

基于改进 MP 稀疏分解的语音识别方法

高显忠, 侯中喜

(国防科学技术大学 航天与材料工程学院, 长沙 410073)

(bestwolf@126.com)

摘要:在限定输入词汇量的条件下进行语音识别的过程中,结合 MP 稀疏分解的原子结构特性,提出把语音信号经 MP 稀疏分解所得的最佳原子时频参数作为匹配参数对语音进行识别。用基于遗传算法和原子库划分的策略对 MP 稀疏分解的寻优过程进行改进以提高 MP 稀疏分解的效率。在 Matlab 环境下进行仿真试验时,采用 LGB 算法对测试信号进行判别。试验结果表明,MP 稀疏分解算法经改进后运行速度得到提高,采用 10 个原子的时频参数可有效识别长度约为 6 000 的语音信号。

关键词:语音识别;匹配追踪稀疏分解算法;遗传算法;库划分

中图分类号: TN912.34 **文献标志码:** A

Speech recognition based on improved MP sparse decomposition algorithm

GAO Xian-zhong, HOU Zhong-xi

(College of Astronautics and Material Engineering, National University of Defense Technology, Changsha Hunan 410073, China)

Abstract: Based on the limit of input glossaries and the characters of atoms structure in Matching Pursuit (MP) sparse decomposition, a new speech recognition algorithm was proposed through using the best obtained atoms' time-frequency parameters as matching parameters among test signals and swatch signals. In order to enhance the computing efficiency of MP sparse decomposition, the strategy based on dictionary partitioning and Genetic Algorithm (GA) was applied. In Matlab, the LGB algorithm was used to distinguish test signals. The simulation results show that the speech signal with length 6 000 can efficiently be recognized through using ten atoms' time-frequency parameters, while the running speed of improved MP sparse decomposition algorithm is enhanced.

Key words: speech recognition; Matching Pursuit (MP) sparse decomposition; Genetic Algorithm (GA); dictionary partition

0 引言

语音识别 (Speech Recognition) 是机器通过识别和理解过程把人类的语音信号转变为相应的文本或命令的技术。随着计算机和人工智能机器的广泛应用,语音识别技术已广泛地应用到信息处理、通信与电子系统、自动控制等领域。由于人类语音信息结构非常复杂、内容极其丰富,处理难度很大,目前,大多数语音识别系统的应用都是在受限条件下进行的,如限定说话人,限定识别词汇量等^[1]。

在限定输入词汇量条件下,语音识别的关键技术为如何快速准确地从带有各种噪声污染的语音信号中寻找模板库中与输入词汇匹配的模板。目前常用的语音识别技术包括:基于矢量量化的识别技术,动态时间归正识别技术以及隐马尔可夫技术。但随着语音识别技术向嵌入式系统方向的发展,这些技术都存在辨识过程复杂,匹配参数多,计算量大等问题^[2]。这就迫切地需要寻找一种以较少参数进行模式识别,同时能减少硬件计算开销的新型语音识别方法。

信号在过完备库上进行分解以获得稀疏表达的 Matching Pursuit (MP) 方法最早由 Mallat 和 Zhang 于 1993 年提出^[3],该方法的核心思想便是根据信号自身的特点选择基本波形,对波形进行平移、反相、拉伸、压缩等变换构造过完备原子库,然后将信号在原子库的各个原子上进行投影,选择最佳原子的

线性组合来重构信号。信号的 MP 稀疏分解算法为信号的稀疏表达提供了一个有效的方法^[4]。利用该方法可以用尽可能少的参数来表达信号的结构特征,这就为减少信号识别参数提供了一条途径。

为了减少语音识别参数,减小硬件计算开销,本文运用基于遗传算法 (Genetic Algorithm, GA) 和库划分方法对 MP 稀疏分解算法找寻最优原子过程的改进,提出利用语音信号稀疏分解的最佳原子参数作为语音信号的特征参数对语音进行识别。由于 MP 稀疏分解算法所得的最佳原子是信号在该方向上投影能量最大的基^[5],故用较少的原子个数就可以有效地表示信号的结构特征。仿真结果表明该方法仅用较少的参数和计算量便具有在强噪声环境中正确识别语音信号的能力。

