一种自学习的特定人语音识别驾驶助手设计方法

技术领域

本发明涉及计算机应用和汽车技术领域，具体涉及一种基于语音识别、云计算和机器学习的汽车控制的特定人识别方法

背景技术

近几年来，语音识别技术有了较大的突破，已经出现了较为成熟的商用化的语音助手，如苹果公司的Siri，谷歌公司的Google Now，微软公司的Cortana等。而在国内，各种语音助手也是层出不穷，如智能360、小i 机器人、虫洞和灵犀等，为各种设备和装置的人机交互提供了便利条件。然而在汽车上，人机交互还是通过旋钮、按键及触屏来实现的，驾驶员还不能用自然语言直接发出各种操控命令。其主要原因是现有的语音识别系统还不能足够准确地识别说话人的身份，不能准确无误地区分驾驶员与乘客发出的语音信号，如果对乘客发出的与行车安全密切相关的语音操作命令（如加速、制动等），汽车控制系统也会执行，就很容易引发交通事故，给驾驶员、乘客和行人的安全造成极大的威胁，这是绝对不允许的。

专利申请号为201410435118的技术方案公开了一种基于声纹识别的特定人控制汽车助手，虽然免除了手动控制的麻烦，但是存在声纹信息库是预先设定、识别精度不能调整、只能本地语音控制（即坐在驾驶室内进行语音控制）等缺陷。本发明结合云计算和机器学习这两种计算机应用技术，不仅能动态更新声纹信息库，而且能够利用云端强大的计算能力和存储能力不断提高识别精度，并且若结合无人车技术可在远端进行语音控制，进而准确、高效的为用户提供进一步服务。

发明内容

针对现有技术的上述问题，本发明要解决的技术问题是提供一种能够动态添加声纹识别训练素材、不断自学习来进一步提高声纹识别和控制指令识别精度、支持远程控制，以及更为安全、更为方便的汽车驾驶语音助手解决方法。

为了解决上述问题，本发明采用的技术方案为一种基于语音识别、云计算和机器学习的汽车控制的特定人识别方法，其实施步骤如下：

⑴建立声纹训练模型：采集特定人的发音语音作为声纹模型的训练素材，然后对这些特定人的语音素材进行预处理，预处理之后提取语音中的声纹特征参数，利用这些特征参数值对声纹模型进行训练。由于考虑到人在各种状态下发声可能有些许的而不同，如高兴时声音轻快、痛苦时可能比较沉重、感冒时鼻音比较重，在训练时最好把各种情况都考虑到，录制各种状态下的语音，对声纹模型进行训练，提高声纹识别的准确度。

⑵建立指令识别训练模型：采集特定人发出的控制语音作为指令识别模型的训练素材，然后对这些特定人的语音素材进行预处理，预处理之后提取语音中跟语音指令识别有关的特征参数，然后利用这些特征参数值使用有监督的机器学习算法对指令识别模型进行训练。由于考虑到人在各种状态下发声可能有些许的不同，如高兴时声音轻快、痛苦时可能比较沉重、感冒时鼻音比较重，在训练时最好把各种情况都考虑到，录制各种状态下的语音，对指令识别模型进行训练。

⑶利用自然语言处理和机器学习对语音指令进行分析：采集测定人发出的语音指令，首先对语音指令进行预处理，然后提取出该语音信息的声纹特征，接着采用声纹模式匹配算法，如概率统计、动态时间规整、矢量量化、隐马尔科夫模型方法或人工神经网络方法来进行声纹模式匹配。如果判定的结果为预先设定人发出的语音信息，则将该片段语音信息交由语音指令识别模型进行处理。语音指令识别模型利用自然语言处理技术和机器学习算法（隐马尔科夫方法、支持向量机、深度神经网络等）把获取的语音指令转为相应的操作，通过云端将指令回传给汽车接收端，接收端将获取的指令交由执行单元操作控制汽车的驾驶。

本发明具有以下技术效果：本发明结合云技术强大的处理、运算和存储能力，能够将声纹模型和指令识别模型的训练过程在云端进行处理，从而避免个人终端运行速度慢、执行时间长、指令操作反应过度延迟，并且支持特定人随时随地录制自己的声音，控制指令训练声纹模型和指令识别模型。由于识别模型的训练过程采用的是机器学习算法，因此训练的样本越多，训练的次数越多，模型的实际执行效果越好，识别的准确度也越高，起到一个很好的自学习效果。实验数据表明，通过这种大数据、大样本的反复训练，能够将模型的正确率趋近于百分之百。

