一种基于 MFCC 与 PCP 联合特征的和弦识别方法

李 锵,秦媛媛
(天津大学 电子信息工程学院,天津 300072)

摘 要:结合乐理理论和信号处理理论,针对传统和弦识别仅考虑音高特性的音级轮廓特征 PCP (pitch class profile) 造成正确识别率较低的问题,提出一种以反映听觉特性的 MFCC (mel frequency cepstral coefficent) 与 PCP 的联合特征和稀疏表示分类器 (sparse representation classification,SRC) 的和弦识别方法. 通过对两特征矢量的叠加构成新的和弦特征,然后利用 SRC 进行和弦识别. 实验结果表明,与传统方法的识别率相比,本方法的识别率大幅提高.

关键词:和弦识别;MFCC;PCP;MFCC+ PCP;稀疏表示分类器
中图分类号:TP391.4 **文献标志码**:A **文章编号**:1671-024X (2015) 01-0050-05

A chord recognition method based on joint feature of MFCC and PCP

LI Qiang, QIN Yuan-yuan
(School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Combined with music and signal process theory, the paper proposes a new chord recognition approach which utilizes the MFCC (Mel Frequency Cepstral Coefficent) reflecting auditory perception properties jointly with the traditional PCP (Pitch Class Profile) as the combined feature and SRC (Sparse Representation classification) to improve the recognition rate. The experimental results show that the recognition rata of the new method is much better in average than that of the traditional methods.

Key words: chord recognition; MFCC; PCP; MFCC+PCP; sparse representation classification

音乐信号处理是近年来人工智能与模式识别领域的研究热点,和弦作为音乐信号重要的中层特征之一,是由 3 个或 3 个以上的不同音按照一定规则组合并同时发音形成的. 不同和弦组成的和弦序列通过音符之间的和谐程度及高低差别表征不同的旋律,充分表达了一段音乐的内容和特征[1],对于实现音乐信息检索、乐曲分割与匹配以及歌曲自动翻唱具有重要作用. 因此,和弦识别的研究具有很广泛的应用价值. 音乐和弦识别主要包括和弦特征提取和识别模型的确定. 比较有代表性的研究工作是 Brown[2]首次将音乐识别与音乐理论结合,提出恒 Q 变换;Fujishima[3]在 1999 年率先提出 12 维音级轮廓 (pitch class profile, PCP),将音乐信号能量映射到 12 个音级上,重建音级谱,最后利用模板匹配法识别和弦,取得了一定效果. Gomez[4]在此基础上提出 HPCP (harmonic PCP) 特征用于和弦识别的键估计系统中并取得了 66.7%的正确键估计;Lee[5]使用谐波产物谱 (harmonic product spectrum,HPS) 提出一种增强型的 PCP 特征,与传统的 PCP 特征相比,增强型 PCP 对具有相同根音的和弦具有更高的识别率. Sheh 和 Ellis[6]提出将统计学方法即隐马尔可夫模型 (hidden markov model,HMM) 模型运用于和弦的分割与识别. Wang[7]结合人耳听觉特性和音乐理论提出了新的识别特征 MPCP (Mel PCP),克服了 PCP 特征在低频段特征模糊和峰值处容易发生混淆的缺陷,但采用了条件随机场分割方法,运行时间长;文献[8]则采用卷积神经网络进行和弦识别,可以有效避免噪声对和弦识别率的影响,但该方法能识别的音频数量较小. 稀疏表示[9]是最小一范数[10]的优化方法,在模式识别领域的相关研究中取得了很多可观的成果. 本文将稀疏表示方法引入和弦识别模型学习与分类. 传统的 PCP 特征没有考虑到人耳听觉特性,在低频段比较模糊,而 MFCC[11]特征恰好能够弥补这一缺陷,充分描述

收稿日期:2014-11-07
基金项目:国家自然科学基金项目 (61471263,61101225,60802049);天津大学自主创新基金 (60302015)
通信作者:李 锵 (1974—),男,博士,教授,主要研究方向为音乐信号处理、模式识别、医学图像处理. E-mail:lqiang@tju.edu.cn

了和弦旋律的低频段. 本文将传统恒 Q 变换的 PCP 特征与梅尔倒谱系数 (MFCC) 相结合，提出一种基于 MFCC 与 PCP 的联合特征，并引入稀疏表示分类器，根据最小一范数实现对待测和弦的类型识别.