1 信号 MP 稀疏分解方法分析

1.1 信号 MP 稀疏分解方法

设采样信号 X 的长度为 N ,则采样信号在信号空间中可表示为 $X = [x(1), x(2), \dots, x(N)]$,用于信号稀疏分解的过完备原子库 $D = \{g_{\gamma_i}\}_{\gamma_i \in \Gamma}$ 其中,原子需做归一化处理,即 $\|g_{\gamma_i}\| = 1$, Γ 为参数组 γ_i 的集合。MP 方法分解信号的过程如下^[6] (初始值 $k = 0$)。

第 1 步 从过完备原子库中寻找原子 g_k , 满足:

收稿日期:2008-12-24;修回日期:2009-02-21。

作者简介:高显忠(1985-),男,重庆人,硕士研究生,主要研究方向:控制工程、统计信号处理;侯中喜(1973-),男,陕西宝鸡人,教授,主要研究方向:飞行器总体设计。

$$|\langle X, g_k \rangle| = \sup_{\gamma_i \in \Gamma} |\langle X, g_{\gamma_i} \rangle| \quad (1)$$

第2步 解算残余信号:

$$R^k X = X - \langle X, g_k \rangle g_k \quad (2)$$

其中 $R^k X$ 是用最佳原子对原信号进行最佳匹配后的残余。

第3步 判断 $\|R^k X\| < \varepsilon, \varepsilon > 0$ 为设定的信号残余阈值。是,则转第4步;否,则令 $X = R^k X, k = k + 1$,转第1步。

第4步 输出重构信号:

$$X \approx \sum_{k=0}^L \langle R^k X, g_k \rangle g_k \quad (3)$$

其中 L 为迭代总次数,也即分解信号的原子数,有 $L \ll N$ 。式(3)和 $L \ll N$ 体现了信号稀疏表示和稀疏分解的思想。

针对语音信号的特点,可由 Gabor 原子形成原子库。Gabor 原子由一个经过调制的高斯窗函数构成:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) \cos(vt+w) \quad (4)$$

其中 $g(t) = e^{-\pi t^2}$ 为高斯窗函数, $\gamma = (s, u, v, w)$ 为时频参数, s 为尺度因子, u 为位移因子, v 为频率因子, w 为相位因子。时频参数可以按以下方法离散化: $\gamma = (\alpha^j, p\alpha^j \Delta u, ka^j \Delta v, i\Delta w)$, 其中 $\alpha = 2, \Delta u = 1/2, \Delta v = \pi, \Delta w = \pi/6, 0 < j \leq \lg N, 0 \leq p \leq N2^{-j+1}, 0 \leq k < 2^{j+1}, 0 \leq i \leq 12$ 。

1.2 基于 GA 和库划分改进的 MP 稀疏分解方法

由式(1)可以看出,在基于 MP 的信号稀疏分解中,每一步都要完成信号或信号分解的残余在过完备库中的每一个原子上的投影计算^[5],由 1.1 节中原子库的形成方法可看出每一步分解实际上要进行的内积计算 $\langle X, g_{\gamma_i} \rangle$ 是在一个很高维 (N) 的空间内进行,而且要进行 L_D 次,其中 L_D 为原子库中原子的个数。其计算公式为:

$$L_D = \sum_{j=1}^{\lg N} \sum_{p=0}^{N2^{-j+1}} \sum_{k=0}^{2^{j+1}-1} \sum_{i=0}^{12} 1 \quad (5)$$

