

基于SVM和增强型PCP特征的和弦识别

闫志勇, 关欣, 李 铨

(天津大学电子信息工程学院, 天津 300072)

摘 要: 和弦识别是自动音乐标注的基础, 在歌曲翻唱识别、音乐分割及音频匹配等领域具有重要作用。针对不同乐器之间相同和弦识别率较低的问题, 提出一种基于瞬时频率提取音级轮廓(PCP)特征的改进算法。该算法结合音高频率倒谱系数, 将增强型 PCP 特征作为新的和弦识别特征, 把音频信号输入到节拍跟踪器, 依据动态规划算法提取信号的节拍信息, 计算音频信号每一个节拍内的增强型 PCP 特征, 采用结构化支持向量机分类方法实现对音乐和弦的识别。实验结果表明, 与传统 PCP 特征相比, 采用增强型 PCP 特征的和弦识别率提高了 2.5%~6.7%。

关键词: 和弦识别; 音级轮廓; 节拍跟踪; 音高频率倒谱系数; 支持向量机

Chord Recognition Based on Support Vector Machine and Enhanced PCP Feature

YAN Zhi-yong, GUAN Xin, LI Qiang

(College of Electronic Information Engineering, Tianjin University, Tianjin 300072, China)

【Abstract】 Chord recognition is the base of automatic music label, which plays an important role in the fields of song cover recognition, audio segmentation and audio matching etc. Among the different instruments, the recognition rate of the same chord is low. This paper proposes an improved chord recognition algorithm which combines the Pitch-frequency Cepstral Coefficients(PFCC) with Instantaneous-Frequency-based(IF) Pitch Class Profile(PCP) and uses the improved PCP as the new chord recognition feature. It inputs the audio signal into the beat tracker to extract the beat information of the signal which is based on dynamic programming algorithm, and calculates the improved PCP feature of the audio signal within each beat and realizes chord recognition by the structured Support Vector Machine(SVM). Results show that the ratios of chord recognition increases by 2.5%~6.7% after using the improved PCP than using the traditional PCP.

【Key words】 chord recognition; Pitch Class Profile(PCP); beat tracking; Pitch-frequency Cepstral Coefficients(PFCC); Support Vector Machine(SVM)

DOI: 10.3969/j.issn.1000-3428.2014.07.035

1 概述

音乐信号处理是计算机智能信息处理和模式识别中一个新兴的研究领域, 其中和弦识别是音乐信号处理的重要方面, 它在计算机自动伴奏、歌曲翻唱识别以及音频匹配等领域具有重要作用。人们能够从复杂的音乐演奏中提取丰富并且有意义的信息, 但是截至目前, 使用计算机处理这些信号是有难度的, 特别是在进行和弦识别或者和弦转录时更是一项具有挑战的任务^[1], 因而和弦识别是本文研究的重点。

和弦识别是音乐信息检索的中心任务之一, 其中, 文献[2]依据音乐理论知识于1999年将信号能量压缩到乐曲的12个音级上, 提出使用12维的音级轮廓特征音级轮廓(Pitch

Class Profile, PCP)作为和弦特征进行模式匹配来识别和弦。在和弦特征方面, 文献[3]提出了HPCP(Harmonic PCP)特征, 将其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; 应用谐波积光谱与PCP特征相结合的EPCP(Enhanced PCP)特征^[4]同传统PCP特征相比, 具有相同根音的和弦具有更好的识别率; 文献[5]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等。对于和弦分类模型的选择, 除模板匹配方法^[2]外, 文献[5]采用了条件随机场分割方法, 该方法对观测序列没有独立性要求, 具有表达长距离依赖性和交叠性特征的能力, 缺点是运行时间长; 文献[6]采用神经网络进行和弦识别, 可

基金项目: 国家自然科学基金资助项目(61101225, 60802049); 天津大学自主创新基金资助项目(60302015)。

作者简介: 闫志勇(1987-), 男, 硕士研究生, 主研方向: 音乐信号处理, 模式识别; 关欣, 讲师、博士; 李铨, 教授、博士。

收稿日期: 2013-06-05 **修回日期:** 2013-07-31 **E-mail:** guanxin@tju.edu.cn

以有效地避免噪声对和弦识别率的影响, 但该方法只能识别有限的歌曲; 此外, 文献[7]将统计学方法隐马尔可夫模型(Hidden Markov Model, HMM)应用于和弦的分割和识别中, 最终的和弦识别率达到了75%, 但HMM是一种产生式模型需要严格的独立性假设; 近年来, 有研究者提出将延时的隐马尔可夫模型(Duration HMM)应用于和弦识别[8]。