1　基于 MFCC+PCP 联合特征的和弦识别系统

本文和弦识别算法的特征提取部分包括 MFCC 和 PCP 特征提取. 和弦的具体识别过程如图 1 所示.

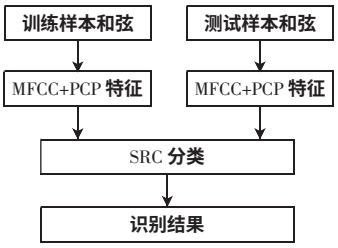


图 1　基于 MPCP 特征的和弦识别流程图

Fig.1　Flow chart of chord recognition based on MPCP

对训练样本和测试样本的每个和弦音频均提取 MFCC 和 PCP 2 种特征，然后对 2 种特征向量相加得到训练样本特征集和测试样本特征集的 MFCC+PCP 特征,再将 2 特征集矩阵输入到 SRC 分类器中,得到和弦识别结果.

1.1　和弦的特征提取

1.1.1　MFCC 特征提取

Mel 频率倒谱系数 (MFCC) 由 Davis 和 Mermelstein[10]于 1980 年基于人耳听觉特性和语音生成原理提出,MFCC 特征被广泛的应用到语音识别研究中.

对于音频信号而言,MFCC 特征具体的计算步骤如下.

第 1 步 :将时域离散和弦音乐信号进行预加重,分帧和加窗处理. 预加重滤波器是一阶的,系统函数为 $H(z) = 1 - uz^{-1}$;取帧长为 N,帧移为 $N/2$;所加窗的窗函数类型为汉明窗 (Hamming) .

第 2 步 :经过快速傅里叶变换 (FFT) 转化为频域信号,得到其频谱 $X(k)$. 计算其能量谱

$$S(k) = |X(k)|^2, k \in 1,2,\cdots,N \tag{1}$$

第 3 步 :用 M 个 Mel 频率带通滤波器 $H_m(k)$ 进行滤波 ;滤波器输出值为 $P_m(k)$, $m = 1,2,\cdots,M$.

$$P_m(k) = H_m(k)\,S(k)\ , k = 1,2,\cdots,N \tag{2}$$

第 4 步 :将每个滤波器的输出值 $P_m(k)$ 取自然对数,得到 $M_m(k)$,$m = 1,2,\cdots,M$.

$$M_m(k) = \ln(P_m(k))\ , k = 1,2,\cdots,N \tag{3}$$

第 5 步 :对第 4 步所得的结果作离散余弦变换 (DCT) ,对于每一帧信号,得到 M 个 $MFCC$ 系数.

$$MFCC_m = \sqrt{\frac{2}{N}} \sum_{k=1}^{N} M_m(k) \cos\left(\frac{\pi m}{M}(k - 0.5)\right)$$
$$1 \leqslant m \leqslant M \tag{4}$$

第 6 步 :Mel 滤波器的通道个数设置为 M 个,每个和弦样本得到的 $MFCC$ 系数矩阵的大小为 $M*L$,L 为帧数. 对每一帧第 m $(1 \leqslant m \leqslant M)$ 个滤波器的输出值 $MFCC_m$ 取平均值,公式为 :

$$MFCC_m = \frac{1}{L} \sum_{l=1}^{L} MFCC_m(l)\ , m = 1,2,\cdots,M \tag{5}$$

其中 $MFCC_m(l)$ 代表第 l 帧第 m 个滤波器的输出值. 这样得到的每个和弦的 MFCC 统计平均值的大小为 $M \times 1$.

