

嵌入式基于 GMM 说话人确认系统的研究

林江云, 张彩虹, 黄关维

(厦门大学 信息科学与技术学院计算机系 福建 厦门 365001)

【摘 要】: 本文主要介绍在嵌入式系统上进行说话人确认的研究。本文中的说话人确认系统采用提取 MFCC 和基于 GMM 模型匹配方法建立起来的。我们在文中着重介绍如何将该系统从 PC 机上移植到嵌入式系统。

【关键字】: 说话人识别; 说话人确认; MFCC 提取; GMM; 定点运算; 嵌入式系统

0. 引言

说话人识别也称声纹识别,属于生物特征识别技术的一种。说话人识别又可以分为说话人辨认和说话人确认。前者用以判断某段语音是已知若干人中的哪一个所说的。而后者是用以确认某段语音是否是说话人所声称的身份,是“一对一判别”问题。一个典型的说话人确认系统建立主要包括 3 个阶段: 前端预处理,训练说话人模型和模型匹配。如图 1 所示^[1]。前端处理又可以分成语音信号的采集和特征向量的提取。

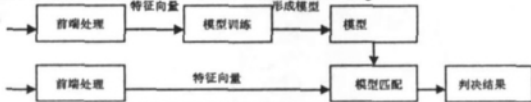


图 1 语音识别系统框图

以前说话人确认系统大多数都是运行在 PC 机上的,随着后 PC 时代出现了信息电器,如掌上电脑,可视电话,数码相机,移动电话等嵌入式设备随着信息技术的高速发展,在嵌入式系统上(ARM)实现说话人确认系统成为一个新的发展方向。该研究应用包括实时服务系统,声控操作设备,安全模型电子设备等。

1. 系统设计

系统由华恒科技教学实验平台 HHARM9-EDU 的核心板和底板组成。核心板上集成了 Sumsung S3C2410 处理器,16M FLASH 和 64M SDRAM。S3C2410 16/32-bit RISC 采用 203MHZ 的 ARM920T 内核,集成了 16KB 指令缓存和 16KB 数据缓存,利用 MMU 实现对虚拟内存的管理,LCD 控制器支持 STN 屏或 FTT 屏支持 NAND FLASH。底板配有 A/D, D/A 电路,控制电路,接口电路,存储电路,四线 RS-232 串口(COM1)一个,音频输入输出接口和麦克风接口,从而实现说话人识别的训练和识别过程。

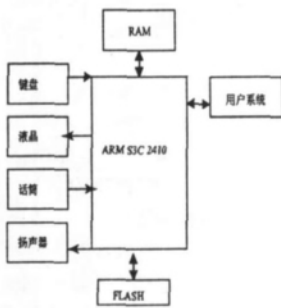


图 2 系统硬件框图

系统控制^[2]由 ARM920T 处理器负责控制语音信号的特征提取和训练算法。语音信号训练的结果数据存放在 16M FLASH 中,将采样数据存放在 64M RAM 中。系统硬件框图如图 2 所示:

3. 系统软件设计

3.1 GMM 模型

GMM (高斯混合模型)本质上是一种多维概率密度函数,GMM 目前最常用于与“文本无关的说话人识别”,具有良好的识别性能^[3]。一个具有 M 个混合成分的 D 维 GMM,可以用 M 个高斯成员的加权和来表示,即:

$$P(x_i|\lambda) = \sum_{i=1}^M w_i P[x_i|\mu_i, \Sigma_i] \quad (1)$$

其中, x_i 是一个 D 维观测矢量, $i(i=1,2, \dots, M)$ 为混合权值,相当于每个高斯成员出现的概率,且 $\sum_{i=1}^M w_i = 1$, $P[x_i|\mu_i, \Sigma_i]$ 为 D 维高斯函数,即:

$$P[x_i|\mu_i, \Sigma_i] = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i)\right\} \quad (2)$$

式中, μ_i 为均值矢量, Σ_i 为协方差矩阵。这里共有 M 个高斯分布函数,其参数为 μ_i 和 Σ_i 。每个函数受 w_i 加权后,取和得到 x_i 的概率分布。

