

◎数据库、信号与信息处理◎

一种基于K-SVD的说话人识别方法

马 振¹, 张雄伟², 杨吉斌²MA Zhen¹, ZHANG Xiongwei², YANG Jibin²

1.解放军理工大学 通信工程学院, 南京 210007

2.解放军理工大学 指挥自动化学院, 南京 210007

1.Institute of Communication Engineering, PLA University of Science & Technology, Nanjing 210007, China

2.Institute of Command Automation, PLA University of Science & Technology, Nanjing 210007, China

MA Zhen, ZHANG Xiongwei, YANG Jibin. Speaker recognition method based on K-SVD. Computer Engineering and Applications, 2012, 48(34): 112-115.

Abstract: In order to extract the personal characteristics, a speaker recognition method based on K -means Singular Value Decomposition (K -SVD) is proposed. The personal characteristics in voice can be well preserved in the dictionary trained from the K -SVD. With this feature, the dictionary which contains the personal characteristics is extracted from training data through the K -SVD algorithm. Then the trained dictionary is used for the speaker recognition. Compared to traditional methods, the personal characteristics in voice can be better preserved based on the proposed method through the sparse nature of voice and can reduce the reconstruction error. Experimental results show that the proposed method outperforms the VQ based methods for too many speakers in the view of recognition rate, so the proposed method has more practical value.

Key words: speaker recognition; K -mean Singular Value Decomposition (K -SVD); dictionary; sparse

摘 要: 为了充分提取语音中的个人特征信息, 类比矢量量化, 提出了一种基于 K -均值奇异值分解(K -SVD)的说话人识别方法。利用 K -SVD训练得到的字典可较好地保存语音信号中的个人特征信息。利用这一特性, 通过 K -SVD从训练数据中提取包含说话人个人特征信息的字典, 利用该字典实现说话人识别。相对于传统方法, 该方法能够更好地利用语音的稀疏性保存语音中的个人特征信息并减小重构误差。实验仿真结果表明, 与基于矢量量化的说话人识别方法相比, 该方法在多说话人的情况下具有更好的识别率, 具有更高的实用价值。

关键词: 说话人识别; K -均值奇异值分解(K -SVD); 字典; 稀疏性

文献标识码: A **中图分类号:** TN912.3 **doi:** 10.3778/j.issn.1002-8331.1207-0400

1 引言

说话人识别根据输入语音确定发音者的身份, 即用待识别语音和预先提取的说话人特征来确定或鉴别说话人的身份^[1]。说话人识别技术在信息安全及多媒体娱乐等领域都有着广阔的应用前景。

当前较为常用的说话人识别技术有以下几种: 矢量量化(VQ)^[2], 高斯混合模型(GMM)^[3], 隐性马尔

可夫模型(HMM)^[4], 人工神经网络(ANN)^[5]等。其中, 矢量量化法能够大大压缩语音信息量; 而且无需考虑复杂的统计模型和时间归整问题, 其运算过程也较为简单, 因此在说话人识别领域有着广泛的应用。但是由于其大大压缩了语音信息量, 减小了个人特征信息, 因而只适用于基于小词汇量孤立词的说话人识别^[6], 对于在说话人数目较多的情况下的识

作者简介: 马振(1989—), 男, 在读硕士研究生, 研究领域为语音转换, 语音识别; 张雄伟(1965—), 男, 博士, 教授, 研究方向为多媒体信息处理, 智能计算机, 压缩感知; 杨吉斌(1978—), 男, 博士, 副教授, 研究方向为信号参数估计, 语音识别和转换, 智能信息处理。E-mail: mazhen1989@126.com