1.1.2　PCP 特征提取

音级轮廓 (PCP) 特征是由 Fujishima[3]于 1999 年提出,将频谱重建为音级谱,然后将音乐信号能量映射到 12 个音级上. 由于 FFT 和 STFT 估计音阶频率时的频率线是按线性分布的,所以两者频率点不能完全对应，致使某些音阶频率的估计值产生错误. 因此在时频变换阶段采用了一种谱线频率与音阶频率具有相同指数分布规律的视频变换方法—CQT (Const-Q Transform,恒 Q 变换) [2]. 将经过 CQT 变换的 PCP 特征作为新的 PCP 特征,该特征包含丰富的音乐谐波结构.

PCP 统计平均值特征的步骤如下.

第 1 步 :时域离散和弦音乐信号 $x(m)$ 分帧,加窗,进行恒 Q (品质因数) 变换 (Constant Q Transform, CQT) 将时域变换到频域. 取帧长为 N,帧移为 $N/2$,所加窗的类型为汉明窗 (hamming) .

$$X_n^{cqt}(k) = \frac{1}{N_k} \sum_{m=1}^{N_k} x(m)\, w_{N_k}(m)\, e^{\frac{-2\pi jmQ}{N_k}}$$
$$k = 1,2,\cdots,M \tag{6}$$

式 (6) 表示第 n 帧十二平均律中第 k 个半音的频谱,故通常 M 值为 12. 式中 :$x(m)$ 为输入的时域离散和弦音乐信号 ;$N_k = Qf_s/f_k$ 表示第 k 个半音对应的窗长 ;f_s 表示采样频率 ;f_k 表示第 k 个半音的频率 ;$w_{N_k}[m]$ 表示窗长 N_k 为的 hamming 窗.

第 2 步 :频谱映射. 将频谱 $X_n^{cqt}(k)$ 映射为音级域的 $p(k)$,它由 12 维向量组成,每维向量代表一个半音音级强度. 按照乐理知识中的十二平均律以对数方式将频率映射到音级上,$X_n^{cqt}(k)$ 中的 k 被映射为 PCP 中的 p,映射公式如下 :

$$p(k) = 12 \cdot [\log_2(f_s/N \cdot k/f_0)] \bmod 12 \tag{7}$$

式中 :$f_0 = 130.8$ Hz 为参考频率 ;f_s 为采样率 ;mod12 为对 12 的求余运算.

第 3 步 通过累加所有与某一特定音级相对应的频率点的频率幅度平方值，得到每一帧信号的各个 *PCP* 分量的值. 具体公式如下：

$$PCP(p) = \sum_{k;p(k)=p} |X^{cqt}(k)|^2, p = 1,2,\cdots,12 \quad (8)$$

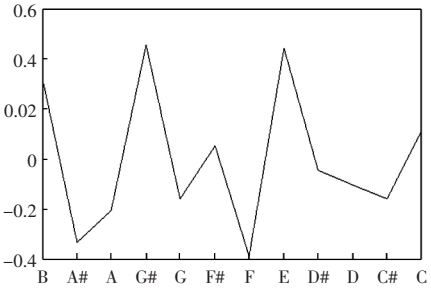
第 4 步：经过上面的计算得到一个 12*L 的矩阵音色图（chromagram），其中 L 代表帧数. 计算每一个音级（行）的均值，公式如下：

$$PCP(p) = \frac{1}{L}\sum_{l=1}^{L} PCP_l(p), p = 1,2,\cdots,12 \quad (9)$$

经过上面的计算，得到一个 12*1 维的矢量，这就是所求的每个和弦样本的 *PCP* 统计平均值.

以大 E 和弦为例，其音色图和 PCP 图如图 2 所示.

(a) 大 E 和弦的音色图



(b) 大 E 和弦的 PCP 图

图 2　大 E 和弦的音色图和 PCP 图

Fig.2　Chromagram and PCP of E major

1.1.3　MPCP 特征提取

MFCC 作为和弦特征，虽然考虑了人耳听觉特性，但由于没有考虑到乐理知识、计算量和精度要求高，抑制了音高（pitch）信息，其识别效果并不好.