音乐是一个典型的高度结构化的艺术形式, 它频繁使用了某些和弦进程, 因此本文采用结构化预测模型来译码音乐的结构化信息, 完成对和弦的识别[1]。结构化支持向量机SVM是近年来广泛使用的一种机器学习方法, 同HMM相比, 它能够有效减少由于增加特征数据而过度拟合带来的误差。本文对基于瞬时频率光谱的PCP特征的提取提出了一种改进算法, 即增加对提取的瞬时频率光谱进行音高频率倒谱系数的处理[9], 之后再压缩信号从而提取基于节拍跟踪的新的PCP特征。新的PCP特征对乐器类型以及音色的变化更具鲁棒性。最后将新的PCP特征输入结构化的SVM分类器中完成和弦的识别。

2 增强型PCP特征的和弦识别系统

和弦的自动识别主要包括2个阶段: 特征提取与和弦识别。如图1所示, 使用节拍跟踪同步提取基于瞬时频率的音级特征PCP, 然后将这些特征输入到结构化支持向量机SVM中, 输出和弦标签, 最后同标准和弦标签比对完成和弦的识别。

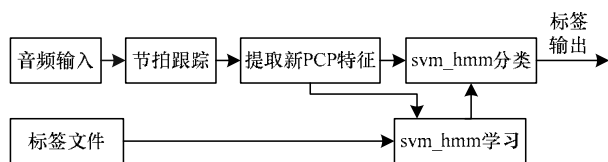


图1 增强型PCP特征的和弦识别系统流程

2.1 信号特征提取

信号特征的计算由节拍跟踪和PCP特征计算2个部分组成。首先通过节拍跟踪获取与节拍构成相符的音频节拍点信息, 然后计算出每个节拍间隔内的音级特征PCP。

2.1.1 节拍跟踪

节拍跟踪用于获取音乐信号中的节拍信息, 确定每个节拍在信号中的大致时间位置。它是自动伴奏及转录、计算机辅助音频编辑以及音乐相似性等应用的基础, 同步节拍特征也是归一化不同音乐版本之间速率变化的一种很好的途径。本文采用了文献[10]提出的动态规划(DP)节拍跟踪算法, 该算法流程如图2所示。

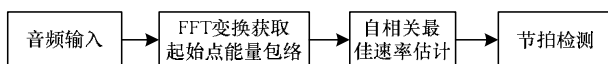


图2 节拍跟踪检测流程

首先需将音频信号进行快速傅里叶变换(Fast Fourier

Transform, FFT)变换, 通过一个权重矩阵将傅立叶频谱变换成40维Mel频谱。计算每一维频谱沿时间轴的一阶差分(只有起始点信息), 将整个频谱内的值相加, 滤除直流后得到起始能量包络。

然后对整个信号的起始能量进行自相关运算, 并在其的对数域上加高斯窗获取单一主峰值, 将自相关中最大的滞后值作为最佳的速率估计。

最后将起始能量包络和全局速率输入节拍检测模块, 通过动态规划算法获取节拍时间位置信息, 完成节拍跟踪。

2.1.2 基于节拍的增强型PCP特征的提取

音级特征PCP是音乐信息检索中广泛使用的一种信号特征。在基于瞬时频率IF的音级频谱中通过估计瞬时频率IF的光谱区域可以从噪声中区分音调元素, 同时处理调谐中的差异。本文结合音高频率倒谱系数PFCC[9], 改进了基于IF的PCP特征的提取算法, 提出了一种对乐器类型以及音色变化更具鲁棒性的音级特征PCP。

(1) 将经分帧处理后的音频信号进行FFT变换得到每一帧的信号能量 E , 对差分相位进行求导, 得到代表音调的瞬时频率IF。这时可以采用类似于Mel尺度的音调尺度 $\lg(C \times E + 1)$ 代替 E , C 表示音高正常数, 在实验中取 $C=100$ 。然后对使用对数表示的音调进行DCT变换获取PFCC系数, 只保留上层PFCC系数并对其进行DCT逆变换得到变换后的音调向量, 记做 $\lg E$ 。对能量进行对数压缩是考虑到声音强度的对数感。