经化简可得:

$$L_D = 52(N \lg N + N - 1) \quad (6)$$

对一般长度 N 为 5000 ~ 10000 的语音信号而言,原子库的数目是非常庞大的。这也是阻碍 MP 分解算法推广应用的原因。为此研究稀疏分解的快速算法是十分有必要的,已有学者提出用遗传算法^[7],并行算法^[8]等多种算法加快 MP 稀疏分解算法的速度,其中加速效果最为明显的是高瑞、徐华南等提出的基于 GA 和过完备原子库划分的 MP 信号稀疏分解算法^[9]。该算法融合了遗传算法的群体并行搜索技术,其独特性在于引入“原子波形相同”的等价关系,将过完备原子库划分为:

$$\begin{cases} D = D_{\beta_1} \cup D_{\beta_2} \cup D_{\beta_3} \cup \dots \\ D_{\beta_i} \cap D_{\beta_j} = \emptyset, \quad i \neq j \end{cases} \quad (7)$$

其中 $\beta_i = (s, v, w)_i$, 则 $\Gamma = \{\beta_i | i = 1, 2, \dots\}$, 从而过完备原子库中每一个等价原子库为:

$$D_{\beta_i} = \{g_\gamma | \gamma = (s, u, v, w) \in \Gamma, (s, v, w)_i = \beta_i\} \quad (8)$$

在具体搜索最佳原子时采用先搜索最佳子库,再从最佳子库中搜索最佳原子的策略,从而极大地提高了 MP 稀疏分解算法的速度。

2 用改进 MP 稀疏分解方法进行语音识别

在语音识别方法中,语音辨识参数的选择对识别语音的计算开销和正确率有重要的影响。在常用的基于矢量量化的识别技术中,将 N 维信号空间划分为 R 个区域边界,然后将输

入信号的矢量与这些区域的边界进行比较,并被量化为距离最小的区域的中心矢量值。其参数选择与信号长度 N 有关,随着信号长度的增加,辨识参数也越来越多,增大了辨识的计算难度。

由于稀疏分解重构的是信号与原子结构特性相吻合的部分,而噪声是随机的,不相关的,没有结构特性,所以用稀疏分解重构后的样本信号与测试信号进行匹配可以很好消除噪声的影响。如图 1 所示,仅用 25 个原子便将原始信号的主体结构重构出来,残余信号则表现为与信号结构无关的随机噪声特性。利用信号 MP 稀疏分解的这一有利性质,本文提出:采用信号 MP 稀疏分解生成最佳原子的时频参数作为信号的匹配参数。由于最佳原子对信号的结构特性有极强的表现力,因此在进行模板匹配的时候,利用少数几个原子就可以将含有强噪声的测试语音主体结构重构出来,并与模板库中的模板进行精确匹配。其具体操作过程如下:

1) 对样本语音进行稀疏分解,提取重构样本语音最佳原子的时频参数 $\gamma_i (i = 1, 2, \dots, k, k$ 为样本个数),形成样本语音的模板库。

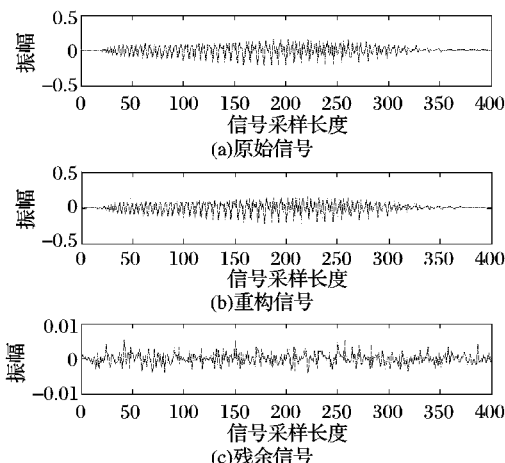


图1 25 个原子分解后的重构信号与残余信号

2) 对待测语音进行稀疏分解,提取重构待测语音最佳原子的时频参数 γ_i 。

3) 将 γ_i 分别与 γ_j 做距离运算,寻找与 γ_i 最佳匹配的模板。

从以上步骤中可以看出,MP 稀疏分解算法将原本 N 维的信号向量匹配运算转换为有限个稀疏的原子时频参数之间的模式匹配问题,可有效地降低计算成本。由于 MP 稀疏分解算法能很好地滤除信号中混杂的非结构噪声,用该方法可得到较高的匹配准确率。

