

基于 MFCC 和双重 GMM 的鸟类识别方法

王恩泽, 何东健⁺

(西北农林科技大学 机械与工程学院, 陕西 杨凌 712100)

摘要: 针对鸟类鸣声信号变化丰富和复杂的特点, 提出一种基于 MFCC 和鸣叫、鸣唱声 GMM 模型的鸟类识别方法。该方法拟采用将鸟鸣声分为鸟叫声和鸟唱声的策略, 分别提取其特征参数 MFCC, 提出双重 GMM 模型进行训练和识别。用 8 种鸟的鸣叫声和鸣唱声 1077 个样本进行实验, 实验结果表明, 双重 GMM 模型的识别率达到 90% 以上, 与单一鸣声模型相比具有更高的识别率。

关键词: 鸟类识别; 梅尔倒谱系数; 鸣叫; 鸣唱; 双重高斯混合模型

中图分类号: TN912.34 **文献标识号:** A **文章编号:** 1000-7024 (2014) 05-1868-04

Bird recognition based on MFCC and dual-GMM

WANG En-ze, HE Dong-jian⁺

(College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China)

Abstract: In accordance with the flexibility and the complexity of bird chirps, a bird recognition method is raised on the basis of MFCC and GMM model. By classifying bird chirps into call and sing, and extracting their feature parameter MFCC respectively, a dual-GMM model is put forward for training and recognizing. Experiments are carried out with 1077 bird chirp samples of 8 kinds of birds. Results show that the recognition rate of dual-GMM model is over 90%, much higher than that of the single vocal model.

Key words: bird identification; Mel-frequency cepstrum coefficient (MFCC); call; song; double Gaussian mixture model (dual-GMM)

0 引言

湿地鸟的种类识别对保护鸟类、了解湿地生态系统和评测湿地环境质量具有重要意义^[1]。传统的鸟类识别方法有视频监测和人工识别方法, 但视频监测成本高、监测目标范围有限^[2]; 而人工识别实施困难、效率低^[3]。声音是鸟类的重要生物学特征, 也是识别鸟类的重要依据^[4]。通过声音识别鸟的种类, 对确定湿地鸟的种类、判断是否有外来鸟类以及研究鸟的迁徙规律具有重要的意义^[5]。

在声音识别中, 特征参数^[6,7]的选择和分类模型^[8,9]的构建, 直接影响识别性能和复杂度。通常采用的特征参数有线性预测倒谱系数 (linear prediction cepstrum coefficient, LPCC) 和 Mel 频率倒谱系数 (mel-frequency cepstral coefficient, MFCC), 分类模型多用动态时间归整 (dynamic time warping, DTW)、支持向量机 (support vector

machine, SVM)、向量量化 (vector quantization, VQ)、隐马尔可夫模型 (hidden Markov model, HMM) 和高斯混合模型 (Gaussian mixture model, GMM)。Jinkui Cheng 等人^[10]通过提取 4 种雀形目鸟类的 MFCC 特征, 用 GMM 进行种内个体识别, 该方法仅在种内进行个体识别, 未考虑不同种类间的识别。薛宇志^[11]采用二阶层式识别架构, 先判断鸣声类型, 然后输入相应的判断单元进行识别, 但需要事先判断鸟类的鸣声类型, 不仅增加了复杂度, 而且人工判断鸣声类型的准确度亟待提高。

为提高湿地鸟类识别率, 本文拟采用将鸟声分为鸣叫声和鸣唱声的策略, 分别提取其特征参数 MFCC, 构建双重 GMM 模型进行训练与识别, 以提高识别准确度。

1 声音样本获取与特征参数提取

1.1 声音样本获取

考虑到声音样本的多寡、地域的合理性、科目的差异

收稿日期: 2013-08-28; **修订日期:** 2013-11-08

基金项目: 西北农林科技大学校长基金项目 (A213020901)

作者简介: 王恩泽 (1989-), 男, 山东泰安人, 硕士研究生, 研究方向为智能化检测技术; 十通讯作者: 何东健 (1957-), 男, 陕西汉中, 教授, 研究方向为图像处理, 智能化检测。E-mail: hdj168@nwsuaf.edu.cn