本发明若与无人车技术相结合，将能够替代目前的滴滴打车之类的软件，通过特定人声纹信息绑定，通过强大的云端技术、移动网络和 LBS 定位功能，能够将车自动开到特定人身边，并且载特定人去想去的地方，掀起一场新的社会变革，市场价值不可估量。

附图说明

图 1 是本发明基于自学习的特定人识别的语音驾驶助手的实施例的原理框图。

图 2 是本发明中双层声纹识别模型示意图

图 3 GMM-UBM 系统框图

图 4 HMM-SVM 的语音识别系统结构

具体实施方式

以下结合附图及实施例对本发明作进一步说明。

1.建立声纹训练模型：由于一般系统随着目标说话人个数的增加很难满足实时性的要求，因此这里采用一种基于VQ-VPT(Vector Quantization-Vantage Point Tree)和GMM-UBM(Gaussian Mixture Model-Universal Background Model)相结合的双层识别声纹模型.在第一层识别模型中，采用基于VQ-VPT模型进行快速匹配，挑选出与测试者声纹特征最相近的K个目标说话人声纹模型。在第二层识别模型中，采用GMM-UBM模型，精确匹配上层模型得到的K个目标说话人声纹模型，并做出最终的判决。这种声纹模型不仅识别准确度高，而且速度快。双层识别模型将决策分为两步进行，首先对待识别的特征向量进行一次快速匹配，挑选出与其最接近的K个声纹模型，淘汰掉不可能的声纹模型，减小第二步的计算量。然后将得到的K个声纹模型进行精确匹配，得出最终结果。 在第一层模型中采用基于VQ的方式将目标说话人的声纹模型构建为码书的形式,并采用VPT的形式将码书中的码字索引为平衡二叉树的结构，快速识别的过程类似矢量量化，但决策依据并不是量化误差，而是查询与测试特征向量最近的若干个码字中哪一个码书的命中率最高，以此挑选出K个声纹模型。然后利用GMM-UBM计算

K个声纹模型的似然度, 选择似然度最大的声纹模型作为识别结果.

1.1基于 VQ-VPT 的快速识别模型的训练过程

具体的训练步骤如下：

①对目标说话人训练语音进行特征提取, 获得特征向量集.

②对特征向量集进行训练, 生成代表每个目标说话人声纹特征的码书.

③采用VPT 的构建算法, 对所有目标说话人码书中的码字进行索引, 生成VPT.

采用GMM-UBM的精确识别模型，即高斯混合模型-通用背景模型思路是UBM由所有目标说话人的语音训练得到,代表所有说话人的声纹特征, 实质上是一个大型的GMM. 一般而言, UBM 的训练语句时长为1 个小时,混合度为1024. 通常目标说话人的训练数据是有限的,训练得到的GMM 不能真实反映说话人特征. 必须还要采用基于MAP(Maximum A Posteriori)自适应UBM 得到目标说话人的GMM, 用来弥补训练数据的不足。.

2.建立指令识别训练模型：在指令识别过程中，最为关键的模块就是语音特征提取和相似性度量，相似性度量通过分类器来实现，本发明采用Mel频率倒谱系数(Mel Frequency Cepstrum Coefficient，MFCC)提取语音特征参数。采用HMM和SVM对语音特征相似性进行度量，即语音指令识别。HMM—SVM的语音识别算法包括两个训练和识别两个阶段。每一个阶段都要进行语音信号预处理和特征提取过程。语音自动识别通过计算机来实现，因此首先需要将采集的语音模拟信号转换为计算机能够处理的数字信号，然后对语音信号进行滤波处理，防止信号混叠干扰和抑制50Hz的电源干扰。最后进行端点检测和预加重，语音信号加重是语音信号加以提升，保证提取的语音特征可以准确反映人的声道参数。

指令识别模型的训练过程如下：

1)对语音信号进行采集并转换成为计算机能够识别的数字信号。

2)对语音信号进行滤波、端点检测和加重处理，消除一些干扰信息。

3)通过MFCC对语音信号的特征参数进行提取，并去掉一个冗余和噪声数据。

4)采用HMM模型对语音信号进行训练。得到HMM参数库。

5)对输人语音信号采用HMM进行时序处理，并通过Viterbi算法获得该语音信号的输出概率矢量(v)。

6)对输出概率矢量(v)进行归一化处理．具体公式为：

$$u_i = \frac{v_i - v_{min}}{v_{max} - v_{min}}$$

7)将归一化处理后的输出概率矢量(v)输入到SVM进行学习．直到满足SVM收敛精度满足要求为止。此时HMM—SVM算法的训练阶段完成。

3.利用自然语言处理和机器学习对语音指令进行分析：首先采集语音信号，对其声纹信息进行验证，看他是否是预设的特定人。由于本发明采用的是双层声纹识别模型，因此声纹信息的识别过程分下面两部进行，如图2所示。

3.1、快速识别模型的识别过程如下:

① 初始化Scores[i] = 0 , i =1,2,...,n

②对测试语音进行特征提取, 获得特征向量集.