3 仿真结果分析

为了验证利用语音信号 MP 稀疏分解得到的最佳原子参数进行语音识别的有效性,实验选取了在正常语速下读的 0~9 共 10 个限定识别词汇量,每个词汇采集 10 个样本。利用这 100 个语音信号建立模板库,采用 LGB 算法对判别函数进行训练,然后对在有噪声环境中采集的 100 个测试语音进行判别。

图 2 和图 3 显示了长度约为 6000 的样本信号与测试信号在重构原子个数为 25 时的重构结果。由图中可以看出,25 个原子虽不足以重构长度约 6000 信号的所有细节特征,但样本信号与具有强噪声背景的测试信号经 MP 稀疏分解后重构

的图像却基本一致。这说明利用信号的 MP 稀疏分解方法可有效地消除噪声的影响,提取出信号最主要的结构特征信息。

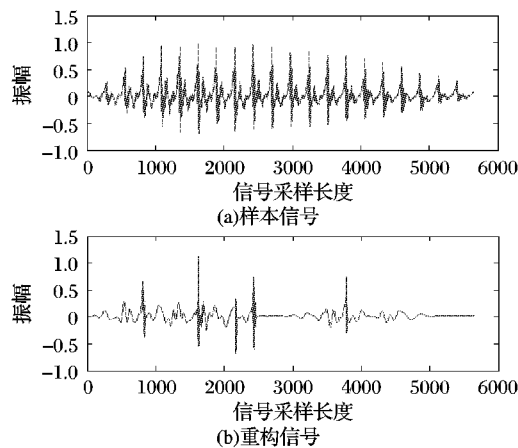


图2 样本信号重构

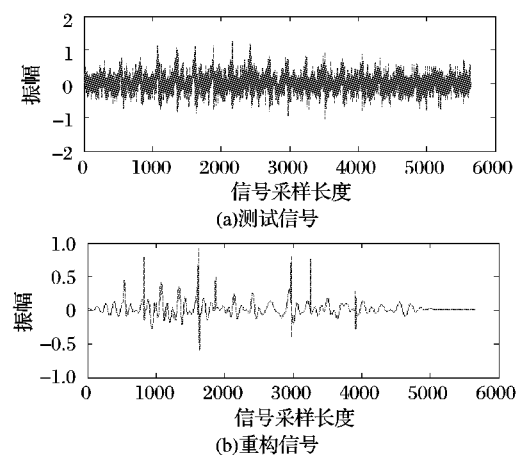


图3 测试信号重构

表1列举了重构原子选择为3,5,10的情况下,判别一个语音信号所花费的计算时间以及判别正确率。由于原子的时频参数集为 $\gamma = (s, u, v, w)$,所以一个原子只需4个参数就可以确定。

表1 不同原子个数下的匹配性能对比

原子个数	参数个数	计算时间/s	判别正确率/%
3	12	0.0045	71
5	20	0.0087	79
10	40	0.0161	94

由表1可以看出,对此类限定输入词汇量的问题,只需要

少数的几个原子便足以区分信号的结构特征,达到较高的判别正确率。

4 结语

本文针对目前语音匹配算法普遍存在的判别形式复杂,匹配参数多,计算复杂的问题,提出利用信号的 MP 稀疏分解方法来构造信号的匹配参数。由于 MP 稀疏分解方法可有效地滤除信号中的非结构误差,用较少的原子便可表现信号的结构特点,因此该方法在模式匹配中具有研究潜力和应用价值。文章对该方法进行了仿真试验,测试程序中 MP 稀疏分解的寻优算法采用的是基于 GA 和库划分的寻优策略,这增加了算法的运行效率。仿真试验表明以信号 MP 稀疏分解的最佳原子参数进行语音识别在基本的 LGB 判别算法下是有效的。如何从最佳原子参数中选取最优的判决向量以及选用一种什么样的判别算法以进一步提高判别准确率、减小运算量是需要继续研究的问题。