和鸣声类型等因素, 选择陕西省关中地区常见黄臀鹌、矛纹草鹌、北红尾鹌、绿背山雀、方尾鹌、红嘴相思鸟、黄喉鹀和淡尾鹌等 8 种鸟类。8 种鸟类声音样本均来自 XC 网站 (www. xeno-canto. org) 在陕西关中地区采集的数据, 声音中夹杂着自然环境噪音。原始鸟类鸣声声音文件为 mp3 格式, 采样频率为 48000Hz, 比特率为 320000bps。每段声音样本的长度从几秒到十几秒不等。为了使模型训练数据充分, 鸟类鸣声声音长度较长的用于模型训练, 较短的用于模型测试, 从而提高识别的速度。

8 种鸟类鸣声鸣叫类型和鸣唱类型的样本数见表 1。

表 1 各类鸟类鸣声的样本数		
鸟类名称	鸣叫样本数	鸣唱样本数
黄臀鹌	48	108
矛纹草鹌	36	54
北红尾鹌	144	72
绿背山雀	18	36
方尾鹌	72	63
红嘴相思鸟	99	39
黄喉鹀	81	45
淡尾鹌	54	108
合计	552	525

其中每种鸟类鸣声的 2/3 的样本数用于训练模型, 1/3 的样本数用于测试模型。

1.2 预处理

为便于处理, 首先将 mp3 声音文件转化为 wav 格式; 每段声音通常会包括风声、水声和其它环境杂音, 为了提高识别率, 采用维纳滤波器消除环境杂音; 预备实验表明, 用一段连续声音序列的鸣声, 可以提高识别率, 故裁剪掉图 1 声音波形中的杂音和静音部分, 只保留虚线框内声音波形部分, 并顺序链接成 1 个波形文件, 如图 2 所示。

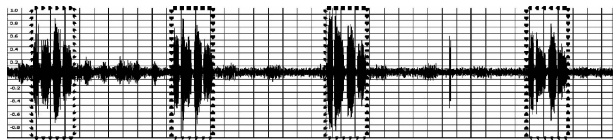


图 1 声音波形剪裁

1.3 MFCC

为识别鸟类, 首先需要提取鸟鸣声和鸟唱声的特征参数。MFCC 根据人耳的听觉特性提出, 对不同频率的敏感度呈非线性, 广泛应用于声音识别领域。MFCC 的计算过程如图 3 所示^[12]。

(1) 预加重: 将输入声音信号通过一个高通滤波器, 提高高频部分, 使信号的频谱变得平坦, 便于分析。高通滤波器表示为

万方数据
$$H(z) = 1 - \lambda z^{-1}$$
 (1)

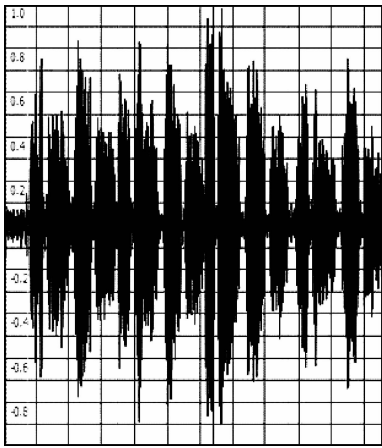


图 2 剪裁后顺序链接的声音波形

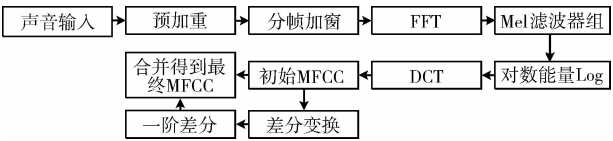


图 3 MFCC 计算过程

本文 λ 取值为 0.9375。

(2) 分帧加窗: 对信号加汉明窗进行重叠分帧, 使其变为短时平稳信号, 减少吉布斯效应^[13]的影响。帧长 N 取 512 个采样点, 帧移取 256 个采样点。

(3) FFT: 对帧信号做快速傅立叶变换 (FFT), 将时域信号转化成频域信号。

(4) Mel 滤波器组: 将 FFT 得到的频谱系数用三角滤波器进行滤波处理, 得到一组系数 m_1, m_2, \dots 。三角滤波器的跨度在 Mel 轴上平均分布, 0Hz 到 Nyquist 频率被所有的滤波器所覆盖, 即采样频率的 1/2。在系统中, Mel 滤波器的阶数为 24。

(5) 对数能量 Log: 对每个滤波器的输出取对数, 得到相应的对数功率谱。

(6) DCT: 利用离散余弦变换 (DCT) 将对数功率谱变换到时域, 所得谱的幅值即为原始的 MFCC 系数。计算公式为

$$C(n) = \sum_{m=1}^M S(m) \cos(\frac{\pi m(m-0.5)}{M}), 0 \leq n \leq M$$
 (2)