收稿日期: 2012-07-25 **修回日期:** 2012-08-27 **文章编号:** 1002-8331(2012)34-0112-04

别率大大降低。一些学者提出了矢量量化的改进算法^[7],虽然识别率有了一定程度的提高,但对于在说话人数目较多的情况下的识别效果并不是十分理想。

矢量量化对语音信号的压缩率很高,语音中的相关个人特征信息保留的不够充分,这些特征在对多人语音进行说话人识别时其性能会受到影响。为了能保留更多的语音个人特征信息,适当减小对语音信号的压缩率,可以考虑将矢量量化进行扩展。矢量量化方法是将提取到的语音特征利用一个码本来表示,稀疏度为1。而语音信号是多变的,单个小尺度的码本设计并不能包含所有的语音变化。如果利用信号稀疏表示方法,增加稀疏度,利用多个码本的组合来表示语音特征,语音的个人特征信息将更加充分,识别效果将得到提高。

信号稀疏表示的基本思想是在某些基函数的基础上,对信号实施线性展开,从而实现利用较少的特征数据(即基函数)来表示信号,选择使用的基函数称为原子。信号稀疏表示方法将信号表示为原子的稀疏组合,其中大部分组合系数为零,只有少数的较大非零系数,具有非零系数的原子(即信号的稀疏成分)揭示了信号的主要特征与内在结构^[8-9]。

K-均值奇异值分解(K-means Singular Value Decomposition, K-SVD)是一种性能优良的信号稀疏分解方法,在语音信号去噪^[10]、图像去噪、特征提取等方面已有成功应用^[11]。该方法可以将语音声道谱分解为一个字典和对应的稀疏矩阵,分解得到的字典可认为是承载了信号特征的一个子空间,而稀疏矩阵则是声道谱参数在此子空间上的投影^[12-13]。

基于以上考虑,根据语音信号的稀疏表示和说话人识别中语音的特点,本文采用K-SVD算法^[14]进行说话人识别。

2 K-SVD算法

一个字典矩阵 $D \in R^{n \times K}$ 共有 K 个列向量 $\{d_j\}_{j=1}^K$, 它的每列表示一个原子信号的原子量。假设信号 $y \in R^n$ 能够表示成这些原子的线性组合

$$(P_{0,x}) \min_x \|x\|_0 \text{ s.t. } \|y - Dx\|_p \leq \varepsilon \tag{1}$$

其中向量 $x \in R^K$ 包含表示信号 $y \in R^n$ 的系数,约束条件中的范数 p 可以选择为1、2范数或无穷大范数。

通常 $n < K$, 同时 D 是满秩矩阵,此时对这种表示问题有无穷多个解,在不同的约束条件下,可以得到不同的优化解,其中包含最少非零系数的解是最

优的表示方法。

基于K-SVD的稀疏表示方法就是在式(1)的基础上,构建一个目标函数,针对目标函数进行最优化解,从而获取信号的稀疏化表示。该算法是在如下目标函数的基础上展开的:

$$\min_{D,X} \|Y - DX\|_F^2 \text{ 依据 } \forall i, \|x_i\|_0 < T \tag{2}$$

其中, Y 为需要进行稀疏表示的信号集, $Y = \{y_i | i \in [1, K], y_i \in R^n\}$, $X = \{x_i\}$ 是一组列向量集,它的每一个元素均为 Y 中某元素(信号)的稀疏表示。 $\|\cdot\|_F^2$ 表示Frobenius范数的平方。

K-SVD算法对目标式(2)进行迭代计算。假设 D 是已知且固定的,那么式(2)则演变成为求取稀疏表示的系数矩阵集 X , 变为如下形式:

$$\|Y - DX\|_F^2 = \sum_{i=1}^N \|y_i - Dx_i\|_2^2 \tag{3}$$

因此式(3)也可分解成为 N 个互不相同的计算问题:

$$\min_{x_i} \{ \|y_i - Dx_i\|_2^2 \} \text{ 约束条件 } \|x_i\|_0 < T_0, i = 1, 2, \dots, N \tag{4}$$

假设 X 固定, D 除了第 k 列 d_k 待求解外,其他均固定已知。设与 d_k 相对应的表示系数向量为矩阵 X 中的第 k 行,记为 x_k^r ,则有:

$$\begin{aligned} \|Y - DX\|_F^2 &= \|Y - \sum_{j=1}^K d_j x_j^r\|_F^2 = \\ &= \|(Y - \sum_{j \neq k} d_j x_j^r) - d_k x_k^r\|_F^2 = \\ &= \|E_k - d_k x_k^r\|_F^2 \end{aligned} \tag{5}$$

