去掉了规一化系数。

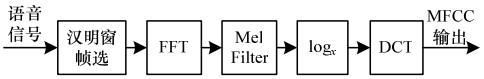

图 1 MFCC 的计算过程

在说话人识别系统中，MFCC 参数取 M=24, L=12。

2.2 一维 DCT 算法的分解

N 点一维 DCT[2]可表示为式(2)，其形式上是向量的内积：

$$Y(k)=\sqrt{\frac{2}{N}}\sum_{m=0}^{N-1}c(m)x(n)\cos\frac{(2m+1)k\pi}{2N} \tag{2}$$

其中，当 $c(0)=1/2$, $m\neq 0$ 时，$c(m)=1$。当 N 为偶数时，N 点一维 DCT 根据变换矩阵的奇偶对称性，可将式(2)进行如下分解：

$$\begin{bmatrix}y_0\\y_2\\\vdots\\y_{N-2}\end{bmatrix}=\sqrt{\frac{2}{N}}\begin{bmatrix}c_{0,0} & c_{0,1} & \cdots & c_{0,\frac{N}{2}}\\ c_{1,0} & c_{1,1} & \cdots & c_{1,\frac{N}{2}}\\ \vdots & \vdots & & \vdots\\ c_{(N-2),0} & c_{(N-2),1} & \cdots & c_{(N-2),\frac{N}{2}}\end{bmatrix}\begin{bmatrix}x_0+x_{N-1}\\x_1+x_{N-2}\\\vdots\\x_{\frac{N}{2}-1}+x_{\frac{N}{2}}\end{bmatrix},$$

$$\begin{bmatrix}y_1\\y_3\\\vdots\\y_{N-1}\end{bmatrix}=\sqrt{\frac{2}{N}}\begin{bmatrix}c_{1,0} & c_{1,1} & \cdots & c_{1,\frac{N}{2}}\\ c_{3,0} & c_{3,1} & \cdots & c_{3,\frac{N}{2}}\\ \vdots & \vdots & & \vdots\\ c_{(N-1),0} & c_{(N-1),1} & \cdots & c_{(N-1),\frac{N}{2}}\end{bmatrix}\begin{bmatrix}x_0-x_{N-1}\\x_1-x_{N-2}\\\vdots\\x_{\frac{N}{2}-1}-x_{\frac{N}{2}}\end{bmatrix} \tag{3}$$

因此，N(N 为偶数)点一维 DCT 只需将输入数据进行预

基金项目：国家“863”计划基金资助项目(2006AA01Z425)
作者简介：孔维功(1980－)，男，助理工程师、硕士研究生，主研方向：集成电路设计；张国杰，教授；张效军，讲师、硕士
收稿日期：2008-09-30 **E-mail：**kongwg_365@163.com

先求和求差的处理后，可转化为计算 $\frac{N}{2}\times\frac{N}{2}$ 和 $\frac{N}{2}\times1$ 的矩阵乘法，即向量的内积，这样可以显著降低乘的次数。

3 MFCC 中 DCT 的实现

在多种 DCT 变换硬件实现结构中，DA 算法的实现结构最为合理。DA 算法实现结构的思想是：将 DCT 系数可能组合的值预先算出并存储在 ROM 表中，以输入数据的比特位为地址进行查表累加，实现向量内积的功能。因此，采用 DA 算法的实现结构只需设置计算 1~L 维所需要的 ROM 表，经过查表累加即可得到 DCT 的 1~L 维，无需设置计算其他各维的 ROM 表，从而减小硬件开销。

当采用传统 DA 算法实现内积时，ROM 表的大小为 2^N，且随 N 成指数增长，因此，如何减小 ROM 表的大小是运用 DA 算法的关键。

3.1 采用偏移二进制编码的 DA 算法

若采用偏移二进制编码(OBC)[3]的形式将传统 DA 算法加以改进，可将 ROM 的大小减为原来的一半，即 2^{N-1}，其推导如下[3]：

2 个长度为 N 的矢量 $\boldsymbol{C}$ 和 $\boldsymbol{X}$ 的内积：

$$Y=\sum_{i=0}^{N-1}c_i\cdot x_i \tag{4}$$

补码形式：

$$x_i=-x_{i,W-1}+\sum_{j=1}^{w-1}x_{i,W-1-j}2^{-j} \tag{5}$$

采用偏移二进制编码，将 x_i 表示为

$$x_i=\frac{1}{2}[x_i-(-x_i)]=\frac{1}{2}[-(x_{i,W-1}-\overline{x_{i,W-1}})+\sum_{j=1}^{W-1}(x_{i,W-1-j}-\overline{x_{i,W-1-j}})-2^{-(W-1)}] \tag{6}$$