式中: $C(n)$ ——第 n 个 MFCC 系数, n ——MFCC 数量, $S(m)$ ——对数功率谱, M ——三角滤波器的个数。

(7) MFCC: 上述得到的特征参数为静态特征参数, 而一阶差分 MFCC 即 Δ MFCC 能反应声音信号的动态特性, 具有较好的鲁棒性。 Δ MFCC 计算公式为

$$d(n) = \frac{\sum_{i=-k}^k i \times C(n+i)}{\sqrt{\sum_{i=-k}^k i^2}}$$
 (3)

式中: k 一般取 2, 为常数。

故将 MFCC 和 Δ MFCC 的组合作为识别特征参数, 既包含了静态特征参数, 又能反应信号的动态特性。特征参数的维数为 24 维。

2 鸟类鸣声模型的建立

2.1 高斯混合模型 GMM

GMM 模型用多个高斯分布的组合来表示鸟类鸣声特征参数在特征空间的分布。一个 GMM 模型可以用均值矢量, 协方差矩阵和混合权重表示, 用 λ 来表示为^[14]

$$\lambda = \{ \omega_i, \mu_i, \Sigma_i \}, i = 1, 2, \dots \tag{4}$$

式中: ω_i ——混合权重; μ_i ——均值矢量; Σ_i ——协方差矩阵。

对于某一种鸟类 N_i , 则它对应的 GMM 模型用 λ 表示为 λ_{i1} 和 λ_{i2} , λ_{i1} 为 N_i 的鸣叫声 GMM 模型, λ_{i2} 为 N_i 的鸣唱声 GMM 模型。

2.2 模型参数估计

在进行模型训练时, 用最大释然估计法^[15] (maximum likelihood, ML) 估计出 GMM 参数集合 λ 。因为 λ 是非线性函数, 所以用最大期望 (expectation-maximization, EM) 算法^[16] 迭代计算出最大可能性的 GMM 参数 λ 。

2.3 鸟类种类识别模型

鸟类的鸣叫声短促而单调, 鸣唱声响亮而富有变化, 鸣叫声和鸣唱声差异很大, 若用单 GMM 进行训练与识别, 识别效果不佳。表 2 为用单 GMM 对 8 类样本 (黄臀鹌、矛纹草鹌、北红尾鹌、绿背山雀、方尾鹌、红嘴相思鸟、黄喉鹌和淡尾鹌莺等) 进行识别的结果。由表 2 可知, 8 类样本的平均正确识别率为 71.90%, 黄臀鹌正确识别率最高也仅为 78.85%。这是因为用单 GMM 进行训练, 相当于 2 个不同特征参数类混在一起进行训练, 会破坏各自特征参数的独特性, 故正确识别率低, 难以达到应用所需的识别效果。

为提高鸟类的正确识别率, 本文采用将鸟类鸣声分为鸣叫声和鸣唱声的策略, 分别提取各自特征参数并训练对应的 GMM 模型, 构建双重 GMM 模型。GMM 的数量比使

用单 GMM 模型增加一倍, 参数 λ 的数量也增加一倍。最后根据最大后验概率准则^[17], 找出未知鸟类所属模型。

图 4、图 5 分别为单 GMM 模型和双重 GMM 模型鸟类识别基本架构图。从两图对比可以看出双重 GMM 模型将模型分的更为细致和精确, GMM 模型的数量是单 GMM 模型的 2 倍, 从而提高了模型识别的准确度。