参考文献:

- [1] 林琳, 王树勋, 郭纲. 短语音说话人识别新方法的研究[J]. 系统仿真学报, 2007, 19(10): 2272-2275.
- [2] LI DENG, DONG YU. Use of differential cepstra as acoustic features in hidden trajectory modeling for phonetic recognition [C]// ICASSP 2007. Honolulu: IEEE Press, 2007: 445-448.
- [3] MALLAT S, ZHANG Z. Matching pursuit with time-frequency dictionaries [J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397-3415.
- [4] SHASHANKA M V, RAJ B, SMARAGDIS P. Sparse overcomplete decomposition for single channel speaker separation [C]// IEEE International Conference on Audio and Speech Signal Processing. Honolulu: IEEE Press, 2007: 15-20.
- [5] 王建英, 尹忠科, 张春梅. 信号与图像的稀疏分解及初步应用[M]. 成都: 西南交通大学出版社, 2006.
- [6] 王建英, 尹忠科. 基于稀疏分解的弱电信号检测方法[J]. 铁道学报, 2007, 29(2): 114-117.
- [7] ROSA M, VENTURA F, VANDERGHEYNSE P. Matching pursuit through genetic algorithms [EB/OL]. [2008-10-10]. <http://lts2www.epfl.ch/~figueras/GeneticAlgorithm.pdf>.
- [8] ARTHUR P L, PHILIPES C L. Voiced/unvoiced speech discrimination in noise using gabor atomic decomposition[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Hong Kong: IEEE Press, 2003: 820-823.
- [9] 高瑞, 徐华楠, 胡钢. 基于 GA 和过完备原子库划分的 MP 信号稀疏分解算法[J]. 科学技术与工程, 2008, 8(4): 914-917.

(上接第1553页)

返回充电功能,所以对产品的开发有一定的推广意义。值得提出的是,本文在分块时没有利用横向划分的优势,使得地图中宽阔的自由区域分块太多,增大了深度优先遍历的开销。如果利用横向划分的优势,对分块进一步优化组合,来减少自由区域划分的数量,便可以减少最终树枝的深度,优化深度优先算法。其次,房间中的物品摆放是建立在静态的基础上,如何应对动态的摆放,以及分区更少和返回路径更直等难题,这些将是下一步的研究工作。

参考文献:

- [1] 艾延廷, 杨明媛, 崔静, 等. 智能吸尘器控制系统软件设计[J]. 测控技术, 2007, 26(2): 73-75, 80.

- [2] 胡跃明, 丁维中, 吴析生. 吸尘机器人的研究现状与展望[J]. 计算机测量与控制, 2002, 10(10): 631-634.
- [3] 肖海荣, 富文军, 张晓军, 等. 基于多传感器数据融合的移动机器人导航[J]. 系统工程与电子技术, 2001, 23(7): 66-99.
- [4] 李群明, 能蓉, 褚健. 室内移动机器人定位方法研究综述[J]. 机器人, 2003, 25(6): 560-567, 573.
- [5] 阮晓刚, 徐绍敏. 一个室内清洁机器人的区域遍历与地图绘制[J]. 机器人技术与应用, 2006(4): 47-52.
- [6] 赵国玲, 任文娟. 数据结构[M]. 北京: 机械工业出版社, 2005.
- [7] 殷剑宏, 吴开亚. 图论及其算法[M]. 合肥: 中国科学技术大学出版社, 2003.
- [8] 周忠荣. 离散数学及其应用[M]. 北京: 清华大学出版社, 2007.