(2) 进行频谱映射。按照音乐理论中的平均律以对数方式将变换后的音调向量 $\lg E$ 映射为新的音级特征PCP, 映射公式为[11]:

$$p(k) = [12 \times \lg(k \times f_s / N \times f_r)] \bmod 12 \quad (1)$$

其中, f_s 为采样率; f_s / N 为频域间隔; f_r 为参考频率, 对应于十二平均律中的C1音; $k \times f_s / N$ 则表示频域每一个分量的频率。忽略八度关系只考虑把频率能量分为12个频带, 将所有与某一特定音级相对应的频率值的音级分量根据式(2)相加得到一个12维的音级向量PCP:

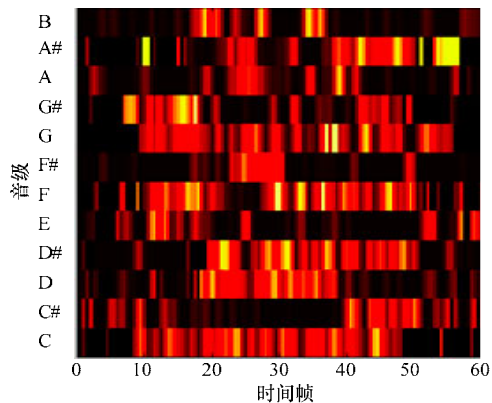
$$PCP = \sum_{k:p(k)=p} |\lg E(k)|, \quad p = 1, 2, \dots, 12 \quad (2)$$

(3) 根据节拍跟踪得到的节拍点计算每个节拍间隔内PCP的平均值作为基于节拍的音级特征[12]:

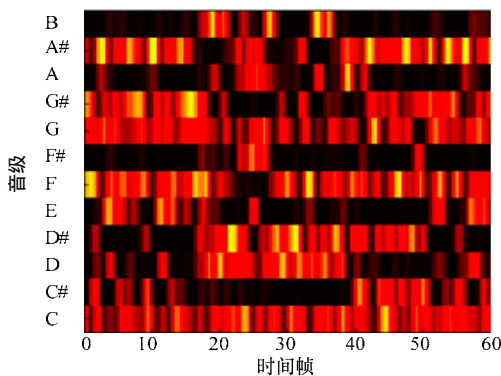
$$\overline{PCP} = \frac{1}{M} \sum_{m=1}^M PCP(p), \quad p = 1, 2, \dots, 12 \quad (3)$$

其中, M 表示每个节拍内的帧数。选用不同乐器演奏同一首乐曲时得到的音级PCP光谱特征图如图3所示, 其中, 图3(a)和图3(c)分别是使用A乐器弹奏曲目时提取的传统PCP特征和增强型PCP光谱特征图, 图3(b)和图3(d)分别是使用B乐器弹奏曲目时提取的传统PCP特征和增强型PCP光谱特征图。通过图3(a)、图3(b)与图3(c)、图3(d)的比较可以看出, 图3(c)与图3(d)光谱特征之间的差异更小, 说明了增强