音级轮廓（PCP）作为和弦特征，虽然体现了音乐理论，但是没有充分考虑到人耳特性，在低频段特征比较模糊，在峰值附近容易发生混淆影响了识别效率. 所以本文将 M 维 MFCC 统计平均值和 12 维 PCP 统计平均值连接，得到一个 M + 12 维联合和弦特征值. 和弦特征提取的具体流程如图 3 所示.

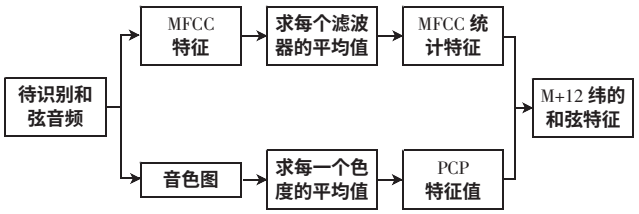


图 3　提取和弦特征的流程图

Fig.3　Flow chart of feature extraction

1.2　基于稀疏表示的和弦识别

稀疏表示分类方法是在最小一范数基础上提出的，是模式识别领域热点研究课题，其分类思想是：在训练数据空间足够大的情况下，测试数据可以由训练数据空间中同类数据线性组合，找到最佳的稀疏向量.

在理想情况下，如果测试数据是训练数据中的某一类，则这个测试数据的线性组合就只能包含该类训练数据，即稀疏系数中只有一小部分是非零值. 本文利用稀疏表示分类模型实现和弦识别.

1.2.1　稀疏表示模型

(1) 稀疏表示方法. 假设第 i 类训练样本的数据 $\boldsymbol{A}_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}] \in R^{m\times n_i}$，其中表示第 i 类样本数. 完备训练样本数据矩阵 $\boldsymbol{A}$ 是由 g 类训练样本组成：

$$\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_g] = [v_{1,1}, v_{1,2}, \cdots, v_{g,n_g}] \quad (10)$$

例如一个待测样本 y 属于训练样本的第 k 类，则由训练矩阵 $\boldsymbol{A}$ 构成的线性空间表示为

$$y = \boldsymbol{A}\boldsymbol{x}_0, y \in \boldsymbol{R}^m \quad (11)$$

式中：$\boldsymbol{x}_0 = [0,\cdots,0,a_{k,1},a_{k,2},\cdots,a_{k,n_k},0,\cdots,0]^{\mathrm{T}} \in \boldsymbol{R}^n$ 为稀疏系数向量. 在理想的情况下，除了该测试样本所属类别的系数不为零，其余的系数均为零.

(2) 利用最小一范数求稀疏解. 由压缩感知理论和稀疏表示[11]研究表明，若 x_0 是稀疏的，则利用 l^1 最小化范数求解式 (11) 可得

$$\hat{x}_1 = \arg\min\|x\|_1 \quad (12)$$

式中：$\hat{x}_1$ 为 x_0 的近似解.

(3) 基于稀疏表示的分类算法. 通常情况下，由于存在噪声和建模误差，除 k 类以外，$\hat{x}_1$ 在的其他类上的映射系数也会出现少量非零值. 这时需要建立一个非零元素成分仅与和 $\hat{x}_1$ 第 i 类相关的新的向量来准确判断 y 的类别. 所以，判断 y 的类别公式为

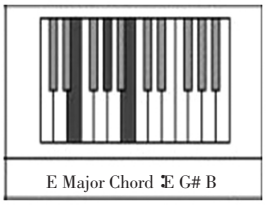
$$identity(y) = \arg\min_i \|y - A\delta_i(\hat{x})\|_2 \quad (13)$$

1.2.2　基于稀疏表示的和弦识别

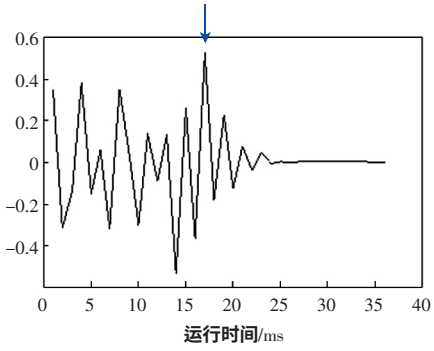
本文提出的基于稀疏表示分类器的和弦识别算法分为如下 5 个步骤：①建立含有 g 类和弦的训练特征矩阵 $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_g] \in R^{m\times n}$，其中 $\boldsymbol{A}_i$ 为第 i 类和弦的特征矩阵，m 为特征个数，n 为样本个数；②$y \in R^m$ 为