③特征向量集中选取一个特征向量, 在VPT 中查找与其距离最近的M 个码字.

④对M 个码字分别查找其对应的码书，并对其所对应的目标说话人进行加分，Scores[i] = Scores[i]+1.
⑤重复③至④直到遍历完测试特征向量集中的所有码字.
⑥在Scores[i]中挑选出得分最高的 K 个目标说话人用于精确识别.

3.2、精确识别阶段对在快速识别阶段筛选出的K 个最可能的说话人模型采用GMM -UBM 进行精确识别.GMM-UBM 的系统框图如图3所示:
由上图3可知，测试语音的特征矢量序列X = {Xt },t =1,2,...,T 的对数似然比可以由式(4)来计算:

$$S(X)=\frac{1}{T}\sum_{t=1}^{T}\left(log\left(p\left(X_t \mid \lambda_s\right)\right)-log\left(p\left(X_t \mid \lambda_{UBM}\right)\right)\right) \tag{4}$$

其中,λ_s是目标说话人的GMM模型参数,λ_{UBM}是UBM的模型参数。采用似然比的方式打分是一种归一化处理，可以对不同的目标说话人设置统一的判决阈值. 在精确识别时，分别计算测试特征向量与快速识别阶段筛选出的K个目标识别模型之间的相似度，并选取最大似然度值所对应的目标说话人模型作为识别结果. 由于不需要计算测试特征向量与所有N个说话人模型的似然度，减小了计算量，提高了识别速度.

在声纹识别为预设的特定人之后，进行指令识别，由于本发明采用的是HMM—SVM混合语音指令识别，具体步骤流程如图4所示。将待识别语音信号输入到HMM系统，得到Viterbi评分，然后再由SVM对Viterbi评分进行非线性映射，得到该语音识别信号，最后由两次识别信息综合完成语音指令识别过程。云端将识别的语音指令通过移动网络回传给汽车上的接收设备，由接收设备将指令信号交由执行单元进行处理。

综上所述，本实施例通过预先训练声纹识别模型和语音指令识别模型，然后将采集端的语音信号进行预处理，提取出特征信息，对其先进行声纹识别，然后再进行语音指令识别，最终将识别结果通过移动网络传送给执行单元进行执行。

以上所述仅是本发明的优选实施方式，本发明的保护范围并不仅局限于上述实施例，凡属于本发明思路下的技术方案均属于本发明的保护范围。应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理前提下的若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