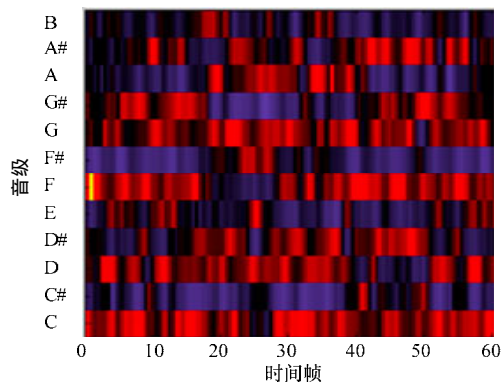
型的PCP 特征对不同乐器的鲁棒性更强。



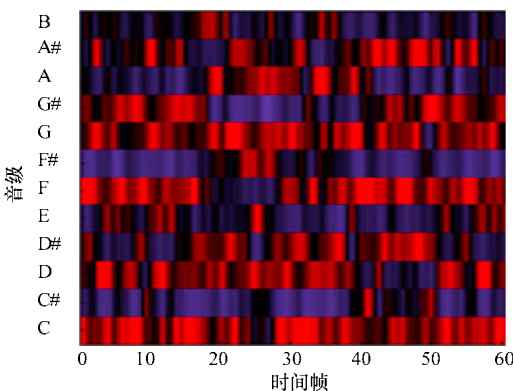
(a)A 乐器弹奏时的传统PCP 光谱图



(b)B 乐器弹奏时的传统PCP 光谱图



(c)A 乐器弹奏时的增强型PCP 光谱图



(d)B 乐器弹奏时的增强型PCP 光谱图

图3 音级PCP 光谱特征图

2.2 结构化的支持向量机

同生成模型HMM 不同,支持向量机SVM 属于判别方法模型,它通过估计模型参数获得精确的输入输出映射,并且与序列标签问题直接相关。和弦转录的目标是将观测特征映射为和弦标签状态 $x \rightarrow y$ 。使用结构化的支持向量机能够很好地完成和弦转录,实现和弦识别。

最大间隔分类器通过使用引入松弛变量的软间隔允许分类超平面(例如SVM)即使当训练数据不可分时也能找到一个最佳的分割。近年来最大间隔方法已经扩展到结构分类问题例如序列标签,尽管存在许多难以控制的约束,结构化的SVM 算法使用切平面的方法使问题得以解决^[1]。

结构化SVM 的目的是估计一个从特征输入空间 X 到离散标签输出空间 Y 的函数 $f: X \rightarrow Y$ ^[1]。通过寻找一个适当的函数 $\psi: X \times Y \rightarrow R$ 以及参数向量 w 来满足:

$$f(x) = \operatorname{argmax}_{y \in Y} \psi(x, y; w) \quad (4)$$

ψ 取线性核,因此可以写成 $\psi(x, y; w) = \langle \phi(x, y), w \rangle$, $\phi(x, y)$ 是一个依赖于应用的联合特征映射,假设训练集为 (x_i, y_i) , 其中, $i \in N = \{1, 2, \dots, n\}$ 。

当数据可分时,在标准的SVM 中根据对偶原理,最小化 $\|w\|$ 来获取最大间隔,因此将式(4)转换成:

$$\begin{aligned} \min_w \quad & \frac{1}{2} \langle w, w \rangle \\ \text{s.t.} \quad & \langle \phi(x_i, y_i), w \rangle - \langle \phi(x_i, y), w \rangle \geq 1 \end{aligned} \quad (5)$$

当数据不可分时,引入了松弛变量 ξ_i , 因此, SVM 结构化公式写为:

$$\begin{aligned} \min_{w, \xi} \quad & \frac{1}{2} \langle w, w \rangle + \frac{A}{n} \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \langle \phi(x_i, y_i), w \rangle - \langle \phi(x_i, y), w \rangle \geq 1 - \xi_i \\ & \forall i \in N, y \neq y_i, \xi_i \geq 0 \end{aligned} \quad (6)$$

其中, $A > 0$ 是一个调谐参数,用来控制间隔宽度与松弛变量之间的折中。

3 实验与结果分析

3.1 实验环境与数据

本文的实验环境为安装有Matlab2010A 的PC 机,采用 Beatles 乐队12 张专辑的180 首歌曲来评估分类模型,并且使用了来自学者Chris Harte^[13]对这些歌曲手动标记的标签文件。在本文中将和弦分为25 类,12 个根音音级(C, C#, D, ..., B)分别对应一个大调和一个小调,再加上一个无和弦的类型。

根据发行时间将专辑分成4 个集合,每个集合包含3 张专辑,以确保每个集合中有一系列的风格。分类集合为A: {Please Please Me, Help, Magical Mystery Tour}, B: {With The Beatles, Rubber Soul, The White Album}, C: {A Hard Day's Night, Revolver, Abbey Roadg} 以及 D: {Beatles For Sale, Sgt. Pepper's Lonely Hearts Club Band, Let It Be}。在

实验中首先对集合 B, C, D 中的乐曲提取增强型的音级特征 PCP, 并将这些特征作为支持向量机的训练数据进行模型训练, 随机选择专辑 {Please Please Me} 中的 4 首歌曲作为测试数据进行测试。最后同基于传统 PCP 特征并分别采用 HMM 和支持向量机方法来实现和弦识别的结果进行比较, 如表 1 所示, 其中, 训练数据为 Beatles 乐曲分类后的集合 B, C, D 。