待识别和弦样本的特征矢量，求出满足 $y = Ax$，并使 $\|x\|_1$ 最小的解 $\hat{x}$，其中 $\hat{x} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_K]^T$，$\hat{x}_i$ 与 A_i 对应，$i = 1, 2, \cdots, K$；③分别保留 K 个和弦对应的系数 $\hat{x}_i$，构建 K 个矢量 $\delta_i(\hat{x}) = [0, \cdots, 0, \hat{x}_i, 0, \cdots, 0]^T$，$i = 1, 2, \cdots, K$，矢量 $\delta_i(\hat{x})$ 的维数与 $\hat{x}$ 相同；④计算冗余值，即二范数为 $r_i(y) = \| y - A\delta_i(\hat{x}) \|_2$，$i = 1, 2, \cdots, k$；⑤由最小冗余值对应的 i 确定 y 所对应的和弦.

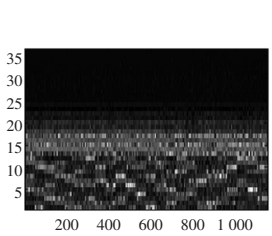
以大 E 和弦为例，其最小一范数解和冗余值的求解过程，如图 4 所示.



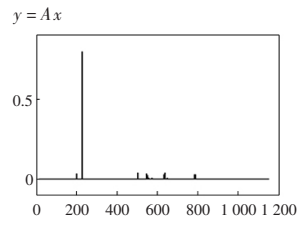
(a) 大 E 和弦的钢琴键盘显示



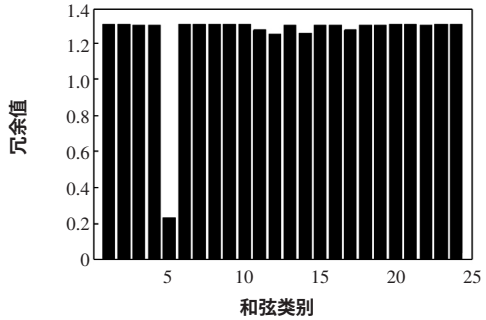
(b) 大 E 和弦 36 维特征图



(c) 特征矩阵图



(d) 最小一范数解



(e) 测试和弦关于 24 类和弦的冗余图

图 4　和弦类型识别的全过程

Fig.4　Whole process of chord recognition

2　实验与分析

2.1　实验数据

本文选用的数据库是 Beatles 乐队的 13 部专辑的 180 首歌曲，Harte 等[12]已经对这些歌曲中的和弦做了正确标注. 实验中，输入的音乐文件格式是采样率为 11 025 Hz，16 bit，单声道的 wav 格式. 然后按所标注的和弦边界和类型从这 180 首歌中截取所需的大三和弦和小三和弦共 24 类，1 152 个样本组成训练数据，288 个样本组成测试数据，数据几乎涵盖了该乐队的所有演奏风格.

2.2　结果分析

本文实验先对所截取的训练数据和测试数据分别提取 MFCC、PCP、MFCC+PCP 联合特征，然后将特征分别输入到 SRC 识别模型中，并与经典的统计模型隐马尔可夫（hidden markov model，HMM）的识别方法作对比，实验结果表明，提取的 MFCC+PCP 特征与 SRC 模型结合效果最好，识别结果对比如表 1 所示.