因此整个高斯混合模型使可以由各均值矢量,协方差矩阵及混合分量的权值来描述,于是我们将一个 GMM 模型参数,表示如下三元式:

$$\lambda = \{i, \mu_i, \Sigma_i\} \quad i=1,2, \dots, M \quad (3)$$

其中协方差矩阵 Σ_i 可以用满矩阵,在实现中为减少运算量,可以将其转化成对角矩阵,给运算带来很大的方便。

3.2 特征参数提取

常用于说话人识别的语音特征参数主要有 LPC 和 MFCC 参数,两者都属于倒谱系数。根据汉语测试表明,MFCC 参数表现比 LPC 参数要好。而且 MFCC 更精确地描述了人对声音频率的感知特性。MFCC 计算过程如图 3 所示:

预处理语音输入



图 3 MFCC 提取框图

3.3 模板匹配和识别判决

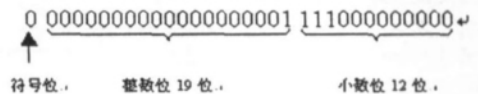
在识别阶段,设集合内有 S 个讲话人,对应的说话人模型为 M_1, M_2, \dots, M_S 。对于待测的说话人语音特征序列,分别对单个声称说话人模型和背景模型进行打分,这里采用对数似然比,然后计算两种似然比的差值,将差值与预定义的判定阈值进行比较来决定确认结果。表达式如下:

$$\sum_{t=1}^T \log [P(x_t|M_k)] - \sum_{t=1}^T \log [P(x_t|U)] \quad (4)$$

其中: $x_t(t=1,2, \dots, T)$ 表示特征矢量, M_k 为声称说话人模型, U 为背景模型。

3.4 浮点动算转化定点实现处理

一般处理器处理浮点数基本上把它分为指数和尾数来表示,然后再在这个基础上进行加减乘除处理。但是在嵌入式系统中,处理器一般是定点处理器,没有浮点运算单元,浮点运算通常用定点运算模拟。而这样会影响速度,因此要采用定点实现系统^[4]。在定点处理器上用 32 位的数来表示一个浮点数^[5]。32 位定点数表示浮点数主要有三部分组成:符号位,整数位,小数位。其中分配给整数的位数个数称为整数字长(IWL)其决定值的大小,分配给小数的位数个数称为小数字长(FWL)其决定量化步骤。例如,一个浮点数用定点数表示中。小数(FWL)字长为 12,则整数字长(IWL)=32-1-12=19。则浮点数 1.875 转化为定点数表示为:



4. 仿真与分析

本文实验语音数据库分成 5 男 5 女,采样率为 16KHz,字长 16 位,帧长 256,帧移 128。录音 48 秒,前 24 秒用于训练,后 24 秒用于识别。MFCC 取 12 阶。高斯混合数为 3。实(下转第 150 页)

问方法的支持即可。实现的方式可以多种多样,但是向外界暴露的 API 接口,必须是符合 XML 数据模型规范中的定义的。

我们这里主要利用数据模型隔离 XQuery 和 XML 存储之间的依赖关系,事实上,使用数据模型,还可以隔离查询语言与 XML 文档之间的依赖关系,换言之即当我们采用不同于 XQuery 的查询语言来查询 XML 时,如果查询是基于数据模型实现的,那么 XML 数据的存储方式可以不加任何改变。

3.2 数据模型方法中 Xquery 的执行流程

XQuery 语言的表达式首先经词法分析、语法分析后得到表层语法树,表层语法树要由 XQuery 查询处理器为核心语法,核心语法中访问实际的 XML 数据的操作通过 XQuery 数据模型规范中的访问器以及 XQuery 函数与操作规范中的函数与操作进行,以关系形式存储的 XML 数据的 RDBMS 和 XQuery 处理器之间的接口负责执行涉及到数据访问的访问器和函数调用,XQuery 处理器接受关系数据库查询的结果,转换为 Sequence 类型,完成涉及数据访问的操作。

3.3 基于模式方法支持 XQuery 的系统结构图

使用基于模式方法支持 XQuery 的系统结构图,如图 1 所示:



图 1 基于模式方法支持 XQuery 的系统结构图

XQuery 的系统结构图

对图 1 的解释如下:

- 1) 一个 XML 文件,如果没有模式定义,则由模式推断算法推导出一个模式,如果有则直接使用。
- 2) 对 XML 模式进行 DTD 简化。
- 3) 对简化后的 DTD 图根据转换规则生成关系模式。
- 4) 将 XML 数据存储到生成的关系模式中。
- 5) XML 模式、XML 数据和 XML 视图分别表示存储的 XML 模式、存储的 XML 数据和生成的 XML 模式。这些信息需要提供给 XQuery 关系接口,该接口实现 XML 的数据模型访问器和函数与操作。在接口中要维护接口状态。
- 6) 关系接口提供的数据模型访问器,由 XQuery 执行机构调用。