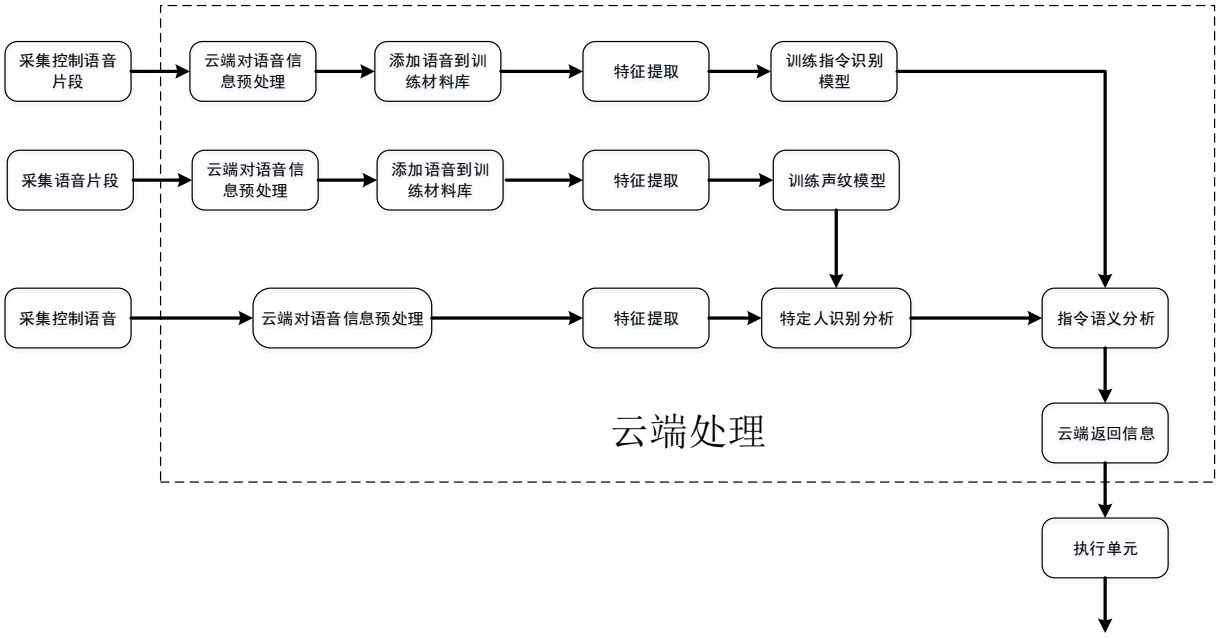

图 1 实施例的原理框图

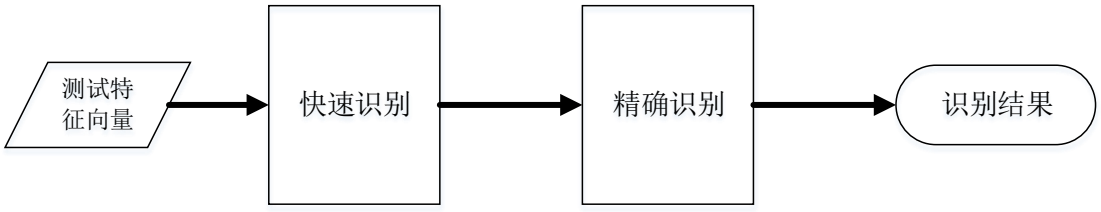

图 2 双层识别模型示意图

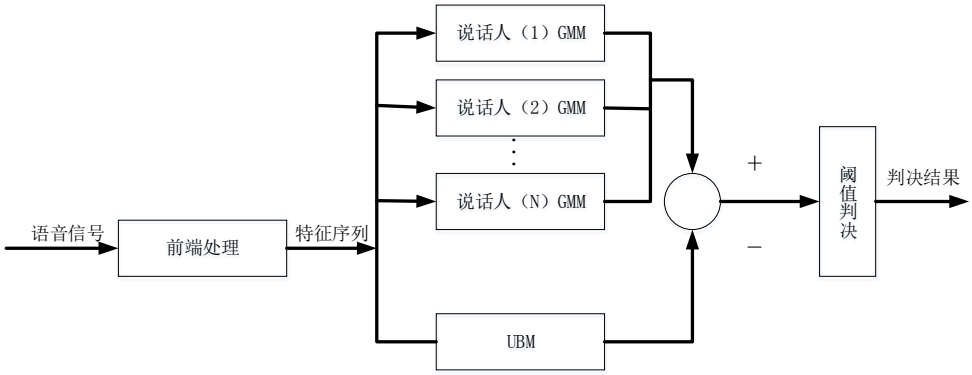

图 3 GMM-UBM 系统框图

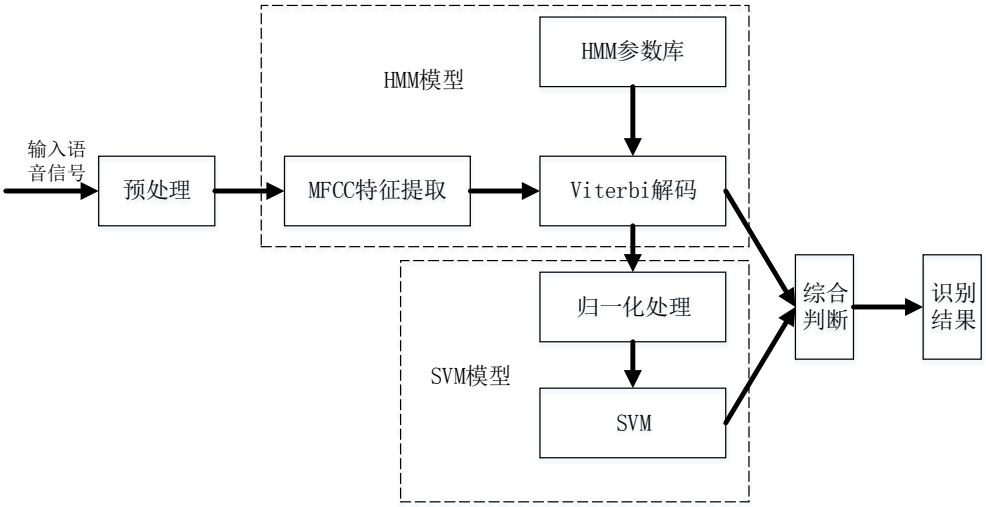

图 4 HMM-SVM 的语音识别系统结构