表 1　识别结果对比

Tab.1　Contrast result

训练数据	测试数据	特征+识别模型	识别率/%
Beatles 乐队中截取的 24 种 1152 个和弦样本	Beatles 乐队中截取的 24 种 288 个和弦样本	MFCC+SRC	27.43
		MFCC+HMM	26.89
		PCP+SRC	69.01
		PCP+HMM	67.34
		MFCC+PCP+ SRC	85.07
		MFCC+PCP+ HMM	82.61

通过对表 1 的和弦识别结果分析发现，MFCC+SRC 与 MFCC+HMM 组合模型识别率最低，由于 MFCC 没有考虑音乐乐理特征，抑制了音频的音高（Pitch）信息，导致和弦识别率低；PCP+SRC 与 PCP+HMM 识别模型虽然比 MFCC+SRC 与 MFCC+HMM 模型识别率高出 40%左右，但识别率也只有 69%和 67%左右，因为 PCP 特征无法识别空和弦与具有相同根音和弦的情况，所以识别率不高；MFCC+PCP+SRC 识别模型充分考虑了人耳听觉特性和音乐乐理特征，能够达到 85.07%的识别效果，同时从表 1 中可以得出 SRC 模型要比 HMM 识别率高 2%~3%左右，因为 SRC 可以有效地避免由于增加数据特征集而影响和弦识别率的问题.

3　结束语

本文提出一种基于 MFCC 与 PCP 联合特征和

SRC 分类器的和弦识别方法. 实验结果表明,MFCC 与 PCP 联合特征既符合人耳听觉特性,又符合和弦乐理上的特性, 与传统基于 MFCC 和 PCP 单一特征和弦识别高出近 20%和 60%方法. 同时,对于分类器的选择,SRC 比 HMM 的识别率高出 3%左右. 下一步将研究如何融入更加丰富的乐理知识来进一步提高和弦识别率.

参考文献:

[1] 董丽梦,关欣,李锵.基于稀疏表示分类器的和弦识别研究[J].计算机工程与应用,2012, 48 (29) :133-219.

[2] BROWN J. Calculation of a constant Q spectral transform [J]. J Acoust Soc Amer, 1991, 89 (1) : 425-434.

[3] FUJISHIMA T. Realtime chord recognition of musical sound: A system using common lisp music[C]//ICMC. 1999 :464-467.

[4] GOMEZ E, HERRERA P. Automatic extraction of tonal metadata from polyphonic audio recordings[R]//London : Audio Engineering Society, 2004.

[5] LEE K. Automatic chord recognition from audio using enhanced pitch class profile[C]//Proc Int Comput Music Conf (ICMC) . New Orleans :LA, 2006.

[6] SHEH A, ELLIS D. Chord segmentation and recognition using EM-trained hidden Markov models[C]//Proc Int Conf Music Inf Retrieval (ISMIR) . Baltimore : MD, 2003:185-191.

[7] WANG Feng, ZHANG Xueying, LI Bingnan. Research of Chord Recognition based on MPCP[C]//Proc The 2nd International Conference on Computer and Automation Engineering (ICCAE) . IEEE Press, 2010 :76-79.

[8] HUMPHREY Eric J, BELLO Juan P. Rethinking Automatic Chord Recognition with Convolutional Neural Networks [C]// Proc The IEEE 11th International Conference on Machine Learning and Applications (ICMLA) . Washington, DC, 2012 : 357-362.

[9] DUAN GangLong, WEI Long, LI Ni. A multiple sparse representation classification approach based on weighted residuals [C]//The IEEE Ninth International Conference on Natural Computation (ICNC) . 2013 :995-999.

[10] 徐星. 基于最小一范数的稀疏表示音乐流派与乐器分类算法研究[D]. 天津 :天津大学,2011.

[11] 王峰. 美尔音级轮廓特征在音乐和弦识别算法中的应用研究[D]. 太原 :太原理工大学,2010.

[12] DAVIS B, MERMELSTEIN P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences [C]//IEEE Transactions on Acoustics, Speech, and Signal Processing. 1980 :357-366.

[13] DONOHO D. For most large underdetermined systems of linear equations the minimal -norm solution is also the sparsest solution[J]. Comm on Pure and Applied Math, 2006, 59 (6) :797-829.

[14] HARTE C, SANDLER M, ABDALLAH S, et al. Symbolic representation of musical chords: A proposed syntax for text annotations[C]//Proc Int Conf Music Inf Retrieval (ISMIR) . 2005: 66-71.