对式(5)中的 E_k 直接实行奇异值分解(Singular Value Decomposition, SVD),就可以得到 D , 因此也能计算出 X , 但这样的结果并不能保证 X 是稀疏的。因此,可以对 E_k 进行补偿^[13],得到新的 E_k^r , 使得

$$\|Y - DX\|_F^2 = \|E_k^r - d_k x_k^r\|_F^2 \tag{6}$$

对新的 E_k^r 进行SVD分解得 $E_k^r = U \Delta V$ 。将其中 U 的第一列作为需要求解的量 d_k 的近似值 \tilde{d}_k , 而 V 的第一列乘以 $\Delta(1, 1)$ 后,作为表示系数向量 x_k^r 的解。因而就得到了信号稀疏表示,以及相应的字典矩阵。如果稀疏度 $L = 1$, 此种方法即为K-means方法。

3 基于K-SVD的说话人识别

由以上分析可以知道,通过K-SVD对语音信号进行学习训练,可得到承载说话人个人特征信息的

字典及其相应的稀疏矩阵。因此,很自然会想到通过类比VQ,实现说话人识别。

基于K-SVD训练得到原子字典和稀疏矩阵并实现说话人识别的流程如图1中所示。首先提取说话人语音信号的特征参数,之后利用参数进行说话人识别。

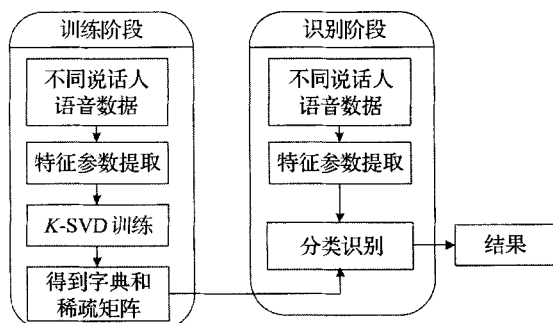


图1 基于K-SVD的说话人识别流程示意图

说话人识别过程可分为训练和识别两个阶段,其中训练阶段由特征参数提取和基于K-SVD的字典提取两部分组成,识别阶段包括特征提取和分类识别两部分。

3.1 特征参数提取

用于训练的语音和用于识别的语音是具有相同说话人的不同的内容。首先需要提取语音的特征参数。常用的语音识别参数有线性预测参数(LPC),线性预测倒谱参数(LPCC)和Mel尺度倒谱参数(MFCC)等。

MFCC参数表示人对声音高低的感受是一种主观感受,用客观度量来表征这种主观感受就采用了Mel标度。Mel滤波器组是一组采用Mel刻度的线性相位FIR带通滤波器,这组滤波器中的每一个中心频率按Mel刻度在讨论的频带上均匀分布,每个滤波器的带宽都为临界带宽。主要是MFCC参数考虑了听觉系统的非线性特点,能够有效地提高系统的性能。已有研究表明使用Mel倒谱参数(MFCC)的识别率确比使用线性预测参数(LPC)的识别率高4个百分点^[15],因此本文选择MFCC(Mel倒谱参数)作为基本识别参数。

3.2 说话人识别

在训练阶段将用于训练的语音经过预处理得到的MFCC参数后,通过K-SVD算法提取用于说话人识别的字典和稀疏矩阵。

在识别阶段首先将用于识别的语音经过预处理得到MFCC参数,然后与训练得到的字典和稀疏矩阵的分类乘积进行比较判断从而进行识别。

4 实验仿真

4.1 实验条件

实验中所用的语音库的数据在普通实验室环境下录制,采样频率为16 kHz,单声道录音,每个采样点采用16 bit量化。整个语音库包含100个说话人,每个说话人两句话,内容不同,长度均为30 s,分别用于训练和识别。

对于基于VQ的说话人识别,参照文献[16],与基于K-SVD的说话人识别方法进行比较。为了便于比较,在实验中令矢量量化方法的码书中的码字与K-SVD算法的字典中原子的个数相同,均为16。