现做如下定义：

$$d_{i,j}=\begin{cases}x_{i,j}-\overline{x_{i,j}}, & j\neq W-1\\ -(x_{i,W-1}-\overline{x_{i,W-1}}), & j=W-1\end{cases} \tag{7}$$

$$x_i=\frac{1}{2}[\sum_{j=0}^{W-1}d_{i,W-1-j}2^{-j}-2^{-(W-1)}] \tag{8}$$

$$Y=\sum_{i=0}^{N-1}c_i\cdot\frac{1}{2}[\sum_{j=0}^{W-1}d_{i,W-1-j}2^{-j}-2^{-(W-1)}]=\sum_{j=0}^{W-1}(\sum_{i=0}^{N-1}\frac{1}{2}c_id_{i,W-1-j})2^{-j}-(\frac{1}{2}\sum_{i=0}^{N-1}c_i)2^{-(W-1)} \tag{9}$$

$$D_j=\sum_{i=0}^{N-1}\frac{1}{2}c_id_{i,j},\quad 0\leqslant j\leqslant W-1 \tag{10}$$

$$D_{extra}=-\frac{1}{2}\sum_{i=0}^{N-1}c_i \tag{11}$$

则可以得到：

$$Y=\sum_{j=0}^{W-1}D_{W-1-j}2^{-j}+D_{extra}2^{-(W-1)} \tag{12}$$

因此，采用偏移二进制编码后，ROM 表中存储 D_j 的值，D_{extra} 作为累加的初值。由式(7)可知，$d_{i,j}\in\{-1,1\}$，故只需存储 $d_{i,j}$=1 时 D_j 的值，当 $d_{i,j}$=-1 时的值可由查表后取相反数得到，ROM 表的大小可减小到原来的一半，即 2^{N-1}。此方法虽然需要增加控制逻辑和译码逻辑，但其减小 ROM 表大小的作用较明显，尤其是当 N 较大时。

3.2 DA 算法的 ROM 分解

由式(12)可知，D_j 与 Y 为线行关系，可以考虑采用 ROM 分解技术进一步减小 ROM 表的大小。所谓 ROM 分解技术是指：对于线性 ROM 表，将其 N 个地址位分成 N/K 个 K 位组，即将 2^N 的 ROM 分解为 N/K 个 2^K 的 ROM，使用多输入累加器将这些 ROM 的输出相加即得到真正的查表值。

3.3 DA 算法的改进

将偏移二进制编码与 ROM 分解技术有机结合可以有效减小 ROM 表的大小，但 2 种方法运用顺序的不同会使其效果也有所不同。若先用 OBC 编码再进行 ROM 分解，ROM 表的大小为：$2^{K-1}+((N/K)-1)2^{K-1}\times2^K$(因为此时 ROM 表的地址位为 N-1 位)；若先进行 ROM 分解再用 OBC 编码，ROM 表的大小为：$(N/K)2^{K-1}$。由此可见，先进行 ROM 分解再用 OBC 编码来减小 ROM 表大小的效果更好。

4 硬件实现结构设计

4.1 说话人识别系统

在说话人识别系统中，采用的特征参数为 MFCC，识别算法采用 VQ，系统能同时识别多路语音信号，每路语音以 256 个采样点为 1 帧数据进行 MFCC 参数提取，数据为定点有符号数，以补码形式存储，位宽 32 bit，其中，整数部分为 8 bit，小数部分 23 bit。MFCC 参数中取 M=24, L=12。定义帧标签为 SR_Tag，以指示该帧数据的话路号及数据性质。

先根据 DCT 变换矩阵的奇偶对称性进行分解，再采用 ROM 分解与 OBC 编码相结合的 DA 算法结构实现 DCT。DCT 模块的输入数据为 24 个 Mel 滤波器输出，并取对数后的数据 x_0~x_{23}，格式为“SR_Tag＋x_0~x_{23}”；输出数据为 12 个 MFCC 参数，写入到 FIFO 中，由 VQ 模块取出进行匹配识别，其格式为“SR_Tag＋$MFCC_1$~$MFCC_{12}$”。