4. 系统的验证

以下是一个基于模型存储的 XML 支持 XQuery 的例子:

1) XML 文档

```

<book year="1994">
  <title>TCP/IP Illustrated</title>
  <author><last>Stevens</last><first>W.</first></author>
  <publisher>Addison- Wesley</publisher>
  <price>65.95</price>
</book>
<book year="1992">
  <title>Advanced Programming in the Unix environment</title>
  <author><last>Stevens</last><first>W.</first></author>
  <publisher>Addison- Wesley</publisher>

```

(上接第 127 页)

验结果表明该系统比基线系统(base line)快 10 倍的识别速度。

5. 结论

本系统在嵌入式系统资源受限的环境下,采用优化算法,达到了较好的识别效果,通过对该系统的测试,系统识别率较高,能够满足实际应用的需要。

参考文献:

1. 易克初.语音信号处理[M].北京:国防工业出版社,2000.
2. 马忠梅,马广云.ARM 嵌入式处理器结构与应用基础[M].北京:北京航空航天大学出版社,2002.

```
</price>65.95</price>
```

```
</book>
</bib>
```

2) DTD 定义:

```

<! ELEMENT bib(book*) >
<! ELEMENT book(title, author+, publisher, price) >
<! ATTLIST book year CDATA #REQUIRED>
<! ELEMENT author(last, first) >
<! ELEMENT title(#PCDATA) >
<! ELEMENT last(#PCDATA) >
<! ELEMENT first(#PCDATA) >
<! ELEMENT publisher(#PCDATA) >
<! ELEMENT price(#PCDATA) >

```

3) 关系模式

```

book(pk_book, fk_author, publisher, price, year)
author(pk_author, last, first)

```

4) XQuery 语句

```

<bib>
FOR$b IN document(/bib.xml)//book
WHERE $b/publisher="Addison- Wesley" AND $b/year>1991
RETURN
  <book year=$b/@year>
    $b/title
</book>
</bib>

```

5) 查询结果

```

<bib>
<book year="1994">
  <title>TCP/IP Illustrated</title>
</book>
<book year="1992">
  <title>Advanced Programming in the Unix environment </title>
</book>
</bib>

```

在上面的实例中,基于 DTD 文件,首先生成关系模式,然后将数据存储到生成的关系中。在查询运行时,将查询的结果以数据模型需要的 Sequence 的形式,返回给 XML 查询处理器,再由 XQuery 处理器绑定到相应的变量。该实例证明我们的研究方法是可行的。

5. 结束语

XML 数据库是一个较新的研究领域。本文在实现 XML 数据存储和查询方面进行了讨论,但这对于整个 XML 数据库系统的研究只是很小的一部分,有待做更深入和广泛的研究。

参考文献:

1. El- Sharkawi, El- Hadi El Tazi. LNV: relational database storage structure for XML documents[J]. Computer Systems and Applications, 2005. The 3rd ACS/IEEE International Conference on 2005: 49
2. 蔡小芳,张永胜.关系模式与 XML 模式的相互转换[J]. 计算机系统应用,2005 (10): 83-85
3. Vakai, Catania, Maddalena. XML data stores emerging practices [J]. Internet Computing. 2005, 9 (2): 62 ~ 69
4. Kroenke. Beyond the relational database model [J]. Computer. 2005, 38 (5): 89 ~ 90
5. 谢荣传,谢铨洋. XQuery 语言的实现 [J]. 安徽大学学报. 2005, 29(1): 23-26

3. Reynolds D.A., Speaker identification and verification using Gaussian mixture speaker models Speech Commun.17(1995),91- 108.
4. K.K. Shin, J.C.H.Poon, K.C.Li,"A Fixed- point DSP Based Cantonese Recognition System,"Proceedings of IEEE International Symposium on Industrial Electronics,PP.390- 393,1995.
5. Y.S.Moon, F.T.Luk,T.Y.Tang,K.C.Chan,C.W.Leung, "Fixed - Point Arithmetic for Mobile Devices- A Fingerprint Verification Case Study,"Proceedings of the SPIE 2002 Seattle, Advanced Signal Processing Algorithms, Architectures, and Implementation XII, vol.4791, PP.144- 149.