4.2 参数分析精度

图2所示为语音库中的一个原始语音,图3给出了在以上条件下两种训练方法对图2所示语音的重构误差结果。

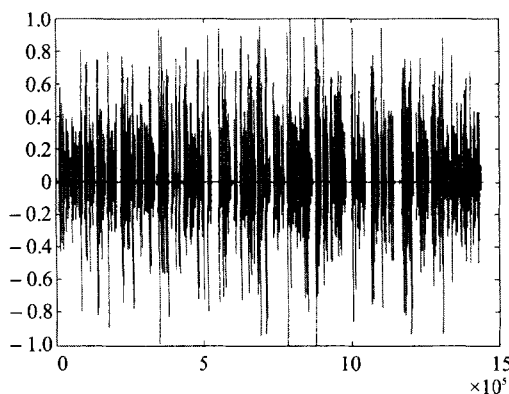


图2 原始语音

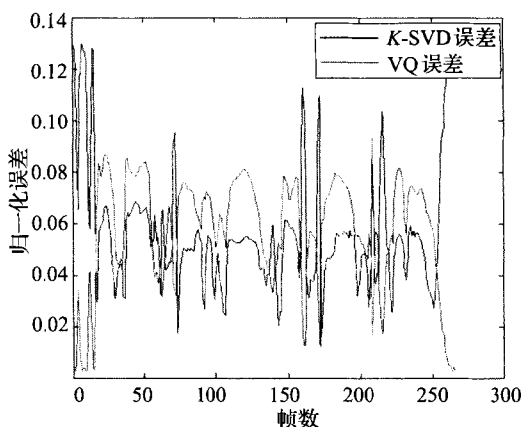


图3 两种不同方法的误差比较结果

通过重构误差结果对比可知,K-SVD方法的重构误差要明显小于采用VQ处理时的重构误差。当利用这两种方法进行说话人识别时,K-SVD得到的字典更准确地包含了与说话人相关的个人特征信息。

接下来将通过实验,研究K-SVD中字典个数对识别性能的影响,并在不同条件下对比本文提出方法与基于VQ的说话人识别方法的性能差异,以验证

本文提出方法的有效性。

4.3 K-SVD 识别的性能

图4给出了字典中原子个数对识别性能的影响结果。仿真结果表明识别性能随着字典中原子个数的增加会有一定的提高。但是为了改善识别效果,训练和识别所用的时间也会相应增多。

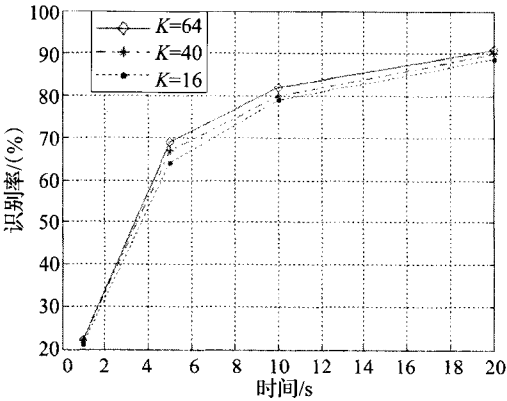


图4 字典中原子个数对识别结果的影响

图5给出了不同识别人数情况下两种方法识别的结果(每句长度为20 s、字典中原子个数 $K=16$)。实验结果说明在说话人数目相对较少的情况下,两种方法的识别率都很高,但随着说话人数目的增多,基于K-SVD的说话人识别方法相对于基于矢量量化方法的说话人识别的识别率会有明显的提升,优势就会显得比较明显。