4.2 MFCC 参数提取中 DCT 变换硬件结构

MFCC 参数提取中 DCT 变换硬件结构如图 2 所示，分为 2 个模块：DCT_Ctl 和 DCT 模块(图中虚线框部分)。

(1)控制模块 DCT_Ctl：完成输入、输出数据的控制，帧标签判别、缓存，同时向 DCT 模块发送待处理的数据、接收处理后的数据；

(2)DCT 模块：完成数据的 DCT 运算。它分为 Pre(预加减移位模块)、RAC(Recycle Add Cell 累加单元)2 个模块。Pre 完成 DCT 模块输入数据的预加减并移位输出，为 RAC 提供 ROM 表的查表地址；RAC 进行查表累加完成向量内积的功能，并将计算结果串行输出。

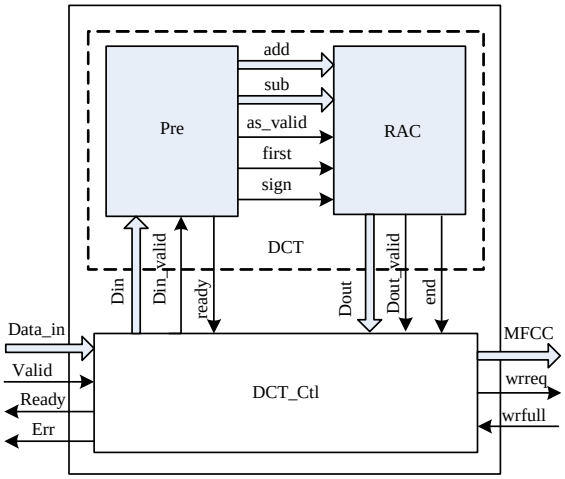

图 2 DCT 硬件结构

4.2.1 预加减移位模块 Pre

Din, Din_Valid 分别为输入数据和有效标志，ready＝1

表示 Pre 可以接收数据；add, sub 为和值、差值移位输出，其位宽均为 12 bit，为 RAC 中 ROM 表的查表地址；as_valid 为移位输出有效标志；first 为第 1 组移位输出标志；sign 为移位输出为符号位的标志。为提高速度，Pre 采用 2 级流水线结构，即第 1 级流水线完成 24 个输入数据 x_0~x_{23} 接收；第 2 级流水线完成预加减、32 次移位操作。这使移位输出与接收下一组数据协调进行，时序更加紧凑。

4.2.2 累加单元 RAC

RAC 完成 12 个 MFCC 参数的计算，由 12 个并行 RAC 子单元来实现，以差值为地址查表得到 $MFCC_1$~$MFCC_{11}$，以和值为地址查表得到 $MFCC_2$~$MFCC_{12}$。在本设计中，将 RAC 子单元的 12 bit 查表地址分为 3 组，每组再采用偏移二进制编码减小 ROM 表大小，结构如图 3 所示。

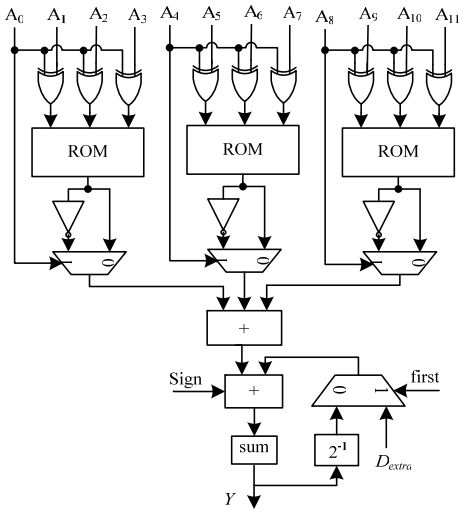

图 3 RAC 子单元结构

将每组中最低位与其他位的异或作为真正的查表地址，同时该位也是查表的原值、负值输出的选择信号，将 3 组查得的值相加才得到真正的查表值。

查表累加的工作过程为：(1)当 Pre 第 1 组移位输出时，first=1，将累加初始值 D_{extra} 与查表值相加；(2)当 Pre 第 2 组~第 31 组移位输出时，上次累加结果 sum 算术右移一位并与查表值相加；(3)当 Pre 第 32 组移位输出时 sign=1，同时将上次累加结果 sum 算术右移一位并减去查表值，得到的 Y 值即为 MFCC 的值。