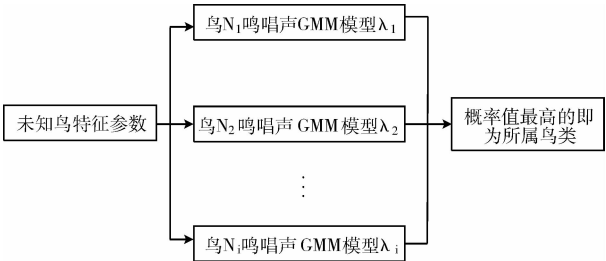


图 4 单 GMM 模型鸟类识别基本架构

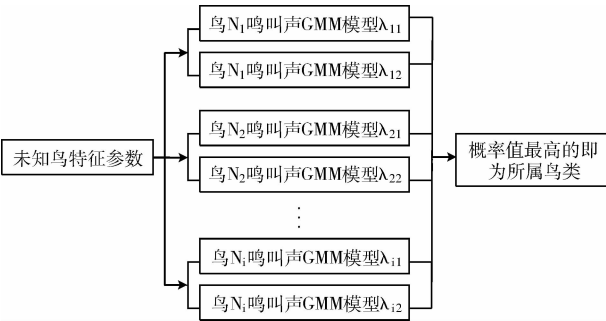


图 5 双重 GMM 模型鸟类识别基本架构

3 实验结果与分析

实验计算机处理器为奔腾双核处理器, 主频为 2.5GHz, 内存为 2GB, 操作系统为 Windows 7, 采用 MATLAB R2012a 进行编程。

单 GMM 模型的正确识别率见表 2。

用本文构建的双重 GMM 模型, 对 8 类鸟进行识别实验, 识别结果见表 3。

表 2 单 GMM 模型的正确识别率 (%)

输入 \ 结果	黄臀鹌	矛纹草鹌	北红尾鹌	绿背山雀	方尾鹌	红嘴相思鸟	黄喉鹌	淡尾鹌莺
黄臀鹌	78.85	7.69	3.84	0	0	0	0	9.62
矛纹草鹌	0	68.89	0	11.11	16.67	0	3.33	0
北红尾鹌	2.78	11.11	69.44	0	0	2.78	0	8.33
绿背山雀	7.41	0	9.26	74.07	0	3.70	0	5.56
方尾鹌	3.69	0	9.63	0	66.67	5.19	8.15	6.67
红嘴相思鸟	0	6.52	7.25	0	0	79.71	0	6.52
黄喉鹌	8.73	0	9.52	10.32	0	0	63.49	7.94
淡尾鹌莺	7.41	0	3.70	5.56	0	0	9.26	74.07
平均正确识别率	71.90							

表 3 双重 GMM 模型的正确识别率 (%)

输入 \ 结果	黄臀鹌	矛纹草鹌	北红尾鹌	绿背山雀	方尾鹌	红嘴相思鸟	黄喉鹀	淡尾鹌鹑
黄臀鹌	94. 23	1. 92	0	0	0	0	0	3. 85
矛纹草鹌	0	92. 59	0	1. 85	3. 71	0	1. 85	0
北红尾鹌	0	5. 56	93. 06	0	0	0	0	1. 38
绿背山雀	3. 70	0	1. 86	90. 74	0	0	0	3. 70
方尾鹌	0	0	2. 96	0	94. 81	0	1. 48	0. 75
红嘴相思鸟	0	0	0	0	0	100. 00	0	0
黄喉鹀	0	1. 59	2. 38	0	0	0	96. 03	0
淡尾鹌鹑	4. 45	0	0	1. 11	0	0	1. 11	93. 33
平均正确识别率	94. 35							

由表 3 可知,红嘴相思鸟正确识别率最高,达 100.00%,绿背山雀正确识别率最低,但也达到 90.74%,8 类鸟的平均正确识别率为 94.35%。对比表 2 与表 3 可以看出,本文提出的双重 GMM 模型的正确识别率比单 GMM 模型最高提高了 42.21 个百分点,平均正确识别率提高了 29.76 个百分点,达到了理想的识别效果。绿背山雀正确识别率低于其它鸟类,是因为其训练样本相对较少(鸣叫和鸣唱样本数分别为 18 个和 36 个),可见训练样本数的多少也是影响正确识别率的一个重要因素。

效率方面,因为声音样本总数是一样的,所以本文提出的双重 GMM 模型训练时间和使用单 GMM 模型训练时间相差无几;使用双重 GMM 模型时单个声音样本的平均识别时间为 2.35s,比用单 GMM 模型平均识别时间 1.72s 多用了 0.63s,但平均正确识别率大大提高,且 2s 左右的识别时间在可接受的范围之内。

4 结束语

本文提出基于 MFCC 和鸣叫声 GMM 与鸣唱声 GMM 的双重 GMM 模型的识别方法,将变化丰富的鸟鸣声信号细分为鸣叫和鸣唱 2 种,解决了 GMM 模型训练时输入信号的多而杂乱的问题。该方法正确识别率高,对既有鸣叫又有鸣唱的鸟类正确识别率达 90% 以上,但识别效率受到一定程度的影响。本文受限于鸟种声音的收集数量有限,只选取了陕西关中地区常见的 8 鸟类,对于系统的鲁棒性和实用性尚需要用更多的鸟种进行验证。

参考文献:

[1] LUO Lei, ZHAO Hongfeng, GAO Xuebin, et al. Waterbirds diversity in Shaanxi province [J]. Sichuan Journal of Zoology, 2008, 27 (4): 579-584 (in Chinese). [罗磊, 赵洪峰, 高学斌, 等. 陕西水鸟的地理分布 [J]. 四川动物, 2008, 27 (4): 579-584.]