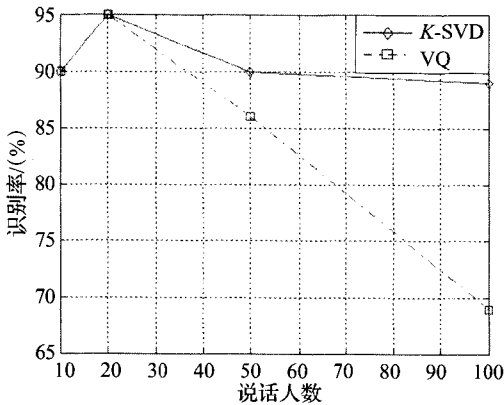


图5 不同识别人数情况下两种方法识别的结果

从图中可以看出,在说话人数为20时,识别率有一定提升,而后再下降。这是由于本实验所选取的语音库中的语音没有一定规律,并且与计算识别率时的句子数有关。

图6给出了相同识别人数(100人)下两种方法不同时长识别的结果。通过仿真结果可知,两种方法的识别率随着语音时长的增加即训练数据量的增加都进一步增加,但随着说话人数目的增加,识别率都有一定的减小。

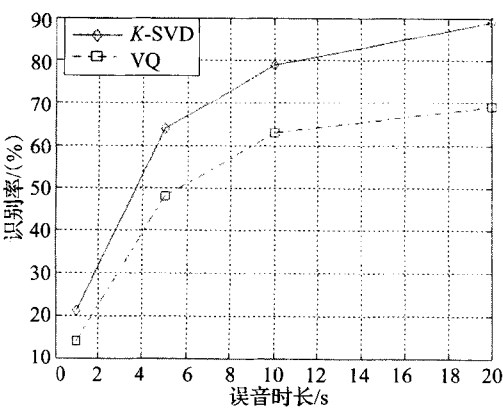


图6 两种识别方法在不同时长的识别率比较

通过以上的仿真实验可知,基于K-SVD的说话人识别方法相对于传统基于VQ的说话人识别方法来说可以获得更好的识别率,且在多说话人的情况下识别结果提高更为明显。

5 结束语

本文提出了一种基于K-SVD的说话人识别方法。由K-SVD训练得到的字典可承载语音的个人特征信息,利用K-SVD算法的这一特点,本文通过字典训练实现说话人识别。实验仿真结果表明,本文提出的方法在说话人识别率等性能上优于基于VQ的说话人识别方法,这种优势在说话人数目较多的情况下尤为明显,因此具有更高的实用价值。

虽然给出了一种说话人识别的新方法,且得到了较好的实验仿真结果,但应注意到,基于K-SVD的字典学习本质上仍旧与VQ相同,因而可能对语音信号中一些特性表征不足,从而造成识别率有一定的损失。因此,下一步的研究中将尝试在K-SVD中稀疏度对识别效果的影响,以进一步提升识别的效果。

参考文献:

[1] 张雄伟,陈亮,杨吉斌.现代语音处理技术及应用[M].北京:机械工业出版社,2003.

[2] 徐利敏,唐振民,何可可,等.说话人识别中基于聚类特征的矢量量化技术[J].计算机工程与应用,2007,43(27):196-198.

[3] 蒋晔,唐振民.GMM文本无关的说话人识别系统研究[J].计算机工程与应用,2010,46(11):179-182.

[4] 冯松,张述清.隐马尔科夫模型在说话人识别中的应用[J].计算机科学,2006,33(9).

[5] 全学海,丁宣浩,蒋英春.基于EMD和概率神经网络的说话人识别[J].桂林电子科技大学学报,2010,30(2).

小的。

由图6可以看出硬件压缩的时间远远比软件所用的时间少。对比三组测试数据,在压缩时间上数据软件压缩是硬件压缩的将近十几倍,但是压缩率却相差的很少。因此对系统的实时性得到巨大的提高。

6 总结

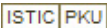
本文将数据压缩理论引入到远程故障诊断领域,最终在FPGA上加以实现,从而解决了故障诊断系统在远程终端产生的巨大数据量而无法有效传输的问题。通过对LZW无损压缩编码的硬件实现,不仅可以使采样数据得到巨大的压缩,而且也可以极大地提高系统实时性,从而为故障的快速诊断,赢得了时间。但是由于时间的原因,本文只实现了数据压缩的核心,也并没有对算法针对FPGA的特点进行优化。接下来的工作将在这方面做进一步的研究和系统的优化,使故障诊断系统的数据压缩效果进一步提高。