5 仿真及 FPGA 验证

5.1 功能仿真

图 4 为调用 ModelSim 进行的功能仿真。

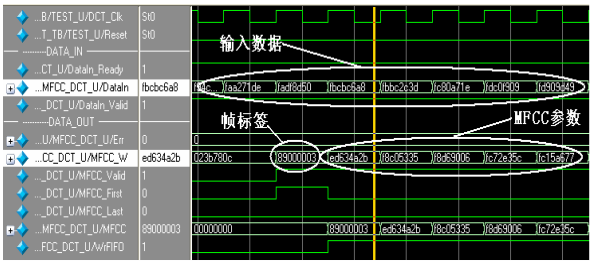

图 4 ModelSim 中功能仿真

遵循自顶向下的设计方法，在 Quartus II 7.1 中对本设计的电路结构用 Verilog HDL 进行编译，并调用 ModelSim SE 6.1f 进行功能仿真。为检验以上提出的硬件算法的正确性和精度，用真实语音数据得到的多组 24 点数据进行测试，并与 Matlab 中采用浮点的计算结果相比较。

5.2 时序仿真

本系统采用 Altera 公司 Stratix Ⅱ系列的 FPGA 实现说话人识别算法，所选芯片为 EP2S60F1020C3。在 Quartus II 上进行综合布线，调用 ModelSim 进行时序仿真，其计算结果与功能仿真的计算结果完全相同。消耗资源为：3 116 个 ALUT, 3 842 个 Register, 9 216 bit Block Memery。值得注意的是，$((12/4)\times 2^{4-1})\times 12\times 32=9\ 216$，这与 3.3 节中 ROM 表的大小 $(N/K)2^{K-1}$ 相吻合。

5.3 FPGA 测试

在时钟频率为 100 MHz 的条件下，对本设计进行 FPGA 测试，使用 Quartus II 中的 SignalTap II 对 FPGA 的内部数据进行采集分析。为测试本设计的正确性、精度及鲁棒性，考虑多种可能出现的情况。整个测试所用的数据均为真实语音数据经上级模块处理后的数据，从而使测试更加贴近实际情况。图 5 为使用 SignalTap II 采集 FPGA 内部数据的截图，另外，图 4、图 5 截取的是同一帧数据的处理结果。采集到的是 32 bit 定点二进制补码，小数部分为 23 bit。将采集到的数据，通过 Matlab 转换成实数的形式，并与 Matlab 中双精度浮点计算结果相比较，经对比分析，其误差数量级为 10^{-6}。

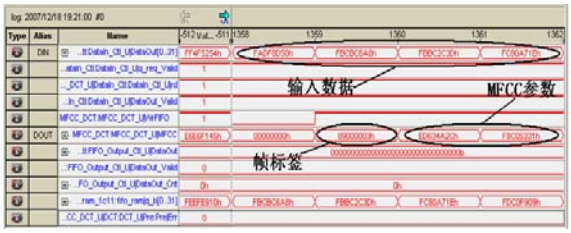

图 5 SignalTap II 采集 FPGA 内部数据

因此，本设计可以在 100 MHz 时钟下稳定工作，其计算误差数量级为 10^{-6}，能够满足 MFCC 提取的实时性以及计算精度要求。

6 结束语

针对 MFCC 中 DCT 的特点，采用 DA 算法结构加以实现，综合运用 ROM 分解和偏移二进制编码来改进 DA 算法，将 ROM 表的大小由 2^N 减小到 $(N/K)2^{K-1}$，同时给出其硬件实现结构，并对本设计进行仿真以及 FPGA 测试，对误差原因进行分析。实验结果验证了本设计的正确性，满足说话人识别中 MFCC 参数提取的实时性要求和精度要求。

参考文献

[1] 王炳锡. 实用语音识别基础[M]. 北京: 国防工业出版社, 2005.

[2] 吴乐南. 数据压缩[M]. 南京: 东南大学出版社, 2005.

[3] Seung C. Efficient ROM Size Reduction for Distributed Arithmetic[C]//Proc. of IEEE Int'l Symp. on Circuits and Systems. Geneva, Switzerland: [s. n.], 2000.

编辑 陈 文