表1 识别结果对比 %

测试数据	特征+识别模型	识别率
I Saw Her Standing There	PCP+HMM	86.01
	PCP+SVM	89.71
	增强型PCP+SVM	93.94
Anna Go To Him	PCP+HMM	74.44
	PCP+SVM	73.96
	增强型PCP+SVM	76.48
Boys	PCP+HMM	73.39
	PCP+SVM	76.39
	增强型PCP+SVM	83.12
Twist And Shout	PCP+HMM	78.20
	PCP+SVM	75.71
	增强型PCP+SVM	81.02

3.2 结果分析

通过表 1 可以看出, 采用本文改进的增强型 PCP 特征在和弦识别率上较之前采用的传统 PCP 特征有所提高, 同时实验结果还与采用 HMM 进行和弦识别的结果进行了比较, 识别率也有所提高。当使用 SVM 分类方法时, 对于歌曲 Anna Go To Him 和 Boys, 使用增强型 PCP 特征的和弦识别率要比使用传统 PCP 特征的和弦识别率分别高 2.5% 和 6.7%, 同时由表 1 还可以看出, 使用 SVM 分类方法比使用 HMM 模型时的和弦识别要分别高 2% 和 9.7%, 这是因为同 HMM 算法相比, 结构化的 SVM 能够有效减少由于增加特征数据而过度拟合带来的误差。

通过对和弦识别结果的分析可知, 识别错误的和弦主要集中在具有相同根音的和弦之间。例如容易将 C 大调和 C# 大调区分开来, 但是对于 C 大调和 A 小调的区分却十分困难, 因为 A 小调三和弦根音的组成 A-C-E 同 C 大调三和弦根音的组成 C-E-G 具有 2 个相同的音符, 所以在实际的识别中常常会将这样的和弦混淆, 造成识别的错误。

4 结束语

本文结合 PFCC 系数, 提出一种基于瞬时频率提取音级轮廓特征 (PCP) 的改进算法, 采用支持向量机分类方法完成和弦识别。实验结果表明, 使用增强型 PCP 特征的和弦识别率要比使用传统 PCP 特征的和弦识别率高 2.5%~6.7%。同时还与采用 HMM 和弦识别的结果进行了比较, 识别率也有所提高。下一步将对识别错误原因进行分析, 改进和寻找新的特征提取方法, 从而提高和弦的整体识别率。

参考文献

[1] Weller A, Daniel E, Jebara T. Structured Prediction Models for Chord Transcription of Music Audio[C]//Proc. of Interna-

tional Conference on Machine Learning and Applications. [S. 1.]: IEEE Press, 2009: 590-595.

[2] Fujishima T. Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music[C]//Proc. of International Computer Music Conference. Beijing, China: [s. n.], 1999.

[3] Gomez E, Herrera P. Automatic Extraction of Tonal Metadata from Polyphonic Audio Recordings[C]//Proc. of Audio Engineering Society. London, UK: [s. n.], 2004: 362-371.

[4] Lee K. Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile[C]//Proc. of International Computer Music Conference. New Orleans, USA: [s. n.], 2006: 225-236.

[5] Wang Feng, Zhang Xueying, Li Bingnan. Research of Chord Recognition Based on MPCP[C]//Proc. of the 2nd International Conference on Computer and Automation Engineering. [S. 1.]: IEEE Press, 2010: 76-79.

[6] Su B, Jeng S. Multitimber Chord Classification Using Wavelet Transform and Selforganized Map Neural Networks[C]//Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA: [s. n.], 2001: 3377-3380.

[7] Sheh A, Ellis D P. Chord Segmentation and Recognition Using EM-trained Hidden Markov Models[C]//Proc. of International Conference on Music Information Retrieval. Baltimore, USA: [s. n.], 2003: 185-191.

[8] Chen R F. Chord Recognition Using Duration-explicit Hidden Markov Models[C]//Proc. of the 13th International Society for Music Information Retrieval Conference. [S. 1.]: IEEE Press, 2012: 548-559.

[9] Muller M. Towards Timbre-invariant Audio Features for Harmony-based Music[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(3): 649-662.

[10] Ellis D P W, Poliner G. Identifying Cover Songs with Chroma Features and Dynamic Programming Beat Tracking[C]//Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing. [S. 1.]: IEEE Press, 2007: 1429-1432.

[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] 董丽梦, 李 锐, 关欣. 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): 133-136.

[13] Harte C. Towards Automatic Extraction of Harmony Information from Music Signals[D]. London, UK: Department of Electronic Engineering, University of London, 2010.

编辑 索书志