[2] LV Xiaoyun, WANG Hongxia. Abnormal audio recognition algorithm based on MFCC and short-term energy [J]. Journal of Computer Applications, 2010, 30 (3): 796-798 (in Chi-

nese). [吕霄云, 王宏霞. 基于 MFCC 和短时能量混合的异常声音识别算法 [J]. 计算机应用, 2010, 30 (3): 796-798.]

[3] ZHU Leqing, ZHANG Zhen. Automatic recognition of insect sounds using MFCC and GMM [J]. Acta Entomologica Sinica, 2012, 55 (4): 466-471 (in Chinese). [竺乐庆, 张真. 基于 MFCC 和 GMM 的昆虫声音自动识别 [J]. 昆虫学报, 2012, 55 (4): 466-471.]

[4] XIAO Hua, ZHANG Yanyun. The research of birds chirp [J]. Bulletin of Biology, 2009, 44 (3): 11-13 (in Chinese). [肖华, 张雁云. 鸟类鸣声研究 [J]. 生物学通报, 2009, 44 (3): 11-13.]

[5] SU Xiu, ZHU Xi. Research progress of birdsong [J]. Journal of Zhejiang Forestry College, 2006, 23 (3): 323-327 (in Chinese). [苏秀, 朱曦. 鸟声研究进展 [J]. 浙江林学院学报, 2006, 23 (3): 323-327.]

[6] CHEN Weidong, WANG Xiaoya, XIE Jing. Speech recognition based on LPCC [J]. Radio Engineering, 2009, 39 (9): 16-18 (in Chinese). [陈卫东, 王晓亚, 解静. 基于 LPCC 的多语种识别算法 [J]. 无线电工程, 2009, 39 (9): 16-18.]

[7] GONG Xiaomei, WANG Huaiyang. MFCC feature extraction in noisy environments [J]. Microcomputer Information, 2007, 23 (22): 247-249 (in Chinese). [宫晓梅, 王怀阳. 噪声环境下 MFCC 特征提取 [J]. 微计算机信息, 2007, 23 (22): 247-249.]

[8] Somervuo P, Harma A, Fagerlund S. Parametric representations of bird sounds for automatic species recognition [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14 (6): 2252-2263.

[9] Lee C H, Han C C, Chuang C C. Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16 (8): 1541-1550.

[10] Cheng J, Sun Y, Ji L. A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines [J]. Pattern Recognition, 2010, 43 (11): 3846-3852.