参考文献:

- [1] 陈幼平,张国辉.远程故障诊断系统体系结构研究[J].计算机应用研究,2005(12):88-90.
- [2] Taleb S A A.Improving LZW image compression[J].European Journal of Scientific Research,2010,44(3):502-509.
- [3] 王振玺,乐嘉锦.列存储数据区级压缩模式与压缩策略选择方法[J].计算机学报,2010,8(33):1523-1530.
- [4] Jou J M,Chen P Y.A fast and efficient lossless data-compression method[J].IEEE Transactions on Communication,2006,47(9):1278-1283.
- [5] Malvern J.Hardware-based LZW data compression co-processor:United States Patent,US6624762[P].2003.
- [6] Wang Quan,Qi Chun,Luo Xinmin.Modified LZW algorithm and its parameters optimization[J].Journal of Chongqing University of Posts and Telecommunications,2005,17(3):351-355.
- [7] Barr K C,Asanovic K.Energy-aware lossless data compression[J].ACM Transactions on Computer Systems,2006,24(3).
- [8] 张凤林,刘思峰.LZW*:一个改进的LZW数据压缩算法[J].小型微型计算机系统,2006,26(10):1897-1899.
- [9] Klein S T,Wiseman Y.Parallel Lempel Ziv coding[J].Discrete Applied Mathematics,2005,164:180-191.
- [10] Ferrer L,Shriberg E,Kajarekar S S,et al.The contribution of cepstral and stylistic features to SRI's 2005 NIST speaker recognition evaluation system[C]//Proc Int'l Conf Acoust,Speech Signal Process(ICASSP),Toulouse,France,2006:101-103.
- [11] 张庆芳,赵鹤鸣.基于改进VQ算法的文本无关的说话人识别[J].计算机工程与应用,2006,42(10):65-68.
- [12] Zheng H,Hellwich O.Adaptive data-driven regularization for variational image restoration in the BV Space[C]//Proceedings of VISAPP'07, Barcelona, Spain, 2007: 53-60.
- [13] Aharon M.Overcomplete dictionaries for sparse representation of signals[D].Thesis, Computer Science Department, the Senate of the Technion-Israel Institute of Technology,2006.
- [14] Gemmeke J F,Cranen B.Using sparse representations for missing data imputation in noise robust speech recognition[C]//European Signal Processing Conf(EUSIPCO), Lausanne, Switzerland, August 2008.
- [15] Elad M,Aharon M.Image denoising via sparse and redundant representations over learned dictionaries[J].IEEE Transactions on Image Processing,2006,15(2):3736-3744.
- [16] ZHAO Nan,XU Xin,YANG Yi.Sparse Representations for Speech Enhancement[J].Chinese Journal of Electronics,2011,19(2).
- [17] Jafari M G,Plumbley M D.Fast dictionary learning for sparse representations of speech signals[J].IEEE Journal of Selected Topics in Signal Processing, Special issue on Adaptive SparseRepresentation of Data and Applications in Signal and Image Processing,2011.
- [18] Aharon M,Elad M,Bruckstein A M.The K-SVD:an algorithm for designing of overcomplete dictionaries for sparse representation[J].IEEE Trans on Signal Processing,2006,54(11):4311-4322.
- [19] 胡征.矢量量化原理及应用[M].西安:西安电子科技大学出版社,1998.
- [20] 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.

(上接115页)

一种基于K-SVD的说话人识别方法

作者: [马振](#), [张雄伟](#), [杨吉斌](#), [MA Zhen](#), [ZHANG Xiongwei](#), [YANG Jibin](#)
作者单位: [马振, MA Zhen\(解放军理工大学通信工程学院, 南京, 210007\)](#), [张雄伟, 杨吉斌, ZHANG Xiongwei, YANG Jibin\(解放军理工大学指挥自动化学院, 南京, 210007\)](#)
刊名: [计算机工程与应用](#) 
英文刊名: [Computer Engineering and Applications](#)
年, 卷(期): 2012, 48(34)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_jsjgcyyy201234022.aspx