- based progressive photon mapping [J]. Computer Engineering and Design, 2012, 33 (1): 219-223 (in Chinese). [李睿, 陈彦云, 刘学慧. 基于自适应光子发射的渐进式光子映射 [J]. 计算机工程与设计, 2012, 33 (1): 219-223.]
- [4] Light structural problems [EB/OL]. [2011-09-27]. <http://www.xhclub.net/forum/viewthread.php?tid=127487>(in Chinese). [光子的结构问题[EB/OL]. [2011-09-27]. <http://www.xhclub.net/forum/viewthread.php?tid=127487>.]
- [5] LIANG Fei, YAN Hongyin. Research of dynamic self-adaption intrusion detection model based of clustering [J]. Computer Engineering and Design, 2013, 34 (3): 814-820 (in Chinese). [梁飞, 闫宏印. 基于聚类分析的动态自适应入侵检测模式研究 [J]. 计算机工程与设计, 2013, 34 (3): 814-820.]
- [6] LIU Shuren, SONG Yaqi, ZHU Yongli. Research on classification and identification of neuronal morphology [J]. Computer Engineering and Design, 2013, 34 (2): 670-675 (in Chinese). [刘树仁, 宋亚奇, 朱永利. 神经元形态分类和识别方法的研究 [J]. 计算机工程与设计, 2013, 34 (2): 670-675.]
- [7] TANG Chunlei, DONG Jiaqi, ZHU Boya. A clustering algorithm for transaction sequences based on growth patterns [J]. Pattern Recognition and Artificial Intelligence, 2013, 26 (5): 467-473 (in Chinese). [汤春蕾, 董家麒, 朱博雅. 一种基于增长模式的交易序列聚类算法 [J]. 模式识别与人工智能, 2013, 26 (5): 467-473.]
- [8] ZHAO Dongfang. Mathematical model and calculation [M]. Wuhan: Science Press, 2007 (in Chinese). [赵东方. 数学模型与计算 [M]. 武汉: 科学出版社, 2007.]
- [9] RUAN Lifeng, TANG Zhilie. Investigation on light intensity distribution of Raman-nath acousto-optic diffraction based on fourier analysis [J]. Acta Optica Sinica, 2013 (3): 90-95 (in Chinese). [阮立锋, 唐志列. 基于傅里叶分析的拉曼-奈斯声光衍射光强分布的研究 [J]. 光学学报, 2013 (3): 90-95.]
- [10] QIAN Guolin, WU Jianhong. Laser pulse pattern influenced by mosaic grating gap [J]. High Power Laser and Particle Beam, 2011, 23 (12): 3177-3182 (in Chinese). [钱国林, 吴建宏. 拼接光栅缝隙对激光脉冲光强分布的影响 [J]. 强激光与粒子束, 2011, 23 (12): 3177-3182.]
- [11] Light diffraction phenomenon [EB/OL]. [2011-09-27]. http://www.china001.com/show_hdr.php?xname=PPDDMV0&dname=L7PRV31&xpos=9(in Chinese). [光的衍射现象[EB/OL]. [2011-09-27]. http://www.china001.com/show_hdr.php?xname=PPDDMV0&dname=L7PRV31&xpos=9.]
-
- (上接第 1871 页)
- [11] XUE Yuzhi. Automatic identification of bird species by their calls and songs [D]. Taipei: MSc Thesis, National Taipei University of Technology, 2010 (in Chinese). [薛宇志. 依照鸟类鸣叫声和鸣唱声识别其种类 [D]. 台北: 台北科技大学硕士学位论文, 2010.]
- [12] CHEN Yong, QU Zhiyi, LIU Ying, et al. The extraction and application of phonetic characteristic parameter MFCC [J]. Journal of Hunan Agricultural University (Natural Sciences), 2009, 35 (10X): 106-107 (in Chinese). [陈勇, 屈志毅, 刘莹, 等. 语音特征参数 MFCC 的提取及其应用 [J]. 湖南农业大学学报: 自然科学版, 2009, 35 (10X): 106-107.]
- [13] YUAN Zhengwu, XIAO Wanghui. Improved speech recognition algorithm based on MFCC feature [J]. Computer Engineering and Applications, 2009, 45 (33): 108-110 (in Chinese). [袁正午, 肖旺辉. 改进的混合 MFCC 语音识别算法研究 [J]. 计算机工程与应用, 2009, 45 (33): 108-110.]
- [14] YU Qingqing, LI Ying, LI Yong. Natural sounds recognition using GMM distribution [J]. Computer Engineering and Applications, 2011, 47 (25): 152-155 (in Chinese). [余清清, 李应, 李勇. 基于高斯混合模型的自然环境声音的识别 [J]. 计算机工程与应用, 2011, 47 (25): 152-155.]
- [15] ZHAI Jiyong, ZHANG Peng. Optimization of parameter estimation based on Gaussian mixture model [J]. Computer Technology and Development, 2011, 21 (11): 145-148 (in Chinese). [翟继友, 张鹏. 高斯混合模型参数估值算法的优化 [J]. 计算机技术与发展, 2011, 21 (11): 145-148.]
- [16] WANG Aiping, ZHANG Gongying, LIU Fang. Research and application of EM algorithm [J]. Computer Technology and Development, 2009, 19 (9): 108-110 (in Chinese). [王爱平, 张功营, 刘方. EM 算法研究与应用 [J]. 计算机技术与发展, 2009, 19 (9): 108-110.]
- [17] KANG Guangyu, GUO Shize, SUN Shenghe. Band energy based GMM speech with noise classification algorithm [J]. Chinese Journal of Scientific Instrument, 2009, 30 (9): 1950-1955 (in Chinese). [康广玉, 郭世泽, 孙圣和. 基于子带能量的 GMM 含噪语音分类算法 [J]. 仪器仪表学报, 2009, 30 (9): 1950-1955.]